Clinical and Experimental Medicine (2025) 25:167
https://doi.org/10.1007/s10238-025-01690-3

RESEARCH

Machine learning and single-cell analysis uncover distinctive characteristics of CD300LG within the TNBC immune microenvironment: experimental validation

Baoxi Zhu[1,2,3] · Hong Wan[1,2] · Zichen Ling[1,2] · Han Jiang[1,2] · Jing Pei[1,2]

Received: 23 March 2025 / Accepted: 14 April 2025

Abstract

Investigating the essential function of CD300LG within the tumor microenvironment in triple-negative breast cancer (TNBC). Transcriptomic and single-cell data from TNBC were systematically collected and integrated. Four machine learning algorithms were employed to identify distinct target genes in TNBC patients. Specifically, CIBERSORT and ssGSEA algorithms were utilized to elucidate immune infiltration patterns, whereas TIDE and TCGA algorithms predicted immune-related outcomes. Moreover, single-cell sequencing data were analyzed to investigate the function of CD300LG-positive cells within the tumor microenvironment. Finally, immunofluorescence staining confirmed the significance of CD300LG in tumor phenotyping. After machine learning screening and independent dataset validation, CD300LG was identified as a unique prognostic biomarker for triple-negative breast cancer. Enrichment analysis revealed that CD300LG expression is strongly linked to immune infiltration and inflammation-related pathways, especially those associated with the cell cycle. The presence of $CD8^+$ T cells and M1-type macrophages was elevated in the CD300LG higher group, whereas the abundance of M2-type macrophage infiltration showed a significant decrease. Immunotherapy prediction models indicated that individuals with low CD300LG expression exhibited better responses to PD-1 therapy. Additionally, single-cell RNA sequencing and immunofluorescence analyses uncovered a robust association between CD300LG and genes involved in tumor invasion. CD300LG plays a pivotal role in the tumor microenvironment of TNBC and represents a promising therapeutic target.

Keywords TNBC · Machine learning · CD300LG · Tumor microenvironment · Single cell

Background

Triple-negative breast cancer (TNBC) represents a highly diverse subset of breast cancer, distinguished by the lack of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression [1, 2]. TNBC accounts for 10–20% of all breast cancer cases and is characterized by its aggressive behavior, high likelihood of recurrence, and unfavorable prognosis [3, 4]. Due to the lack of specific therapeutic targets, treatment primarily relies on chemotherapy, which has limited efficacy and is prone to drug resistance, leading to lower survival rates [5, 6]. Consequently, an in-depth analysis of the molecular mechanisms underlying TNBC and the identification of effective therapeutic targets have emerged as key priorities and challenges in current research [7, 8].

With the rapid advancement of bioinformatics and artificial intelligence technologies, machine learning algorithms have demonstrated significant potential in disease marker identification and target prediction [9–11]. Numerous studies have employed four classical machine learning algorithms—Random Forest, Support Vector Machine, Gradient Boosting Machine, and XGBoost—to systematically identify key genes and potential therapeutic targets by integrating multi-omics data [12–14]. The combined evaluation of these algorithms not only improves the precision of target recognition but also successfully reduces

Baoxi Zhu, Hong Wan and Zichen Ling have contributed equally to this work.

✉ Jing Pei
peijing@ahmu.edu.cn

[1] Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

[2] Department of Breast Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

[3] Department of Thyroid and Breast Surgery, Anhui No.2 Provincial People's Hospital, Hefei, Anhui, China

Published online: 17 May 2025

the inherent bias present in single algorithms, thus establishing a solid theoretical basis for further experimental verification [15–17].

CD300LG (Cluster of Differentiation 300 Ligand G) is a member of the immunoglobulin superfamily and belongs to the CD300 family of molecules, which plays a critical role in immune regulation and cell–cell interactions [18]. This protein is broadly expressed on the surface of endothelial cells, dendritic cells, and specific epithelial cells, and regulates inflammatory responses, immune cell activation, and vascular barrier function through interactions with its receptors [19]. Research has demonstrated that CD300LG is essential for maintaining vascular endothelial integrity, suppressing excessive inflammatory reactions, and modulating leukocyte migration, processes that are closely associated with the pathogenesis of sepsis, atherosclerosis, and other diseases [20]. Moreover, CD300LG has been implicated in tumor microenvironment modulation and tumor immune escape mechanisms [21, 22]. Its distinctive structural and functional characteristics render it a promising therapeutic target, particularly in the context of inflammatory diseases and cancer immunotherapy [23].

To further elucidate the functions and mechanisms of action of the screened genes in disease, it is crucial to resolve the heterogeneity of the tumor microenvironment at single-cell resolution using single-cell RNA sequencing technology [24, 25]. Single-cell technology enables precise identification of gene expression profiles across distinct cell subpopulations, uncovers interactions between tumor cells, immune cells, and stromal cells, and offers novel insights into disease progression and drug resistance mechanisms [26, 27].

Through the synergistic integration of machine learning algorithms and single-cell technologies, this study aims to identify potential therapeutic targets for TNBC and thoroughly investigate their functions in tumorigenesis, progression, and metastasis. The findings are anticipated to provide novel targets and strategies for precision treatment of TNBC, advance the development of personalized medicine, and enhance the clinical prognosis of patients.

Method

Data collation

Clinical and expression profiling data for TNBC patients were obtained from the official TCGA database (https://portal.gdc.cancer.gov/). The counts data were screened for differentially expressed genes using DESeq2, with a threshold of $|\log2FC| > 1$ and FDR < 0.05. FPKM data were initially assessed for expression differences between samples to determine the necessity of normalization. The processed FPKM data were subsequently utilized for functional enrichment analysis, immune infiltration analysis, and survival analysis. Samples with missing

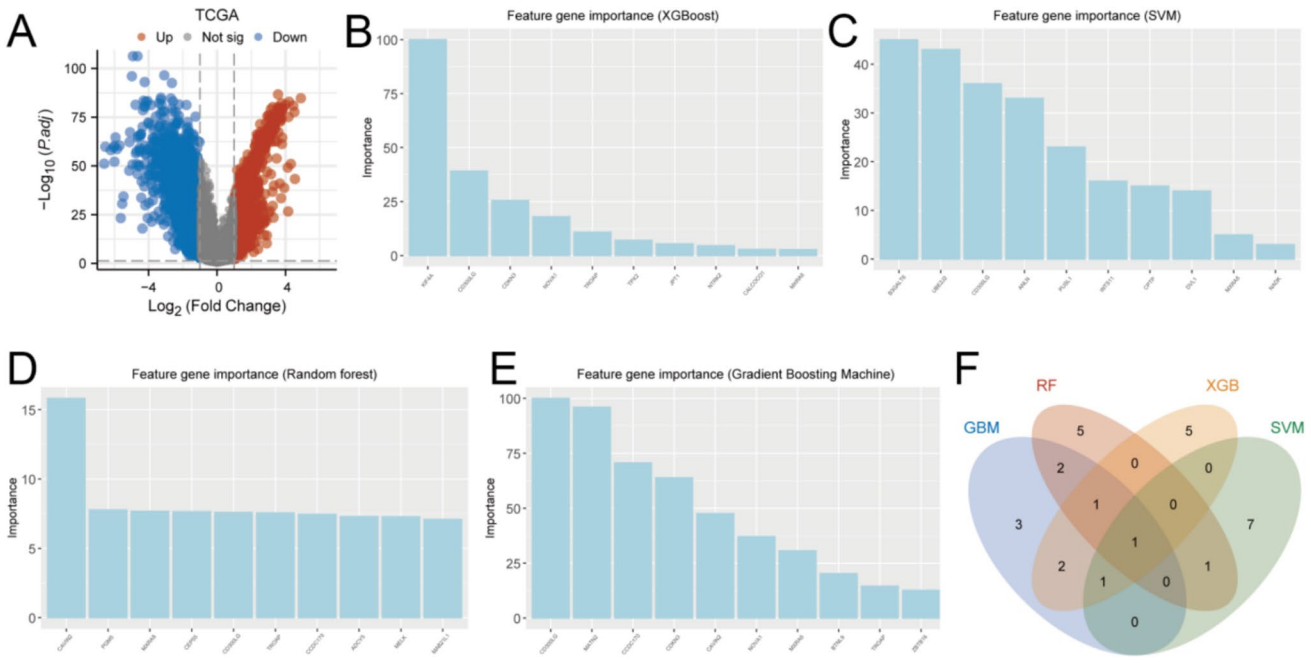


Fig. 1 Machine learning screening for unique prognostic targets in TNBC patients. **A** Volcano plot of the results of gene expression differential analysis between TNBC patients and normal tissues. **B**, **C**, **D**, **E** Unique prognostic genes in TNBC patients were explored using four of the machine learning methods. **F** Gene intersection-CD300LG screened by four machine learning methods

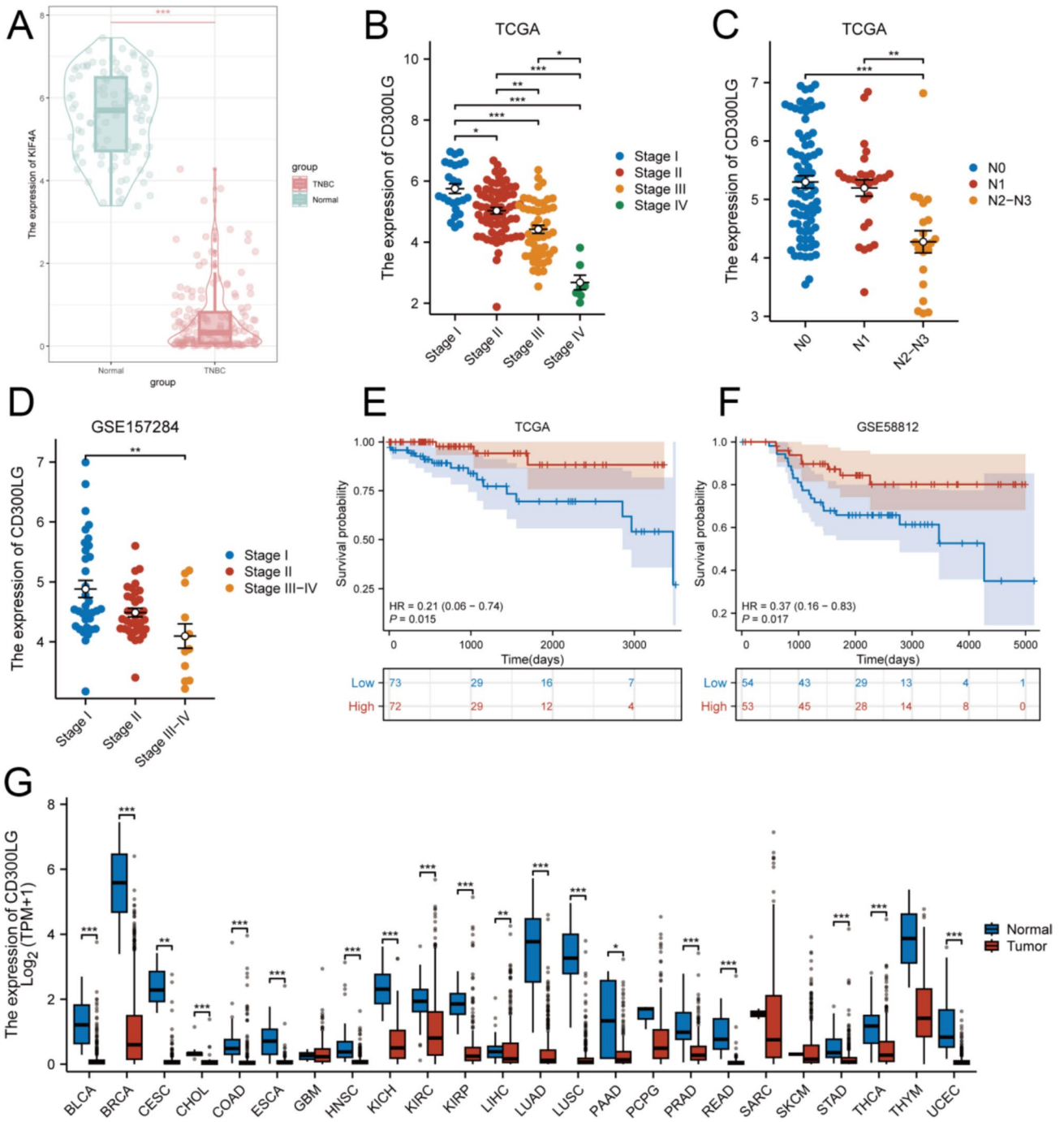
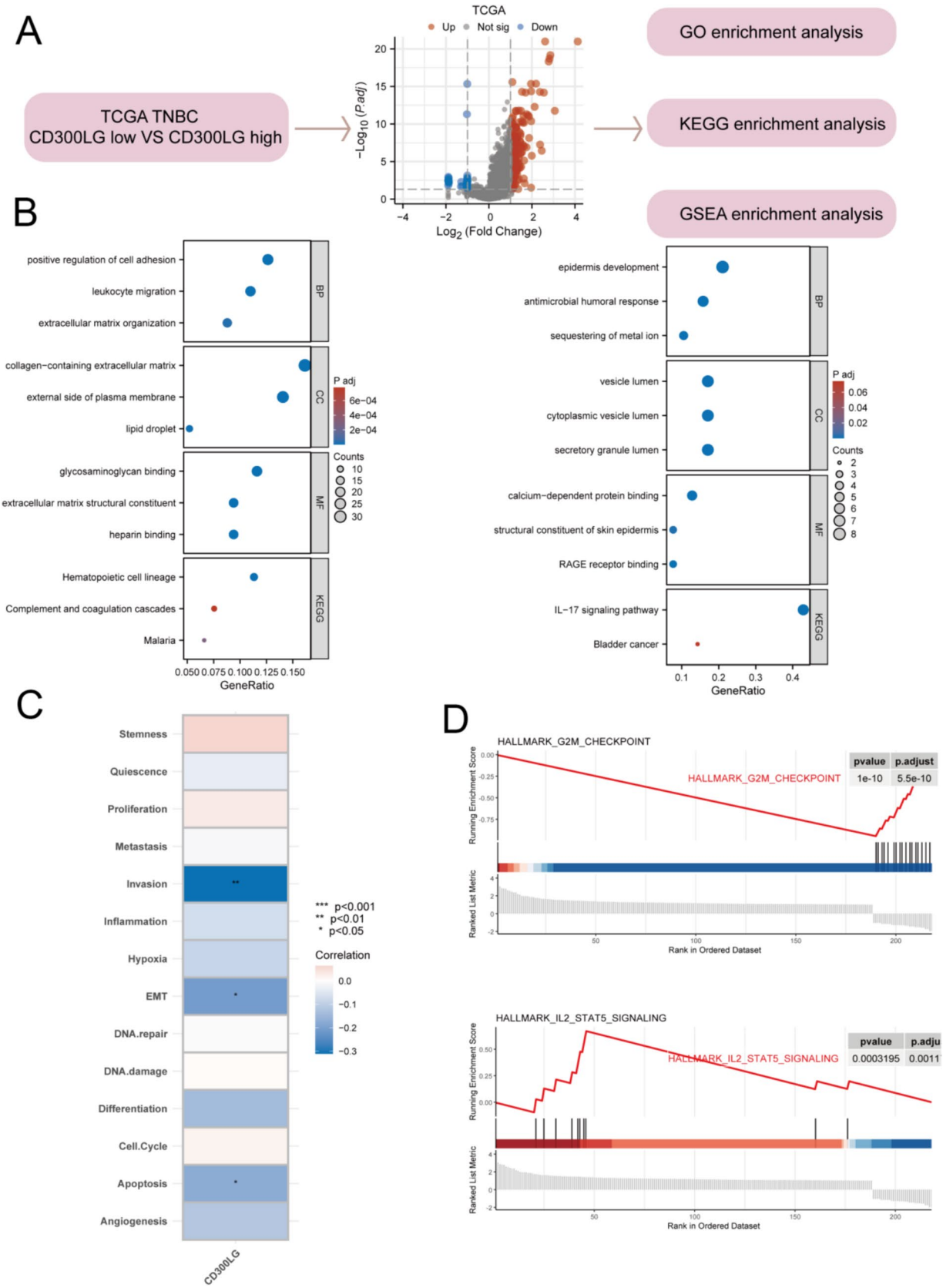


Fig. 2 Correlation between CD300LG expression and clinical features in TNBC. **A** CD300LG expression is significantly elevated in TNBC patients. **B**, **C**, **D** CD300LG expression progressively increases with advancing tumor stage. **E**, **F** Higher CD300LG expression correlates with poorer patient prognosis. **G** Pan-cancer analysis of CD300LG expression

clinical features were excluded, and after screening, a total of 145 samples were included in the final analysis. The baseline characteristics of these samples are provided in Supplementary Information 1. Additionally, we obtained transcriptome and single-cell sequencing data of TNBC from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Specifically, GSE58812 transcriptome data served as an independent validation set, while GSE176078 single-cell dataset was utilized for single-cell profiling of TNBC. Baseline information for incorporation is provided in the supplementary material.


A
TCGA TNBC
CD300LG low VS CD300LG high
TCGA
Up
Not sig
Down
−Log10 (P.adj)
Log2 (Fold Change)
GO enrichment analysis
KEGG enrichment analysis
GSEA enrichment analysis
B
positive regulation of cell adhesion
leukocyte migration
extracellular matrix organization
collagen−containing extracellular matrix
external side of plasma membrane
lipid droplet
glycosaminoglycan binding
extracellular matrix structural constituent
heparin binding
Hematopoietic cell lineage
Complement and coagulation cascades
Malaria
BP
CC
MF
KEGG
P adj
Counts
GeneRatio
epidermis development
antimicrobial humoral response
sequestering of metal ion
vesicle lumen
cytoplasmic vesicle lumen
secretory granule lumen
calcium−dependent protein binding
structural constituent of skin epidermis
RAGE receptor binding
IL−17 signaling pathway
Bladder cancer
C
Stemness
Quiescence
Proliferation
Metastasis
Invasion
Inflammation
Hypoxia
EMT
DNA.repair
DNA.damage
Differentiation
Cell.Cycle
Apoptosis
Angiogenesis
CD300LG
*** p<0.001
** p<0.01
* p<0.05
Correlation
D
HALLMARK_G2M_CHECKPOINT
pvalue
p.adjust
1e-10
5.5e-10
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
HALLMARK_IL2_STAT5_SIGNALING
pvalue
p.adju
0.0003195
0.0011

◂**Fig. 3** Functional Enrichment Analysis of CD300LG in TNBC. **A** Functional enrichment analysis associated with CD300LG based on high versus low-expression groups. **B** GO and KEGG enrichment analyses were conducted for both high and low-expression groups. **C** Correlation analysis between CD300LG and tumor phenotypes. **D** GSEA enrichment analysis indicated that CD300LG predominantly influences inflammatory response pathways and cell cycle pathways

Machine learning

Machine learning has become a crucial tool in the field of life sciences. In this research, we utilized four different machine learning approaches: Gradient Boosting Machine (GBM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM). These were applied to identify potential targets that may influence triple-negative breast cancer patients. GBM iteratively trains weak models, such as decision trees, optimizing the residuals at each step to gradually reduce error and ultimately combine them into a robust model. RF integrates multiple decision trees, each trained on random samples and features, thereby reducing overfitting risk through voting or averaging predicted outcomes. XGBoost, an efficient implementation of gradient boosting, incorporates regularization, parallel computation, and missing-value processing to enhance speed and performance, making it popular in life sciences. SVM identifies the optimal hyperplane to maximize classification margins, handling nonlinear data via kernel functions, and is particularly suitable for high-dimensional spaces and small sample sizes. Details of machine learning are provided in the supplementary material.

Function enrichment

Up- and down-regulated genes were initially analyzed for pathway and functional enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology. To explore differential gene enrichment phenotypes within the Hallmark gene set, Gene Set Enrichment Analysis (GSEA) was utilized. Additionally, Gene Set Variation Analysis (GSVA) was performed to clarify the association between potential therapeutic targets and 13 tumor-related phenotypes Supplementary Table 1.

Prediction of immune cell infiltration

In the area of oncology, various algorithms are commonly used to analyze immune infiltration. To assess the differences in immune cell infiltration analysis, we applied the CIBERSORT method along with single-sample Gene Set Enrichment Analysis (ssGSEA).

Immunotherapy predictions

TIDE (http://tide.dfci.harvard.edu) and TCIA (https://www.cancerimagingarchive.net/) were utilized to evaluate the efficacy of immunotherapy. Higher TIDE scores suggest that tumors may exhibit greater immune evasion capabilities, potentially rendering them less responsive to immune checkpoint inhibition treatments. TCIA was employed to assess the therapeutic potential of targeting PD-1 and CTLA-4 in the new model. Additionally, the analysis focused on evaluating the role of immune checkpoint inhibitory genes in therapy. Finally, the functional changes in immune cells during tumor progression were examined.

Single-cell sequencing data processing

Single-cell datasets were processed by loading the raw data and constructing Seurat objects for quality control at the single-cell level. Cells with insufficient RNA content or disproportionately high mitochondrial gene expression were excluded from subsequent analyses. Following this, normalization of the data and correction for batch effects were performed using Seurat. Highly variable genes were identified, and dimensionality reduction was achieved through principal component analysis (PCA). The resulting embeddings were visualized using UMAP. Cell types were annotated based on differential gene expression analysis and known marker genes. Cell trajectory analysis aims to reveal the temporal progression of cell state changes. To accomplish this, the Monocle3 package applies a time-series approach that includes preprocessing steps such as batch effect removal and data normalization. Topological relationships among cells can be inferred by reducing dimensionality using highly variable genes or gene sets associated with developmental processes via UMAP. Trajectories are constructed in Monocle3 by leveraging nearest-neighbor graphs within UMAP space. Determining the root node of the trajectory is an essential step, which can be achieved either by identifying early-stage cell markers or by employing algorithms based on transcriptome entropy. After pseudotime values are computed, branch-specific genes are detected using a branching expression analysis model. The expression patterns of key genes along the trajectory can be displayed using heatmaps or violin plots [28–30].

Clinical validation

We collected clinical samples from patients with TNBC to validate the expression of target genes and perform

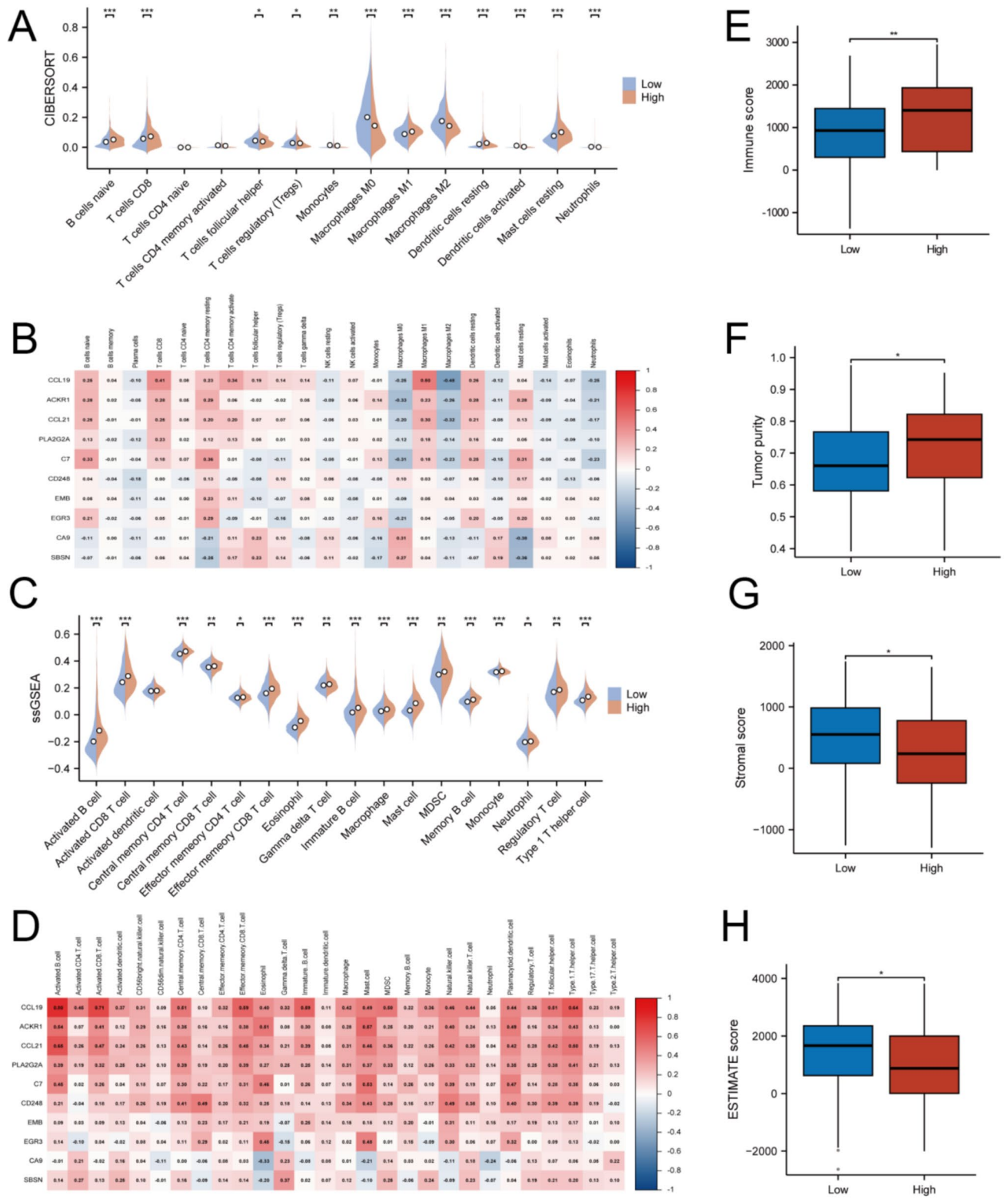


Fig. 4 Analysis of tumor microenvironment infiltration in TNBC patients. **A**, **B** The CIBERSORT algorithm identified a higher proportion of CD8 + T cells and M2-type macrophages in the CD300LG high-expression group compared to the low-expression group. **C**, **D** ssGSEA enrichment analysis indicated an increased abundance of suppressor cell infiltration in the CD300LG high-expression group. **E**, **F**, **G**, **H** The ESTIMATE algorithm revealed significant differences in immune-related scores between the CD300LG high- and low-expression groups

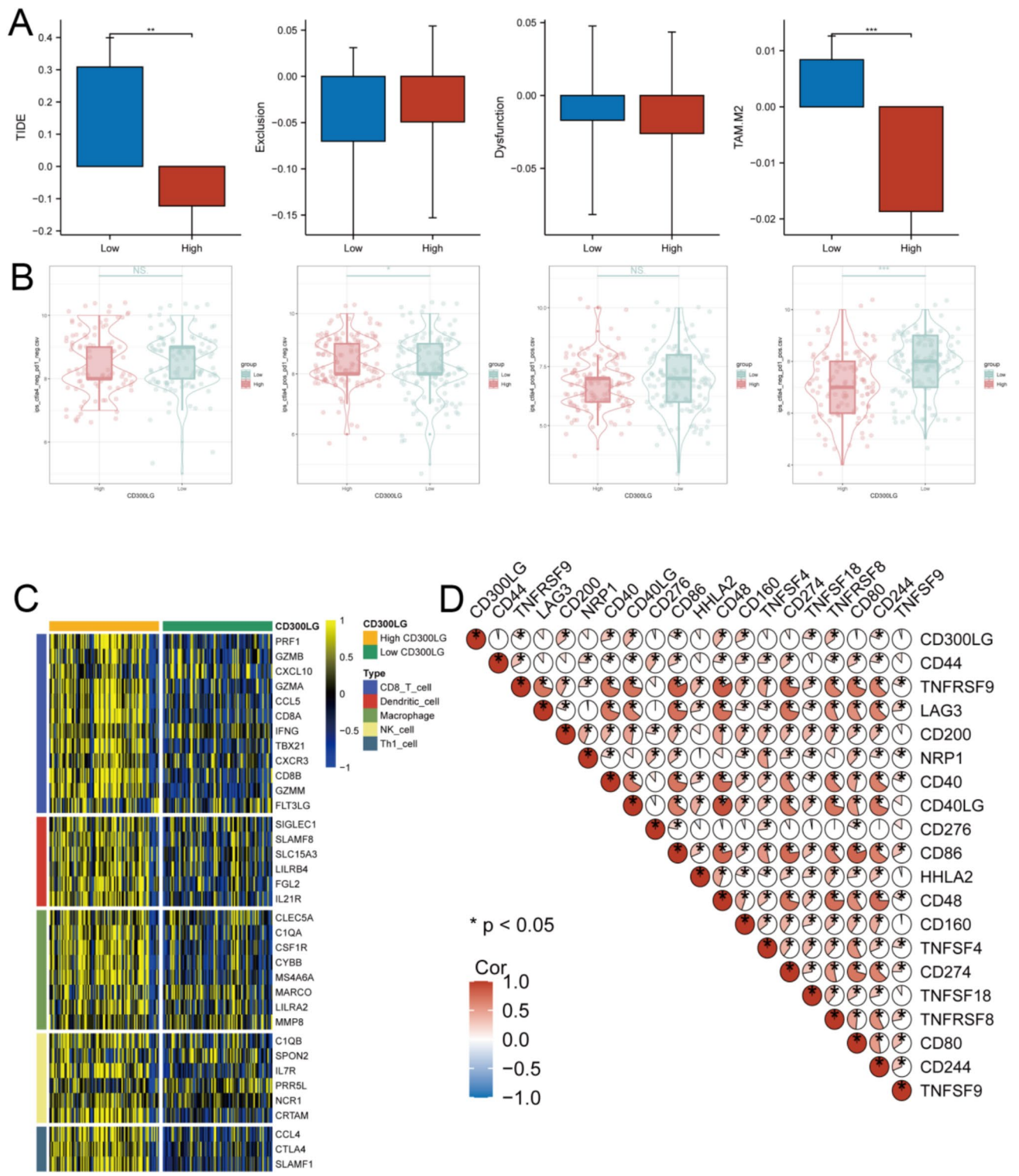


Fig. 5 Immunotherapy Prediction Analysis. **A** Application of the TIDE tool to predict the therapeutic potential of CD300LG in TNBC patients. **B** Use of the Tumor Immune Estimation Resource (TIMER) tool to evaluate the therapeutic value of PD-1 and CTLA4 in relation to CD300LG expression in TNBC patients. **C** Differences in immune cell function between high and low CD300LG expression groups. **D** Correlation between CD300LG expression and immune checkpoint inhibitor genes

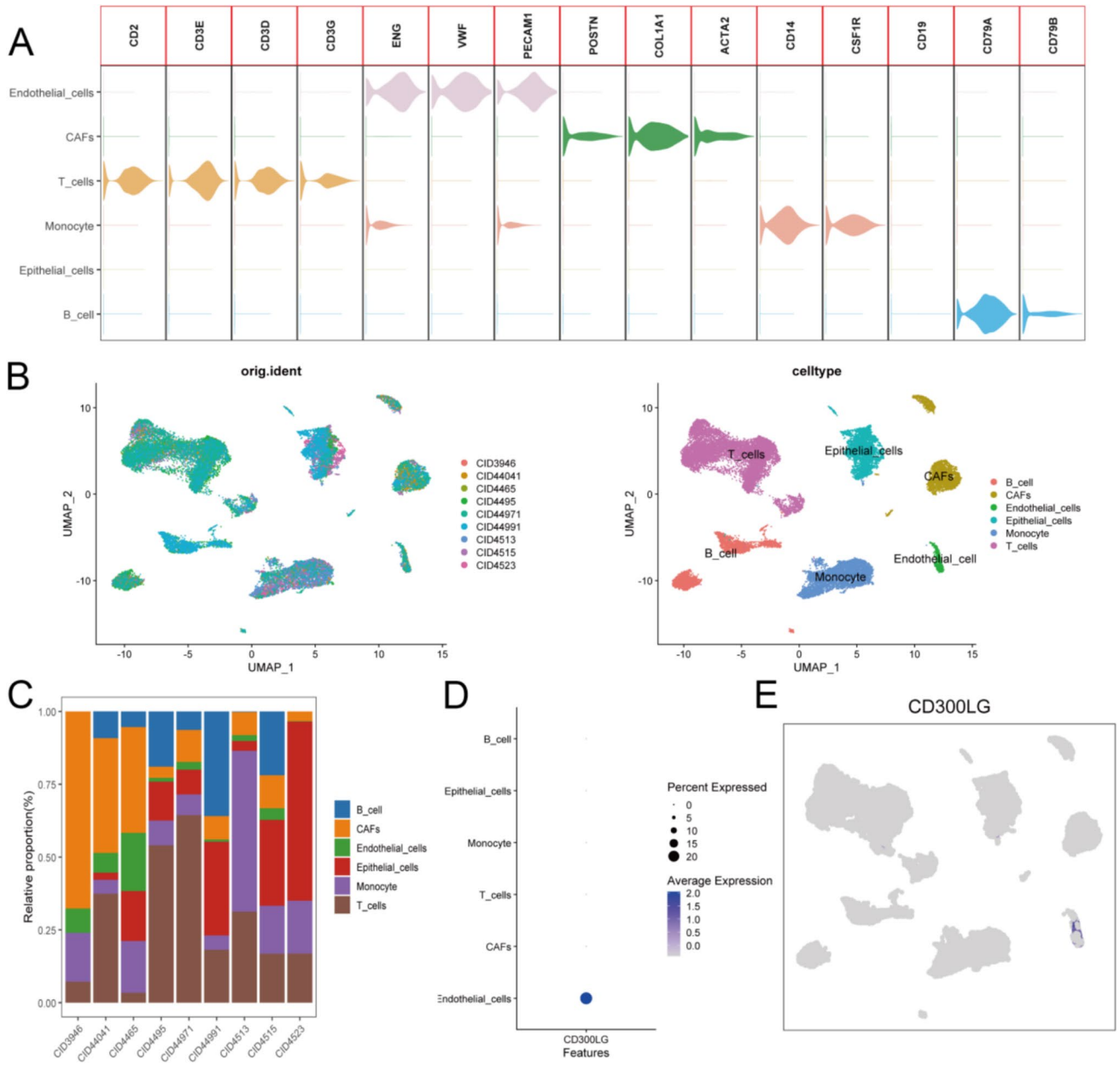
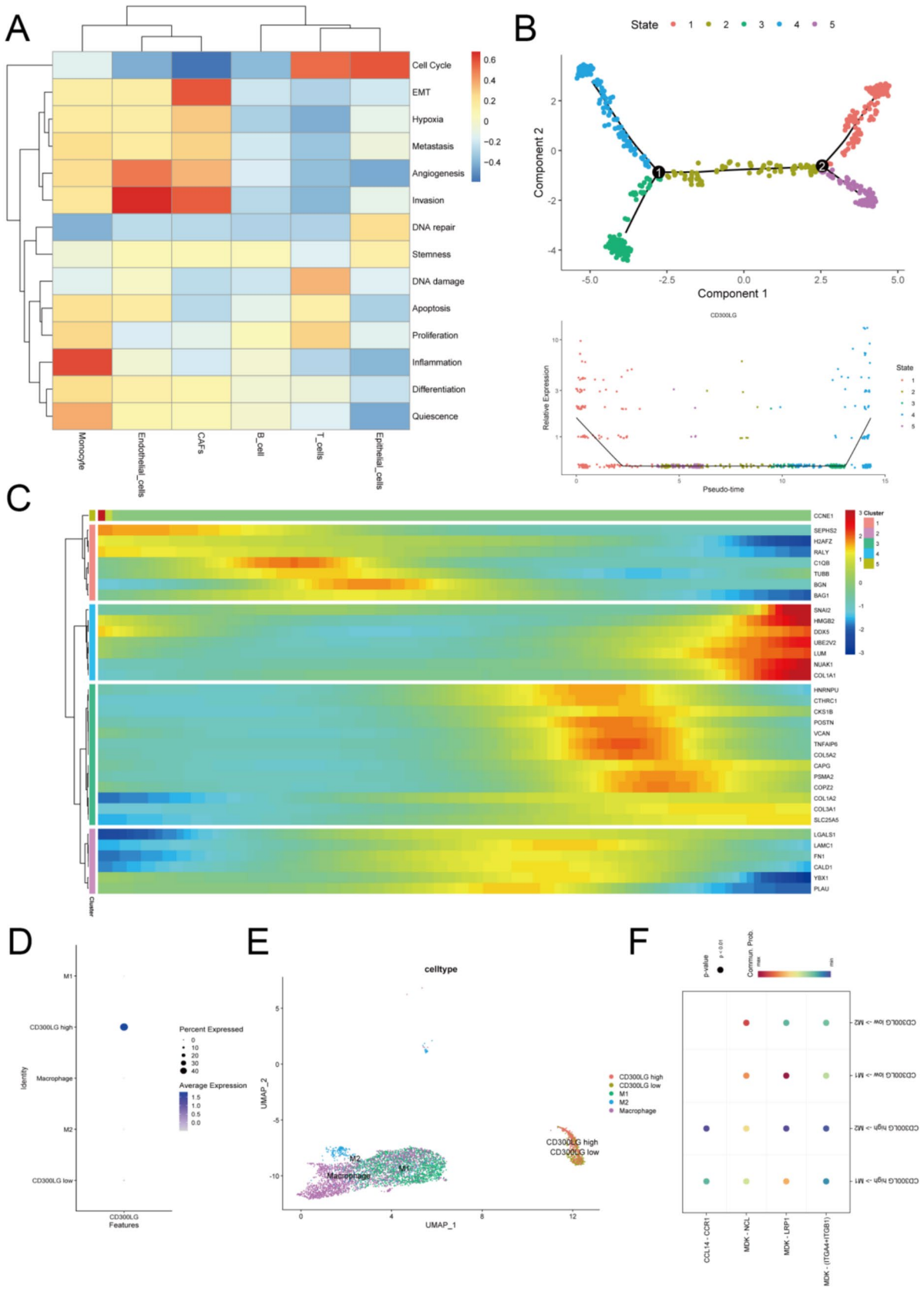


Fig. 6 Single-cell analysis of CD300LG in TNBC. **A** Gene marker plot for cell type annotation in single-cell RNA sequencing data analysis. **B** Dimensionality reduction clustering of single-cell sequencing data from GSE176078 TNBC patients. **C** Proportion of cell types in TNBC patient samples. **D**, **E** CD300LG is mainly expressed in endothelial cells

phenotypic characterization using immunofluorescence assays. In the present investigation, immunofluorescence staining was utilized to evaluate the expression and localization of target proteins within tissue samples. Specimens were fixed via perfusion with 4% paraformaldehyde, subsequently dehydrated in a series of sucrose solutions (10%, 20%, and 30%, each for 24 h), embedded in OCT compound, and stored at -80 °C until further processing. Cryostat

Fig. 7 Presents single-cell enrichment analysis, cell differentiation trajectory, and cell communication analysis. **A** Endothelial cells exhibit a strong correlation with tumor invasion. **B**, **C** CD300LG expression varies significantly during cell differentiation, with invasive genes showing progressively higher expression as differentiation proceeds. **D**, **E**, **F** The interaction between M2-type macrophages and other cells was significantly more frequent in the high CD300LG expression group compared to the low CD300LG expression group


A
Cell Cycle
EMT
Hypoxia
Metastasis
Angiogenesis
Invasion
DNA repair
Stemness
DNA damage
Apoptosis
Proliferation
Inflammation
Differentiation
Quiescence
Monocyte
Endothelial_cells
CAFs
B_cell
T_cells
Epithelial_cells
B
State 1 2 3 4 5
Component 2
Component 1
CD300LG
Relative Expression
Pseudo-time
C
CCNE1
SEPHS2
H2AFZ
RALY
C1QB
TUBB
BGN
BAG1
SNAI2
HMGB2
DDX5
UBE2V2
LUM
NUAK1
COL1A1
HNRNPU
CTHRC1
CKS1B
POSTN
VCAN
TNFAIP6
COL5A2
CAPG
PSMA2
COPZ2
COL1A2
COL3A1
SLC25A5
LGALS1
LAMC1
FN1
CALD1
YBX1
PLAU
Cluster
D
M1
CD300LG high
Macrophage
M2
CD300LG low
Identity
Percent Expressed
Average Expression
CD300LG
Features
E
celltype
UMAP_2
UMAP_1
M2
Macrophage
M1
CD300LG high
CD300LG low
F
p-value
Commun. Prob.
CD300LG low -> M2
CD300LG low -> M1
CD300LG high -> M2
CD300LG high -> M1
CCL14 - CCR1
MDK - NCL
MDK - LRP1
MDK - (ITGA4+ITGB1)

Sects. (5 μm thick) were prepared at -20 °C, adhered onto poly-L-lysine-coated slides, and dried at room temperature for 2 h prior to experimental procedures. Prior to staining, the sections were rinsed three times with PBS (pH 7.4; 5 min per wash) to eliminate residual embedding material. Antigen retrieval was achieved by heating the sections in 10 mM sodium citrate buffer (pH 6.0) at 95 °C for 15 min, followed by cooling to room temperature. Endogenous peroxidase activity was inhibited by incubation in a solution of 3% H_2O_2 in methanol for 10 min. To reduce non-specific binding, the sections were treated with a blocking solution consisting of 5% donkey serum and 1% BSA in PBS for 1 h at room temperature. Primary antibodies, diluted to their optimal concentrations based on preliminary experiments (e.g., anti-XX monoclonal antibody at 1:200), were applied and incubated overnight (16–18 h) at 4 °C in a humidified environment. On the subsequent day, after extensive washing with PBS (three times for 10 min each), fluorescently conjugated secondary antibodies specific to the respective species were added and incubated for 1.5 h at room temperature in the dark. Nuclei were stained with DAPI (1 μg/mL) for 5 min, followed by a PBS rinse. Finally, the sections were mounted with an anti-fade medium to maintain fluorescence integrity. All participants provided written informed consent. This study was approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University (R 2024–007).

Statistical analysis

To examine the differences between the two groups, the Wilcoxon rank sum test was utilized. For prognostic analysis, the Kaplan–Meier method combined with the log-rank test was employed. All statistical assessments were conducted using R software (version 4.1.2; accessible at https://www.r-project.org/) [31–35]. Statistical significance was defined as a two-tailed P value less than 0.05.

Results

Machine learning algorithm screens CD300LG as a key target for TNBC

We conducted variance analysis to compare TNBC and paracancerous tissues in the TCGA database, identifying a total of 4,369 differentially expressed genes with |log2(fold change)|> 1 and a false discovery rate (FDR) < 0.05. These 4,369 genes were then analyzed using four machine learning algorithms: SVM, GBM, Random Forest, and XGBoost, to screen for key targets associated with TNBC. After intersecting the results from all four algorithms, CD300LG was consistently ranked among the top ten genes (Fig. 1).

CD300LG as a prognostic factor associated with adverse outcomes in TNBC

We examined the predictive importance of CD300LG in patients with triple-negative breast cancer. Our findings indicated that the expression of CD300LG was substantially reduced in TNBC tissues. Additionally, the expression levels of CD300LG decreased gradually as the tumor stage progressed. The survival analysis revealed that individuals with lower CD300LG expression tended to have worse outcomes compared to those with higher expression levels. This conclusion was corroborated by an independent dataset (GSE58812), which similarly indicated that patients exhibiting elevated CD300LG expression had improved survival rates (Fig. 2).

CD300LG was implicated in cell cycle regulation and inflammatory response pathways

CD300LG was categorized into two groups based on its expression levels, and differential analysis was conducted. Subsequently, differential gene expression and enrichment analyses were performed. GO and KEGG enrichment analyses indicated that CD300LG is associated with the cell cycle, extracellular matrix organization, and inflammatory response pathways. GSVA enrichment analysis highlighted a strong association between CD300LG and tumor invasion. GSEA enrichment analysis revealed that the differentially expressed genes were predominantly enriched in the G2/M checkpoint signaling pathway and the IL-2 signaling pathway (Fig. 3).

CD300LG affected immune cell infiltration

The analysis of the CIBERSORT algorithm demonstrated that the infiltration levels of $CD8^+$ T cells and M1-type macrophages were notably elevated in the group with high CD300LG expression compared to the low-expression group. In contrast, M2-type macrophage infiltration was more significant in the low-expression cohort. According to the ssGSEA algorithm, the high-expression group exhibited increased levels of activated B cells, activated $CD8^+$ T cells, activated dendritic cells, central memory $CD4^+$ T cells, central memory $CD8^+$ T cells, and macrophages. Furthermore, the assessment method revealed greater immune scores and tumor purity scores in the high CD300LG expression group. On the other hand, mesenchymal and overall scores were observed to be higher in the low CD300LG expression group (Fig. 4).

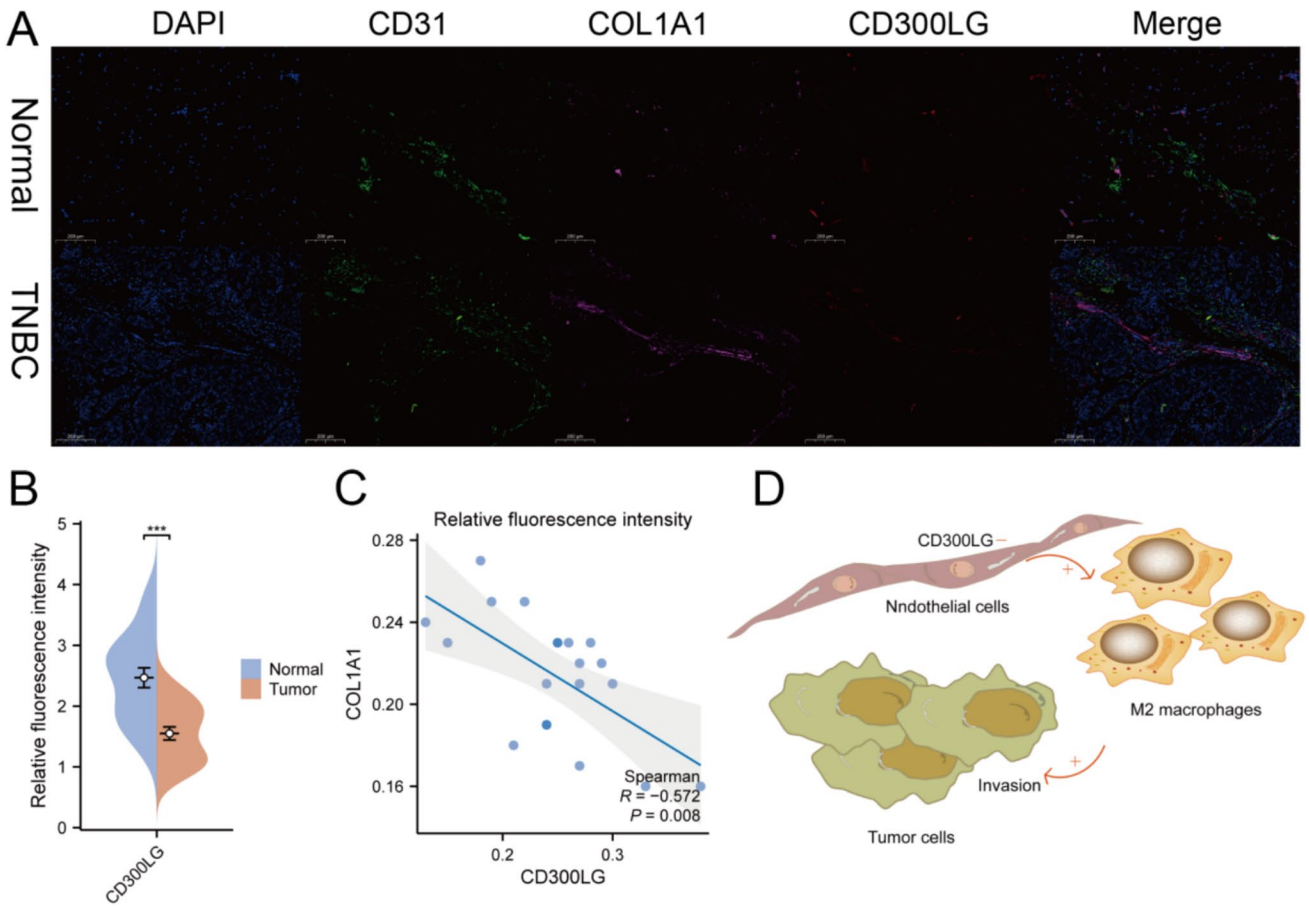


Fig. 8 The experiment confirmed that CD300LG participated in TNBC invasion. **A** The expression levels and locations of CD300LG and invasion genes were verified by immunofluorescence assay. **B**, **C** Fluorescence intensity relationship between CD300LG and COL1A1. **D** Schematic diagram of CD300LG's influence on TNBC invasion. (DAPI: a marker for nucleus. CD31: a marker for endothelial cells. COL1A1: a gene associated with invasion.)

CD300LG influences the efficacy of immunotherapy in patients with TNBC

TIDE analysis demonstrated that the scores were elevated in the CD300LG low-expression group relative to the high-expression group, implying reduced effectiveness of immunotherapy for individuals with lower CD300LG expression. According to TCGA data, patients within the CD300LG high-expression group were more likely to respond favorably to CTLA-4 therapy, whereas those in the low-expression group exhibited improved outcomes following PD-1 treatment. In line with these findings, the expression levels of functional genes in immune cells were markedly greater in the CD300LG high-expression group compared to the low-expression group. Furthermore, a negative association was identified between CD300LG expression and various immune checkpoint inhibitory genes (Fig. 5).

CD300LG was predominantly expressed in tumor-associated endothelial cells

Single-cell sequencing data from nine triple-negative breast cancers were quality-controlled and integrated, resulting in a dataset comprising 19,324 genes and 21,064 cells for subsequent analysis (Fig. 6). CD300LG was found to be predominantly expressed in triple-negative breast cancers, particularly in association with tumor-associated endothelial cells, guiding the focus of our study. We conducted GSVA enrichment analysis across various cell types, which corroborated the transcriptome data analysis results, revealing that tumor-associated endothelial cells exhibited the highest correlation with tumor invasion. Cell differentiation trajectory analysis demonstrated a progressive increase in the expression of invasion-related genes in tumor-associated endothelial cells as they differentiated. Additionally, intercellular communication patterns differed significantly between the CD300LG high-expression group and the CD300LG low-expression

group, as well as between M1 and M2-type macrophages (Fig. 7).

Experimental validation

The fluorescence intensity of CD300LG exhibited a significant negative correlation with that of COL1A1. Additionally, the expression patterns of CD300LG and CD31 were highly coincident. Our study demonstrated that CD300LG is predominantly expressed in endothelial cells and plays a role in influencing the invasive characteristics of TNBC (Fig. 8).

Discussion

TNBC is a highly aggressive form of breast cancer defined by the lack of expression of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 [36, 37]. Presently, the main therapy for TNBC involves chemotherapeutic drugs like paclitaxel and anthracyclines. However, these agents demonstrate restricted effectiveness and a significant tendency toward drug resistance [38]. In the past few years, immune checkpoint blockers, especially those targeting PD-1/PD-L1, have shown a degree of efficacy in specific patients with TNBC. Nevertheless, the percentage of individuals who benefit remains relatively low [39, 40]. Additionally, PARP inhibitors have shown promise in patients with BRCA mutations, but their applicability is restricted [41]. Overall, therapeutic options for TNBC are limited, lacking specific targeted therapies and resulting in poor patient outcomes [42, 43]. Current treatment strategies face significant challenges, including chemoresistance, low response rates to immunotherapy, and narrow applicability of targeted therapies [44]. An urgent requirement exists for the discovery of novel therapeutic targets and the creation of personalized treatment strategies [45].

Machine learning offers substantial advantages in medical research by efficiently processing large-scale, high-dimensional biomedical data [46, 47]. Through algorithmic models, it can identify potential biomarkers and therapeutic targets, thereby enhancing the accuracy of disease prediction and diagnosis [48]. Its capabilities in automated feature selection and pattern recognition offer novel insights into the molecular mechanisms of complex diseases. Single-cell technology, meanwhile, elucidates cellular heterogeneity at single-cell resolution, providing precise characterization of the tumor microenvironment, immune cell subpopulations, and their interactions [49]. This technology reveals unprecedented detail for disease mechanism research [50]. The integration of machine learning with single-cell technology enables a more comprehensive analysis of disease molecular characteristics, facilitating the discovery of new therapeutic targets and advancing precision medicine [51]. This multidisciplinary approach provides robust technical support for disease diagnosis, treatment, and drug development [52].

CD300LG influences the activation of immune cells by interacting with its receptors, such as CD300a and CD300c. Studies have shown that in sepsis models, reduced expression of CD300LG intensifies systemic inflammation, leading to increased vascular permeability and multi-organ dysfunction [53]. Additionally, in atherosclerosis, CD300LG suppresses macrophage inflammatory polarization, specifically by decreasing M1-type pro-inflammatory macrophages, which helps slow down plaque progression [54]. This highlights its potential as a therapeutic target for combating atherosclerosis. Furthermore, CD300LG exhibits abnormal expression patterns in tumor-associated vascular endothelial cells, potentially affecting tumor immune evasion pathways. For example, in melanoma and colorectal cancer, lower levels of CD300LG are associated with irregular tumor vascular growth and an immunosuppressive microenvironment [55]. At a mechanistic level, CD300LG may regulate anti-tumor immunity by altering interactions between endothelial cells and immune cells, such as by inhibiting the recruitment of Treg cells [56]. In certain cancers, such as breast cancer, higher CD300LG expression correlates with improved patient outcomes, suggesting its utility as a prognostic biomarker or a novel immunotherapy target [57, 58]. These findings underscore the multifaceted role of CD300LG in immune regulation and its implications in various diseases.

In the past few years, the integration of machine learning and single-cell technology into breast-related disease studies has led to substantial advancements. This approach provides fresh insights and techniques for exploring the molecular processes, as well as the diagnosis and therapy of breast cancer. For instance, Wang et al. utilized Random Forest and Support Vector Machine algorithms to analyze breast cancer transcriptome data, successfully distinguishing between different subtypes and predicting patient prognosis [59]. Additionally, Zhang et al. demonstrated that the XGBoost algorithm could more accurately predict breast cancer recurrence risk when integrating multi-omics data [60]. Wu et al. identified tumor-associated fibroblasts and immune cell subpopulations via single-cell analysis, revealing their roles in tumor progression [61]. Another study utilized single-cell technology to identify a subpopulation of drug-resistant cells in triple-negative breast cancer, offering novel insights into overcoming chemotherapy resistance [62]. Liu et al. integrated single-cell transcriptomic data with deep learning models to uncover key genes associated with immune escape in TNBC [63]. Similarly, Chen et al. applied single-cell technology and machine learning algorithms to discover specific markers for breast cancer stem cells [64]. Another study identified gene networks linked to chemotherapy

sensitivity and resistance via single-cell analysis, providing a foundation for personalized treatment strategies [65].

Our study identified CD300LG as a converging gene in the results of four machine learning algorithms. Further analysis demonstrated that CD300LG expression was significantly reduced in TNBC and progressively decreased with advancing tumor stages. Functional enrichment analysis indicated that CD300LG influences the cell cycle and modulates immune responses within the tumor microenvironment. Single-cell analysis showed that CD300LG is primarily present in tumor-associated endothelial cells. Furthermore, significant variations were observed in the interaction patterns between the high- and low-expression groups of CD300LG with M1 and M2 macrophages.

Our research is subject to certain limitations. First, we did not examine the impact of CD300LG on the cell cycle at the cellular level. Second, we failed to develop an animal model to analyze the role of CD300LG within the tumor microenvironment. Lastly, we did not delve into the molecular mechanisms that govern the function of CD300LG.

In summary, CD300LG was pinpointed using machine learning as a key factor affecting prognosis and the tumor immune microenvironment in triple-negative breast cancer. This molecule is expected to become a promising therapeutic target for this specific type of breast cancer.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10238-025-01690-3.

Acknowledgements None

Author's contributions The research is designed by JP, and drafted by BZ and HW, while JP performed the final revision. HW performed the experiments. Data collection and collation are performed by BZ, while ZL and HJ performed data analysis.

Funding None.

Data availability The relevant code and data can be obtained through correspondence with the corresponding author in scientific research.

Declarations

Conflict of interest All authors declare that there is no conflict of interest.

Ethical approval This study was approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University. All the patients involved signed informed consent forms.

References

1. Jamdade VS, Sethi N, Mundhe NA, Kumar P, Lahkar M, Sinha N. Therapeutic targets of triple-negative breast cancer: a review. Br J Pharmacol. 2015;172(17):4228–37.
2. Li Y, Zhang H, Merkher Y, Chen L, Liu N, Leonov S, et al. Recent advances in therapeutic strategies for triple-negative breast cancer. J Hematol Oncol. 2022;15(1):121.
3. De Laurentiis M, Cianniello D, Caputo R, Stanzione B, Arpino G, Cinieri S, et al. Treatment of triple negative breast cancer (TNBC): current options and future perspectives. Cancer Treat Rev. 2010;36(Suppl 3):S80–6.
4. Keenan TE, Tolaney SM. Role of immunotherapy in triple-negative breast cancer. J Natl Compr Canc Netw. 2020;18(4):479–89.
5. Li Z, Qiu Y, Lu W, Jiang Y, Wang J. Immunotherapeutic interventions of triple negative breast cancer. J Transl Med. 2018;16(1):147.
6. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res. 2020;22(1):61.
7. Wang Z, Jiang Q, Dong C. Metabolic reprogramming in triple-negative breast cancer. Cancer Biol Med. 2020;17(1):44–59.
8. Cao L, Niu Y. Triple negative breast cancer: special histological types and emerging therapeutic methods. Cancer Biol Med. 2020;17(2):293–306.
9. Huang X, Ali A, Yachioui DEI, Le Dévédec SE, Hankemeier T. Lipid dysregulation in triple negative breast cancer: Insights from mass spectrometry-based approaches. Prog Lipid Res. 2025;98: 101330.
10. Hwang SY, Park S, Kwon Y. Recent therapeutic trends and promising targets in triple negative breast cancer. Pharmacol Ther. 2019;199:30–57.
11. Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. Cancer Res. 2013;73(7):2025–30.
12. Ahuja A, Al-Zogbi L, Krieger A. Application of noise-reduction techniques to machine learning algorithms for breast cancer tumor identification. Comput Biol Med. 2021;135: 104576.
13. Das SC, Tasnim W, Rana HK, Acharjee UK, Islam MM, Khatun R. Comprehensive bioinformatics and machine learning analyses for breast cancer staging using TCGA dataset. Briefings Bioinform. 2024. https://doi.org/10.1093/bib/bbae628.
14. Lu X, Gou Z, Chen H, Li L, Chen F, Bao C, et al. Gene panel predicts neoadjuvant chemoimmunotherapy response and benefit from immunotherapy in HER2-negative breast cancer. J Immunother Cancer. 2024;12(8):e009587.
15. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. NPJ Breast Cancer. 2023;9(1):21.
16. Wu X, Chen M, Liu K, Wu Y, Feng Y, Fu S, et al. Molecular classification of geriatric breast cancer displays distinct senescent subgroups of prognostic significance. Mol Ther Nucleic Acids. 2024;35(4): 102309.

17. Zhang C, Zhai W, Ma Y, Wu M, Cai Q, Huang J, et al. Integrating machine learning algorithms and multiple immunohistochemistry validation to unveil novel diagnostic markers based on costimulatory molecules for predicting immune microenvironment status in triple-negative breast cancer. Front Immunol. 2024;15:1424259.
18. Li YJ, Nuytemans K, La JO, Jiang R, Slifer SH, Sun S, et al. Identification of novel genes for age-at-onset of Alzheimer's disease by combining quantitative and survival trait analyses. Alzheimers Dement. 2023;19(7):3148–57.
19. Yang Z, Liu L, Zhu Z, Hu Z, Liu B, Gong J, et al. Tumor-associated monocytes reprogram CD8(+) T cells into central memory-like cells with potent antitumor effects. Adv Sci (Weinh). 2024;11(16): e2304501.
20. Lee-Ødegård S, Hjorth M, Olsen T, Moen G-H, Daubney E, Evans DM, et al. Serum proteomic profiling of physical activity reveals CD300LG as a novel exerkine with a potential causal link to glucose homeostasis. Elife. 2024. https://doi.org/10.7554/eLife.96535.3.
21. Li Q, Xie D, Yao L, Qiu H, You P, Deng J, et al. Combining autophagy and immune characterizations to predict prognosis and therapeutic response in lung adenocarcinoma. Front Immunol. 2022;13: 944378.
22. Wei Q, Jiang X, Miao X, Zhang Y, Chen F, Zhang P. Molecular subtypes of lung adenocarcinoma patients for prognosis and therapeutic response prediction with machine learning on 13 programmed cell death patterns. J Cancer Res Clin Oncol. 2023;149(13):11351–68.
23. Adriaens ME, Lodder EM, Moreno-Moral A, Šilhavý J, Heinig M, Glinge C, et al. Systems genetics approaches in rat identify novel genes and gene networks associated with cardiac conduction. J Am Heart Assoc. 2018;7(21): e009243.
24. Ding S, Chen X, Shen K. Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Commun (Lond). 2020;40(8):329–44.
25. So JY, Ohm J, Lipkowitz S, Yang L. Triple negative breast cancer (TNBC): Non-genetic tumor heterogeneity and immune microenvironment: emerging treatment options. Pharmacol Ther. 2022;237: 108253.
26. Asleh K, Riaz N, Nielsen TO. Heterogeneity of triple negative breast cancer: current advances in subtyping and treatment implications. J Exp Clin Cancer Res. 2022;41(1):265.
27. Yamashita N, Long M, Fushimi A, Yamamoto M, Hata T, Hagiwara M, Bhattacharya A, Qiang H, Wong K-K, Liu S, Kufe D. MUC1-C integrates activation of the IFN-γ pathway with suppression of the tumor immune microenvironment in triple-negative breast cancer. J ImmunoTherapy Cancer. 2021;9(1):e002115. https://doi.org/10.1136/jitc-2020-002115.
28. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell. 2018;174(5):1293-308.e36.
29. Shiao SL, Gouin KH 3rd, Ing N, Ho A, Basho R, Shah A, et al. Single-cell and spatial profiling identify three response trajectories to pembrolizumab and radiation therapy in triple negative breast cancer. Cancer Cell. 2024;42(1):70-84.e8.
30. Zhang Y, Chen H, Mo H, Hu X, Gao R, Zhao Y, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. Cancer Cell. 2021;39(12):1578-93.e8.
31. Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, et al. Using clusterProfiler to characterize multiomics data. Nat Protoc. 2024;19(11):3292–320.
32. Cao Y, Dong Q, Wang D, Zhang P, Liu Y, Niu C. microbiomeMarker: an R/Bioconductor package for microbiome marker identification and visualization. Bioinformatics. 2022;38(16):4027–9.
33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
34. Lim CN, Liang S, Feng K, Chittenden J, Henry A, Mouksassi S, et al. Phxnlme: an R package that facilitates pharmacometric workflow of phoenix NLME analyses. Comput Methods Programs Biomed. 2017;140:121–9.
35. Nikolaienko O, Lønning PE, Knappskog S. epialleleR: an R/bioconductor package for sensitive allele-specific methylation analysis in NGS data. GigaScience. 2022. https://doi.org/10.1093/gigascience/giad087.
36. Leon-Ferre RA, Goetz MP. Advances in systemic therapies for triple negative breast cancer. BMJ. 2023;381: e071674.
37. Lyons TG. Targeted therapies for triple-negative breast cancer. Curr Treat Options Oncol. 2019;20(11):82.
38. Ye F, Dewanjee S, Li Y, Jha NK, Chen ZS, Kumar A, et al. Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. Mol Cancer. 2023;22(1):105.
39. Liu Y, Hu Y, Xue J, Li J, Yi J, Bu J, et al. Advances in immunotherapy for triple-negative breast cancer. Mol Cancer. 2023;22(1):145.
40. Sharma P. Biology and management of patients with triple-negative breast cancer. Oncologist. 2016;21(9):1050–62.
41. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment landscape of triple-negative breast cancer—expanded options, evolving needs. Nat Rev Clin Oncol. 2022;19(2):91–113.
42. Agostinetto E, Eiger D, Punie K, de Azambuja E. Emerging therapeutics for patients with triple-negative breast cancer. Curr Oncol Rep. 2021;23(5):57.
43. Hu Y, Wang C, Liang H, Li J, Yang Q. The treatment landscape of triple-negative breast cancer. Med Oncol. 2024;41(10):236.
44. Jia H, Truica CI, Wang B, Wang Y, Ren X, Harvey HA, et al. Immunotherapy for triple-negative breast cancer: existing challenges and exciting prospects. Drug Resist Updat. 2017;32:1–15.
45. Yi H, Li Y, Tan Y, Fu S, Tang F, Deng X. Immune checkpoint inhibition for triple-negative breast cancer: current landscape and future perspectives. Front Oncol. 2021;11: 648139.
46. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55.
47. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng. 2022;6(12):1346–52.
48. Li LS, Yang L, Zhuang L, Ye ZY, Zhao WG, Gong WP. From immunology to artificial intelligence: revolutionizing latent tuberculosis infection diagnosis with machine learning. Mil Med Res. 2023;10(1):58.
49. Liu C, Xie J, Lin B, Tian W, Wu Y, Xin S, et al. Pan-cancer single-cell and spatial-resolved profiling reveals the immunosuppressive role of APOE+ macrophages in immune checkpoint inhibitor therapy. Adv Sci (Weinh). 2024;11(23): e2401061.
50. Gradeci D, Bove A, Charras G, Lowe AR, Banerjee S. Single-cell approaches to cell competition: high-throughput imaging, machine learning and simulations. Semin Cancer Biol. 2020;63:60–8.
51. Dohmen J, Baranovskii A, Ronen J, Uyar B, Franke V, Akalin A. Identifying tumor cells at the single-cell level using machine learning. Genome Biol. 2022;23(1):123.
52. Li J, Kong Z, Qi Y, Wang W, Su Q, Huang W, et al. Single-cell and bulk RNA-sequence identified fibroblasts signature and CD8 + T-cell - fibroblast subtype predicting prognosis and immune therapeutic response of bladder cancer, based on machine learning: bioinformatics multi-omics study. Int J Surg. 2024;110(8):4911–31.

53. Clark GJ, Ju X, Tate C, Hart DN. The CD300 family of molecules are evolutionarily significant regulators of leukocyte functions. Trends Immunol. 2009;30(5):209–17.
54. Rozenberg P, Reichman H, Moshkovits I, Munitz A. CD300 family receptors regulate eosinophil survival, chemotaxis, and effector functions. J Leukoc Biol. 2018;104(1):21–9.
55. Umemoto E, Takeda A, Jin S, Luo Z, Nakahogi N, Hayasaka H, et al. Dynamic changes in endothelial cell adhesion molecule nepmucin/CD300LG expression under physiological and pathological conditions. PLoS ONE. 2013;8(12): e83681.
56. Moshkovits I, Reichman H, Karo-Atar D, Rozenberg P, Zigmond E, Haberman Y, et al. A key requirement for CD300f in innate immune responses of eosinophils in colitis. Mucosal Immunol. 2017;10(1):172–83.
57. Kitaura J, Murakami M. Positive and negative roles of lipids in mast cells and allergic responses. Curr Opin Immunol. 2021;72:186–95.
58. Nissim Ben Efraim AH, Karra L, Ben-Zimra M, Levi-Schaffer F. The inhibitory receptor CD300a is up-regulated by hypoxia and GM-CSF in human peripheral blood eosinophils. Allergy. 2013;68(3):397–401.
59. Wang Y, Wang N, Xu M, Yu J, Qin C, Luo X, et al. Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. IEEE Trans Med Imaging. 2020;39(4):866–76.
60. Zhang H, Lin F, Zheng T, Gao J, Wang Z, Zhang K, et al. Artificial intelligence-based classification of breast lesion from contrast enhanced mammography: a multicenter study. Int J Surg. 2024;110(5):2593–603.
61. Wu Y, Yi Z, Li J, Wei Y, Feng R, Liu J, et al. FGFR blockade boosts T cell infiltration into triple-negative breast cancer by regulating cancer-associated fibroblasts. Theranostics. 2022;12(10):4564–80.
62. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. Cell. 2018;173(4):879-93.e13.
63. Liu H, Paddock MN, Wang H, Murphy CJ, Geck RC, Navarro AJ, et al. The INPP4B tumor suppressor modulates EGFR trafficking and promotes triple-negative breast cancer. Cancer Discov. 2020;10(8):1226–39.
64. Chen W, Zeng YC, Achinger-Kawecka J, Campbell E, Jones AK, Stewart AG, et al. Machine learning enables pan-cancer identification of mutational hotspots at persistent CTCF binding sites. Nucleic Acids Res. 2024;52(14):8086–99.
65. Li S, Wen Z, Ghalandari B, Zhou T, Warden AR, Zhang T, et al. Single-cell immunoblotting based on a photoclick hydrogel enables high-throughput screening and accurate profiling of exogenous gene expression. Adv Mater. 2021;33(22): e2101108.