SOD2 Expression in Patients with Triple-Negative Breast Cancer is Associated with Tumor-infiltrating Lymphocytes and Prognosis

Yoon Young Kim[a], GunHee Lee[a], Yeonjin Jeon[a], Gyungyub Gong[a], Kyunggon Kim[b], JiSun Kim[c], Byung-Kwan Jeong[a,*], Hee Jin Lee[a,d,*]

*** Corresponding authors:**

Byung-Kwan Jeong, M.D., Ph.D.

Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine,

88 Olympic-ro, 43-gil, Songpa-gu, Seoul 05505, Republic of Korea.

Tel; +82-2-3010-1450; E-mail: goodididid@gmail.com

Hee Jin Lee, M.D., Ph.D.

Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine,

88 Olympic-ro, 43-gil, Songpa-gu, Seoul 05505, Republic of Korea.

Tel; +82-2-3010-5889; E-mail: backlila@gmail.com

Supplementary materials and methods

Immunohistochemistry

Tissue microarrays from 229 surgical samples and 144 pre-NAC biopsy samples, were immunostained using SOD2 mouse monoclonal antibodies (sc-137254, Santacruz; dilution 1:100). Immunostaining of the tumor samples was performed utilizing the autoimmunostainer Benchmark XT (Ventana Medical Systems, Tucson, AZ, USA) with Optiview Dab Detection Kit (Ventana Medical Systems) according to the manufacturer's instructions and using the provided reagents. In brief, 4 μm sections were mounted on silanized slides, dried for 10 min at RT, and incubated for 20 min in an incubator at 65°C. Heat-induced epitope retrieval (HIER) using CC1 was performed for 32 min, followed by incubation for 32 min with anti-SOD2 in the autoimmunostainer. The slides with immunostained samples were digitally scanned using a PANNORAMIC 250 Flash III (3DHISTECH Ltd., Budapest, Hungary) with PANNORAMIC Scanner Software (3DHISTECH Ltd.). SOD2 immunoexpression was evaluated in invasive carcinoma cells showing granular cytoplasmic expression using the semi-quantitative H-score method. The staining intensity was graded as follows: Supplementary Figure 5; 0, no; 1 + , weak; 2 + , moderate; and 3 + , strong. The percentage of cells exhibiting expression was multiplied by each intensity category, and the total scores were determined by adding together the values of each category (H-score = $3\times P_3 + 2\times P_2 + 1\times P_1$; P_3, P_2 and P_1 represent the percentages of strongly, moderately, and weakly stained cells, respectively; H-score range from 0 to 300). Immunoreactive scores of the other antigens with immunologic features such as MHC1, MHC2, MX1, and PD-L1 were obtained as previously described [1]. The count of cells positive for CD4 and CD8 in tissue microarray cores was assessed using the NuclearQuant module of the Pannoramic Viewer 1.15.2 (3DHISTECH Ltd) [1].

Proteomics analysis in 336 biopsy samples of patients with TNBC before neoadjuvant chemotherapy (NAC)

For protein extraction, 403 FFPE samples underwent deparaffinization using 100% heptane. The specimens were then incubated in an extraction buffer at 100℃ for 20 min followed by 80℃ for 4 h. Quantification of the extracted protein was performed using the Bradford approach. This process was conducted utilizing the QProteome FFPE Tissue Kit (Qiagen, Hilden, Germany). Subsequently, proteins were digested into peptides using S-Trap^{TM} (ProtiFi, Long Island, New York, USA). For Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS), a TripleTOF® 5600+ system (AB SCIEX, Framingham, Massachusetts, USA) coupled with a 450 LC system was employed, with a 55 min run time. Post-data analysis utilized the Pan-Human Library and data-independent acquisition-neural networks (DIA-NN) for proteomic qualitative and quantitative analysis.

We quantified a total of 5,547 proteins, with notable variation observed between samples. The least identified case yielded quantification for 111 proteins, while the most identified case encompassed 2,596 proteins. On average, each sample yielded quantification for 1,249 proteins. For stable statistical analysis, quantile analysis was performed and samples with insufficient protein variation were excluded. Among the initial 403 samples, 67 were excluded due to failing quality control (QC) or lacking detection of SOD2 in the protein quantitative results. Finally, a total of 336 samples were included in the analysis.

To confirm that genes correlated with SOD2 in TNBCs, a protein interaction network was constructed for 194 proteins exhibiting correlation at both RNA (from TCGA dataset) and protein levels using Cytoscape (ver 3.9.0) with stringApp (ver 2.0.3), integrating proteins with high-confidence interactions from STRING (edge score, >0.7) [2-4]. The generated network was clustered using the Markov clustering algorithm with a granularity parameter of 2.

Functional enrichment analysis of the largest cluster was conducted using Reactome Pathways, GO Biological Process, and KEGG Pathways in stringApp.

Single-cell RNA-seq data analysis

To analyze the single-cell transcriptome data, the Seurat package (version 4.1.1) was used [5]. The NormalizeData function was employed to normalize the data, followed by scaling with 2000 features selected using the FindVariableFeatures function. Principal components analysis (PCA) was then executed on the scaled data using the RunPCA function (dims = 1:20). To address batch effects, Harmony (version 0.1.0) [6] was employed, assigning each sample as a batch for correction. Harmony embedding was utilized in downstream analysis (reduction = "harmony"). Cell clustering was performed using the FindNeighbors and FindClusters functions. Uniform manifold approximation and projection (UMAP) was generated with the RunUMAP function. Subsequently, 981 doublets were identified and removed using scDblFinder (version 1.17.0) [7]. The remaining 109,168 cells were rescaled, batch effect corrected, and clustered for further analysis. Marker genes for each cluster were identified using the FindAllMarkers function, and differential expression analyses were performed using the FindMarkers function with default parameters. Each cluster was annotated by canonical markers, then the TIL groups were divided based on the cell types annotated. Given the relatively high proportions of TILs in TNBC, the TIL-high group was defined as cases where the proportion of lymphocytes (T cells, B cells, and plasma cells) relative to total cells exceeded the first percentile (12.8%), with the remainder categorized as TIL-low.

Cell–cell communication within the TME was investigated using the R package CellChat (version 1.6.0). Employing default parameters, we inferred 1,199 paracrine or autocrine signaling interactions from CellchatDB.human [8]. Initially, TNBC samples were categorized based on TIL groups, and cell–cell communication analyses were conducted for each group.

Subsequently, the results were consolidated and compared according to the TIL groups. The identification and visualization of differential signaling between the two groups were executed using the compareInteractions, netVisual_diffInteractions, and netVisual_heatmap functions. Visualizations depicting the disparity in signaling among fibroblasts and signaling from fibroblasts to immune cells were achieved using the netVisual_aggregate and netVisual_bubble functions. Upregulated and down-regulated signaling ligand–receptor pairs in the aforementioned interactions were identified by the identifyOverExpressedGenes function, based on the results of the differential gene expression analysis. Visualization of the distinct ligand–receptor pairs was performed using the netVisual_chord_gene function.

Functional enrichment analysis

Functional enrichment analyses of significantly correlated genes of proteomics and TCGA RNA-seq data, and marker genes of single-cell RNA-seq data, were conducted using the MsigDB hallmark gene database with EnrichR [9, 10].

Supplementary Tables

Supplementary Table 1. Clinicopathologic features of 336 patients with TNBC received neoadjuvant chemotherapy.

Variables	Proteomics (total n=336)	IHC (total n=144)
Age (yrs)		
Mean ± SD	45.63 ± 10.7	46.9 ± 10.1
Range	23–79	23–71
Sex (n, %)		
Male	0	0
Female	336 (100%)	144 (100%)
Histologic type (n, %)		
IBC-NST	325 (96.7)	141 (97.9)
Metaplastic carcinoma	11 (3.3)	8 (2.1)
Tumor size (cm)		
Mean ± SD	4.8 ± 2.3	5.0 ± 2.3
Range	0.6–15.1	1.2–15.1

TNBC Triple negative breast cancer, *IBC-NST* Invasive breast carcinoma of no special type

Supplementary Table 2. Clinicopathologic features of 229 patients with surgically resected TNBC

Variables	TNBC samples (total n=229)
Age (yrs)	
Mean ± SD	47.6 ± 9.7
Range	25–72
Sex (n, %)	
Male	0
Female	229 (100%)
Histologic type (n, %)	
IBC-NST	191 (83.4)
Metaplastic carcinoma	26 (11.4)
Carcinoma with apocrine differentiation	9 (3.9)
Invasive micropapillary carcinoma	3 (1.3)
Tumor size (cm)	
Mean ± SD	2.4 ± 1.16
Range	0.5–9

IBC-NST Invasive breast carcinoma of no special type

Supplementary Table 3. Correlation between clinicopathologic variables and SOD2 expression in invasive tumor of biopsied samples from 144 patients with TNBC before receiving neoadjuvant chemotherapy

Variables	SOD2 expression		
	Low (n=97)	**High (n=47)**	***P***
Age (yrs) (n, %)			
<50	52 (53.6)	26 (55.3)	0.861
≧ 50	45 (46.4)	21 (44.7)	
Histologic grade (n, %)			
1–2	39 (40.2)	16 (34.0)	0.584
3	58 (59.8)	31 (66.0)	
Nuclear grade (n, %)			
1–2	40 (41.2)	17 (36.2)	0.590
3	57 (58.8)	30 (63.8)	
cT stage (n, %)			
1	7 (7.2)	4 (8.5)	0.600
2	64 (66.0)	29 (61.7)	
3	18 (18.6)	7 (14.9)	
4	8 (8.2)	7 (14.9%)	

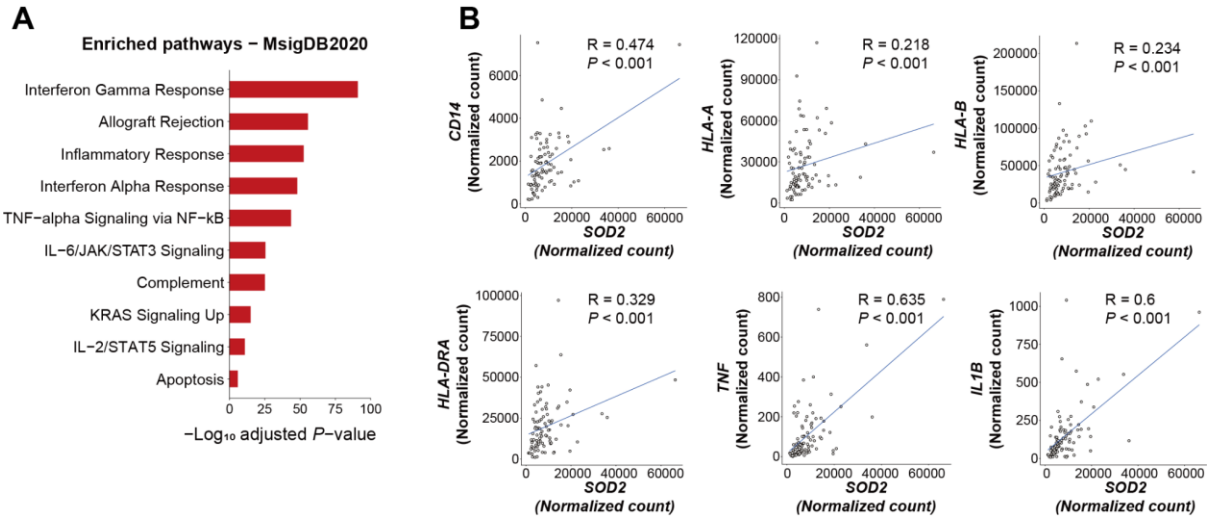
cN stage (n, %)			
0	26 (26.8)	14 (29.8)	0.652
1	38 (39.2)	22 (46.8)	
2	8 (8.2)	2 (4.3)	
3	25 (25.8)	9 (19.1)	
ypT stage (n, %)			
0	27 (27.8)	17 (36.2)	0.533
1	34 (35.1)	15 (31.9)	
2	29 (29.9)	10 (21.3)	
3	7 (7.2)	5 (10.6)	
ypN stage (n, %)			0.613
0	58 (59.8)	31 (66.0)	
1	23 (23.7)	12 (25.5)	
2	7 (7.2)	2 (4.3)	
3	9 (9.3)	2 (4.3)	

Supplementary Table 4. Correlation between clinicopathologic variables and SOD2 expression in invasive tumor of 229 patients with surgically resected TNBC

Variables	SOD2 expression		
	Low (n=139)	**High (n=90)**	***P***
Age (yrs) (n, %)			
<50	78 (56.1)	63 (70.0)	0.038
≧ 50	61 (43.9)	27 (30.0)	
Histologic grade (n, %)			
1–2	37 (26.6)	17 (18.9)	0.204
3	102 (73.4)	73 (81.1)	
Nuclear grade (n, %)			
1–2	39 (28.1)	17 (18.9)	0.120
3	100 (71.9)	73 (81.1)	
pT stage (n, %)			
1	61 (43.9)	40 (44.4)	0.550
2	70 (50.4)	49 (54.4)	
3	8 (5.8)	1 (1.1)	
pN stage (n, %)			
0	95 (68.3)	64 (71.1)	0.161

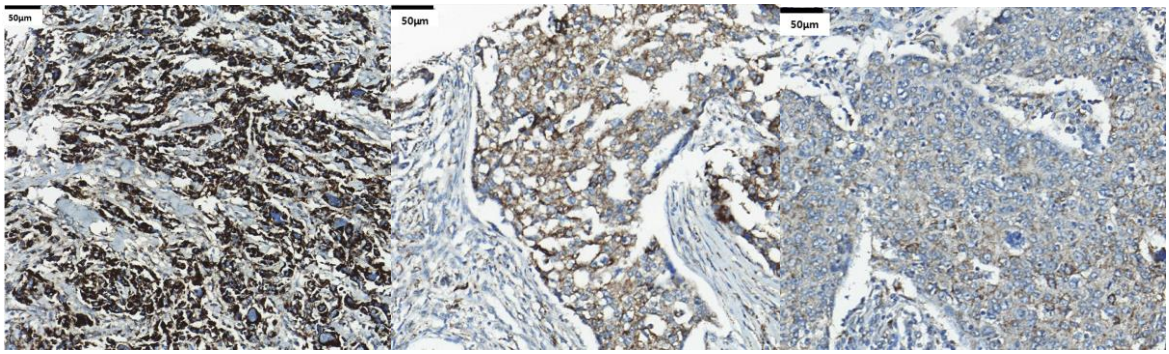
1	22 (15.8)	17 (18.9)	
2	6 (4.3)	7 (7.8)	
3	16 (11.5)	2 (2.2)	
Lymphovascular invasion			
(n, %)			
Not identified	102 (73.4)	75 (83.3)	0.106
Present	37 (26.6)	15 (16.7)	

Supplementary Figures

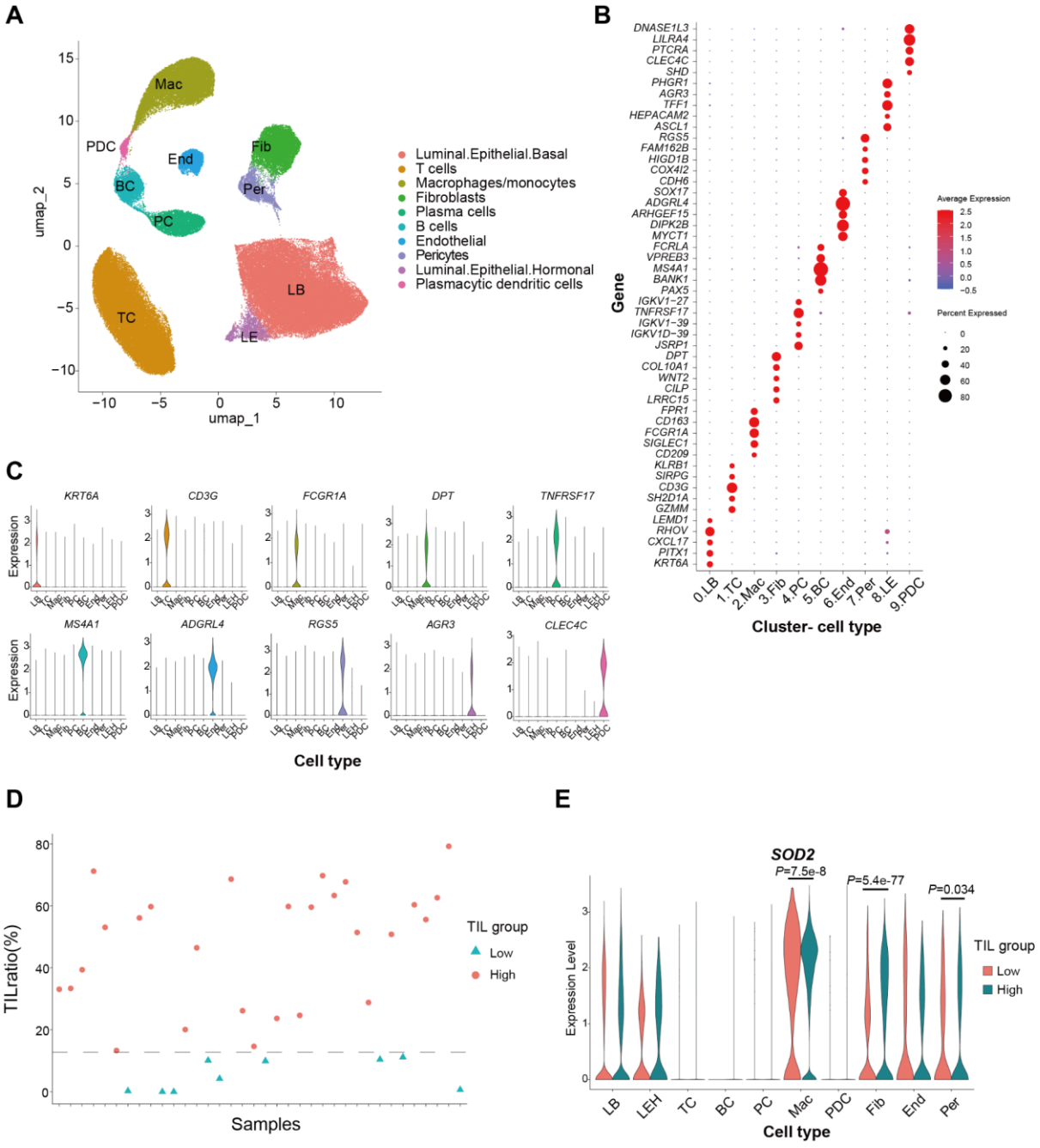


Supplementary Fig. 1 Genes positively correlated with SOD2 in TCGA RNA-seq data are associated with immune responses.

A Gene set enrichment analysis results for positively correlated genes with SOD2 expression in TCGA RNA-seq data using MsigDB Hallmark gene sets **B** Scatter correlation plots of SOD2 RNA level with the levels of genes related to immune response (*CD14*, *HLA-A*, *HLA-B*, *HLA-DRA*, *TNF*, and *IL1B*)



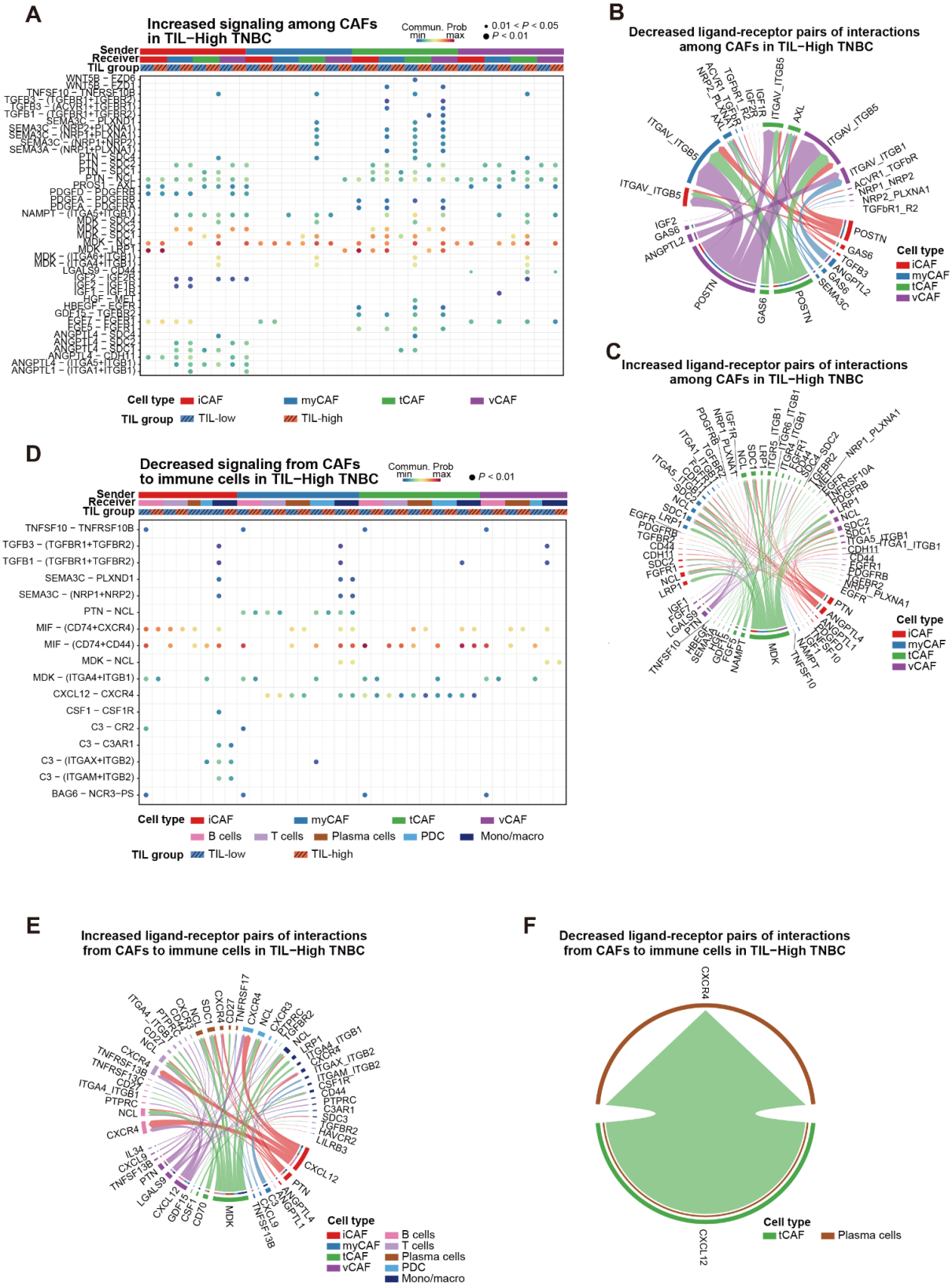
Supplementary Fig. 2 Representative immunohistochemical staining results of SOD2 in the invasive carcinoma component of triple-negative breast cancer. Strong, moderate, and weak expression from left to right. 400× magnification.



Supplementary Fig. 3 Single-cell RNA-seq revealed SOD2 expression in cancer-associated fibroblasts is associated with a higher ratio of lymphocytes in triple-negative breast cancer (TNBC)

A UMAP plot shows unsupervised clustered cells from 36 TNBC samples. **B** Dot plot showing expression of top 5 marker genes in each unsupervised cluster **C** Violin plot showing expression of representative marker for each cell type **D** Scatter plot showing the ratio of

lymphocytes (including T cells, B cells, and plasma cells) in each sample, grouped by whether it is higher than the first quartile (12.79 %, TIL-high) or not (TIL-low) **E** Violin plot showing expression of SOD2 in each cell type depending on TIL group, with P-values calculated using Wilcoxon Rank Sum test


A
Increased signaling among CAFs in TIL−High TNBC
Commun. Prob
min
max
• 0.01 < P < 0.05
● P < 0.01
Sender
Receiver
TIL group
WNT5B − FZD6
WNT5B − FZD1
TNFSF10 − TNFRSF10B
TGFB3 − (ACVR1+TGFBR1)
TGFB3 − (TGFBR1+TGFBR2)
TGFB1 − (TGFBR1+TGFBR2)
SEMA3C − PLXND1
SEMA3C − (NRP2+PLXNA1)
SEMA3C − (NRP1+PLXNA1)
SEMA3C − (NRP1+NRP2)
SEMA3A − (NRP1+PLXNA1)
PTN − SDC4
PTN − SDC3
PTN − SDC1
PTN − NCL
PROS1 − AXL
PDGFD − PDGFRB
PDGFA − PDGFRB
PDGFA − PDGFRA
NAMPT − (ITGA5+ITGB1)
MDK − SDC4
MDK − SDC2
MDK − SDC1
MDK − NCL
MDK − LRP1
MDK − (ITGA6+ITGB1)
MDK − (ITGA4+ITGB1)
LGALS9 − CD44
IGF2 − IGF2R
IGF1 − IGF1R
IGF1 − IGF1R
HGF − MET
HBEGF − EGFR
GDF15 − TGFBR2
FGF7 − FGFR1
FGF5 − FGFR1
ANGPTL4 − SDC4
ANGPTL4 − SDC3
ANGPTL4 − SDC2
ANGPTL4 − CDH11
ANGPTL4 − (ITGA5+ITGB1)
ANGPTL1 − (ITGA1+ITGB1)
Cell type iCAF myCAF tCAF vCAF
TIL group TIL-low TIL-high
B
Decreased ligand-receptor pairs of interactions among CAFs in TIL−High TNBC
ITGAV_ITGB5
AXL
ITGAV_ITGB1
ACVR1_TGFBR
NRP1_NRP2
NRP2_PLXNA1
TGFbR1_R2
POSTN
GAS6
TGFB3
ANGPTL2
SEMA3C
IGF2
Cell type
iCAF
myCAF
tCAF
vCAF
C
Increased ligand-receptor pairs of interactions among CAFs in TIL−High TNBC
PTN
MDK
ANGPTL4
TNFSF10
Cell type
iCAF
myCAF
tCAF
vCAF
D
Decreased signaling from CAFs to immune cells in TIL−High TNBC
Commun. Prob
min
max
● P < 0.01
Sender
Receiver
TIL group
TNFSF10 − TNFRSF10B
TGFB3 − (TGFBR1+TGFBR2)
TGFB1 − (TGFBR1+TGFBR2)
SEMA3C − PLXND1
SEMA3C − (NRP1+NRP2)
PTN − NCL
MIF − (CD74+CXCR4)
MIF − (CD74+CD44)
MDK − NCL
MDK − (ITGA4+ITGB1)
CXCL12 − CXCR4
CSF1 − CSF1R
C3 − CR2
C3 − C3AR1
C3 − (ITGAX+ITGB2)
C3 − (ITGAM+ITGB2)
BAG6 − NCR3−PS
Cell type iCAF myCAF tCAF vCAF B cells T cells Plasma cells PDC Mono/macro
TIL group TIL-low TIL-high
E
Increased ligand-receptor pairs of interactions from CAFs to immune cells in TIL−High TNBC
CXCL12
PTN
MDK
Cell type
iCAF
myCAF
tCAF
vCAF
B cells
T cells
Plasma cells
PDC
Mono/macro
F
Decreased ligand-receptor pairs of interactions from CAFs to immune cells in TIL−High TNBC
CXCR4
CXCL12
Cell type
tCAF
Plasma cells

Supplementary Fig. 4 Comparative intercellular communication analysis among cancer-associated fibroblasts (CAFs) and from CAFs to immune cells using communication probability and gene expression of ligand-receptor pairs

A Dot plot illustrating increased interactions among CAFs in TIL-high TNBC compared with TIL-low. Dot color reflects communication probability. **B, C** Chord diagrams displaying ligand-receptor pairs with **B** decreased or **C** increased expression among CAFs in TIL-high TNBCs compared with TIL-low **D** Dot plot showing decreased interactions from CAFs to immune cells in TIL-high TNBC compared with TIL-low. Dot color reflects communication probability. **E, F** Chord diagrams displaying ligand-receptor pairs with **E** increased or **F** decreased expression from CAFs to immune cells in TIL-high TNBCs compared to TIL-low

References in supplementary files

[1] Park IA, Hwang SH, Song IH, Heo SH, Kim YA, Bang WS, et al. Expression of the MHC class II in triple-negative breast cancer is associated with tumor-infiltrating lymphocytes and interferon signaling. PLoS One. 2017;12(8):e0182786.

[2] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-46.

[3] 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):2498-504.

[4] Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019;18(2):623-32.

[5] Wang YS, Chiu WT, Chang FP, Chen YL. Decline of chlorinated hydrocarbon insecticides residues in the tea-garden soils of Taiwan. Proc Natl Sci Counc Repub China B. 1988;12(1):9-13.

[6] Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96.

[7] Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. F1000Res. 2021;10:979.

[8] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.

[9] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.

[10] Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021;1(3):e90.