Title: Integrated Analysis of Single-Cell and Bulk RNA-Seq Data reveals that Ferroptosis-Related Genes Mediated the Tumor Microenvironment predicts Prognosis, and guides Drug Selection in Triple-Negative Breast Cancer

We kindly request the opportunity to deposit our work as a preprint on bioRxiv or medRxiv, depending on its suitability.

Xuantong Gong[1#], Lishuang Gu[2#], Di Yang[1], Yu He[1], Qian Li[3*], Hao Qin[4*], Yong Wang[1*]

First authors: Xuantong Gong and Lishuang Gu contributed equally to this work.

Corresponding authors: Correspondence to Yong Wang, Hao Qin or Qian Li

[1]Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Department of Ultrasound, Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing, China

[3]Department of Ultrasound, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China

[4]Key Laboratory of Cancer and Microbiome, State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Email: Xuantong Gong: gongxuantong@163.com

Lishuang Gu: 13026300487@163.com

Di Yang: yangdi4134@163.com

Yu He: 1282302858@qq.com

*: Correspondence to Yong Wang, Hao Qin, Qian Li

Yong Wang

Department of Ultrasound, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China.

Email address: drwangyong77@163.com

Hao Qin

Key Laboratory of Cancer and Microbiome, State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Email address: haoqin@cicams.ac.cn

Qian Li

Department of Ultrasound, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China

Email address: 754427296@qq.com

Declarations

•Ethics approval and consent to participate: Not applicable

•Consent for publication: Not applicable

•Availability of datasets：The data used in this study are available in the public datasets. Nine single-cell RNA-seq TNBC samples were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi). The data for external validation were obtained from the GEO database. The data of 471 genes associated with ferroptosis were collected

from the FerrDb website (http://www.zhounan.org/ferrdb/current/).

•Competing Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

•Authors' contributions: YW：Conceptualization, Methodology, Formal analysis, Investigation, Writing-original draft preparation, Writing-review & editing, Project administration and Funding acquisition. XTG: Conceptualization, Methodology, Formal analysis, Investigation, Writing-original draft preparation, Writing-review & editing, Resources and Data curation. LSG: Conceptualization, Methodology, Formal analysis, Investigation, Writing-original draft preparation, Writing-review & editing, Resources and Data curation. DY: Writing-original draft preparation and Writing-review & editing. YH: Writing-original draft preparation and Writing-review & editing. QL: Conceptualization, Methodology, Formal analysis, Investigation, Writing-original draft preparation, Writing-review & editing, Visualization and Supervision. HQ: Conceptualization, Methodology, Formal analysis, Investigation, Writing-original draft preparation, Writing-review & editing, Visualization and Supervision. All authors: reading and approval of the final manuscript.

•Impact Statement: Integrating single-cell and bulk RNA-seq data elucidates the role of ferroptosis in TNBC, offering a prognostic model and personalized therapeutic insights.

Key words: RNA-seq; ferroptosis; tumor microenvironment; triple-negative breast cancer

Abstract

Background

TNBC is aggressive, lacking methods to predict recurrence and drug sensitivity. Ferroptotic heterogeneity varies in TNBC subtypes. However, the TME mediated by ferroptosis genes is unclear. Our study aims to integrate single-cell and bulk RNA-seq data to reveal the ferroptosis-mediated TME in TNBC, predicting prognosis and guiding treatment.

Methods

The single-cell RNA-seq (scRNA-seq) and bulk RNA-seq data of TNBC were sourced from the Gene Expression Omnibus (GEO) database. Using these data, a machine learning algorithm was employed to integrate and analyze the characteristics of the TME mediated by ferroptosis-related genes in TNBC. Prediction models for TNBC survival prognosis and drug treatment response were established and then validated in an independent set.

Results

At the individual cell level, T cells were categorized into three distinct subpopulations, and local macrophages into two subpopulations. The infiltration degree of these different cell subpopulations was closely associated with prognosis and treatment outcomes. Based on this, the risk score model we developed effectively predicted recurrence-free survival in TNBC patients, with independently validated pooled predicted 3-, 4-, and 5-year Area Under the Curves(AUCs) of 0.65, 0.67, and 0.71, respectively. Additionally, we found that patients in the high-risk group may be more responsive to 27 drugs.

Conclusions

We have uncovered the tumor immune cell clusters in TNBC mediated by ferroptosis. A risk score model was constructed to identify high-risk TNBC patients, which can assist physicians in disease monitoring and precision therapy. The genes identified hold significant potential as therapeutic targets for TNBC patients.

Funding

This project is funded by the National Natural Science Foundation of China (81974268, 82304151), the Talent Incentive Program of Cancer Hospital Chinese, Academy of Medical Sciences (801032247), the Cancer Hospital of Chinese Academy of Medical Sciences-Shenzhen Hospital Cooperation Fund (CFA202202023), and the open project of Beijing Key Laboratory of Tumor Invasion and Metastasis Mechanism, Capital Medical University(2023ZLKF03).

1. Background

Global cancer statistics show that breast cancer is the most common cancer in women and the leading cause of cancer deaths [1]. Triple-negative breast cancer (TNBC) is a subtype of breast cancer that lacks expression of estrogen receptor, progesterone receptor and human epidermal growth factor receptor-2. Compared with other subtypes, TNBC is highly aggressive, with a poor overall prognosis for patients and a median survival of less than 1 year after recurrent metastasis [2-5]. Due to the lack of targets for endocrine therapy and targeted therapeutic agents, chemotherapy and emerging immunotherapy are the main therapies. TNBC is highly heterogeneous and only

109 some patients benefit from treatment [6-8]. Unfortunately, there is no effective method that can
110 effectively predict the prognostic risk and drug sensitivity of TNBC, which is an urgent problem in
111 the clinic.
112
113 TME is composed of various cell types such as cancer-associated fibroblasts, tumor-associated
114 macrophages, T cells, NK cells, B cells, endothelial cells, etc., which interact with cancer cells and
115 influence various aspects such as tumor progression, metastasis, and response to therapy [9-11]. The
116 composition and functional status of the tumor microenvironment (TME) varies greatly among
117 different patients with breast cancer, and revealing the TME of each patient is crucial for selecting
118 a reasonable treatment and controlling tumor progression in the long term [12-13].
119
120 Ferroptosis is an iron-dependent form of non-apoptotic, oxidative form of regulated cell death
121 involving lipid hydroperoxides [14- 15]. Studies have shown that drug-resistant cancer cells are more
122 sensitive to ferroptosis [16-18]. Therefore, ferroptosis is more recognized as a potential target for
123 cancer therapy. Ferroptotic heterogeneity exists in different subtypes of TNBC [19]. Since
124 ferroptosis is regulated by multiple metabolic pathways, the ferroptosis landscape of TNBC
125 remains unexplored and its relationship with patient prognosis and treatment response is uncertain
126 [20]. Therefore, exploring ferroptosis-related genes in TNBC and their correlation with the immune
127 microenvironment may provide a new treatment trend for TNBC.
128
129 In this work, we collected and integrated data from single-cell and bulk RNA sequencing (RNA-
130 seq) and bulk RNA-seq to characterize the ferroptosis-related gene mediated TME landscape of

TNBC on multiple scales. furthermore, based on the findings in TME, predictive models of patient prognosis and treatment response were constructed using machine learning algorithms. We hope to provide a new way of individualized risk management for triple-negative breast cancer patients and assist their precision treatment.

2. Methods

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
software, algorithm	R package Seurat	Microsoft	N/A	Version 4.1.1

2.1 Patient data collection and processing

The data used in this study were collected from public datasets. Nine single-cell RNA-seq TNBC samples were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) under the accession number GSE176078 (Table 1).

Table 1 Sample information for triple-negative breast cancer in GSE176078

Case ID	Gender	Age	T_Stage	Subtype by IHC
3946	**Female**	**52**	**T2**	**TNBC**
44041	**Female**	**35**	**T2**	**TNBC**
4465	**Female**	**54**	**T2**	**TNBC**
4495	**Female**	**63**	**T1**	**TNBC**
44971	**Female**	**49**	**T2**	**TNBC**
44991	**Female**	**47**	**NA**	**TNBC**
4513	**Female**	**73**	**T3**	**TNBC**
4515	**Female**	**67**	**T1**	**TNBC**
4523	**Female**	**52**	**T2**	**TNBC**

To construct the model to predict prognosis and immunotherapy efficacy, bulk RNA sequencing datasets and clinicopathological information of TNBC were obtained from the GEO database under the accession number GSE25066. The data for external validation were obtained from the GEO database under the accession number GSE86166.

2.2 Single-cell RNA-seq data processing

Data were processed using the R package Seurat (version 4.1.1) for quality control (QC). Based on QC metrics suggested in Scanpy tutorial, cells with less than 200 genes expressed or more than 20% mitochondrial genes counts were filtered out. After quality control, the expression data of each cell was normalized separately using the NormalizeData function and the top 2000 highly variable genes (HVGs) of each cell were identified using the FindVariableFeatures function. To remove batch effects among different samples, we utilized the FindIntegrationAnchors function to find anchor genes and then the reclassified datasets were integrated utilized the IntegrateData function. Data normalization was performed using the ScaleData function in Seurat. The RunPCA function was utilized to reduce the dimension of principal component analysis (PCA) for the first 2000 HVGs screened above.

2.3 Genes associated with ferroptosis acquiring

The data of 471 genes associated with ferroptosis were collected from the FerrDb website (http://www.zhounan.org/ferrdb/current/). Then, from the above genes, screened for genes with

expression information in the single-cell matrix of TNBC patients.

2.4 Cell subtype identification associated with ferroptosis

Because information of cell type annotation was listed in the original literature of the GSE176078 dataset, the metadata of the original literature was used directly for the next analyses. The R package Seurat (version 4.1.1) was used to identify cell types and presented results via t-SNE. Non-negative matrix decomposition (NMF) is an algorithm based on high-throughput data to identify and cluster out different molecular functional patterns [21]. In this work, subtypes of macrophages and T cells based on genes associated with ferroptosis using the "NMF" R package (v 0.26).

2.5 Cell-cell communication analysis

Cell-cell communication mediated by ligand-receptor complexes plays an important role in multiple biological processes [22]. In this study, the "iTALK" R package was utilized to construct the cellular communication network.

2.6 Pseudotime analysis

Pseudotime analysis with scRNA-seq data helps to obtain an approximate landscape of gene expression dynamics [23]. The "Monocle" R package (v 2.28.0) was applied for pseudotime analysis to conduct cellular trajectory. The reduceDimensio function based on the DDRTree algorithm is used to reduce the dimensions of the data. Based on the Progenitor Cell Biology

Consortium (PCBC) database (https://www.synapse.org), the stemness signature was identified via the one-class logistic regression (OCLR) algorithm, then the stemness index of each TNBC cell was calculated by scaling the Spearman correlation coefficients to be between 0 and 1. Eventually, the order Cells function was utilized to sort cells and complete construction of trajectory.

2.7 Transcriptional factor analysis

The Single-cell Regulatory Network Inference and Clustering tool (SCENIC) enables simultaneous gene regulatory network reconstruction and cell-state identification from scRNA-seq data [24-25]. We utilized the "SCENIC" R package (v1.3.1) to establish the transcription factor (TF) regulatory network. Specifically, first, coexpression modules are inferred using the "GENIE3/GRNBOOST" function to identify gene set with co-expressed TFs. Next, the indirect targets are pruned from these modules using cis-regulatory motif discovery (cisTarget). Finally, the AUCell algorithm was utilized to evaluate the activity of regulons.

2.8 Hallmarks gene set enrichment analysis

Hallmarks gene set enrichment analysis was utilized the "irGSEA" R package (v2.1.5). The ssGSEA algorithm was utilized to conduct differential pathway score between ferroptosis-related immune cell subtypes.

2.9 Construction and validation of the prognostic model

First, based on the recurrence-free survival (RFS) data of 178 TNBC patients in the GSE25066 database, this study utilized the "survival" (v3.2-7) and "survminer" (v0.4.8) R package to perform univariate Cox proportional hazards regression analyses on ferroptosis-related genes

signature of different immune cell subtypes.

The genes screened for significant association with RFS in the regression analysis ($p<0.01$) were utilized to construct a risk factor-based model by the Least Absolute Shrinkage and Selection Operator (LASSO) method implemented in the "glmnet" R package (v4.0-2). TNBC patients were divided into low- and high-risk groups according to the median risk score, and Kaplan-Meier survival curves were utilized to compare the RFS rates between the two groups, p value < 0.05 was considered to significance. To validate the performance of the model, receiver operating characteristic (ROC) curves were demonstrated and the AUC values were calculated for evaluating 3-year, 4-year, and 5-year RFS rates of TNBC patients in GSE25066 database. In addition, the GSE86166 as well as the TCGA database were utilized as external validation sets to further verify the robustness of the performance of the constructed model in this study.

2.10 Immune cells infiltration and drug sensitivity analysis

Immune cells infiltration is important to the anti-tumor response, and these cells are diverse among patients [26-27]. The "CIBERSORT", "GSVA" and "TIMER" R package was utilized to determine the distribution of different immune cell types between low- and high-risk groups. Next, the 50% inhibitory concentration (IC50) of 138 chemotherapeutic drugs was calculated for each patient using the "pRRophetic" R package (v 0.5), and drugs significantly associated with the risk score were screened. Finally, we calculated the tumor immune dysfunction and exclusion (TIDE) score (http://tide.dfci.harvard.edu.) for each patient and thus analyzed its difference between low- and high-risk groups.

2.11 Statistical analysis

The R software (version 4.1.3) was utilized for all data analysis. Wilcoxon-rank sum test was utilized to analyze associations of continuous variables. Log-rank test was utilized to analyze differences in survival curves between groups. P value less than 0.05 was considered statistically significant.

3. Results

3.1 The single-cell landscape of TNBC samples

In this study, a total of 9 TNBC single-cell samples were included. After QC, 38985 genes from 29733 cells per sample were finally selected for subsequent analysis. Based on original annotations of data, these cells were divided into 9 major cell clusters and 29 minor cell clusters (Figure 1a-c). Major cell clusters include B-cells, Cancer-associated Fibroblasts (CAFs), Cancer Epithelial, Endothelial, Myeloid, Normal Epithelial, Plasmablasts, Perivascular-like Cells (PVL), T-cells. Of these, Myeloid cluster is further divided into Cycling_Myeloid, Dendritic Cells (DCs), Macrophage, and Monocyte clusters. As shown in Figure 1C, the immune microenvironment of TNBC is characterized by a high proportion of T cells and macrophages. On this basis, ligand-receptor interactions between different major cell clusters are frequent (Figure 1d).

Of the 471 ferroptosis-related genes from the FerrDb website, 391 were present in the single-cell expression matrix of TNBC. We found that there were significant differences in the average expression of these genes in different clinicopathological classifications, as shown in Figure 1e. Moreover, there were significant differences in the expression of ferroptosis-related genes in 12 major cell clusters of TNBC (Figure 1f).

Fig. 1 Integration and clustering of scRNA-Seq data from triple-negative breast cancer. (a) t-SNE plot of the 9 major cell clusters. (b) t-SNE plot of the 29 minor cell clusters. (c) t-SNE plot of the 12 major cell clusters. (d) The number of ligand-receptor interactions of different major cell clusters in cell-cell communication network, different colors represent different cell clusters and arrows represent ligand-receptor orientation. (e) Heat map showing the average expression of ferroptosis-related genes in different clinicopathological classifications. (f) Heat map showing the average expression of ferroptosis-related genes in different major cell clusters.

3.2 Ferroptosis-related subpopulations of T cells in the TNBC

A total of 11,784 T cells in the scRNA-seq data of this study, including cycling T cells, NK cells, NKT cells, T cells CD4+, and T cells CD8+ identified clusters (Figure 2a). Based on the NMF algorithm (rank=3), each T cell cluster was further categorized into 3 ferroptosis-related subpopulations (T_C1, T_C2, T_C3). Figure 2b shows that differentially expressed top 50 ferroptosis-related genes differed significantly among these 3 subpopulations. Among them, the proportion of NK cells in the T_C2 subpopulation was significantly more than the other two subpopulations (Figure 2c). According to pseudotime trajectories, we found that all three T cell subpopulations were involved at different periods of differentiation (Figure 2d-f). Further, Figure 2g shows the average expression of signature genes associated with 8 functions in 15 subpopulations of T-cells, and we found that gene expression was significantly higher in the Cycling T-cells. Moreover, different T-cell subpopulations play an important role in the anti-tumor process (Figure 2h).

TFs are key regulators in cellular signal transduction [28]. As a result, RFX5, EOMES, TBX21, CEBPB, RUNX3 and IKZF3 showed higher activities in NKT cells. Besides, IRF, NFATC2, STAT and PRDM1 showed higher activities in $CD8^+$ T cells (Figure 2i). The HALLMARK analysis results revealed a prominent enrichment in pathways such as pathways in DNA−Repair and

adipogenesis in cycling T cells, whereas in other clusters of T cells, pathways such as epithelial-mesenchymal transition and KRAS signaling had higher enrichment. However, the difference in enrichment scores between ferroptosis-related subpopulations of T cells was not significant.

Fig. 2 Ferroptosis-related subpopulations of T cells in triple-negative breast cancer. (a) t-SNE plot of the 5 identified clusters of T cells. (b) Heat map showing the expression of ferroptosis-related genes (Top 50) in 3 subpopulations of T cells. (c) Demonstration of the proportion of clusters of T cells in ferroptosis-related subpopulations. (d) Pseudotime trajectories showing the developmental time course of T cells. (e) Pseudotime trajectories of the 5 identified clusters of T cells. (f) Pseudotime trajectories of ferroptosis-related subpopulations of T cells. (g) Heat map showing the average expression of signature genes associated with 8 functions in 15 subpopulations of T-cells. (h) Network diagram of the cell-cell communication between cancer epithelial cells and T cells. (i) Heat map for differential analysis of transcription factors activity.

3.3 Ferroptosis-related subpopulations of macrophages in the TNBC

A total of 3671 macrophages in the scRNA-seq data of this study. Based on the NMF algorithm (rank=4), 2 ferroptosis-related subpopulations of macrophages (M_C1, M_C2) were finally obtained after clustering cell clusters with similar markers, the t-SNE plot was shown in Figure 3a. The pseudotime trajectories showed that M_C2 cells belong to the early stage of macrophages and then gradually develop into M_C1 cells (Figure 3b-c). Figure 3d shows that differentially expressed top 50 ferroptosis-related genes differed significantly among subpopulations.

Further, we found that FOLR2, SEPP1, MRC1, LYVE1, SLC40A1, and CD163 were highly expressed in M_C1 cells and named M_C1 as FOLR2+MAC. whereas, TREM2, FN1, CXCR4, C3, S100A8, IFI6, and SPP1 were highly expressed in M_C2 and named M_C2 as TREM2+MAC. Figure 3e-f shows the marker genes in each subpopulation of macrophages. Enrichment scores of the 2 subpopulations in each cell are shown by t-SNE plots(Figure 3g-h).

Moreover, the enrichment score of TREM2+MAC and FOLR2+MAC were significantly negatively correlated (r=-0.779) (Figure 3i). Communication between these two subpopulations and cancer epithelial cells is vigorous (Figure 3j). In addition, detection of TFs activity in each macrophage cell revealed that most of TFs had higher activity in FOLR2+ MAC cells (Supplementary Figure 4).

Fig. 3 Ferroptosis-related subpopulations of macrophages in triple-negative breast cancer. (a) t-SNE plot of 2 ferroptosis-related subpopulations of macrophages. (b) Pseudotime trajectories showing the developmental time course of macrophages. (c) Pseudotime trajectories of 2 ferroptosis-related subpopulations of macrophages. (d) Heat map showing the average expression of ferroptosis-related genes in 2 subpopulations of macrophages. (e) t-SNE plot showing the expression patterns of marker genes in M_C2. (f) t-SNE plot showing the expression patterns of marker genes in M_C1. (g) t-SNE plot showing the enrichment score of the TREM2+MAC in each cell. (h) t-SNE plot showing the enrichment score of the FOLR2+MAC in each cell. (i) Bubble plot showing the correlation of enrichment score between TREM2+MAC and FOLR2+MAC. (j) Network diagram of the cell-cell communication between cancer epithelial cells and ferroptosis-related subpopulations of macrophages.

3.4 Construction and validation of predictive survival models based on ferroptosis-related genes

Based on the RFS data of TNBC in the GSE25066 database, univariate Cox regression analysis was performed on 8371 specific markers of ferroptosis-related subpopulations, and 41 genes significantly associated with RFS were screened ($p < 0.01$) (Supplementary Table 4, Supplementary Figure 6). These genes were further analyzed by Least Absolute Shrinkage and Selection Operator(LASSO) regression, and 23 genes were finally selected for the construction of the risk model. We calculate the risk score using the following formula: risk score = TMEM160 * (-0.4689) + EWSR1 * (-0.4237) + BCAT2 * (-0.2346) + PNKP * (-0.1592) + MLEC * (-0.1567) + SAP30BP * (-0.1469) + TBL1XR1 * (-0.0894) + STAG1 * (-0.0695) + NR4A1 * (-0.0630) + TNFRSF9 + (-0.0503) + PSD3 * (-0.0153) + BAD * (-0.0094) + MIIP * 0.0104 + HIST3H2A *

0.0294 + CDC25B * 0.0528 + TCEB1 * 0.0670 + HMGCS1 * 0.0988 + SPC25 * 0.1317 + TKT * 0.2234 + PTTG1 * 0.2960 + ADA * 0.3097 + AK1 * 0.4003 + AIMP2 * 0.4081 (Figure 4a-c).All patients were categorized into low- and high-risk groups according to the median value of the risk score (Figure 4d-f). Survival curves showed that patients in the high-risk group had a shorter RFS compared to patients in the low-risk group ($p<0.05$, Figure 4g). In addition, the risk score had an excellent predictive effect in predicting RFS in TNBC patients, its AUCs for 3-, 4-, and 5 years RFS were 0.87, 0.88, and 0.88 (Figure 4h).

In external validation set (GSE86166), patients in the high-risk group also had a shorter RFS compared to patients in the low-risk group ($p<0.05$). This risk model also performed well in predicting RFS in TNBC patients, its AUCs for 3-, 4-, and 5 years RFS were 0.65, 0.67, and 0.71 (Figure 4i).

Fig. 4 Prognostic differences in triple-negative breast cancer with ferroptosis-related subpopulations. (a) LASSO regression of 23 ferroptosis-related genes. (b) Cross-validation for optimizing the parameter in LASSO regression. (c) Demonstration of regression coefficients corresponding to 23 genes. (d) Graph showing risk scores for all samples. (e) Scatterplot showing recurrence-free survival for all samples. (f) Heat map showing the average expression of 23 genes in the low- and high-risk groups. (g) Kaplan–Meier curves of survival analysis in the low- and high-risk groups. (h) Receiver operating characteristic curves for predicting the recurrence-free survival at 3, 4 and 5 years in training set. (i) Receiver operating characteristic curves for predicting the recurrence-free survival at 3, 4 and 5 years in external validation set.

3.5 Analysis of independent prognostic factors

Further, we performed univariate and multivariate Cox analyses to determine whether the risk score could serve as an independent prognostic factor for TNBC patients compared to other common clinicopathologic factors. In the GSE25066 database, Figure 5a shows that both the risk

factor and stage were significantly associated with TNBC patients' RFS and were independent prognostic factors ($p<0.05$). In the external validation set, Figure 5b shows that only the risk factor can be considered as independent prognostic factor in TNBC patients ($p<0.05$).

Fig. 5 Analysis of independent prognostic factors for triple-negative breast cancer patients. (a) The forest plot showing the results of univariate and multivariate COX regression analysis of risk score, age, grade, and stage in the GSE25066 database. (b) The forest plot showing the results of univariate and multivariate COX regression analysis of risk score, grade, and stage in the GSE86166 database.

3.6 Correlation of the risk score with immune microenvironment

Next, we calculated the ImmuneScore, StromalScore, ESTIMATEScore, and TumorPurity for samples in the low- and high-risk groups and compared the differences in scores between the two groups. The results showed that these scores did not differ significantly between the low- and high-risk groups (Figure 6a-f). Then, based on the CIBERSORT algorithm, we found that T cells CD4 memory activated, NK cells resting, and Monocytes had significantly higher proportions in the high-risk group, and the NK cells activated, and mast cells resting had significantly higher proportions in the low-risk group (Figure 6g). Infiltration of these immune cells may play an important role in the clinical course of TNBC patients.

*Fig. 6 Analysis of immune microenvironment in low- and high-risk groups. (a-d) Correlation of the risk score with the ImmuneScore, StromalScore, ESTIMATEScore, and TumorPurity. (e) Difference of ImmuneScore, StromalScore, and ESTIMATEScore in low- and high-risk groups. (f) Difference of TumorPurity in low- and high-risk groups. (g) Difference of immune infiltration score between low- and high-risk groups calculated by CIBERSORT. * $P < 0.05$, ** $P < 0.01$, ns: $p>0.05$.*

3.7 Analysis of clinical response to drugs with the risk score

Based on the “pRRophetic” R package, we explored the relationship between the risk score and clinical response to 138 drugs and calculated IC50. As a result, 16 of 50 drugs whose clinical response was significantly associated with the risk score were positively associated with the risk score, and the remaining 34 were negatively associated with the risk score. Further, we found that 27 of drugs negatively associated with the risk score had significantly lower IC50 in the high-risk group, suggesting that patients in the high-risk group may be more sensitive to these drugs, favoring the choice of clinical medication (Figure 7a). And 13 of drugs positively associated with the risk score had significantly higher IC50 in the high-risk group than in the low-risk group, suggesting that patients in the high-risk group may be less sensitive to these drugs (Figure 7b). Finally, we calculated the TIDE scores of patients in the low- and high-risk groups; unfortunately, the TIDE scores were not significantly different between the two groups (Figure 7c).

Fig. 7 Analysis of immune microenvironment in low- and high-risk groups. (a) Box plot showing differences of IC50 for drugs negatively associated with risk scores in different groups. (b) Box plot showing differences of IC50 for drugs positively associated with risk scores in different groups. (c) Box plot showing differences of TIDE scores in different groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns: $p>0.05$.

Discussion

TNBC is the most difficult subtype of breast cancer to treat, and survival and efficacy prediction for each individual is a critical step in precision tumor therapy. Several studies have identified

405 ferroptosis-related genes as novel therapeutic targets to enhance treatment efficacy and improve
406 patient prognosis [29-32]. However, to the best of our knowledge, there is a lack of reports on the
407 association of ferroptosis-related genes with prognosis and clinical response to treatment in
408 TNBC. Here, firstly, we revealed ferroptosis-related immune cell clustering in TNBC patients at
409 the single-cell level, demonstrating the ferroptosis heterogeneity in TNBC, which was also
410 supported by other cohorts [19]. On the basis of the above results, we constructed an ferroptosis-
411 related risk score model based on bulk RNA-seq data, which successfully predicted the
412 recurrence-free survival of TNBC patients and validated the robustness of the results in various
413 datasets. This is important for the individualized management of TNBC patients. More
414 importantly, the model can also predict the sensitivity to drugs for TNBC patients with different
415 risk stratification to guide physicians in the selection of drugs for different patients, and it has
416 significant clinical application value.
417
418 Understanding the TME in TNBC is essential for improving prognosis and guiding treatment.
419 scRNA-seq enables transcriptomic analysis of individual cells to characterize the cellular diversity
420 in the TME in detail, thereby improving understanding of the disease [33-35]. In this study, our
421 analysis of scRNA-seq data from TNBC revealed that there are multiple immune cells infiltrating
422 in the TME and interacting frequently with each other, and they may impede or promote tumor
423 progression [36]. Since, we found that the average expression of ferroptosis-related genes was
424 significantly different in different T-stages of tumors, which might be strongly correlated with
425 patient prognosis. Further, based on ferroptosis-related genes, a higher proportion of T cells and
426 macrophages in the TME were categorized into different subpopulations to reveal the

427 heterogeneity of ferroptosis-related in TNBC. The T_C2 subpopulation was accompanied by a
428 greater infiltration of NK cells.NK cells have rapid and efficient anti-tumor immunity and their
429 activity is negatively correlated with breast cancer progression [37-39]. The relevant results were
430 validated in the dataset GSE25066. FOLR2, SEPP1, MRC1, LYVE1, SLC40A1 and CD163 were
431 highly expressed in the M_C1 subpopulation. Studies have shown that these markers, which are
432 normally present in mammary macrophages from healthy humans and mice, correlate with a
433 favorable prognosis in breast cancer, with FOLR2 being associated with anti-tumor immunity
434 players of $CD8^+$ T cells, DCs, B cells, and tertiary lymphoid structures [40-42]. In the M_C2
435 subpopulation, TREM2, FN1, C3, CXCR4, and SPP1 are highly expressed. It has been shown that
436 these markers are lowly expressed in macrophages of healthy breast tissues and highly expressed
437 in tumor tissues [42-44]. In addition, S100A8 and IFI6 were also highly expressed in the M_C2
438 subpopulation. Studies have reported that a high percentage of $S100A8^+$ myeloid cell infiltration
439 has been suggested as a potential mechanism for worsening the prognosis of TNBC [45-46].
440 Furthermore, IFI6 is an interferon-stimulated gene with anti-apoptotic and metastasis-promoting
441 effects [47]. These findings demonstrate the heterogeneity of the function of iron death-associated
442 macrophage subpopulations, which may play different roles in the anti-tumor process. The
443 different distributions of T cell and macrophage subtypes in different TNBC patients in this study
444 revealed that ferroptosis-related TME in TNBC patients may be valuable for patient survival and
445 outcome prediction. In the future, altering the degree of macrophage or T-cell infiltration in
446 different subpopulations could be potentially valuable in improving the prognosis of TNBC
447 patients.
448

449 Intratumor heterogeneity is one of the most important factors contributing to poor clinical
450 outcomes in breast cancer, and ferroptosis is one of the major pathways of activation in highly
451 heterogeneous tumors [48-50]. In order to better address the problem of accurately predicting the
452 prognosis of TNBC, we firstly performed the correlation analysis of ferroptosis-related genes with
453 RFS and constructed a risk score model in the publicly available dataset of bulk RNA-seq. TNBC
454 patients were categorized into low- and high-risk groups based on the risk score. In the high-risk
455 group, TNBC patients had higher expression levels of the MIIP, HIST3H2A, CDC25B, TCEB1,
456 HMGCS1, SPC25, TKT, PTTG1, ADA, AK1, and AIMP2 genes. Among them, the cell division
457 cycle 25 (CDC25) family of proteins is dual-specific tyrosine phosphatases responsible including
458 three subtypes, CDC25A, CDC25B and CDC25C, which are used to regulate cell cycle transitions
459 [51-52]. CDC25B dephosphorylates and activates cyclin-dependent kinase/cyclin complex
460 (CDK1/Cyclin B), which have been shown to be overexpressed in breast cancer and promote cell
461 cycle progression [52-53]. Spindle component 25 (SPC25) is a key component of the nuclear
462 division cycle 80 (NDC80) complex, and its high expression promotes tumor cell proliferation by
463 inducing mitotic disorder [54]. Study shows that SPC25 promotes the proliferation of breast cancer
464 cells, and high levels of SPC25 mRNA levels are associated with high recurrence rates and low
465 survival rates in breast cancer patients [55]. Pituitary tumor transforming gene 1 (PTTG1) has been
466 demonstrated to promote proliferation, migration and invasion of cancer cells and is a gene that
467 promotes breast cancer development [56]. At the mRNA level, PTTG1 is more present in patients
468 with high histologic grade and with lymph node metastasis, is significantly associated with low
469 patient survival, and is a biomarker suggestive of poor prognosis in breast cancer, which may be
470 related to the regulation of the cell cycle by PTTG1 to promote the development of breast cancer

471 cells [57]. In the low-risk group, TNBC patients had higher expression levels of the TMEM160,
472 EWSR1, BCAT2, PNKP, MLEC, SAP30BP, TBL1XR1, STAG1, NR4A1, TNFRSF9, PSD3, and
473 BAD genes. These genes have different roles in regulating the TME in colorectal cancers, bladder
474 cancers, lymphomas, etc., but they have been rarely reported in TNBC [58-60]. In future studies,
475 these genes are valuable to investigate in the regulation of anti-TNBC.
476
477 In TNBC, we found that the risk score model we constructed had independent predictive power
478 for RFS and the predictive performance of this score was stable in an external validation set.
479 Considering the impact of the tumor immune microenvironment on the prognosis of TNBC
480 patients, we further explored the differences in immune cell infiltration in different subgroups of
481 patients. The results showed that T cells CD4 memory activated, NK cells resting, and Monocytes
482 had higher proportions in the high-risk group. This is in line with previous studies, such as that
483 CXCL7 expressed and released by monocytes can stimulate cancer cell migration, invasion and
484 metastasis [61]. While in the low-risk group, the proportion of NK cells activated and mast cells
485 resting was higher. It has been suggested that the effect of mast cells on tumor invasiveness may
486 vary among different breast cancer subtypes [62-63]. Our findings suggest that infiltration of mast
487 cells resting in the TME is more favorable to the prognosis of patients.
488
489 This study has some limitations. First, this study is based on the analysis of transcriptomic data,
490 and the integration of multi-omics data is needed to improve the predictive performance of the
491 model. Second, this study is a retrospective analysis of a public dataset, and the results obtained
492 need to be further validated by multiple prospective trials.

Conclusions

In conclusion, this study revealed the TNBC ferroptosis-mediated tumor immune cell clustering based on scRNA-seq data, and the degree of infiltration of different subpopulations of cells had different roles in the prognosis of TNBC patients. On this basis, combined with bulk RNA-seq data, the survival prediction model for TNBC patients was established with excellent predictive performance and stability. The high-risk TNBC patients screened by this prediction model and their sensitive therapeutic drugs can help to guide physicians in disease monitoring and precise treatment. The related genes screened also provide important value for TNBC patients to find potential therapeutic targets to improve their prognosis.

Acknowledgements

Our thanks go to the Gene Expression Omnibus (GEO) database and FerrDb for providing essential data resources. We extend our appreciation to the participating institutions and our dedicated colleagues for their contributions and expertise. Special recognition is given to the patients and clinical staff for their invaluable participation and assistance. We also thank the reviewers for their insightful feedback.

List of abbreviations

Triple-negative breast cancer (TNBC); tumor microenvironment (TME); RNA sequencing (RNA-

seq); single-cell RNA-seq (scRNA-seq); Area Under the Curves(AUCs); Gene Expression Omnibus (GEO); The Cancer Genome Atlas (TCGA); quality control (QC); highly variable genes (HVGs); principal component analysis (PCA); t-distributed stochastic neighbor embedding (t-SNE); Non-negative matrix decomposition (NMF); Progenitor Cell Biology Consortium (PCBC); one-class logistic regression (OCLR); Single-cell Regulatory Network Inference and Clustering tool (SCENIC); transcription factor (TF); recurrence-free survival (RFS); Least Absolute Shrinkage and Selection Operator (LASSO); receiver operating characteristic (ROC); tumor immune dysfunction and exclusion (TIDE); Cancer-associated Fibroblasts (CAFs); Perivascular-like Cells (PVL); Dendritic Cells (DCs); Least Absolute Shrinkage and Selection Operator(LASSO); cell division cycle 25 (CDC25); cyclin-dependent kinase/cyclin complex (CDK1/Cyclin B); Spindle component 25 (SPC25); nuclear division cycle 80 (NDC80); Pituitary tumor transforming gene 1 (PTTG1).

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263.
2. Keenan TE, Tolaney SM. Role of Immunotherapy in Triple-Negative Breast Cancer. J Natl Compr Canc Netw. 2020;18(4):479-489.
3. Yam C, Yen EY, Chang JT, et al. Immune Phenotype and Response to Neoadjuvant Therapy in Triple-Negative Breast Cancer. Clin Cancer Res. 2021;27(19):5365-5375.

537 4. Howlader N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by
538 Molecular Subtypes in the United States. Cancer Epidemiol Biomarkers Prev.
539 2018;27(6):619-626.
540 5. Bardia A, Mayer IA, Vahdat LT, et al. Sacituzumab Govitecan-hziy in Refractory Metastatic
541 Triple-Negative Breast Cancer. N Engl J Med. 2019;380(8):741-751.
542 6. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical
543 benefit in breast cancer: the CTNeoBC pooled analysis [published correction appears in
544 Lancet. 2019 Mar 9;393(10175):986]. Lancet. 2014;384(9938):164-172.
545 7. Shepherd JH, Ballman K, Polley MC, et al. CALGB 40603 (Alliance): Long-Term Outcomes
546 and Genomic Correlates of Response and Survival After Neoadjuvant Chemotherapy With or
547 Without Carboplatin and Bevacizumab in Triple-Negative Breast Cancer. J Clin Oncol.
548 2022;40(12):1323-1334.
549 8. Kim C, Gao R, Sei E, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer
550 Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-893.e13.
551 9. Mayer S, Milo T, Isaacson A, et al. The tumor microenvironment shows a hierarchy of cell-
552 cell interactions dominated by fibroblasts. Nat Commun. 2023;14(1):5810.
553 10. Mao X, Xu J, Wang W, et al. Crosstalk between cancer-associated fibroblasts and immune
554 cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer.
555 2021;20(1):131.
556 11. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune
557 microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541-550.
558 12. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to

559 metastatic outgrowth. Cancer Cell. 2023;41(3):374-403.

560 13. Yang W, Xu M, Xu S, et al. Single-cell RNA reveals a tumorigenic microenvironment in the
561 interface zone of human breast tumors. Breast Cancer Res. 2023;25(1):100.

562 14. Fang X, Ardehali H, Min J, Wang F. The molecular and metabolic landscape of iron and
563 ferroptosis in cardiovascular disease. Nat Rev Cardiol. 2023;20(1):7-23.

564 15. Viswanathan VS, Ryan MJ, Dhruv HD, et al. Dependency of a therapy-resistant state of
565 cancer cells on a lipid peroxidase pathway. Nature. 2017;547(7664):453-457.

566 16. Kim R, Hashimoto A, Markosyan N, et al. Ferroptosis of tumour neutrophils causes immune
567 suppression in cancer. Nature. 2022;612(7939):338-346.

568 17. Hangauer MJ, Viswanathan VS, Ryan MJ, et al. Drug-tolerant persister cancer cells are
569 vulnerable to GPX4 inhibition. Nature. 2017;551(7679):247-250.

570 18. Tsoi J, Robert L, Paraiso K, et al. Multi-stage Differentiation Defines Melanoma Subtypes
571 with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. Cancer
572 Cell. 2018;33(5):890-904.e5.

573 19. Yang F, Xiao Y, Ding JH, et al. Ferroptosis heterogeneity in triple-negative breast cancer
574 reveals an innovative immunotherapy combination strategy. Cell Metab. 2023;35(1):84-
575 100.e8.

576 20. Lu YJ, Gong Y, Li WJ, Zhao CY, Guo F. The prognostic significance of a novel ferroptosis-
577 related gene model in breast cancer. Ann Transl Med. 2022;10(4):184.

578 21. Ding X, Qin J, Huang F, Feng F, Luo L. The combination of machine learning and untargeted
579 metabolomics identifies the lipid metabolism -related gene CH25H as a potential biomarker in
580 asthma. Inflamm Res. 2023;72(5):1099-1119.

581 22. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-
582 cell communication from combined expression of multi-subunit ligand-receptor complexes.
583 Nat Protoc. 2020;15(4):1484-1506.
584 23. Sugihara R, Kato Y, Mori T, Kawahara Y. Alignment of single-cell trajectory trees with
585 CAPITAL. Nat Commun. 2022;13(1):5972.
586 24. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network
587 inference and clustering. Nat Methods. 2017;14(11):1083-1086.
588 25. Van de Sande B, Flerin C, Davie K, et al. A scalable SCENIC workflow for single-cell gene
589 regulatory network analysis. Nat Protoc. 2020;15(7):2247-2276.
590 26. Iglesias-Escudero M, Arias-González N, Martínez-Cáceres E. Regulatory cells and the effect
591 of cancer immunotherapy. Mol Cancer. 2023;22(1):26. Published 2023 Feb 4.
592 27. Pérez-Romero K, Rodríguez RM, Amedei A, Barceló-Coblijn G, Lopez DH. Immune
593 Landscape in Tumor Microenvironment: Implications for Biomarker Development and
594 Immunotherapy. Int J Mol Sci. 2020;21(15):5521.
595 28. Gong TQ, Jiang YZ, Shao C, et al. Proteome-centric cross-omics characterization and
596 integrated network analyses of triple-negative breast cancer. Cell Rep. 2022;38(9):110460.
597 29. Li H, Yang P, Wang J, et al. HLF regulates ferroptosis, development and chemoresistance of
598 triple-negative breast cancer by activating tumor cell-macrophage crosstalk. J Hematol Oncol.
599 2022;15(1):2.
600 30. Zhang Z, Lu M, Chen C, et al. Holo-lactoferrin: the link between ferroptosis and radiotherapy
601 in triple-negative breast cancer. Theranostics. 2021;11(7):3167-3182. Published 2021 Jan 1.
602 31. Zhou TJ, Zhang MM, Liu DM, et al. Glutathione depletion and dihydroorotate dehydrogenase

603 inhibition actuated ferroptosis-augment to surmount triple-negative breast cancer.
604 Biomaterials.
605 32. Desterke C, Cosialls E, Xiang Y, et al. Adverse Crosstalk between Extracellular Matrix
606 Remodeling and Ferroptosis in Basal Breast Cancer. Cells. 2023;12(17):2176.
607 33. Li J, Hubisz MJ, Earlie EM, et al. Non-cell-autonomous cancer progression from
608 chromosomal instability. Nature. 2023;620(7976):1080-1088.
609 34. Van de Sande B, Lee JS, Mutasa-Gottgens E, et al. Applications of single-cell RNA
610 sequencing in drug discovery and development. Nat Rev Drug Discov. 2023;22(6):496-520.
611 35. Zhai Y, Zhang J, Huang Z, et al. Single-cell RNA sequencing integrated with bulk RNA
612 sequencing analysis reveals diagnostic and prognostic signatures and immunoinfiltration in
613 gastric cancer. Comput Biol Med. 2023;163:107239.
614 36. Deepak KGK, Vempati R, Nagaraju GP, et al. Tumor microenvironment: Challenges and
615 opportunities in targeting metastasis of triple negative breast cancer. Pharmacol Res.
616 2020;153:104683.
617 37. Gong Y, Chen W, Chen X, et al. An Injectable Epigenetic Autophagic Modulatory Hydrogel
618 for Boosting Umbilical Cord Blood NK Cell Therapy Prevents Postsurgical Relapse of Triple-
619 Negative Breast Cancer. Adv Sci (Weinh). 2022;9(23):e2201271.
620 38. Guillerey C, Huntington ND, Smyth MJ. Targeting natural killer cells in cancer
621 immunotherapy. Nat Immunol. 2016;17(9):1025-1036.
622 39. Bouzidi L, Triki H, Charfi S, et al. Prognostic Value of Natural Killer Cells Besides Tumor-
623 Infiltrating Lymphocytes in Breast Cancer Tissues. Clin Breast Cancer. 2021;21(6):e738-e747.
624 40. Jäppinen N, Félix I, Lokka E, et al. Fetal-derived macrophages dominate in adult mammary

625 glands. Nat Commun. 2019;10(1):281.

626 41. Wang Y, Chaffee TS, LaRue RS, et al. Tissue-resident macrophages promote extracellular
627 matrix homeostasis in the mammary gland stroma of nulliparous mice. Elife. 2020;9:e57438.

628 42. Nalio Ramos R, Missolo-Koussou Y, Gerber-Ferder Y, et al. Tissue-resident FOLR2+
629 macrophages associate with CD8+ T cell infiltration in human breast cancer. Cell.
630 2022;185(7):1189-1207.e25.

631 43. Molgora M, Esaulova E, Vermi W, et al. TREM2 Modulation Remodels the Tumor Myeloid
632 Landscape Enhancing Anti-PD-1 Immunotherapy. Cell. 2020;182(4):886-900.e17.

633 44. Müller A, Homey B, Soto H, et al. Involvement of chemokine receptors in breast cancer
634 metastasis. Nature. 2001;410(6824):50-56.

635 45. Drews-Elger K, Iorns E, Dias A, et al. Infiltrating S100A8+ myeloid cells promote metastatic
636 spread of human breast cancer and predict poor clinical outcome. Breast Cancer Res Treat.
637 2014;148(1):41-59.

638 46. Rigiracciolo DC, Nohata N, Lappano R, et al. Focal Adhesion Kinase (FAK)-Hippo/YAP
639 transduction signaling mediates the stimulatory effects exerted by S100A8/A9-RAGE system
640 in triple-negative breast cancer (TNBC). J Exp Clin Cancer Res. 2022;41(1):193.

641 47. Davenport AM, Morris M, Sabti F, et al. G1P3/IFI6, an interferon stimulated protein,
642 promotes the association of RAB5+ endosomes with mitochondria in breast cancer cells. Cell
643 Biol Int. 2023;47(11):1868-1879.

644 48. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer
645 heterogeneity. Nat Rev Clin Oncol. 2015;12(7):381-394.

646 49. Su GH, Xiao Y, You C, et al. Radiogenomic-based multiomic analysis reveals imaging

647 intratumor heterogeneity phenotypes and therapeutic targets. Sci Adv. 2023;9(40):eadf0837.

648 50. Wu SZ, Al-Eryani G, Roden DL, et al. A single-cell and spatially resolved atlas of human
649 breast cancers. Nat Genet. 2021;53(9):1334-1347.

650 51. Boutros R, Lobjois V, Ducommun B. CDC25 phosphatases in cancer cells: key players? Good
651 targets?. Nat Rev Cancer. 2007;7(7):495-507.

652 52. Cairns J, Ly RC, Niu N, et al. CDC25B partners with PP2A to induce AMPK activation and
653 tumor suppression in triple negative breast cancer. NAR Cancer. 2020;2(4):zcaa039.

654 53. Guo YQ, Wang Q, Wang JG, et al. METTL3 modulates m6A modification of CDC25B and
655 promotes head and neck squamous cell carcinoma malignant progression. Exp Hematol
656 Oncol. 2022;11(1):14.

657 54. Yang J, Huang Y, Song M, et al. SPC25 promotes proliferation and stemness of hepatocellular
658 carcinoma cells via the DNA-PK/AKT/Notch1 signaling pathway. Int J Biol Sci.
659 2022;18(14):5241-5259.

660 55. Wang Q, Zhu Y, Li Z, et al. Up-regulation of SPC25 promotes breast cancer. Aging (Albany
661 NY). 2019;11(15):5689-5704.

662 56. Qi L, Zhou B, Chen J, et al. Significant prognostic values of differentially expressed-
663 aberrantly methylated hub genes in breast cancer. J Cancer. 2019;10(26):6618-6634.

664 57. Meng C, Zou Y, Hong W, Bao C, Jia X. Estrogen-regulated PTTG1 promotes breast cancer
665 progression by regulating cyclin kinase expression. Mol Med. 2020;26(1):33.

666 58. Dai X, Wu Z, Ruan R, et al. TMEM160 promotes tumor immune evasion and radiotherapy
667 resistance via PD-L1 binding in colorectal cancer. Cell Commun Signal. 2024;22(1):168.

668 59. Lei MZ, Li XX, Zhang Y, et al. Acetylation promotes BCAT2 degradation to suppress BCAA

catabolism and pancreatic cancer growth. Signal Transduct Target Ther. 2020;5(1):70.

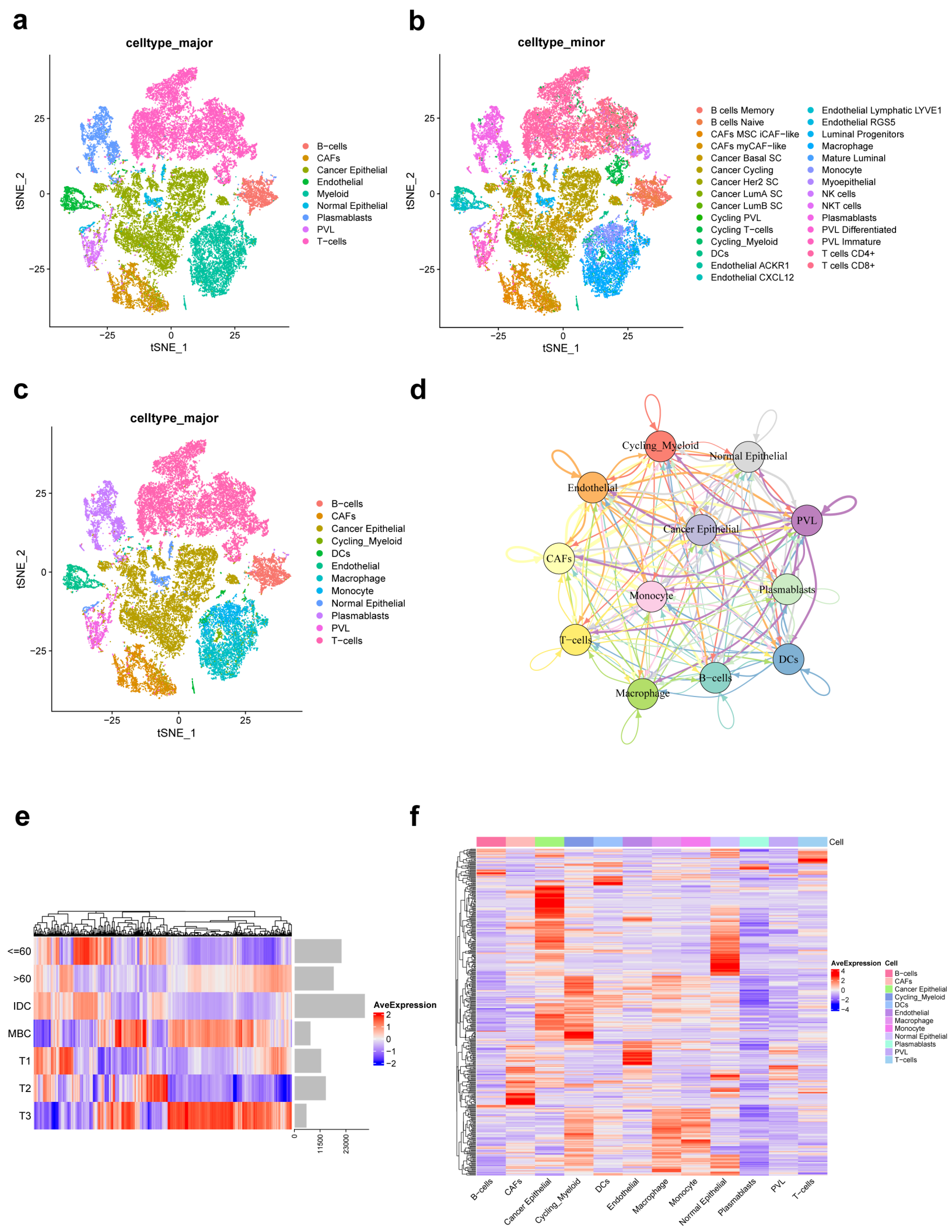
60. Venturutti L, Teater M, Zhai A, et al. TBL1XR1 Mutations Drive Extranodal Lymphoma by Inducing a Pro-tumorigenic Memory Fate. Cell. 2020;182(2):297-316.e27.

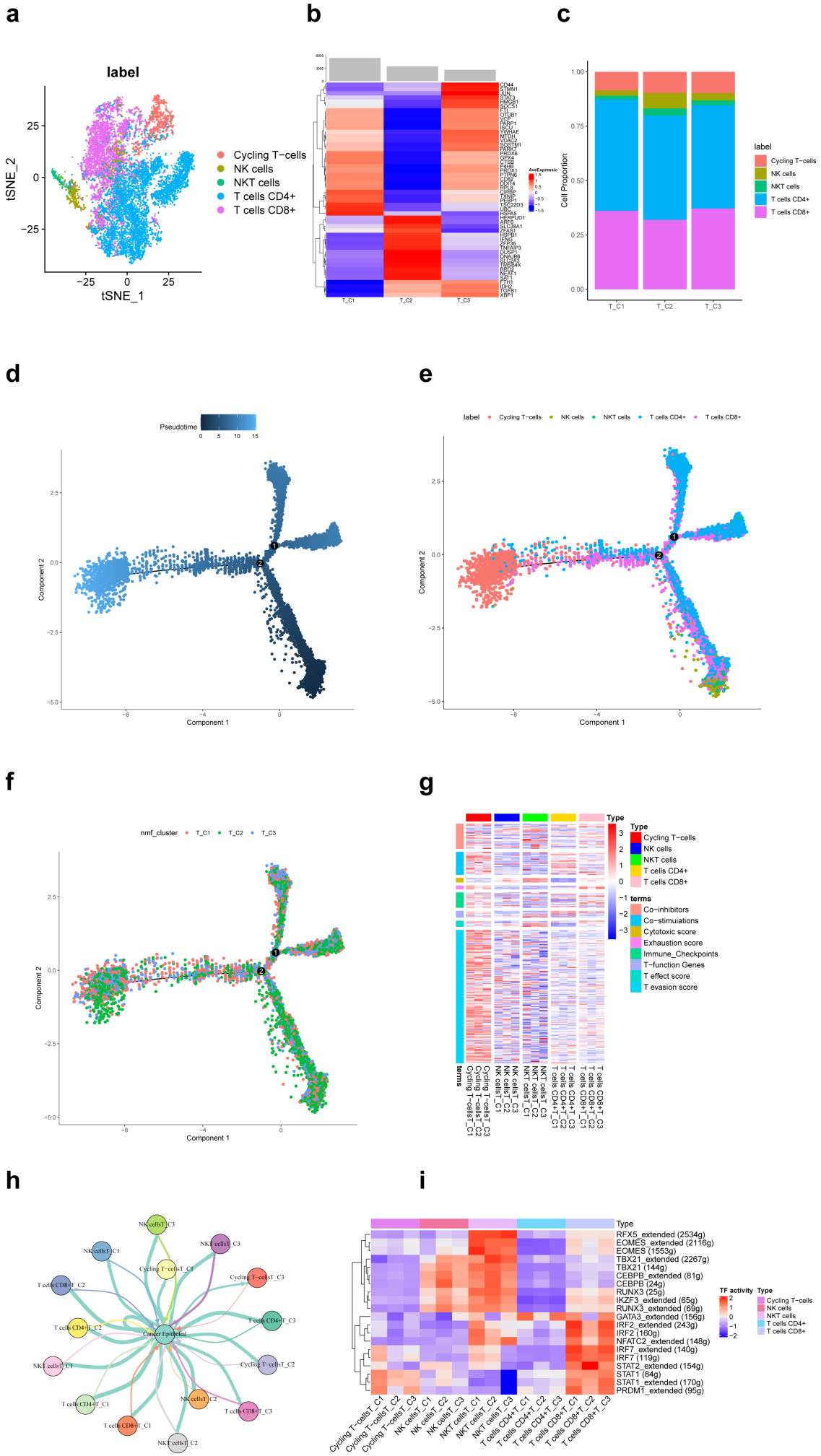
61. Wang YH, Shen CY, Lin SC, et al. Monocytes secrete CXCL7 to promote breast cancer progression. Cell Death Dis. 2021;12(12):1090. Published 2021 Nov 17.

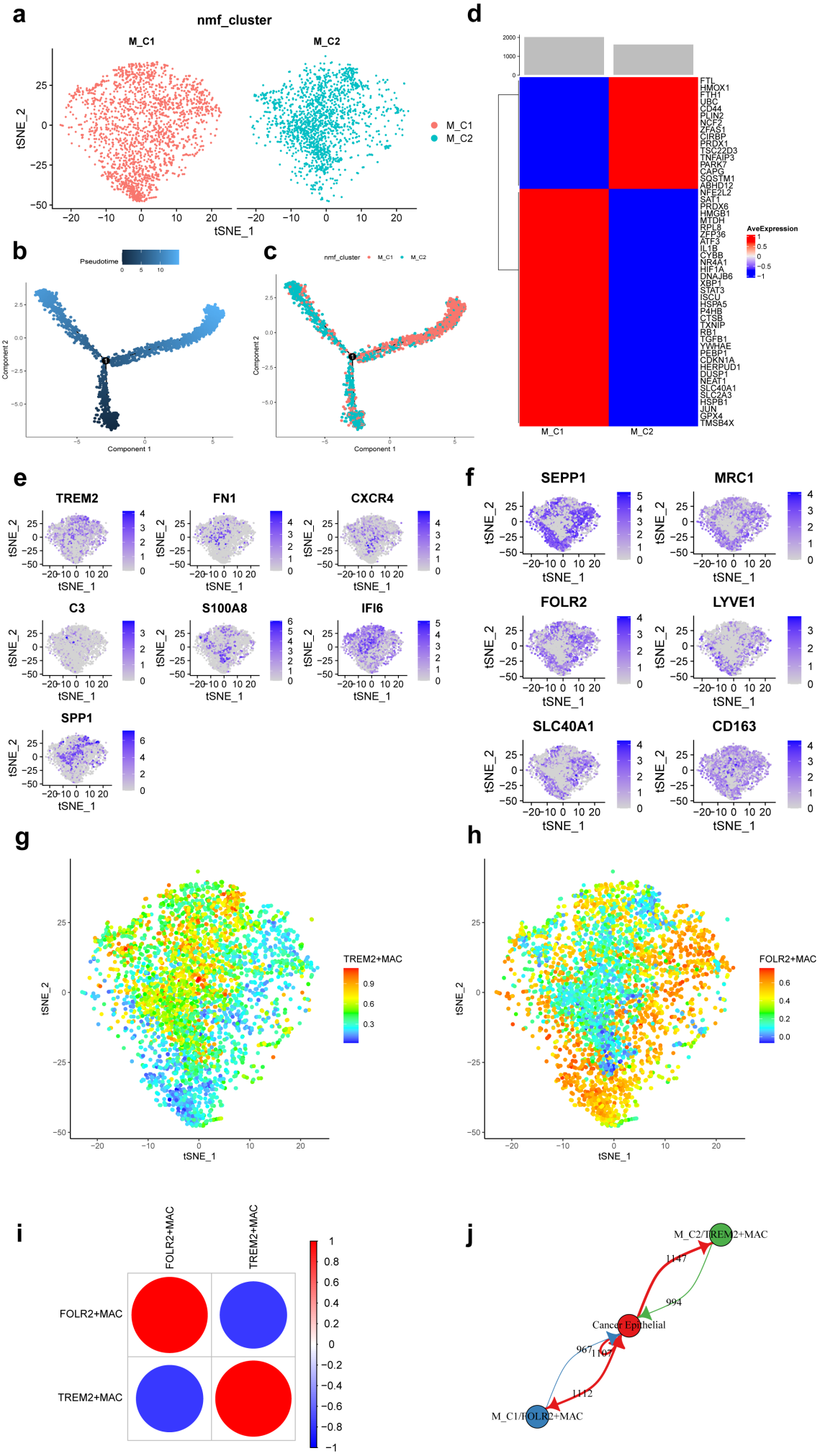
62. Bense RD, Sotiriou C, Piccart-Gebhart MJ, et al. Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer. J Natl Cancer Inst. 2016;109(1):djw192.

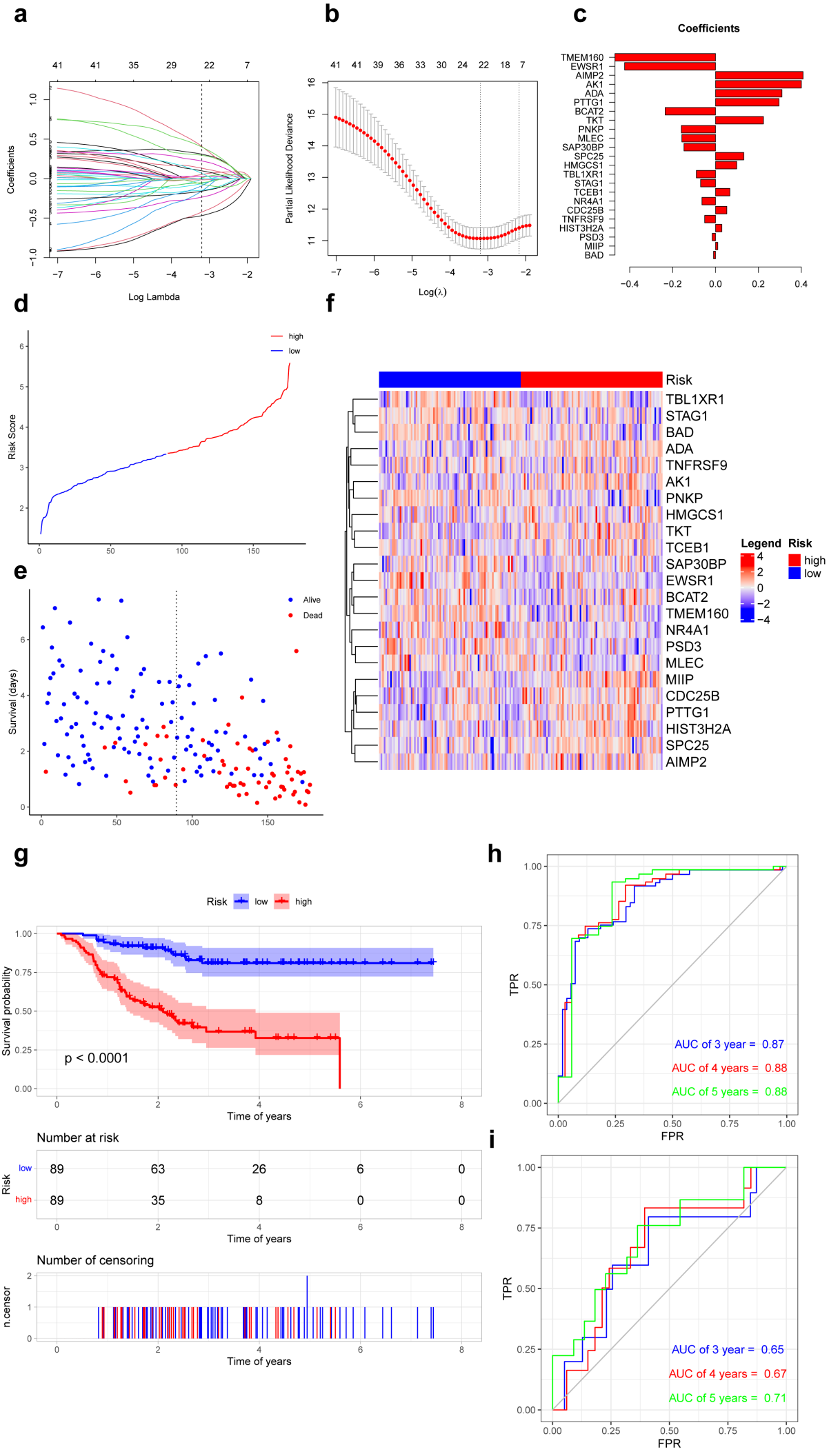
63. Majorini MT, Cancila V, Rigoni A, et al. Infiltrating Mast Cell-Mediated Stimulation of Estrogen Receptor Activity in Breast Cancer Cells Promotes the Luminal Phenotype. Cancer Res. 2020;80(11):2311-2324.

• **Additional files File:** Appendix 1-figure 1.TIF. Appendix 1-figure 2.TIF. Appendix 1-figure 3.TIF. Appendix 1-figure 4.TIF. Appendix 1-figure 5.TIF. Appendix 1-figure 6.TIF. Appendix 1-figure 7.TIF. Figure 1–figure supplement.pdf. Figure 2–figure supplement.pdf. Figure 3–figure supplement.pdf. Figure 4–figure supplement.pdf. Figure 5–figure supplement.pdf. Figure 6–figure supplement.pdf. Figure 7–figure supplement.pdf. Appendix 2-ICMJE.zip. Appendix 3-collaborators.xlsx.


a
celltype_major
tSNE_2
tSNE_1
25
0
−25
B−cells
CAFs
Cancer Epithelial
Endothelial
Myeloid
Normal Epithelial
Plasmablasts
PVL
T−cells
b
celltype_minor
B cells Memory
B cells Naive
CAFs MSC iCAF−like
CAFs myCAF−like
Cancer Basal SC
Cancer Cycling
Cancer Her2 SC
Cancer LumA SC
Cancer LumB SC
Cycling PVL
Cycling T−cells
Cycling_Myeloid
DCs
Endothelial ACKR1
Endothelial CXCL12
Endothelial Lymphatic LYVE1
Endothelial RGS5
Luminal Progenitors
Macrophage
Mature Luminal
Monocyte
Myoepithelial
NK cells
NKT cells
Plasmablasts
PVL Differentiated
PVL Immature
T cells CD4+
T cells CD8+
c
celltype_major
B−cells
CAFs
Cancer Epithelial
Cycling_Myeloid
DCs
Endothelial
Macrophage
Monocyte
Normal Epithelial
Plasmablasts
PVL
T−cells
d
Cycling_Myeloid
Normal Epithelial
Endothelial
PVL
Cancer Epithelial
CAFs
Plasmablasts
Monocyte
T−cells
DCs
B−cells
Macrophage
e
<=60
>60
IDC
MBC
T1
T2
T3
AveExpression
2
1
0
−1
−2
0
11500
23000
f
Cell
AveExpression
4
2
0
−2
−4
B−cells
CAFs
Cancer Epithelial
Cycling_Myeloid
DCs
Endothelial
Macrophage
Monocyte
Normal Epithelial
Plasmablasts
PVL
T−cells


a
label
tSNE_2
tSNE_1
Cycling T−cells
NK cells
NKT cells
T cells CD4+
T cells CD8+
b
AveExpressio
T_C1
T_C2
T_C3
c
Cell Proportion
label
Cycling T−cells
NK cells
NKT cells
T cells CD4+
T cells CD8+
d
Pseudotime
Component 1
Component 2
e
label
Cycling T−cells
NK cells
NKT cells
T cells CD4+
T cells CD8+
Component 1
Component 2
f
nmf_cluster
T_C1
T_C2
T_C3
Component 1
Component 2
g
Type
Cycling T−cells
NK cells
NKT cells
T cells CD4+
T cells CD8+
terms
Co−inhibitors
Co−stimuiations
Cytotoxic score
Exhaustion score
Immune_Checkpoints
T−function Genes
T effect score
T evasion score
h
Cancer Epithelial
i
Type
RFX5_extended (2534g)
EOMES_extended (2116g)
EOMES (1553g)
TBX21_extended (2267g)
TBX21 (144g)
CEBPB_extended (81g)
CEBPB (24g)
RUNX3 (25g)
IKZF3_extended (65g)
RUNX3_extended (69g)
GATA3_extended (156g)
IRF2_extended (243g)
IRF2 (160g)
NFATC2_extended (148g)
IRF7_extended (140g)
IRF7 (119g)
STAT2_extended (154g)
STAT1 (84g)
STAT1_extended (170g)
PRDM1_extended (95g)
TF activity
Type
Cycling T−cells
NK cells
NKT cells
T cells CD4+
T cells CD8+

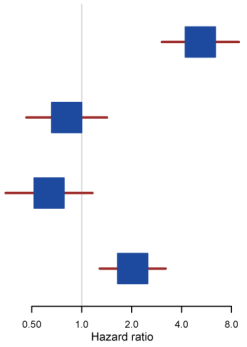

a
nmf_cluster
M_C1
M_C2
tSNE_2
tSNE_1
b
Pseudotime
Component 1
Component 2
c
nmf_cluster
d
AveExpression
FTL
HMOX1
FTH1
UBC
CD44
PLIN2
NCF2
ZFAS1
CIRBP
PRDX1
TSC22D3
TNFAIP3
PARK7
CAPG
SQSTM1
ABHD12
NFE2L2
SAT1
PRDX6
HMGB1
MTDH
RPL8
ZFP36
ATF3
IL1B
CYBB
NR4A1
HIF1A
DNAJB6
XBP1
STAT3
ISCU
HSPA5
P4HB
CTSB
TXNIP
RB1
TGFB1
YWHAE
PEBP1
CDKN1A
HERPUD1
DUSP1
NEAT1
SLC40A1
SLC2A3
HSPB1
JUN
GPX4
TMSB4X
e
TREM2
FN1
CXCR4
C3
S100A8
IFI6
SPP1
f
SEPP1
MRC1
FOLR2
LYVE1
SLC40A1
CD163
g
TREM2+MAC
h
FOLR2+MAC
i
FOLR2+MAC
TREM2+MAC
j
M_C2/TREM2+MAC
Cancer Epithelial
M_C1/FOLR2+MAC
1147
994
967
1107
1112


a
41 41 35 29 22 7
Coefficients
Log Lambda
b
41 41 39 36 33 30 24 22 18 7
Partial Likelihood Deviance
Log(λ)
c
Coefficients
TMEM160
EWSR1
AIMP2
AK1
ADA
PTTG1
BCAT2
TKT
PNKP
MLEC
SAP30BP
SPC25
HMGCS1
TBL1XR1
STAG1
TCEB1
NR4A1
CDC25B
TNFRSF9
HIST3H2A
PSD3
MIIP
BAD
d
high
low
Risk Score
e
Alive
Dead
Survival (days)
f
Risk
TBL1XR1
STAG1
BAD
ADA
TNFRSF9
AK1
PNKP
HMGCS1
TKT
TCEB1
SAP30BP
EWSR1
BCAT2
TMEM160
NR4A1
PSD3
MLEC
MIIP
CDC25B
PTTG1
HIST3H2A
SPC25
AIMP2
Legend
Risk
high
low
g
Risk
low
high
Survival probability
p < 0.0001
Time of years
Number at risk
low 89 63 26 6 0
high 89 35 8 0 0
Number of censoring
n.censor
h
TPR
FPR
AUC of 3 year = 0.87
AUC of 4 years = 0.88
AUC of 5 years = 0.88
i
TPR
FPR
AUC of 3 year = 0.65
AUC of 4 years = 0.67
AUC of 5 years = 0.71

a

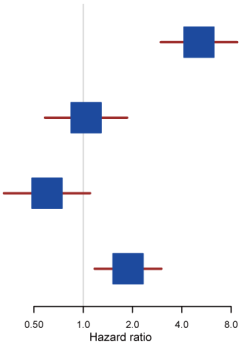
GSE25066 Univariate cox regresson analysis

Risk factors	pvalue	Hazard ratio
Risk (high vs. low)	<0.001	5.194(3.049−8.848)
age (>60 vs. <=60)	0.538	0.811(0.464−1.419)
grade (G3 vs. G1-2)	0.216	0.636(0.348−1.161)
stage (III vs. I-II)	0.011	2.030(1.286−3.205)



GSE25066 Multivariate cox regresson analysis

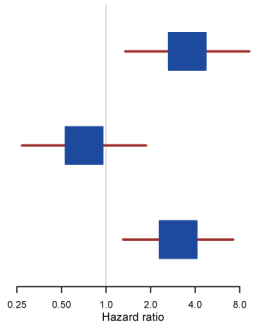
Risk factors	pvalue	Hazard ratio
Risk (high vs. low)	<0.001	5.091(2.979−8.701)
age (>60 vs. <=60)	0.909	1.041(0.585−1.853)
grade (G3 vs. G1-2)	0.166	0.601(0.328−1.100)
stage (III vs. I-II)	0.027	1.878(1.175−3.001)



b

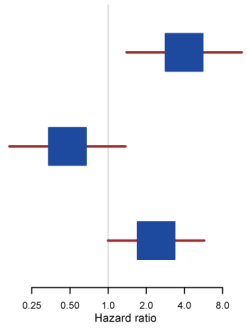
GSE86166 Univariate cox regresson analysis

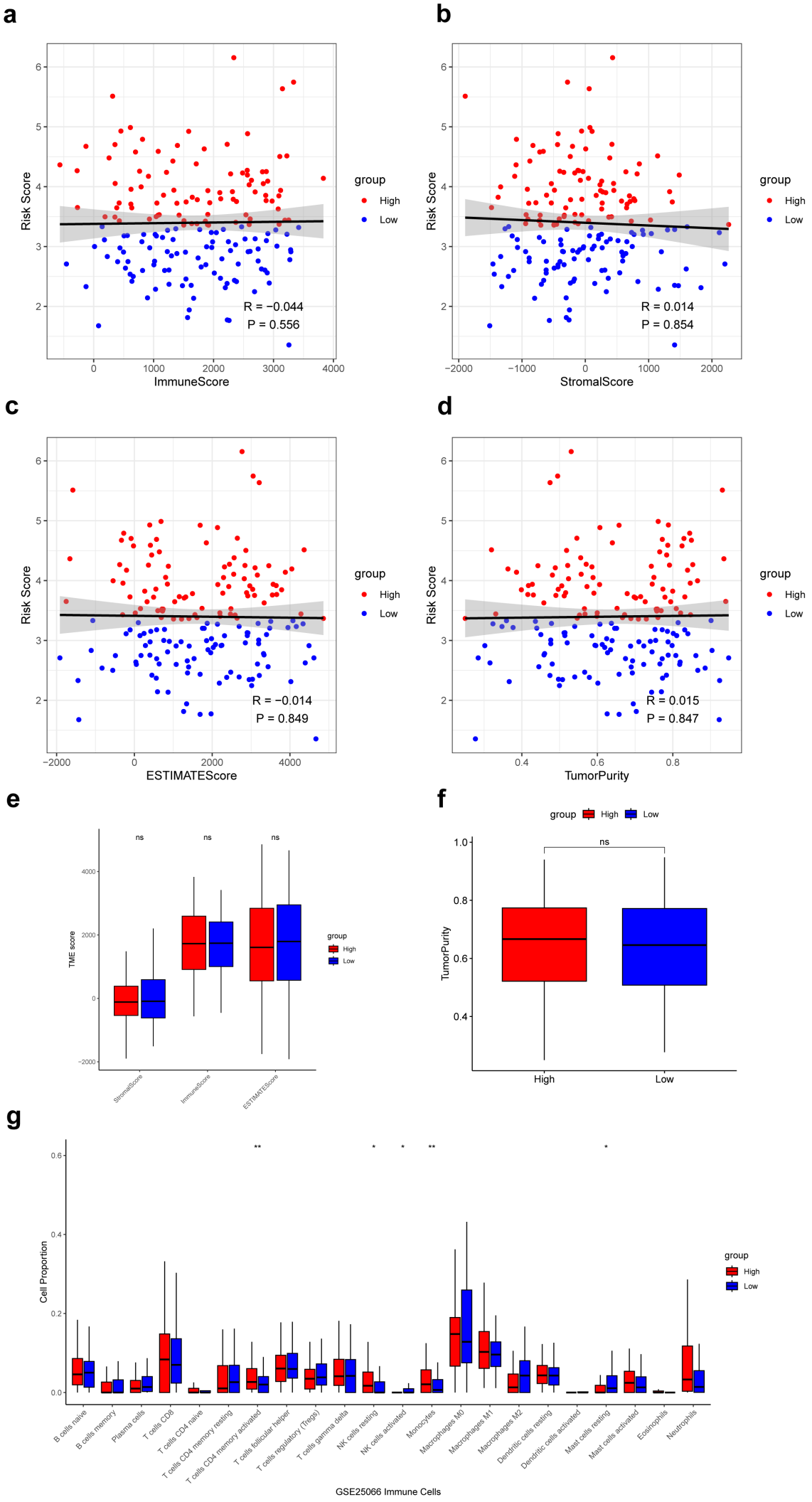
Risk factors	pvalue	Hazard ratio
Risk (high vs. low)	0.031	3.536(1.349−9.270)
grade (G3 vs. G1-2)	0.558	0.710(0.271−1.857)
stage (III vs. I-II)	0.031	3.068(1.305−7.215)

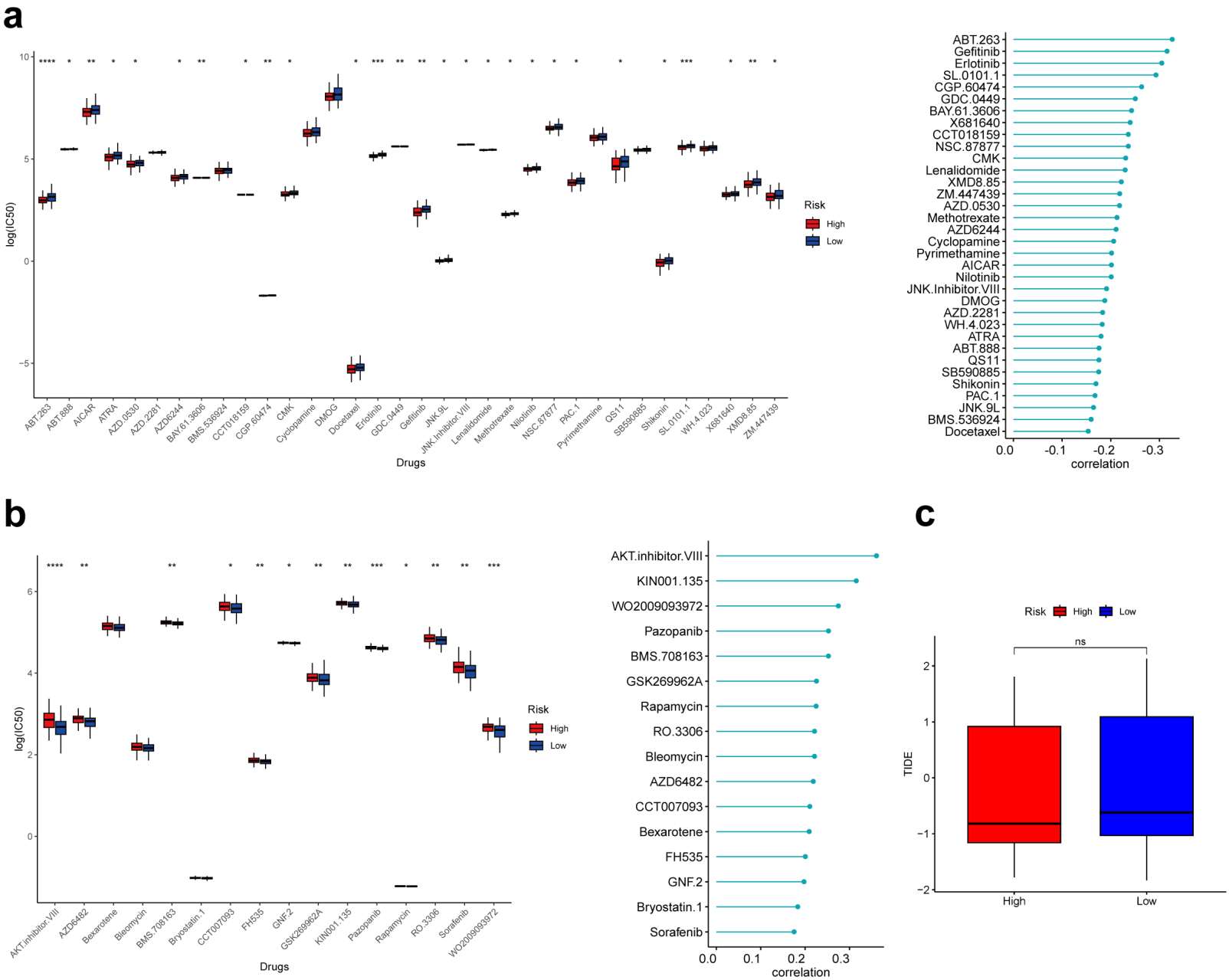


GSE86166 Multivariate cox regresson analysis

Risk factors	pvalue	Hazard ratio
Risk (high vs. low)	0.030	3.982(1.398−11.338)
grade (G3 vs. G1-2)	0.249	0.478(0.167−1.370)
stage (III vs. I-II)	0.100	2.392(0.999−5.731)




a
Risk Score
ImmuneScore
R = −0.044
P = 0.556
group
High
Low
b
Risk Score
StromalScore
R = 0.014
P = 0.854
group
High
Low
c
Risk Score
ESTIMATEScore
R = −0.014
P = 0.849
group
High
Low
d
Risk Score
TumorPurity
R = 0.015
P = 0.847
group
High
Low
e
TME score
ns
ns
ns
StromalScore
ImmuneScore
ESTIMATEScore
group
High
Low
f
group
High
Low
TumorPurity
ns
High
Low
g
Cell Proportion
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
group
High
Low
GSE25066 Immune Cells


a
log(IC50)
Drugs
Risk
High
Low
ABT.263
ABT.888
AICAR
ATRA
AZD.0530
AZD.2281
AZD6244
BAY.61.3606
BMS.536924
CCT018159
CGP.60474
CMK
Cyclopamine
DMOG
Docetaxel
Erlotinib
GDC.0449
Gefitinib
JNK.9L
JNK.Inhibitor.VIII
Lenalidomide
Methotrexate
Nilotinib
NSC.87877
PAC.1
Pyrimethamine
QS11
SB590885
Shikonin
SL.0101.1
WH.4.023
X681640
XMD8.85
ZM.447439
correlation
b
AKT.inhibitor.VIII
AZD6482
Bexarotene
Bleomycin
BMS.708163
Bryostatin.1
CCT007093
FH535
GNF.2
GSK269962A
KIN001.135
Pazopanib
Rapamycin
RO.3306
Sorafenib
WO2009093972
c
TIDE
ns
High
Low