Original Article

Mining the prognostic significance and immune infiltration of STAT family members in human breast cancer by bioinformatics analysis

Yudong Zhou^, Siyuan Jiang, Shibo Yu, Lizhe Zhu, Yang Liu, Shouyu Li, Na Hao, Yu Ren

Department of Breast Surgery, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Contributions: (I) Conception and design: Y Zhou, N Hao; (II) Administrative support: N Hao, Y Ren; (III) Provision of study materials or patients: Y Zhou, Y Liu, S Jiang, S Li; (IV) Collection and assembly of data: Y Zhou, S Jiang, L Zhu, S Yu; (V) Data analysis and interpretation: Y Zhou, S Yu, L Zhu, Y Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Na Hao; Yu Ren. 277 West Yanta Road, Xi'an 710061, China. Email: haona0318@163.com; renyyyyy@126.com.

Background: Growing evidence proved that signal transducer and activators of transcription (*STAT*) proteins are cytoplasmic transcription factors known to play key roles in many cellular biological processes and may be prognostic predictors of some cancers. However, the role of each *STAT* family members in breast cancer (BRCA) is diverse and controversial. This study aimed to systematic mine the prognostic significance and immune infiltration of *STAT* family member in human BRCA.

Methods: Based on The Cancer Genome Atlas (TCGA) database, we used the Oncomine, Gene Expression Profiling Interactive Analysis (GEPIA) and The Human Protein Atlas to analyze the expression of *STAT* family members in normal human breast and tumor tissues. The Kaplan-Meier Plotter, GEPIA and PrognoScan were utilized to assess the prognostic value of different *STATs* in BRCA. Then we used the cBioPortal, STRING, GeneMANIA and Metascape to make further mutation analysis, protein-protein interaction (PPI) analysis and subsequent functional enrichment analysis. Finally, the "ESTIMATE" and "ggcorrplot" package of R 17 software were used for immune infiltration analysis.

Results: *STAT2* [P<0.01, hazard ratio (HR) =1.23, 95% confidence interval (CI): 1.07–1.42] and *STAT3* (P=0.018, HR =0.69, 95% CI: 0.51–0.94) could be an independent risk factor for predicting overall survival (OS). *STAT4* could be used as an independent predictor of distant metastasis-free survival in BRCA based on both GSE19615 (P=0.021, HR =0.21, 95% CI: 0.06–0.79) and GSE2034 (P=0.015, HR =0.57, 95% CI: 0.37–0.90) datasets. Meanwhile, *STAT5A*, *STAT5B* and *STAT6* also have been shown to independently predict the prognosis of BRCA. Additionally, the functional mechanisms of *STAT4* co-expressed genes were mainly focused on immune-related pathways and its expression was associated with immune checkpoint-associated genes and immunomodulators in BRCA.

Conclusions: Our study mined the prognostic significance of *STAT* family members in BRCA and their correlation with immune infiltration. The results suggest that individual *STATs*, except *STAT1*, may act as a prognostic biomarker for BRCA and provide a reference for further potential immunotherapies.

Keywords: Signal transducer and activator of transcription (*STAT*); overall survival (OS); immune infiltration; breast cancer (BRCA)

Submitted Feb 22, 2022. Accepted for publication Apr 15, 2022.
doi: 10.21037/gs-22-189
View this article at: https://dx.doi.org/10.21037/gs-22-189

^ ORCID: 0000-0002-6314-1463.

Introduction

Breast cancer (BRCA) is currently one of the leading causes of cancer-related death in women worldwide, and the incidence rate is still increasing (1). At present, there are multiple of treatments for BRCA, including surgery, chemotherapy, radiotherapy, and endocrine therapy (2,3). Also, increasing numbers of researches have shown that the immune system plays an important role in BRCA, such as detecting the response to standard treatment or predicting the long-term survival for BRCA patients (4). Meanwhile, with the discovery of immune checkpoints, which regulate immune activation, has led to new advances for the treatment of BRCA (5). Besides, many prognostic models based on multigene testing already exist in clinical practice, such as OncotypeDX, EndoPredict, the Breast Cancer Indicator, etc. (6). However, problems of high cost and poor reproducibility continue to limit their practical clinical application. Therefore, identifying some dependable prognostic biomarkers and performing a comprehensive analysis to assess its roles in the process of tumorigenesis is still necessary for BRCA patients.

Signal transducer and activator of transcription (*STAT*) family members are transcriptional factors that were first discovered by Wegenka *et al*. in 1994 (7). Seven *STAT* family members have been confirmed in the human genome: *STAT1* (chromosomal location: 2q32.2), *STAT2* (12q13.3), *STAT3* (17q21.2), *STAT4* (2q32.2), *STAT5A* (17q21.2), *STAT5B* (17q21.2), and *STAT6* (12q13.3) (8). *STATs* need to be triggered by extracellular signaling ligands such as hormones, cytokines, and growth factors, and then perform biological functions via the Janus tyrosine kinase (JAK)-STAT signaling pathway (9,10). Previous studies have reported that *STATs* are engaged in numerous normal physiological cellular processes, including cell proliferation, differentiation, apoptosis, and immune system regulation (11-13). However, many irregular actions and mechanisms can occur in different *STAT* family members, resulting in aberrant *STAT* regulation, which leads to different pathological events in cancer, such as malignant cell conversion and metastasis. At present, several *STAT* family members are even regarded as oncogenes, while others are not (10,14). Thus, the relevance of immune infiltration and prognostic value of *STAT* family members in BRCA still requires further study.

In this research, we aim to conduct an in-depth analysis of the *STAT* family using multiple databases, such as The Cancer Genome Atlas (TCGA) and Kaplan-Meier plotter, and summarize the role of *STATs* in the clinical prognosis and immune infiltration in BRCA, including gene expression, survival status, genetic mutation, immune infiltration, and related cellular pathways. This will provide a comprehensive insight into the role of different *STAT* members in cancer initiation and progression, and will help us to better understand the prospect of their potential clinical implications. We present the following article in accordance with the REMARK reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-22-189/rc).

Methods

Datasets source and patients selection

To facilitate our analysis, we mainly used bioinformatics techniques to assist in the prognostic analysis of *STAT* family members. TCGA (https://tcga-data.nci.nih.gov/) is a public database that aims to provide researchers with available datasets to improve oncology-related diagnostic methods, treatment standards, and in order to facilitate cancer prevention (15). In this study, we included 1,880 BRCA patients who had complete survival data in the TCGA database and removed patient information with missing data. Meanwhile, we downloaded and analyzed several datasets from the PrognoScan platform and directly collated into a table format for our interpretation. Finally, we downloaded the RNA-sequencing (RNA-seq) expression data of *STAT* family members from the Xena system (https://xenabrowser.net/datapages/) for immune profiling. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Oncomine

The messenger RNA (mRNA) expression levels of the *STAT* family members in different types of cancers were analyzed through the Oncomine database. Oncomine is a publicly available online oncology-related database and a web-based data mining platform, which integrates many bioinformatics resources such as LocusLink, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Biocarta, etc. (16-18). In this study, genes with a fold change (FC) of 2 and a P value <0.01were ranked in the top 10%, and the data types were set as mRNA.

Gene Expression Profiling Interactive Analysis (GEPIA)

GEPIA is a time-saving tool that can unlock the value of large genome data in TCGA and Genotype-Tissue Expression (GTEx). It also can provide an interactive web platform for gene expression analysis, and is a good complement to other existing gene analysis tools, such as cBioPortal and Expression Atlas (19). In this study, we used a |log2FC| cutoff of 1, q-value cutoff of 0.01, and the log scale was set as yes. We then performed comparisons between the tumor and normal breast tissues, analyzing different gene expression, associated prognostic value, and pathological stage by GEPIA.

The Human Protein Atlas (HPA)

The HPA database can map human tissues, cells, and organs based on proteomic, transcriptomics, and systems biology data. Forty-four major normal tissues and some cancer tissues were included by immunohistochemistry. In our study, the protein expression of *STAT* family members in normal human breast and tumor tissues were confirmed by this database (20).

Kaplan-Meier plotter

The Kaplan-Meier plotter is a website that includes numerous modules, such as gene expression information and clinical outcome data. This website was utilized to assess the prognostic value of different *STATs* in BRCA patients, which were constructed based on gene chips and RNA-seq data from public databases including Gene Expression Omnibus (GEO), European Genome-phenome Archive (EGA), and TCGA. The log rank test was used to compare the Kaplan-Meier survival curves. All statistical comparisons were considered significant at P<0.05 (21).

PrognoScan

Based on a large collection of publicly available cancer gene microarray datasets, the PrognoScan database use the minimal P value method to assess the relationship between gene expression and prognosis in various cancer patients (22). We extracted some *STAT* family-related datasets in BRCA and analyzed the expression of *STATs* in relation to overall survival (OS), relapse-free survival (RFS), disease-free survival (DFS), and distant metastasis-free survival (DMFS) in BRCA patients. All statistical comparisons were considered significant at P<0.05.

Survival analysis

To assess the prognostic value of individual *STATs*, the patients was divided into two groups based on their median gene expression (high *vs.* low expression) in the datasets from TCGA and PrognoScan platform. OS was utilized to assess the prognosis of BRCA patients in the datasets from both sources, whereas DFMS, DFS and RFS were only used to evaluate the prognosis of patients in the PrognoScan platform. Using the Kaplan-Meier method, survival analysis was carried out and gene expression groups were compared using the log-rank test. Cox regression analysis was used to screen the independent prognostic factors by calculating the Cox P value, adjusted hazard ratio (HR) and 95% confidence interval (CI).

cBioPortal

cBioPortal is an interactive resource that facilitates access to cancer genomic datasets (23). Based on genetic mutations-related data, the modules of *STATs* were provided by cBioPortal. We used this database to select the BRCA datasets, with 817 cases (TCGA, *Cell* 2015) (24) selected for further analysis. The Oncoprint, OS, and DFS of *STAT* family members were analyzed online using this method.

GeneMANIA

GeneMANIA is a user-friendly website that can predict gene function, analyze gene lists, and prioritize genes for functional assays (25). We used it to explore the relationship among different *STAT* family members.

Search Tool for the Retrieval of Interacting Genes (STRING)

STRING can provide important evaluation clues and integrate the protein-protein interactions (PPIs), including both direct and indirect associations (26). In this study, protein-protein associations in the *STATs* family from co-expression data were analyzed by STRING.

Metascape

Metascape is an efficient functional annotation analysis tool comprising interactome analysis, gene annotation,

functional enrichment, and membership search (27). The Metascape database was used to explore the enrichment pathways associated with the *STAT4* co-expressed genes.

Tumor Immune Estimation Resource (TIMER)

The TIMER database can be utilized to analyze the infiltration of immune-related cells in various types of tumors (28). In this study, TIMER database analysis allowed us to investigate the relationship between *STAT4* expression levels and the degree of infiltration of various immune-related cells in different BRCA subtypes.

Follow-up

Follow-up was defined as the time between the diagnosis of BRCA and the occurrence of the outcome event. In this study, the primary clinical outcome endpoint was OS, which was defined as the time from the diagnosis of BRCA to death from any cause. The secondary study endpoints were DFMS, DFS and RFS. The survival status of patients lost to follow-up was recorded as "blank" in the TCGA database and excluded accordingly.

Statistical analysis

All gene expression data were standardized by log2 transformation. Normal and cancer tissues were compared using two sets of *t*-test. Relevance analysis between two variables was performed using Spearman or Pearson tests. All statistical analyses were processed using the "ESTIMATE" and "ggcorrplot" package of R 17 software (https://CRAN.R-project.org, version 4.0.2). A two-tailed $P<0.05$ was considered to be statistically significant.

Results

Divergent expression of STATs in BRCA patients

To investigate the expression levels of different *STATs* in BRCA patients, we analyzed the mRNA expression of *STATs* through the Oncomine database. The results suggested that, compared to normal tissues, the mRNA expression levels of *STAT1* and *STAT2* were elevated in BRCA tissues (*Figure 1A*), while those of *STAT3*, *STAT4*, *STAT5A*, and *STAT5B* were reduced in BRCA. We also used the GEPIA database to compare the mRNA expression levels of *STATs* between BRCA and normal breast tissues (*Figure 1B*). The results suggested that, compared to normal tissues, the expression level of *STAT1* was elevated in BRCA tissues, while those of *STAT5A* and *STAT5B* were reduced. These results, which were obtained via the GEPIA database, were almost identical to those previously obtained through the Oncomine database.

Additionally, some typical immunohistochemical images were selected from the HPA database to analyze the protein expressions of the *STAT* family in human breast normal and tumor tissues. The immunohistochemical images illustrated that *STAT1*, *STAT2*, *STAT3*, and *STAT6* had low coloring in normal breast tissues, but high staining in BRCA (Figure S1). Also, *STAT5A* and *STAT5B* had lower levels of protein expression in BRCA tissues compared with normal tissues, indicating a similar result to mRNA transcriptional levels.

We then explored the relationship between the pathological stage of BRCA patients and the expression of *STATs* using the GEPIA database. As shown in *Figure 2*, we found that the *STAT3* and *STAT6* groups were obviously variable, whereas the *STAT1*, *STAT2*, *STAT4*, *STAT5A*, and *STAT5B* groups were not obviously different. These results indicated that *STAT* expressions were markedly different in BRCA compared with normal tissues, and that *STAT3* may play an important role in the development and progression of BRCA.

Prognostic value of STATs in BRCA patients

To analyze the prognostic value of *STATs* in BRCA patients, we used Kaplan-Meier plotter website to conduct a subgroup analysis. As shown in *Figure 3*, the mRNA expression levels of *STAT4* (HR =0.78, 95% CI: 0.65–0.94, P=0.0099), *STAT5A* (HR =0.74, 95% CI: 0.61–0.89, P=0.0017), and *STAT6* (HR =0.76, 95% CI: 0.63–0.92, P=0.0043) were significantly correlated with long OS in BRCA patients.

Furthermore, we used the GEPIA database to validate the value of the variable expression of *STATs* in the progression of BRCA. The OS and DFS curves are displayed in *Figure 4*. As for BRCA patients, there was no obvious association between any of the *STAT* family members and the DFS rates. Meanwhile, we found that the patients with high transcriptional levels of *STAT4* (HR =0.56, P=0.00062) were significantly related to long OS. With the exception of the *STAT4* group, all other members of the *STAT* family did not appear to show a remarkable effect on OS. Therefore, we subsequently extracted

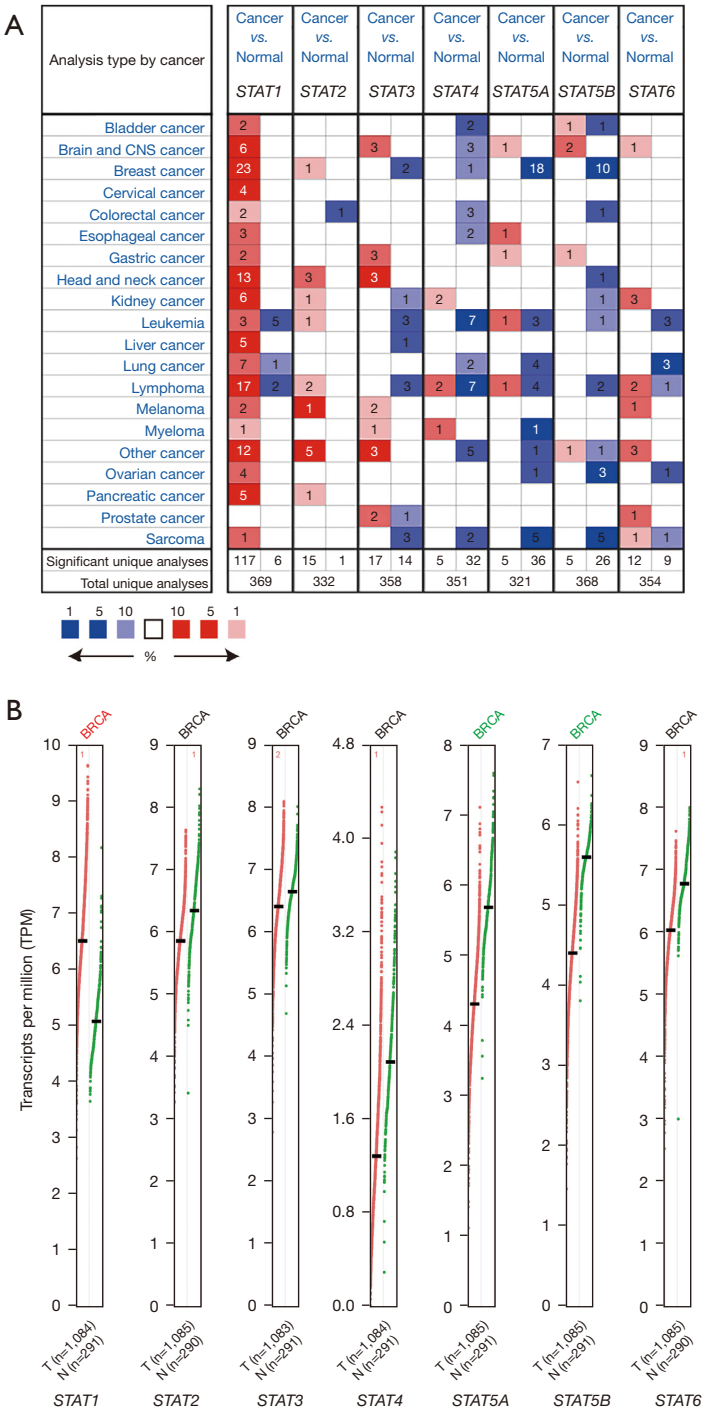

Figure 1 Expression of STATs in BRCA. (A) mRNA expression of *STAT* family members in different cancer types (Oncomine). The graph shows the number of datasets with obvious mRNA expression among the *STAT* family genes: upregulated (red) and downregulated (blue). The criteria were as follows: P value: 0.01, fold change: 2, gene rank: 10%, date type: mRNA, and analysis type: cancer *vs.* normal tissue. (B) Expression of *STATs* in BRCA (GEPIA). As shown in this graphic: expression levels of *STAT1* were higher in BRCA tissues, while the expression levels of *STAT5A* and *STAT5B* were lower in BRCA *vs.* normal tissues (P<0.05). BRCA, breast cancer; CNS, central nervous system; GEPIA, Gene Expression Profiling Interactive Analysis; mRNA, messenger RNA; *STAT*, signal transducer and activator of transcription; N, normal tissue; T, tumor tissue.

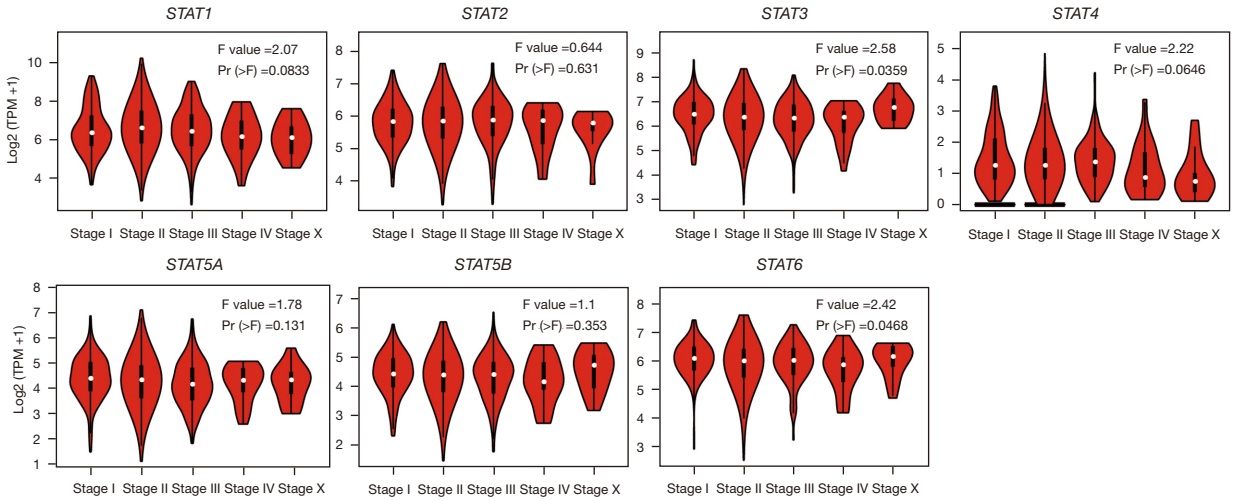

Figure 2 Correlation between *STAT* expression and tumor pathological stage in breast cancer patients (GEPIA). The expression of *STAT3* and *STAT6* were related to the pathological stage of BRCA patients (P<0.05). TPM, transcripts per million; BRCA, breast cancer; GEPIA, Gene Expression Profiling Interactive Analysis; *STAT*, signal transducer and activator of transcription.

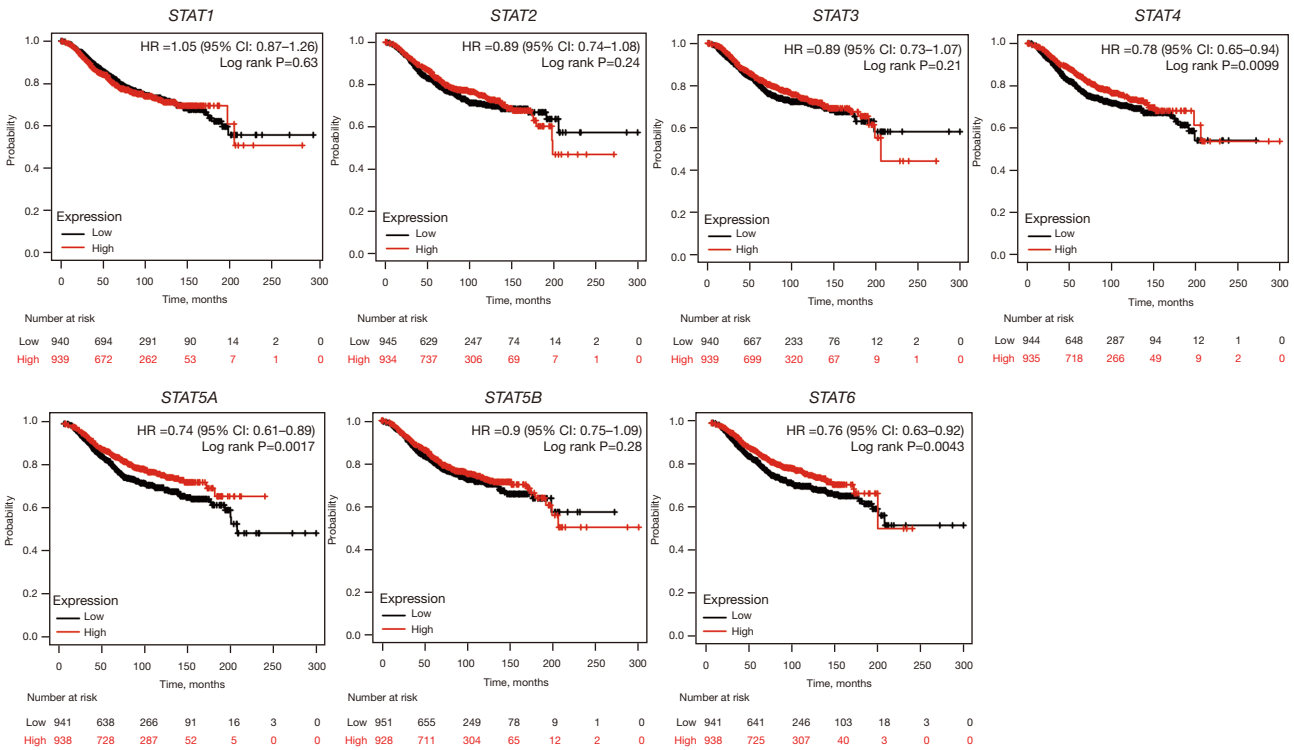

Figure 3 Prognostic value of *STATs* in breast cancer (Kaplan-Meier plotter). High levels of mRNA expression in *STAT4*, *STAT5A*, and *STAT6* were significantly correlated with long OS in breast cancer patients. HR, hazard ratio; mRNA, messenger RNA; OS, overall survival; *STAT*, signal transducer and activator of transcription.

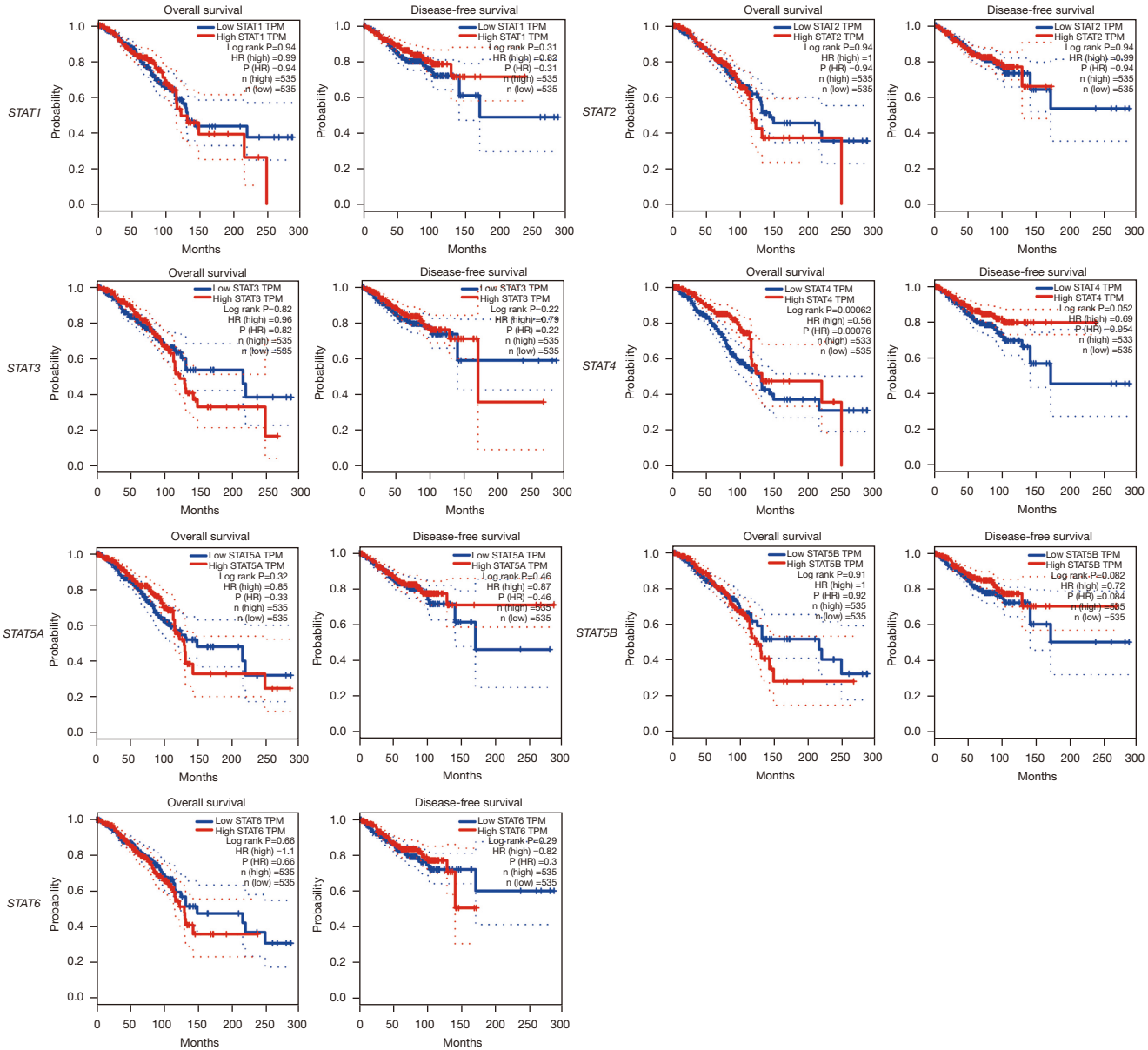

Figure 4 Prognostic value of the mRNA expression of *STAT* family members in BRCA (GEPIA). BRCA patients with a high transcriptional level of *STAT4* (P=0.00062) were obviously related to long OS. BRCA, breast cancer; GEPIA, Gene Expression Profiling Interactive Analysis; HR, hazard ratio; mRNA, messenger RNA; OS, overall survival; *STAT*, signal transducer and activator of transcription; TPM, transcripts per million.

several datasets from the PrognoScan platform and further analyzed different *STATs* expression in relation to OS, RFS, DFS, and DMFS in BRCA patients (*Table 1*), aiming to further evaluate our previous analysis. The results showed that *STAT2* (P<0.01, HR =1.23, 95% CI: 1.07–1.42) and *STAT3* (P=0.018, HR =0.69, 95% CI: 0.51–0.94) could be an independent risk factor for predicting OS. *STAT4* could be used as an independent predictor of DMFS in BRCA based on both GSE19615 (P=0.021, HR =0.21, 95% CI: 0.06–0.79) and GSE2034 (P=0.015, HR =0.57, 95% CI: 0.37–0.90) datasets. Meanwhile, *STAT5A*, *STAT5B* and *STAT6* also have been shown to independently predict the

Table 1 STAT family gene expression was related to the prognosis of breast cancer in PrognoScan

Gene	Dataset	Endpoint	Cutpoint	Cox P value	Ln(HR)	HR (95% CI)
STAT1	GSE7378	OS	158	0.734006	0.422818	1.53 (0.96–2.42)
	E-TABM-158	OS	198	0.690165	0.102520	1.11 (0.88–1.39)
	GSE19615	DFS	76	–1.49087	–0.231171	0.79 (0.44–1.42)
	GSE3143	DFS	54	–15.3825	0.034494	1.04 (0.48–2.22)
	GSE7849	DFS	249	0.543037	0.223837	1.25 (0.88–1.78)
	GSE3494-GPL96	RFS	204	0.380639	0.121462	1.13 (0.88–1.44)
	GSE4922-GPL96	RFS	87	0.570071	0.024605	1.02 (0.61–1.72)
	GSE2990	RFS	159	0.856193	0.462988	1.59 (0.88–2.88)
	GSE7390	RFS	125	0.729704	0.215864	1.24 (0.89–1.73)
	GSE9195	DMFS	87	0.843094	0.069123	1.07 (0.80–1.43)
	GSE12093	DMFS	136	1.48027	0.413829	1.51 (0.92–2.50)
	GSE11121	DMFS	200	0.904803	0.225317	1.25 (0.88–1.77)
	GSE1378	DMFS	286	–1.27494	–0.236758	0.79 (0.56–1.12)
	GSE9893	DMFS	117	0.59709	0.058867	1.06 (0.76–1.48)
	GSE2034	DMFS	125	–1.23639	–0.072966	0.93 (0.66–1.31)
STAT2	GSE3143	OS	158	0.123157	0.722624	2.06 (0.82–5.16)
	GSE9893	OS	155	0.00447014	0.209301	1.23 (1.07–1.42)
	GSE1456-GPL97	OS	159	0.741445	–0.133722	0.87 (0.40–1.94)
	E-TABM-158	OS	117	0.604802	–0.220720	0.80 (0.35–1.85)
	GSE7390	OS	198	0.671654	0.116688	1.12 (0.66–1.93)
	GSE7849	DFS	76	0.89763	–0.128759	0.88 (0.12–6.25)
	GSE7378	DFS	54	0.710306	–0.322114	0.72 (0.13–3.97)
	GSE4922-GPL97	DFS	249	0.40129	0.303484	1.35 (0.67–2.75)
	GSE12276	RFS	204	0.7863	0.055055	1.06 (0.71–1.57)
	GSE6532-GPL570	RFS	87	0.706449	–0.199444	0.82 (0.29–2.31)
	GSE9195	RFS	77	0.784835	0.233873	1.26 (0.24–6.77)
	GSE1379	RFS	60	0.688543	0.178464	1.20 (0.50–2.86)
	GSE1456-GPL96	RFS	159	0.90511	–0.027443	0.97 (0.62–1.53)
	GSE1456-GPL97	RFS	159	0.391313	0.341762	1.41 (0.64–3.07)
	E-TABM-158	RFS	117	0.604802	–0.220720	0.80 (0.35–1.85)
	GSE2990	RFS	125	0.904064	0.032345	1.03 (0.61–1.75)
	GSE7390	RFS	198	0.245853	–0.126170	0.88 (0.71–1.09)
	GSE19615	DMFS	115	0.153716	–0.868095	0.42 (0.13–1.38)
	GSE6532-GPL570	DMFS	87	0.695463	–0.202646	0.82 (0.30–2.25)
	GSE9195	DMFS	77	0.255748	–2.201030	0.11 (0.00–4.93)

Table 1 (*continued*)

Table 1 (*continued*)

Gene	Dataset	Endpoint	Cutpoint	Cox P value	Ln(HR)	HR (95% CI)
	GSE12093	DMFS	136	0.956642	−0.017426	0.98 (0.52–1.84)
	GSE11121	DMFS	200	0.444782	0.458926	1.58 (0.49–5.13)
	GSE2034	DMFS	286	0.0981089	−0.424718	0.65 (0.40–1.08)
	E-TABM-158	DMFS	117	0.190977	1.181818	3.26(0.55-19.17)
	GSE2990	DMFS	125	0.767727	0.271797	1.31 (0.22–7.97)
	GSE7390	DMFS	198	0.304212	0.277772	1.32 (0.78–2.24)
STAT3	GSE3143	OS	158	0.695188	−0.147051	0.86 (0.41–1.80)
	GSE9893	OS	155	0.0177795	−0.366907	0.69 (0.51–0.94)
	GSE1456-GPL97	OS	159	0.840228	0.104908	1.11 (0.40–3.08)
	E-TABM-158	OS	117	0.918875	0.020681	1.02 (0.69–1.52)
	GSE7390	OS	198	0.907517	−0.030143	0.97 (0.58–1.61)
	GSE7849	DFS	76	0.0141442	1.593150	4.92 (1.38–17.56)
	GSE7378	DFS	54	0.456313	−0.241100	0.79 (0.42–1.48)
	GSE4922-GPL96	DFS	249	0.971322	0.007425	1.01 (0.67–1.51)
	GSE12276	RFS	204	0.687901	−0.072622	0.93 (0.65–1.33)
	GSE6532-GPL570	RFS	87	0.70237	−0.344585	0.71 (0.12–4.15)
	GSE9195	RFS	77	0.0217172	1.335230	3.80 (1.22–11.89)
	GSE1379	RFS	60	0.894997	−0.060561	0.94 (0.38–2.31)
	GSE1456-GPL97	RFS	159	0.89862	−0.065406	0.94 (0.34–2.56)
	E-TABM-158	RFS	117	0.918875	0.020681	1.02 (0.69–1.52)
	GSE2990	RFS	62	0.0108244	−1.112200	0.33 (0.14–0.77)
	GSE2990	RFS	125	0.837496	−0.048294	0.95 (0.60–1.51)
	GSE2990	RFS	62	0.0378638	−0.593717	0.55 (0.32–0.97)
	GSE7390	RFS	198	0.646099	0.082683	1.09 (0.76–1.55)
	GSE19615	DMFS	115	0.468102	−0.499143	0.61 (0.16–2.34)
	GSE6532-GPL570	DMFS	87	0.70237	−0.344585	0.71 (0.12–4.15)
	GSE9195	DMFS	77	0.135511	1.010010	2.75 (0.73–10.34)
	GSE9195	DMFS	77	0.144656	0.890767	2.44 (0.74–8.07)
	GSE12093	DMFS	136	0.0164555	−0.532524	0.59 (0.38–0.91)
	GSE11121	DMFS	200	0.0203471	−0.822380	0.44 (0.22–0.88)
	GSE2034	DMFS	286	0.0157849	−0.254330	0.78 (0.63–0.95)
	E-TABM-158	DMFS	117	0.117872	−0.417841	0.66 (0.39–1.11)
	GSE2990	DMFS	125	0.905535	−0.036515	0.96 (0.53–1.76)
	GSE2990	DMFS	54	0.0858196	−0.982659	0.37 (0.12–1.15)
	GSE7390	DMFS	198	0.0827706	0.387547	1.47 (0.95–2.28)

Table 1 (*continued*)

Table 1 (*continued*)

Gene	Dataset	Endpoint	Cutpoint	Cox P value	Ln(HR)	HR (95% CI)
STAT4	GSE3143	OS	158	0.638658	−0.158958	0.85 (0.44–1.66)
	GSE9893	OS	155	0.0693526	0.316011	1.37 (0.98–1.93)
	GSE1456-GPL96	OS	159	0.337157	−0.435590	0.65 (0.27–1.57)
	E-TABM-158	OS	117	0.731327	−0.069909	0.93 (0.63–1.39)
	GSE7390	OS	198	0.899221	−0.021423	0.98 (0.70–1.36)
	GSE7849	DFS	76	0.898074	−0.056164	0.95 (0.40–2.23)
	GSE7378	DFS	54	0.0599192	−1.718460	0.18 (0.03–1.07)
	GSE4922-GPL96	DFS	249	0.531574	0.183205	1.20 (0.68–2.13)
	GSE12276	RFS	204	0.370653	−0.093580	0.91 (0.74–1.12)
	GSE6532-GPL570	RFS	87	0.753865	0.064398	1.07 (0.71–1.60)
	GSE9195	RFS	77	0.376928	−0.541672	0.58 (0.17–1.93)
	GSE1456-GPL96	RFS	159	0.598349	0.245888	1.28 (0.51–3.19)
	E-TABM-158	RFS	117	0.731327	−0.069909	0.93 (0.63–1.39)
	GSE2990	RFS	62	0.686218	−0.069183	0.93 (0.67–1.31)
	GSE19615	DMFS	115	0.0210692	−1.539200	0.21 (0.06–0.79)
	GSE6532-GPL570	DMFS	87	0.753865	0.064398	1.07 (0.71–1.60)
	GSE9195	DMFS	77	0.142219	−1.103740	0.33 (0.08–1.45)
	GSE12093	DMFS	136	0.774322	−0.096607	0.91 (0.47–1.76)
	GSE11121	DMFS	200	0.057605	−0.581530	0.56 (0.31–1.02)
	GSE2034	DMFS	286	0.0147119	−0.556861	0.57 (0.37–0.90)
	E-TABM-158	DMFS	117	0.986949	−0.003965	1.00 (0.62–1.60)
	GSE2990	DMFS	125	0.254872	−0.531531	0.59 (0.24–1.47)
	GSE7390	DMFS	198	0.992948	−0.001456	1.00 (0.72–1.38)
STAT5A	GSE3143	OS	158	0.210806	−0.2375	0.79 (0.54–1.14)
	GSE9893	OS	155	0.440632	−0.185188	0.83 (0.52–1.33)
	GSE1456-GPL96	OS	159	0.158685	−0.706401	0.49 (0.18–1.32)
	E-TABM-158	OS	117	0.501951	−0.213658	0.81 (0.43–1.51)
	GSE7390	OS	198	0.71127	−0.082341	0.92 (0.60–1.42)
	GSE7849	DFS	76	0.543889	0.238656	1.27 (0.59–2.74)
	GSE7378	DFS	54	0.0314761	−2.14162	0.12 (0.02–0.83)
	GSE4922-GPL96	DFS	249	0.0172677	−0.825435	0.44 (0.22–0.86)
	GSE12276	RFS	204	0.167323	−0.228702	0.80 (0.58–1.10)
	GSE6532-GPL570	RFS	87	0.00227358	−1.144590	0.32 (0.15–0.66)
	GSE9195	RFS	77	0.923972	−0.074396	0.93 (0.20–4.28)
	GSE1378	RFS	60	0.248078	−0.225438	0.80 (0.54–1.17)

Table 1 (*continued*)

Table 1 (*continued*)

Gene	Dataset	Endpoint	Cutpoint	Cox P value	Ln(HR)	HR (95% CI)
	GSE1379	RFS	60	0.452682	−0.228444	0.80 (0.44–1.44)
	GSE1456-GPL96	RFS	159	0.185957	−0.683025	0.51 (0.18–1.39)
	E-TABM-158	RFS	117	0.501951	−0.213658	0.81 (0.43–1.51)
	GSE2990	RFS	125	0.284496	−0.312392	0.73 (0.41–1.30)
	GSE7390	RFS	198	0.763745	0.053695	1.06 (0.74–1.50)
	GSE19615	DMFS	115	0.00884115	−1.729470	0.18 (0.05–0.65)
	GSE6532-GPL570	DMFS	87	0.00227358	−1.144590	0.32 (0.15–0.66)
	GSE9195	DMFS	77	0.48064	−0.659811	0.52 (0.08–3.23)
	GSE12093	DMFS	136	0.0598301	−1.15754	0.31 (0.09–1.05)
	GSE11121	DMFS	200	0.0358548	−0.681225	0.51 (0.27–0.96)
	GSE2034	DMFS	286	0.00786138	−0.590146	0.55 (0.36–0.86)
	E-TABM-158	DMFS	117	0.406866	−0.327397	0.72 (0.33–1.56)
	GSE2990	DMFS	125	0.319595	−0.384966	0.68 (0.32–1.45)
	GSE7390	DMFS	198	0.996648	−0.000910	1.00 (0.65–1.53)
STAT5B	GSE12417-GPL150	OS	158	0.502612	0.243754	1.28 (0.63–2.60)
	GSE12417-GPL191	OS	155	0.0739188	0.306504	1.36 (0.97–1.90)
	GSE12417-GPL197	OS	159	3.15E−07	−2.43983	0.09 (0.03–0.22)
	GSE12417-GPL207	OS	117	0.314488	0.276225	1.32 (0.77–2.26)
	GSE12417-GPL243	OS	198	0.51792	0.142256	1.15 (0.75–1.77)
	GSE12417-GPL155	DFS	76	0.633366	−0.238224	0.79 (0.30–2.10)
	GSE12417-GPL204	DFS	54	0.131027	−0.90912	0.40 (0.12–1.31)
	GSE12417-GPL222	DFS	249	0.0659483	−0.661075	0.52 (0.26–1.04)
	GSE12417-GPL160	RFS	204	0.182215	−0.092228	0.91 (0.80–1.04)
	GSE12417-GPL171	RFS	77	0.899623	−0.089730	0.91 (0.23–3.69)
	GSE12417-GPL188	RFS	60	0.033972	0.459418	1.58 (1.04–2.42)
	GSE12417-GPL195	RFS	159	0.0667261	−0.967095	0.38 (0.14–1.07)
	GSE12417-GPL211	RFS	117	0.163118	−0.805688	0.45 (0.14–1.39)
	GSE12417-GPL226	RFS	62	0.0297399	−0.866101	0.42 (0.19–0.92)
	GSE12417-GPL231	RFS	125	0.419839	−0.240821	0.79 (0.44–1.41)
	GSE12417-GPL242	RFS	198	0.675771	−0.083283	0.92 (0.62–1.36)
	GSE12417-GPL149	DMFS	115	0.0255209	−1.57939	0.21 (0.05–0.82)
	GSE12417-GPL163	DMFS	87	0.066610	0.964541	2.62 (0.94–7.35)
	GSE12417-GPL180	DMFS	77	0.326454	−0.909009	0.40 (0.07–2.48)
	GSE12417-GPL181	DMFS	136	0.496678	−0.283021	0.75 (0.33–1.70)
	GSE12417-GPL186	DMFS	200	0.00553487	−1.30901	0.27 (0.11–0.68)

Table 1 (*continued*)

Table 1 (*continued*)

Gene	Dataset	Endpoint	Cutpoint	Cox P value	Ln(HR)	HR (95% CI)
	GSE12417-GPL192	DMFS	286	0.0750032	−0.465471	0.63 (0.38–1.05)
	GSE12417-GPL214	DMFS	117	0.372841	−0.613899	0.54 (0.14–2.09)
	GSE12417-GPL228	DMFS	125	0.458577	−0.291181	0.75 (0.35–1.61)
	GSE12417-GPL229	DMFS	54	0.694294	−0.129606	0.88 (0.46–1.68)
	GSE12417-GPL236	DMFS	125	0.76281	−0.215466	0.81 (0.20–3.27)
	GSE12417-GPL239	DMFS	198	0.289715	0.252804	1.29 (0.81–2.06)
STAT6	GSE3143	OS	158	0.993998	−0.002656	1.00 (0.50–1.99)
	GSE9893	OS	155	0.00582855	−0.217493	0.80 (0.69–0.94)
	GSE1456-GPL96	OS	159	0.0581333	−0.846861	0.43 (0.18–1.03)
	E-TABM-158	OS	117	0.997088	0.0010673	1.00 (0.56–1.78)
	GSE7390	OS	198	0.189067	−0.177569	0.84 (0.64–1.09)
	GSE7849	DFS	76	0.232317	1.06851	2.91 (0.50–6.81)
	GSE7378	DFS	54	0.286081	0.703717	2.02 (0.55–7.36)
	GSE4922-GPL96	DFS	249	0.0879759	−0.313232	0.73 (0.51–1.05)
	GSE12276	RFS	204	0.0735396	−0.165275	0.85 (0.71–1.02)
	GSE6532-GPL570	RFS	87	0.0561677	−0.971252	0.38 (0.14–1.03)
	GSE9195	RFS	77	0.573397	−0.384807	0.68 (0.18–2.60)
	GSE1378	RFS	60	0.361895	−0.272733	0.76 (0.42–1.37)
	GSE1456-GPL96	RFS	159	0.948818	−0.013512	0.99 (0.65–1.49)
	E-TABM-158	RFS	117	0.959771	0.022162	1.02 (0.43–2.42)
	GSE2990	RFS	125	0.742506	−0.136701	0.87 (0.39–1.97)
	GSE7390	RFS	198	0.488795	−0.142638	0.87 (0.58–1.30)
	GSE19615	DMFS	115	0.309889	−0.678707	0.51 (0.14–1.88)
	GSE6532-GPL570	DMFS	87	0.0561677	−0.971252	0.38 (0.14–1.03)
	GSE9195	DMFS	77	0.706615	−0.232473	0.79 (0.24–2.66)
	GSE12093	DMFS	136	0.205107	−0.73496	0.48 (0.15–1.49)
	GSE11121	DMFS	200	0.000547358	−0.603841	0.55 (0.39–0.77)
	GSE2034	DMFS	286	0.0381629	−0.47926	0.62 (0.39–0.97)
	E-TABM-158	DMFS	117	0.300091	0.550172	1.73 (0.61–4.91)
	GSE2990	DMFS	125	0.936266	−0.044824	0.96 (0.32–2.87)

OS, overall survival; DFS, disease-free survival; RFS, relapse-free survival; DMFS, disease metastasis-free survival; HR, hazard ratio; CI, confidence interval. *STAT*, signal transducer and activator of transcription.

prognosis of BRCA.

Genetic mutations, clinical relevance, and protein-protein network of STATs in BRCA

We utilized the cBioPortal online tool to explore the genetic mutations of *STATs* in BRCA. As shown in *Figure 5A*, two or more mutations were observed in different BRCA isoforms, with deep depletion alterations being more prevalent in mixed ductal and lobular breast carcinoma samples. As shown in *Figure 5B*, *STATs* were mutated in 72 samples of 816 patients in BRCA, accounting for 9%. In addition, *STAT1*, *STAT2*, *STAT3*, *STAT4*, *STAT5A*, *STAT5B*, and *STAT6* were altered in the queried BRCA samples in the order of 1.7%, 0.7%, 3.0%, 2.2%, 2.9%, 3.0%, and 1.5%, respectively. Furthermore, the results of this database also showed a statistically significant relationship between the existence of mutations in *STATs* and OS (*Figure 5C*, P=0.0350) in BRCA patients, but not with DFS (*Figure 5D*, P=0.835).

Moreover, we further explored the potential interaction of these differentially expressed *STATs* using the STRING database. Several lines and points were acquired in the network (*Figure 5E*). The variably-expressed *STATs* were related to cell growth, signal transduction, and transcriptional activation. As shown in *Figure 5F*, the GeneMANIA analysis also indicated that the functions of the *STAT* family member proteins and their relevant molecules (*CLNK*, *SH2D2A*, *DAPP1*, *SH2D1B*, *STAP1*, *SHB*, *PIK3R3*, *SH2D1A*, *SH3BP2*, and *SH2B2*) were mainly related to the JAK-STAT cascade, protein binding and bridging, SH3/SH2 adaptor activity, and cellular response to growth hormone stimulus.

Co-expressed genes of STAT4 and their associated enrichment pathways

Given the remarkable prognostic correlation between *STAT4* and BRCA, we searched the GEPIA database for the top 100 related genes that are co-expressed with *STAT4*, and performed an online data enrichment analysis of Gene Ontology (GO) and KEGG for these co-expressed genes using the Metascape database (*Figure 6*). The results indicated that the co-expressed genes of *STAT4* were mainly enriched in the adaptive immune response, T cell activation, T cell receptor signaling pathway, regulation of cytokine production, negative regulation of immune system processes, and other biological processes. In addition, the main molecular functions, which were regulated by the co-expressed genes of *STAT4*, were T cell receptor binding, major histocompatibility complex (MHC) protein binding, SH2 domain binding, and non-membrane spanning protein tyrosine kinase activity. Also, the KEGG enrichment pathway, which was influenced by co-expressed genes of *STAT4*, was mainly focused on T cell receptor signaling pathway, T helper (Th)1, and Th2 cell differentiation, cell adhesion molecules (CAMs), natural killer cell-mediated cytotoxicity, as well as the nuclear factor-kappa B (NF-κB) and chemokine signaling pathways.

Immune infiltration relevance of STAT4 in BRCA patients

Based on the fact that *STAT4* was significantly enriched in the adaptive immune response and T cell receptor signaling pathways, we investigated the relationship between *STAT4* gene expression and the degree of immune cell infiltration in BRCA. According to previous studies (29,30), we mainly selected six tumor-infiltrating immune cell types, including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells, to analyze the data from TCGA database using R 17 software (version 4.0.2). The results suggested that *STAT4* expression was remarkably positively related to the level of infiltration of B cells (P=1.31e−26, r=0.31, 95% CI: 0.26–0.37), $CD4^+$ T cells (P=2.98e−166, r=0.71, 95% CI: 0.67–0.74), $CD8^+$ T cells (P=1.89e−131, r=0.65, 95% CI: 0.61–0.69), neutrophils (P=1.67e−171, r=0.71, 95% CI: 0.68–0.74), macrophages (P=7.14e−18, r=0.26, 95% CI: 0.20–0.31), and dendritic cells (P=3.7e−178, r=0.72, 95% CI: 0.69–0.75) (*Figure 7A*).

To further depict the underlying immune pattern of *STAT4*, we calculated ESTIMATE, immune, and stromal for all BRCA samples in TCGA database using the ESTIMATE package, in order to assess their correlation with *STAT4* expression. As shown in *Figure 7B*, we observed positive correlations between *STAT4* and the stromal, immune, and ESTIMATE scores in the BRCA dataset.

As shown in Figure S2, we further analyzed the correlation between *STAT4* expression and immune cell infiltration in various BRCA subtypes. The results indicated that *STAT4* exhibited a distinctive connection with the level of infiltration of B cells (r=0.277, P=2.14e−04), $CD4^+$ T cells (r=0.305, P=4.30e−05), neutrophils (r=0.495, P=3.96e−12), and dendritic cells (r=0.256, P=6.66e−04) in the basal-like BRCA group. Additionally, the infiltration of $CD8^+$ T cells (r=0.263, P=2.59e−02), $CD4^+$ T cells (r=0.448, P=7.85e−05), neutrophils (r=0.558, P=3.55e−07),

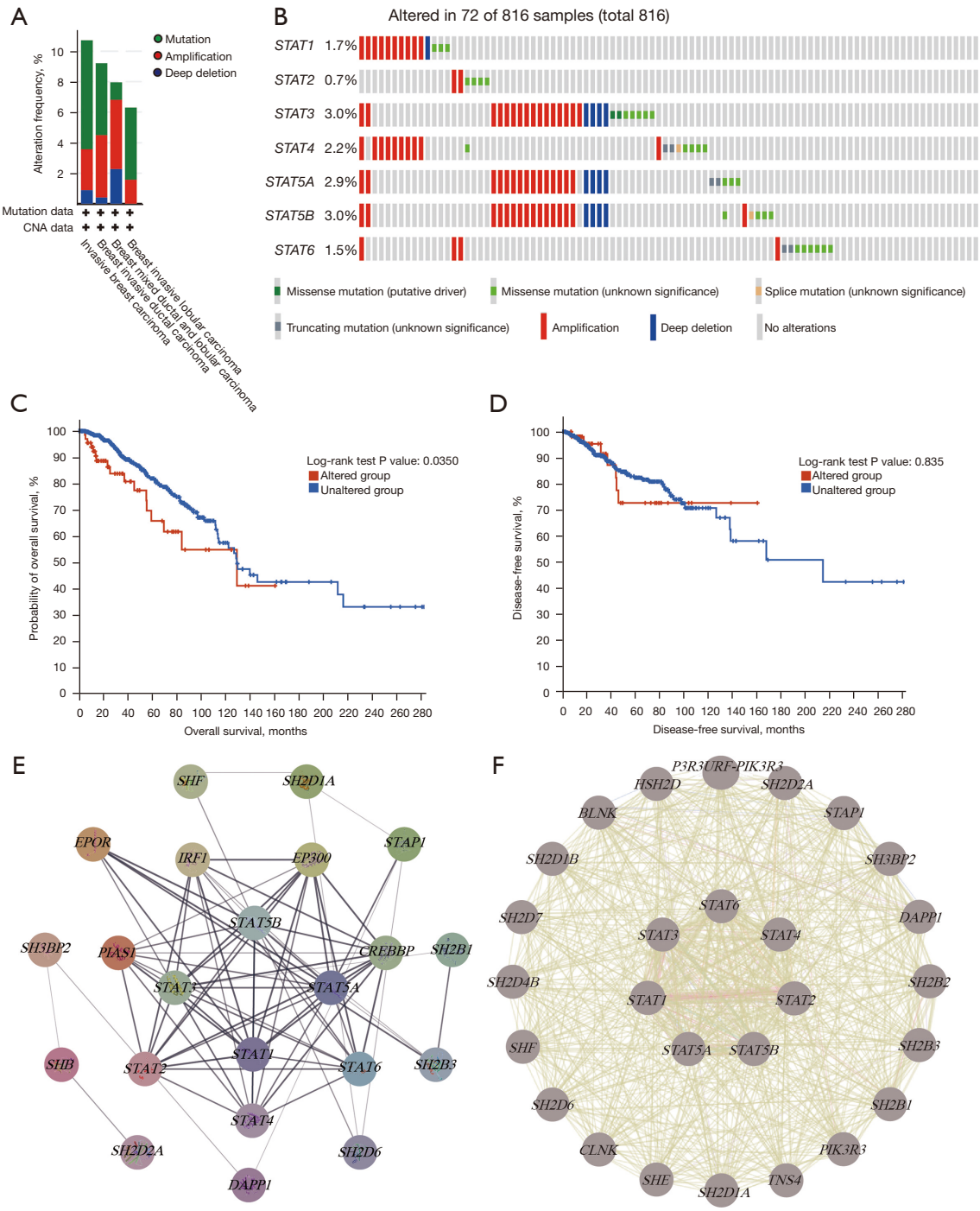

Figure 5 Mutational profile and the protein-protein interaction network of STAT family in breast cancer. (A,B) The *STAT* gene mutation and expression analyses in BRCA (cBioPortal). "+" means the data was included in the analysis. (C,D) *STATs* were altered in 72 samples of 816 patients with BRCA, accounting for 9%. Alterations in *STATs* were correlated with short OS in breast cancer patients, but not with DFS. (E,F) A protein-protein interaction network was constructed using the STRING and GeneMANIA databases, and shows the *STATs* and *STATs*-interacting proteins. BRCA, breast cancer; DFS, disease-free survival; *STAT*, signal transducer and activator of transcription; OS, overall survival.

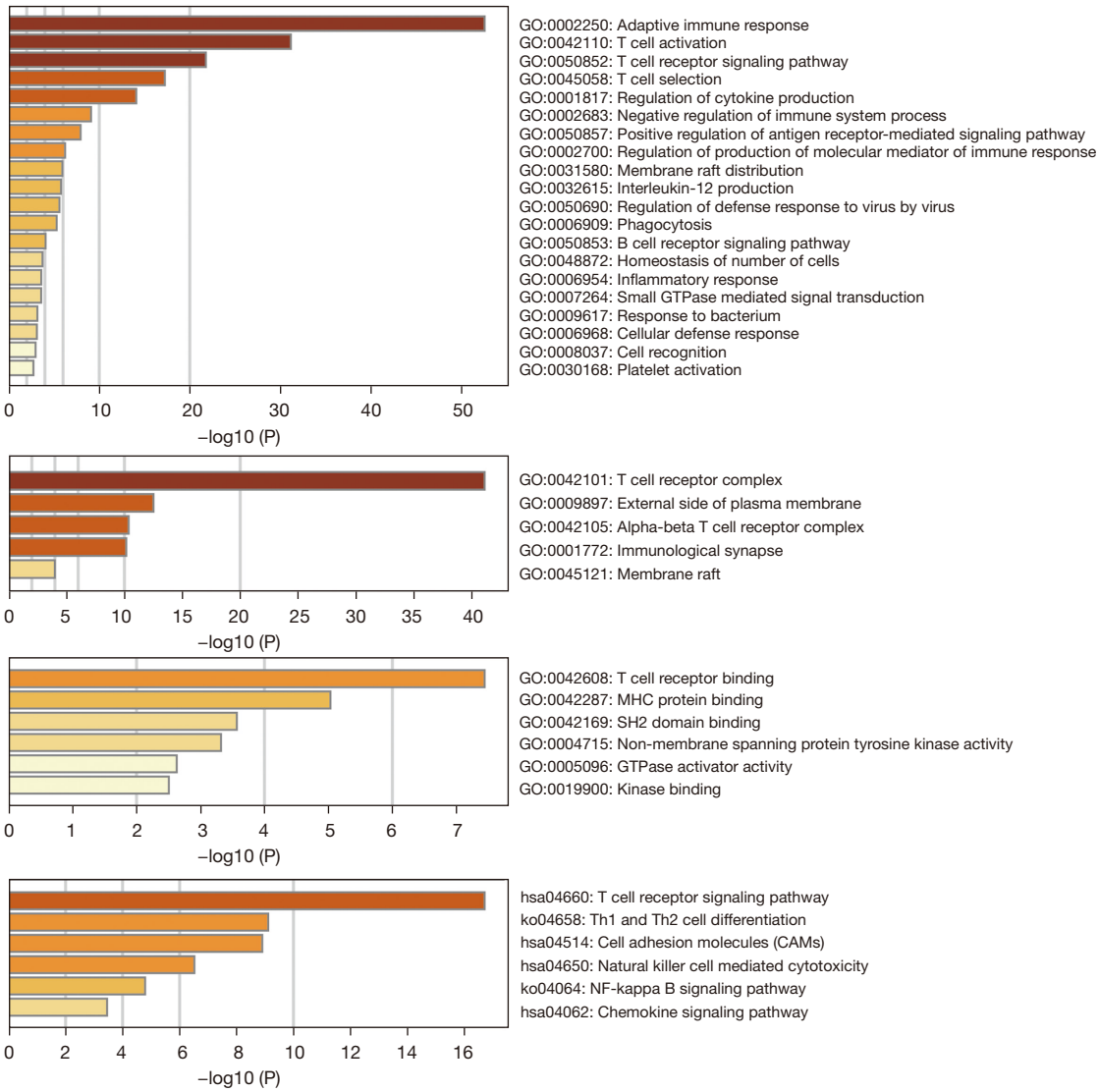

Figure 6 Co-expressed genes of *STAT4* and their associated enrichment pathways analyzed using the Metascape database demonstrated that they were mainly enriched in numerous immune-related biological processes, such as the adaptive immune response, T cell activation, T cell receptor signaling pathway, and negative regulation of immune system processes. *STAT*, signal transducer and activator of transcription.

and dendritic cells (r=0.363, P=1.71e–03) were also related to the high expression level of *STAT4* in the HER2 BRCA group. In the luminal A BRCA patient group, *STAT4* expression was positively correlated with the infiltration of $CD8^+$ T cells (r=0.291, P=1.59e–11), $CD4^+$ T cells (r=0.328, P=1.95e–14), neutrophils (r=0.592, P=3.21e–50), and dendritic cells (r=0.448, P=6.73e–27). In the luminal B group, it was associated with the infiltration of $CD4^+$ T cells (r=0.249, P=5.05e–04), neutrophils (r=0.589, P=2.58e–19), and dendritic cells (r=0.418, P=1.54e–09).

Correlations between STAT4 expression and immune checkpoints and immunomodulators

Immune checkpoint is a regulatory molecule that plays a suppressive role in the immune system. It can control the time and intensity of the immune response to maintain self-tolerance, prevent autoimmune reactions, and minimize tissue damage (31). In this study, we initially selected eight immune checkpoint-associated genes (*CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCD1LG2*, *TIGIT*, and

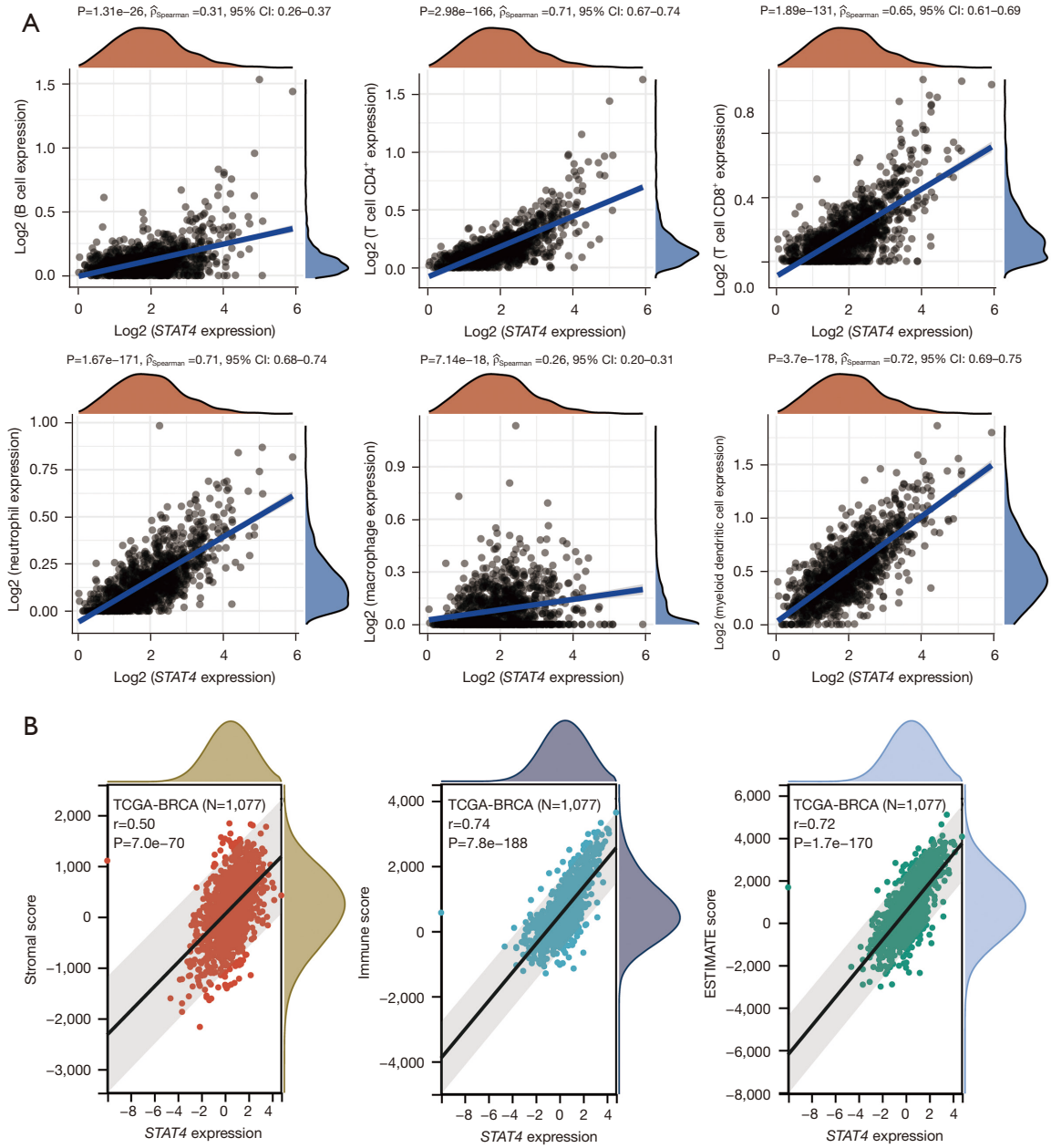

Figure 7 Relationship between *STAT4* and immune cell expression and analysis of immune infiltration scores. (A) The positive relationship between *STAT4* gene expression and the degree of immune cell infiltration in breast cancer. (B) Based on three algorithms, the correlation between immune score and *STAT4* expression was calculated, and the results showed that *STAT4* expression was significantly correlated with immune infiltration. BRCA, breast cancer; CI, confidence interval; *STAT*, signal transducer and activator of transcription; TCGA, The Cancer Genome Atlas.

SIGLEC15) to analyze their expression difference between BRCA and normal tissues (*Figure 8A*). We then further extracted *STAT4* and the expression information of two kinds of immune checkpoint pathway genes from the TCGA database. As shown in *Figure 8B*, we observed that *STAT4* had a relationship with immune checkpoint pathway genes. Furthermore, as shown in *Figure 9*, our finding also revealed that *STAT4* was positively correlated with the

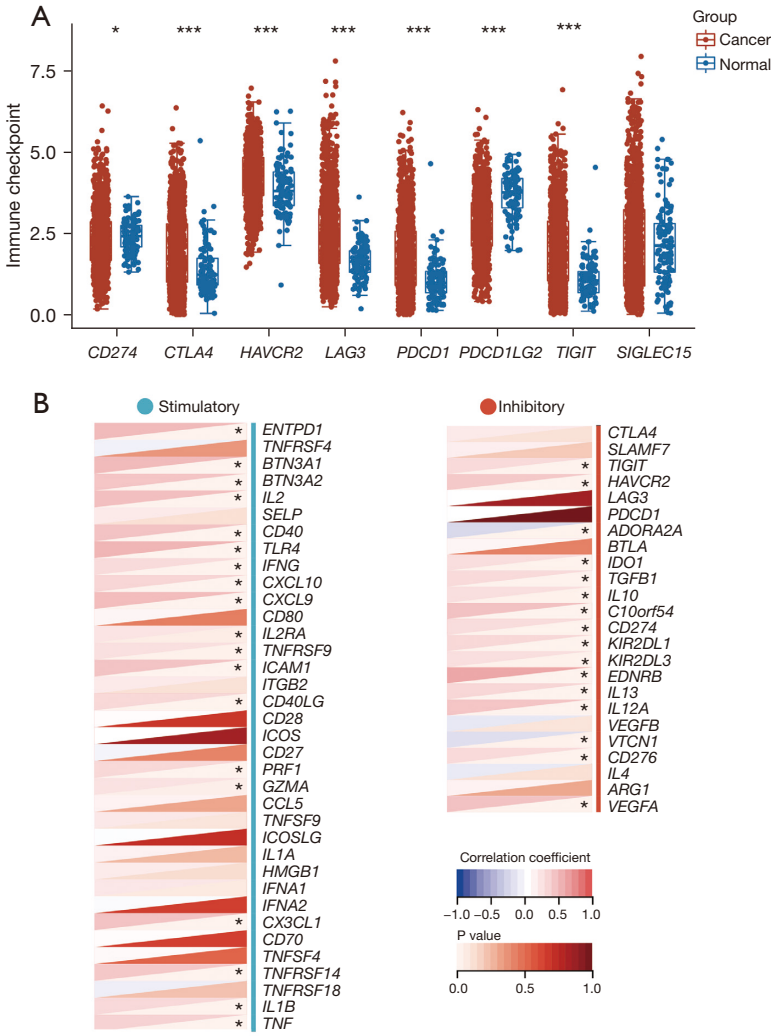

Figure 8 Expression of immune checkpoint genes in breast cancer and their relationship with *STAT4*. (A) Expression of eight immune checkpoint genes in breast tumor tissues and normal tissues, where the horizontal axis represents different genes and the vertical axis represents the distribution of expression of that gene. The different colors represent different groups, and the asterisks represent the levels of significance: *, P<0.05; ***, P<0.001. (B) Relationship between *STAT4* and two kinds of immune checkpoint pathway genes (inhibitory, stimulatory). *STAT*, signal transducer and activator of transcription.

majority of immunomodulators in BRCA.

Discussion

In this study, we mainly carried out an in-depth analysis of the mRNA expression of *STATs* in BRCA, and explored the relationship between *STAT* family genes and BRCA using multiple online databases. The result suggested that compared to normal tissues, the expression levels of *STAT1* and *STAT2* were higher in BRCA tissues, while those of *STAT3*, *STAT4*, *STAT5A*, and *STAT5B* were lower. We also found that *STAT3* was significantly associated with the pathological stage of BRCA, which indicates that *STAT3* might play an important role in the occurrence and development of BRCA (32,33). More importantly, through mutual verification by several database tools, we observed that BRCA patients with a high transcriptional level of *STAT4* were apparently related with long OS. This indicates that *STAT4* could have the potential to become a prognostic marker for BRCA.

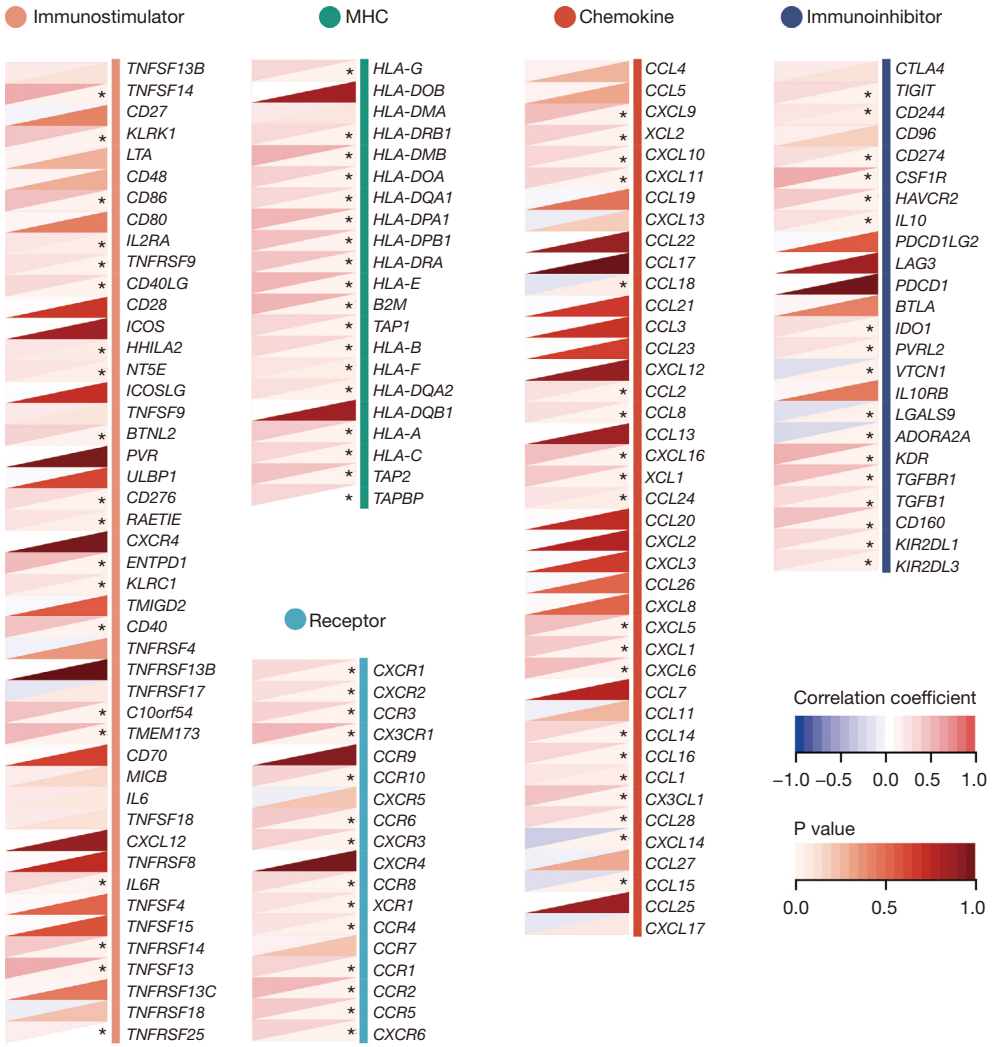

Figure 9 Correlation between *STAT4* and 150 immunomodulators (immunostimulators, MHC, chemokines, immunoinhibitors, and receptors). The asterisks represent levels of significance: *, P<0.05. MHC, major histocompatibility complex; *STAT*, signal transducer and activator of transcription.

To explain the molecular mechanism of *STAT4* on the onset and progression of BRCA, we further selected the top 100 related genes that were co-expressed with *STAT4* and analyzed their associated enrichment pathways. The results showed these co-expressed genes were mainly enriched in the adaptive immune response, T cell activation, T cell receptor signaling pathway, regulation of cytokine production, and negative regulation of immune system processes. Furthermore, the association between the immune system and *STAT4* expression was investigated in depth using several databases, and revealed that *STAT4* was associated with the degree of infiltration of multiple immune cells, the expression of immune checkpoint-related genes, and a majority of immunomodulators in BRCA. Therefore, mining the prognostic significance and immune infiltration of *STAT* family members in BRCA might facilitate the discovery of new biomarkers and further our understanding of important immunotherapies.

As an important component of signal transduction, *STAT* family proteins are associated with the development of a variety of cancers. Previous study has confirmed that activated *STAT1* acts as a tumor suppressor in cancer cells (34). Also, the anti-carcinogenic effect of *STAT1* primarily depends on the phosphorylation of tyrosine

synthase (35). However, the role of *STAT1* is puzzling due to the fact that Greenwood *et al.* observed that *STAT1* positive triple-negative tumors are more aggressive (36). This is probably attributable to the fact that *STAT1* can upregulate the action of various pro-inflammatory cytokines, enzymes, and chemokines, such as tumor necrosis factor α (TNF-α), cyclooxygenase 2 (COX2), and inducible nitric oxide synthase (iNOS) (37-39). In addition, the mRNA expression of *STAT1* was obviously higher in BRCA tissues, but no correlation of *STAT1* on the OS rate of BRCA patients was observed in our study.

Unlike *STAT1*, *STAT2* is a cardinal and specific effector of type I interferon (IFN-I) signaling, which plays an important role in tumor immune surveillance (40). *STAT2* can also regulate the expression of genes involved in the differentiation and recruitment of immune effector cells to the tumor sites (41). Therefore, it is also crucial for *STAT2* to be studied. In recent years, inflammation has attracted attention with regards to tumorigenesis and malignant progression. There is some evidence that *STAT* family proteins, particularly *STAT3*, play a key role in the selective induction and maintenance of an oncogenic inflammatory microenvironment, both at the onset of malignant transformation and during cancer progression (42). Furthermore, *STAT3* has been shown to be particularly active in BRCA, where it promotes cancer progression by stimulating cell proliferation, increasing angiogenesis, influencing escape to the immune system, and providing resistance to apoptosis (33,43). In our research, *STAT3* expression was related to the pathological stage of BRCA, which could perhaps be explained by the mechanism described above.

STAT5A and *STAT5B*, two other important members of the *STAT* family, are two isoforms of *STAT5*. Similar to *STAT3*, structural activation of *STAT5* is a major cause of tumorigenesis (44). However, emerging data has proposed that activated *STAT5* could act as a predictor of better prognosis in BRCA patients (45). According to the database exploration in our study, the enhanced mRNA expression of *STAT5A* portended better OS in all BRCAs. This appears to be associated with the fact that *STAT5A* can inhibit BRCA cells via E-cadherin-mediated junctions (46). As for *STAT6*, we determined that the elevated mRNA expression of *STAT6* was correlated with better OS using the Kaplan-Meier plotter. Also, the findings of Wang *et al.* (47) are generally consistent our results. Nevertheless, further investigation of larger populations is necessary.

Immunotherapy is becoming increasingly popular in cancer prevention and treatment research, and immune cells in the tumor microenvironment are regarded as an important determinant of clinical outcomes and the reaction to immunotherapy (48). In our study, we found that *STAT4* was highly expressed in BRCA compared to normal breast tissue, and its higher expression was associated with better prognosis. Through bioinformatics analysis, we also found that *STAT4* co-expressed genes were focused on immune-related pathways, and the *STAT4* expression level was associated with the infiltration level of multiple immune cells in BRCA. Previous studies have also shown that *STAT4* is engaged in a series of immune response processes, including differentiation and activation of immune cells, interferon-gamma (IFN-γ) production, interleukin (IL)-12 and -23, and IFN signaling in immune cells, and so on (49,50). In recent years, Anderson *et al.* found that *STAT4* deficiency decreases the anti-tumor immune response and promotes the accumulation of immunosuppressive myeloid cell populations to promote tumor metastasis in head and neck squamous cell carcinoma (51). In contrast, knockdown or inhibition of *STAT4* can suppress cancer cell migration, invasion, and metastasis in gastric and colorectal cancers (50,52). However, in terms of BRCA, *STAT4* was still poorly studied in terms of immunity. Although there are multiple treatments for BRCA (2), it is important that more effective treatments are continually being explored. This study may offer some detailed information about immunization to support the design of new immunotherapies.

There were some limitations in this study that should be noted. Firstly, all of the data analyzed were obtained from online databases, and thus, we our findings need to clarify through more cellular experiments and long-term clinical studies. In addition, it is also necessary to further explore the potential mechanisms of action and clinically relevant applications of different *STATs* in BRCA.

Conclusions

In summary, based on the validation of a large amount of data, our study mined the prognostic significance and immune infiltration relevance of *STAT* family members. The results suggest that individual *STATs*, except *STAT1*, may act as a prognostic biomarker for BRCA and provide a reference for further potential immunotherapies.

Acknowledgments

We thank The Cancer Genome Atlas (TCGA) Database

for sharing a large amount of data and the convenience provided by several online database tools.
Funding: This study was funded by the National Natural Science Fund of China (No. 82003183).

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-22-189/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-22-189/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Britt KL, Cuzick J, Phillips KA. Key steps for effective breast cancer prevention. Nat Rev Cancer 2020;20:417-36.
2. Tong CWS, Wu M, Cho WCS, et al. Recent Advances in the Treatment of Breast Cancer. Front Oncol 2018;8:227.
3. Veronesi U, Boyle P, Goldhirsch A, et al. Breast cancer. Lancet 2005;365:1727-41.
4. Savas P, Salgado R, Denkert C, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. Nat Rev Clin Oncol 2016;13:228-41.
5. Esteva FJ, Hubbard-Lucey VM, Tang J, et al. Immunotherapy and targeted therapy combinations in metastatic breast cancer. Lancet Oncol 2019;20:e175-86.
6. Chen H, Pu S, Yu S, et al. A nomogram based on CENPP expression for survival prediction in breast cancer. Gland Surg 2021;10:1874-88.
7. Wegenka UM, Lütticken C, Buschmann J, et al. The interleukin-6-activated acute-phase response factor is antigenically and functionally related to members of the signal transducer and activator of transcription (STAT) family. Mol Cell Biol 1994;14:3186-96.
8. Berg T. Signal transducers and activators of transcription as targets for small organic molecules. Chembiochem 2008;9:2039-44.
9. Darnell JE Jr. STATs and gene regulation. Science 1997;277:1630-5.
10. Shao F, Pang X, Baeg GH. Targeting the JAK/STAT Signaling Pathway for Breast Cancer. Curr Med Chem 2021;28:5137-51.
11. Miklossy G, Hilliard TS, Turkson J. Therapeutic modulators of STAT signalling for human diseases. Nat Rev Drug Discov 2013;12:611-29.
12. Pencik J, Pham HT, Schmoellerl J, et al. JAK-STAT signaling in cancer: From cytokines to non-coding genome. Cytokine 2016;87:26-36.
13. Villarino AV, Kanno Y, O'Shea JJ. Mechanisms and consequences of Jak-STAT signaling in the immune system. Nat Immunol 2017;18:374-84.
14. Sasidharan Nair V, Toor SM, Ali BR, et al. Dual inhibition of STAT1 and STAT3 activation downregulates expression of PD-L1 in human breast cancer cells. Expert Opin Ther Targets 2018;22:547-57.
15. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015;19:A68-77.
16. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1-6.
17. Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 2001;29:137-40.
18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
19. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45:W98-W102.
20. Chen W, Dai X, Chen Y, et al. Significance of STAT3 in Immune Infiltration and Drug Response in Cancer. Biomolecules 2020;10:834.
21. Voutsadakis IA. 8p11.23 Amplification in Breast Cancer: Molecular Characteristics, Prognosis and Targeted

Therapy. J Clin Med 2020;9:3079.
22. Mizuno H, Kitada K, Nakai K, et al. PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics 2009;2:18.
23. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
24. Ciriello G, Gatza ML, Beck AH, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 2015;163:506-19.
25. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. Nucleic Acids Res 2018;46:W60-4.
26. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
27. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
28. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 2017;77:e108-10.
29. Li Q, Pan Y, Cao Z, et al. Comprehensive Analysis of Prognostic Value and Immune Infiltration of Chromobox Family Members in Colorectal Cancer. Front Oncol 2020;10:582667.
30. Sorrelle N, Ganguly D, Dominguez ATA, et al. Improved Multiplex Immunohistochemistry for Immune Microenvironment Evaluation of Mouse Formalin-Fixed, Paraffin-Embedded Tissues. J Immunol 2019;202:292-9.
31. Li B, Chan HL, Chen P. Immune Checkpoint Inhibitors: Basics and Challenges. Curr Med Chem 2019;26:3009-25.
32. Segatto I, Baldassarre G, Belletti B. STAT3 in Breast Cancer Onset and Progression: A Matter of Time and Context. Int J Mol Sci 2018;19:2818.
33. Eroglu M, Kokenek-Unal TD, Akin-Bali DF, et al. STAT3 expression is correlated with pathological stage in luminal subtypes of breast carcinoma. Bratisl Lek Listy 2020;121:51-61.
34. Koromilas AE, Sexl V. The tumor suppressor function of STAT1 in breast cancer. JAKSTAT 2013;2:e23353.
35. Meissl K, Macho-Maschler S, Müller M, et al. The good and the bad faces of STAT1 in solid tumours. Cytokine 2017;89:12-20.
36. Greenwood C, Metodieva G, Al-Janabi K, et al. Stat1 and CD74 overexpression is co-dependent and linked to increased invasion and lymph node metastasis in triple-negative breast cancer. J Proteomics 2012;75:3031-40.
37. Samardzic T, Jankovic V, Stosic-Grujicic S, et al. STAT1 is required for iNOS activation, but not IL-6 production in murine fibroblasts. Cytokine 2001;13:179-82.
38. Akram M, Kim KA, Kim ES, et al. Selective inhibition of JAK2/STAT1 signaling and iNOS expression mediates the anti-inflammatory effects of coniferyl aldehyde. Chem Biol Interact 2016;256:102-10.
39. Kaur T, Mukherjea D, Sheehan K, et al. Short interfering RNA against STAT1 attenuates cisplatin-induced ototoxicity in the rat by suppressing inflammation. Cell Death Dis 2011;2:e180.
40. Arimoto KI, Löchte S, Stoner SA, et al. STAT2 is an essential adaptor in USP18-mediated suppression of type I interferon signaling. Nat Struct Mol Biol 2017;24:279-89.
41. Yue C, Xu J, Tan Estioko MD, et al. Host STAT2/type I interferon axis controls tumor growth. Int J Cancer 2015;136:117-26.
42. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer 2009;9:798-809.
43. Wang T, Niu G, Kortylewski M, et al. Regulation of the innate and adaptive immune responses by Stat-3 signaling in tumor cells. Nat Med 2004;10:48-54.
44. Bowman T, Garcia R, Turkson J, et al. STATs in oncogenesis. Oncogene 2000;19:2474-88.
45. Nevalainen MT, Xie J, Torhorst J, et al. Signal transducer and activator of transcription-5 activation and breast cancer prognosis. J Clin Oncol 2004;22:2053-60.
46. Sultan AS, Xie J, LeBaron MJ, et al. Stat5 promotes homotypic adhesion and inhibits invasive characteristics of human breast cancer cells. Oncogene 2005;24:746-60.
47. Wang S, Yu L, Shi W, et al. Prognostic roles of signal transducers and activators of transcription family in human breast cancer. Biosci Rep 2018;38:BSR20171175.
48. Sun J, Zhang Z, Bao S, et al. Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. J Immunother Cancer 2020;8:e000110.
49. Korman BD, Kastner DL, Gregersen PK, et al. STAT4: genetics, mechanisms, and implications for autoimmunity. Curr Allergy Asthma Rep 2008;8:398-403.
50. Cheng JM, Yao MR, Zhu Q, et al. Silencing of stat4 gene inhibits cell proliferation and invasion of colorectal cancer cells. J Biol Regul Homeost Agents 2015;29:85-92.
51. Anderson K, Ryan N, Volpedo G, et al. Immune

Suppression Mediated by STAT4 Deficiency Promotes Lymphatic Metastasis in HNSCC. Front Immunol 2019;10:3095.
52. Zhou X, Xia Y, Su J, et al. Down-regulation of miR-141 induced by helicobacter pylori promotes the invasion of gastric cancer by targeting STAT4. Cell Physiol Biochem 2014;33:1003-12.

(English Language Editor: A. Kassem)

Cite this article as: Zhou Y, Jiang S, Yu S, Zhu L, Liu Y, Li S, Hao N, Ren Y. Mining the prognostic significance and immune infiltration of STAT family members in human breast cancer by bioinformatics analysis. Gland Surg 2022;11(4):720-741. doi: 10.21037/gs-22-189

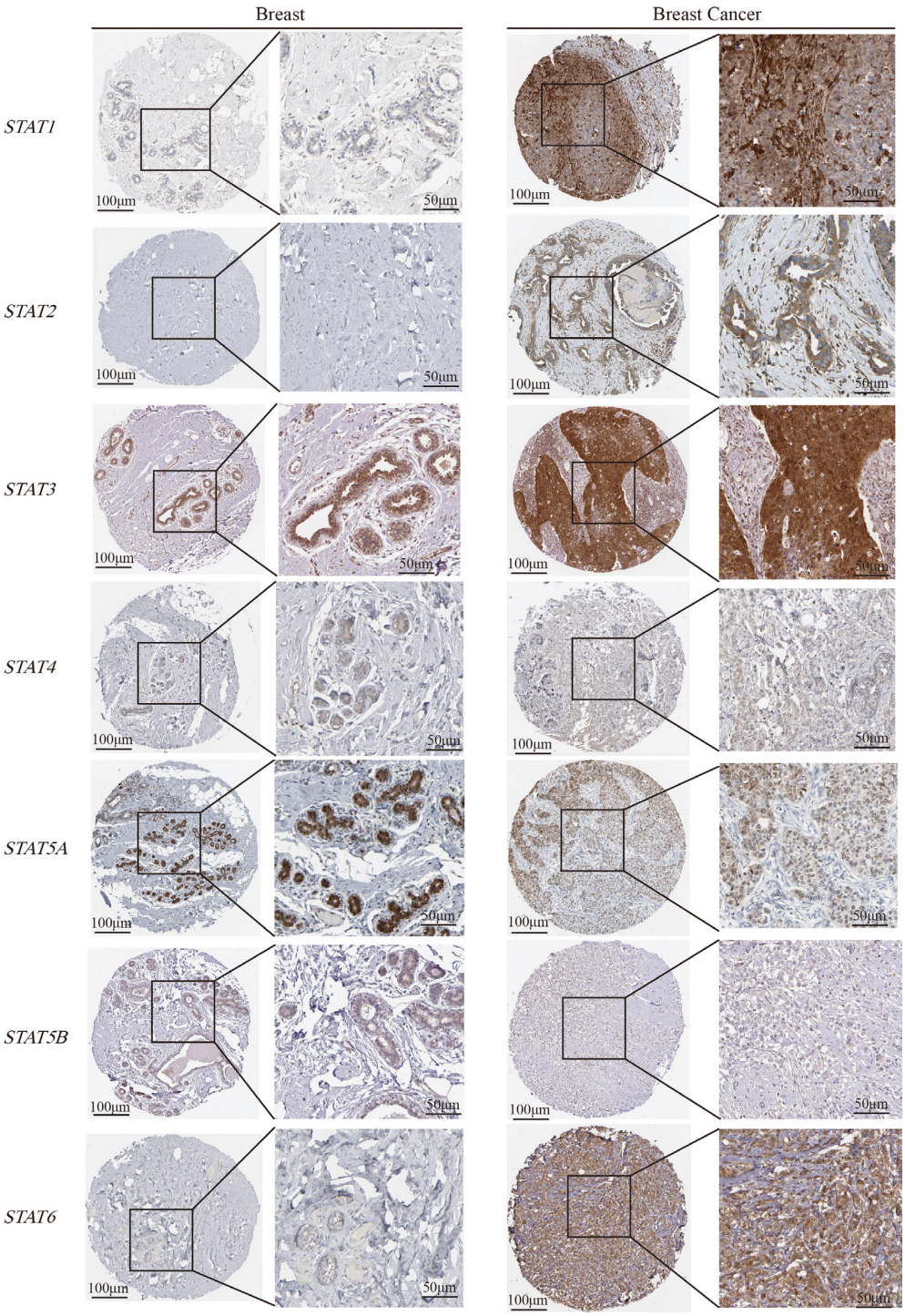

Figure S1 Typical immunohistochemistry images of the protein expression of the *STAT* family from the HPA database. *STAT1*, *STAT2*, *STAT3*, and *STAT6* showed low staining in normal breast tissues, but high staining in breast cancer (P<0.05). Breast cancer tissues exhibited lower protein expression levels of *STAT5A* and *STAT5B* compared with normal tissues. HPA, Human Protein Atlas; *STAT*, signal transducer and activator of transcription.

https://dx.doi.org/10.21037/gs-22-189

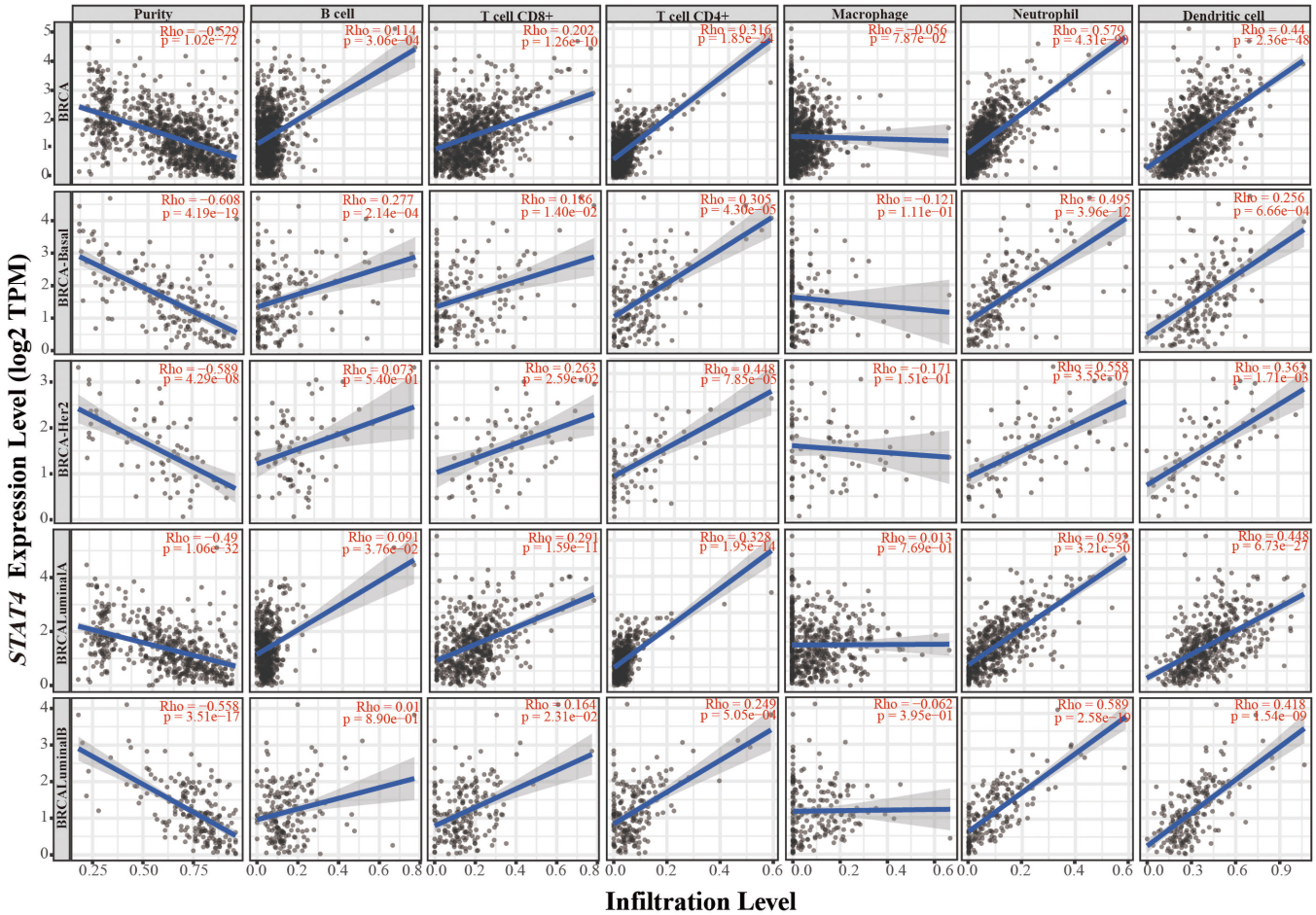

Figure S2 Correlations between differential immune cell infiltration (TIMER) and the expression level of STAT4 among the different breast cancer subtypes. STAT, signal transducer and activator of transcription; TIMER, Tumor Immune Estimation Resource; TPM, transcripts per million.

https://dx.doi.org/10.21037/gs-22-189