



Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated With Prognosis

Jie Li¹, Cun Liu¹, Yi Chen², Chundi Gao¹, Miyuan Wang³, Xiaoran Ma¹, Wenfeng Zhang⁴, Jing Zhuang⁵, Yan Yao⁴ and Changgang Sun^{6*}

¹ College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China, ² Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden, ³ College of Management, Beijing University of Chinese Medicine, Beijing, China, ⁴ Clinical Medical Colleges, Weifang Medical University, Weifang, China, ⁵ Department of Oncology, Weifang Traditional Chinese Medicine Hospital, Weifang, Shandong, China, ⁶ Innovative Institute of Chinese Medicine and Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, China

OPEN ACCESS

Edited by:

Meng Zhou,
Wenzhou Medical University,
China

Reviewed by:

Emiel Janssen,
University of Stavanger,
Norway
Toshiharu Yamamoto,
Kanagawa Dental University,
Japan

*Correspondence:

Changgang Sun
scgdoctor@126.com

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Genetics

Received: 25 June 2019

Accepted: 16 October 2019

Published: 12 November 2019

Citation:

Li J, Liu C, Chen Y, Gao C, Wang M, Ma X, Zhang W, Zhuang J, Yao Y and Sun C (2019) Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated With Prognosis. *Front. Genet.* 10:1119. doi: 10.3389/fgene.2019.01119

There has been increasing attention on immune-oncology for its impressive clinical benefits in many different malignancies. However, due to molecular and genetic heterogeneity of tumors, the activities of traditional clinical and pathological criteria are far from satisfactory. Immune-based strategies have re-ignited hopes for the treatment and prevention of breast cancer. Prognostic or predictive biomarkers, associated with tumor immune microenvironment, may have great prospects in guiding patient management, identifying new immune-related molecular markers, establishing personalized risk assessment of breast cancer. Therefore, in this study, weighted gene co-expression network analysis (WGCNA), single-sample gene set enrichment analysis (ssGSEA), multivariate COX analysis, least absolute shrinkage, and selection operator (LASSO), and support vector machine-recursive feature elimination (SVM-RFE) algorithm, along with a series of analyses were performed, and four immune-related genes (*APOD*, *CXCL14*, *IL33*, and *LIFR*) were identified as biomarkers correlated with breast cancer prognosis. The findings may provide different insights into prognostic monitoring of immune-related targets for breast cancer or can be served as reference for the further research and validation of biomarkers.

Keywords: breast cancer, immune-related genes, prognosis, predictive biomarker, characterization

INTRODUCTION

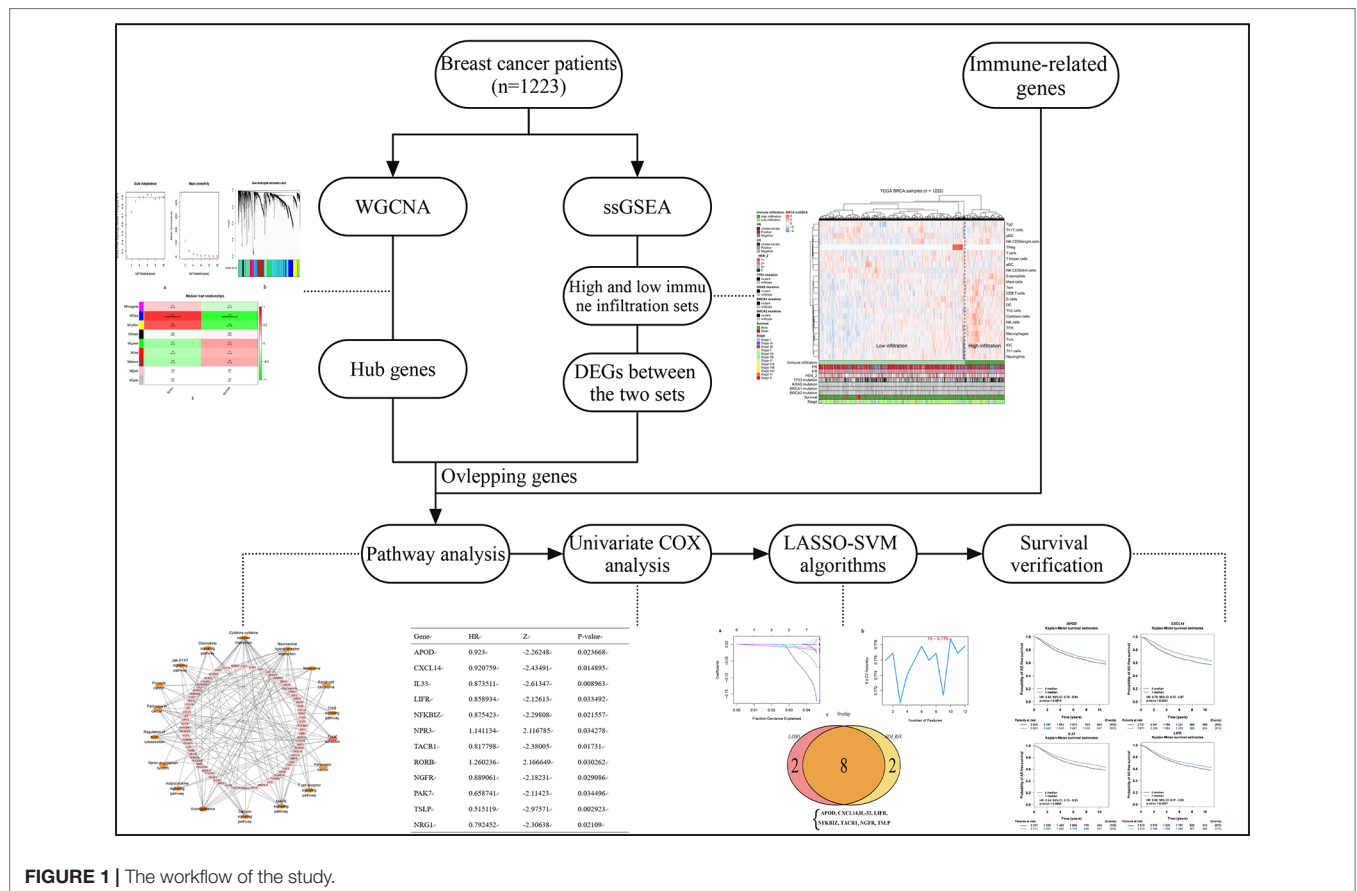
Recent analyses of the single-cell genome and transcriptome of breast cancer (BC) have provided insights into the prognosis between tumors and immune cells (Wagner et al., 2019). Large-scale atlas deepened our understanding of the recognition of tumors and their immune environments in breast cancer ecosystems. The immune system has been shown to be a determining factor in the development and progression of cancer. In fact, many reports have confirmed that immune cells and immune-related genes are attractive targets for regulating cancer progression (Elinav et al., 2013; Topalian et al., 2016; Nagarsheth et al., 2017). Notably, innate and adaptive immune cell infiltration are associated with responses to treatments and clinical outcomes, breast cancers with high immune infiltration to have

better prognosis (Atreya and Neurath 2008; Fridman et al., 2012; Fridman et al., 2017; Karn et al., 2017). This suggests that the application of immune-based prognostic features in breast cancer is a potential. Based on this, furthermore, it is also of high priority to consider the collaborative efforts to develop immunobiology research and search for prognostic indicators, so as to bring about the best prognostic evaluation (Nik-Zainal and Morganello 2017; Yates and Desmedt 2017).

Computational techniques applied to a large number of tumor gene expression profiles, based on immunological features, can rapidly provide a broader scale of intratumoral immune landscape (Bindea et al., 2013). Previously, tumor immune infiltration was primarily characterized by tissue-based methods such as flow cytometry and immunohistochemistry (IHC), both of which are limited by the amount of tissue required and the number of cell types assayed simultaneously. Currently, emerging bioinformatics resources are assisting these types of analyses. Large-scale public data with gene expression and clinical information, complete biological databases, and sophisticated high-throughput data analysis methods together provide opportunities for identifying broader prognostic features in breast cancer biology. With these approaches, Li et al. (2017) discovered individualized immune prognostic signature in Early-Stage Non-squamous Non-Small Cell Lung Cancer. Shen et al. (2019) have developed a generalized, individualized

immune prognostic signature that can stratify and predict ovarian cancer survival. Besides, in the research of Bindea et al. (2013), *CXCL13* and *IL21* are likely associated with long-term survival of patients and absence of tumor recurrence in human colorectal cancer (CRC).

Previous studies have revealed the relationship between immune cell levels and breast cancer prognosis, therefore, on this basis, the current study aimed to assess the association between immune infiltration and the genome in BC to reveal the effect of immune-relevant genes on the prognosis of breast cancer. Specifically, we explored the relationship of immune infiltration (measured by the expression characteristics of immune genes) in molecular level rather than cellular level with patients' prognosis. In this study, we performed a multi-perspectives, multi-dimensional analysis of a large number of breast cancer samples using massive bioinformatics and machine learning methods, such as weighted gene co-expression network analysis (WGCNA), single-sample gene set enrichment analysis (ssGSEA), differential analysis, univariate COX analysis, least absolute shrinkage and selection operator (LASSO), and support vector machine-recursive feature elimination (SVM-RFE) algorithm, functional analysis, and prognostic verification. Finally, four biomarkers of breast cancer were identified that were thought to be probably important prognostic features in breast cancer. The workflow is shown in **Figure 1**.



MATERIALS AND METHODS

Data Collection and Preprocessing

RNA-sequencing data, updated clinical data, and sample information of breast invasive cancer (BRCA) cohort were downloaded from the TCGA data portal (<https://portal.gdc.cancer.gov/>); fragments per kilobase million (FPKM) values were transformed into transcripts per kilobase million (TPM) values. Immune-related genes were collected from the ImmPort database (<http://www.immport.org>) and by tracking Tumor Immunophenotype (<http://biocc.hrbmu.edu.cn/TIP/index.jsp>); 24 immune cell type-specific gene signatures were adopted from Bindea et al. (2013).

Definition of BC-Related Genes by Weighted Gene Co-Expression Network Analysis (WGCNA)

First, Cluster 3.0 was used to remove the discrete values of transcripts with less than 50% expression, and differential expressed analysis was performed by R package limma ($\log_{2}FC \geq 2$ or < -0.5 and $p\text{-value} < 0.05$). Pearson's correlation analysis of the top 4,000 genes, based on standard deviation, was performed to construct a matrix of similarity (Ivliev et al., 2010). According to the power value which main affects the independence and the average degree of connectivity (k) of the co-expression modules, an adjacency matrix (AM) and a Topological overlap matrix (TOM) are obtained. The gradient method was used here and the power values ranged from 1 to 10. When the correlation between k and $p(k)$ reached 0.8, the optimal power value was determined to construct a scale-free topology network. Then TOM that measured the network connectivity of genes was transformed from AM (Botia et al., 2017). Modules were defined as branches of the hierarchical cluster tree generated based on TOM dissimilarity (Langfelder and Horvath 2008). For any given module, since the module Eigengenes (ME) offered the most appropriate interpretation of gene expression profile, we associated the ME with clinical features that refer to tumor or normal status in this study. Modules displaying high correlation (according to Module-trait relationships) were selected as further research objects, and genes of which were named BC-related genes.

Implementation of Single-Sample Gene Set Enrichment Analysis (ssGSEA)

Related infiltration and activity levels for 24 immune cell types, obtained from published signature gene lists across all tumor and normal samples, were quantified using the ssGSEA in R package GSVA (Hanzelmann et al., 2013). The signatures used in this study comprised of innate immunity, including natural killer (NK) cells, CD56dim NK cells, CD56bright NK cells, dendritic cells (DCs), plasmacytoid dendritic cells (pDC), immature DCs (iDC), activated DCs (aDC), neutrophils, mast cells, eosinophils, and macrophages, and adaptive immunity, including B cells, T cells, T central memory cells (T_{cm}), T effector memory (T_{em}), CD8 T cells, cytotoxic cells, T follicular helper (TFH), Th1, Th2, Th17, and Treg cells. The ssGSEA scores for each individual immune cell

type were standardized, and the qualitatively different tumor microenvironment cell infiltration patterns were grouped into high-infiltration and low-infiltration ones using hierarchical agglomerative clustering based on Euclidean distance and Ward's linkage. To identify genes associated with tumor microenvironment cell infiltration patterns, we calculated the differential expressed genes (DEGs) in the two sets using limma package.

Pathway Enrichment Analysis for the Molecular Function

Overlapping ones among the immune-related genes, BC-related genes and DEGs in high-infiltration versus low-infiltration groups, were subjected to perform enrichment function analysis using R package clusterProfiler to assess the molecular function.

Feature Selection by Least Absolute Shrinkage and Selection Operator (LASSO) and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) Algorithms

Univariate regression analysis was performed on overlapping genes obtained in *Section Implementation of Single-Sample Gene Set Enrichment Analysis (ssGSEA)* to select the survival-related genes. SVM-RFE and LASSO logistic regression were used to screen the characteristic genes. The LASSO algorithm was applied with the glmnet package (Friedman et al., 2010). Furthermore, SVM-RFE is a machine learning method based on support vector machine, which is used to find the best variables by deleting SVM-generated eigenvectors. SVM module was established to further identify the diagnostic value of these biomarkers in BC by e1071 package (Huang et al., 2014). Ultimately, we combined the genes from either LASSO or SVM-RFE algorithm for further analysis. A two-sided P value < 0.05 was considered to be statistically significant.

Verification of Survival Prediction

Next, we used external data to validate the survival results in Breast Cancer Gene-Expression Miner v4.2 (Jezequel et al., 2012), a database based on a large number of published studies. Kaplan–Meier survival curves with 2-sided log-rank test were drawn to evaluate OS differences between the two groups divided based on the median quantile expressions of genes.

GO Analysis of Single Gene

We carried out a holistic GO analysis of the screened biomarkers through the clusterProfiler package. Further, in order to understand the potential molecular function of each biomarker more specifically, guilt by association approach was used in this study, which means that the function of biomarkers could be inferred by performing a functional analysis of the genes that were significantly correlated with biomarkers (Oliver 2000). Pearson's correlation coefficient threshold was >0.5 .

RESULTS

Data Downloading and Collection

A total of 1,222 specimens, consisting of 1,109 cancer samples and 113 normal samples, were obtained from TCGA. A total 2,211 immune-related genes were collected from ImmPort database, TTI database, and from the report by Bindea et al. (2013).

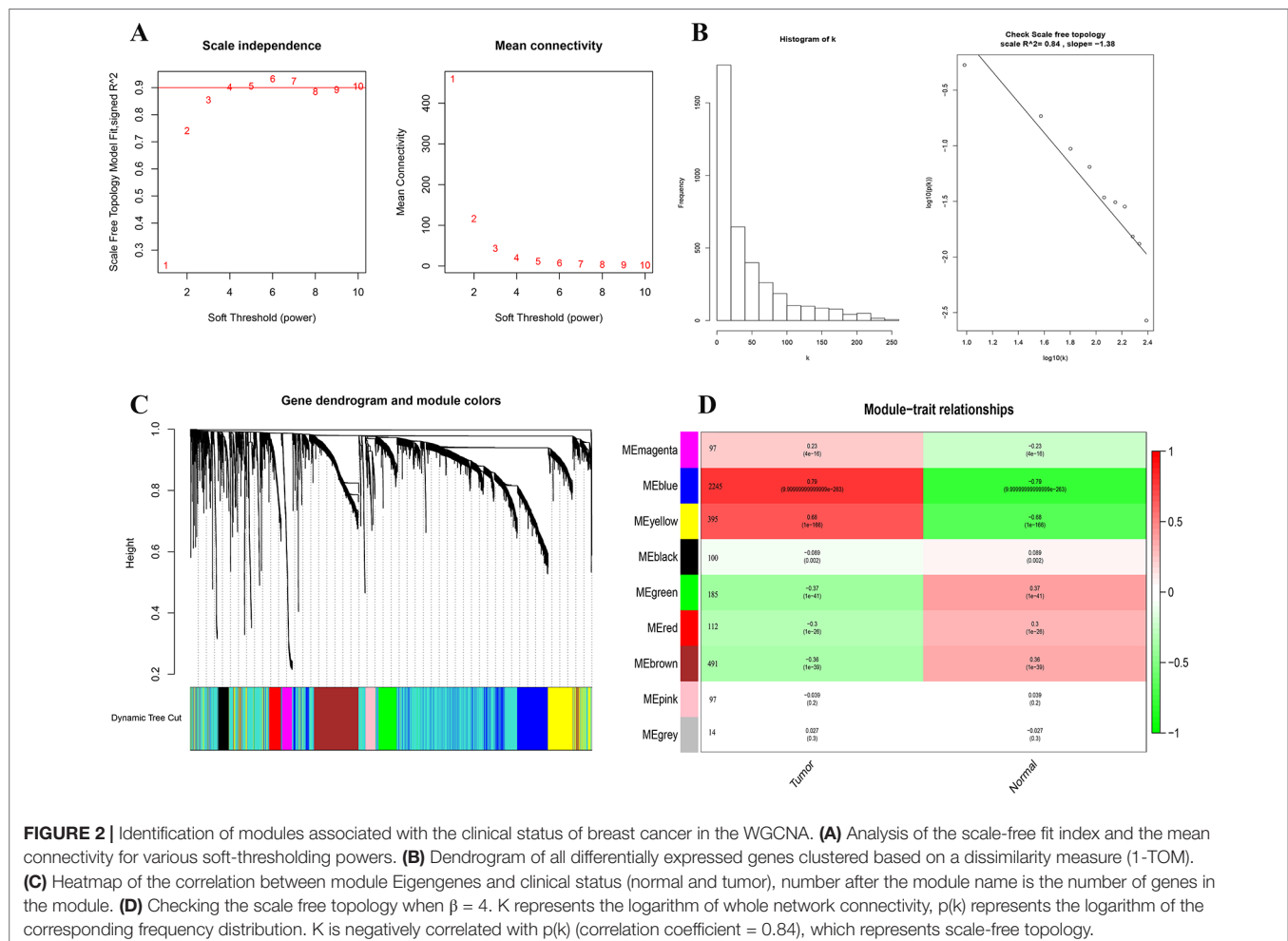
Determination of the Most Relevant Module Genes for BC

We first identified 5,058 differential expressed genes in 1,222 samples, and selected the top 4,000 (of 4,385) after sorting by standard deviation. Co-expression analysis was performed to construct a co-expression network. A total of 9 modules were identified *via* the average linkage hierarchical clustering. To achieve scale-free co-expression network, power of $\beta = 4$ was selected; to merge the highly similar modules, we chose a cut line of < 0.25 and minimum module size of 50, using the dynamic hybrid tree cut method. Blue and yellow modules (Module-trait relationships = 0.79 and 0.68, respectively) were found to have the highest association with tumor status (Figure 2), and hence, 2,629 genes in the two modules were considered to be significant module

genes for further analysis. Notably, our research differs from other studies about breast cancer using WGCNA method in that our aim is to apply WGCNA to discover hub genes related to immune.

High-Immune Infiltrated Genes Identified by ssGSEA

ssGSEA analysis was used to assess the immune infiltration level of every gene set, for every sample, by calculating separate enrichment scores for each sample and genome. We used ssGSEA scores from 24 immune-related gene signature extension groups to perform unsupervised clustering on breast cancer samples with transcriptome analysis data and clinical features (Figure 3A). This analysis clearly revealed two different clusters referred to herein as the high immune infiltration and low immune infiltration groups. Interestingly, TReg group are significantly different from the other groups. It showed that TReg has less quantified infiltration levels than other immune cells, so the calculated enrichment score in many samples tends to zero. However, focal point of our study is not on immune cells, so we are only focusing on their high and low infiltration groups divided and calculated by these immune feature genes of immune cells here. For further characterization, we performed



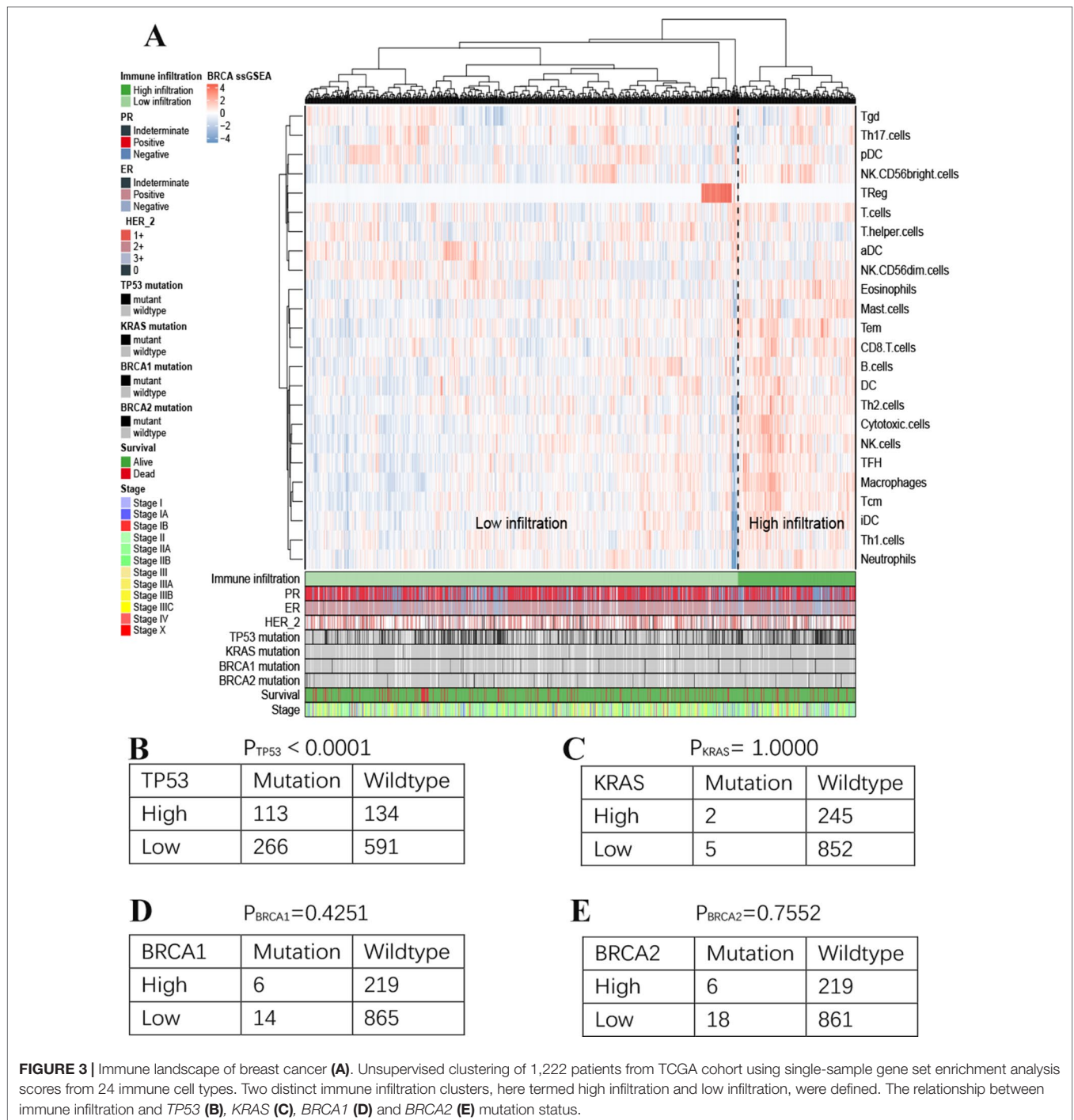


FIGURE 3 | Immune landscape of breast cancer (A). Unsupervised clustering of 1,222 patients from TCGA cohort using single-sample gene set enrichment analysis scores from 24 immune cell types. Two distinct immune infiltration clusters, here termed high infiltration and low infiltration, were defined. The relationship between immune infiltration and *TP53* (B), *KRAS* (C), *BRCA1* (D) and *BRCA2* (E) mutation status.

differential expressed analysis of the genes in high versus low immune infiltration, 2,951 genes were obtained and considered to be potentially associated with tumor immune microenvironment and prognostic effects probably.

In addition, we assessed the association between immune cell infiltration and the mutation status of *TP53*, *KRAS*, *BRCA1* and *BRCA2* (Figures 3B–E). Significant difference was observed in *TP53* (H = 113; L = 266 vs H = 134; L = 591, *TP53*-mutant vs

wildtype; $p < 0.001$). However, there was no relationship between other mutation status and immune cell infiltration.

Functional Annotation and Analysis

In total, 131 genes overlapped among BC-related genes, immune-related genes and differential immune infiltrated genes. To improve biological understanding of the genes identified in this study, we conducted enrichment analysis. As shown in Figure 4,

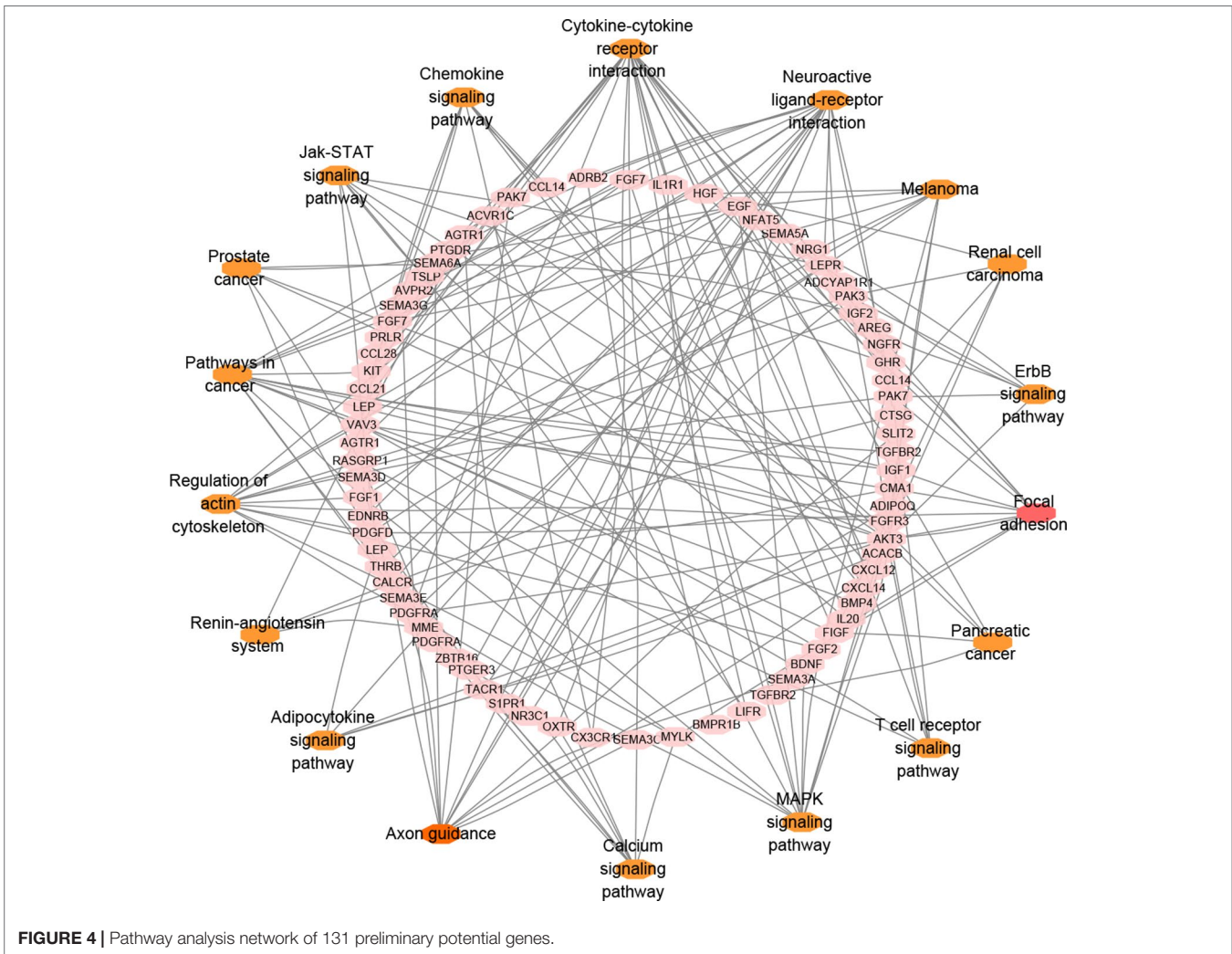


FIGURE 4 | Pathway analysis network of 131 preliminary potential genes.

some pathways like cytokine-cytokine receptor interaction, neuroactive ligand-receptor interaction, and other signaling pathways, were all quite significant in key cascades in the basic biology of cancer, such as initiation, growth, and recurrence of breast cancer (Hanahan and Weinberg 2011; Dees et al., 2012). Interestingly, “MAPK signaling pathway” has been reported to be essential for cancer-immune evasion in human cancer cells (Sumimoto et al., 2006). Moreover, there were many other pathways involved in specific cancers, such as in melanoma, pancreatic cancer, and renal cell carcinoma, many of which have shown optimistic response on immunotherapy.

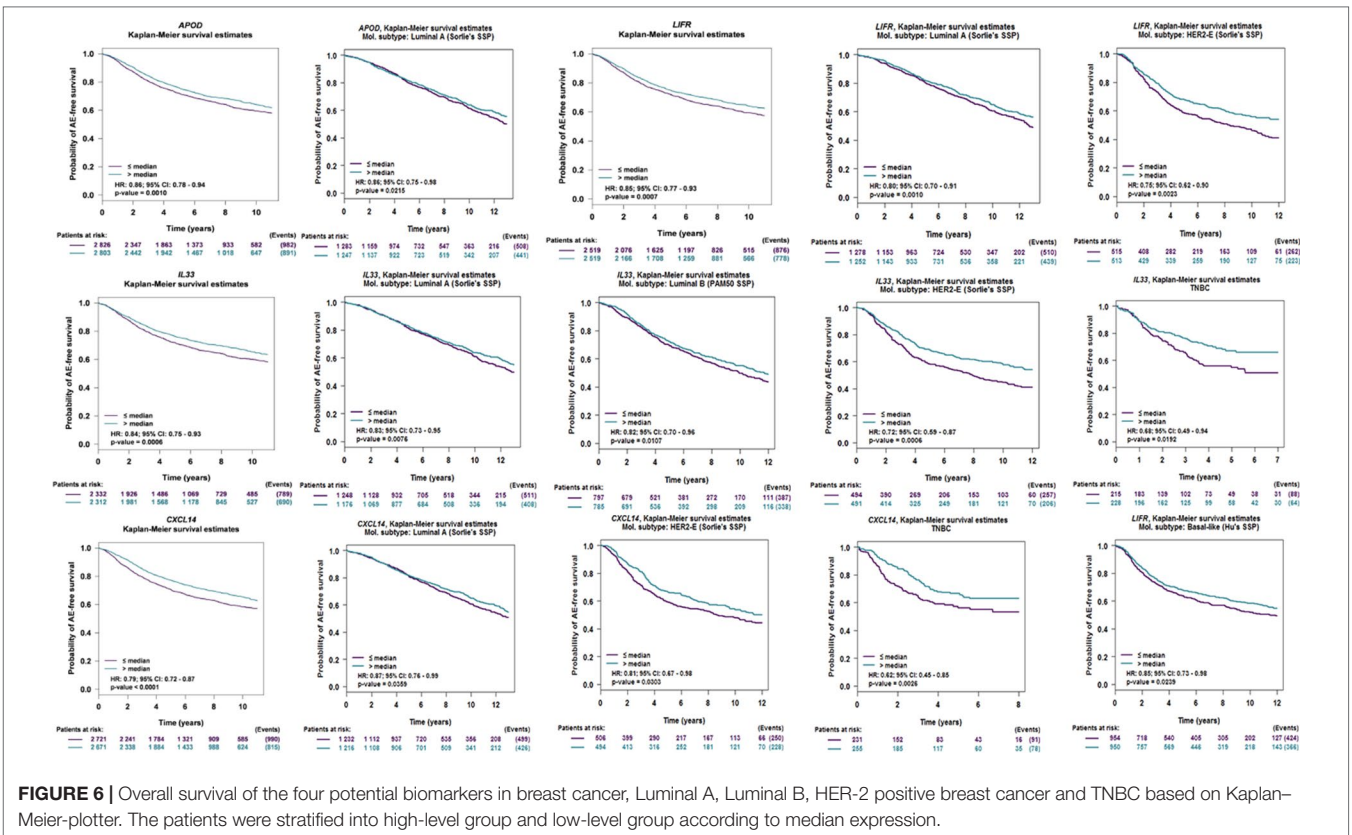
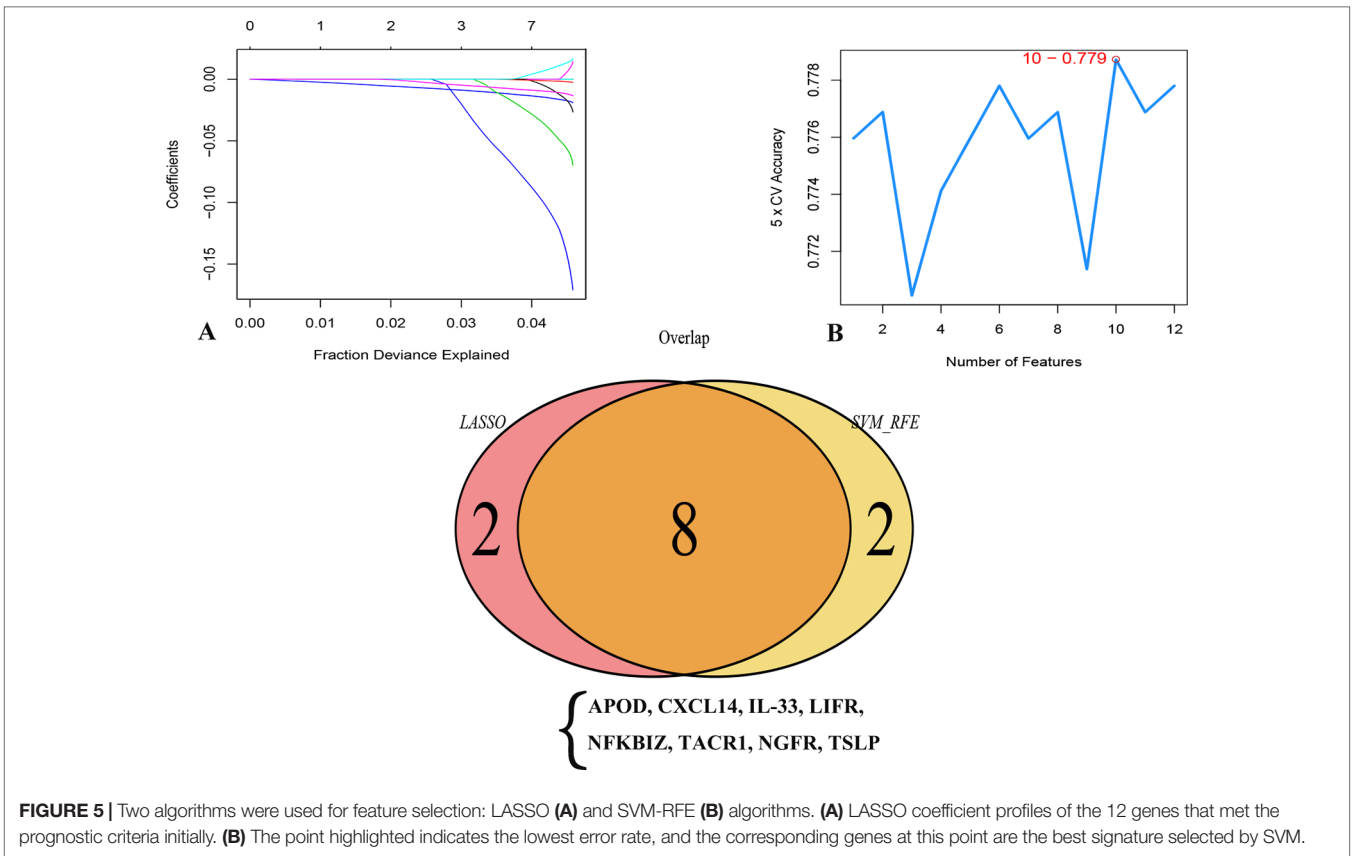
Preliminary Identification of Optimal Prognostic Biomarkers

In an attempt to confirm the independent prognostic impact of individual genes, we performed univariate COX regression of the 131 previously screened variables, and obtained 12 genes that met the prognostic criteria (Table 1); 10 genes were identified with the LASSO and SVM algorithms, respectively. Eight shared biomarkers for BC were defined by overlapping the biomarkers

TABLE 1 | The Univariate COX analysis of the signature.

Gene	HR	Z	P-value
<i>APOD</i>	0.923	-2.26248	0.023668
<i>CXCL14</i>	0.920759	-2.43491	0.014895
<i>IL33</i>	0.873511	-2.61347	0.008963
<i>LIFR</i>	0.858934	-2.12613	0.033492
<i>NFKBIZ</i>	0.875423	-2.29808	0.021557
<i>NPR3</i>	1.141134	2.116785	0.034278
<i>TACR1</i>	0.817798	-2.38005	0.01731
<i>RORB</i>	1.260236	2.166649	0.030262
<i>NGFR</i>	0.889061	-2.18231	0.029086
<i>PAK7</i>	0.658741	-2.11423	0.034496
<i>TSLP</i>	0.515119	-2.97571	0.002923
<i>NRG1</i>	0.792452	-2.30638	0.02109

derived from these two algorithms (Figure 5). These 8 genes, associated with the prognosis of BC, are considered to be the best features, and included apolipoprotein D (*APOD*), Chemokine (C-X-C motif) ligand 14 (*CXCL14*), Interleukin-33 (*IL33*), leukemia inhibitory factor receptor (*LIFR*), nuclear factor kappa B inhibitor zeta (*NFKBIZ*), tachykinin-1 receptor (*TACR1*),



nerve growth factor receptor (*NGFR*), and thymic stromal lymphopoietin genes (*TSLP*).

Survival Verification by External Data

We put the 8 genes into a web-site (bc-GenExMiner V 4.1) to perform survival analysis. Based on the validation of larger external sample data, our results showed that *APOD*, *CXCL14*, *IL33*, and *LIFR* displayed good prognostic significance (Figure 6). Further, we performed prognostic analysis of breast cancer subtypes, and the results indicated that *APOD* shows good prognostic value in luminal A breast cancer, *IL33* is related to prognosis of Luminal A, Luminal B, HER-2 positive breast cancer and Triple negative breast cancer (TNBC). *CXCL14* and *LIFR* are associated with the prognosis of Luminal A, HER-2 positive breast cancer and TNBC. Therefore, we used these 4 genes as targets for further analyses.

Potential Significance of the Biomarkers Identified

To ascertain the possible mechanisms of genes affecting BC progression, GO analysis of each biomarker was conducted separately. Molecular functions of the four genes are mainly related to the production, binding, or migration of chemokines, lymphocytes, cytokines, and receptors, which are closely associated not only with the growth process of cancer cells, but also with the immune environment.

DISCUSSION

Although breast cancer is not considered to be a particularly immunogenic tumor compared to melanoma and renal cell carcinoma, molecular profiling of breast tumors suggests that it does show certain levels of immunoregulatory gene activation (Ascierto et al., 2012). Multiple investigators have reported that immune-related genes can be used as prognostic indicators for breast cancer (Iwamoto et al., 2011; Gingras et al., 2015). The increasing body of evidence demonstrating the importance of biomarkers, especially genes, in determining cancer outcomes provides new opportunities to integrate this information into therapeutic algorithms (Li et al., 2017).

Previous studies about tumors and immune infiltration mostly focused on immune cell types. For example, Li B et al. (2016) infers the abundance of the six immune cell types (B cells, CD4 T cells, CD8 T cells, neutrophil, macrophage, and dendritic cells) using approach of constrained least squares fitting and found many significant associations between immune cell abundance and outcome of 23 cancer types patients. For instance, except for association with prolonged survival of patients, CD8 T cells may also play an important role in preventing tumor recurrence (in melanoma and colorectal cancer and cervical cancer). Notably, the focus of our research is to estimate the degree of 24 immune cells infiltration using ssGSEA, and finally determined the four most promising immune-related genes associated with prognosis.

The original design of this study was to use various bioinformatics tools and databases to identify useful and

potential immune-related targets for breast cancer associated with prognostic outcomes. Based on that, in this study, ssGSEA was implemented to calculate the immune cell infiltration levels for each sample in order to determine the two infiltration groups. Since previous studies had shown immune-infiltration to have better prognosis in breast cancer, we analyzed the differential genes in the high and low groups. To identify the genes most relevant to breast cancer, we performed a WGCNA analysis to select the modules with the strongest correlation between the module traits and genes in the modules. At the same time, we collected immune-related genes from various databases and literatures; the overlapping 131 genes of the above three clusters were identified as the immune strongly related genes associated with breast cancer. Functional analysis showed these 131 genes to be closely associated with the development of breast cancer and immune escape, such as *via* cytokine–cytokine receptor interaction and MAPK signaling pathway. Next, unlike previous studies that selected markers only by one algorithm, our current study chose a combination strategy to minimize the possibility of losing important markers through incorporating genes from three distinct algorithms. Based on univariate regression analysis determining the genes associated with prognosis, LASSO and SVM-RFE algorithms were performed to screen eight characteristic variables. After extensive survival validation tests with external data, four of eight genes (*APOD*, *CXCL14*, *IL33*, and *LIFR*) were successfully identified as prognostic markers and the other four were filtered out because they did not show good prognostic differences in validation.

Besides, we tested whether the infiltration of immune cells is associated distinct mutations, and the results indicated that *TP53* showed relatively obvious difference between the mutant and the wild type. These differences may illustrate that *TP53* mutation may affect the immune microenvironment of breast cancer to a certain extent (the specific mechanism is not clear here), this is consistent with the conclusion of a recent study (Liu Z et al., 2019).

Further, functional analysis of the four identified biomarkers showed them to be mainly related to the production, metastasis, or regulation of chemokines, cytokine–cytokine interaction, lymphocytes, and leukocytes, which contribute to the progress of cancer and immunity. Specifically, chemokines and cytokines can lead to recruitment of multiple cell types, including different immune cell subsets, into tumors, thereby involving in angiogenesis, inflammation, tumor growth, invasion, metastasis, and anti-tumor immunity (Motz and Coukos 2011).

CXCL14 (earlier designated as *BRAK*, *MIP-2γ*, *BMAC*, or *KSI*), was identified in this study as one of the chemokines that had been previously shown to play a key role in breast cancer. For example, in the animal models of Allinen et al. (2004) *CXCL14* was shown to be overexpressed in tumor myoepithelial cells; it binds to receptors located on epithelial cells and increases the proliferation, migration, and invasion of BC. However, in patients, *CXCL14* tends to show a more favorable correlation between protein level and overall BC survival without considering the cell type in charge of *CXCL14* expression, as was consistent with our findings (Gu et al., 2012). In addition, some studies have suggested that it displays chemotactic activity

towards various immune cells, including B-cells, NK cells, and monocytes (Shellenberger et al., 2004). This also consolidates and complements our functional analysis results. *IL33* is a member of the *IL1* cytokine family, and has been suggested to play a pro-tumoral role in BC. For example, several research found that *IL33* facilitates the expansion of immunosuppressive myeloid, ILC2, and Treg cells, and reduces activated NK cells, thereby enhancing the growth of BC, and development of lung and liver metastases (Xiao et al., 2016; Afferni et al., 2018). Another study pointed that high *IL33* value in 4T1 breast cancer cells reduces tumor growth and metastasis *in vivo* (Jovanovic et al., 2014).

Our findings are particularly notable considering that recent studies have reported *LIFR*, the receptor of leukemia inhibitory factor (LIF), as a prognostic marker of BC for clinical outcomes. This protein is upregulated in normal breast tissues, but its high expression in tumor tissues has inverse association with lymph node metastasis. Notably, down regulation or loss of *LIFR* is related to poor prognosis in most nonmetastatic stage I–III breast cancer (Chen et al., 2012). *APOD* is a member of the lipocalin family of proteins, seen in patients with BC; analysis of the relationship between *APOD* expression and clinical outcome demonstrates significant relevance of the low expression of *APOD* with a poor overall survival (Diez-Itza et al., 1994). In previously two small-scale studies of Soiland H et al., the association between *APOD* and recurrence is only found in specific age groups, especially disease-specific survival in elderly, comorbid patients not receiving chemotherapy (Soiland et al., 2009a; Soiland et al., 2009b). However, in a large study, the results of Soiland H et al. could not be verified. At the same time, Klebaner et al. (2017). believe that the true association between *APOD* expression and recurrence may be ineffective in ER+ breast cancer patients treated with tamoxifen. Our study did not consider the patients' different treatment, and it showed *APOD* expression has a good prognosis value in Luminal A breast cancer. This also suggests that all the markers including *APOD* may need to be further explored in patients receiving different treatments.

Interestingly, there is no report of *IL33* as a prognostic marker of BC, but consistency of our findings regarding *CXCL14*, *LIFR*, and *APOD* with previous studies indicates our method to be reliable, and thus supports the reliability of our conclusions to a large extent.

At present, there are several treatment methods for targeting tumor-associated immune genes, and have achieved good results in clinical or clinical trial. For example, blocking macrophage

chemokines such as *CXCL12*, and preventing macrophages from entering tumors, or fighting for macrophages to kill tumor cells (Gholamin and Mitra 2017). Although focal point of our research is not on the mechanism of action of immune cells, it adds strong evidence that tumor-associated immune genes may become potential targets for cancer therapy. Previous studies have been based on specific layers of immune cells such as tumor infiltrating lymphocytes or developing prognostic models. Our study focuses on immune-related genes and uses strict standard layer screening to obtain genes which may be potential prognosis targets of BC. But our research also has certain limitations, clinical trials of checkpoint inhibitors have indicated immune infiltration to be critical for tumor regression, and that the quality of immune response is a key factor for successful treatment, so substantial clinical trials still need to be carried out.

In conclusion, using bioinformatics and machine learning methods, we focused on immune-related genes of breast cancer based on large data of real samples, and screened four potential prognostic targets, three of which were confirmed in previous studies. Although the four prognostic markers identified in our current study may still need a lot of clinical trials for validation, they may provide some clues and landscape for the prognosis assessment of breast cancer.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The data can be found in <https://portal.gdc.cancer.gov/>, <http://www.immport.org>, <http://biocc.hrbmu.edu.cn/TIP/index.jsp>.

AUTHOR CONTRIBUTIONS

JL and CS conceived of the presented idea. CL, XM, and WZ contributed to data collection in consultation with JL. CG, JZ, and YY carried out the analysis. JL and CS wrote the manuscript. YC and MW contributed to the manuscript revision. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work is supported by the grants from National Natural Science Foundation of China (81673799) and National Natural Science Foundation of China Youth Fund (81703915).

REFERENCES

- Afferni, C., Buccione, C., Andreone, S., Galdiero, M. R., Varricchi, G., Marone, G., et al. (2018). The pleiotropic immunomodulatory functions of IL-33 and its implications in tumor immunity. *Front. Immunol.* 9, 2601. doi: 10.3389/fimmu.2018.02601
- Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., et al. (2004). Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 6, 17–32. doi: 10.1016/j.ccr.2004.06.010
- Ascierto, M. L., Kmiecik, M., Idowu, M. O., Manjili, R., Zhao, Y., Grimes, M., et al. (2012). A signature of immune function genes associated with recurrence-free survival in breast cancer patients. *Breast Cancer Res. Treat* 131, 871–880. doi: 10.1007/s10549-011-1470-x
- Atreya, I., and Neurath, M. F. (2008). Immune cells in colorectal cancer: prognostic relevance and therapeutic strategies. *Expert Rev. Anticancer Ther.* 8, 561–572. doi: 10.1586/14737140.8.4.561
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal

- the immune landscape in human cancer. *Immunity* 39, 782–795. doi: 10.1016/j.immuni.2013.10.003
- Botia, J. A., Vandrovicova, J., Forabosco, P., Guelfi, S., D'Sa, K., Hardy, J., et al. (2017). An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* 11, 47. doi: 10.1186/s12918-017-0420-6
- Chen, D., Sun, Y., Wei, Y., Zhang, P., Rezaeian, A. H., Teruya-Feldstein, J., et al. (2012). LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat. Med.* 18, 1511–1517. doi: 10.1038/nm.2940
- Dees, N. D., Zhang, Q., Kandath, C., Wendl, M. C., Schierding, W., Koboldt, D. C., et al. (2012). MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* 22, 1589–1598. doi: 10.1101/gr.134635.111
- Diez-Itza, I., Vizoso, F., Merino, A. M., Sanchez, L. M., Tolivia, J., Fernandez, J., et al. (1994). Expression and prognostic significance of apolipoprotein D in breast cancer. *Am. J. Pathol.* 144, 310–320.
- Elinav, E., Nowarski, R., Thaiss, C. A., Hu, B., Jin, C., and Flavell, R. A. (2013). Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat. Rev. Cancer* 13, 759–771. doi: 10.1038/nrc3611
- Fridman, W. H., Pages, F., Sautes-Fridman, C., and Galon, J. (2012). The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* 12, 298–306. doi: 10.1038/nrc3245
- Fridman, W. H., Zitvogel, L., Sautes-Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. *Nat. Rev. Clin. Oncol.* 14, 717–734. doi: 10.1038/nrclinonc.2017.101
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Software* 33, 1–22. doi: 10.18637/jss.v033.i01
- Gholamin, S., and Mitra, S. S. (2017). Disrupting the CD47-SIRPalpha anti-phagocytic axis by a humanized anti-CD47 antibody is an efficacious treatment for malignant pediatric brain tumors 9 (381), eaf2968. doi: 10.1126/scitranslmed.aaf2968
- Gingras, I., Azim, H. A. Jr., Ignatiadis, M., and Sotiriou, C. (2015). Immunology and breast cancer: toward a new way of understanding breast cancer and developing novel therapeutic strategies. *Clin. Adv. Hematol. Oncol.: H&O* 13, 372–382.
- Gu, X. L., Ou, Z. L., Lin, F. J., Yang, X. L., Luo, J. M., Shen, Z. Z., et al. (2012). Expression of CXCL14 and its anticancer role in breast cancer. *Breast Cancer Res. Treat* 135, 725–735. doi: 10.1007/s10549-012-2206-2
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013
- Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 14, 7. doi: 10.1186/1471-2105-14-7
- Huang, M. L., Hung, Y. H., Lee, W. M., Li, R. K., and Jiang, B. R. (2014). SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *Sci. World J.* 2014, 795624. doi: 10.1155/2014/795624
- Ivliev, A. E., t Hoen, P. A., and Sergeeva, M. G. (2010). Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Res.* 70, 10060–10070. doi: 10.1158/0008-5472.can.10-2465
- Iwamoto, T., Bianchini, G., Booser, D., Qi, Y., Coutant, C., Shiang, C. Y., et al. (2011). Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. *J. Natl. Cancer Institute* 103, 264–272. doi: 10.1093/jnci/djq524
- Jezequel, P., Campono, M., Gouraud, W., Guerin-Charbonnel, C., Leux, C., Ricolleau, G., et al. (2012). bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res. Treat* 131, 765–775. doi: 10.1007/s10549-011-1457-7
- Jovanovic, I. P., Pejnovic, N. N., Radosavljevic, G. D., Pantic, J. M., Milovanovic, M. Z., Arsenijevic, N. N., et al. (2014). Interleukin-33/ST2 axis promotes breast cancer growth and metastases by facilitating intratumoral accumulation of immunosuppressive and innate lymphoid cells. *Int. J. Cancer* 134, 1669–1682. doi: 10.1002/ijc.28481
- Karn, T., Jiang, T., Hatzis, C., Sanger, N., El-Balat, A., Rody, A., et al. (2017). Association between genomic metrics and immune infiltration in triple-negative breast cancer. *JAMA Oncol.* 3, 1707–1711. doi: 10.1001/jamaoncol.2017.2140
- Klebaner, D., Hamilton-Dutoit, S., Ahern, T., Crawford, A., Jakobsen, T., Cronin-Fenton, D. P., et al. (2017). Apolipoprotein D expression does not predict breast cancer recurrence among tamoxifen-treated patients. *PLoS One* 12, e0171453. doi: 10.1371/journal.pone.0171453
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559
- Li, B., Severson, E., Pignon, J. C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 17, 174. doi: 10.1186/s13059-016-1028-7
- Li, B., Cui, Y., Diehn, M., and Li, R. (2017). Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol.* 3, 1529–1537. doi: 10.1001/jamaoncol.2017.1609
- Liu, Z., Jiang, Z., Gao, Y., Wang, L., Chen, C., and Wang, X. (2019). TP53 mutations promote immunogenic activity in breast cancer. *J. Oncol.* 2019, 5952836. doi: 10.1155/2019/5952836
- Motz, G. T., and Coukos, G. (2011). The parallel lives of angiogenesis and immunosuppression: cancer and other tales. *Nat. Rev. Immunol.* 11, 702–711. doi: 10.1038/nri3064
- Nagarsheth, N., Wicha, M. S., and Zou, W. (2017). Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat. Rev. Immunol.* 17, 559–572. doi: 10.1038/nri.2017.49
- Nik-Zainal, S., and Morganella, S. (2017). Mutational signatures in breast cancer: the problem at the DNA level. *Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res.* 23, 2617–2629. doi: 10.1158/1078-0432.ccr-16-2810
- Oliver, S. (2000). Guilt-by-association goes global. *Nature* 403, 601–603. doi: 10.1038/35001165
- Shellenberger, T. D., Wang, M., Gujrati, M., Jayakumar, A., Strieter, R. M., Burdick, M. D., et al. (2004). BRAK/CXCL14 is a potent inhibitor of angiogenesis and a chemotactic factor for immature dendritic cells. *Cancer Res.* 64, 8262–8270. doi: 10.1158/0008-5472.can-04-2056
- Shen, S., Wang, G., Zhang, R., Zhao, Y., Yu, H., Wei, Y., et al. (2019). Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. *EBioMedicine* 40, 318–326. doi: 10.1016/j.ebiom.2018.12.054
- Soiland, H., Janssen, E. A., Korner, H., Varhaug, J. E., Skaland, I., Gudlaugsson, E., et al. (2009a). Apolipoprotein D predicts adverse outcome in women > or = 70 years with operable breast cancer. *Breast Cancer Res. Treat* 113, 519–528. doi: 10.1007/s10549-008-9955-y
- Soiland, H., Skaland, I., Varhaug, J. E., Korner, H., Janssen, E. A., Gudlaugsson, E., et al. (2009b). Co-expression of estrogen receptor alpha and Apolipoprotein D in node positive operable breast cancer—possible relevance for survival and effects of adjuvant tamoxifen in postmenopausal patients. *Acta Oncol.* 48:4, 514–521. doi: 10.1080/02841860802620613
- Sumimoto, H., Imabayashi, F., Iwata, T., and Kawakami, Y. (2006). The BRAF-MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *J. Exp. Med.* 203, 1651–1656. doi: 10.1084/jem.20051848
- Topalian, S. L., Taube, J. M., Anders, R. A., and Pardoll, D. M. (2016). Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat. Rev. Cancer* 16, 275–287. doi: 10.1038/nrc.2016.36
- Wagner, J., Rapsomaniki, M. A., Chevrier, S., Anzeneder, T., Langwieder, C., Dykgers, A., et al. (2019). A single-cell atlas of the tumor and immune ecosystem of human breast. *Cancer Cell* 177, 1330–1345. doi: 10.1016/j.cell.2019.03.005
- Xiao, P., Wan, X., Cui, B., Liu, Y., Qiu, C., Rong, J., et al. (2016). Interleukin 33 in tumor microenvironment is crucial for the accumulation and function of myeloid-derived suppressor cells. *Oncoimmunology* 5, e1063772. doi: 10.1080/2162402x.2015.1063772
- Yates, L. R., and Desmedt, C. (2017). Translational genomics: practical applications of the genomic revolution in breast cancer. *Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res.* 23, 2630–2639. doi: 10.1158/1078-0432.ccr-16-2548

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Liu, Chen, Gao, Wang, Ma, Zhang, Zhuang, Yao and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.