Novel biomarkers and prediction model for the pathological complete response to neoadjuvant treatment of triple-negative breast cancer

Yiqun Han
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College  https://orcid.org/0000-0001-5338-3058

Jiayu Wang
Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. No. 17, Panjiayuan Nanli, Chaoyang District, Beijing 100021, China

Binghe Xu (✉ xubingheBM@163.com)
Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. No. 17, Panjiayuan Nanli, Chaoyang District, Beijing 100021, China  https://orcid.org/0000-0002-0234-2747

Research

Keywords: triple-negative breast cancer, neoadjuvant chemotherapy, pathological complete response, nomogram, molecular heterogeneity

DOI: https://doi.org/10.21203/rs.3.rs-54094/v1

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

**Background:** To develop and validate a prediction model for the pathological complete response (pCR) to neoadjuvant chemotherapy (NCT) of triple-negative breast cancer (TNBC).

**Methods:** We systematically searched Gene Expression Omnibus, ArrayExpress, and PubMed for the gene expression profiles of operable TNBC accessible to NCT. The molecular heterogeneity was detected with hierarchical clustering method, and the biological profiles of differentially expressed genes were investigated by Gene Ontology, Kyoto Encyclopedia of Genes and Genomes analyses, and Gene Set Enrichment Analysis (GSEA). Next, the machine-learning algorithms including random-forest analysis and least absolute shrinkage and selection operator (LASSO) analysis were synchronously performed and, then, the intersected proportion of genes were undergone binary logistic regression to fulfill variables selection. The predictive response score (pRS) system was built as the product of the gene expression and coefficient obtained from logistic analysis. Last, the cohorts were randomly divided in a 7:3 ratio into training cohort and validation cohort for the introduction of robust model, and a nomogram was constructed with the independent predictors for pCR rate.

**Results:** A total of 217 individuals from four cohort datasets (GSE32646, GSE25065, GSE25055, GSE21974) with complete clinicopathological information were included. Based on the microarray data, a six-gene panel (ATP4B, FBXO22, FCN2, RRP8, SMERK2, TET3) was identified. A robust nomogram, adopting pRS and clinical tumor size stage, was established and the performance was successively validated by calibration curves and receiver operating characteristic curves with the area under curve 0.704 and 0.756, respectively. Results of GSEA revealed that the biological processes including apoptosis, hypoxia, mTORC1 signaling and myogenesis, and oncogenic features of EGFR and RAF were in proactivity to attribute to an inferior response.

**Conclusions:** This study provided a robust prediction model for pCR rate and revealed potential mechanisms of distinct response to NCT in TNBC, which were promising and warranted to further validate in the perspective.

**Background**

Triple-negative breast cancer (TNBC) is a special subtype of breast cancer, of which accounts for around 15% proportion, and marked by the absent expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) [1]. The biological profiles of TNBC tend to be aggressive, which is characterized by the early relapse and distant metastases in addition to the inferior prognosis with 5-year survival rate less than 30% [2]. Systemic treatment has been considered as the mainstay, which cytotoxic agents are widely applied in the overall course of therapeutics for TNBC. Considering the acknowledged aggressiveness, the neoadjuvant chemotherapy (NCT) has been recommended for TNBC within a broader range, in comparisons with the other subtypes, and pathological...
complete response (pCR) rate was used to assess the efficacy in associations with a better prognosis of TNBC undergone NCT [3].

Although systemic treatment is the essential composition of the therapeutic introduction for triple-negative breast cancer, the pCR rate is just 35–40% of the whole group of patients based on the application of standard-of-care protocols. Besides, several efforts have been put in the field of pCR prediction including both clinical and experimental management, however, except for a few predictors such as BRCA1 deficiency for TNBC treated with platinum-containing NCT, no efficient biomarkers have been yet generally recommended [4–9]. This dilemma could be the result of molecular heterogeneities of TNBC. Previous studies have managed to explore the heterogenous subtypes of TNBC and propose different classifications with distinct biological profiles, which some specific subtypes tend to chemo-sensitive to neoadjuvant therapy [10, 11]. Additionally, current findings remain still in the translational or even experimental phases, indicating the fulfillment of general application in clinical practice is challenging.

Recently, the dissection of multi-omics data through bioinformatic tools have been used to provide precise insight on the cancer biology and novel parameters for cancer therapy [12, 13]. Howbeit, it is crucial to systematically integrate the molecular data and clinicopathological characteristics for the broad application. Herein, we carried out this study to construct and validate a prediction model for the pCR rate of neoadjuvant chemotherapy for TNBC patients, with the aim of the accumulation of solid evidence for clinical practice and the potential survival benefit.

Method

Datasets selection and preprocessing

The gene expression profiles of operable TNBC accessible to NCT were systematically searched on Gene Expression Omnibus (GEO), ArrayExpress, and PubMed. Cohorts datasets were included if adopted operable TNBC patients, complete clinicopathological information, and definite clinical efficacy of neoadjuvant treatment. Raw gene expression matrix and supplementary materials were downloaded and data from TNBC patients were identified through subtype selection. The multiple gene panel corresponding to a single probe were divided into individuals, and only the maximum expression values of genes was preserved for the following analysis. Each cohort dataset was independently processed and further collected with the removal of batch effect using R package limma (version 3.42.2) [14].

Exploration of subtypes and biologic profiles

To identify the TNBC subtypes with the potentially distinct response to therapeutics, the consensus clustering method using R package ConsensusClusterPlus (version 1.50.0) was performed based on hierarchical clustering algorithms. The differential expressed genes (DEGs) were retrieved using R package limma, which the absolute of fold change (FC) more than 1 and P value adjusted by Benjami-
Hochberg method less than 0.05 were considered as the criteria for significant DEGs. Biological profiles of both up-regulated and down-regulated genes were respectively elucidated through enrichment analyses with the basis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using Cytoscape software (version 3.8.0). The STRING database (http://string.db.org/) was utilized to clarify the interactive correlations among DEGs with protein-protein interaction (PPI) networks established and further visualized by Cytoscape software.

**Identification of predictive biomarkers**

The retrieved DEGs were prepared for machine-learning algorithms to facilitate the dimensionally reduction and selection of features with the simultaneous performance of random-forest analysis and least absolute shrinkage and selection operator (LASSO) analysis using R package RandomForest (version 4.6–14) and glmnet (version 4.0–2), respectively [15]. Then, the significant genes identified from the couple methods were collected for the intersected proportions. Next, binary logistic regression analysis was carried out to discover the genes with the predictive values for pCR rate using R package survival (version 3.2-3). To quantitatively assess the promising values of biomarkers, the predictive response score (pRS) was defined as:

$$pRS = \sum_k \text{exprs}^k \text{coef}^k + \sum_i \text{exprs}^i \text{coef}^i$$

The exprsk and coefk was the gene expression and regression coefficient for DEGs of which odds ratio (OR) more than 1, while exprsi and coe was for the genes of which OR less than 1. The pRS of each patient was calculated and recorded for the following analysis.

**Construction and validation of prediction model for pCR rate**

The included datasets were randomly in a 7:3 ratio divided into training cohort and validation cohort. Subsequently, the predictive model was constructed in training cohort with the combination of pRS and the clinicopathological factors through logistic regression analysis, and the characteristics with independent predictive values for pCR rate were adopted for nomogram using R package rms (version 6.0–0). Receiver operating characteristic (ROC) curves with the calculated area under curve (AUC), using R package pROC (version 1.16.2) were utilized to validate the discriminative power of this model, while calibration plot was adopted for calibrating capability.

**Quest for promising clinical response**
Eligible patients were classified into high pRS group and low pRS group with the borderline of median value, which was followed by differential analysis to investigate the potential mechanisms of distinct response. After the DEGs were identified, the Gene Set Enrichment Analysis (GSEA) were performed with 1000 permutations referred to gene sets of h: hallmarks (h.all.v7.1.symbols) and c6: oncogenic signatures (c6.all.v7.1.symbols) downloaded from Molecular Signatures Database (MSigDB, https://gsea-msigdb.org/) using R package ClusterProfiler (version 3.14.3).

**Statistical analysis**

In this study, statistical tests were two-sided and P value less than 0.05 was considered as significance. All the statistical analyses were accomplished by SPSS (version 26.0) and R software (version 3.6.4).

**Results**

A total of eighteen cohort datasets were initially retrieved, and 217 TNBC patients from four GEO datasets (GSE32646, GSE25065, GSE25055, GSE21974) were eligible with 193 patients finally adopted. The procedure of population selection and flow diagram was presented in Figure S1, and the detailed information were listed in Table S1 and Table S2.

**Three phenotypes and differential signatures of TNBC**

On the basis of gene expression profiles, three stable phenotypes of TNBC subject to neoadjuvant chemotherapy were identified (Fig. 1a), with the assumption that the response to therapeutics and clinical evolvements were intrinsically heterogenous. Differential analysis was performed to clarify the potential mechanisms attributed from DEGs, and there were 1487 DEGs identified including 326 up-regulated genes and 381 down-regulated genes with statistical significance (Fig. 1b; Table S3). Among this proportion of DEGs, SMARCC2, VHL, KANSL3, SOX3, and NPAS3 were the utmost significant up-regulated genes, while SYN2, FGF4, ADAM21, KLF7, and PPP1R11 were the foremost down-regulated genes.

To illustrate the biologic profiles of the DEGs, the significant up-regulated genes and down-regulated genes were undergone enrichment analyses, respectively. Results from GO analysis demonstrated that the up-regulated DEGs were remarkably mapped to the GO terms including outer membrane, lymph vessel morphogenesis, regulation of mitochondrial organization and the G protein-coupled receptor signaling pathway, and the downregulated DEGs were substantially enriched in plasma membrane region, cell part morphogenesis and positive regulation of protein kinase B signaling pathway (Fig. 2a). Results of KEGG analysis suggested that the up-regulated DEGs were significantly functional at the biologic pathways of focal adhesion and glucagon signaling, and the downregulated DEGs mainly enrolled in the biologic profiles of regulation of action skeleton and GABAergic synapse (Fig. 2b). The interactive correlations among functional products of gene expression were measured in degree, and visualized in circular output as protein-protein interactive networks (Fig. 2c). It was evident that the core functional expression lay in the panels comprising CASP8, ITGA1, ITGAX, PAK1, and EXOSC10.
Predictive biomarkers and score system

To retrieve the most significant variables with predictive values, we dimensionally reduced the volume and selected the foremost characteristics based on machine-learning algorithms, which random forest analysis and Lasso regression analysis were concurrently carried out. There were 131 genes and 94 genes were respectively chosen, and 22 genes in total intersected which were considered to carry the foremost predictive power for pCR rate (Fig. 3a-c). To further achieve the shrinkage of variables, the binary logistic regression analysis was performed toward each gene, of which the results revealed a total of 6 genes were statistically significant (Fig. 4a; Table S4). Next, the pRS of each patient was calculated as the product of gene expression and the corresponding coefficient obtained from the logistic analysis. ROC curve was used to depict the predictive power of this score system and the computational AUC was 0.696 (Fig. 4b).

Prediction model for pCR rate

Populations were randomly divided in a 7:3 ratio to training cohort and validation cohort to establish and validate model. The integrity of characteristics consisted of clinicopathological factors, including age at diagnosis, clinical stage, histopathological grade, tumor size, nodal status, and pRS was adopted to construct the predictive model for pCR rate of neoadjuvant chemotherapy. Through binary logistic regression analysis, the significant variables were identified, which pRS ($P < 0.0001$) and tumor size ($P = 0.024$) were in statistical significance to this predictive value (Table S5). Then, nomogram was established and validated through ROC curve for the discrimination power and calibration curve for the calibrating capability, respectively. The AUC of ROC curves from training cohort was 0.704 and from validation cohort was 0.756, while the calibration curve presented a slope in around 45° angle. Collectively, this predictive model for the pCR rate was well-performed.

Potential mechanisms of distinct response

To further investigate the potential mechanisms of distinct response to neoadjuvant therapy in TNBC, differential analysis was performed between high pRS group and low pRS group, and followed by enrichment analysis based on the rendered gene sets. A total of 98 genes were considered as statistically significant, which GIT2, DYNC1H1, FN1, CNPY2, PTPN11 were evidently up-regulated, and AFF4, FGD2, MYO16, GLS2, CCDC132 were of the foremost down-regulated proportions (Table S6). Results of GSEA revealed that the biological processes associated with apoptosis, hypoxia, mTORC1 signaling, and myogenesis were up enriched, while the oncogenic signatures including EGFR and RAF were up regulated and CTIP and RELA were downregulated.

Discussion
Overall, this study identified a prediction score system curated from a six-gene panel, and constructed a predictive model adopting clinicopathological characteristics for pCR rate of NCT in TNBC. Given the distinct response to systemic therapy, potential mechanisms were investigated and the promising signatures were identified.

The heterogeneity of molecular features in TNBC has been long investigated, and it is well recognized that this kind of divergence could result in distinct clinical profiles and survival outcomes [10, 11, 16]. With the consideration of molecular heterogeneities, we identified three distinct subtypes of TNBC with the receipt of NCT through consensus clustering method, and further investigated the genomic difference based on gene expression profiles. Then, enrichment analyses were carried out to comprehensively explore the cellular component, molecular functions, and biological process. Both the up-regulated genes and down-regulated genes were undergone analyses, respectively, of which the results revealed that the increased proactivity of mitochondrion organization, lymph vessel morphogenesis, and focal adhesion could attribute to this kind of molecular heterogeneity. The core functional group of genes was presented in PPI network, suggestive of the significant roles in the biological course.

To facilitate the introduction of a practical prediction model, we conducted analyses on the basis of random forest analysis and Lasso analysis to retrieve the intersected proportion for dimensionally reduction and features selection, which has been widely used for characteristics of cancer diagnosis and therapy [17–19]. After this group of genes identified, the respective prediction values for pCR was successively evaluated by logistic analysis, and the a six-gene panel comprising ATP4B, FBXO22, FCN2, RRP8, SMERK2, TET3 were finally recognized. With the application of mathematic formula, pRS system was quantitatively established with a decent value for pCR rate. To optimize this prediction method, the clinicopathological characteristics were integrated and assessed the predictive values to select the valuable multi-variables, which pRS and clinical stage of tumor size was ultimately determined.

With the selected gene panel and variables, the nomogram which was considered as an intuitive method was constructed to quantitatively assess the predicted pCR rate in TNBC [20]. Results from validation analyses was suggestive of the well performance of this model and the rationale of broad application in clinical practice. Several have managed to establish prediction models for pCR of NCT in TNBC which was estimated to perform well in clinical practice [21–23]. However, most of them basically focused on the characteristics from imaging or laboratory indexes. This nomogram took full consideration of the molecular heterogeneities detected from transcriptome information and clinicopathological characteristics to provide practical prediction of pCR rate for TNBC patients planned to NCT. Besides, we also discussed the potential mechanisms of benefit and non-benefit phenotypes in order to provide promising evidence for practice. Biological processes including apoptosis, hypoxia, mTORC1 signaling and myogenesis, and oncogenic features of EGFR and RAF were in proactivity to attribute to an inferior response. In fact, the potential triggers of distinct response to NCT remain undetermined, considering, these findings to some extent enlighten the quest for improvement of efficacy, which were in accordance with the previous studies [24–27]. Current trials have exerted efforts and assess the efficacy of this kind
of targeted therapies in TNBC [28, 29], however, these results were controversial and remained to be updated through randomized controlled trials with a large sample cohort.

Indeed, there were some inevitable limitations of this study. Firstly, this was a retrospective analysis with the adoption of the identified datasets from publicly databases, and the heterogeneities among populations from different cohorts could not be removed. Secondly, a few characteristics, such as the Ki67 index and therapeutic protocols, could not be taken into considerations due to the lack of records in publicly available database, which could potentially weaken the prediction power of this model. Thirdly, the cutoff value of pRS was determined as the median which was practical yet less precise, which was necessary to be validated and optimized. Last, both experimental and clinical research are supposed to conducted and validate these findings obtained this study.

Conclusion

In conclusions, our study established a robust prediction model based on the transcriptome signatures and clinicopathological features, taking molecular heterogeneities into consideration, for the pCR rate, and discussed the potential mechanisms of distinct response to NCT in TNBC, which were promising for the exploration of novel therapeutics. This prediction model is warranted to be further applied and validated among large-scale cohorts in the upcoming future.

Abbreviations
| **triple-negative breast cancer** | **TNBC** |
|---------------------------------|----------|
| estrogen receptor                | ER       |
| progesterone receptor            | PR       |
| human epidermal growth factor receptor 2 | HER2     |
| neoadjuvant chemotherapy         | NCT      |
| pathological complete response   | pCR      |
| Gene Expression Omnibus          | GEO      |
| differential expressed genes     | DEGs     |
| fold change                      | FC       |
| Gene Ontology                    | GO       |
| Kyoto Encyclopedia of Genes and Genomes | KEGG   |
| protein-protein interaction      | PPI      |
| least absolute shrinkage and selection operator | LASSO |
| the predictive response score    | pRS      |
| odds ratio                       | OR       |
| receiver operating characteristic | ROC     |
| area under curve                 | AUC      |
| Gene Set Enrichment Analysis     | GSEA     |

**Declarations**

**Availability of data and materials**

The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s)).

**Authors’ contributions**

Conception and design: Binghe Xu, Jiayu Wang, Yiqun Han

Development of methodology: Yiqun Han

Acquisition of data: Yiqun Han
Analysis and interpretation of data: Yiqun Han, Binghe Xu, Jiayu Wang

Writing of the manuscript: Yiqun Han

Review and revision of the manuscript: Binghe Xu, Jiayu Wang, Yiqun Han

Study supervision: Binghe Xu, Jiayu Wang

**Funding**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests

**Ethics approval and consent to participant**

Not applicable.

**Consent for publication**

Not applicable.

**Acknowledgements**

Not applicable.

**Supplementary Information**

**Supplementary tables**

- **Table S1.** Detailed information of eligible cohort datasets
- **Table S2.** Clinicopathological features of included patients
- **Table S3.** Differentially expressed genes of molecular phenotypes of TNBC
- **Table S4.** Significant genes for the prediction of pCR rate
- **Table S5.** Results of binary logistic regression analysis
Table S6. Differentially expressed genes of high pRS low pRS group

Supplementary figures

Figure S1. Flow diagram of selection and identification of cohort datasets.

References

1. Cao L, Niu Y. Triple negative breast cancer: special histological types and emerging therapeutic methods. Cancer biology medicine. 2020;17:293–306.

2. Bonotto M, Gerratana L, Poletto E, Driol P, Giangreco M, Russo S, Minisini AM, Andreetta C, Mansutti M, Pisa FE, et al. Measures of outcome in metastatic breast cancer: insights from a real-world scenario. Oncologist. 2014;19:608–15.

3. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2008;26:1275–81.

4. Gamucci T, Pizzuti L, Sperduti I, Mentuccia L, Vaccaro A, Moscetti L, Marchetti P, Carbognin L, Michelotti A, Iezzi L, et al. Neoadjuvant chemotherapy in triple-negative breast cancer: A multicentric retrospective observational study in real-life setting. Journal of cellular physiology. 2018;233:2313–23.

5. Cocco S, Piezzo M, Calabrese A, Cianniello D, Caputo R, Lauro VD, Fusco G, Gioia GD, Licenziato M, Laurentiis M. Biomarkers in Triple-Negative Breast Cancer: State-of-the-Art and Future Perspectives. International journal of molecular sciences 2020, 21.

6. Jung YY, Hyun CL, Jin MS, Park IA, Chung YR, Shim B, Lee KH, Ryu HS. Histomorphological Factors Predicting the Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. Journal of breast cancer. 2016;19:261–7.

7. Li Z, Zhang Y, Zhang Z, Zhao Z, Lv Q. A four-gene signature predicts the efficacy of paclitaxel-based neoadjuvant therapy in human epidermal growth factor receptor 2-negative breast cancer. Journal of cellular biochemistry. 2019;120:6046–56.

8. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, Sun K, Tang Z, Jiang H, Li H, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. Clin Cancer Res. 2019;25:3538–47.

9. Silver DP, Richardson AL, Eklund AC, Wang ZC, Szallaszi Z, Li Q, Juul N, Leong CO, Calogrias D, Buraimoh A, et al. Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2010;28:1145–53.

10. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Investig. 2011;121:2750–67.
11. Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, Valero V, Lehmann BD, Pietenpol JA, Hortobagyi GN, et al. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clinical cancer research: an official journal of the American Association for Cancer Research. 2013;19:5533–40.

12. Milan T, Wilhelm BT. Mining Cancer Transcriptomes: Bioinformatic Tools and the Remaining Challenges. Mol Diagn Ther. 2017;21:249–58.

13. Boegel S, Castle JC, Kodysh J, O'Donnell T, Rubinsteyn A. Bioinformatic methods for cancer neoantigen prediction. Prog Mol Biol Transl Sci. 2019;164:25–60.

14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W. GK Smyth: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43:e47.

15. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software. 2010;33:1–22.

16. Lee JS, Yost SE, Yuan Y. Neoadjuvant Treatment for Triple Negative Breast Cancer: Recent Progresses and Challenges. Cancers 2020, 12.

17. Shen WC, Chen SW, Wu KC, Lee PY, Feng CL, Hsieh TC, Yen KY, Kao CH. Predicting pathological complete response in rectal cancer after chemoradiotherapy with a random forest using F-fluorodeoxyglucose positron emission tomography and computed tomography radiomics. Annals of translational medicine. 2020;8:207.

18. Xu L, Wu Y, Che X, Zhao J, Wang F, Wang P, Qu X, Liu Y, Li Z. Cox-LASSO Analysis Reveals a Ten-IncRNA Signature to Predict Outcomes in Patients with High-Grade Serous Ovarian Cancer. DNA cell biology. 2019;38:1519–28.

19. Xie G, Dong C, Kong Y, Zhong JF, Li M, Wang K. Group Lasso Regularized Deep Learning for Cancer Prognosis from Multi-Omics and Clinical Features. Genes 2019, 10.

20. Gross JP, Whelan TJ, Parulekar WR, Chen BE, Rademaker AW, Helenowski IB, Donnelly ED, Strauss JB. Development and Validation of a Nomogram to Predict Lymphedema After Axillary Surgery and Radiation Therapy in Women With Breast Cancer From the NCIC CTG MA.20 Randomized Trial. Int J Radiat Oncol Biol Phys. 2019;105:165–73.

21. Zhang F, Huang M, Zhou H, Chen K, Jin J, Wu Y, Ying L, Ding X, Su D, Zou D. A Nomogram to Predict the Pathologic Complete Response of Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer Based on Simple Laboratory Indicators. Ann Surg Oncol. 2019;26:3912–9.

22. Xie X, Xiong Z, Li X, Huang X, Ye F, Tang H, Xie X. Nomogram to Predict Internal Mammary Lymph Nodes Metastasis in Patients With Breast Cancer. Frontiers in oncology. 2019;9:1193.

23. Diao JD, Ma LX, Sun MY, Wu CJ, Wang LJ, Liu YL, Yang YJ. Construction and validation of a nomogram to predict overall survival in patients with inflammatory breast cancer. Cancer medicine. 2019;8:5600–8.

24. Nakai K, Hung MC, Yamaguchi H. A perspective on anti-EGFR therapies targeting triple-negative breast cancer. American journal of cancer research. 2016;6:1609–23.
25. Montero JC, Esparís-Ogando A, Re-Louhau MF, Seoane S, Abad M, Calero R, Ocaña A, Pandiella A. Active kinase profiling, genetic and pharmacological data define mTOR as an important common target in triple-negative breast cancer. Oncogene. 2014;33:148–56.

26. Zheng S, Zou Y, Liang JY, Xiao W, Yang A, Meng T, Lu S, Luo Z, Xie X. Identification and validation of a combined hypoxia and immune index for triple-negative breast cancer. *Molecular oncology* 2020.

27. Blomme A, Costanza B, de Tullio P, Thiry M, Van Simaeys G, Boutry S, Doumont G, Valentin ED, Hirano T, Yokobori T, et al. Myoferlin regulates cellular lipid metabolism and promotes metastases in triple-negative breast cancer. Oncogene. 2017;36:2116–30.

28. Jovanović B, Mayer IA, Mayer EL, Abramson VG, Bardia A, Sanders ME, Kuba MG, Estrada MV, Beeler JS, Shaver TM, et al. A Randomized Phase II Neoadjuvant Study of Cisplatin, Paclitaxel With or Without Everolimus in Patients with Stage II/III Triple-Negative Breast Cancer (TNBC): Responses and Long-term Outcome Correlated with Increased Frequency of DNA Damage Response Gene Mutations, TNBC Subtype, AR Status, and Ki67. Clinical cancer research: an official journal of the American Association for Cancer Research. 2017;23:4035–45.

29. Basho RK, Gilcrease M, Murthy RK, Helgason T, Karp DD, Meric-Bernstam F, Hess KR, Herbrich SM, Valero V, Albarracin C, et al. Targeting the PI3K/AKT/mTOR Pathway for the Treatment of Mesenchymal Triple-Negative Breast Cancer: Evidence From a Phase 1 Trial of mTOR Inhibition in Combination With Liposomal Doxorubicin and Bevacizumab. *JAMA oncology*. 2017;3:509–15.

30. Appendices.

**Figures**
Identification and differential analysis of triple-negative breast cancer subtypes. (a) Three stable phenotypes of TNBC subject to neoadjuvant chemotherapy were identified through consensus clustering algorithms based on hierarchical clustering method. (b) A total of 1487 differentially expressed genes were recognized by differential analysis, which included 326 up-regulated genes and 381 down-regulated genes with statistical significance.
Figure 2

Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of differentially expressed genes. (a) The respective results of GO analyses for up-regulated genes and down-regulated genes. (b) The respective results of KEGG analyses for up-regulated genes and down-regulated genes. (c) Protein-protein network of identified differentially expressed genes.
Figure 3

The process of variables selection through machine-learning algorithms. (a) The top 30 significant genes recognized from random forest analysis. (b) The performance and validation of least absolute shrinkage and selection operator (LASSO) analysis. (c) The intersected genes of these two analyses were selected.
Figure 4

The six-gene panel and validation of predictive response score (pRS) system. (a) The six-gene panel including ATP4B, FBXO22, FCN2, RRP8, SMERK2, TET3 were identified through binary logistic analysis. (b) Receiver operating characteristic curve was built with the area under curve of 0.696.

Figure 5
Development and validation of nomogram for the prediction of pCR rate. (a) The nomogram was established based on the independent predictors for the pCR rate in TNBC. (b) Calibration curve was presented with a slope in around 45° angle. (c) Receiver operating characteristic curves were built with the area under curve of 0.704 in training cohort and of 0.756 in validation cohort.

Figure 6

Exploration of the mechanisms for distinct pCR rate. (a) Differential analysis was performed between high pRS group and low pRS group. (b) Gene set Enrichment Analysis were conducted with the reference to gene lists of cancer hallmarks signatures and oncogenic signatures.

Supplementary Files

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

- Supplementaryfigures.pdf
- Supplementarytables.xlsx