Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer

Yi Xiao,1,† Ding Ma,1,‡ Shen Zhao,1 Chen Suo,2 Jinxiu Shi,3 Mengzhu Xue,4 Miao Ruan,5 Hai Wang1, Jingjing Zhao,6 Qin Li,6 Peng Wang,7 Leming Shi,2 Wentao Yang,5 Wei Huang,3 Xin Hu,1 Ke-Da Yu,1 Shenglin Huang,6 François BERTUCCI,8 Yi-Zhou Jiang,1,* Zhi-Ming Shao,1,* Written on behalf of AME Breast Cancer Collaborative Group

† These authors contributed equally to this work

1. Department of Breast Surgery, Fudan University Shanghai Cancer Center, Shanghai, 200032, China
2. State Key Laboratory of Genetic Engineering, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, 200433, China
3. Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center and Shanghai Industrial Technology Institute (SITI), Shanghai, 201210, China
4. SARI center for Stem Cell and Nanomedicine, Shanghai Advanced Research Institute, Chinese Academy of Sciences, 200031, Shanghai, China
5. Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, 200032, China
6. Fudan University Shanghai Cancer Center and Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai 200032, China
7. Bio-Med Big Data Center, Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological sciences, Chinese Academy of Sciences, 200031, Shanghai, China
8. Predictive Oncology team, CRCM, Department of Medical Oncology, Institut Paoli-Calmettes, 232 Bd de Ste-Marguerite, 13273, Marseille, France

Running title: Microenvironment landscape of triple-negative breast cancer

* Address for correspondence and reprints:
Dr. Yi-Zhou Jiang
Fudan University Shanghai Cancer Center, No. 270 Dong’an Road, Shanghai, 200032, P.R. China
Email: yizhoujiang@fudan.edu.cn
Prof. Zhi-Ming Shao
Fudan University Shanghai Cancer Center, No. 270 Dong’an Road, Shanghai, 200032, P.R. China
Email: zhimin_shao@yeah.net

Conflicts of interests: The authors declare no potential conflicts of interest.
Based on the multi-omics data of the largest original TNBC cohort, our study characterized the landscape of the TNBC microenvironment from an immunogenomics perspective. We revealed the heterogeneity and significant prognostic efficacy of TNBC microenvironment phenotypes. According to the heterogeneous microenvironment and immune escape mechanisms of TNBC, we proposed personalized immunotherapy for TNBC patients. Specifically, we suggest the selective application of immune checkpoint inhibitors (ICIs) to patients with the “immune-inflamed” cluster. Tumor-infiltrating lymphocytes and expression of immune checkpoint molecules, but not the tumor mutation burden, are potential biomarkers for predicting the therapeutic efficacy. Moreover, the transformation of “cold tumors” into “hot tumors” and the combination of ICIs could be considered alternative therapeutic strategies for patients in the “immune-desert” and “innate immune-inactivated” clusters. Some oncogenic pathways, such as the MYC-related and PI3K-AKT pathways, were identified as potential targets for this transformation.
Abstract

**Purpose:** The tumor microenvironment has a profound impact on prognosis and immunotherapy. However, the landscape of the triple-negative breast cancer (TNBC) microenvironment has not been fully understood.

**Experimental Design:** Using the largest original multi-omics dataset of TNBC (n = 386), we conducted an extensive immunogenomic analysis to explore the heterogeneity and prognostic significance of the TNBC microenvironment. We further analyzed the potential immune escape mechanisms of TNBC.

**Results:** The TNBC microenvironment phenotypes were classified into three heterogeneous clusters: cluster 1, the “immune-desert” cluster, with low microenvironment cell infiltration; cluster 2, the “innate immune-inactivated” cluster, with resting innate immune cells and nonimmune stromal cells infiltration; and cluster 3, the “immune-inflamed” cluster, with abundant adaptive and innate immune cells infiltration. The clustering result was validated internally with pathological sections and externally with TCGA and METABRIC cohorts. The microenvironment clusters had significant prognostic efficacy. In terms of potential immune escape mechanisms, cluster 1 was characterized by an incapability to attract immune cells, and MYC amplification was correlated with low immune infiltration. In cluster 2, chemotaxis but inactivation of innate immunity and low tumor antigen burden might contribute to immune escape, and mutations in the PI3K-AKT pathway might be correlated with this effect. Cluster 3 featured high expression of immune checkpoint molecules.

**Conclusions:** Our study represents a step towards personalized immunotherapy for TNBC patients. Immune checkpoint inhibitors might be effective for “immune-inflamed” cluster, and the transformation of “cold tumors” into “hot tumors” should be considered for “immune-desert” and “innate immune-inactivated” clusters.
Introduction

Triple-negative breast cancer (TNBC) is the most aggressive breast cancer subtype, which is defined as no expression of estrogen receptor (ER) and progesterone receptor (PR) and no amplification or overexpression of human epidermal growth factor receptor 2 (HER2) (1, 2). Early relapse and lack of therapeutic targets are major problems in TNBC treatment (3, 4). Recent research has demonstrated that TNBC had higher immunogenicity than other subtypes do, suggesting immunotherapeutic strategies for TNBC patients (5-9). However, recent clinical trials indicated that immunotherapy, such as immune checkpoint inhibitors (ICIs), showed low efficacy in the whole population of TNBC patients (10-12). A lack of patient selection based on tumor microenvironment landscapes might be the major reason for these disappointing results. Meanwhile, the whole landscape of TNBC microenvironment phenotypes remains unknown. Previous studies have focused on only one or two microenvironment cell subtypes of TNBC (13-20), which might result in the biased understanding of TNBC microenvironment. As microenvironment cells have intensive crosstalk, it is more rational to consider them all as a whole.

The development of next generation sequencing provides opportunities to systemically explore the tumor microenvironment. As RNA sequencing of tumor tissue usually contains microenvironment cells, researchers have developed some expression profile-based estimation of the abundance of microenvironment cells in tumor tissue (21-24). Furthermore, researchers have analyzed the connection between DNA-level alternations and immune filtration to explore genomic alterations that drive the low immune infiltration (25-27). However, owing to small sample size and the lack of multi-omics data, few studies conducted genomic analysis of TNBC from the perspective of immunology.

Collectively, we questioned whether TNBC has heterogeneous microenvironment phenotypes and what genomic events drive the formation of these phenotypes. With multi-omics data for the largest single-center TNBC cohort, we successfully classified 386 TNBC samples into three microenvironment clusters with distinct potential immune escape
mechanisms and genomic drivers.
Materials and Methods

Tumor and normal samples and data sets

We retrospectively selected 386 consecutive TNBC patients who underwent surgeries at the Department of Breast Surgery, Fudan University, Shanghai Cancer Center (FUSCC), from 1 January 2007 to 31 December 2014. Detailed inclusion criteria for the 386 samples were as follows: 1) female patients; 2) unilateral invasive ductal carcinoma; 3) pathological examination of the ER, PR and HER2 status performed by the Department of Pathology at FUSCC through immunochemical analysis and in situ hybridization (for HER2 status only); 4) patients with no evidence of metastasis at the time of diagnosis; and 5) sufficient frozen tissue for further research. In summary, RNA-seq data (tumor tissues: n = 245; paired normal tissues: n = 90), HTA 2.0 microarray data (n = 141), whole exome sequencing data (n = 268), OncoScan microarray copy number data (n = 335), H.E. sections data (n = 300) and tissue microarray data (n = 181) were obtained. In addition, we invited two pathologists to evaluate the stromal and intratumoral tumor-infiltrating lymphocytes (sTILs and iTILs, respectively) and fibrosis and necrosis in H.E. pathological sections based on established guidelines (28, 29). The studies were conducted in accordance with the Declaration of Helsinki. All the tissue samples included in the present study were obtained with approval from the independent ethics committee/institutional review board at Fudan University Shanghai Cancer Center Ethical Committee, and each patient provided written informed consent. The follow-up of our cohort was completed on 30 June 2017. The median follow-up length was 47 months (interquartile range, 29.3-73.5 months). The events included in the relapse-free survival (RFS) analysis were defined as the first recurrence of locally, regionally, or distantly invasive disease, a diagnosis of contralateral breast cancer, or death from any cause. Patients without events were censored at the last follow-up.

Detailed information regarding the sample processing and sequencing data generation is provided in the supplementary materials and methods.

Calculation of microenvironment cell abundance
To construct a compendium of microenvironment genes related to specific microenvironment cell subsets, we considered three aspects in the gene selection: first, the gene is specifically expressed in one specific microenvironment cell subset; second, other normal tissues do not express the gene; third, the compendium contains major cell types of the TNBC microenvironment. After researching papers, we modified two gene signatures, CIBERSORT (21) and MCP-Counter (23), to construct our compendium. First, we filtered the gene list of CIBERSORT (see their supplement 2). A 1.5-fold change cutoff from the highest expression cell type to the next highest expression cell type was considered as a criterion for cell type-specific expression. As a result, 324 genes were selected. As CIBERSORT do not contain signatures of fibroblasts and endothelial cells, which were important component of TNBC microenvironment, we added extra 40 genes for these cells (32 for endothelial cells and 8 for fibroblasts) from MCP-Counter gene list to our compendium. In all, our TNBC compendium contained 364 genes representing 24 microenvironment cell types (Supplementary Table 1 and Supplementary Figure 1). Subsequently, we used single sample gene set enrichment analysis (ssGSEA, “GSVA” function in R) to calculate the abundance of each cell subset in each sample with expression data. Similarly, we also referred to another published paper (22) to construct immune cell signatures for calculating the abundance of types 1, 2 and 17 T cells (Th1, Th2 and Th17) and myeloid-derived suppressor cells (MDSCs).

**Calculation of signature score distributions of microenvironment cell subsets**

To describe the constituent pattern (i.e. relative cell proportions) of microenvironment cell subsets within clusters, we first used the tumor purity, which was calculated by ASCAT (30), to adjust the enrichment scores of each microenvironment cell subset. The adjusted enrichment score was calculated as the enrichment score divided by (1 - tumor purity). We then illustrated the enrichment score distributions of each microenvironment cell subset to represent the relative cell proportions (“density” function in R).

**Microenvironment phenotypes clustering**

We performed k-means (“kmeans” function in R) clustering and Nbclust testing
(“NbClust” function in R, index = “all”) to determine the optimal number of stable TNBC microenvironment subtypes. To cluster samples based on the constituent pattern of each microenvironment cell type, we scaled each sample before clustering. Silhouette analysis was performed to confirm the stability of the clustering. For heatmap plotting (“pheatmap” function in R), we utilized the k-means clustering result to reorder the samples and scaled the original ssGSEA results before plotting.

**Prognostic analysis of microenvironment phenotypes**

We developed the univariate and multivariate Cox proportional hazards model to analyze the prognostic significance of microenvironment phenotypes. Age, tumor size, the number of positive lymph nodes and PAM50 subtypes were first analyzed in univariate Cox proportional hazards model. All significant variables were then included as covariates in the multivariate Cox proportional hazards model. We further validated the prognostic value of microenvironment clusters by comparing the predictive efficacy between two models: one consisting of tumor size and the number of positive lymph nodes as covariates; and the other adding microenvironment phenotypes as the third covariates. The area under the curve (AUC) of time-dependent ROC curves (“timeROC” package in R) was set as the indicator of prognostic efficacy. Statistical tests of the difference in AUC between the two models were conducted in the first three years of follow-up (“compare” function in R). We also evaluated the prognostic value of each cell subsets in the whole cohort and within each microenvironment cluster. In each analysis process, the abundance of cells was divided into high or low categories according to the optimal cutoff (“cutp” function in R). A univariate Cox proportional hazards model was then performed for the categorical variables of cell abundance.

**Calculation of immunogenomic indicators**

The detailed calculation of several bioinformatic indicators, such as neoantigens, cancer testis antigens (CTA), homologous recombination deficiency (HRD) scores and intratumoral heterogeneity (ITH), were described in supplementary materials and methods.

**Comparison of enriched oncogenic pathways**
We referred to a published paper (31) to establish a signature of ten oncogenic pathways containing 256 genes. We then used the ssGSEA method on these gene sets to generate enrichment scores for each pathway in each sample. Subsequently, we compared the ssGSEA score of each pathway among the three clusters.

**Comparison of mutations among clusters**

Genes with mutation frequencies greater than 2.5% were included in our comparison. To assess cluster-specific mutated genes, clusters were modeled (using logistic regression) as a function of the gene’s mutational status (ignoring silent events) and mutation load. The latter one was included to diminish confounding effects. P value less than 0.05 after adjusting for mutation load was considered significant. Since logistic regression model only receipt dependent variable with binomial distribution, we performed comparison in every two clusters. We conducted permutation test (“glmperm” function in R program) in logistic analysis. Cluster-specific mutations were defined as significantly mutated genes from the comparison of the given cluster with one of the other two clusters (Supplementary Table 6, column 13).

**Comparison of somatic copy number variations (SCNVs) among clusters**

The ASCAT algorithm was used to adjust the copy number of genes based on ploidy and purity. Samples were not included in further analysis when ASCAT-evaluated ACF score equaled 1 or if ASCAT failed to evaluate the purity. A regression approach similar to mutation analysis was then used to test for CNV associations. To test a given gene, clusters were modeled as a function of the gene’s copy number (categorized by the median value) and SCNV load. In order to explicitly define amplification and deletion events, we applied this linear regression approach twice. The first run was amplification-centric, and thus, the negative copy number values were set to zero. The second run was deletion-centric, and thus, the positive copy number values were set to zero. The cluster-specific amplified “peak” was defined as a continuous stretch of genes (arranged in genomic order) with a nominal P value less than 0.05 and log2 ratio greater than log2(2.5/2). Similarly, the cluster-specific deleted
“peak” was defined as a continuous stretch of genes (arranged in genomic order) with a nominal P value less than 0.05 and a log2 ratio less than log2(1.5/2). The comparison methods mainly referred to some previous immunological papers (25, 26) and the cutoff values, log2(2.5/2) and log2(1.5/2), were also obtained from one published high-quality paper of this field (32). Cluster-specific peaks were defined as significant peaks in each comparison between the given cluster and every one of the other two clusters.

Statistical analysis

Student’s t test, Wilcoxon’s test and Kruskal-Wallis test were utilized to compare continuous variables and ordered categorical variables, such as mutation load, neoantigen load, HRD score, CTAs number and ITH. Prior to the comparisons, the normality of the distributions was tested with Shapiro-Wilk test before comparison. Pearson’s chi-square test or Fisher’s exact test were employed for the comparison of unordered categorical variables. Permutation test was conducted in the comparison of gene mutation frequencies among clusters. Correlation matrices were created with Pearson’s or Spearman’s correlation. Survival analysis was performed using the Kaplan-Meier method, and the survival of the clusters was compared using the log rank test. All the tests were two sided, and $P < 0.05$ was regarded as indicating significance, unless otherwise stated. The false discovery rate (FDR) correction was utilized in multiple tests to decrease false positive rates. All of the analyses were performed with R software (version 3.4.2, http://www.R-project.org).
Results

Landscape of the microenvironment phenotypes in TNBC

We first established a reference microenvironment compendium that included 364 genes representing 24 microenvironment cell subsets, to systematically characterize the microenvironment phenotypes of TNBC (Materials and Methods, Supplementary Figure 1 and Supplementary Table 1). We then estimated the abundance of 24 microenvironment cell subsets in each sample and confirmed the accuracy of our results by comparing them with those of other microenvironment signatures (Supplementary Table 2). Subsequently, we performed k-means clustering of the TNBC microenvironment phenotypes. All 386 TNBC microenvironment phenotypes were classified into three heterogeneous clusters (Figure 1A). Our analysis revealed that three was the optimal and stable clustering number (Supplementary Figure 2). Cluster 1, the “immune-desert” cluster (type 1 “cold tumor”), was characterized by relatively low microenvironment cells infiltration, and thus, we selected a “red light” for this cluster, which represented low immune infiltration. Cluster 2, the “innate immune-inactivated” cluster (type 2 “cold tumor”), was characterized by the infiltration of inactivated innate immune cells, fibroblasts and endothelial cells, and thus, we chose a “yellow light” to indicate moderate immune infiltration. Cluster 3, the “immune-inflamed” cluster (“hot tumor”), was characterized by relatively high innate and adaptive immune cells infiltration, and we therefore selected a “green light” for this cluster to denote abundant immune infiltration. Moreover, we described the distributions of the abundance of microenvironment cell subsets to analyze the relative proportions of cell subsets within clusters. Adaptive immune cells and inactivated innate immune cells represented the major proportions in the microenvironments of the three clusters. The relative weights of innate immune cells and nonimmune cells were increased in cluster 2, whereas the relative proportion of adaptive immune cells was increased in cluster 3 (Figure 1B).

Validation of TNBC microenvironment clustering

To validate the expression profile-based clustering, we first evaluated the stromal and
intratumoral tumor-infiltrating lymphocytes (sTILs and iTILs) and CD8$^+$ cells in pathological sections among the clusters. Cluster 3 had significantly higher sTILs (mean, 19.86 vs 18.06 vs 32.65, P < 0.001), iTILs (mean, 5.27 vs 4.58 vs 11.02, P < 0.001) and CD8$^+$ cells (mean 13.36% vs 12.61 vs 25.02, P < 0.001) than the other two clusters (Figure 1C). We further analyzed other immune signatures to validate our microenvironment clustering. The distributions of types 1, 2 and 17 T helper cells (Th1, Th2 and Th17) and myeloid-derived suppressor cells (MDSCs) were also coincident with our established microenvironment clusters (Figure 1D). Moreover, we compared the distributions of the TNBC intrinsic subtypes among clusters. Cluster 1 mainly consisted of the classical basal-like and immune-suppressed (BLIS) subtype; cluster 2 was mainly composed of tumors with mesenchymal stem-like (MSL) features, which were mostly classified as the non-basal subtype in the PAM50 clustering; and clusters 3 was primarily made up of the immunomodulatory (IM) subtype, a non-classical basal-like subtype.

The luminal androgen receptor (LAR) TNBC subtype was dispersed among all three clusters (Figure 1E). Furthermore, we utilized the expression profiles of the METABRIC and TCGA cohorts to externally validate the repeatability of our clustering result (Supplementary Figure 3 and 4). Overall, we demonstrated that TNBC had three heterogeneous microenvironment phenotypes that could not be fully explained by mRNA-based TNBC intrinsic subtyping. The features of the three clusters are summarized in Figure 1F.

**Prognostic significance of microenvironment cells in TNBC**

Considering the important role of the tumor microenvironment in prognosis, we investigated the clinical relevance of the microenvironment clusters. The three clusters had similar clinicopathological characteristics (Supplementary Table 3). Cluster 3 had significantly better relapse free survival (RFS) (log rank P = 0.04) and overall survival (OS) (log rank P = 0.04) than the other two clusters (Figure 2A). The multivariate Cox proportional hazards model also revealed that cluster 3 independently predicted better RFS in TNBC (HR = 0.45, 95% CI: 0.21-0.97, P = 0.04, Figure 2B). The time-dependent AUC demonstrated that the addition of microenvironment clusters into the Cox proportional hazards model significantly increased the prognostic efficacy of one- (AUC: 0.72 vs 0.81, P = 0.0037) and
two-year (AUC: 0.74 vs 0.81, P < 0.001) recurrence (Figure 2C). Furthermore, we explored the prognostic significance of each cell subset (Figure 2D and Supplementary Table 4). In the whole cohort, a higher infiltration of immune cells, even the infiltration of immune suppressive cells, predicted better prognosis. However, a within-clusters analysis revealed that the prognostic significance of the cell subsets was diverse, even opposite, among the three clusters. Similar results were exhibited in METABRIC cohort (Supplementary Figure 3).

Potential extrinsic immune escape mechanisms of TNBC

The heterogeneity of the TNBC microenvironment phenotypes led to the question of whether different clusters of TNBC had distinct tumor immune escape mechanisms. Based on the immunoediting theory of previous publications (33), we first researched the extrinsic immune escape mechanism. This concept indicates that microenvironment components other than tumor cells contribute to the immune escape of tumor cells. The extrinsic immune escape mechanism consists of four major aspects: lack of immune cells, presence of immunoinhibitory cells (such as type 2 macrophages [M2Φ] and regulatory T-cells [Treg]), high concentrations of immunoinhibitory cytokines (such as IL10 and TGF-β), and fibrosis (34).

We compared the estimated number of microenvironment cells between the tumor site and a paired normal site (Figure 3A). Cluster 1 had almost no more microenvironment cells in the tumor site than in the paired normal site, suggesting an inability to attract innate immune cells (resulting in failure of adaptive immunity). The comparison of the expression of STING, the factor of spontaneous initiation of innate immunity, and other molecules potentially involved in initiation of innate immunity among the three clusters also suggested this theory (Figure 3C and Supplementary Figure 3F, 4D and 5) (35). Cluster 2 had more resting and activated innate immune cells in the tumor site, suggesting chemotaxis but inactivation of innate immune cells (also resulting in failure of adaptive immunity). Cluster 3 had not only abundant active innate and adaptive immune cells, but also immunosuppressive cells, such as Treg, in the tumor site, suggesting a role of immunosuppressive cells in immune escape. The expression of chemokines was consistent with these results (Figure 3B). Clusters 2 and 3 had...
higher expression of chemokines, including CCL4, CXCL9 and CXCL10, which have been proved to attract dendritic cells (DCs) and CD8^+ T cells. In addition, cluster 3 had both higher secreted immunostimulatory and immunoinhibitory cytokines; cluster 2 had increased secretion of immunoinhibitory cytokines only; and these cytokines were all relatively low in cluster 1. Notably, the difference in the expression level of these cytokines was not derived from SCNVs (all P > 0.05, Figure 3B). Furthermore, the fibrosis existed in all three clusters, but it did not differ significantly among them (P = 0.22, Figure 3D and 3E). Overall, the inability to attract innate immune cells, the chemotaxis but inactivation of innate immunity, the increase of immunoinhibitory factors after immune stimulation might contribute to the extrinsic immune escape of the three clusters, respectively.

**Tumor immunogenicity of TNBC**

We further investigated the potential intrinsic immune escape mechanisms of TNBC. Intrinsic immune escape indicates that tumor cells directly mediate their own immune escape. There are at least two aspects of intrinsic immune escape: tumor immunogenicity and immune checkpoint molecules expression (33).

We first compared some potential factors determining tumor immunogenicity among the three clusters: mutation load (Figure 4A), neoantigen load (Figure 4B and Supplementary Figure 6), chromosomal instability level (Figure 4C and Supplementary Figure 7), cancer-testis antigen (CTA) level (Figure 4D and Supplementary Figure 8), necrosis level (Figure 4E), intratumoral heterogeneity (ITH) (Figure 4F) and tumor antigen-presenting capability (Figure 4G, left). The first five factors were the main source of tumor antigens. In general, the difference in the tumor antigen burden among the three clusters of TNBC was not as large as that between microsatellite-instable (MSI) and microsatellite-stable (MSS) colorectal cancers. Cluster 1 (highest mutation load, homologous recombination deficiency (HRD) score, CTA burden and necrosis level, all P < 0.05) and cluster 2 (lowest mutation load, HRD score and CTA load, all P < 0.05) of TNBC had relatively high and low tumor antigen burdens, respectively. Moreover, the two clusters both had lower MHC I-related antigen-presenting molecules expression than cluster 3 (all P < 0.001), contributing to their...
low immunogenicity. The difference in the expression of antigen-presenting molecules among clusters was also not explained by SCNVs (all P > 0.05, Figure 4G, bottom left). In addition, the three clusters had no significant differences in ITH, although cluster 3 showed a non-significant tendency toward having lower ITH (mean fraction of subclonal mutations, 20.27% vs 21.26% vs 15.97%, P = 0.096, Figure 4F). Overall, the difference in tumor immunogenicity among the TNBC clusters might be relatively small, with clusters 1 and 2 having relatively low immunogenicity.

Regulation of immunomodulators in TNBC

The expression of immune checkpoint molecules after immune stimulation is another potential important intrinsic immune escape mechanism. Therefore, we referred to a database of co-stimulatory and co-inhibitory molecules (https://www.rndsystems.com/cn/research-area/co-stimulatory-and-co-inhibitory-molecules) to compare these immunomodulators among clusters. We demonstrated that cluster 3 had a higher expression of co-stimulating (most P < 0.05) and immune checkpoint molecules (all P < 0.05) than the other clusters (Figure 4G, right). This result suggested that cluster 3 tumors expressed immune checkpoint molecules to avoid immune killing after immune stimulation.

The high expression of one co-stimulator in cluster 3, TNFSF8 (mean log2 fold change relative to the paired normal tissue, 0.05 vs 1.41 vs 1.70, P < 0.001), could be explained by SCNVs (mean log2 ratio of somatic copy numbers, -0.25 vs -0.14 vs 0.03, P = 0.002, Figure 4I). In addition, we noticed that a co-inhibitor, VTCN1 (B7-H7), exhibited higher expression in cluster 1 (mean log2 fold change relative to the paired normal tissue, 0.57 vs -1.10 vs -0.67, P = 0.014) and was negatively correlated with immune infiltration (Spearman’s correlation = -0.27, P = 0.039, Figure 4H). Furthermore, we investigated the relationship among immune infiltration (TILs and cytolytic activity [CYT] (25)), immunogenicity and expression of immune checkpoint molecules. We demonstrated that immune infiltration and the expression of most checkpoint molecules were positively correlated, whereas the mutation load, neoantigen load, HRD score, necrosis and ITH seemed not to be correlated with these factors (Figure 4J).
Correlation of genomic alterations with low immune infiltration in TNBC

We further investigated the genomic alterations that could be correlated with the low immune infiltration in clusters 1 and 2. We aimed to identify some potential targets to reverse the absence of adaptive immune infiltration in these clusters.

We first referred to published signatures (Supplementary Table 5) to calculate the enrichment scores of ten common oncogenic pathways among the three clusters (31). The Hippo, MYC and Wnt pathways had higher scores in cluster 1 (all P < 0.001); the Notch, TGF-β and RTK/RAS pathways were enriched in cluster 2 (all P < 0.01); and the scores of the cell cycle, NRF2, PI3K and TP53 pathways were higher in cluster 3 (all P < 0.01, Figure 5A). Gene set enrichment analysis (GSEA) validated some of these results (Supplementary Figure 9). When selecting cluster-specific mutated genes (Materials and Methods, Figure 5B and Supplementary Table 6), we found that mutations among the PI3K-AKT pathway members (PI3KCA [30.8%], AKT1 [7.7%], PIK3R1 [6.1%] and PKD1 [6.1%]) were most frequently detected in cluster 2. After adjusting for the mutation load, PI3KCA and AKTI mutations remained significantly increased in this cluster (all adjusted P < 0.05). Our cluster-specific SCNVs analysis (Materials and Methods, Figure 5C and Supplementary Table 7) demonstrated that the amplification of 8q24.13-8q24.3 (MYC) was more frequent in cluster 1 (mean log2 ratio of somatic copy numbers, 3.99 vs 3.06 vs 3.19, adjusted P < 0.05) consistent with the results of the expression profile analysis. Other cluster 1-specific amplifications included 1p34.2, 3q26.1-3q26.2, 12p11.22 and 18q11.2-18q12.1. 1p36.22-1p36.21 (TNFRSF8, TNFRSF1B). 19p13.3 were cluster 1-specific deletions. Moreover, the amplification of 20p13-20p11.21 and the deletion of 9p24.3-9p22.3 (CD274, PDCD1LG2) were specific to cluster 2. We annotated genes located in copy number peaks with the GO and KEGG pathways. Notably, genes located in the cluster 2-specific amplification peaks were annotated to various pathways, including the innate immune response (Supplementary Table 8). Overall, our analysis revealed that some genomic alterations might drive the low immune infiltration in clusters 1 and cluster 2.
Using multi-omics data from the largest single-center TNBC cohort, our study revealed three heterogeneous TNBC microenvironment phenotypes and their clinical significance. Although clusters 1 and cluster 2 were both so-called “cold tumors”, they had distinct microenvironment phenotypes. Cluster 3 was the so-called “hot tumor”. We emphasized the following characteristics of the clusters that might contributed to immune escape: defects in the attraction of innate immune cells in cluster 1, chemotaxis but inactivation of innate immune cells and low tumor antigen burden in cluster 2, and high expression of immune checkpoint molecules in cluster 3. Furthermore, we found some genomic alterations that might drive the low immune infiltration of clusters 1 and 2 (Table 1). To the best of our knowledge, this study constitutes the first systemic analysis of the microenvironment heterogeneity of TNBC. The clustering results are in accordance with the immunological principles described by previous papers (36).

Our study has important implications for clinical translations. First, our results might facilitate the selection of suitable patients for ICIs treatment. We revealed a cluster of “hot tumors” in TNBC (cluster 3, approximately 28% of all TNBCs) and demonstrated that high expression of immune checkpoint molecules might lead to the immune escape of this cluster. Although most clinical trials have shown that the efficacy of ICIs in TNBC was less than 10% (10, 11), these patients usually received several lines of treatment without detection of the real status of immune checkpoint proteins before immunotherapy. Notably, the efficacy of PD-L1 inhibitors in first-line monotherapy could be up to 25% (10), which was consistent with the percentage of “hot tumors” (28%) in our research. We hypothesized that some TNBC cells, which were sensitive to ICIs, were also more sensitive to chemotherapy. As a result, after several lines of chemotherapy, ICI-sensitive cells were eliminated. The level of immune checkpoint proteins after surgery could not represent this fact after several lines of chemotherapy. Considering the more targeted efficacy of ICIs than chemotherapy, we speculated that ICIs might need earlier lines application in “hot tumors” of TNBC. Second,
we identified potential indicators to predict ICI efficacy in TNBC. Previous studies have suggested that the tumor mutation burden, checkpoint molecules expression and TILs score might individually and collectively predict ICI efficacy in other tumor types (37). However, we illustrated that the tumor mutation burden of TNBC was not associated with the TIL score or checkpoint molecules expression, whereas the latter two were highly correlated (Figure 4J), which suggests that the latter two, but not the mutation load, might be ideal predictors.

Recently, IMpassion130, the first phase III clinical trial of ICI in metastatic TNBC, also demonstrated that PD-L1 might be a suitable biomarker to predict ICI efficacy in TNBC (12). Moreover, the non-significant association between T cell infiltration and TMB (or neoantigen burden) in TNBC has been reported in previous papers (25, 26). Several factors in addition to TMB, such as local tumor microenvironment, host immune state and neoantigen quality, might influence the T cell infiltration in the TME as well (38-40).

Our study could also help facilitate research on the relationship between oncogenic pathways and immune infiltration in TNBC. The transformation from a “cold tumor” to a “hot tumor” is currently a popular topic in cancer research. Recent studies have suggested that the activation of oncogenic pathways might decrease immune infiltration (41). However, few studies have focused on the mechanisms of impaired immune infiltration in TNBC. Our study demonstrated that at least two phenotypes of “cold tumors” exist in TNBC. Cluster 1 was characterized by almost no immune cell infiltration and a high percentage of MYC amplification. Previous studies have indicated that MYC amplification could induce the expression of CCL5, CCL23, IL1β, CD47 and PD-L1, inactivate DCs and macrophages; and limit NK, T, and B cells recruitment (42-44). Therefore, we speculated that MYC-induced low innate immune cells chemotaxis might be the reason for the poor immune infiltration in cluster 1. In addition, the features of cluster 2 were chemotaxis but inactivation of innate immunity and higher fibroblast and endothelial cells infiltration, with a higher mutation frequency in PI3K-AKT pathway members. The TGF-β-related and cancer-associated fibroblasts (CAF)-related pathways were enriched in cluster 2. Therefore, we speculated that hyper-activated PI3K-AKT pathway in tumor cells might play a major role in suppressing
immune infiltration in cluster 2, such as through TGF-β secretion and CAF differentiation. In articular, previous studies have revealed that pan-PI3K inhibition enhances anti-tumor immunity and anti-PD1 response in breast cancer (45).

Our research had some limitations. First, there are more than 24 stromal cells types, and it is difficult to precisely include all phenotypes. We classified these cell subsets into three categories, namely, adaptive and activated innate immune cells, inactivated innate immune cells and non-immune cells, which solved this problem to some extent. In addition, immunogenomic analysis could not reflect the cause and consequence effect. Potential driver molecules in our research, such as MYC and the PI3K-AKT pathway, require further functional validation. Experimental studies are ongoing in our laboratory.

In conclusion, our study revealed that the microenvironment phenotypes of TNBC could be classified into three heterogeneous clusters with distinct potential immune escape mechanisms. Specific oncogenic pathways might drive the formation of these microenvironment phenotypes.
Acknowledgements

The authors wish to thank Prof. McGranahan for providing us with R code to calculate ITH score in TNBC. This work was supported by the National Natural Science Foundation of China (81502278, 81572583, 81372848, 81370075, 31720103909, 31671368, and 31671380), the Training Plan of Excellent Talents in Shanghai Municipality Health System (2017YQ038), the “Chen Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (17CG01), the Training Plan of Excellent Talents of Fudan University Shanghai Cancer Center (YJYQ201602), the Municipal Project for Developing Emerging and Frontier Technology in Shanghai Hospitals (SHDC12010116), the Cooperation Project of Conquering Major Diseases in Shanghai Municipality Health System (2013ZYJB0302), the Innovation Team of Ministry of Education (IRT1223), the Shanghai Key Laboratory of Breast Cancer (12DZ2260100), the National Key R&D Project of China (2016YFC0901704, 2017YFC0907502 and 2017YFC0907503), and the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01, 18JC1420100 and 15411953300). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conception and design: Zhi-Ming Shao and Yi-Zhou Jiang

Development of methodology: Yi Xiao, Ding Ma, Shen Zhao and Yi-Zhou Jiang

Acquisition of data: Yi Xiao, Ding Ma, Chen Suo, Jinxiu Shi and Mengzhu Xue and Miao
Analysis and interpretation of data: Yi Xiao, Ding Ma, Shen Zhao, Hai Wang, Jingjing Zhao and Qin Li

Writing, review, and/or revision of the manuscript: Yi Xiao, Ding Ma, Shen Zhao, Hai Wang, François BERTUCCI and Yi-Zhou Jiang

Study supervision: Peng Wang, Leming Shi, Wentao Yang, Wei Huang, Xin Hu, Ke-Da Yu, Shenglin Huang, François BERTUCCI, Yi-Zhou Jiang and Zhi-Ming Shao

Data availability

All data can be viewed in The National Omics Data Encyclopedia (NODE) (http://www.biosino.org/node) by pasting the accession (OEP000155) into the text search box or through the URL: http://www.biosino.org/node/project/detail/OEP000155. The sequencing data is also available in GSE118527 (OncoScan), GSE76250 (HTA 2.0) and SRP157974 (WES and RNAseq).
References

15. Su S, Liu Q, Chen J, Chen J, Chen F, He C, et al. A positive feedback loop between
mesenchymal-like cancer cells and macrophages is essential to breast cancer metastasis.


Table 1. Summary of immune escape mechanisms and cluster-specific genomic alterations of triple-negative breast cancer

<table>
<thead>
<tr>
<th>Category</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrinsic immune escape mechanisms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unable to attract innate immune cells</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>unable to attract adaptive immune cells</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>attract immunosuppressive cells</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>fibrosis</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>low expression of immunostimulating cytokines</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>high expression of immunoinhibiting cytokines</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Intrinsic immune escape mechanisms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low mutation load</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>low neoantigen load</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>low cancer testis antigen load</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>low somatic copy number variations load</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>low necrosis</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>low MHC I associated antigen presenting</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>low MHC II associated antigen presenting</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>high intratumoral heterogeneity</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>high expression of checkpoint molecules</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Cluster-specific DNA-level alterations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutations</td>
<td>TTN/TNR/PKHD1L1/SPTA1/NCKAP5/COL15A1/ANKRD11/MYLK</td>
<td>PIK3CA/NF1/akt1/FBN3/ABCC1/DNHD1</td>
<td>IGsf10/DNAH1/CDH23/AHNAK2/GTF3C1</td>
</tr>
<tr>
<td>somatic copy number gains/amplifications</td>
<td>1p34.2/3q26.1-3q</td>
<td>20p13-20p11.21</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>26.2/8q24.13-8q2</td>
<td>4.3/12p11.22/18q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.2-18q12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>somatic copy number loss/deletions</td>
<td>1p36.22-1p36.21/19p13.3</td>
<td>9p24.3-9p22.3</td>
<td>-</td>
</tr>
</tbody>
</table>
**Figure Legends**

**Figure 1. Landscape of the microenvironment phenotypes in triple-negative breast cancer**

(A) K-means clustering of triple-negative breast cancer (TNBC) microenvironment phenotypes based on the estimated numbers of 24 microenvironment cell subsets calculated by single sample gene set enrichment analysis (ssGSEA). (B) Signature scores distributions of four cell subsets among the three clusters after adjustment for tumor purity. (C) Pathological scores of stromal and intratumoral tumor infiltrating lymphocytes and immunohistochemistry (IHC) scores of CD8 cells among clusters. (D) Signature scores of types 1, 2 and 17 T helper cells (Th1, Th2 and Th17) and myeloid-derived suppressor cells (MDSCs) among clusters. In the violin plots, the mean values are plotted as red dots, and the boxplot is drawn inside of the violin plot. (E) Distribution of TNBC intrinsic subtypes (Lehmann TNBC subtypes, Fudan University Shanghai Cancer Center [FUSCC] TNBC subtypes and PAM50 subtypes) among the clusters. (F) Summary of characteristics of each microenvironment cluster. *** P < 0.001; ** 0.001 < P < 0.01; * 0.01 < P < 0.05; ns P > 0.05.

Abbreviations: TMA, tissue microarray; BL1, basal-like 1; BL2, basal-like 2; M, mesenchymal; MSL, mesenchymal stem-like; UNS, unstable; LAR, luminal androgen receptor; IM, immunomodulatory; BLIS, basal-like and immune-suppression; MES, mesenchymal-like; M1 and M2, type 1 and 2 macrophages, respectively.

**Figure 2. Prognostic significance of microenvironment cells in triple-negative breast cancer**

(A) Kaplan–Meier curves of relapse free survival (RFS) and overall survival (OS) among the
clusters. (B) Hazard ratios (HRs) and P values of the covariates in the univariate and multivariate Cox proportional hazards model for RFS. (C) Area under the curve (AUC) of the time-dependent ROC curve with two Cox proportional hazards models for RFS; one included two covariates (tumor size and number of positive lymph nodes), and the other added microenvironment clusters as covariates. The significant difference in the AUC was estimated at one, two and three years. (D) Estimation of the prognostic value of each cell subset by using a univariate Cox proportional hazards model for RFS in all cohorts and each microenvironment cluster. The color represents the hazard ratio, and the size of the circles represents -log10 (FDR). Larger circles represent smaller FDR values. *** P  < 0.001; ** 0.001 < P < 0.01; * 0.01 < P < 0.05. ns P > 0.05.

Abbreviations: CI: confidence interval.

Figure 3. Potential extrinsic immune escape mechanisms of triple-negative breast cancer

(A) Volcano plots of enriched (red) and depleted (blue) microenvironment cell subsets compared with the paired normal samples (n = 90) for each cluster. (B) Log2-fold change in mRNA expression in the tumor site relative to the paired normal tissue and log2 ratio of copy number values of chemokines, interleukins, interferons and other important cytokines and their receptors for each cluster. Molecules with significantly differential expression between the tumor site and the paired normal site (P < 0.01) were illustrated. (C) STING mRNA expression among clusters. (D-E) Typical H.E. sections of fibrosis (D) and distributions of fibrosis levels among clusters (E). *** P < 0.001; ** 0.001 < P < 0.01; * 0.01 < P < 0.05; ns P > 0.05.

Figure 4. Potential intrinsic immune escape mechanisms of triple-negative breast cancer
(A-F) Comparison of mutation loads (A), neoantigen load (B), homologous recombination deficiency (HRD) scores (C), cancer testis antigen (CTA) numbers (D), necrosis (E) and intratumoral heterogeneity (ITH) scores (F) among the three clusters. In the violin plots, the mean values are plotted as red dots, and the boxplot was drawn inside the violin plot. (G) Comparison of the log2-fold changes in mRNA expression at the tumor sites relative to the paired normal tissue and log2 ratio of the copy number values of the MHC molecules, co-stimulators and co-inhibitors for each cluster. For co-stimulators and co-inhibitors, only molecules having significantly differential expression between the tumor site and the paired normal site (P < 0.01) were illustrated. (H) Negative correlation between VTCN1 mRNA expression and cytolytic toxicity (CYT). (I) Distributions of somatic copy number variations (SCNVs) categories of CD274 and TNFSF8. (J) Correlations between tumor immunogenicity indicators, immune infiltration and expression of immune checkpoint molecules. *** P < 0.001; ** 0.001 < P < 0.01; * 0.01 < P < 0.05; ns P > 0.05.

Abbreviations: sTILs and iTILs: stromal and intratumoral tumor infiltrating lymphocytes, respectively.

Figure 5. Correlation of genomic alterations with low immune infiltration in TNBC

(A) Heatmap of normalized enrichment scores of ten oncogenic pathways among clusters. (B) Genes showing a significant difference (adjusted P < 0.05) in the comparison between every two clusters among the three clusters for non-silent mutations are illustrated. The color represents the cluster in which the gene had the highest mutation frequency, and the color saturation represents the mutation frequency. (C) Comparison of the somatic copy number variations (SCNVs) between every two clusters among the three clusters. The upper plot
illustrates the frequency of the amplification (dark red), gain (light red), loss (light blue) and deletion (dark blue) of each gene in each cluster, and the lower plot illustrates the -log10 adjusted P value of each gene when compared between every two clusters among the three clusters in the amplification-centric (light yellow) or deletion-centric (light green) calculations.
Figure 2

A

B

Univariate and multivariate Cox proportional hazards model for RFS

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.98 (0.96-1.01)</td>
<td>0.164</td>
</tr>
<tr>
<td>Tumor size (cm)</td>
<td>1.35 (1.11-1.65)</td>
<td>0.0034</td>
</tr>
<tr>
<td>Positive lymph nodes (number)</td>
<td>1.11 (1.06-1.15)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PAM50 subtypes</td>
<td>non basal 1.30 (0.73-2.3)</td>
<td>0.378</td>
</tr>
<tr>
<td>Microenvironment</td>
<td>cluster 1 0.96 (0.52-1.77)</td>
<td>0.894</td>
</tr>
<tr>
<td>cell subtype</td>
<td>cluster 2 0.48 (0.23-1.01)</td>
<td>0.0425</td>
</tr>
</tbody>
</table>

C

Time dependent ROC curve

D

Cluster analysis

- CD8+ T cells
- Regulatory T cells
- Naive CD4+ T cells
- Follicular helper T cells
- Naive B cells
- Memory B cells
- γδ T cells
- Activated dendritic cells
- M1 type Macrophages
- Activated NK cells
- Plasma cells
- Resting memory CD4+ T cells
- Activated memory CD4+ T cells
- Activated mast cells
- Resting NK cells
- M0 type Macrophages
- Monocytes
- Eosinophils
- M2 type Macrophages
- Resting mast cells
- Resting dendritic cells
- Neutrophils
- Endothelial cells
- Fibroblasts

Legend:
- High infiltration suggests a better prognosis
- High infiltration suggests a worse prognosis

FDR

- >0.1 (n.s.)
- 0.1-0.01 Significant
- 0.01-0.001 Highly significant
Clinical Cancer Research

Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer

Yi Xiao, Ding Ma, Shen Zhao, et al.

Clin Cancer Res Published OnlineFirst March 5, 2019.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-18-3524

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2019/03/05/1078-0432.CCR-18-3524.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/early/2019/03/05/1078-0432.CCR-18-3524. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.