International Clinical and Medical Case Reports Journal
Research Article (ISSN: 2832-5788)

PPARG: A New Prognostic Biomarker and Correlates with Immune Infiltrates in Breast Invasive Carcinoma Based on Bioinformatics Analysis

Yaohui Jiang[†], Xiaoyu Zhou[†], Yalin Ju, Yuxin Ma,Yujiao Dong, Shaobo Wu[*]

Health Science Center, Xi'an Jiaotong University, Xi'an 710061, Shaanxi Province, PR China

[†]These authors contributed equally to this work and should be considered co-first authors

Citation: *Yaohui Jiang, Xiaoyu Zhou, Yalin Jv, Yuxin Ma,Yujiao Dong, Shaobo Wu. PPARG: A New Prognostic Biomarker and Correlates with Immune Infiltrates in Breast Invasive Carcinoma Based on Bioinformatics Analysis. Int Clinc Med Case Rep Jour. 2023;2(4):1-15.*

Received Date: 30 January, 2023; **Accepted Date:** 03 February, 2023; **Published Date:** 05 February, 2023

[*]**Corresponding author:** Shaobo Wu. Health Science Center, Xi'an Jiaotong University, Xi'an 710061, Shaanxi Province, PR China

1. ABSTRACT

PPARG (Peroxisome Proliferator Activated Receptor Gamma) is a Protein Coding gene. This gene encodes a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. The protein encoded by this gene is PPAR-gamma and is a regulator of adipocyte differentiation. However, the relationship between PPARG expression, tumor immunity, and prognosis in Breast invasive carcinoma (BRCA) remain unclear. PPARG expression and its influence on tumor prognosis were analyzed by the ONCOMINE, Tumor Immune Estimation Resource (TIMER) and Kaplan-Meier plotter. The relationship between BRCA expression and tumor immunity were analyzed by TIMER and Gene Expression Profiling Interactive Analysis (GEPIA). PPARG expression in BRCA tissues correlated with prognosis. PPARG expression was significantly lower in several human cancers, including HCC, than in corresponding normal tissues. Moreover, we screened hub genes by constructing a protein-protein interaction (PPI) network. These findings demonstrate that PPARG is a potential prognostic biomarker that determines cancer progression and correlated with tumor immune cells infiltration in BRCA.

Keywords: PPARG; Biomarker; Immune infiltrates; Breast cancer

2. INTRODUCTION

Breast cancer is the most frequently diagnosed cancer in women and ranks second among causes for cancer related death in women.[1] Breast cancer is the most frequent malignancy in women worldwide and is curable in ~70–80% of patients with early-stage, non-metastatic disease. Advanced breast cancer with distant organ metastases is considered incurable with currently available therapies. On the molecular level, breast cancer is a heterogeneous disease; molecular features include activation of human epidermal growth factor receptor 2 (HER2, encoded by ERBB2), activation of hormone receptors (oestrogen receptor and progesterone receptor) and BRCA mutations. Treatment strategies differ according to molecular subtype. Breast cancer is the second

highest incidence rate of cancer.[2] Therefore, a deeper understanding of the molecular mechanisms involved in BRCA pathogenesis is critical for developing new treatments to improve survival rates.

The Genechip, a high-throughput platform and effective technique for gene expression is widely used in multiple disease research areas, and can detect the expression of thousands of genes and the association between disease and genes.[3-5] Although several bioinformatical studies on BRCA have been reported in recent years, we could sieve through different target genes through analyzing distinct databases, which could assist us in further exploring and better studying the underlying mechanisms.

In this study, we downloaded two gene expression profiles (GSE15852, GSE42568) from the Gene Expression Omnibus (GEO) database. A total of 207 breast tissue samples were included in this study, consisting of 60 normal breast tissue samples and 147 BRCA samples. Then, the GEO2R online tool, volcano map software, and Venn diagram online tool were applied for DEGs in the four data sets above. Furthermore, Gene Ontology (GO) analysis and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis were performed to analyze these DEGs' molecular function (MF), cellular component (CC), biological process (BP), and KEGG pathways using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) and KOBAS, respectively. We constructed a protein-protein interaction (PPI) network with Cytoscape MCODE (Molecular Complex Detection) for identifying 11 hub genes with high connectivity degrees. Moreover, we verified the differential expression of 11 screened hub genes and prognosis for survival of patients with BRCA in the large-scale gene data sets regarding BRCA from The Cancer Genome Atlas (TCGA) database. As a result, only PPARG qualified for the prognosis target gene. Finally, we carried out gene set enrichment analysis (GSEA) for these three genes and potential molecular mechanisms in BRCA. In conclusion, this study provides some additional effective biomarkers for BRCA patients and contributes to the understanding of the molecular mechanisms of BRCA progression. Then, we comprehensively analyzed the expression of PPARG, its correlation with prognosis in different types of tumors including BRCA, and the status of different tumor-infiltrating immune cells based on expression of specific markers using the ONCOMINE, Kaplan-Meier plotter, Gene Expression Profiling Interactive Analysis (GEPIA), and Tumor Immune Estimation Resource (TIMER) databases. Our results shed light on the important role of PPARG in BRCA prognosis and provided an underlying mechanism that PPARG expression might modulate tumor immunity by regulating the infiltration of immune cells in BRCA.

3. MATERIALS AND METHODS

3.1 Microarray data information

The GEO (http://www.ncbi.nlm.nih.gov/geo/) database is a free public database of microarrays and is used for gene expression data sets and platform records. The gene expression profiles of GSE15852 and GSE42568 were chosen from the GEO database. GSE15852 was based on the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array, containing 43 breast cancer tissues and 43 normal breast tissues. GSE42568 was based on the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, containing 104 breast cancer tissues and 17 normal breast tissues. The downloaded data was processed using the R package, then calibrated, standardized, and log2 transformation was performed on all data of gene expression.

3.2 Screening for DEGs

DEGs between BRCA samples and normal endometrium samples were identified using the GEO2R web tool (http://www.ncbi.nlm.nih.gov/geo/geo2r/).[6] GEO2R is an interactive online tool, which compares two groups of samples in most GEO series to obtain genes with different expressions under the same experimental conditions. The DEGs with |log FC|≥ 1 and adjust P< 0.05 were considered as the cut-off criteria. Then, we applied for Venn software online (http://bioinfogp.cnb.csic.es/tools/venny/) to obtain the common DEGs in all four in-dependent cohorts.

3.3 GO term and KEGG pathway enrichment analysis of DEGs

GO analysis, as a common approach, was conducted to identify the unique biological properties of DEGs, including biological processes, cellular components, and molecular functions.[7] KEGG pathway enrichment analysis was performed to explore key pathways to initiation and progression of BRCA.[8] The GO annotation enrichment of DEGs were performed using DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/),[9] which consisted of systematic and integrative functional annotation tools to integrate functional genomic annotations. The KEGG pathway analysis of DEGs was performed using the KOBAS 3.0 online analysis tool (http://kobas.cbi.pku.edu.cn/).[10] We regarded P<0.05 and counts more than 2 as a statistically significant difference and significant enrichment, respectively.

3.4 PPI network and module analysis

The PPI network is essential to interpret the key genes and important gene modules in cancer development. The network analysis software Cytoscape (www.cytoscape.org; version 3.2.1)[11] was used to visualize networks and screen hub genes from these DEGs according to degrees. Subsequently, the MCODE (Molecular Complex Detection) app in Cytoscape software was used to screen modules within the PPI network with the default parameters “Degree Cutoff = 7,” “Node Score Cutoff = 0.2,” “K-Core = 2,” and “Max.Depth = 100.” Finally, the interactions of all hit genes were constructed using String (https://string-db.org, version 11.0).[12]

3.5 Validation of the hub genes in TCGA database

The expression of upregulated and downregulated hub genes was validated in the TCGA_BRCA data set after hub genes were determined from these four GEO datasets. Furthermore, we analyzed the association between expression of hub genes and biological processes/pathways and phenotypes using GSEA (GSEA v2.0, available online: http://www.broad.mit.edu/gsea/).

3.6 Statistical analysis

The results of RNA seq v2 of unpaired samples were analyzed using an independent sample t test. Hub genes levels from G1 to G4 was studied using a two-sided Student's test for two-group comparisons and by one-way analysis of variance, followed by a Bonferroni post hoctest, for multiple group comparisons. The Kaplan-Meier (KM) curve was conducted to assess the association between the expression level of hub genes and survival time of patients with BRCA. All statistical analyses were performed with the GraphPad Prism 5 (GraphPad Software, Inc, La Jolla, CA). Data were presented as the mean ±standard deviation. Statistical significance was considered significant when P< 0.05. Gene expression data from the Oncomine database were analyzed using the P-values, fold changes, and ranks. Survival curves were produced by the Kaplan-Meier plots and GEPIA database. The correlation of gene expression was evaluated in the TIMER and GEPIA databases using Spearman’s correlation analysis. P-values <0.05 were considered as statistically significant.

3.7 CCL14 gene expression analysis

The mRNA levels of PPARG in several cancers including BRCA were identified from the Oncomine database (https://www.oncomine.org/resource/login.html).[13] The threshold was determined as follows: fold change of 1.5, P-value of 0.001, and gene ranking of all.

3.8 Kaplan-Meier survival curve analysis

Kaplan-Meier survival curve analysis was performed to assess the correlation between the expression of the 54,000 genes on the survival rates in 21 different cancers using more than 10,000 cancer samples, including 371 liver, 1,440 gastric, 3,452 lung, 2,190 ovarian, and 6,234 breast cancer samples. Kaplan-Meier plots (http://kmplot.com/analysis/) were used to analyze the relationship between CCL14 gene expression and survival rates in liver, gastric, breast, pancreatic, ovarian, and lung cancers based on the hazard ratios (HR) and log-rank P-values.[14]

3.9 TIMER analysis

TIMER database was used to systematically analyze the tumor-infiltrating immune cells (TIICs) in 32 cancer types using more than 10,000 samples from The Cancer Genome Atlas (TCGA) (https://cistrome.shinyapps. io/timer/) database.[15] TIMER determines the abundance of tumor-infiltrating immune cells (TIICs) based on the statistical analysis of gene expression profiles.[16] We analyzed the association between the level of CCL14 gene expression and the abundance of infiltrating immune cells, including CD4+ T cells, CD8+ T cells, B cells, neutrophils, dendritic cells and macrophages based on expression of specific marker genes in different cancers including HCC. The marker genes used for analysis of tumor-infiltrating immune cells including T cells, B cells, TAMs, monocytes, M1 macrophages, M2 macrophages, natural killer (NK) cells, neutrophils, dendritic cells (DCs), T-helper (Th) cells, T-helper 17 (Th17) cells, follicular helper T (Tfh) cells, exhausted T cells, and Tregs were based on data from previous studies.[17] CCL14 gene was on the x-axis and related marker genes are on the y-axis.

3.10 GEPIA analysis

The Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/index. html) was used to analyze the RNA sequencing expression data from 113 normal and 1109 tumor tissue samples from the TCGA and GTEx projects.[18] We also used GEPIA to generate survival curves and determine OS and DFS rates and their correlation to specific gene expression in 33 different types of cancer to further confirm the significantly correlated genes in the TIMER analysis.

4. RESULTS

4.1 Microarray data information and identification of DEGs in BRCA

Data from each GEO data set was respectively analyzed using GEO2R online tools to screen DEGs (|log FC|≥1 and adjust P<0.05). In this study, 104 BRCA tissues and 17 normal breast tissues from four GEO data sets were involved. A total of 1,731 upregulated and 1,688 downregulated genes were filtered from GSE42568; 70 upregulated and 150 downregulated genes from GSE15852. The DEGs of each gene expression profile data were visualized using a volcano plot (Figure 1A). Subsequently, 52 overlapping upregulated genes and 146 overlapping downregulated genes were screened in the BRCA group compared with the normal control group using Venn software online (Figure 1B).

4.2 GO functional enrichment analysis

Aiming at explore the biological functions of the integrated DEGs in BRCA, GO term enrichment analysis of the 198 DEGs (52 upregulated genes and 146 downregulated genes) was performed using the DAVID online analysis tool. In biological process terms, the DEGs were significantly enriched in multicellular organismal homeostasis (Figure 1C). For molecular functions, the DEGs were mainly enriched in sulfur compound binding (Figure 1D). In cellular component terms, the DEGs were mainly involved in the collagen−containing extracellular matrix (Figure 1E). The significant GO enrichment terms are shown in bubble maps using the ggplot2.R package with $P< 0.05$.

4.3 KEGG pathway analysis

As shown in Figure 1F, KEGG pathway analysis of the 198 aberrant DEGs was conducted in BRCA using the KOBAS 3.0 online analysis tool, of which the top was Fatty acid degradation. The analysis indicated that the 122 dysregulated genes were mainly related to human cancer.

4.4 PPI network construction, module analysis, and hub gene selection

The PPI network was constructed for further investigation of the interaction among the integrated DEGs via Cytoscape software, based on the STRING database. The PPI network consisted of 155 nodes and 461 edges. We regarded the top 11 DEGs with a high degree of connectivity as the hub genes of BRCA (Figure 1G). The identified hub genes were all downregulated (PCK1, FABP4, LPL, CEBPA, CIDEC, ADIPOQ, CFD, LEP PLIN1, CIDEA and PPARG) genes. The String database was used to construct an interactive network of all hit genes. Interestingly, all the hub genes formed a network of interactions (Figure 1H).

4.5 Hub gene validation

To further evaluate the 11 candidate hub genes that were dysregulated in BRCA, publicly available transcriptome data were downloaded from the TCGA_BRCA database. We analyzed RNA-Seq profiles of 113 normal and 1109 cancer samples from the TCGA_BRCA data set. All hit genes were remarkably differently expressed in BRCA compared with control group, their expression tendency in TCGA data set was consistent with the messenger RNA (mRNA) level from GEO data set (Figure 2A-K). Then we download pathological section of BRCA from human protein atlas (https://www.proteinatlas.org/search/HAMP) and we found that the PPARG expression products level was obviously decreased in breast cancer tissues compared with normal lung tissues (Figure 1I).

4.6 Prognostic significance of hubgenes in patients with BRCA

To further elucidate whether these hub genes were potentially prognostic markers for BRCA, we analyzed the overall survival (OS) and disease-free survival (DFS) for each hub gene by an online KM survival analysis tool (KMplot, http://kmplot.com/analysis/) and the patients were split by auto select best cutoff. Among those 11 hub genes, only the downregulated gene PPARG was associated with BRCA prognosis. The low expression of PPARG mRNA level was significantly associated with worse OS and DSS in patients with BRCA (Figures 3A and 3B). Therefore, PPARG could be prognostic markers for BRCA.

4.7 Clinical stage analyses

To clarify the clinical significance of PPARG expression in BRCA, we analyzed their expression in different stages from TCGA. As shown in Figure 3F, increased PPARG expression were significantly associated with advanced tumor stage. We observed PPARG could distinguish patients with poor prognosis for DSS, PFI

(Figure 3C and 3E). These results demonstrated that might be associated with BRCA progression and might have prognostic significance for BRCA patients.

4.8 Hub gene GSEA analysis

To further identify the possible mechanism of these hub genes in BRCA, GSEA was conducted to obtain the biological pathway from a database to a gene set. The patients from the TCGA-BRCA data set were divided into high and low-expression groups based on the median value of PPARG. As shown in Figure 2L, GSEA analysis revealed that the low expression of these hub genes was enriched in the "cytokine receptor interaction" and "cell surface interactions at the vascular" genesets from TCGA_BRCA data sets.

4.9 The levels of PPARG expression correlate with the infiltration levels of immune cells in BRCA

The survival times of patients in several cancers is determined by the quantity and activity status of tumor infiltrating lymphocytes. Therefore, we analyzed the association of PPARG expression with prognosis and immune infiltration in BRCA. The analysis of immune infiltration by genomic methods is highly influenced by tumor purity in clinical samples. Moreover, GEPIA and TIMER databases contain most of the homologous TCGA. Therefore, we selected BRCA in the TIMER database that show significant correlation between PPARG expression and tumor purity, and in the GEPIA database that PPARG expression is relevant to tumor prognosis. We observed that low PPARG expression correlated with poorer prognosis and high infiltration of most immune cell types in BRCA (Figure 4). The level of PPARG expression negatively correlated with the infiltration levels of B cells (r=-0.039, P=2.28e-01), CD4+ T cells (r=0.25, P=3.18e-15), CD8+ T cells (r=0.279, P=5.96e-19), macrophages (r=-0.226, P=2.10e-17), neutrophils (r=0.176, P=4.75e-08), and DCs (r=-0.186, P=8.11e-09) in BRCA tissues. The data suggest that CCL14 expression modulates infiltration of immune cells into tumor tissues.

4.10 Correlation analysis between mRNA levels of PPARG and markers of different subsets of immune cells

Next, we investigated the correlation between PPARG expression and the status of tumor-infiltrating immune cells based on the levels of immune marker gene expression in BRCA tissues using the TIMER and GEPIA databases. The immune cells analyzed in BRCA tissues included CD8+ T cells, T cell (general), Monocyte, tumor-associated macrophages (TAMs), monocytes, M1 and M2 macrophages, neutrophils, DCs, and natural killer (NK) cells. Moreover, different subsets of T cells, namely, T helper 1 (Th1), Th2, follicular helper T (Tfh), Th17, regulatory T (Tregs), and exhausted T cells were also analyzed. Since tumor purity of clinical samples influences the analysis of immune infiltration, the correlation analysis was adjusted for purity (Table 1). Analysis of the TIMER and GEPIA databases showed that PPARG expression in BRCA tissues significantly correlated with the expression of marker genes from tumor-infiltrating M1 and M2 Macrophage, Monocyte, Th1, Tfh, Neutrophils, B cell, CD8+ T cell (Figure 5 and Table 2).

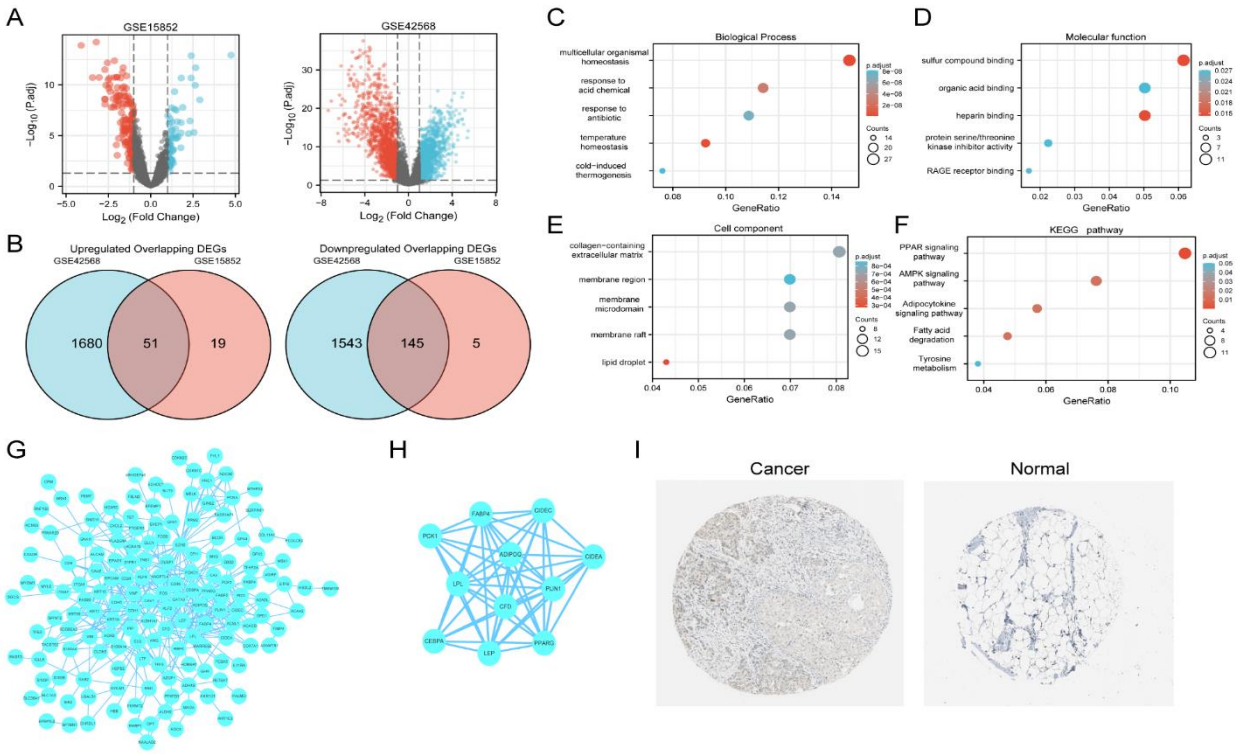

Figure 1: Identification of overlapping DEGs, functional analysis and interaction network construction in BRCA. A, Volcano plots of gene expression profile data in GSE15852 and GSE42568, respectively. B, Venn plots of upregulated overlapping DEGs and Venn plots of downregulated overlapping DEGs. C, Biological process of the DEGs. D, Molecular function of the DEGs. E, Cell component of the DEGs. F, KEGG pathway of the DEGs. G, The PPI network of the DEGs. H, Top 11 hub genes PPI network constructed by STRING online database. I, Immunohistochemical staining of PPARG was performed in breast cancer and normal tissue. DEG, differentially expressed gene; BRCA, Breast invasive carcinoma; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Gene and Genome; PPI, protein-protein interaction.

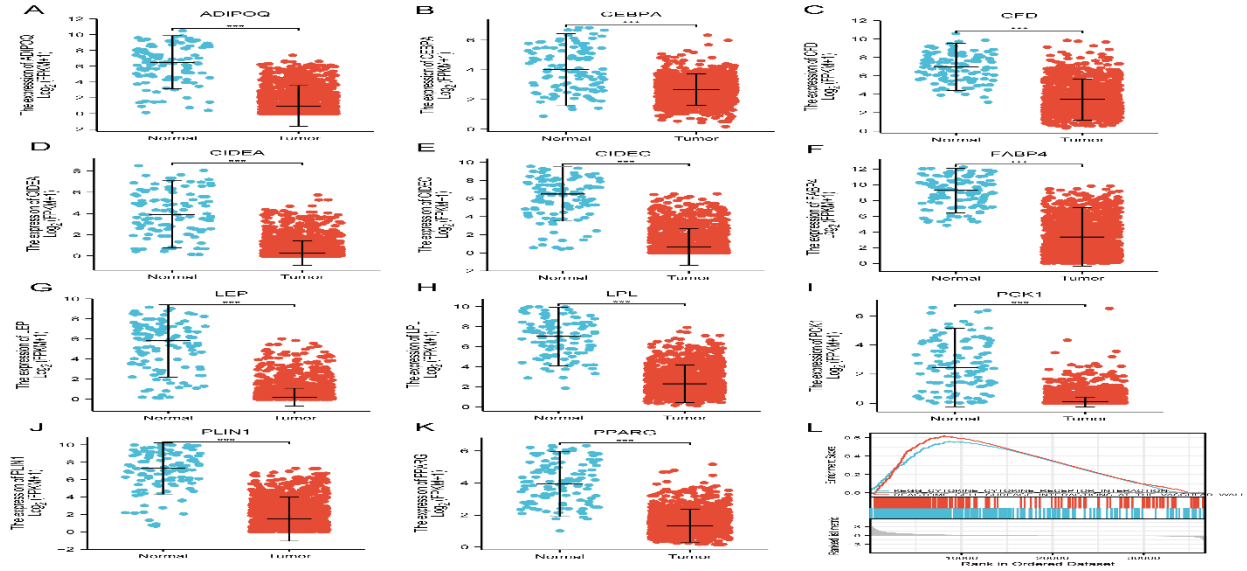

Figure 2: Expression validation of nine hub targets in BRCA compared with adjacent tissues from TCGA data sets. A, ADIPOQ. B, CEBPA. C, CFD. D, CIDEA. E, CIDEC. F, FABP4. G, LEP. H, LPL. I, PCK1. J, PLIN1. K, PPARG. L, Enrichment plots from GSEA. "cytokine receptor interaction" and "cell surface interactions at the vascular" were significantly enriched. BRCA, Breast invasive carcinoma; TCGA, The Cancer Genome Atlas; GSEA, Gene Set Enrichment Analysis.

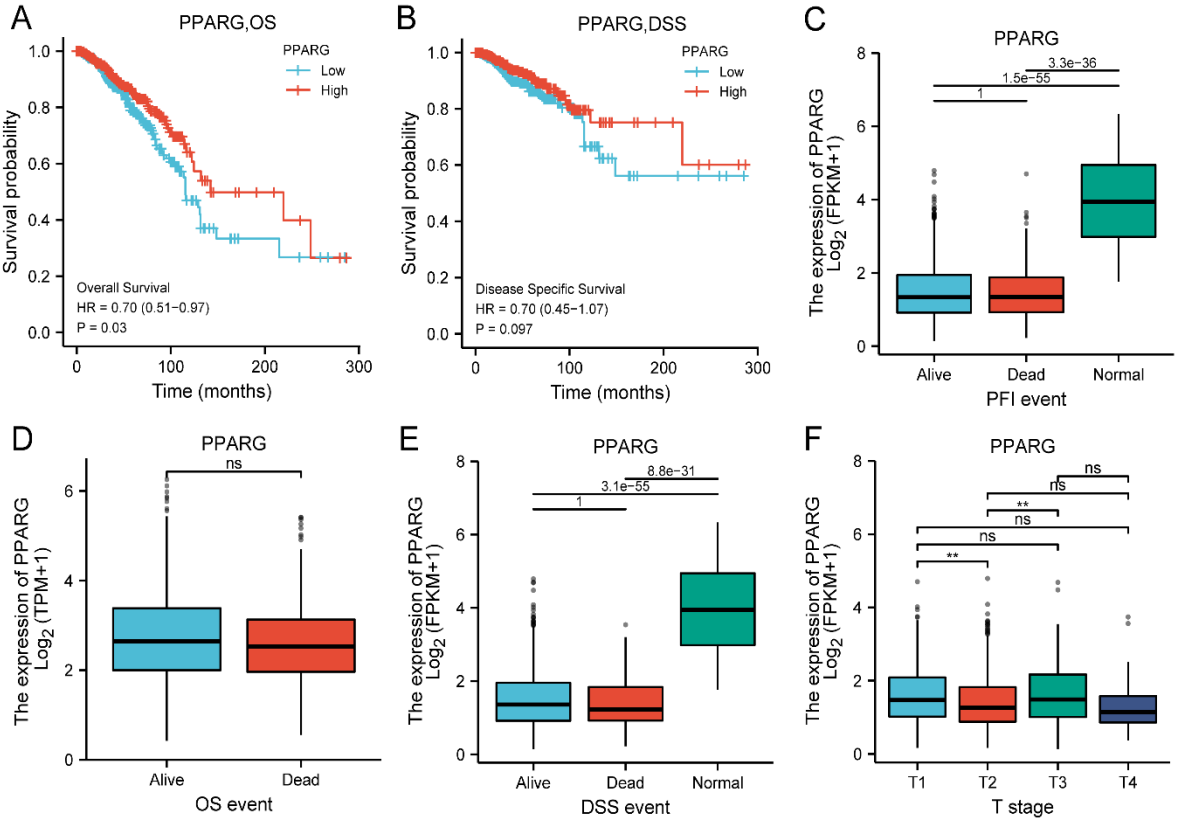

Figure 3: Prognostic gene signature in the BRCA patients. A, Association between PPARG expression and overall survival (OS) were performed with the online Kaplan-Meier survival analysis (KMplot, http://kmplot.com/analysis/) (P<0.05). Low expression of PPARG was significantly associated with shorter OS time of BRCA patients. B, Online Kaplan-Meier survival analysis revealed a shorter DSS time with low PPARG expression level .C, the mRNA levels of the potential prognostic markers (PPARG) is extracted from the TCGA_BRCA data set and mRNA levels between alive and dead patients for PFI, OS(D) and DFS(E). F, The clinical significance of PPARG mRNA with T stage in TCGA. DSS, Disease Specific Survival; PFI, platinum-free interval; BRCA, Breast invasive carcinoma; mRNA, messenger RNA; OS, overall survival; TCGA, The Cancer Genome Atlas.

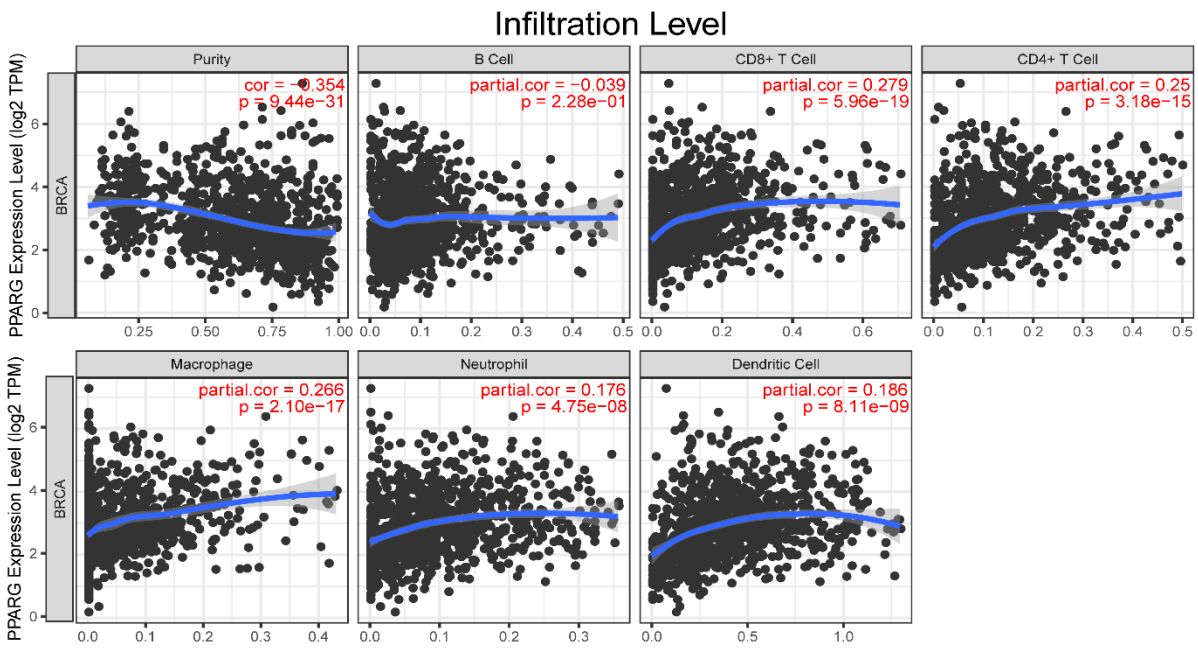

Figure 4: Correlation analysis of PPARG expression and infiltration levels of immune cells in BRCA tissues using the TIMER database. PPARG expression in BRCA tissues negatively correlates with tumor purity and infiltration levels of B cells, and positively correlates with CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells.

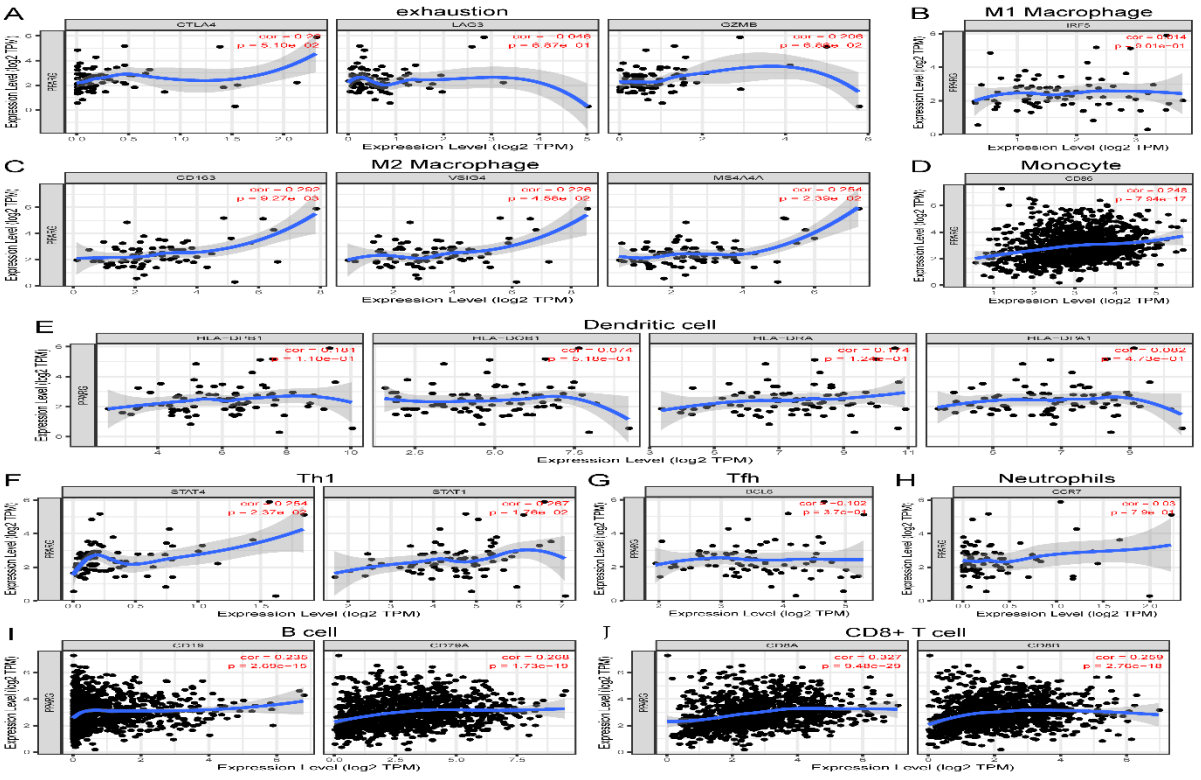

Figure 5: Correlation analysis of PPARG expression and the expression of marker genes of infiltrating immune cells in BRCA (A-J) using the TIMER database. The scatter plots show correlation between CCL14 expression and the gene markers of (A) exhaustion (CTLA4, LAG3, GZMB); (B) M1 Macrophage (IRF5); (C) M2 Macrophage (CD163, VSIG4 and MS4A4A); (D) Monocyte (CD86); (E) DCs (HLA-DPB1, HLA-DQB1, HLA-DRA, HLA-DPA1); (F) Th1 (STAT4 and STAT1); (G) Tfh (BCL6); (H) Neutrophils (CCR7); (I) B cell (CD19 and CD79); (J) CD8+ T cell (CD8A and CD8B).

Table 1: Correlation analysis between CCL14 and relate genes and markers of immune cells in TIMER.

Description	Gene markers	BRCA			
		None		Purity	
		Core	P	Core	P
CD8+ T cell	CD8A	0.326564317	9.48E-29	-0.5067317	5.10E-66
	CD8B	0.258754319	2.76E-18	-0.47988955	1.91E-58
T cell (general)	CD3D	0.307295929	1.74E-25	-0.54908489	1.96E-79
	CD3E	0.339940259	3.70E-31	-0.5603989	2.37E-83
	CD2	0.312522543	2.39E-26	-0.52754995	2.25E-72
B cell	CD19	0.235242905	2.69E-15	-0.46527842	1.35E-54
	CD79A	0.267567835	1.73E-19	-0.50369749	3.96E-65
Monocyte	CD86	0.247581312	7.94E-17	-0.37684359	6.28E-35
TAM	CCL2	0.03093963	0.78624198	-0.50282623	4.98E-06
	CD68	0.313096397	0.00513428	-0.52888717	1.27E-06
	IL10	0.333982849	0.00262912	-0.51947624	2.11E-06
M1 Macrophage	IRF5	0.014167478	0.90123127	-0.40577632	0.00033475
M2 Macrophage	CD163	0.291869523	0.00927245	-0.45948528	3.82E-05
	VSIG4	0.225608569	0.04579146	-0.56315055	1.76E-07
	MS4A4A	0.25443038	0.02390235	-0.63009022	1.80E-09
Neutrophils	CCR7	0.030410298	0.79020385	-0.53658766	8.31E-07
Natural killer cell	KIR2DL1	0.238523044	0.03426791	-0.33926974	0.00310585
	KIR2DL3	-0.02237195	0.84484404	-0.35673653	0.0018094
	KIR2DL4	0.143086097	0.20839522	-0.27139321	0.01933807
	KIR3DL1	0.235183607	0.03694237	-0.30169822	0.00899363
	KIR3DL2	0.074558624	0.51373139	-0.28297404	0.01456782
	KIR3DL3	0.242804862	0.03107783	-0.39410039	0.00051237
	KIR2DS4	0.161574264	0.15485819	-0.27414051	0.01809974
Dendritic cell	HLA-DPB1	0.180988315	0.11035796	-0.54989018	3.88E-07
	HLA-DQB1	0.073734177	0.51763721	-0.32689638	0.00447273
	HLA-DRA	0.174415774	0.12407141	-0.55066147	3.71E-07
	HLA-DPA1	0.081840312	0.47257042	-0.5204474	2.00E-06
Th1	STAT4	0.254346728	0.02369814	-0.49580665	7.06E-06

	STAT1	0.267039922	0.01760704	0.050445782	0.66949946
Th2	GATA3	0.112925998	0.321043	0.180305478	0.12422739
	STAT6	0.028042843	0.80584879	0.146576072	0.21270128
	STAT5A	0.271859786	0.01561052	-0.22128804	0.05812601
	IL13	-0.08696117	0.44602828	0.417275542	0.00021674
Tfh	BCL6	-0.10209348	0.36986771	0.249677699	0.03192522
Th17	STAT3	0.147541383	0.19402644	0.062341553	0.59773119
	IL17A	0.132057967	0.24599291	-0.2383423	0.04085836
Treg	FOXP3	-0.17397761	0.12503023	-0.19561644	0.09485549
	CCR8	0.086648285	0.44767255	0.082533923	0.4844955
	STAT5B	0.160053554	0.15855396	-0.03629545	0.75883507
exhaustion	CTLA4	0.220388116	0.05097393	-0.50365828	4.78E-06
	LAG3	-0.04598345	0.68680385	-0.25857728	0.02611645
	GZMB	0.205964336	0.06859925	-0.40592504	0.0003329

Table 2: Correlation analysis between CCL14 and marker genes of immune cells in GEPIA.

Description	**Gene markers**	**HCC**			
		Tumor		**Normal**	
		R	**P**	**R**	**P**
exhaustion	CTLA4	0.19	8.2e−10	−0.066	0.49
	LAG3	0.003	0.91	−0.22	0.019
	GZMB	0.16	5.6e−08	0.36	8.6e−05
M1 Macrophage	IRF5	0.15	5e−07	−0.064	0.5
M2 Macrophage	CD163	0.25	1.4e−16	0.6	2.4e−12
	VSIG4	0.3	2.4e−23	0.56	9.8e−11
	MS4A4A	0.39	1.1e−39	0.63	1.2e−13
Monocyte	CD86	0.25	1.2e−16	0.21	0.029
Th1	STAT4	0.36	3.7e−34	0.098	0.3
	STAT1	0.065	0.032	0.19	0.043
Tfh	BCL6	0.27	7.1e−20	0.31	9e−04
Neutrophils	CCR	0.32	1.8e−27	−0.22	0.019
B cell	CD19	0.23	5.1e−14	−0.29	0.0019
	CD79A	0.24	1.7e−15	−0.45	6.2e−07
CD8+ T cell	CD8A	0.32	4.9e−27	−0.57	3.9e−11
	CD8B	0.25	4e−17	−0.57	6.1e−11

DISCUSSION

Recently, treatment of BRCA is still an enormous challenge because of its aggressiveness and recurrence. BRCA needs to be managed effectively after diagnosis due to it being not appropriate for screening. Hence, it is essential to explore the pathogenesis and biomarkers of BRCA. Understanding the molecular level dysfunction of BRCA could provide effective treatment and more predictive and diagnostic value. Bioinformatics analysis has been widely used to seek genes that are related with various types of cancer progression. Moreover,

bioinformatics analysis also has been widely applied in the field of BRCA. Candidate genes and pathways in BRCA were determined using synthesized bioinformatics methods.

In this study, to identify the potential molecular mechanism and seek biomarkers of BRCA, we obtained the gene expression patterns from two GEO database (GSE15852 and GSE42568), which were differently expressed in BRCA samples compared with controls. As a result, 52 upregulated and 146 downregulated genes were regarded as DEGs. The functional enrichment analysis indicated that the DEGs were mainly enriched in multiple biological processes including multicellular organismal homeostasis, collagen−containing extracellular matrix, sulfur compound binding. KEGG pathway analysis indicated that the DEGs were significantly associated with pathways in Fatty acid degradation. Furthermore, 11 hub genes among the overlapping DEGs were identified by constructing a PPI network, containing PCK1, FABP4, LPL, CEBPA, CIDEC, ADIPOQ, CFD, LEP PLIN1, CIDEA and PPARG, which were all downregulated. Subsequently, these hub genes were validated and all targets expression tendency in TCGA data set were consistent with mRNA level from the GEO data set. In addition, aiming to find potential prognostic markers of BRCA, the OS and DFS of these targets were analyzed. Interestingly, the downregulation of PPARG mRNA level were significantly associated with poor survival of BRCA. Therefore, PPARG were selected as prognostic markers of BRCA. Additionally, we performed GSEA using TCGA data to further investigate the functions of PPARG. The GSEA results revealed the cytokine receptor interaction and cell surface interactions at the vascular were differentially enriched in the low expression phenotype of PPARG.

PPARG (Peroxisome Proliferator Activated Receptor Gamma) is a Protein Coding gene. This gene encodes a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. The protein encoded by this gene is PPAR-gamma and is a regulator of adipocyte differentiation.[19] However, the relationship between PPARG expression, tumor immunity, and prognosis in Breast invasive carcinoma (BRCA) remain unclear.

This study also demonstrates that PPARG expression correlated with the infiltration status of immune cells in BRCA. There was a strong negative correlation between PPARG expression with infiltration of M1 and M2 Macrophage, Monocyte, Th1, Tfh, Neutrophils, B cell, CD8+ T cell. This suggests that PPARG plays an important role in regulating tumor immunity, and therefore influences BRCA prognosis. We observed correlation between the levels of PPARG mRNA and the expression of the M1 Macrophage marker, IRF5, M2 Macrophage markers, CD163 and VSIG4 and MS4A4A, and Monocyte marker, CD86, and Th1 markers, STAT4 and STAT1, and so on. This suggests that PPARG regulates infiltration and activity of tumor-associated macrophages (TAM). PPARG expression also correlates with the expression of markers of different subsets of T helper (Th) cells, including Th1 (T-bet, STAT-1, IFN-γ and TNF-α), Th2 (STAT6), Tfh (IL-21), and Tregs (CCR8, STAT5B and TGF-β).[20] This suggests a role for PPARG in regulating tumor-infiltration of T-helper cells. Moreover, expression of exhausted T cells markers, CTLA4, LAG3 and GZMB, which are critical inhibitory immune checkpoint proteins negatively correlate with PPARG expression. Most cancers, including BRCA, overexpress inhibitory ligands to evade immune response by dampening T cell function, thus contributing to cancer progression. The expression of inhibitory immune checkpoint proteins is altered in the tumor microenvironment. PPARG can bind to chemokine receptors, such as CCR1, CCR3, and CCR5, and regulate activation and migration of different leukocytes by mobilizing Ca2+ influx. Altered Ca2+ flux in the T

cell subsets promotes cytokine production and downregulates CTLA-4 and PD-1 expression. We postulate that low PPARG expression in the tumor microenvironment diminishes the Ca2+ influx and upregulates the expression of inhibitory immune checkpoint proteins, PD-1, CTLA-4 and TIM-3 on the exhausted T cells. These mechanistic changes can alter anti-tumor function of T cells and result in poorer prognosis of BRCA.[21-23] Besides the regulation of cell proliferation associated with the procession of BRCA, PPARG could also regulate protein polyubiquitination and ubiquitination, which has become increasingly recognized as a controller to regulate the function and signaling of a profusion of proteins.[24] However, this hypothesis needs to be further investigated. Taken together, our findings indicate that PPARG plays an important role in regulating tumor infiltration of immune cells in BRCA.

Our study shows that low PPARG expression is associated with poorer prognosis in BRCA, and infiltration of various types of immune cells, including B cells, DCs, macrophages, neutrophils, CD4+, and CD8+ T cells. PPARG expression also correlates with infiltration of Th, Treg, and exhausted T cells. Hence, our study suggests that PPARG is a potential independent biomarker for BRCA prognosis and the status of tumor immunity.

In conclusion, our study identified 11 hub genes that might be involved in the progression of BRCA with multiple gene expression profile data sets and a series of comprehensive analyses of bioinformatics. Moreover, only PPARG might be regarded as potential prognostic biomarkers. In addition, the target IS primarily involved in fatty acid metabolism, which revealed they might stimulate the occurrence and development of BRCA. However, a series of molecular biology experiments and numerous clinical specimens are required to validate these key genes and pathways in the future, so as to conclusively identify the beneficial biological targets for BRCA.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.J.; formal analysis, X.Z.; methodology, Y.J., Y.M. and Y.D.; writing—original draft, Y.J.; writing—review and editing, X.Z. and S.W. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Mohammad Fahad Ullah. Breast cancer: current perspectives on the disease status. Adv Exp Med Biol . 2019;1152:51-64.
2. Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin . 2018;68(6):394-424.
3. Donna G Albertson, Daniel Pinkel. Genomic microarrays in human genetic disease and cancer. Hum Mol Genet. 2003;12 Spec No 2:R145-52.
4. Xuesong Lu, Xuegong Zhang. The effect of GeneChip gene definitions on the microarray study of cancers. Bioessays. 2006;28(7):739-46.
5. Bert Vogelstein, Nickolas Papadopoulos, Victor E Velculescu, Shibin Zhou, Luis A Diaz Jr, Kenneth W Kinzler. Cancer genome landscapes. Science. 2013;339(6127):1546-58.

6. Davis S. Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23(14):1846-7.
7. The Gene Ontology (GO) project in 2006. Nucleic acids research 2006;34 (Database issue):D322-6.
8. Kanehisa M. Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
9. Glynn Dennis Jr, Brad T Sherman, Douglas A Hosack, Jun Yang, Wei Gao, H Clifford Lane, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4(5):P3.
10. Chen Xie, Xizeng Mao, Jiaju Huang, Yang Ding, Jianmin Wu, Shan Dong, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue):W316-22.
11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11):2498-2504.
12. Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362-D368.
13. Daniel R Rhodes, Shanker Kalyana-Sundaram, Vasudeva Mahavisno, Radhika Varambally, Jianjun Yu, Benjamin B Briggs, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia. 2007;9(2):166-80.
14. Lánczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, et al. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. Breast Cancer Res Treat. 2016;160(3):439-446.
15. Taiwen Li, Jingyu Fan, Binbin Wang, Nicole Traugh, Qianming Chen, Jun S Liu, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017;77(21):e108-e110.
16. Bo Li, Eric Severson, Jean-Christophe Pignon, Haoquan Zhao, Taiwen Li, Jesse Novak, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17(1):174.
17. Patrick Danaher, Sarah Warren, Lucas Dennis, Leonard D'Amico, Andrew White, Mary L Disis, et al. Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer. 2017;5:18.
18. Zefang Tang, Chenwei Li, Boxi Kang, Ge Gao, Cheng Li, Zemin Zhang. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98-W102.
19. Borja Bandera Merchan, Francisco José Tinahones, Manuel Macías-González. Commonalities in the Association between PPARG and Vitamin D Related with Obesity and Carcinogenesis. PPAR Res. 2016;2016:2308249.

20. Barbara Toffoli, Federica Gilardi, Carine Winkler, Magnus Soderberg, Laura Kowalczuk, Yvan Arsenijevic, et al. Nephropathy in Pparg-null mice highlights PPARγ systemic activities in metabolism and in the immune system. PLoS One. 2017;12(2):e0171474.
21. Villa, ALP, Parra RS, Feitosa MR, Camargo HP, Machado VF, Tirapelli D, et al. PPARG expression in colorectal cancer and its association with staging and clinical evolution. Acta Cir Bras. 2020;35(7): e202000708.
22. Liliane Michalik, Béatrice Desvergne, Walter Wahli. Peroxisome-proliferator-activated receptors and cancers: complex stories. Nat Rev Cancer. 2004;4(1):61-70.
23. Huang R, Zhang J, Li M, Yan P, Yin H, Zhai S, et al. The Role of Peroxisome Proliferator-Activated Receptors (PPARs) in Pan-Cancer. PPAR Res. 2020;2020:6527564.
24. William H McBride, Keisuke S Iwamoto, Randi Syljuasen, Milena Pervan, Frank Pajonk. The role of the ubiquitin/proteasome system in cellular responses to radiation. Oncogene. 2003;22(37):5755-73.