Tumor microenvironment subtypes and immune-related signatures for the prognosis of 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
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Binghe Xu (xubingheBM@163.com)
https://orcid.org/0000-0002-0234-2747

Primary research

Keywords:

Posted Date: August 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-53701/v1

Version of Record: A version of this preprint was published at BioMed Research International on June 1st, 2021. See the published version at https://doi.org/10.1155/2021/6650107.

Abstract

To better understand the heterogeneity of tumor microenvironment (TME) and establish a prognostic model for breast cancer in clinical practice, the leukocyte infiltrations of 22 cell types of interest from 2620 breast cancer patients were quantitatively estimated using deconvolution algorithms, and three TME subtypes with distinct molecular and clinical features were identified by unsupervised clustering approach. Then, we carried out systematic analyses to illustrate the contributing mechanisms for differential phenotypes, which suggested that the divergences were distinguished by cell cycle dysfunction, variation of cytotoxic T lymphocytes activity. Next, through dimensionally reduction and selection based on random-forest analysis, least absolute shrinkage and selection operator (LASSO) analysis, and uni- and multivariate COX regression analysis, a total of 15 significant genes were proposed to construct the prognostic immune-related score (pIRS) system and, in combinations with clinicopathological characteristics, a predictive model was ultimately built with well performance for survival of breast cancer patients. Comparative analyses demonstrated that proactivity of CD8 T lymphocytes and hyper-angiogenesis could be attributed to distinct prognostic outcomes. In conclusion, we retrieved three TME phenotypes and the curated prognostic model based on pIRS system for breast cancer. This model is justified for validation and optimized in the coming future.

Introduction

Currently, the landscape of tumor microenvironment (TME) has been generally portraying, of which the components are considered as an essential composition of cancer immunity, with counterpart activities across ‘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, clinical response provided by immunotherapy were inconsistent among populations or even in the changing stage of an individual, which could be the result of heterogeneities exist in this complicated interactive contexture.

Breast cancer is a worldwide leading newly-diagnosed cancer in the female and a heterogenous 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 tumor microenvironment[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 multi-omics 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 well performance for patients with breast cancer.

Methods

Breast cancer datasets and preprocessing

The breast cancer gene expression dataset, as 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) were obtained from the University California of Santa Cruz (UCSC) Xena browser (http://xena.ucsc.edu/). For supplement, gene expression profile and curated clinical data (N = 1525), as validation dataset, were retrieved from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database. Detailed information of included datasets was provided in Table S1.

Infiltrating abundances and TME subtypes

On the basis of gene expression mixture, we applied 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.standford.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 log-rank test.

Identification for differentially expressed genes (DEGs) among TME clusters was accomplished with limma R package[8], which the absolute of fold change (FC) more than 1 and P value adjusted by Benjamini-Hochberg method less than 0.05 were considered as the criteria for significant DEGs with annotations searched from 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 Molecular Signatures Database (http://gsea-msigdb.org/) using ClusterProfiler R package[9, 10]. Through STRING database (http://string.db.org/), the functional analysis of protein-protein interaction (PPI) networks were established and further visualized by Cytoscape software (version 3.8.0)[11].

Prognostic immune-related score (pIRS) and prognostic model

Survival analyses concerning expression of each DEG for overall survival (OS) were performed using KM method by survival and survminer R packages[12], and variables with statistical significance were further through with random-forest analysis and least absolute shrinkage and selection operator (LASSO) analysis, using randomForest package and glmnet package of R software[13], for dimensionally reduction and identification of the overlapping DEGs significant for prognosis. Next, univariate COX

proportional hazard model was utilized to differentiate the foremost groups of DEGs and the prognostic immune-related score (pIRS) was defined as:

$$\mathrm{pIRS} = \sum_{k}^{n} exprs^{k} coef^{k} + \sum_{j}^{n} exprs^{i} coef^{i}$$

The $exprs^{k}$ and $coef^{k}$ was the gene expression and regression coefficient for DEGs of which hazard ratio (HR) more than 1, while $exprs^{i}$ and $coef^{i}$ was for the genes of which HR less than 1.

In combination with clinicopathological characteristics, pIRS was included in the multivariate COX regression analysis to construct nomogram with 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 prognostic model was quantitatively measured by Harrell's concordance index (C-index). The flow diagram of data processing and analysis was presented in Figure S1.

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 independent t-test for normal distribution and Mann-Whitney U test for abnormal distribution, respectively, which was visualized using 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).

Results

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 was exhibited in Fig. 1, Figure S2 and Table S2. Data on the estimated proportions of infiltrating cells indicated an evident heterogeneity, considering, an unsupervised clustering analysis was performed to determine the potential TME subtypes. On the basis of 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) (Fig. 2a-b). With the aim of validation, the bulk tissue gene expression profiles of 1525 patients from METABRIC database was also estimated for immune infiltrations and undergone a clustering analysis (Table S3), from which three clusters were obtained with statistical difference in OS (P < 0.001) (Fig. 2c-d), indicating this partition was stable for TME subtypes in terms of immunological infiltrations.

Signatures of TME phenotypes

To explore the contributing mechanisms for versatile TME phenotypes, we systematically conducted 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 up-regulated and 431 genes down-regulated (Fig. 3a-b, Table S4). Among the identified DEGs, MMP9, SLC16A3, CDT1, CA9 and CDC20 were revealed to be the leading up-regulated, while P2RY12, GPR34, ABCB1, IGF1, and CX3CR1 were estimated to express down-regulated 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 including collagen-containing extracellular matrix, condensed chromosome, centromeric region, and spindle (Fig. 3c). Downstream analysis from KEGG suggested that DEGs significantly involved in the enriched pathways comprising neuroactive ligand-receptor interaction, PI3K-Akt signaling pathways, and cell cycle (Fig. 3d). GSEA results showed the DEGs, in terms of cancer hallmarks, were ensemble of cell cycle dysfunction, including E2F targets and G2M checkpoints, while were significantly muted in metabolism 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 (Fig. 3e). Built on cluster analysis, PPI networks were built to clarify the interactive functions of significant DEGs, and the corresponding relationships were shown with ranked degree (Fig. 3f).

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 were 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 (Fig. 4a-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 (Fig. 5a-b, Table S5).

Subsequently, the pIRS was constructed based on the expression of 15 genes in combinations with the corresponding coefficient for each patient, and all cases were stratified into high pIRS group and 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 ($P < 0.0001$) (Fig. 5c-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 low pIRS group (Fig. 6a).

Distributions of leukocyte infiltration among groups with diverse clinicopathological factors was shown in Figure S3. The expression levels of immunologic modulators were also evaluated between two groups with a total of 73 signature genes included[15] (Figure S4). Interactive correlations were shown in Fig. 6b, and the genes, in particular, associated with the immune checkpoint were undergone comparative analyses (Fig. 6c). Results from paired analyses showed that the expression of PDCD1 (PD-1) and ICAM1 was significantly higher in low pIRS group, while VEGFA presented an increasing trend of expression in high pIRS group.

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 quantitatively estimation of survival. Results from multivariate regression analysis demonstrated that the age, pTNM, and pIRS group were independent factors for prognosis of breast cancer, which were utilized to construct nomogram for the prediction of 3-year, 5-year, and 10-year survival probability (Fig. 7a, Table S6). Time-dependent ROC suggested that the time-dependent under curve area were ranging from 0.77 to 0.78, indicating the curated prognostic model was well performed (Fig. 7b), and the quantified C-index was 0.823 obtained from training cohort and 0.776 from validation cohort, respectively, which were generally higher than those computed from TNM staging system, revealing the robustness in addition to a better accuracy of this prediction model (Table S7).

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 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, which promoting us to perform an analysis for the potential subtypes of TME. Three TME clusters were following 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, which was in accordance with the findings from previous studies[16, 17]. Collectively, the potential mechanisms leading to the differential TME phenotypes 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 tends 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 varies in CD8 T lymphocyte activity, and the repression of metabolite homeostasis.

The construction of prognostic panel was experienced systematic analyses with genes shrinkage and selections. Survival analysis using KM method was initially carried out to identify the significant genes with prognostic values for survival. As the criteria of group established, the median of gene expression was adopted to partition patients into two groups and the total of 20530 DEGs were successively undergone assessment for the statistical significance. Then, the combination of random forest analysis and Lasso analysis was following 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 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 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 well performance.

We also carried out comparative analyses to portray the immune-related features containing in these patients from two groups. Results of tumor-infiltrating cells showed that the abundance of CD8 T lymphocyte, B cells, and monocytes was greatly higher in low IRS group, revealing 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[18–20]. These findings confirmed the improved survival was in positive associations with enriched abundance of CD8 T cells. To further elucidate 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 consisted 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[21–23]. Herein, immune checkpoints were selected to be evaluated between these two pIRS groups. This part of results revealed that PDCD1[24], ICAM1[25], and GZMA[26], which were associated with proactivity of cytotoxic T lymphocytes and cancer immunity, were predictive of prolonged prognosis in low pIRS group, and VEGFA, as the leading molecule of tumor-induced angiogenesis for tumor invasion[27], exhibited an increased expression in breast cancer patients with limited survival of patients from high pIRS group.

Immune infiltration remains an intriguing focus in the field of oncology research, which has been considered as not only an indicator for therapeutic effectiveness but also a promising basis, as the essential step for ‘immunomonitoring’, to drive treatment evolution[28]. The developing techniques have been applied to dissect tumor-related contexture which have 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 with survival benefits for patients[29].

Some inevitable limitations should be stated here. Firstly, the sample size and clinicopathological characteristics, for instance, sequential treatment details and comitant diseases, stored in publicly available database were limited, considering, which could attribute to the potential bias and weaken the power of prediction model. Secondly, the partition method for pIRS groups was adopted the median as a cutoff value, which might lead to the inappropriate division of the cohort and the limited unearth 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 with deep mining on the increasing factors that induced intratumor heterogeneity or divergences from populations was obvious.

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.

Declarations

Acknowledgements Not applicable.

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.

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)).

Ethics approval and consent to participant Not applicable.

Consent for publication Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol. 2004;22:329–60.
2. Maggs L, Ferrone S. **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* 2020.
3. Egen JG, Ouyang W, Wu LC. Human Anti-tumor Immunity: Insights from Immunotherapy Clinical Trials. Immunity. 2020;52(1):36–54.
4. Adams S., Schmid P., Rugo H. S., Winer E. P., Loirat D., Awada A., Cescon D. W., Iwata H., Campone M., Nanda R. et al: **Pembrolizumab monotherapy for previously treated metastatic triple-negative breast cancer: cohort A of the phase II KEYNOTE-086 study**. *Annals of oncology*: *official journal of the European Society for Medical Oncology* 2019, **30**(**3**):**397–404**.
5. Y Lavin, S Kobayashi, A Leader, ED Amir, N Elefant, C Bigenwald, R Remark, R Sweeney, CD Becker, JH Levine et al: **Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses**. *Cell* 2017, **169**(**4**):**750–765.e717**.
6. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015;12(5):453–7.
7. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics (Oxford. England) 2010;26(12):1572–3.
8. 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(7):e47.

9. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7. a journal of integrative biology.
10. A Subramanian, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub, ES Lander 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* 2005, **102**(**43**):**15545–15550**.
11. P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, **T Ideker**: **Cytoscape**: **a software environment for integrated models of biomolecular interaction networks**. Genome research 2003, **13**(**11**):**2498–2504**.
12. Williamson JM, Kim HY, Manatunga A, Addiss DG. Modeling survival data with informative cluster size. Statistics in medicine. 2008;27(4):543–55.
13. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software. 2010;33(1):1–22.
14. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC medical research methodology. 2017;17(1):53.
15. V Thorsson, DL Gibbs, SD Brown, D Wolf, DS Bortone, TH Ou Yang, E Porta-Pardo, GF Gao, CL Plaisier, JA Eddy et al: **The Immune Landscape of Cancer**. *Immunity* 2018, **48**(**4**):**812–830.e814**.
16. C Denkert, G **von** Minckwitz, JC Brase, BV Sinn, S Gade, R Kronenwett, BM Pfitzner, C Salat, S Loi, WD Schmitt et al: **Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers**. *Journal of clinical oncology*: *official journal of the American Society of Clinical Oncology* 2015, **33**(**9**):**983–991**.
17. JH Park, SF Jonas, G Bataillon, C Criscitiello, R Salgado, S Loi, G Viale, HJ Lee, MV Dieci, SB Kim et al: **Prognostic value of tumor-infiltrating lymphocytes in patients with early-stage triple-negative breast cancers (TNBC) who did not receive adjuvant chemotherapy**. *Annals of oncology*: *official journal of the European Society for Medical Oncology* 2019, **30**(**12**):**1941–1949**.
18. J Galon, A Costes, F Sanchez-Cabo, A Kirilovsky, B Mlecnik, C Lagorce-Pagès, M Tosolini, M Camus, A Berger, P Wind et al: **Type, density, and location of immune cells within human colorectal tumors predict clinical outcome**. *Science (New York*, *NY)* 2006, **313**(**5795**):**1960–1964**.
19. NA Giraldo, E Becht, Y Vano, F Petitprez, L Lacroix, P Validire, R Sanchez-Salas, A Ingels, S Oudard, A Moatti 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* 2017, **23**(**15**):**4416–4428**.
20. DW Scott, FC Chan, F Hong, S Rogic, KL Tan, B Meissner, S Ben-Neriah, M Boyle, R Kridel, A Telenius 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* 2013, **31**(**6**):**692–700**.

21. P Schmid, J Cortes, L Pusztai, H McArthur, S Kümmel, J Bergh, C Denkert, YH Park, R Hui, N Harbeck et al: **Pembrolizumab for Early Triple-Negative Breast Cancer**. *The New England journal of medicine* 2020, **382**(**9**):**810–821**.
22. B Burtness, KJ Harrington, R Greil, D Soulières, M Tahara, G de Castro, A Psyrri, N Basté, P Neupane, Å Bratland 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)* 2019, **394**(**10212**):**1915–1928**.
23. TSK Mok, YL Wu, I Kudaba, DM Kowalski, BC Cho, HZ Turna, G Castro, V Srimuninnimit, KK Laktionov, I Bondarenko 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)* 2019, **393**(**10183**):**1819–1830**.
24. M Santarpia, A Aguilar, I Chaib, AF Cardona, S Fancelli, F Laguia, JWP Bracht, P Cao, MA Molina-Vila, N Karachaliou et al: **Non-Small-Cell Lung Cancer Signaling Pathways, Metabolism, and PD-1/PD-L1 Antibodies**. *Cancers* 2020, **12**(**6**).
25. H Feng, DL Stachura, RM White, A Gutierrez, L Zhang, T Sanda, CA Jette, JR Testa, DS Neuberg, DM Langenau et al: **T-lymphoblastic lymphoma cells express high levels of BCL2, S1P1, and ICAM1, leading to a blockade of tumor cell intravasation**. *Cancer cell* 2010, **18**(**4**):**353–366**.
26. Z Zhou, H He, K Wang, X Shi, Y Wang, Y Su, Y Wang, D Li, W Liu, Y Zhang et al: **Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells**. Science (New York, NY) 2020, **368**(**6494**).
27. L Ma, MO Hernandez, Y Zhao, M Mehta, B Tran, M Kelly, Z Rae, JM Hernandez, JL Davis, SP Martin et al: **Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer**. *Cancer cell* 2019, **36**(**4**):**418–430.e416**.
28. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nature reviews Clinical oncology. 2017;14(12):717–34.
29. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nature reviews Cancer. 2012;12(4):298–306.
30. Appendices.

Figures

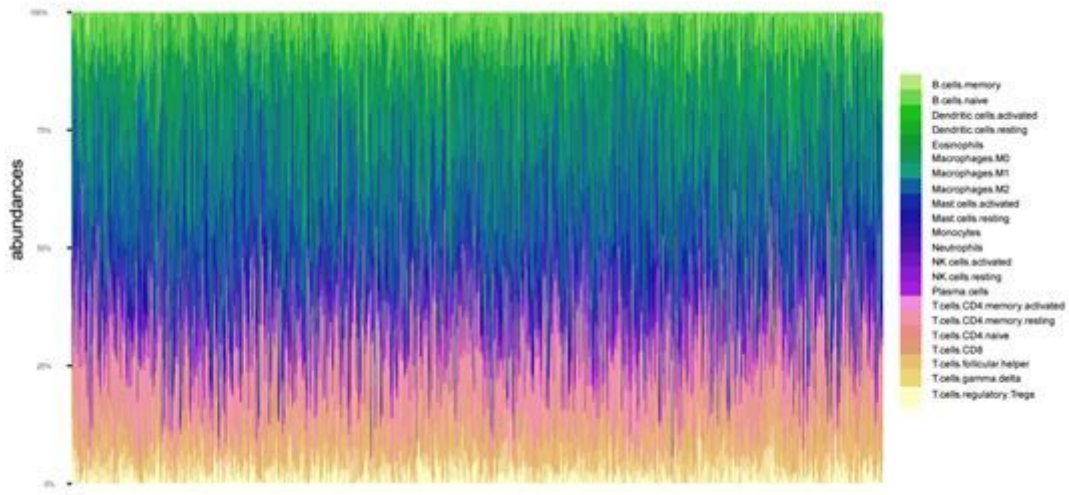

Figure 1

The landscape of immunologic infiltrations in 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.

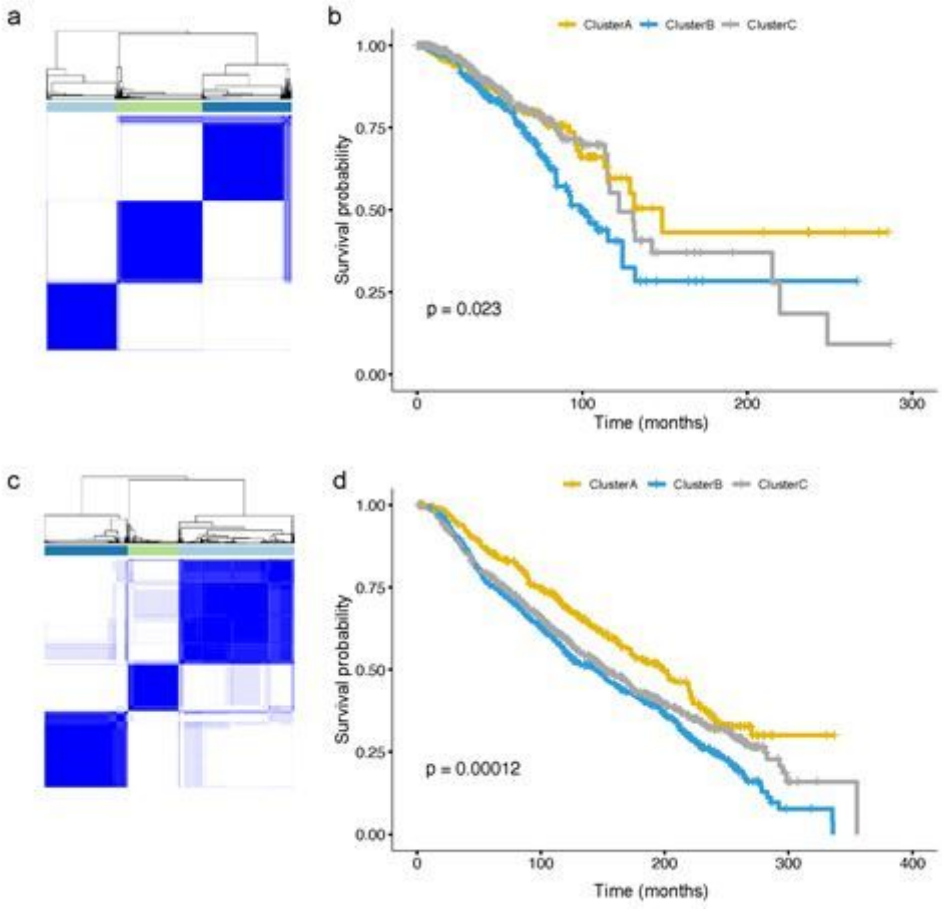

Figure 2

The curated TME subtypes with distinct prognostic outcomes. Three TME phenotypes of TCGA-BRCA patients (N=1096) were retrieved by consensus clustering method (a) with a significant difference in overall survival (P=0.023) (b). Consistent results were obtained in patients from METABRIC database (N=1525) (c) with distinction in prognosis (P=0.00012) (d). The survival analyses were performed with Kaplan-Meier method using log-rank

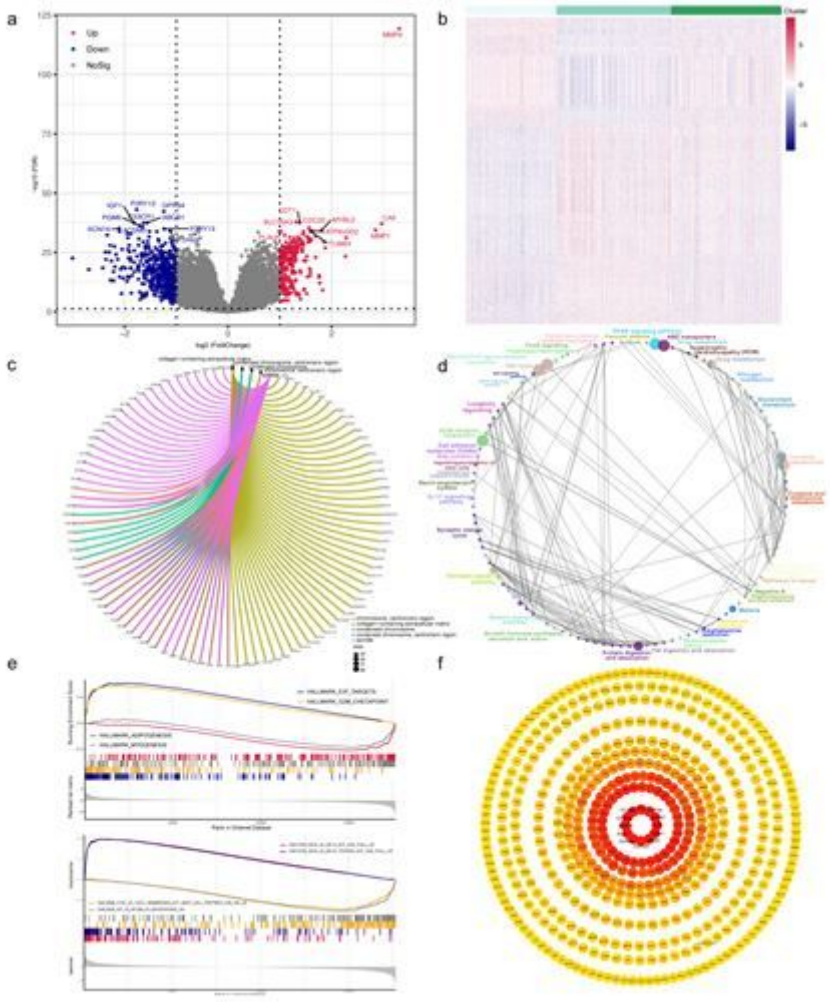

Figure 3

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)

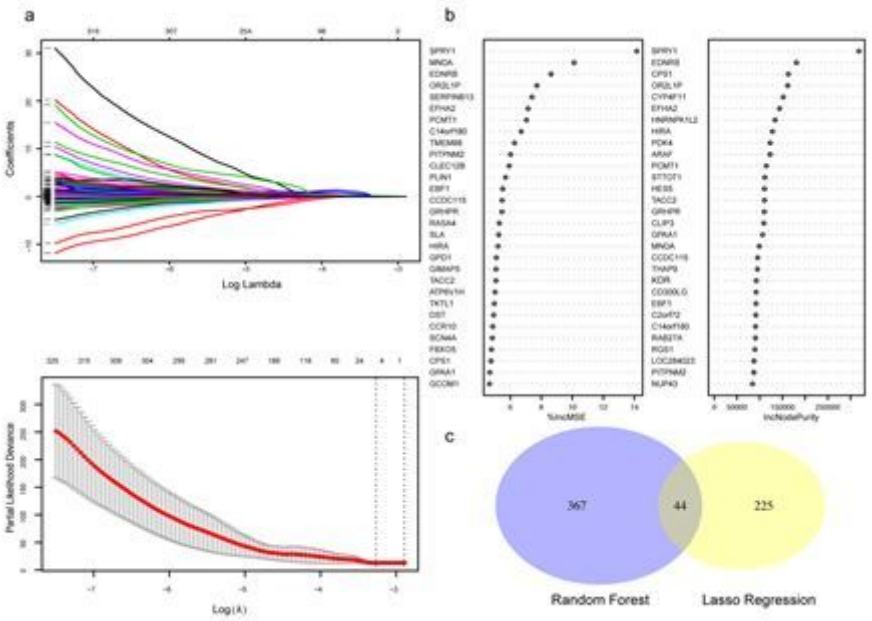

Figure 4

Identification of genes mostly significant for prognosis. Lasso analysis (a) and random forest analysis (b) were synchronously performed for the dimensionally reduction and selection of variables, with 411 and 269 genes identified, respectively. The intersected proportion of 44 genes was undergone the following analysis (c).

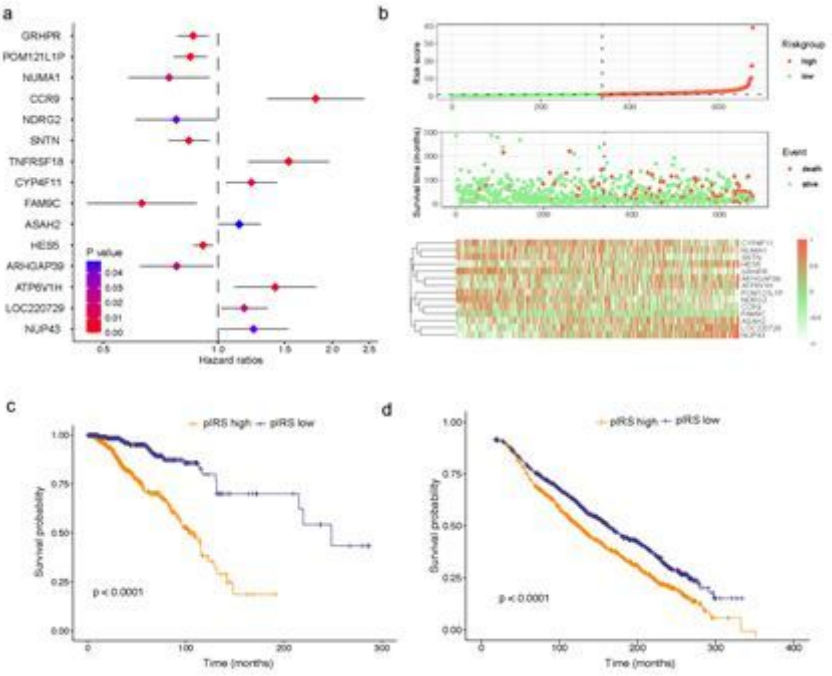

Figure 5

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 Kaplan-Meier method, which revealed significant difference in survival of patients from both TCGA (P<0.0001) (c) and METABRIC database (P<0.0001) (d).

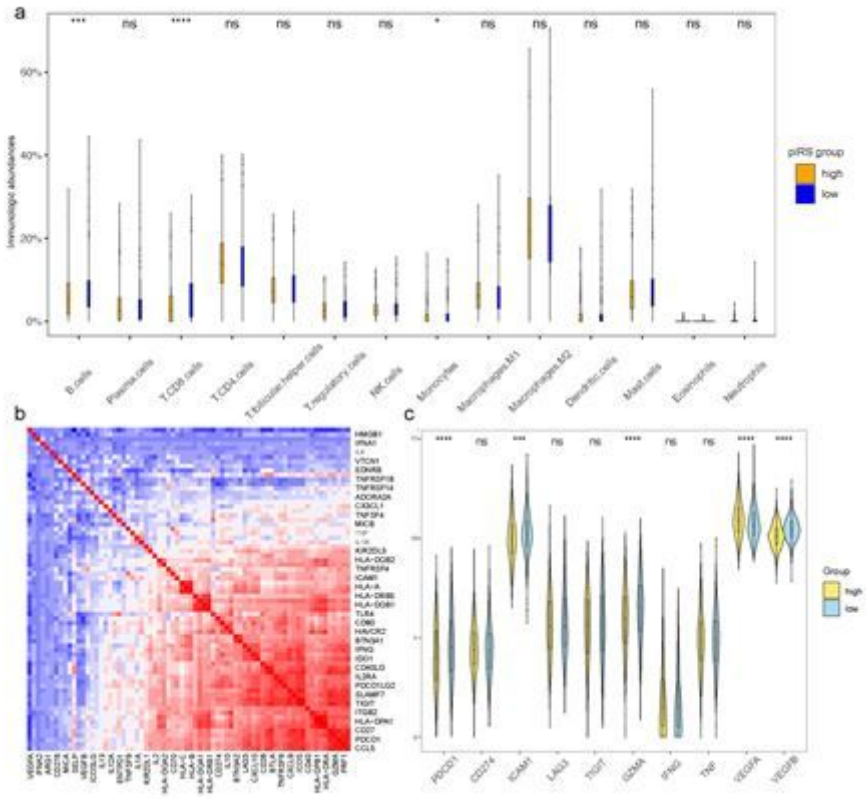

Figure 6

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).

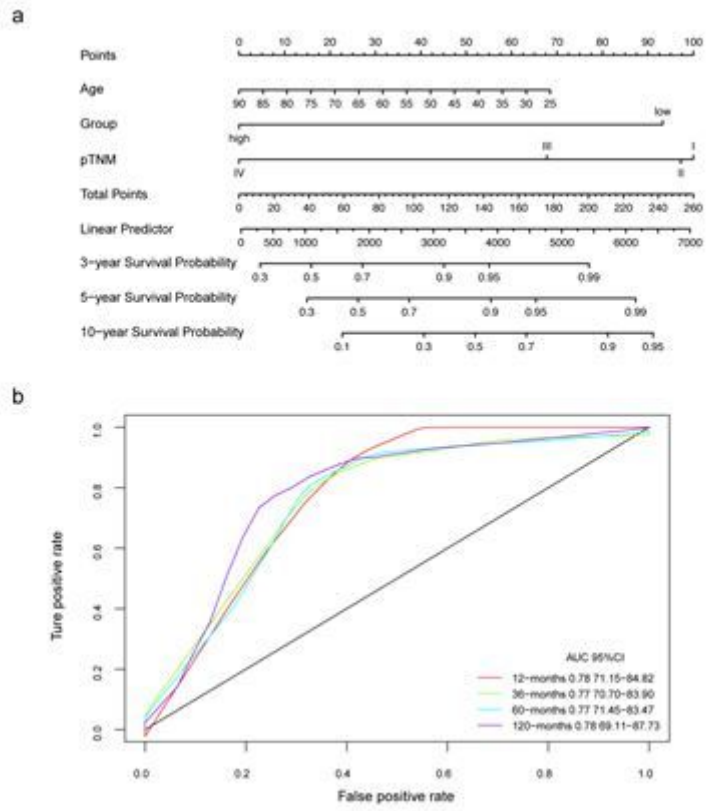

Figure 7

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).

Supplementary Files

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

- Supplementarycontent.pdf