

Research Article

Tumor Microenvironment Subtypes and Immune-Related Signatures for the Prognosis of Breast Cancer

Yiqun Han , Jiayu Wang , and Binghe Xu 

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

Correspondence should be addressed to Jiayu Wang; wangjiayu8778@sina.com and Binghe Xu; xubinghebm@163.com

Received 16 December 2020; Revised 14 April 2021; Accepted 20 April 2021; Published 2 June 2021

Academic Editor: Nagarajan Raju

Copyright © 2021 Yiqun Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To better understand the immune-related heterogeneity of tumor microenvironment (TME) and establish a prognostic model for breast cancer in clinical practice. **Methods.** For the 2620 breast cancer cases obtained from The Cancer Genome Atlas and the Molecular Taxonomy of Breast Cancer International Consortium, the CIBERSORT algorithm was performed to identify the immunological pattern, which underwent consensus clustering to curate TME subtypes, and biological profiles were explored by enrichment analysis. Random forest analysis, least absolute shrinkage, and selection operator analysis, in addition to uni- and multivariate COX regression analyses, were successively employed to precisely select the significant genes with prediction values for the introduction of the prognostic model. **Results.** Three TME subtypes with distinct molecular and clinical features were identified by an unsupervised clustering approach, of which the molecular heterogeneity could be the result of cell cycle dysfunction and the variation of cytotoxic T lymphocyte activity. A total of 15 significant genes were proposed to construct the prognostic immune-related score system, and a predictive model was established in combination with clinicopathological characteristics for the survival of breast cancer patients. For immunological signatures, proactivity of CD8 T lymphocytes and hyperangiogenesis could be attributed to heterogeneous survival profiles. **Conclusions.** We developed and validated a prognostic model based on immune-related signatures for breast cancer. This promising model is justified for validation and optimized in future clinical practice.

1. Introduction

Currently, the landscape of the tumor microenvironment (TME) has been generally portrayed, of which the components are considered an essential composition of cancer immunity, with counterpart activities across the “immunoediting” process [1]. In this tumor-related contexture, the density, activity, and organization of immunological infiltration is crucial and regarded as a promising indicator for both clinical response and prognosis of cancer patients [2]. In the advancing era of immunotherapy, several malignancies have been rendered with unprecedented benefits and durable response for the specific tailing effect of novel mechanisms [3]. However, the clinical response provided by immunotherapy was inconsistent among populations or even in the changing stage of an individ-

ual, which could be the result of heterogeneities existing in this complicated interactive contexture.

Breast cancer is a worldwide leading newly diagnosed cancer in the female and a heterogeneous disease with the possibility of distinct clinical outcomes [4]. Recent studies using bioinformatics tools have provided insight into the deep mining on the dissection of the TME [5]. However, to further achieve the prediction for the prognosis of patients from clinical practice, taking clinical characteristics into consideration is essential to this implementation. Accordingly, it is critical to integrate comprehensive factors including both data on multiomics and clinical parameters to create a precise system for breast cancer. In this study, we used the transcriptome mixture and clinicopathological information of 2620 individuals which were publicly retrieved on databases,

analyzed the potential immune-related mechanisms for divergent TME phenotypes, and constructed a prognostic model with good performance for patients with breast cancer.

2. Materials and Methods

2.1. Breast Cancer Datasets and Preprocessing. The breast cancer gene expression dataset, as a training cohort, was searched on The Cancer Genome Atlas (TCGA), and the RNA-seq by Expectation-Maximization (RSEM) counts with full clinical information of breast cancer patients ($N = 1095$) was obtained from the University of California, Santa Cruz (UCSC) Xena browser (<http://xena.ucsc.edu/>). For a supplement, gene expression profile and curated clinical data ($N = 1525$), as the validation dataset, were retrieved from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database which was downloaded from cBioPortal for the Cancer Genomics database (<https://cbioportal.org>). The process of cohort selection was presented in Supplementary Figure 1, in which the detailed information of included datasets was provided in Supplementary Table 1.

2.2. Infiltrating Abundances and TME Subtypes. On the basis of gene expression mixture, we applied the Cell Type Identification by Estimation Relative Subsets of RNA Transcripts (CIBERSORT) approach to infer the relative proportions of infiltrating components, and the algorithms were performed using LM22 signatures with 1000 permutations (<http://cibersort.stanford.edu/>) [6]. The number of TME subtypes was successively determined by the consensus clustering algorithm on the basis of hierarchical agglomerative clustering methods, using ConsensusClusterPlus package of R software [7], based on the quantitatively immunological infiltrating patterns of cases from TCGA-BRCA and METABRIC, respectively, and validated by survival analysis by the Kaplan-Meier (KM) method using the log-rank test.

Identification for differentially expressed genes (DEGs) among TME clusters was accomplished with limma R package [8], in which the absolute of fold change (FC) more than 1 and P value adjusted by the Benjamini-Hochberg method less than 0.05 were considered the criteria for significant DEGs with annotations searched from the GeneCardsSuite database (<http://genecards.org/>). Enrichment analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), were conducted to describe the molecular function and biological profiles of DEGs, while Gene Set Enrichment Analysis (GSEA) was performed to successively explore the potential mechanisms by virtue of the gene sets of h: hallmarks (h.all.v7.1.symbols) and c7: immunologic signatures (c7.all.v7.1.symbols) rendered by the Molecular Signatures Database (<http://gsea-msigdb.org/>) using the ClusterProfiler R package [9, 10]. Through the STRING database (<http://string.db.org/>), the functional analysis of protein-protein interaction (PPI) networks was established and further visualized by Cytoscape software (version 3.8.0) [11].

2.3. Prognostic Immune-Related Score (piRS) and Prognostic Model. Survival analyses concerning the expression of each

DEG for overall survival (OS) were performed using the KM method by survival and survminer R packages [12], and variables with statistical significance were determined with random forest analysis and least absolute shrinkage and selection operator (LASSO) analysis, using the random-Forest package and glmnet package of R software [13], for dimensional reduction and identification of the overlapping DEGs significant for prognosis. Next, a univariate COX proportional hazard model was utilized to differentiate the foremost groups of DEGs, and the prognostic immune-related score (piRS) was defined as

$$\text{piRS} = \sum_{k=1}^n \text{exprs}_k \cdot \text{coef}_k + \sum_{i=1}^n \text{exprs}_i \cdot \text{coef}_i, \quad (1)$$

where n is the number of genes. exprs_k and coef_k were the gene expression and regression coefficient for DEGs of which the hazard ratio (HR) is more than 1, while exprs_i and coef_i were for the genes of which HR is less than 1. The piRS of each patient was calculated and recorded for the following analysis.

In combination with clinicopathological characteristics, piRS was included in the multivariate COX regression analysis to construct a nomogram with the rms package of R software, of which predictive power for prognosis in terms of sensitivity and specificity was assessed by the time-dependent receiver characteristic curve (ROC) using timeROC R package [14]. The discriminative power of the prognostic model was quantitatively measured by Harrell's concordance index (C-index). The flow diagram of data processing and analysis is presented in Supplementary Figure 2.

2.4. Statistical Analysis. In this study, correlation analysis was carried out and demonstrated with Spearman's coefficients and the corresponding P values. Paired comparative analysis was performed for continuous variables with the independent t -test for normal distribution and the Mann-Whitney U test for abnormal distribution, respectively, which was visualized using the ggplot2 R package. A P value less than 0.05 was regarded as statistically significant. The missing data were excluded against analyses to weaken heterogeneity. All the statistical analyses were 2-sided and conducted by R software (version 3.6.4).

3. Results

3.1. Identification of TME Subtypes. The infiltration patterns were analyzed with CIBERSORT deconvolution algorithms by quantifying the fractions of 22 immune cell types in TME of 1095 TCGA-BRCA patients. The landscape of immunological infiltration is exhibited in Figure 1, Supplementary Figure 3, and Supplementary Table 2. Data on the estimated proportions of infiltrating cells indicated an evident heterogeneity, considering that an unsupervised clustering analysis was performed to determine the potential TME subtypes. On the basis of the consensus clustering method, three robust TME clusters were retrieved, and the prognosis of patients from these curated subtypes was different with statistical significance ($P = 0.023$) (Figures 2(a) and 2(b)).

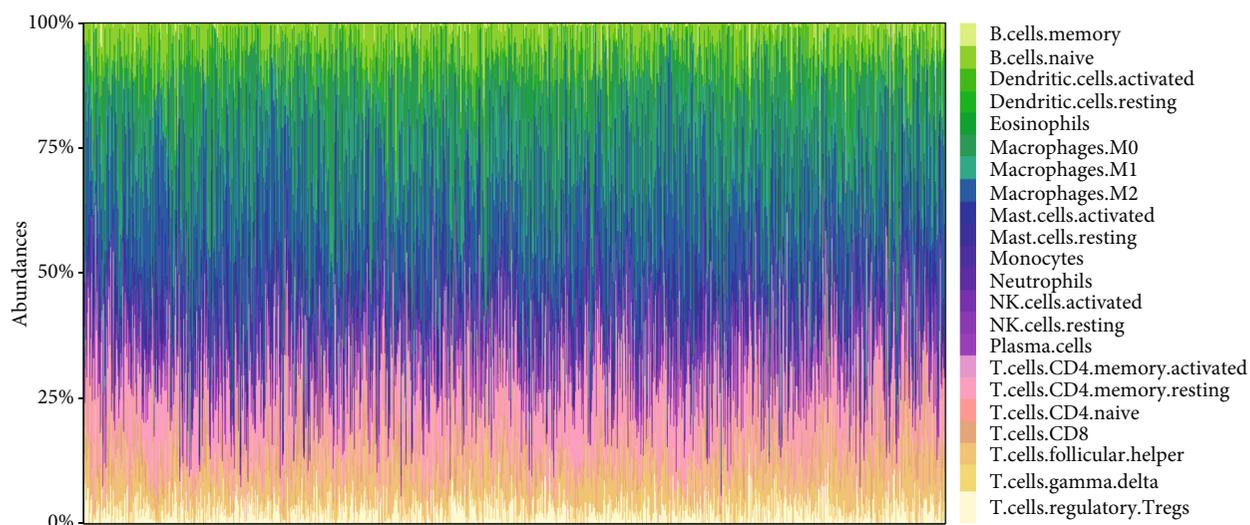


FIGURE 1: The landscape of immunologic infiltrations in the TME of breast cancer. The relative proportions of 22 lymphocytes infiltrating in the TME of TCGA-BRCA patients were portrayed, which demonstrated an evident heterogeneity among individuals.

With the aim of validation, the bulk tissue gene expression profiles of 1525 patients from the METABRIC database were also estimated for immune infiltrations and underwent a clustering analysis (Supplementary Table 3), from which three clusters were obtained with a statistical difference in OS ($P < 0.001$) (Figures 2(c) and 2(d)), indicating that this partition was stable for TME subtypes in terms of immunological infiltrations.

3.2. Signatures of TME Phenotypes. To explore the contributing mechanisms for versatile TME phenotypes, we systematically conducted the differential analysis to identify the DEGs and enrichment analyses to elucidate the biological profiles. A total of 20530 DEGs were identified through differential analysis, of which 664 genes significantly expressed heterogeneity among three immune-related TME subtypes, with 233 genes upregulated and 431 genes downregulated (Figures 3(a) and 3(b), Supplementary Table 4). Among the identified DEGs, MMP9, SLC16A3, CDT1, CA9, and CDC20 were revealed to be the leading upregulated, while P2RY12, GPR34, ABCB1, IGF1, and CX3CR1 were estimated to express downregulation with the foremost significance.

Enrichment analyses, including GO, KEGG, and GSEA, were carried out to portray the landscape of biological profiles of DEGs and illustrate the mechanism contributing to the differential phenotypes. GO analysis demonstrated that the significant DEGs intensively mapped to the GO terms concerning cellular localizations include the collagen-containing extracellular matrix, condensed chromosome, centromeric region, and spindle (Figure 3(c)). Downstream analysis from KEGG suggested that DEGs were significantly involved in the enriched pathways comprising neuroactive ligand-receptor interaction, PI3K-Akt signaling pathways, and cell cycle (Figure 3(d)). GSEA results showed that the DEGs, in terms of cancer hallmarks, were an ensemble of cell cycle dysfunction, including E2F targets and G2M checkpoints, which were significantly muted in metabolisms such

as adipogenesis and myogenesis. Regarding immune-related signatures, the improved effectiveness of CD8 T lymphocyte was expected to facilitate the genesis of TME phenotypes (Figure 3(e)). Built on cluster analysis, PPI networks were built to clarify the interactive functions of significant DEGs, and the corresponding relationships were shown with a ranked degree (Figure 3(f)).

3.3. Prognostic Value of Signature Genes. Survival analysis was carried out using the KM method to identify the significant genes with predictive values for survival. The whole 20530 DEGs were successively assessed based on the corresponding expression from each patient, and a total of 2274 genes exhibited a statistical significance for OS. Then, random forest analysis and LASSO analysis were synchronously performed to further recognize the significant genes for prognosis, which discriminated 411 and 269 genes, respectively, and the intersected 44 genes were prepared for the following analysis (Figures 4(a)–4(c)). Next, we carried out univariate COX regression analysis for identification of the most significant variables for survival, and the panel with a total of 15 genes was finally determined (Figures 5(a) and 5(b), Supplementary Table 5).

Subsequently, the pIRS was constructed based on the expression of 15 genes in combination with the corresponding coefficient for each patient, and all cases were stratified into the high pIRS group and the low pIRS group. Survival analysis suggested that patients from these two groups exhibited a divergent clinical prognosis with statistical significance ($P < 0.0001$), which was in accordance with the result of the validation cohort ($P < 0.0001$) (Figures 5(c) and 5(d)).

Comparative analysis for immunological phenotypes was performed, of which the results indicated that the abundance of B lymphocytes and CD8 T cells was notably higher in the low pIRS group (Figure 6(a)). Distributions of leukocyte infiltration among groups with diverse clinicopathological factors are shown in Supplementary Figure 4.

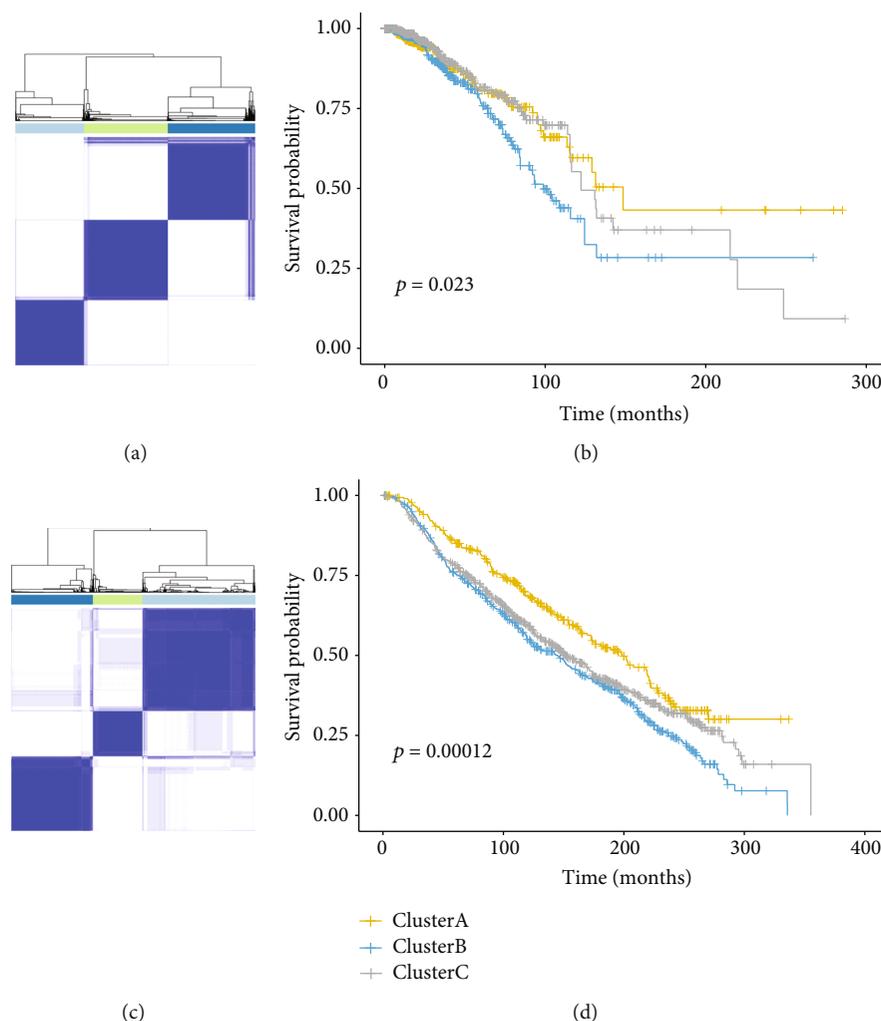


FIGURE 2: The curated TME subtypes with distinct prognostic outcomes. Three TME phenotypes of TCGA-BRCA patients ($N = 1095$) were retrieved by a consensus clustering method (a) with a significant difference in overall survival ($P = 0.023$) (b). Consistent results were obtained in patients from the METABRIC database ($N = 1525$) (c) with distinction in prognosis ($P = 0.00012$) (d). The survival analyses were performed with the Kaplan-Meier method using the log-rank test.

The expression levels of immunologic modulators were also evaluated between two groups with a total of 73 signature genes included [15] (Supplementary Figure 5). Interactive correlations are shown in Figure 6(b), and the genes, in particular, associated with the immune checkpoint underwent comparative analyses (Figure 6(c)). Results from paired analyses showed that the expression of PDCD1 (PD-1) and ICAM1 was significantly higher in the low pIRS group, while VEGFA presented an increasing trend of expression in the high pIRS group.

3.4. Prediction Model for Prognosis of Breast Cancer. A total of 677 breast cancer patients with complete clinicopathological characteristics including age, pathological TNM stage (pTNM), molecular features, PAM50 subtypes, and pIRS group were adopted into the COX proportional model for quantitative estimation of survival. Results from multivariate regression analysis demonstrated that the age, pTNM, and pIRS group were independent factors for the prognosis of breast cancer, which were utilized to construct a nomogram

for the prediction of 3-year, 5-year, and 10-year survival probability (Figure 7(a), Supplementary Table 6). Time-dependent ROC suggested that the time-dependent under curve area was ranging from 0.77 to 0.78, indicating that the curated prognostic model was well performed (Figure 7(b)), and the quantified C-index was 0.823 obtained from the training cohort and 0.776 from the validation cohort, respectively, which were generally higher than those computed from the TNM staging system, revealing the robustness in addition to better accuracy of this prediction model (Supplementary Table 7).

4. Discussion

In this study, we performed an overall analysis of TME immunological profiles based on transcriptome data of breast cancer, discussing the heterogeneity of the stromal contexture in terms of the immune-related subtypes and potential contributing mechanisms, in addition to giving a quantitative

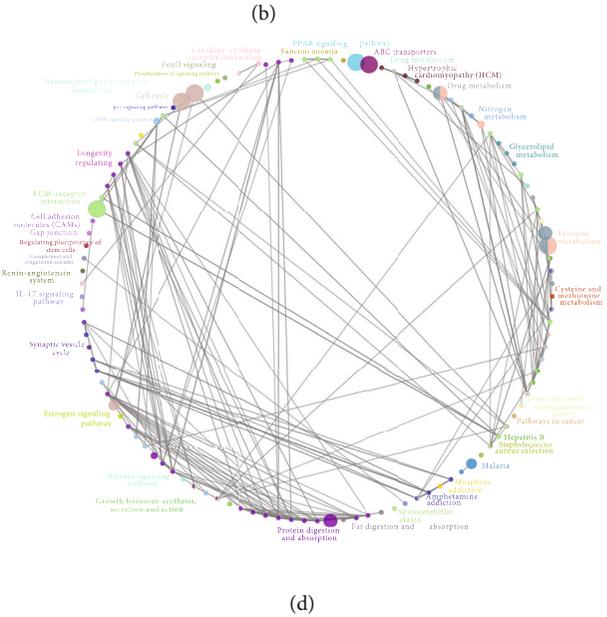
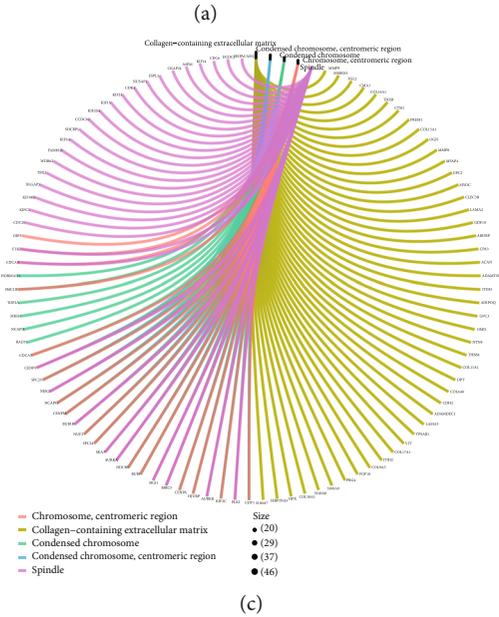
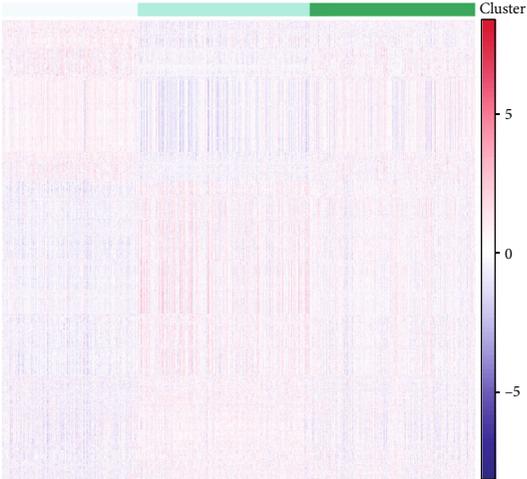
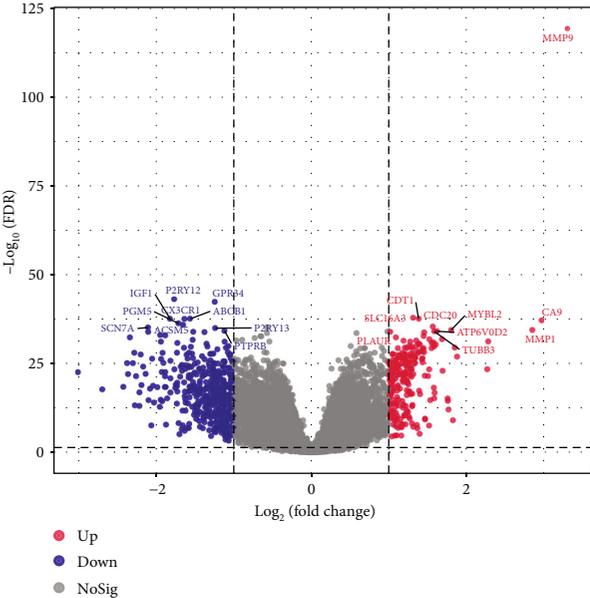


FIGURE 3: Continued.

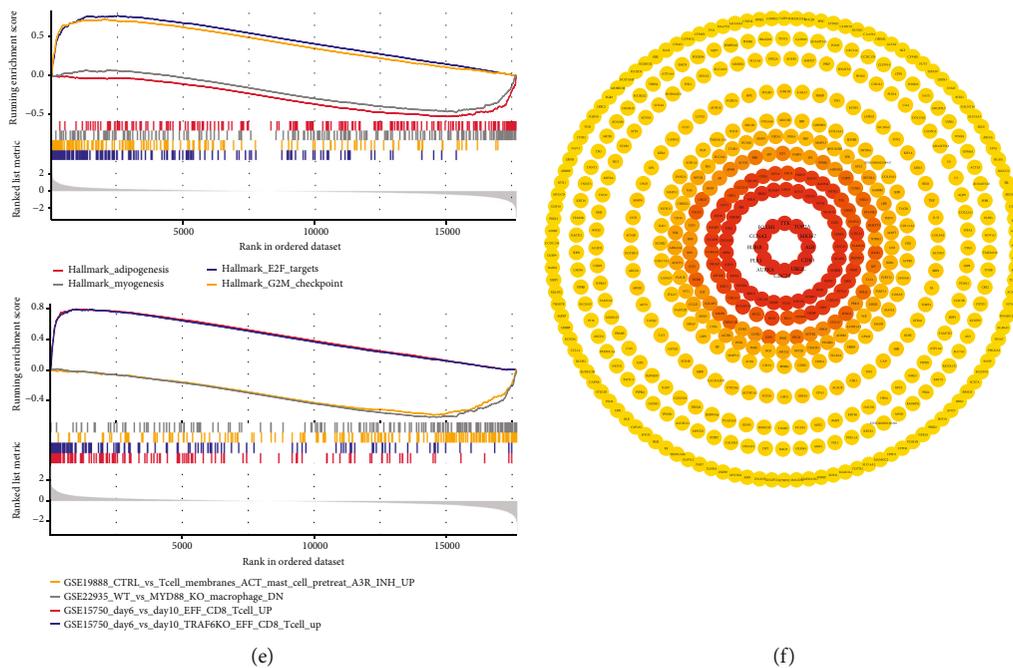


FIGURE 3: Contributing mechanisms of differential TME phenotypes. A total of 664 significant differentially expressed genes (DEGs) identified through differential analysis with 233 genes upregulated (red) and 431 genes downregulated (blue), respectively (a). DEGs were distributed evenly among three TME subtypes (b). GO analysis (c), KEGG pathways analysis (d), and GSEA (e) were successively carried out, and PPI networks were constructed to explore the interactions (f).

estimation of the associations between immune-related parameters and the prognosis of breast cancer.

To quantitatively ascertain the stromal infiltration of TME, the CIBERSORT approach was adopted for computational proportions of immunological cell types of interest [6]. The landscape of immunologic infiltration exhibited a varying tendency among patients, prompting us to perform an analysis for the potential subtypes of TME. Three TME clusters were determined based on consensus clustering algorithms and proven to be stable in the validation cohort. Moreover, results from survival analysis, as the exploratory mining, demonstrated that an apparent difference of survival was detected among the patients from corresponding TME subtypes, indicating that the immunologic heterogeneity of TME could be a promising predictor for the survival of breast cancer. Previous studies have managed to correlate molecular features to the prognosis of breast cancer patients. However, multiomic angles have to be focused which primarily included genomics, epigenetics, and transcriptomic profiles. On the basis of the dataset, Shen and colleagues established an lncRNA panel associated with immunological signatures to stably predict the prognosis of breast cancer [16], which is consistent with several studies [17, 18], and supported that the phenotypes could be the results of molecular heterogeneity. From the immunologic perspective, we indicated that inherent heterogeneity could lead to divergent prognosis and curated three robust TME subtypes for breast cancer, of which the potential mechanisms leading to this kind of differentiation remained to be explored.

Focusing on the illustration of this mechanism, then, we performed systematic analyses for the differential genetic

expressions, of which the versatile phenotypes tended to be the product. Differential analysis among three TME clusters recognized the genes with divergent transcriptome profiles, while the enrichment analyses demonstrated that mapping molecular positions lied in the extracellular matrix and nuclear components, and the ensemble pathways were significantly enriched in the cellular signaling transduction involved in cancer proliferation, migration, and invasion. Results from enrichment analysis with specific genetic references suggested that the differential phenotypes could be the concerted result of cell cycle dysfunction, the variety in CD8 T lymphocyte activity, and the repression of metabolite homeostasis, which could be a potential target for novel therapeutics toward the breast cancer population. PPI analysis identified the core protein consisting of TOP2A, MKI67, and CDK1, which was supposed to function at the process for immunological heterogeneity. Based on the tissue-based research, Zheng et al. suggested that TOP2A expression could independently predict the survival of triple-negative breast cancer [19], which was further confirmed by Xu and colleagues [20]. MKI67 has been regarded as a contributing component for molecular heterogeneity of breast cancer validated by previous studies [21, 22]. As an essential regulator for the cell cycle, the activity of CDK1 has long been considered a potential target, which is associated with cell cycle dysfunction and corresponding distinct molecular profiles of breast cancer [23]. These findings were in accordance with the current researches of TTK [24], CDC20 [25], PLK1 [26], and AURKA [27], which were considered the leading contributors for molecular heterogeneity genesis in the current study.

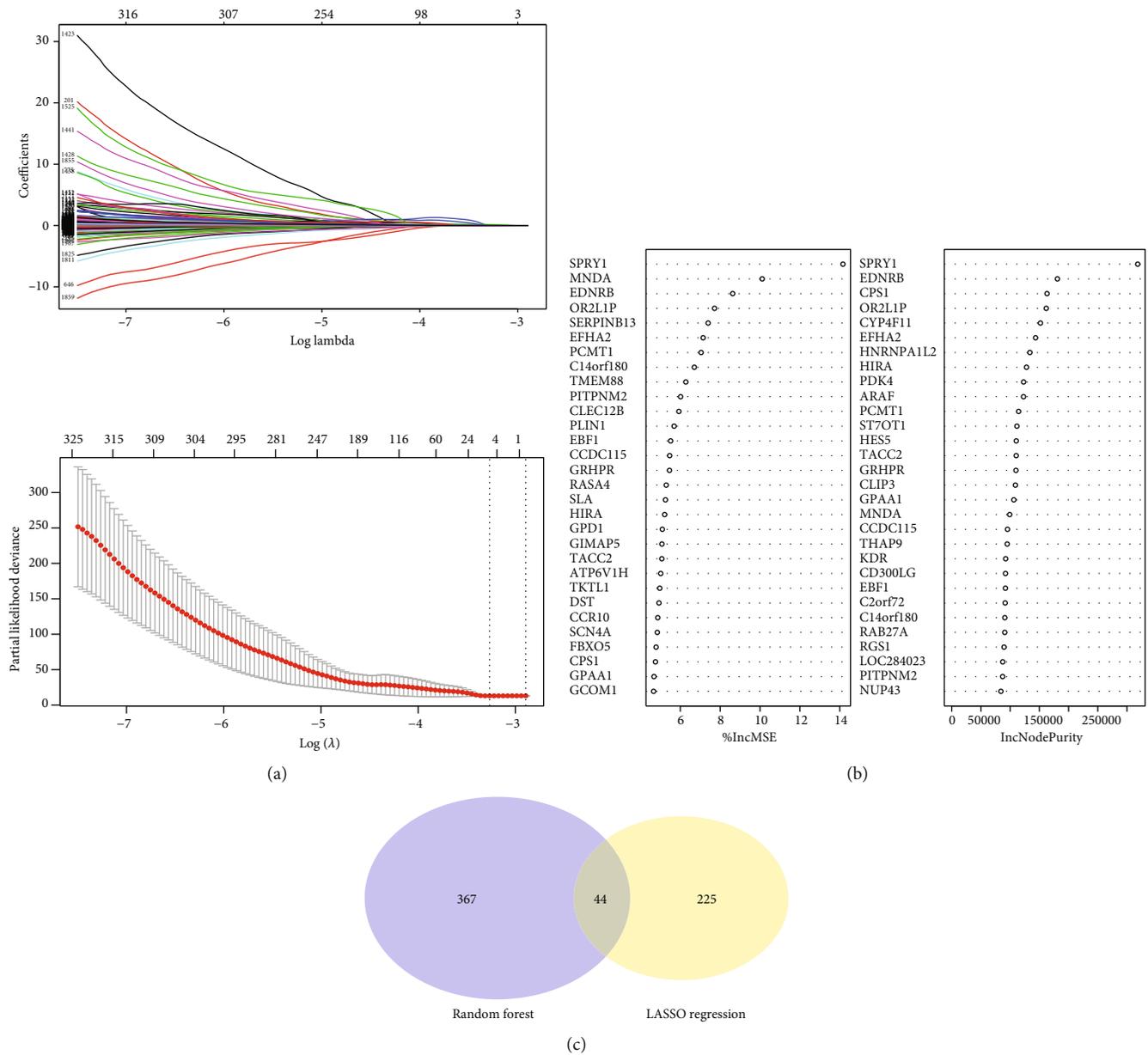


FIGURE 4: Identification of genes mostly significant for prognosis. LASSO analysis (a) and random forest analysis (b) were synchronously performed for the dimensional reduction and selection of variables, with 411 and 269 genes identified, respectively. The intersected proportion of 44 genes underwent the following analysis (c).

The construction of a prognostic panel experienced systematic analyses with gene shrinkage and selections. Survival analysis using the KM method was initially carried out to identify the significant genes with prognostic values for survival. As the criteria of the group established, the median of gene expression was adopted to partition patients into two groups, and a total of 20530 DEGs successively underwent assessment for statistical significance. Then, the combination of random forest analysis and LASSO analysis was performed for variable shrinkage and selection based on the recognized genes, and the intersection proportion was retrieved for precision. The expression of this group of selected ones, as the continuous variable, was further consecutively adopted into a univariate COX proportional model and evaluated

the corresponding significance to survival, and the genetic panel, comprising 15 genes, was finally determined. In combination with regression coefficients, gene expression was quantified as an indicator for OS of breast cancer, and the pIRS system was accordingly constructed as the product. Subsequently, individuals were classified into two pIRS groups based on the immunological contexture. The clinicopathological characteristics and the pIRS group of eligible patients were synchronously assessed by the application of multivariate COX regression analysis to organize a prognostic model, and the independent variables were utilized to create the nomogram validated with good performance.

We also carried out comparative analyses to portray the immune-related features contained in these patients from

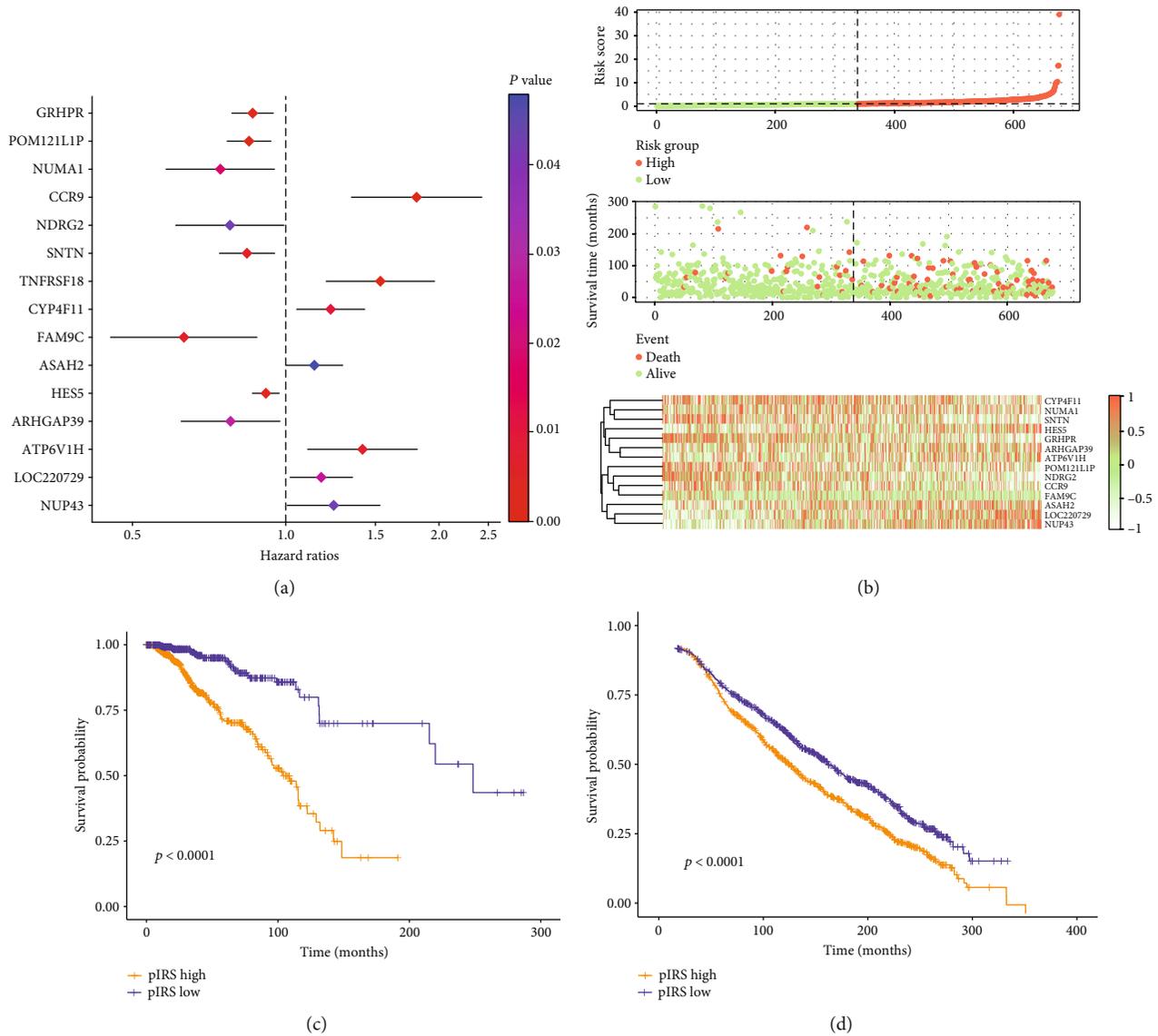
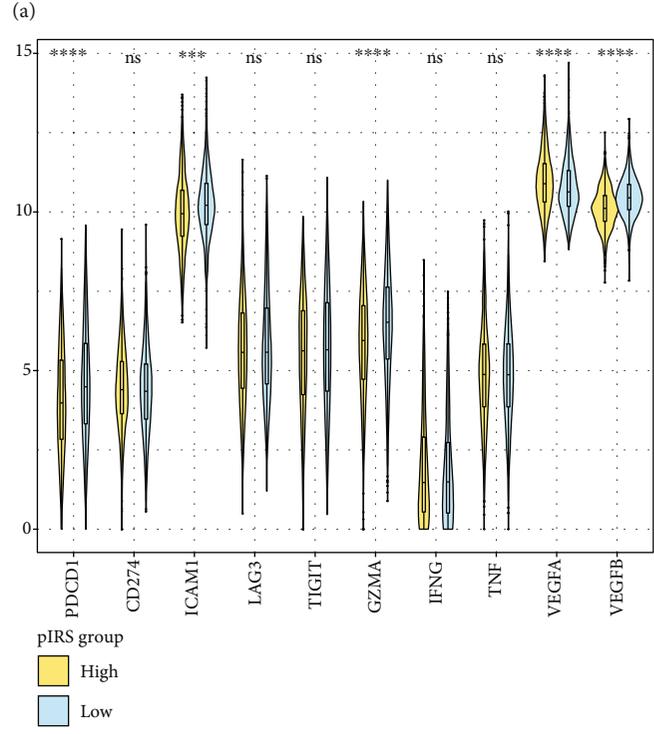
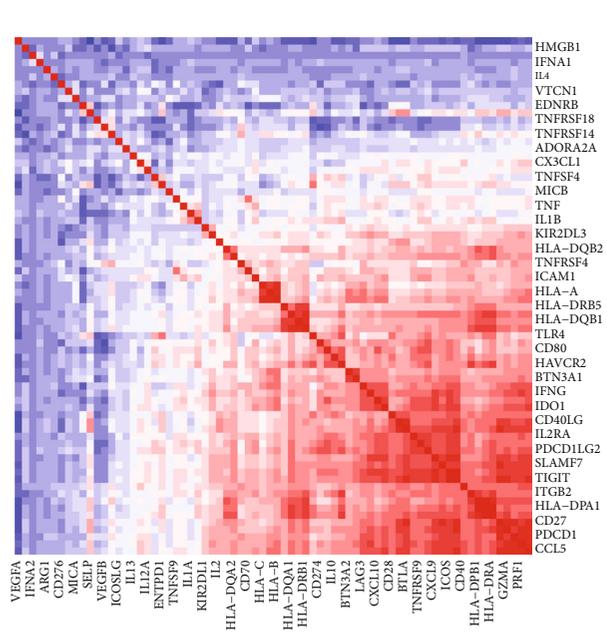
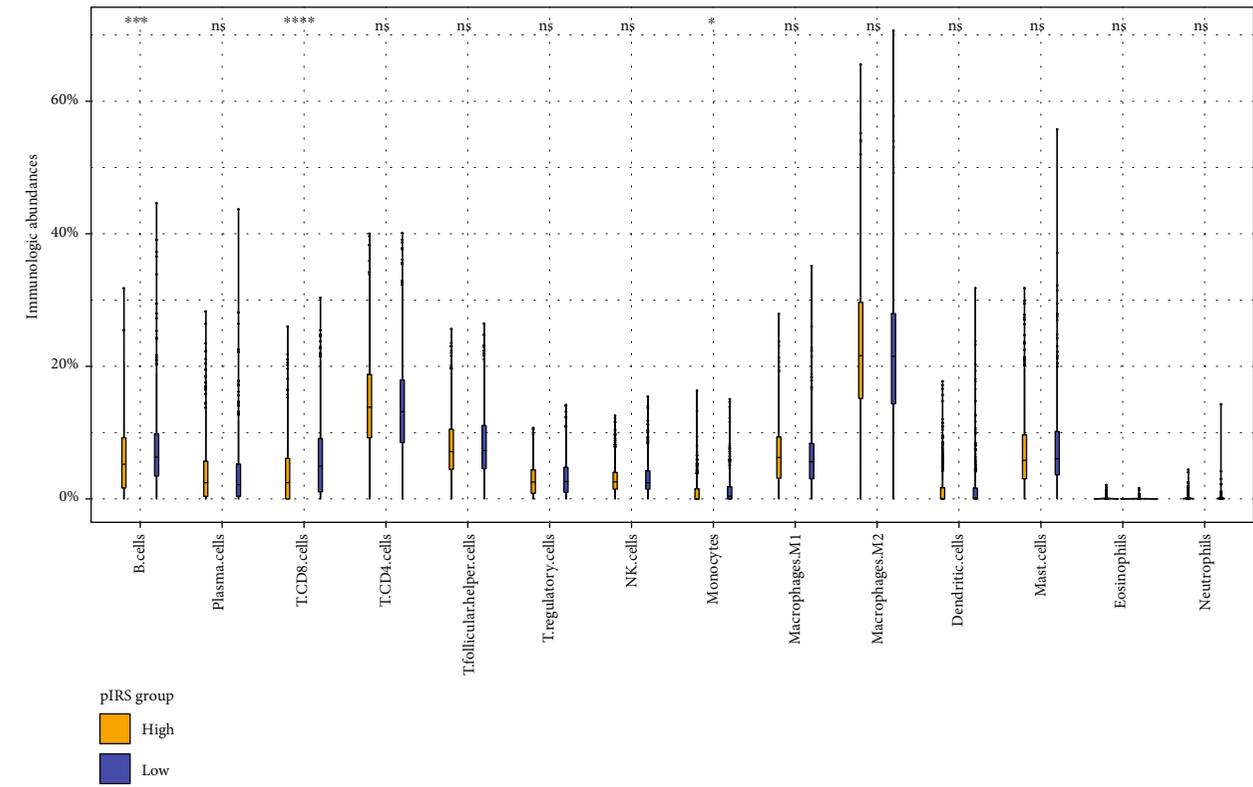


FIGURE 5: Construction and validation of the prognostic panel. The univariate COX regression analysis was carried out to recognize the variables with the utmost statistical significance with 15 genes determined (a). The correlations between gene expression and prognostic outcomes were demonstrated (b). The partition system based on pIRS was validated with survival analysis using the Kaplan-Meier method, which revealed a significant difference in survival of patients from both TCGA ($P < 0.0001$) (c) and METABRIC database ($P < 0.0001$) (d).

two groups. Results of tumor-infiltrating cells showed that the abundance of CD8 T lymphocyte, B cells, and monocytes was greatly higher in the low IRS group, revealing that these sorts of immunologic cells could presage improved survival of breast cancer. Currently, the relation between the densities of CD8 effector T cells and prognosis was reportedly controversial [28–31]. These findings confirmed that the improved survival was in positive associations with an enriched abundance of CD8 T cells. On top of that, recent researches have revealed that the spatially distinct distribution of CD8 T cells constitutes the leading reference to classifying the patterns of TME in the breast cancer subgroup [32]. This kind of association between the heterogeneous patterns of CD8 T cells and the clinical outcomes could be potentially correlated to the heterogeneity of TME in breast cancer as well. To further elu-

cidate the inherent correlations between immunologic phenotypes and distinct prognostic outcomes, the curated gene list of immune modulators was adopted [15], of which the correlogram was suggestive of evident collections consisting of these immune-related components. With the in-depth illustration of immunoediting theory and evolution of immunotherapy, immune checkpoints have been considered with cruciality in cancer immunity response, and the developed inhibitors have been dramatically improving the prognosis of several types of malignancies [33–35]. Herein, immune checkpoints were selected to be evaluated between these two pIRS groups. This part of results revealed that PDCD1 [36], ICAM1 [37], and GZMA [38], which were associated with proactivity of cytotoxic T lymphocytes and cancer immunity, were predictive of prolonged prognosis in the



(b)

(c)

FIGURE 6: Immunologic profiles of the pIRS groups. Comparative analysis of the abundances of infiltrating immunologic cells from the high pIRS group and the low pIRS group (a). The correlations among immune-related modulators in breast cancer tissue (b) and the difference in immune checkpoint between two groups (c).

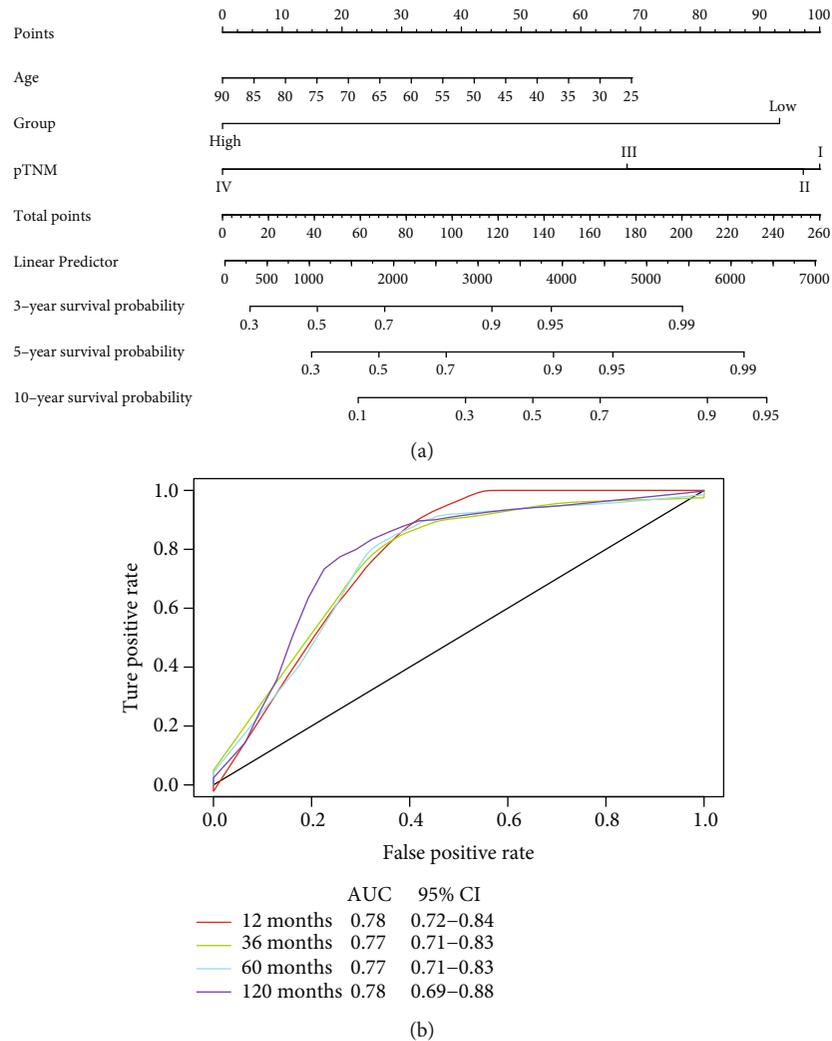


FIGURE 7: Prognostic model for survival probability of breast cancer. Nomogram constructed for the prediction of 3-year, 5-year, and 10-year survival probability of patients in the training cohort (a). The time-dependent ROC validated the performance of this prognostic model (b).

low pIRS group, and VEGFA, as the leading molecule of tumor-induced angiogenesis for tumor invasion [39], exhibited an increased expression in breast cancer patients with limited survival of patients from the high pIRS group.

Immune infiltration remains an intriguing focus in the field of oncology research, which has been considered not only an indicator for therapeutic effectiveness but also a promising basis, as the essential step for “immunomonitoring,” to drive treatment evolution [40]. The developing techniques have been applied to dissect tumor-related contexture which has shed novel light on the understanding of tumor heterogeneity and cancer management. However, current research could not fully interpret some clinical phenomenon that was inconsistent with experimental results, for instance, insufficient biomarkers with predictive values for response and prognosis. It seems that the intricate interactions among tumor-related complex systems, including tumor cells, immune cells, and TME, should be elucidated by integrating technologies and curate increasing findings for cancer therapy followed by survival benefits for patients [41]. Indeed, several genetic panels were constructed for the precise pre-

diction of survival in breast cancer cohorts, such as the Oncotype DX assay [42] and MammaPrint panel [43], and have been extensively validated in clinical practice. However, the current panel tends to be developed and designed for a specified cohort and relatively limited in the general utilization for the entire cohort. From this perspective, the curated panel introduced by our research was promising for future practice and promising benefits to breast cancer patients.

Some inevitable limitations should be stated here. Firstly, the sample size and clinicopathological characteristics, for instance, sequential treatment details and concomitant diseases, stored in the publicly available database were limited, which could attribute to the potential bias and weaken the power of the prediction model. Secondly, the partition method for the pIRS groups adopted the median as a cutoff value, which might lead to the inappropriate division of the cohort and the limited unearthing of promising findings. Under this circumstance, the standard for population selection should be optimized in the upcoming research. Last, the focus of TME phenotypes lay in the differential analysis for mechanism; however, the absence of deep mining on

the increasing factors that induced intratumor heterogeneity or divergences from populations was obvious.

5. Conclusions

In conclusion, we systematically derived the three molecularly and clinically distinct TME subtypes based on immunologic infiltration and explored the potential mechanisms promoting this kind of divergence. A predictive score system was built for population selection, and a prognostic model was built with the combinations with clinicopathological characteristics of cohorts for prediction of survival. Future studies are warranted to absorb increasing factors and optimize this model in favor of clinical practice in the perspective.

Data Availability

The datasets and/or analysis results used during the current study are available from the corresponding authors upon reasonable request.

Disclosure

This paper has been presented as a preprint and could be found in the link below: <https://www.researchsquare.com/article/rs-53701/v1>.

Conflicts of Interest

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

Supplementary Materials

Supplementary Figure 1: flow diagram of patient selection and cohort identification. Supplementary Figure 2: the overview and flow diagram of this study. Supplementary Figure 3: correlations among tumor-infiltrating immunologic cell types. Supplementary Figure 4: distributions of leukocyte infiltrations related to clinicopathological characteristics. Supplementary Figure 5: the gene expression of immunologic modulators in the cohort. Supplementary Table 1: basic information of the included datasets. Supplementary Table 2: the immunologic infiltrations of TCGA-BRCA patients quantitatively evaluated by CIBERSORT algorithms. Supplementary Table 3: the immunologic infiltrations of METABRIC patients quantitatively evaluated by CIBERSORT algorithms. Supplementary Table 4: the significant differentially expressed genes among three TME subtypes. Supplementary Table 5: the identified 15 genes with the foremost significance for prognosis. Supplementary Table 6: clinicopathological characteristics at baseline of training cohort ($N = 677$). Supplementary Table 7: the computational C-index of cohorts. (*Supplementary Materials*)

References

- [1] G. P. Dunn, L. J. Old, and R. D. Schreiber, "The three Es of cancer immunoediting," *Annual Review of Immunology*, vol. 22, no. 1, pp. 329–360, 2004.
- [2] L. Maggs and S. Ferrone, "Improving the clinical significance of preclinical immunotherapy studies through incorporating tumor microenvironment-like conditions," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 26, no. 17, pp. 4448–4453, 2020.
- [3] J. G. Egen, W. Ouyang, and L. C. Wu, "Human anti-tumor immunity: insights from immunotherapy clinical trials," *Immunity*, vol. 52, no. 1, pp. 36–54, 2020.
- [4] N. Harbeck, F. Penault-Llorca, J. Cortes et al., "Breast cancer," *Nature Reviews. Disease Primers*, vol. 5, no. 1, p. 66, 2019.
- [5] Y. Lavin, S. Kobayashi, A. Leader et al., "Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses," *Cell*, vol. 169, no. 4, pp. 750–765.e17, 2017.
- [6] A. M. Newman, C. L. Liu, M. R. Green et al., "Robust enumeration of cell subsets from tissue expression profiles," *Nature Methods*, vol. 12, no. 5, pp. 453–457, 2015.
- [7] M. D. Wilkerson and D. N. Hayes, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking," *Bioinformatics (Oxford, England)*, vol. 26, no. 12, pp. 1572–1573, 2010.
- [8] M. E. Ritchie, B. Phipson, D. Wu et al., "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [9] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics : a journal of integrative biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [10] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [11] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [12] J. M. Williamson, H. Y. Kim, A. Manatunga, and D. G. Addiss, "Modeling survival data with informative cluster size," *Statistics in Medicine*, vol. 27, no. 4, pp. 543–555, 2008.
- [13] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [14] A. N. Kamarudin, T. Cox, and R. Kolamunnage-Dona, "Time-dependent ROC curve analysis in medical research: current methods and applications," *BMC Medical Research Methodology*, vol. 17, no. 1, p. 53, 2017.
- [15] V. Thorsson, D. L. Gibbs, S. D. Brown et al., "The immune landscape of cancer," *Immunity*, vol. 48, no. 4, pp. 812–830.e14, 2018.
- [16] Y. Shen, X. Peng, and C. Shen, "Identification and validation of immune-related lncRNA prognostic signature for breast cancer," *Genomics*, vol. 112, no. 3, pp. 2640–2646, 2020.
- [17] W. Zhou, Y. Pang, Y. Yao, and H. Qiao, "Development of a ten-lncRNA signature prognostic model for breast cancer survival: a study with the TCGA database," *Analytical Cellular Pathology (Amsterdam)*, vol. 2020, article 6827057, pp. 1–9, 2020.
- [18] X. Li, Y. Li, X. Yu, and F. Jin, "Identification and validation of stemness-related lncRNA prognostic signature for breast

- cancer,” *Journal of Translational Medicine*, vol. 18, no. 1, p. 331, 2020.
- [19] H. Zheng, X. Li, C. Chen et al., “Quantum dot-based immunofluorescent imaging and quantitative detection of TOP2A and prognostic value in triple-negative breast cancer,” *International Journal of Nanomedicine*, vol. Volume 11, pp. 5519–5529, 2016.
- [20] Y. C. XU, F. C. ZHANG, J. J. LI et al., “RRM1, TUBB3, TOP2A, CYP19A1, CYP2D6: difference between mRNA and protein expression in predicting prognosis of breast cancer patients,” *Oncology Reports*, vol. 34, no. 4, pp. 1883–1894, 2015.
- [21] Z. Varga, A. Lebeau, H. Bu et al., “An international reproducibility study validating quantitative determination of ERBB2, ESRI, PGR, and MKI67 mRNA in breast cancer using MammaTyper®,” *Breast cancer research : BCR*, vol. 19, no. 1, p. 55, 2017.
- [22] M. Laible, K. Schlombs, K. Kaiser et al., “Technical validation of an RT-qPCR in vitro diagnostic test system for the determination of breast cancer molecular subtypes by quantification of ERBB2, ESRI, PGR and MKI67 mRNA levels from formalin-fixed paraffin-embedded breast tumor specimens,” *BMC Cancer*, vol. 16, no. 1, p. 398, 2016.
- [23] P. W. Xi, X. Zhang, L. Zhu et al., “Oncogenic action of the exosome cofactor RBM7 by stabilization of CDK1 mRNA in breast cancer,” *NPJ breast cancer*, vol. 6, no. 1, p. 58, 2020.
- [24] B. C. Chandler, L. Moubadder, C. L. Ritter et al., “TTK inhibition radiosensitizes basal-like breast cancer through impaired homologous recombination,” *The Journal of Clinical Investigation*, vol. 130, no. 2, pp. 958–973, 2020.
- [25] L. H. Alfarsi, R. E. Ansari, M. L. Craze et al., “CDC20 expression in oestrogen receptor positive breast cancer predicts poor prognosis and lack of response to endocrine therapy,” *Breast Cancer Research and Treatment*, vol. 178, no. 3, pp. 535–544, 2019.
- [26] A. Ueda, K. Oikawa, K. Fujita et al., “Therapeutic potential of PLK1 inhibition in triple-negative breast cancer,” *Laboratory investigation; a journal of technical methods and pathology*, vol. 99, no. 9, pp. 1275–1286, 2019.
- [27] N. J. Taylor, J. T. Bensen, C. Poole et al., “Genetic variation in cell cycle regulatory gene AURKA and association with intrinsic breast cancer subtype,” *Molecular Carcinogenesis*, vol. 54, no. 12, pp. 1668–1677, 2015.
- [28] J. Galon, A. Costes, F. Sanchez-Cabo et al., “Type, density, and location of immune cells within human colorectal tumors predict clinical outcome,” *Science (New York, N.Y.)*, vol. 313, no. 5795, pp. 1960–1964, 2006.
- [29] N. A. Giraldo, E. Becht, Y. Vano et al., “Tumor-infiltrating and peripheral blood T-cell immunophenotypes predict early relapse in localized clear cell renal cell carcinoma,” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 23, no. 15, pp. 4416–4428, 2017.
- [30] D. W. Scott, F. C. Chan, F. Hong et al., “Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma,” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, vol. 31, no. 6, pp. 692–700, 2013.
- [31] T. Jamiyan, H. Kuroda, R. Yamaguchi et al., “Prognostic impact of a tumor-infiltrating lymphocyte subtype in triple negative cancer of the breast,” *Breast cancer (Tokyo, Japan)*, vol. 27, no. 5, pp. 880–892, 2020.
- [32] Y. Bareche, L. Buisseret, T. Gruosso et al., “Unraveling triple-negative breast cancer tumor microenvironment heterogeneity: towards an optimized treatment approach,” *Journal of the National Cancer Institute*, vol. 112, no. 7, pp. 708–719, 2020.
- [33] P. Schmid, J. Cortes, L. Pusztai et al., “Pembrolizumab for early triple-negative breast cancer,” *The New England Journal of Medicine*, vol. 382, no. 9, pp. 810–821, 2020.
- [34] B. Burtness, K. J. Harrington, R. Greil et al., “Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study,” *Lancet (London, England)*, vol. 394, no. 10212, pp. 1915–1928, 2019.
- [35] M. TSK, Y. L. Wu, I. Kudaba et al., “Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial,” *Lancet (London, England)*, vol. 393, no. 10183, pp. 1819–1830, 2019.
- [36] M. Santaripa, A. Aguilar, I. Chaib et al., “Non-small-cell lung cancer signaling pathways, metabolism, and PD-1/PD-L1 antibodies,” *Cancers*, vol. 12, no. 6, p. 1475, 2020.
- [37] H. Feng, D. L. Stachura, R. M. White et al., “T-lymphoblastic lymphoma cells express high levels of BCL2, S1P1, and ICAM1, leading to a blockade of tumor cell intravasation,” *Cancer Cell*, vol. 18, no. 4, pp. 353–366, 2010.
- [38] Z. Zhou, H. He, K. Wang et al., “Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells,” *Science (New York, N.Y.)*, vol. 368, no. 6494, p. eaaz7548, 2020.
- [39] L. Ma, M. O. Hernandez, Y. Zhao et al., “Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer,” *Cancer Cell*, vol. 36, no. 4, pp. 418–430.e6, 2019.
- [40] W. H. Fridman, L. Zitvogel, C. Sautès-Fridman, and G. Kroemer, “The immune contexture in cancer prognosis and treatment,” *Nature Reviews. Clinical Oncology*, vol. 14, no. 12, pp. 717–734, 2017.
- [41] W. H. Fridman, F. Pagès, C. Sautès-Fridman, and J. Galon, “The immune contexture in human tumours: impact on clinical outcome,” *Nature Reviews Cancer*, vol. 12, no. 4, pp. 298–306, 2012.
- [42] S. Paik, S. Shak, G. Tang et al., “A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer,” *The New England Journal of Medicine*, vol. 351, no. 27, pp. 2817–2826, 2004.
- [43] F. Cardoso, L. J. van’t Veer, J. Bogaerts et al., “70-gene signature as an aid to treatment decisions in early-stage breast cancer,” *The New England Journal of Medicine*, vol. 375, no. 8, pp. 717–729, 2016.