

A CTL/M2 Macrophage-Related Four-Gene Signature Predicting Metastasis-Free Survival in Triple-Negative Breast Cancer Treated with Adjuvant Radiotherapy

Yunfei Ye

Daping Hospital & Army Medical center of PLA, Army Medcial University

Jungang Ma

Daping Hospital & Army Medical Center of PLA, Army Medical University

Qin Zhang

Daping Hospital & Army Medical Center, Army Medical University of PLA

Kai Xiong

Daping Hospital & Army Medical Center of PLA, Army Medical University

Zhimin Zhang

Daping Hospital & Army Medical Center of PLA, Army Medical University

Chuan Chen

Daping Hospital & Army Medical Center of PLA, Army Medical University

He Xiao

Daping Hospital & Army Medical Center of PLA, Army Medical University

Dong Wang (✉ dongwang64@hotmail.com)

Daping Hospital & Army Medical Center of PLA, Army Medical University <https://orcid.org/0000-0001-6585-2704>

Research Article

Keywords: triple-negative breast cancer, radiotherapy, cytotoxic T lymphocyte, M2 macrophage, metastasis-free survival, prognosis

Posted Date: June 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-527123/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose This study aimed to develop and validate a prognostic model for metastasis-free survival (MFS) based on genes that may functionally interact with cytotoxic T lymphocytes (CTLs) and M2 macrophages in patients with triple-negative breast cancer (TNBC) who underwent adjuvant radiotherapy.

Methods The transcriptional profiles and phenotypical files of TNBC and other subtypes of breast cancer were downloaded from the Gene Expression Omnibus (GEO). The abundance of infiltrated immune cells was evaluated through CIBERSORTx or MCP-counter. A weighted linear model, the score for MFS (SMFS), was developed by using least absolute shrinkage and selection operator (LASSO) in GSE58812 and validated in GSE2034 and GSE12276. The biological implication of SMFS was explored by evaluating its associations with TNBC molecular subtypes and other radiosensitivity- or immune-related signatures.

Results A model consisting of the gene expression ratios of PCDH12/ELP3, PCDH12/MSRA and FAM160B2/MSRA with nonzero coefficients finally selected by LASSO was developed in GSE58812. In GSE2034 (treatment with adjuvant radiotherapy), SMFS was significantly associated with MFS in TNBC patients ($HR=8.767$, 95% CI: 1.856-41.408, $P=0.006$) and, to a lesser extent, in non-TNBC patients ($HR=2.888$, 95% CI: 1.076-7.750, $P=0.035$). However, the interaction of subtype (TNBC vs non-TNBC) and SMFS tended to be significant ($P_{interaction}=0.081$). In contrast, SMFS was not significantly associated with MFS in either TNBC patients ($P=0.499$) or non-TNBC patients ($P=0.536$) in GSE12276 (treatment without radiotherapy). Among the four TNBC molecular subtypes, the c1 and c4 subtypes exhibited higher CTL infiltration and lower SMFS values than the c2 and c3 subtypes. In addition, SMFS was positively correlated with the abundance of endothelial cells ($r=0.413$, $P<0.001$).

Conclusions The proposed model has the potential to predict MFS in TNBC patients after adjuvant radiotherapy. SMFS may represent a measurement of tumor immune suppression.

Introduction

Breast cancer is a malignant tumor with a high incidence worldwide and in China [1, 2]. Triple-negative breast cancer (TNBC) accounts for 15–20% of all breast cancers. TNBC has the worst prognosis of all breast cancer subtypes, and the 5-year mortality rate can reach 40% after the initial diagnosis compared with luminal breast cancer [3]. TNBC is highly aggressive; nearly 46% of patients with TNBC will develop distant metastasis, with a median survival time of only 13.3 months. Most distant metastases of TNBC occur within 3 years after the initial diagnosis, with a higher probability of brain metastasis and visceral metastasis [4, 5]. According to the National Comprehensive Cancer Network (NCCN) guidelines of breast cancer, due to the young age of onset and the lack of targets for endocrine therapy and anti-HER2 therapy, radiotherapy (RT) and chemotherapy are usually used as adjuvant treatments in TNBC [6]. Although several transcriptomic signatures have been proposed for predicting the efficacy of RT for breast cancer [7–9], there are currently no such indicators specific to TNBC. Therefore, screening the population suitable for RT in TNBC is a problem worthy of study.

TNBC is defined as breast cancer with negative estrogen receptor, progesterone receptor, and HER2 receptor expression [10]. Compared with luminal breast cancer, TNBC has biological characteristics, such as high proliferation, a high overlap ratio with basal-like tumors, more mesenchymal stem cells, and homologous recombination defects with BRCA1/2 inactivation [3, 11, 12]. Moreover, recent studies have demonstrated that TNBC has a higher level of T cell infiltration accompanied by a higher level of immunosuppression than luminal breast cancer [13, 14]. One study screened genes that interact with RT and affect the disease-specific survival (DSS) of breast cancer from antigen presentation process genes and constructed an immune signature (IMS) [15]. Another study showed that high tumor-infiltrating lymphocytes (TILs) in the primary tumor may independently reduce the risk of an ipsilateral breast tumor recurrence (IBTR) in breast cancer, whereas patients with low TILs may attain a greater benefit from RT regarding the risk of IBTR [16]. However, this way of developing indicators using candidate gene sets as discovery sets or using a rough estimate of TILs to reflect immune status does not take into account the subclasses of infiltrated immune cells and their functions in TNBC tumors. Therefore, the performance of these signatures derived from the perspective of immunity to predict benefit from RT should be further tested in TNBC.

Many studies have proposed radioresistant gene signatures or recurrence prediction models after mastectomy or breast-conserving surgery [7–9]. These signatures or models were developed mainly based on luminal breast cancers. Moreover, Sjöström found that the post-RT recurrence score calculated by the same gene signature had inconsistent effects on prognosis for IBTR between the ER- and ER+ groups receiving adjuvant RT. This result suggested that the mechanisms of RT resistance may be different among various breast cancer subtypes with different immune microenvironments and tumor cell characteristics [8]. Considering the high degree of immune infiltration in TNBC and the important impact of TILs on the prognosis of breast cancer patients undergoing adjuvant RT, we speculate that the subclass and number of immune infiltrating cells inherent in tumor masses and their cellular functional status may affect the efficacy of RT, resulting in different prognoses of patients. Therefore, inspired by the method proposed by Jiang [17], a more general and robust method was developed to screen genes that may affect the function of certain immune cells in tumors, thereby affecting the survival of TNBC patients after RT. Then, based on these genes, we developed a new model, the score for metastasis-free survival (SMFS), which contains four genes and can be used as a prognostic marker for distant metastasis of TNBC after RT.

Methods

Dataset retrieval and preprocessing

The Gene Expression Omnibus (GEO) was screened with the keyword “breast cancer”. Only tumor gene expression profiles obtained by arrays containing TNBC primary tissue samples were selected. All datasets used in this study are shown in Table 1. The raw data (CEL files) of each dataset were downloaded from GEO. The *rma* function implemented in the *limma* R package was used to perform background correction, normalization and probe summarization for each dataset [18]. A total of twelve

datasets were merged, and the batch effect was corrected by using the main function *virtualArrayComBat* in the *virtualArray* R package [19]. For genes represented by several probes, the probe with the maximum interquartile range (IQR) was finally selected as the expression of that gene. A total of 12402 genes were finally included. To expand the number of TNBC samples included in the consensus clustering analysis, a classifier discriminating TNBC samples from non-TNBC samples was first developed in the GSE21653 dataset by using the support vector machine (SVM) algorithm. In brief, the expression of 749 differentially expressed genes (DEGs) identified with 87 TNBC samples and 179 non-TNBC samples through *limma* was used as the training expression matrix to construct the SVM classifier through the *penalizedSVM* R package [20]. The classifier was then verified in the GSE12276 dataset. The sensitivity and specificity of the SVM classifier in the training and validation sets were 94.25% and 96.65% and 95.45% and 93.33%, respectively. The number of TNBC samples used by the SVM classifier for GSE20685, GSE7390 and GSE31448 are shown in Table 1. A total of 920 TNBC samples were finally included in the subsequent analysis. We followed Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) guidelines for reporting this study [21].

Evaluation of the abundance of tumor-infiltrating immune cells, consensus clustering and annotation

The abundance of tumor-infiltrating immune cells was evaluated by both MCP-counter [22] and CIBERSORTx [23] in GSE58812 and whole TNBC samples. In the MCP-counter method, the abundance of ten kinds of immune cells was represented as the log₂ geometric mean of these immune cell transcriptomic markers called the MCP-counter score. Although the MCP-counter score cannot be represented as the actual fraction of each immune cell subpopulation in bulk tumor tissues, it has a numerical advantage for downstream statistical analysis. CIBERSORTx was used with the LM22 signature matrix and B-mode batch correction. The absolute score was used as the enumeration of abundance of the 22 kinds of immune cells. Consensus clustering of 4 subtypes was achieved using genes with 5% maximum variation in the whole TNBC samples with the *CancerSubtypes* R package [24]. To characterize the biological features of each subtype, the log fold change (FC) values of all 12402 genes were calculated for each subtype versus the rest, and the genes were subjected to gene set enrichment analysis with the *gsePathway* function implemented in the *clusterProfiler* and *ReactomePA* R package [25].

Development of the transcriptional score for MFS (SMFS) and validation

Having been inspired by the method proposed by Jiang [17], a similar approach was used to identify genes that could influence the function of certain immune cells and affect metastasis-free survival (MFS). In the GSE58812 dataset, the association of MFS with the abundance of immune cells was determined by univariate Cox regression. The abundance of cytotoxic T lymphocytes (CTLs) inferred from MCP-counter and M2 macrophages inferred from CIBERSORTx were found to be significantly positively and negatively associated with MFS, respectively (Supplementary Table 1). Based on these two univariate Cox models, gene expression and the interaction of the abundance of immune cells (CTL or M2 macrophage) and gene expression were added, and a likelihood ratio test was used to evaluate the

significance of these two terms. This analysis was performed for CTLs and M2 macrophages separately. In the Cox model with three terms, only genes in which the significance of immune cells and gene expression was less than 0.05 and the significance of the interaction and likelihood ratio test was less than 0.01 were chosen. In this situation, genes with a z value of the interaction term less than zero indicated that the expression of those genes could improve the antitumor immune effect of CTLs, whereas genes with a z value of the interaction term greater than zero indicated that the expression of those genes could enhance the tumor-promoting effect of M2 macrophages in terms of MFS. The gene expression ratio was calculated by dividing the expression of genes with $z > 0$ in the interaction with M2 macrophages (M2) by the expression of genes with $z < 0$ in the interaction with CTLs. Least absolute shrinkage and selection operator (LASSO) was used to perform penalized Cox regression. Fivefold cross-validation was used to select the lambda value with the *cv.glmnet* function. The final values of the coefficients of the gene expression ratios that were not zeroed out were used as the weights for the linear weight model for the prediction of MFS, which was defined as the score for MFS (SMFS). The linear weight model for MFS was validated in GSE2034.

Computation of previously published immune signature scores and radiosensitivity indices

Because the immune status of the tumor microenvironment has a clear impact on the long-term clinical outcomes of breast cancer patients treated with breast-conserving surgery followed by adjuvant RT, two immune-related signature scores, the IFNy signature [26] and TIDE [17] were used to further characterize SMFS. The IFNy signature score was calculated by averaging 17 genes in the "Expanded immune gene signature" shown in Table 2 published in the original article. Because of the prevalently high CTL infiltration in TNBC tumor tissues (> 1%), only the T cell dysfunction score was used. Ninety-three genes were retained to calculate the TIDE score. The TIDE value itself is a measurement of the degree of the irreversible exhaustion of CTLs. The higher the TIDE value is, the more irreversible the exhaustion of CTLs [17]. In addition to immune-related signatures, the 10-gene radiosensitivity index (RSI) [27] and 27-gene adjuvant radiotherapy intensification classifier (ARTIC) [7] were also included to illustrate differences in the TNBC specificity of these radiosensitivity indices as well as SMFS. However, only 22 genes were retained to calculate the ARTIC score from the 27 originally published genes. The IMS was calculated according to the original formula [15].

Molecular subtype

Lehmann's TNBC subtyping [28] was performed by using a web tool. The z-score-transformed expression matrix of a total of 603 cases was loaded into the web service after removing ER-positive samples from 920 TNBC samples based on the criteria proposed by the authors [28]. Because some signature genes representing four subtypes in Burstein's classification were missed in the expression matrix of 920 TNBC samples, Burstein's molecular classification was achieved through the method of nearest template prediction [29] by using GSE76124 as the training set [30]. Fifty genes were selected to represent each of four original subtypes, and a total of 200 genes comprised the classifier. Pearson correlation coefficients

were used as metrics of similarity. All cases with P values less than 0.1 were retained for subsequent analysis.

Statistical analysis

The abundance of immune cells, signature scores and other continuous variables are represented as the median value and IQR and visualized with scatter plots and box plots. The difference in abundance of immune cells and signature scores among four TNBC subtypes was evaluated by Kruskal-Wallis test accompanied with pairwise comparisons. Chi-square test and Mann-Whitney U test was used to evaluate differences of category and continuous variates between TNBC and non-TNBC subsets respectively. The Kaplan-Meier method and the log-rank test were used to compare the MFS between different subgroups. Univariate and multivariate Cox regression analyses were used to evaluate independent prognostic factors for MFS. All statistical analyses were performed using SPSS 17.0 (IBM SPSS, Chicago, IL, USA). All tests were two-sided, and $P < 0.05$ was considered statistically significant.

Results

Development and characteristics of SMFS

Because almost all TNBC patients enrolled in GSE58812 (96.26%, 103/107) had received adjuvant RT without systemic therapy, GSE58812 was used as the discovery and training set. The z values of the interaction term of CTL or M2 with all genes were subjected to Reactome analysis separately. The Reactome pathways that were enriched for improvement of CTL functions (enrichment in $z < 0$) were involved in "Antigen processing-Cross presentation", "Interferon gamma signaling" and "PD-1 signaling". The "Collagen biosynthesis and modifying enzymes", "Extracellular matrix organization" and "Integrin cell surface interactions" pathways possibly participated in the enhancement of M2 functions (Fig. 1, Supplementary Table 2). Fifty-eight and 115 genes that had $z < 0$ and $z > 0$ in the interaction term for CTLs and M2, respectively, were identified (Supplementary Table 3). For the expression ratios of a total of 6670 (58×115) gene pairs, only PCDH12/ELP3, PCDH12/MSRA and FAM160B2/MSRA gene pairs were selected by LASSO. In GSE58812, 4, 44, 27 and 32 patients were categorized into the c1, c2, c3 and c4 subtypes respectively. Kaplan-Meier analysis showed that patients with the c1 or c4 subtypes had significantly longer MFS than those with the c2 or c3 subtypes (2-year MFS: 85.9% vs. 62.9%, log rank $P = 0.019$) (Fig. 2a). Univariate Cox regression revealed that age, subtype (c1 + c4 vs. c2 + c3), IFNy, RSI, ARTIC and SMFS were significantly associated with MFS. However, only age and SMFS remained significantly adjusted for subtype, abundance of CTLs and M2 and other signatures (Table 2). All patients were divided into low- and high-risk groups according to the 75th percentile of SMFS, and patients with low risk had significantly longer MFS than those with high risk (5-year MFS: 79.9% vs. 32.3%, log rank $P < 0.001$) (Fig. 2b).

Validation of SMFS in GSE2034 and GSE12276

The comparisons of baseline clinical characteristics between TNBC and non-TNBC samples in GSE2034 and GSE12276 are shown in Supplementary Table 4. RT was given to 87% of patients in GSE2034, whereas none of the patients in GSE12276 received adjuvant RT. There were 41 TNBC patients in the GSE2034 dataset. In the subset of TNBC patients in GSE2034, only SMFS was significantly associated with MFS ($HR = 8.767$ 95% CI: 1.856–41.408, $P = 0.006$) (Table 3). The patients in the whole GSE2034 dataset were also categorized into low- and high-risk groups based on the 75th percentile of SMFS. Patients with low risk exhibited profoundly superior MFS compared with patients with high risk in the TNBC subset (Fig. 2c). Although SMFS was also significantly related to prognosis in the non-TNBC subpopulation, the effect was less pronounced with the same cutoff value (Fig. 2d). Although SMFS and proliferation were independent prognostic factors for MFS in the subset of non-TNBC patients in GSE2034 (Table 4), the interaction of TNBC and SMFS tended to be significant in the whole GSE2034 population ($P_{\text{interaction}} = 0.081$). Moreover, patients with short MFS (< 24 months) had significantly higher SMFS than the rest of the patients in the TNBC subset (median: 4.69 vs 4.20, Mann-Whitney $P = 0.001$). However, such a difference was not observed in the subset of non-TNBC patients (median: 4.34 vs 4.28, Mann-Whitney U test, $P = 0.244$). As expected, ARTIC was shown to be significantly associated with MFS only in the non-TNBC subset in GSE2034 (Table 3). These results suggested that SMFS had a more pronounced association with prognosis in TNBC patients who received adjuvant RT than ARTIC, which was developed mainly based on patients with luminal breast cancer. It was noted that SMFS did not exhibit any relationships with prognosis in either the TNBC or non-TNBC subset in the GSE12276 dataset, roughly illustrating the RT specificity of SMFS (Table 3, Fig. 2e, f). In fact, age, T stage and ARTIC remained significantly associated with MFS in multivariate Cox regression in the non-TNBC subset in GSE12276 (Table 4).

Subtype and SMFS

To gain insight into the biological characteristics of SMFS, the associations of SMFS with the subtypes of TNBC and other immune signatures were examined. Consensus clustering separated 920 TNBC samples into four subtypes, c1 ($n = 208$), c2 ($n = 314$), c3 ($n = 206$) and c4 ($n = 192$). The average silhouette width was 0.88 (Fig. 3a). Reactome analysis revealed that the c1 subtype had the properties of highly active mitosis (R-HSA-69620), the c2 subtype had an EMT phenotype (R-HSA-3000171), the nuclear receptor transcription pathway predominated in the c3 subtype (R-HSA-383280), and high immune cell signaling was found in the c4 subtype (R-HSA-909733) (Supplementary Table 5). The abundance of CTLs and M2 macrophages, IFNy, RSI, ARTIC, TIDE, IMS and SMFS were significantly different among the four subtypes (Fig. 3b, Supplementary Table 6). In particular, the abundances of CTLs and IFNy in the c1 and c4 subtypes were significantly higher than those in the c2 and c3 subtypes, respectively, whereas the abundance of M2 macrophages, RSI and SMFS in the c1 and c4 subtypes were significantly lower than those in the c2 and c3 subtypes. SMFS was slightly but significantly correlated with IFNy, TIDE, RSI and ARTIC. However, it was noted that RSI was strongly negatively correlated with IFNy and the abundance of CTLs (Fig. 3c). Nevertheless, SMFS was moderately positively correlated with the abundance of endothelial cells inferred by MCP-counter (Fig. 4). Burstein and Lehmann subtypes

were available for 510 samples from a total of 920 TNBC patients (Supplementary Table 7). It was observed that the c1 subtype almost only comprised of both the basal-like immune-activated (BLIA) and basal-like immunosuppressed (BLIS) subtypes of the Burstein classification, the c2 subtype consisted of mostly BLIS tumors, the c3 subtype almost only included the luminal androgen receptor (LAR) and mesenchymal (MES) subtypes of both the Burstein and Lehmann classifications, and the c4 subtype was similar to c1 and involved a large proportion of the BLIA subtype of the Burstein classification (Fig. 5).

Discussion

In this study, we developed SMFS based on the gene expression ratio that was calculated by dividing the expression of genes enhancing the tumor-promoting effect of M2 macrophages by the expression of genes improving the antitumor immune effect of CTLs. This ratio can take into account the possible interaction between CTLs and M2 macrophages to some extent [31]. Therefore, we speculate that the higher the SMFS value is, the higher the degree of tumor immunosuppression, and the lower the SMFS value is, the stronger the antitumor immune effect. SMFS was validated stratified for TNBC status and RT, creating four groups (TNBC with RT, non-TNBC with RT, TNBC with no RT, and non-TNBC with no RT). In the RT group (GSE2034), SMFS was an independent prognostic factor for MFS in both patients with TNBC and non-TNBC. However, SMFS was not associated with prognosis in terms of MFS in the no RT group (GSE12276), regardless of TNBC or non-TNBC. GSE2034 and GSE12276 were two independent datasets, and the administration of RT was not random, so it was impossible to verify the interaction between SMFS and RT in one cohort in a rigorous manner. However, these results could indirectly indicate the RT specificity of SMFS. In addition, SMFS showed a trend of interaction with the TNBC subtype in GSE2034 ($P_{\text{interaction}}=0.081$), demonstrating that SMFS may be more strongly linked to prognosis in TNBC patients than in non-TNBC patients. Compared with non-TNBC patients, TNBC patients with rapid distant recurrence (MFS < 1 year) had a higher median SMFS ($P=0.001$). These results suggest that SMFS could be a prognostic indicator for MFS specific to TNBC patients treated with adjuvant RT.

We also compared the performance of SMFS with previously published immune or RT signatures in breast cancer. First, we found that proliferation and SMFS were independent prognostic factors for MFS in the non-TNBC with RT group (Table 4). This finding indicates that proliferation is a very important factor affecting the effects of RT on patients with non-TNBC. Previous studies have also demonstrated that the radiosensitivity signature (RSS) and single-sample predictors (SSPs) could predict the prognosis for IBTR in patients with ER + tumors owing to their biological effect on proliferation [8, 9]. Second, previous studies found that RSI was associated with the prognosis for IBTR in the ER-RT + group [8, 27]. However, our results showed that RSI was associated with prognosis in terms of MFS in the training set (TNBC with RT) by univariate Cox regression, but the association was not significant by multivariate analysis (Table 2). Furthermore, the performance of RSI was poor in all groups of the validation sets, probably because RSI and ARTIC were developed with survival fraction at 2 Gy (SF2) [32] and locoregional recurrence (LRR) [7] as endpoints, which mainly reflected the local control effect of RT, that is, the direct effect of RT. However, RT might have an indirect effect on long-term survival through the

immune system, which could not be reflected by ARTIC or RSI. In fact, there is evidence that the RT sensitivity of solid tumors is associated with immune activation [33]. Third, in our analysis, IMS showed a prognostic effect on MFS only in the non-TNBC with RT group (GSE2034) (univariate $P=0.003$, multivariate $P=0.069$). IMS was developed in E-TABM-158 dataset [34], which only contains 15% TNBC patients, which might explain why IMS did not show prognostic efficacy in the TNBC with RT group. Finally, we found that T3/T4 patients had a worse prognosis than T1 patients in the no RT group (GSE12276), but no such phenomenon was observed in the RT group (GSE2034) (Table 3, Table 4). According to the clinical guidelines of breast cancer, T3 (tumor size >5 cm) tumors should be treated with RT after surgery [6], which likely indicates that the T3/T4 patients in GSE12276 could obtain a better prognosis by receiving RT.

Many previous studies have shown that the immune system plays an important role in mediating the antitumor effects of RT [35, 36]. For instance, RT can activate the antitumor immune effect by inducing the maturation of dendritic cells and enhancing the activation of T cells [37, 38]. That is, RT can eliminate the immunosuppressive state of cancer and turn immunologically “cold” tumors “hot” [39]. In our study, patients in the low-risk group (low SMFS) may derive a greater benefit from indirect RT effects through the immune system, leading to a better prognosis. SMFS developed by our approach may be able to select populations of breast cancer, especially TNBC, suitable for RT from the perspective of immunity. RT is generally considered to be an important local control method in breast cancer treatment [40], but the endpoint we used in the training set and validation sets was MFS, rather than the usual endpoints of IBTR or LRR. The reasons are as follows. First, the clinical outcome indicators available in public datasets are limited, especially satisfying the conditions of both RT and a sufficient number of patients with TNBC. Second, at present, the local control effect of RT for breast cancer has been very good, and there are many indicators to predict the local control efficacy of RT [7, 8, 27]. Our focus is on the interaction between immunity and RT in TNBC, and the antitumor immune effect is usually considered to be related to the long-term outcome of cancer [41, 42]. At present, TNBC patients are prone to developing distant metastasis after first-line treatment, and the survival time after distant metastasis is very short [4, 5]. Antitumor immune effects play an important role in the distant metastasis of tumors [43, 44]. Therefore, investigating how to combine the antitumor effects of RT and the immune system and maximize their role in the treatment of TNBC is the fundamental goal of this study. Recently, a study on the application of RT combined with immunotherapy in metastatic TNBC showed encouraging results [45]. However, there are currently no predictive biomarkers for RT combined with immunotherapy. We developed and validated a CTL/M2 macrophage-related four-gene signature (SMFS) that had prognostic value for MFS in TNBC patients undergoing RT, which could provide some information to achieve this goal, but it needs to be further verified in large randomized clinical trials or even trials of checkpoint blockade plus RT.

It was observed that each of our four subtypes of TNBC seemed to be comprised of two or several Burstein subtypes illustrating moderate heterogeneity between these two classification systems. We found that the c1 and c4 subtypes had the highest proportion of BLIA tumors but had a lower SMFS value and longer MFS. This further confirms that the lower the SMFS is, the stronger the antitumor immune activation. We also found that SMFS was significantly positively correlated with the abundance

of endothelial cells inferred by MCP-counter. Single-cell sequencing research of lung cancer showed that endothelial cells in tumors can downregulate immune cell homing and genes correlated with T-cell activity [46]. This suggests that the angiogenesis of tumors may be different from that of normal tissues and may impair the antitumor immune effect. SMFS included 4 genes, among which ELP3 (Gene ID: 55140) and MSRA (4482) were genes that could improve antitumor immune effect of CTLs, while FAM160B2 (64760) and PCDH12 (51294) were genes that could enhance the tumor-promoting effect of M2 macrophages. However, these genes did not overlap with the markers of endothelial cells or CTLs inferred by MCP-counter or the markers of M2 macrophages inferred by CIBERSORTx (data not shown). These four genes were incorporated into the model of SMFS in the form of PCDH12/ELP3, PCDH12/MSRA and FAM160B2/MSRA gene pairs. Therefore, these genes may be expressed on any cells of the tumor bulk and reflect the immunosuppression status of the tumor and immune ecosystem in the form of SMFS.

There are some limitations in our study. First, the number of TNBC cases in the validation sets was too small, and the administration of RT was not randomized which means that it was impossible to verify the predictive effect of SMFS and the interaction effect on RT. Second, the current clinical treatment of invasive breast cancer almost always includes chemotherapy and RT, and even neoadjuvant therapy has been widely used in the treatment of TNBC. However, most patients in GSE58812 and GSE2034 only received adjuvant RT, and patients in GSE12276 received chemotherapy but not RT. Third, in terms of the long-term survival influenced by the antitumor immune effects of RT, overall survival may be more appropriate as the endpoint to further develop classifiers [42]. Fourth, the biological characteristics of SMFS need to be further analyzed with the single-cell sequencing data of TNBC to evaluate the precise intercellular communications by which the functions of TILs or macrophages are impacted.

In conclusion, we developed a CTL/M2 macrophage-related four-gene signature (SMFS) that had prognostic value for MFS in TNBC patients undergoing RT and then validated our SMFS in two independent datasets of patients with or without RT. Our research provides an idea on how to use transcriptional data to screen genes interacting with tumor-infiltrating immune cells to develop prognostic or predictive indicators for RT. This study may provide new ideas for the development of biomarkers guiding the combined use of RT with immunotherapy in the future.

Declarations

Funding

This research was partially supported by the National Natural Science Foundation of China 81802292.

Conflicts of interest/Competing interests

The authors declare that there are no conflicts of interest.

Availability of data and material

The datasets used and/or analysis during the current study are available from the corresponding author on reasonable request.

Code availability

Not available

Authors' contributions

Conception and design: Yunfei Ye, He Xiao, Dong Wang

Development of methodology: Yunfei Ye, He Xiao, Dong Wang

Acquisition of data: Jungang Ma, Qin Zhang, Kai Xiong,

Analysis and interpretation of data: Yunfei Ye, He Xiao, Dong Wang

Writing, review, and/or revision of the manuscript: Yunfei Ye, Jungang Ma, Qin Zhang, Kai Xiong, Zhimin Zhang, Chuan Chen, He Xiao, Dong Wang

Study supervision: Yunfei Ye, He Xiao, Dong Wang

Ethics approval

All the public raw datasets used in this study have been approved by the Ethics Committees at the corresponding institutions.

Consent to participate

Not applicable

Consent for publication

Not applicable

Acknowledgment

We are grateful to Dr. Lei Lin from the Radiotherapy Center of Daping Hospital for her assistance in clinical radiotherapy related issues for breast cancer. This research was partially supported by the National Natural Science Foundation of China 81802292.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA: a cancer journal for clinicians* 2021

2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J (2016) Cancer Statistics in China, 2015. CA-CANCER J CLIN 66(2):115–132
3. Yin L, Duan J, Bian X, Yu S: **Triple-negative breast cancer molecular subtyping and treatment progress.** BREAST CANCER RES 2020, 22(611)
4. Lin NU, Claus E, Sohl J, Razzak AR, Arnaout A, Winer EP (2008) Sites of Distant Recurrence and Clinical Outcomes in Patients With Metastatic Triple-negative Breast Cancer High Incidence of Central Nervous System Metastases. CANCER-AM CANCER SOC 113(10):2638–2645
5. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: Clinical features and patterns of recurrence. CLIN CANCER RES 13(151):4429–4434
6. Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, Elias AD, Farrar WB, Forero A, Giordano SH et al (2017) NCCN Guidelines (R) Insights Breast Cancer, Version 1.2017 Featured Updates to the NCCN Guidelines. J NATL COMPR CANC NE 15(4):433
7. Sjostrom M, Chang SL, Fishbane N, Davicioni E, Zhao SG, Hartman L, Holmberg E, Feng FY, Speers CW, Pierce LJ et al (2019) Clinicogenomic Radiotherapy Classifier Predicting the Need for Intensified Locoregional Treatment After Breast-Conserving Surgery for Early-Stage Breast Cancer. J CLIN ONCOL 37(35):3340
8. Sjostrom M, Staaf J, Eden P, Warnberg F, Bergh J, Malmstrom P, Ferno M, Nimeus E, Fredriksson I: **Identification and validation of single-sample breast cancer radiosensitivity gene expression predictors.** BREAST CANCER RES 2018, 20(64)
9. Speers C, Zhao S, Liu M, Bartelink H, Pierce LJ, Feng FY (2015) Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. CLIN CANCER RES 21(16):3667–3677
10. Denkert C, Liedtke C, Tutt A, von Minckwitz G (2017) Molecular alterations in triple-negative breast cancer-the road to new treatment strategies. LANCET 389(10087):2430–2442
11. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER et al (2012) Comprehensive molecular portraits of human breast tumours. NATURE 490(7418):61–70
12. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MDM, Niu B, McLellan MD, Uzunangelov V et al: **Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin.** CELL 2014, 158(4):929–944
13. Miyan M, Schmidt-Mende J, Kiessling R, Poschke I, de Boniface J: **Differential tumor infiltration by T-cells characterizes intrinsic molecular subtypes in breast cancer.** J TRANSL MED 2016, 14(227)
14. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, Rees M, Ramaswamy A, Muenst S, Soysa SD et al: **A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer.** CELL 2019, 177(5):1330
15. Cui Y, Li B, Pollock EL, Horst KC, Li R (2018) Integrating Radiosensitivity and Immune Gene Signatures for Predicting Benefit of Radiotherapy in Breast Cancer. CLIN CANCER RES 24(19):4754–4762

16. Kovacs A, Tullberg AS, Ronnerman EW, Holmberg E, Hartman L, Sjostrom M, Lundstedt D, Malmstrom P, Ferno M, Karlsson P (2019) Effect of Radiotherapy After Breast-Conserving Surgery Depending on the Presence of Tumor-Infiltrating Lymphocytes: A Long-Term Follow-Up of the SweBCG91RT Randomized Trial. *J CLIN ONCOL* 37(14):1179
17. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B et al (2018) Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *NAT MED* 24(10):1550
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *NUCLEIC ACIDS RES* 43:e477
19. Heider A, Alt R: **virtualArray: a R/bioconductor package to merge raw data from different microarray platforms.** *BMC BIOINFORMATICS* 2013, 14(75)
20. Becker N, Werft W, Toedt G, Lichter P, Benner A (2009) penalizedSVM: a R-package for feature selection SVM classification. *BIOINFORMATICS* 25(13):1711–1712
21. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). *BREAST CANCER RES TR* 100(2):229–235
22. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selvès J, Laurent-Puig P, Sautes-Fridman C, Fridman WH et al: **Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression.** *GENOME BIOL* 2016, 17(218)
23. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. *NAT BIOTECHNOL* 37(7):773
24. Xu T, Thuc DL, Liu L, Su N, Wang R, Sun B, Colaprico A, Bontempi G, Li J: **CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization.** *BIOINFORMATICS* 2017, 33(19):3131–3133
25. Yu G, Wang L, Han Y, He Q (2012) clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* 16(5):284–287
26. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V et al (2017) IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J CLIN INVEST* 127(8):2930–2940
27. Torres-Roca JF, Fulp WJ, Caudell JJ, Servant N, Bollet MA, van de Vijver M, Naghavi AO, Harris EE, Eschrich SA (2015) Integration of a Radiosensitivity Molecular Signature Into the Assessment of Local Recurrence Risk in Breast Cancer. *INT J RADIAT ONCOL* 93(3):631–638
28. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J CLIN INVEST* 121(7):2750–2767
29. Hoshida Y: **Nearest Template Prediction: A Single-Sample-Based Flexible Class Prediction with Confidence Assessment.** *PLOS ONE* 2010, 5(e1554311)

30. Burstein MD, Tsimelzon A, Poage GM, Coyington KR, Contreras A, Fuqua SAW, Sayage MI, Osborne CK, Hilsenbeck SG, Chang JC et al (2015) Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer. *CLIN CANCER RES* 21(7):1688–1698
31. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, Tian T, Wei Z, Madan S, Sullivan RJ et al (2018) Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *NAT MED* 24(10):1545
32. Eschrich S, Zhang H, Zhao H, Boulware D, Lee J, Bloom G, Torres-Roca JF (2009) SYSTEMS BIOLOGY MODELING OF THE RADIATION SENSITIVITY NETWORK: A BIOMARKER DISCOVERY PLATFORM. *INT J RADIAT ONCOL* 75(2):497–505
33. Strom T, Harrison LB, Giuliano AR, Schell MJ, Eschrich SA, Berglund A, Fulp W, Thapa R, Coppola D, Kim S et al (2017) Tumour radiosensitivity is associated with immune activation in solid tumours. *EUR J CANCER* 84:304–314
34. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo W, Lapuk A, Neve RM, Qian Z, Ryder T et al (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *CANCER CELL* 10(6):529–541
35. Hou Y, Liang HL, Yu X, Liu Z, Cao X, Rao E, Huang X, Wang L, Li L, Bugno J et al (2021) Radiotherapy and immunotherapy converge on elimination of tumor-promoting erythroid progenitor cells through adaptive immunity. *SCI TRANSL MED* 13:eabb0130582
36. Weichselbaum RR, Liang H, Deng L, Fu Y (2017) Radiotherapy and immunotherapy: a beneficial liaison? *NAT REV CLIN ONCOL* 14(6):365–379
37. Hou Y, Liang H, Rao E, Zheng W, Huang X, Deng L, Zhang Y, Yu X, Xu M, Mauceri H et al: **Non-canonical NF-kappa B Antagonizes STING Sensor-Mediated DNA Sensing in Radiotherapy.** *IMMUNITY* 2018, **49**(3):490
38. Deng L, Liang H, Burnette B, Beckett M, Darga T, Weichselbaum RR, Fu Y (2014) Irradiation and anti-PD-L1 treatment synergistically promote antitumor immunity in mice. *J CLIN INVEST* 124(2):687–695
39. McLaughlin M, Patin EC, Pedersen M, Wilkins A, Dillon MT, Melcher AA, Harrington KJ (2020) Inflammatory microenvironment remodelling by tumour cells after radiotherapy. *NAT REV CANCER* 20(4):203–217
40. Effects of radiotherapy and surgery in early breast cancer. An overview of the randomized trials. *The New England journal of medicine* 1995, 333(22):1444–1455
41. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, Lin NU, Litiere S, Dancey J, Chen A et al (2017) iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *LANCET ONCOL* 18(3):E143–E152
42. Wang Q, Li P, Wu W: **A systematic analysis of immune genes and overall survival in cancer patients.** *BMC CANCER* 2019, 19(12251)
43. Kitamura T, Qian B, Pollard JW (2015) Immune cell promotion of metastasis. *NAT REV IMMUNOL* 15(2):73–86

44. Ren B, Cui M, Yang G, Wang H, Feng M, You L, Zhao Y: **Tumor microenvironment participates in metastasis of pancreatic cancer.** *MOL CANCER* 2018, 17(108)
45. Ho AY, Barker CA, Arnold BB, Powell SN, Hu ZI, Gucalp A, Lebron-Zapata L, Wen HY, Kallman C, D'Agnolo A et al (2020) A phase 2 clinical trial assessing the efficacy and safety of pembrolizumab and radiotherapy in patients with metastatic triple-negative breast cancer. *CANCER-AM CANCER SOC* 126(4):850–860
46. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwe H, Pircher A, Van den Eynde K et al (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. *NAT MED* 24(8):1277

Tables

Table 1
The list of GEO datasets used in this study

GSE number	Total number of profiled tumor	Number of TNBC	Percentage of patients underwent adjuvant radiotherapy	Usage in this study	Array and Platform used
GSE2034	286	41‡	87.00%	Validation set/Subtyping	Affymetrix GPL96
GSE2603	121	25‡	-	Subtyping	Affymetrix GPL96
GSE12276	204	44‡	-	Subtyping	Affymetrix GPL570
GSE21653	266	87‡	-	Subtyping	Affymetrix GPL570
GSE58812	107	107	96.26%	Discovery set/Subtyping	Affymetrix GPL570
GSE83937	131	131	-	Subtyping	Affymetrix GPL570
GSE76124	198	198	-	Subtyping	Affymetrix GPL570
GSE95700	57	57	-	Subtyping	Affymetrix GPL570
GSE31519	67	67	-	Subtyping	Affymetrix GPL96
GSE20685	327	75 †	-	Subtyping	Affymetrix GPL570
GSE7390	198	52 †	-	Subtyping	Affymetrix GPL96
GSE31448	89	36 †	-	Subtyping	Affymetrix GPL570
Total	2051	920			

Table 2
Univariate and multivariate Cox regression for MFS in the training set GSE58812

Factor	Median (IQR)	Univariate		Multivariate	
		HR (95% C.I.)	P	HR (95% C.I.)	P
Age	57 (49–66)	1.045 (1.013–1.079)	.006	1.039 (1.004–1.075)	.031
Subtype (c1 + c4 vs c2 + c3)	-	0.335 (0.129–0.874)	.025	1.259 (0.349–4.545)	.725
CTLs	5.153 (4.664–5.680)	0.337 (0.171–0.663)	.002	0.487 (0.080–2.974)	.436
Macrophages M2	0.448 (0.3829–0.5313)	33.20 (1.637–673.225)	.023	6.257 (0.161–242.96)	.326
IFNy	7.330 (6.622–8.000)	0.580 (0.388–0.866)	.008	1.107 (0.320–3.825)	.872
TIDE	-0.160 (-0.701–0.769)	0.894 (0.663–1.207)	.465	0.936 (0.624–1.404)	.748
RSI	0.525 (0.422–0.616)	27.666 (1.464–522.706)	.027	4.187 (0.090–193.9)	.464
ARTIC	-12.65 (-13.23–(-12.14))	1.508 (1.003–2.268)	.049	1.530 (0.918–2.548)	.102
IMS	78.94 (75.38–81.45)	0.985 (0.921–1.053)	.654	1.002 (0.895–1.122)	.970
SMFS	3.985 (3.308–4.892)	1.639 (1.284–2.094)	.000	1.624 (1.201–2.198)	.002

Table 3

Univariate Cox regression in the different subset of validation sets GSE2034 and GSE12276

	GSE2034 TNBC		GSE2034 Non-TNBC		GSE12276 TNBC		GSE12276 Non-TNBC	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age	0.981 (0.939–1.025)	0.391	0.987 (0.968–1.006)	0.183	1.022 (0.989–1.056)	0.188	1.019 (1.005–1.032)	0.008
T stage (T2 vs T1)	1.192 (0.390–3.648)	0.758	1.181 (0.731–1.906)	0.496	1.918 (0.713–5.160)	0.197	1.647 (1.041–2.606)	0.033
T stage (T3/T4 vs T1)	-	-	0.405 (0.055–2.957)	0.373	4.479 (1.412–14.205)	0.011	3.373 (1.773–6.416)	0.000
Proliferation	1.251 (0.155–10.108)	0.834	3.826 (2.068–7.080)	0.000	1.380 (0.315–6.048)	0.669	1.128 (0.739–1.723)	0.576
IFNy	0.575 (0.281–1.176)	0.130	0.753 (0.544–1.041)	0.086	0.715 (0.492–1.038)	0.078	0.959 (0.753–1.222)	0.734
TIDE	0.629 (0.359–1.104)	0.106	0.694 (0.535–0.900)	0.006	0.909 (0.674–1.225)	0.530	0.756 (0.591–0.968)	0.026
RSI	1.820 (0.063–52.709)	0.727	4.528 (0.717–28.615)	0.108	1.265 (0.163–9.803)	0.822	1.888 (0.540–6.606)	0.320
ARTIC	0.856 (0.455–1.610)	0.630	1.367 (1.038–1.800)	0.026	1.207 (0.831–1.752)	0.324	1.349 (1.053–1.730)	0.018
IMS	1.078 (0.957–1.216)	0.218	1.073 (1.024–1.125)	0.003	0.983 (0.913–1.057)	0.638	0.972 (0.936–1.009)	0.138
SMFS	8.767 (1.856–41.408)	0.006	2.796 (1.070–7.302)	0.036	0.598 (0.134–2.659)	0.499	1.857 (0.847–4.073)	0.122
Chemotherapy (Yes vs No)	-	-	-	-	1.233 (0.547–2.777)	0.614	0.856 (0.544–1.345)	0.499

	GSE2034 TNBC		GSE2034 Non-TNBC		GSE12276 TNBC		GSE12276 Non-TNBC	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Hormone therapy (Yes vs No)	-	-	-	-	0.646 (0.265–1.574)	0.337	1.373 (0.847–2.227)	0.199

Table 4
Multivariate Cox regression in the subset of Non-TNBC samples in the validation sets GSE2034 and GSE12276

	GSE2034		GSE12276	
	HR (95% CI)	P	HR (95% CI)	P
Age	0.980 (0.960–1.000)	0.054	1.028 (1.009–1.046)	0.003
T stage (T2 vs T1)	0.930 (0.567–1.527)	0.774	1.308 (0.799–2.142)	0.285
T stage (T3/T4 vs T1)	0.355 (0.048–2.608)	0.308	3.153 (1.512–6.573)	0.002
Chemotherapy (Yes vs No)	-	-	0.833 (0.488–1.422)	0.504
Hormone therapy (Yes vs No)	-	-	1.644 (0.904–2.990)	0.103
Proliferation	2.737 (1.316–5.695)	0.007	0.843 (0.491–1.446)	0.535
IFN_gamma	0.829 (0.527–1.305)	0.418	1.345 (0.910–1.989)	0.137
TIDE	0.725 (0.519–1.014)	0.060	0.715 (0.509–1.004)	0.053
RSI	1.995 (0.297–13.395)	0.477	1.401 (0.329–5.969)	0.648
ARTIC	0.887 (0.636–1.238)	0.482	1.457 (1.077–1.972)	0.015
SMFS	2.888 (1.076–7.750)	0.035	0.755 (0.311–1.834)	0.536
IMS	1.050 (0.996–1.106)	0.069	0.984 (0.940–1.030)	0.492

Figures

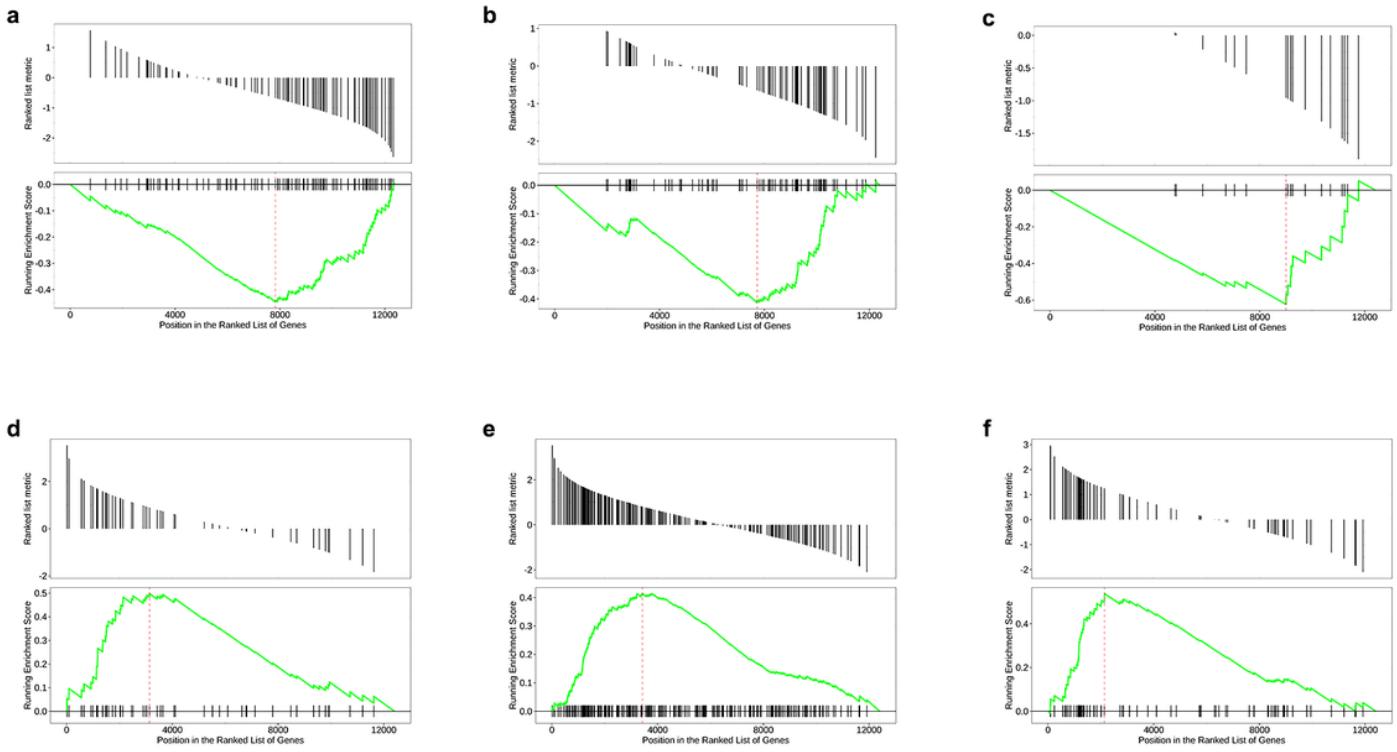


Figure 1

Representative significantly enriched pathways with z values of interaction term in Cox regression as loading quantity in GSEA analysis for CLT (a Antigen processing – Cross presentation, b Interferon gamma signaling, c PD-1 signaling; for M2 macrophage; d Collagen biosynthesis and modifying enzymes, e Extracellular matrix organization, f Integrin cell surface interactions)

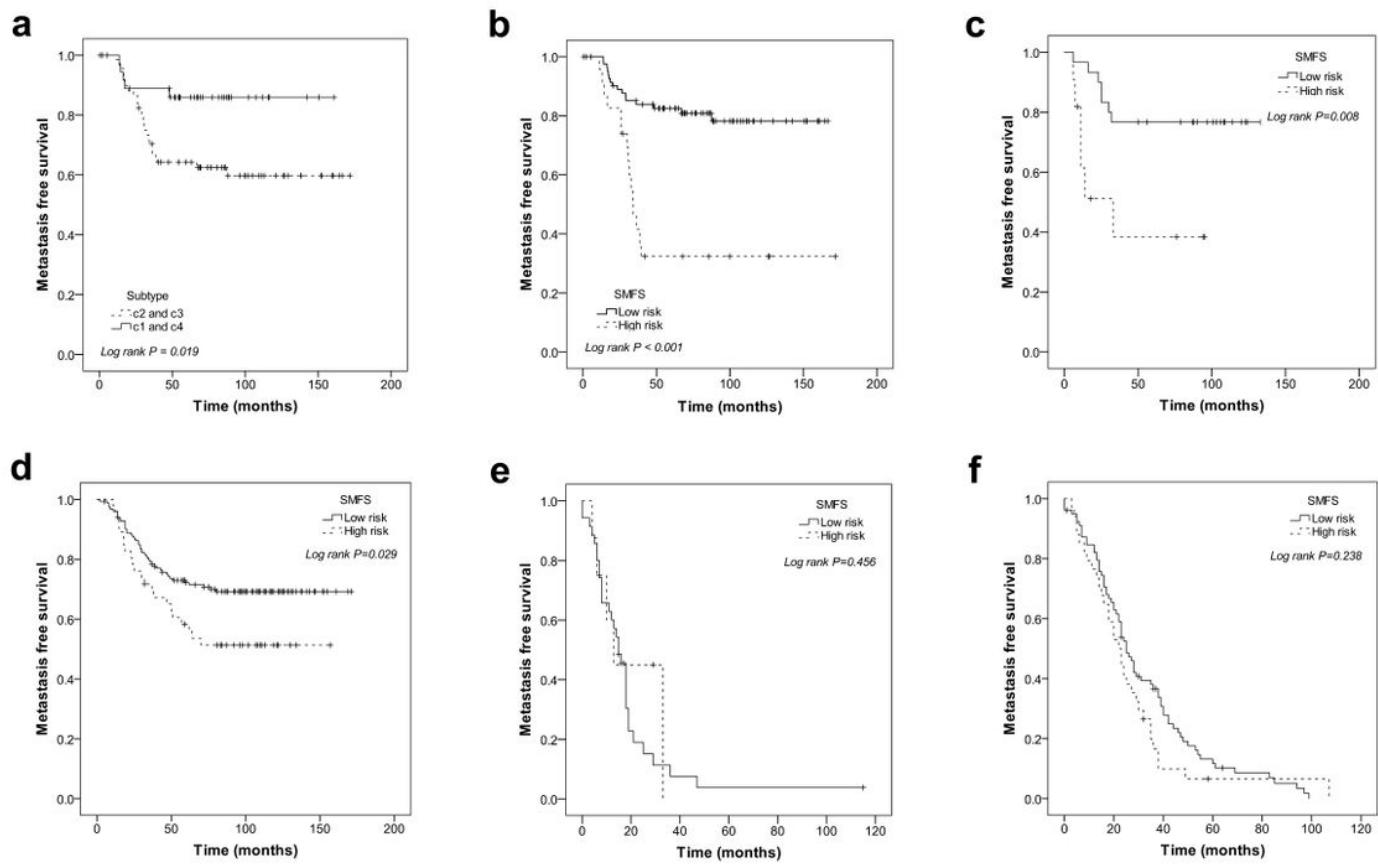


Figure 2

Kaplan-Meier curves showing efficiencies of various signatures or molecular subtypes in different subgroups (a TNBC molecular subtypes in GSE58812, b SMFS in GSE58812; c SMFS in TNBC subset in GSE2034, d SMFS in non-TNBC subset in GSE2034; e SMFS in TNBC subset in GSE12276, f SMFS in non-TNBC subset in GSE12276)

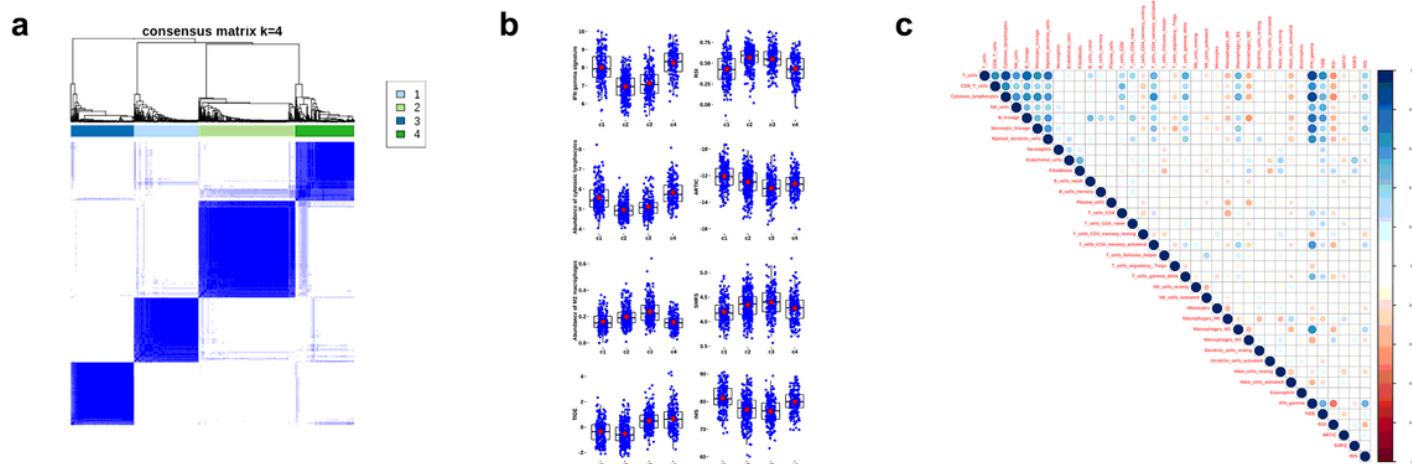


Figure 3

Consensus clustering of 920 TNBC samples (a consensus matrix showing clustering coincidence of individual samples; b distribution of IFN gamma, TIDE, RSI, ARTIC, IMS and SMFS across four subtypes in 920 TNBC; c correlation of IFN gamma, TIDE, RSI, ARTIC, IMS and SMFS in 920 TNBC)

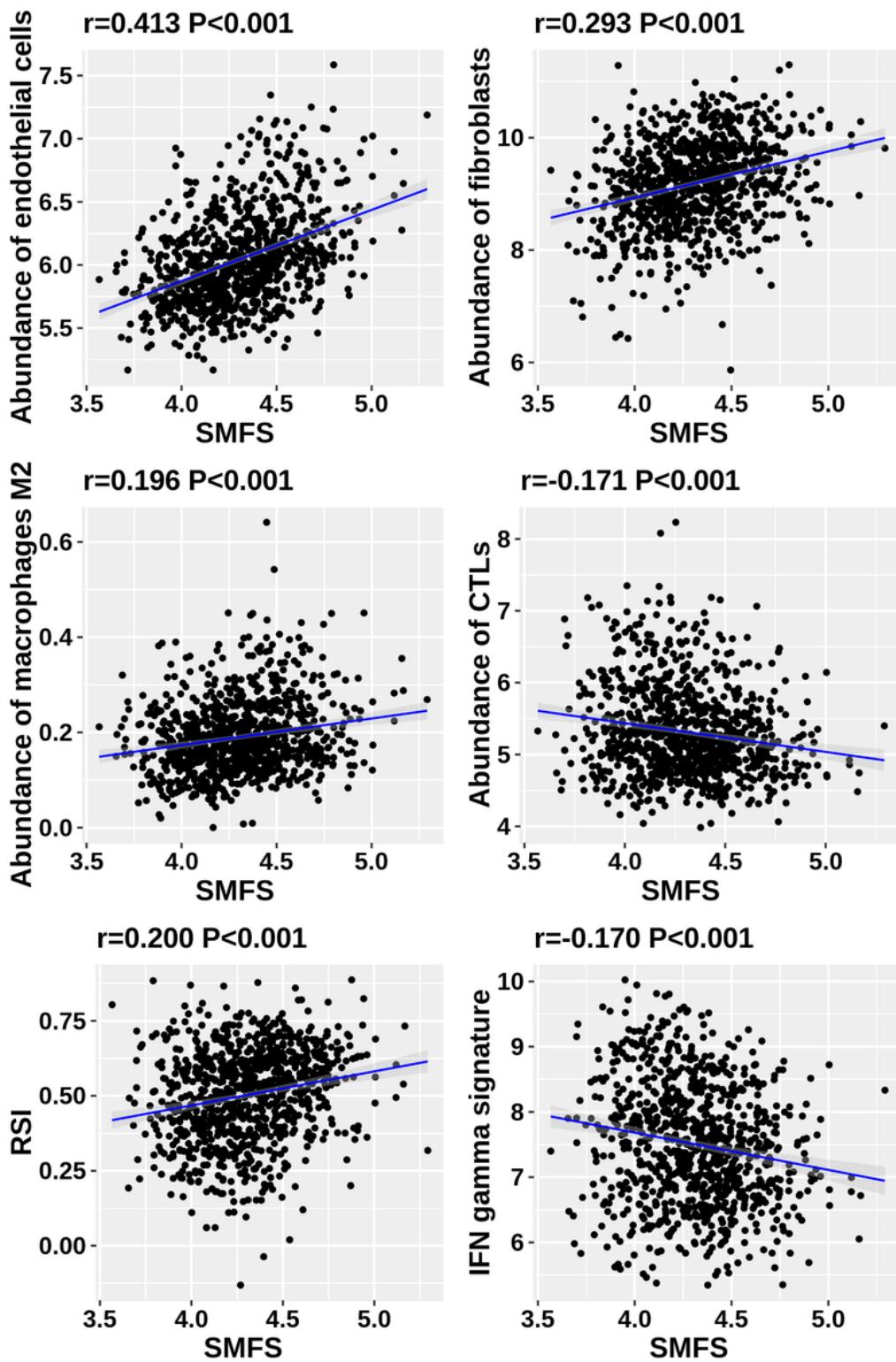


Figure 4

Scatter plots showing correlation of abundance of immune cells, IFN gamma, RSI with SMFS in all TNBC n=920

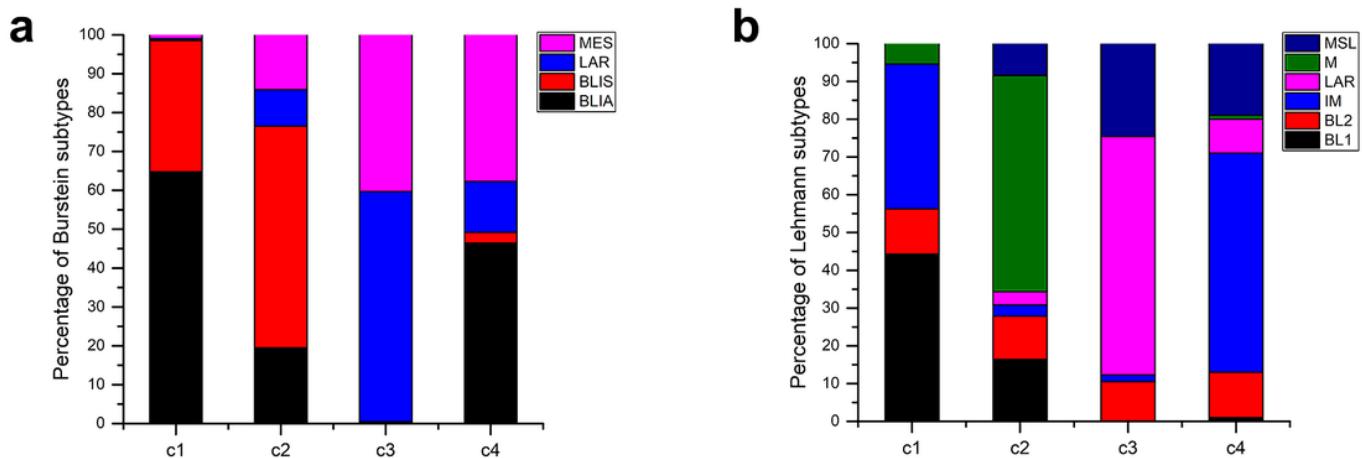


Figure 5

Stacked column charts showing compositions of each of four molecular subtypes in TNBC (a percentages of each subtype of Burstein subtyping, b percentages of each subtype of Lehmann subtyping)