Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating a Hundred Machine Learning Combinations and Single- Cell Transcriptomics to Define a KIAA0408-Driven B-Cell Prognostic Signature in Ovarian Cancer

Xiaoyu Qu
Second Affiliated Hospital of Nanchang University
Hao Wan
Nanchang University
Chenshuo Fang
Nanchang University
Ye Zhu
Nanchang University
Mingyuan Bo
Nanchang University
Ruiyu Zhang
Nanchang University
Shengwei Tang
Nanchang University
Xin Wang
Nanchang University
Ruida Shi
Nanchang University
Zhengchang Liu
Nanchang University
Zixin Chen
Nanchang University
Kangcheng Wang
Nanchang University
Da Huang
Second Affiliated Hospital of Nanchang University
Ye Cao
Nanchang University
Haiyan Wang
Second Affiliated Hospital of Nanchang University

Yanying Zhong

zy_yin_g01@163.com

Second Affiliated Hospital of Nanchang University

Research Article

Keywords: Ovarian cancer, Prognostic model, Immune heterogeneity, Tumor immune microenvironment, Tumor-infiltrating B cells

Posted Date: May 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9681346/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Ovarian cancer is the most lethal gynecologic malignancy, with most patients exhibiting suboptimal responses to T cell-targeted immune checkpoint inhibitors. Despite increasing attention to B cells in tumor immunity, the role of tumor-infiltrating B cells in ovarian cancer remains incompletely elucidated.

Methods

Using single-cell RNA sequencing data, we constructed a high-resolution cellular atlas of the high-grade serous ovarian cancer, integrating inferCNV, CytoTRACE, and hdWGCNA to systematically analyze B cell copy number variations, differentiation trajectories, and co-expression networks. A machine learning-derived prognostic risk model was validated across multiple independent cohorts. Immune microenvironment was assessed by ESTIMATE, gene functions by flow cytometry, colony formation, and CCK-8 assays, and drug candidates were identified by database screening with experimental validation.

Results

Single-cell sequencing reveals ovarian cancer B cell heterogeneity, aberrant T-cell crosstalk, and *KIAA0408* as a core regulator. The prognostic model based on *KIAA0408*, *CEACAM8*, *CAPZA1*, *PTCH1*, and *VPS13A* performed robustly across ovarian cancer subtypes. High-risk patients exhibited an "ineffective inflammation" phenotype, featuring myeloid-derived suppressor cell-predominant infiltration, insufficient memory T cells, and impaired anti-tumor immunity. Mechanistically, *KIAA0408* induced B cell cycle arrest and immunosuppressive/pro-tumorigenic phenotypic transformation by inhibiting the B cell receptor/PI3K-AKT pathway, with the natural compound isoliquiritigenin predicted and preliminarily validated as a pathway target.

Conclusion

This study reveals *KIAA0408*-mediated immune regulation of B cells in ovarian cancer, establishes a multi-gene prognostic model, and identifies isoliquiritigenin as a potential therapeutic agent.

Introduction

Ovarian cancer (OC) is the most lethal gynecologic malignancy and the third most prevalent in women, severely endangering female health[1]. High-grade serous ovarian cancer (HGSOC), the dominant histological subtype, accounts for 80% of OC-related deaths[2]. HGSOC exhibits aggressive progression, diagnostic delay, frequent recurrence, and common chemoresistance[3]; most patients present at stage III-IV, with a 5-year overall survival of only ~ 30%[4]. Standard treatment comprises cytoreductive surgery

plus platinum-based chemotherapy, yet 70%-80% of advanced cases show 5-year survival below 40%[5], underscoring an urgent need for novel molecular targets and personalized therapeutic strategies.

Tumor immunotherapy has revolutionized OC care. Immune checkpoint inhibitors targeting PD-1/PD-L1 yield objective responses in only 10%-15% of patients[6], and research centered on T-cell regulation has shown limited clinical benefit[7]. Increasing evidence highlights the critical role of B cells (BC) in the ovarian tumor microenvironment (TME)[8]. Tumor-infiltrating B cells (TIL-Bs) modulate antitumor immunity via antigen presentation, antibody and cytokine secretion, and cellular crosstalk[9, 10]; plasma cell infiltration positively correlates with prolonged survival[11, 12]. However, the heterogeneity and regulatory mechanisms of TIL-Bs in OC remain poorly defined.

Single-cell RNA sequencing (scRNA-seq) technology enables the resolution of transcriptomic features at single-cell precision and has become a pivotal tool for dissecting TME heterogeneity[13–16]. While this technique has successfully mapped TIL-B heterogeneity in cancers such as breast and lung cancer, its application in OC remains limited, leaving the molecular regulatory networks and prognostic associations of these cells largely unexplored[17].

To address these unmet needs, we performed a systematic characterization of TIL-B heterogeneity in OC using scRNA-seq. Combining bioinformatic analyses and functional assays, we explored the regulatory mechanisms of TIL-Bs and their prognostic implications. Machine learning was applied to build a robust prognostic model and identify core molecular targets, followed by screening for candidate therapeutic agents. This study aims to clarify the regulatory landscape of TIL-Bs in OC, providing a theoretical framework and translational rationale for personalized prognosis and precision therapy.

Material and methods

Data Collection and Processing

All transcriptome data were obtained from the Gene Expression Omnibus (GEO) database. The scRNA-seq dataset GSE184880 contained 59,324 cells from 7 HGSOC patients and 5 normal ovarian samples. Bulk cohorts included GSE26193 (training, n = 107), GSE14764 (validation, n = 80) and GSE63885 (external validation, n = 101). GSE54388 was used for pathway enrichment analysis.

Single-cell RNA-seq data processing

Data processing was performed using Seurat (v5.3.0). Cells were filtered according to gene counts (200–5,000) and mitochondrial proportion (< 10%), and genes expressed in fewer than 10 cells were excluded. Batch correction was conducted via Harmony, followed by dimensionality reduction and visualization with t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP).

Differentially expressed genes (DEGs) identification and functional enrichment analysis

DEGs were identified by the Wilcoxon rank-sum test (adjusted $P < 0.05$) and functionally annotated using clusterProfiler for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment. Gene Set Variation Analysis (GSVA) was applied with Hallmark and Reactome signatures to compare pathway activity between tumor and normal tissues.

High-dimensional Weighted Gene Co-expression Network (hdWGCNA) analysis

hdWGCNA was analyzed separately for BC (B cell) and ASC (antibody-secreting cell) clusters. Soft-thresholding power of 7 yielded 5 modules in BC cells, while power of 3 generated 2 modules in ASC cells. Module activity was scored by UCell, and hub gene networks were constructed and visualized.

Copy Number Variations (CNVs) analysis

Chromosomal CNVs were estimated using inferCNV. Aggregated CNV scores were compared between tumor and normal groups across cell lineages, with further heterogeneity analysis specifically performed in B cell subtypes.

Cell trajectory analysis

CytoTRACE was used to assess transcriptional stemness and differentiation potential. Monocle3 (v1.3.7) was adopted to reconstruct B cell pseudotime trajectories with UMAP dimensionality reduction.

Cell-cell communication analysis

CellChat (v1.6.1) and CellChatDB.human were used to characterize intercellular crosstalk. Low-confidence interactions with fewer than 10 cells were filtered, and credible ligand-receptor pairs were quantified.

Construction of prognostic model driven by machine learning

To develop a robust prognostic model, we integrated 10 machine learning algorithms and their combinations. These included penalized regressions (LASSO, Ridge, and Enet), ensemble learning (Random Survival Forest, CoxBoost, and GBM), and other sophisticated frameworks such as SuperPC, plsRcox, StepCox, and survival-SVM.All models were trained using GSE26193 and verified with GSE14764 and GSE63885.

The validation of the prognostic model's performance

Model performance was assessed by the C-index (0–1), with the model of the highest mean C-index across three cohorts selected for further analysis. Calibration curves validated predicted vs. actual 1-, 3-, 5-year recurrence-free rates; Kaplan-Meier curves (survminer/ggsurvplot) were generated for risk groups. Cox regression determined hazard ratio (HR) and 95% confidence interval (CI); 5-year receiver operating characteristic (ROC), decision curve analysis (DCA), and Schoenfeld residual tests were additionally performed.

Immune infiltration analysis

Bar graph stacks illustrated the immunological infiltration topography of OC utilising the “CIBERSORT” method. Independent t-tests were employed to analyse the prevalence of 24 immune cell types in cancer and normal samples, as well as in high- and low-risk subgroups. Heatmaps created by “ggplot2” and “linkET” visualised correlations between immune cell populations and five prognostic biomarkers. ESTIMATE scores were computed, and Spearman correlation analysis examined the relationship between the prognostic model and immune infiltration.

Cell Culture

Human SKOV3 ovarian cancer cells were cultured in DMEM/F-12 medium supplemented with 10% fetal bovine serum (FBS), penicillin and streptomycin at 37°C with 5% CO_2, and passaged at 90% confluence.

Isolation of human CD20⁺ B Cells and Stable Transfection

CD20⁺ B cells were isolated from clinical ovarian cancer tissues using the CD20 MicroBead Kit (Miltenyi Biotec). *KIAA0408*-overexpressing (OE) and control B cells were generated via lentiviral transduction followed by puromycin selection to ensure stable expression.

Flow Cytometric Analysis of CD79a Expression

Isolated B cells were blocked and stained with anti-CD79a antibodies to evaluate the expression of the B cell receptor B (BCR) signaling subunit. BCR expression levels were quantified by mean fluorescence intensity (MFI) using FlowJo software.

Phospho-Flow Cytometric Analysis of AKT Phosphorylation

B cells were serum-starved (0.5% FBS) and stimulated with anti-IgM $F(ab')_2$ for indicated durations. Following fixation and permeabilization, p-AKT (Ser473) levels were detected by flow cytometry and quantified as geometric MFI (gMFI).

Cell Cycle

B cells were categorized into control, *KIAA0408*-OE, and *KIAA0408*-OE + SC79 (AKT activator) groups. After ethanol fixation and propidium iodide (PI) staining, cell cycle distribution was analyzed via flow cytometry.

B Cell-CM Co-culture Colony Formation Assay

Conditioned media (CM) were harvested from B cells (Control, Vector, or *KIAA0408*-OE) cultured in 1% FBS medium for 24 h. SKOV3 cells were seeded (600 cells/well) and treated with a 1:1 mixture of CM and 20% FBS complete medium. After 10–14 days of incubation, visible colonies (≥ 50 cells) were fixed with 4% PFA, stained with 1% crystal violet, and the total number of colonies per well was quantified.

Compound Screening and Drug Sensitivity Analysis

Connectivity Map (CMap) (https://clue.io/) explores disease-target associations via transcriptomic data[18]. *KIAA0408* expression group DEGs were submitted to CMap for OC therapeutic target and drug identification. HERB (http://47.92.70.12/) and TCMbank (https://tcmbank.cn/) screening of six PI3K/AKT-related genes (*PIK3CA*, *PDPK1*, *AKT1*, *AKT*, *MTOR* and *RPS6KB1*) collectively identified three overlapping compounds[19, 20]:

Molecular Docking Analysis

CB-Dock2 (https://cadd.labshare.cn/cb-dock2/index.php)[21], for protein-ligand blind docking with cavity detection, docking and template fitting, was used for molecular docking of *KIAA0408* and isoliquiritigenin (ISL). PyMOL 2.3.6 dissociated the native ligand and visualized the results.

ISL Sensitivity and Intervention Assays

SKOV3 viability following ISL treatment (24–72 h) was determined using the CCK-8 kit (Dojindo, Japan). Absorbance was measured at 450 nm to evaluate the dose-dependent inhibitory effect of ISL.

SKOV3 cells were cultured in *KIAA0408*-overexpressing B cells (OE-B-CM) with or without ISL for 10–14 days. Visible colonies were quantified after crystal violet staining to verify whether ISL effectively abrogated the CM-triggered cell proliferation.

Statistical Analysis

All statistical analyses were performed using R v4.4.2 and GraphPad Prism 10 Between-group differences were analyzed via Wilcoxon test, and correlations were assessed by Pearson's correlation analysis. Data were presented as mean ± SD. $P < 0.05$ indicated statistical significance.

Results

Single-cell RNA-seq reveals the transcriptomic features of GSE184880 ovarian cancer cells

First, to examine the cellular composition and heterogeneity between OC and normal tissues, we used scRNA-seq on 59,324 cells from the OC dataset GSE184880. Following quality control and batch effect elimination, these cells were categorized into 13 clusters utilizing both t-SNE and UMAP **(**Fig. 1A, 1B**)**. We subsequently observed and compared cells from OC samples and normal ovarian samples **(**Fig. 1C**)**. Subsequently, utilizing recognized marker gene expression levels, we annotated cells within the UMAP and t-SNE dimensionality reduction spaces, identifying seven cell types: T cell, B cell, Macrophage, Mature NK T cell, Endothelial cell, Cancer stem cell, Dendritic cell, and Fibroblast **(**Fig. 1D–1F**)**. Consequently, we re-evaluated the cell distributions across the normal and tumor cohorts **(**Fig. 1G, **Figure S1A)**. Next, we analyzed the distribution of each cell type in normal and tumor samples, together with the ratio of normal to tumor samples across 13 clusters. Analysis demonstrated a notable disparity

in B cell abundance between the two groups (Fig. 1H **and Figure S1B)**. We concurrently presented the absolute numbers and relative proportions of seven cell categories across 12 samples, therefore reaffirming the significant heterogeneity of BC **(Figure S1C, 1D)**. Furthermore, to initially ascertain pivotal genes affecting OC progression, we examined differentially expressed genes among the seven cell types. Simultaneously, we performed FindAllMarkers differential analysis on genes among these seven cell types (Fig. 1I **and Figure S1E)**. Next, we conducted GO, KEGG, and GSVA enrichment analyses on the BC DEGs. GSVA results indicated that OC tissue exhibits significant enrichment of Hallmark pathways closely linked to malignant tumor characteristics (Fig. 1J **and Figure S2E)**. Ultimately, we examined CNVs among cell types and gene expression levels related with proliferation were assessed in all cells with the "AddModuleScore" method, indicating elevated expression in BC **(Figure S1F, 1G and Figure S2A-2D)**. We collectively identified BCs with notable disparities in gene abundance for further investigation.

Single-cell sequencing reveals unique characteristics of B cell subsets

To further examine the makeup and functional attributes of BC subpopulations, we then conducted scRNA-seq on these BC. Both t-SNE and UMAP clustering revealed four unique clusters. We additionally examined the displayed maps of normal and tumor cells obtained after UMAP dimensionality reduction (Fig. 2A-2C). Following that, we computed CNVs in BCs and the CNV scores of each tumor BC cluster were juxtaposed with the normal BC cluster. Cluster 1 exhibited the highest CNV score among the four BC clusters (Fig. 2D **and Figure S2F)**. Next, we identified these four cell clusters utilizing recognized classical markers: naive B cell (Bn), memory B cell (Bm), germinal center B cell (Bgc), and ASC (Fig. 2E-2G). Utilizing this annotation, we analyzed the distribution features of B cell subgroups in normal vs tumor samples (Fig. 2H **and Figure S2G)**. We further assessed the abundance of the four BC clusters in tumor samples compared to nearby normal samples. The results indicated a substantial disparity in ASC proportion (Fig. 2I). We subsequently conducted a statistical analysis of the quantity and relative proportions of the four B cell subgroups across 12 independent samples, yielding consistent results **(Figure S2H)**. Furthermore, we conducted a differential gene expression analysis of four BC subtypes (Fig. 2J **and Figure S2I)**. Thereafter, we conducted GO and KEGG enrichment analyses on the DEGs found in ASCs (Fig. 2K). In summary, ASCs' prevalence is markedly modified in OC relative to normal tissue.

Analysis of Interactions and Stemness in OC-Related Cells

To illustrate the interaction dynamics among cell types inside the TME, we developed a cell-cell communication network. Given the inadequate quantity of mature NK T cells, examining their intercellular connections proved unproductive. Hence, we concentrated exclusively on the other cell types. Initially, we examined the quantity and intensity of interactions among the six cell types. String diagram analysis revealed significant interaction nodes between BC and T cells (TC) in the TME. Analysis of the heatmaps revealed that compared with other cell types, BC accounted for a markedly lower proportion of total cell-cell interactions in tumor tissue, with an overall weaker interaction intensity.

(Fig. 3A, 3B). Subsequently, to identify cell populations acting as key senders or receivers inside the signaling network, we calculated two centrality metrics: out-degree and in-degree using CellChat (Fig. 3C). Simultaneously, we employed NMF clustering analysis to examine the degree to which the six cell types contribute to various outgoing signals **(Figure S3A)**. We subsequently examined ligand-receptor interactions between BC and other cell types, uncovering substantial regulatory influences of BC on TC (Fig. 3D). Next, we generated a bar chart illustrating the proportional contribution of each pathway **(Figure S3B)**. Simultaneously, we discovered that both the quantity and intensity of interactions were elevated in tumor tissues compared to normal tissues **(Figure S3C)**. Subsequently, we developed cell-cell communication networks for the four subtypes of BC and examined the quantity and strength of interactions among these subtypes. The findings indicated that the quantity and intensity of interactions were greater for Bgc and lesser for ASC within the TME (Fig. 3E, F). Next, we computed the out-degree and in-degree centrality metrics for the four BC subtypes (Fig. 3G). We further examined the ligand-receptor interactions between ASC and the three additional cell types **(Figure S3D)**. Furthermore, we illustrated the proportional contributions of each signaling channel to information flow and assessed cell-cell interaction among the four BC subsets in tumor and normal tissues. The findings indicated a pronounced escalation in communication intensity among B-cell subsets in tumor tissues **(Figure S3E, 3F)**. Ultimately, we employed CytoTRACE technology to delineate differentiation potential on the UMAP plot of the initial cluster, revealing that BC are in an advanced, terminally differentiated state (Fig. 3H). Overall, BC are pivotal in the OC interaction network and significantly enhance TC performance synergistically.

Exploring gene modules closely associated with B cells via hdWGCNA

Considering the significant correlation between BC and OC, it is essential to investigate co-expression gene networks that are pivotal in BC. We utilized hdWGCNA to discover co-expressed gene modules and clarify their functions in BC. Employing 7 as the ideal soft threshold, we developed a scale-free co-expression network and delineated five gene co-expression modules (Fig. 4A, 4B). The top ten significant feature genes derived from the five modules were presented (Fig. 4C). The distinctive genes' kME in the brown module exhibits a high score, indicating that its co-expression network possesses robust connections and a stable architecture (Fig. 4D, 4E). Subsequently, we conducted a more detailed analysis of the associations among the modules (Fig. 4F). Figure 4G illustrates the interaction network of distinctive genes across five modules. Figure 4H illustrates the interaction network of distinctive genes inside the brown module. We employed hdWGCNA to thoroughly examine BC co-expressed gene modules in the BC subtype ASC, utilizing a soft threshold power of 3 to construct a scale-free co-expression network, which resulted in the identification of blue and turquoise modules **(Figure S4A, 4B)**. We additionally presented the ten most distinctive genes within these two modules **(Figure S4C)**. Both modules had significant activation in ASC and displayed robust associations **(Figure S4D-4F)**. Simultaneously, we developed interaction networks for the distinctive genes across both modules and independently within each module **(Figure S4G-4M)**. To identify the essential genes driving B cell tumor progression, we conducted an intersection analysis between the distinctive genes from the brown module and the previously found set of differentially expressed B-cell genes, yielding 222 core genes for

further investigation (Fig. 4I). In summary, we identified a distinct brown module in BC and integrated it with DEGs to select critical genes for further examination.

Constructing a scoring model based on a combination of 101 machine learning algorithms

To develop an OC risk model, we analyzed 222 pivotal hub genes employing machine learning. In the training dataset GSE26193 and the validation datasets GSE14764 and GSE63885, 101 predictive models were developed, and their C-index were computed. Of the 101 models, the StepCox[forward]+plsRcox method attained the greatest average C-index during the final model development phase, consisting of five fundamental genes. The principal hub gene was *KIAA0408*, attributed to its elevated coefficient of 0.925. Five genes were designated as prognostic indicators. The formula for the risk model pertaining to OC patients is: RiskScore = 0.925 × Expression_*KIAA0408* + 0.428 × Expression_*CEACAM8* + 0.153 × Expression_*CAPZA1–0*.178 × Expression_*PTCH1–0*.458 × Expression_*VPS13A* (Fig. 5A). Simultaneously, we evaluated this model's predictive capability against ten previously documented prognostic feature sets across three cohorts. Our model demonstrated enhanced prognostic efficacy, exhibiting HR exceeding 1 (GSE26193 and GSE63885: $p < 0.01$) alongside elevated C-index and AUC values (Fig. 5B **and Figure S5A, 5B)**. We subsequently categorized OC samples as high-risk group (HRG) or low-risk group (LRG) according to the median risk score of -0.7196 in the training set GSE26193. Next, in the GSE26193 training set, we generated KM curves indicating diminished HRG survival probabilities. This was corroborated in validation cohorts GSE14764 and GSE63885 (Fig. 5C, 5D **and Figure S5C)**. Ultimately, ROC analysis across three cohorts and a meta-analysis of univariate Cox regression indicated that the model forecasted 3- and 5-year survival rates. A pooled HR of 2.04 demonstrated robust predictive performance from the fixed-effects model (Fig. 5E, 5F). In summary, our five-gene prognostic signature, developed by merging 101 machine learning algorithms, exhibited strong predictive efficacy in both the training set and two validation cohorts.

Model Validation and Development of a Prognostic Assessment Tool for OC

To enhance the validation of model robustness, risk score analysis of the GSE26193 training set indicated that elevated risk scores were significantly correlated with adverse prognosis. *KIAA0408*, *CEACAM8*, and *CAPZA1* were considerably elevated in the HRG, with *KIAA0408* demonstrating the most substantial increase in expression (Fig. 6A). Concurrently, the risk score plots for the validation datasets GSE14764 and GSE63885 produced consistent outcomes, assessing the model's resilience **(Figure S5E, 5F)**. We subsequently computed each patient's survival probability utilizing risk scores and a mixture of five predictive biomarker coefficients through a nomogram. The overall score exhibited an inverse connection with patient survival rates (Fig. 6B). Moreover, the calibration curve analysis conducted on both the training set GSE26193 and the validation set GSE14764 revealed a strong concordance between predicted and observed probabilities, signifying the model's exceptional accuracy (Fig. 6C, 6G). Subsequent decision curve research further validated that, in contrast to the extreme strategy of universally classifying all patients as high-risk or low-risk, our risk scoring model provides a greater net benefit for clinical decision-making (Fig. 6D, 6H). To assess the efficacy of the predictive model, we

developed a time-varying ROC curve for both cohorts. The area under the curve (AUC) values were 0.778 for the training set GSE26193 and 0.780 for the validation set GSE14764 at 5 years, demonstrating commendable model accuracy **(**Fig. 6E, 6I**)**. We subsequently utilized the Schoenfeld residual test to assess the assumptions of proportional risks. No substantial violations were identified for any variable (all p-values > 0.05), confirming the model's dependability **(**Fig. 6F, 6J**)**. We subsequently combined the risk score with clinical-pathological characteristics to develop a nomogram-based prognostic model, which quantifies the expected survival probability at 1-year, 3-year, and 5-year intervals for clinical application **(Figure S5D)**. Heatmaps 6K and 6L comprehensively connected OC risk stratification, clinical-pathological characteristics, and survival outcomes, visually demonstrating the relationship between risk stratification and patient prognosis. The findings indicate that the risk model effectively differentiates prognosis among various risk categories.

Gene set enrichment analysis identifies key pathways

The principal potential oncogene *KIAA0408*, found via machine learning methods, captured our interest because of its remarkably high model coefficients and absence of previous documentation. Consequently, we evaluated its functional importance by synthesizing single-cell and batch transcriptomic data. The Monocle3 pseudo time study of BC effectively delineated a dynamic differentiation trajectory from resting to active states. The UMAP projection distinctly illustrated varied state distributions along the pseudo-time axis **(Figure S6A)**. At the global expression level, the high-weight risk gene *KIAA0408* and *CEACAM8* exhibited consistently low baseline expression throughout the trajectory, perhaps serving as the transcriptional foundation for enduring immunological tolerance in the TME. Conversely, protective genes (e.g., *VPS13A*) demonstrated elevated expression during the mid-phase of the pseudo-time trajectory, indicating their function in preserving immunological homeostasis at pivotal moments of cellular state transitions **(Figure S6B-D)**. To thoroughly evaluate the functional pathways of *KIAA0408*, GSEA analyses were performed on the newly incorporated cohort GSE54388 as well as the previously examined cohorts GSE14764, GSE26193, and GSE63885. In cohort GSE54388, *KIAA0408*-low samples exhibited considerable enrichment in the REACTOME_SIGNALING_BY_THE_B_CELL_RECEPTOR_BCR pathway, along with BCR signaling and cell cycle regulation pathways **(**Fig. 7A-7C**)**. In cohort GSE14764, samples with low expression of *KIAA0408* were highly enriched in pathways including WP_PI3KAKT_SIGNALING and REACTOME_PI3K_AKT_SIGNALING_IN_CANCER **(**Fig. 7B **and Figure S7B)**. In GSE26193 and GSE63885, *KIAA0408*-low samples enriched in pathways such as REACTOME_SIGNALING_BY_THE_B_CELL_RECEPTOR_BCR, thereby reinforcing the robustness of the bidirectional regulation pattern **(Figure S7A, C)**. To gain deeper insight into the mechanism suggested by GSEA, we conducted functional assays on isolated BC. Flow cytometry demonstrated that the overexpression of *KIAA0408* (OE group) markedly diminished the expression of membrane-bound CD79a on BC **(**Fig. 7D**)**. *KIAA0408* overexpression (OE group) inhibited the activation of the BCR-downstream PI3K/AKT signaling pathway. The OE group demonstrated markedly reduced p-AKT levels at 120 minutes post-αIgM stimulation in comparison to the empty vector control group **(**Fig. 7E**)**. To determine the impact of *KIAA0408* on B-cell cycle progression, the cell cycle distribution of the Vector, *KIAA0408*

overexpression (OE), and OE+SC79 (AKT activator) groups was examined. OE cells exhibited a greater proportion of G1 cells and a reduced proportion of S cells compared to Vector cells. The SC79 therapy reversed this phenotype, leading to a reduction in G1 cells and an elevation in S cells. **(Figure. 7F)**. Additionally, we investigated the effects of *KIAA0408*-OE BC on OC. The plate clonogenic assay demonstrated that the proliferation of SKOV3 cell colonies markedly increased in the OE-B-CM group, indicating that BC OE *KIAA0408* facilitated tumour growth **(**Fig. 7G**)**. In conclusion, *KIAA0408* acts as an oncogene in OC by suppressing BCR/PI3K/AKT signaling in BC, leading to immune evasion and tumor progression.

Correlation Analysis Between Model Scores and Immune Infiltration

In order to evaluate the relationship between immune cell characteristics and risk scores, we implemented the single-sample gene set enrichment analysis (ssGSEA) algorithm to evaluate immune cell enrichment within the TME. Heatmap 8A illustrates the distribution of various immune cells among samples and their correlation with model scores. The correlation analysis indicated that model scores exhibited a negative correlation with NK cells, Tcm cells, and Tem cells, but they had a positive correlation with macrophages, neutrophils, and BC **(**Fig. 8B, C**)**. Furthermore, we examined the makeup and proportions of immune cells in high- and low-risk groups, elucidating the correlation between risk stratification and the tumor immune microenvironment **(Figure S7D)**. We subsequently analyzed immune cell infiltration in the samples, demonstrating notable disparities in the infiltration levels of BC, TC, and macrophages between high- and low-risk groups **(**Fig. 8D**)**. The examination of the ESTIMATE algorithm revealed favorable associations between the Immune Score, ESTIMATE Score, and Stromal Score with the risk score, indicating higher scores in the high-risk group. Nonetheless, this heightened abundance was positively associated with the group's unfavorable prognosis **(**Fig. 9A-C**)**. Our comprehensive examination of the immune microenvironment revealed that the unfavorable prognosis and possible resistance to immunotherapy in the high-risk OC cohort may be ascribed to its "highly infiltrated, low-functioning" immunosuppressive condition.

Drug Sensitivity Analysis of Model Gene Expression in OC

To optimize the utilization of traditional Chinese medicine pharmacological resources for identifying treatment options for OC, we initially performed an analysis utilizing the CMAP database. Concurrently, we conducted screening for six major genes in the PI3K/AKT pathway (PIK3CA, PDPK1, AKT1, AKT2, MTOR, RPS6KB1) by amalgamating the HERB and TCMbank databases. Through the intersection, we ultimately found three compounds: ISL, curcumin, and arbutin **(**Fig. 9D**)**. The visualization of the three evaluated chemicals indicated that ISL demonstrated the lowest enrichment score **(**Fig. 9E, F**)**. Molecular docking between ISL and the *KIAA0408* protein resulted in a binding energy of -6.54 kcal/mol, signifying a substantial affinity **(**Fig. 9G**)**. To further investigate the pharmacological role of ISL, we conducted a set of cell function assays. The CCK-8 cell viability experiment indicated that ISL directly suppresses the proliferation of SKOV3 cells in a time- and concentration-dependent fashion **(**Fig. 9H**)**. The plate-based colony formation experiment validated that ISL significantly inhibited the clonogenic development of

SKOV3 cells stimulated by OE-B-CM, indicating its potential therapeutic significance in targeting this regulatory pathway (Fig. 9I). Collectively, ISL is identified as a promising therapeutic agent for OC.

Discussion

OC remains a formidable therapeutic challenge due to its molecular heterogeneity and chemoresistance[22]. While immunotherapy has traditionally focused on TC, TIL-Bs are increasingly recognized as indispensable mediators of anti-tumor immunity in HGSOC[12, 23–26]. Leveraging scRNA-seq, this study seeks to deconstruct B-cell heterogeneity to elucidate novel prognostic mechanisms and identify molecular targets for personalized patient stratification.

While scRNA-seq provides a panoramic view of the HGSOC cellular landscape, the pronounced phenotypic remodeling of

TIL-Bs is particularly striking. These BC exhibit a highly mature, terminally differentiated state. Notably, the intensity of cell-cell communication within B-cell subsets far exceeds the pan-cellular average, with ligand-receptor analysis highlighting that this surging crosstalk is primarily concentrated between BC and TC. Given that B-cell exhaustion is known to drive T-cell dysfunction[23, 27], this may suggest that mature BC act as pivotal immunoregulatory hubs, directly dictating tumor progression by modulating T-cell recruitment and function.

By integrating hdWGCNA modules with B-cell DEGs, we identified core molecular networks influencing prognosis. Leveraging over a hundred machine learning combinations, we constructed a robust 5-gene signature (*KIAA0408*, *CEACAM8*, *CAPZA1*, *PTCH1*, and *VPS13A*) centered on *KIAA0408*. Multi-dimensional validation, including ROC, calibration curves, and DCA, confirmed the model's accuracy across independent cohorts. Furthermore, the signature demonstrated strong predictive performance in other ovarian malignancies, such as endometrioid and clear cell carcinomas (GSE63885), highlighting its potential as a broad-spectrum prognostic tool.

Assessment of the immune landscape revealed a "quantity-over-quality" paradox in high-risk patients: despite elevated immune scores, there was no corresponding survival benefit. ESTIMATE analysis showed that the high-risk microenvironment, while characterized by robust immune recruitment, is skewed toward myeloid infiltration—marked by increased macrophages and neutrophils[28–31], which typically exert immunosuppressive functions—and lacks immunological memory, as evidenced by reduced NK cells and key memory T-cell subsets (Tcm and Tem)[32, 33]. This suggests that *KIAA0408*-driven B-cell dysfunction disrupts the essential B-T crosstalk required for memory formation, creating a niche replete with myeloid suppressors but devoid of long-term surveillance[10, 27, 34]. This "ineffective inflammation" likely underlies the poor immunotherapy response and dismal prognosis in high-risk patients[35].

KIAA0408 is a poorly characterized protein-coding gene whose aberrant expression correlates closely with the prognosis of malignancies, while its exact mechanism driving tumor progression remains

unclarified[36]. The PI3K/AKT signaling axis serves as the core engine for BCR signal transduction, B-cell activation and proliferation[37, 38]. To deconstruct how *KIAA0408* influences the OC microenvironment, we integrated GSEA with in vitro functional assays, identifying a novel *KIAA0408*/BCR/PI3K-AKT inhibitory axis. Mechanistically, *KIAA0408* overexpression significantly downregulates surface mIgM expression and diminishes BCR complex abundance. This receptor depletion results in markedly reduced AKT phosphorylation upon αIgM stimulation, ultimately triggering G1-phase arrest—a phenotype successfully reversed by the AKT activator SC79. Furthermore, in a T-cell-independent co-culture system, conditioned medium from OE-B-CM significantly enhanced tumor cell viability. This dual mechanism—B-cell immune dysfunction and direct paracrine tumor nourishment—likely underpins the poor prognosis of high-risk patients[39, 40].

Accumulating studies have demonstrated that ubiquitination serves as a key post-translational modification for BCR homeostasis and signal transduction[41–43]. Multiple E3 ubiquitin ligases target the BCR complex to restrain its overactivation and inhibit the downstream PI3K/AKT pathway[44, 45], thus maintaining B-cell immune homeostasis[46]. Based on these regulatory mechanisms, we hypothesized that *KIAA0408* may facilitate BCR complex ubiquitination and subsequent degradation, thereby attenuating downstream signaling.

To identify therapeutic candidates, we integrated cMAP transcriptional reversal analysis with TCM-specific screening, identifying ISL—a natural chalcone derived from licorice—as the lead candidate. Previously documented to possess definite anti-tumor activity and PI3K/AKT pathway-modulating capacity[47–50], our functional assays further validated that ISL significantly suppressed OC cell viability. Importantly, ISL neutralized the pro-tumorigenic effects of OE-B-CM, effectively reversing the survival advantage conferred by the dysfunctional B-cell niche. As a known PI3K/AKT modulator, ISL offers a mechanism-anchored strategy for precision treatment in OC.

Despite rigorous bioinformatic and experimental analyses, this study has several limitations. First, all mechanistic exploration and drug screening were conducted in vitro, which cannot fully recapitulate the complex TME. The anti-tumor effect of ISL also requires further validation in animal models. Second, we proposed the hypothesis that *KIAA0408* promotes BCR ubiquitination and degradation, yet this regulatory mechanism lacks direct molecular verification. Third, although the *KIAA0408*/BCR/PI3K-AKT intracellular axis was identified, the extracellular mediators driving B cell paracrine pro-tumor effects remain unclear. This work provides novel insights, and future studies will further elucidate the underlying mechanisms of this regulatory network.

Conclusion

This study identified a *KIAA0408*-driven B-cell subset in OC via scRNA-seq. Mechanistically, *KIAA0408* represses the BCR/PI3K-AKT axis to induce G1-phase arrest and a tumor-promoting paracrine phenotype. By integrating hdWGCNA and machine learning, we constructed a robust prognostic model that characterizes the high-risk immune landscape. Furthermore, ISL was validated as a therapeutic

agent capable of neutralizing *KIAA0408*-mediated pro-tumorigenic effects. Collectively, our findings elucidate the dual role of BC in the TME, providing novel biomarkers and mechanism-anchored therapeutic avenues for personalized OC care.

Declarations

Ethics Approval and Informed Consent

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Second Affiliated Hospital of Nanchang University (Ethics approval number: IIT- 2026-214). Written informed consent was obtained from all patients prior to sample collection. All animal experiments were approved by the Animal Ethics Committee of Second Affiliated Hospital of Nanchang University and were performed in accordance with the Guide for the Care and Use of Laboratory Animals.

Conflict of Interest

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

Disclosure of interest

The authors report there are no competing interests to declare.

Funding

This research was funded by by Jiangxi Provincial Natural Science Foundation - Youth Fund (Grant No. 2024YJ123).

Acknowledgments

we are grateful to all research scientists and patients who participated in the GEO and TCGA cohorts.

Data Availability Statement

The datasets supporting the conclusions of this article are available in the TCGA repository (https://portal.gdc.cancer.gov) and GEO repository (https://www.ncbi.nlm.nih.gov/geo/).

Author Contribution

X.Q. conceived the study. H.W. collected the datasets. X.Q., H.W., and C.F. performed the bioinformatics analyses and drafted the main manuscript. Y.Z. and M.B. prepared the figures and figure layout, and collected clinical samples. R.Z. and Y.C. performed statistical analyses. S.T. managed the data. Z.C.,

K.W., and Z.L. procured experimental materials. X.W., R.S., and D.H. conducted the experiments. Y.Z. and H.W. supervised the study and acquired funding. All authors reviewed the manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. 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.
2. Seidman JD, Horkayne-Szakaly I, Haiba M, et al. The histologic type and stage distribution of ovarian carcinomas of surface epithelial origin. Int J Gynecol Pathol. 2004;23(1):41–4.
3. Narod S. Can advanced-stage ovarian cancer be cured? Nat Rev Clin Oncol. 2016;13(4):255–61.
4. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.
5. Andersen CL, Sikora MJ, Boisen MM, et al. Active Estrogen Receptor-alpha Signaling in Ovarian Cancer Models and Clinical Specimens. Clin Cancer Res. 2017;23(14):3802–12.
6. Ghisoni E, Imbimbo M, Zimmermann S et al. Ovarian Cancer Immunotherapy: Turning up the Heat. Int J Mol Sci 2019;20(12).
7. Brudno JN, Maus MV, Hinrichs CS. CAR T Cells and T-Cell Therapies for Cancer: A Translational Science Review. JAMA. 2024;332(22):1924–35.
8. Ning F, Cole CB, Annunziata CM. Driving Immune Responses in the Ovarian Tumor Microenvironment. Front Oncol. 2020;10:604084.
9. Nelson MA, Ngamcherdtrakul W, Luoh SW, et al. Prognostic and therapeutic role of tumor-infiltrating lymphocyte subtypes in breast cancer. Cancer Metastasis Rev. 2021;40(2):519–36.
10. Bruno TC, Ebner PJ, Moore BL, et al. Antigen-Presenting Intratumoral B Cells Affect CD4(+) TIL Phenotypes in Non-Small Cell Lung Cancer Patients. Cancer Immunol Res. 2017;5(10):898–907.
11. Biswas S, Mandal G, Payne KK, et al. IgA transcytosis and antigen recognition govern ovarian cancer immunity. Nature. 2021;591(7850):464–70.
12. Montfort A, Pearce O, Maniati E, et al. A Strong B-cell Response Is Part of the Immune Landscape in Human High-Grade Serous Ovarian Metastases. Clin Cancer Res. 2017;23(1):250–62.
13. Izar B, Tirosh I, Stover EH, et al. A single-cell landscape of high-grade serous ovarian cancer. Nat Med. 2020;26(8):1271–9.
14. Shih AJ, Menzin A, Whyte J, et al. Identification of grade and origin specific cell populations in serous epithelial ovarian cancer by single cell RNA-seq. PLoS ONE. 2018;13(11):e0206785.
15. Wang Y, Xie H, Chang X, et al. Single-Cell Dissection of the Multiomic Landscape of High-Grade Serous Ovarian Cancer. Cancer Res. 2022;82(21):3903–16.
16. Xu J, Fang Y, Chen K, et al. Single-Cell RNA Sequencing Reveals the Tissue Architecture in Human High-Grade Serous Ovarian Cancer. Clin Cancer Res. 2022;28(16):3590–602.
17. Huang Y, Gui Z, Wu M, et al. Tumor-infiltrating B cell-related lncRNA crosstalk reveals clinical outcomes and tumor immune microenvironment in ovarian cancer based on single-cell and bulk

RNA-sequencing. Heliyon. 2024;10(21):e39496.
18. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437–e14521417.
19. Lv Q, Chen G, He H, et al. TCMBank: bridges between the largest herbal medicines, chemical ingredients, target proteins, and associated diseases with intelligence text mining. Chem Sci. 2023;14(39):10684–701.
20. Gao K, Liu L, Lei S, et al. HERB 2.0: an updated database integrating clinical and experimental evidence for traditional Chinese medicine. Nucleic Acids Res. 2025;53(D1):D1404–14.
21. Yang X, Liu Y, Gan J et al. FitDock: protein-ligand docking by template fitting. Brief Bioinform 2022;23(3).
22. Veneziani AC, Gonzalez-Ochoa E, Alqaisi H, et al. Heterogeneity and treatment landscape of ovarian carcinoma. Nat Rev Clin Oncol. 2023;20(12):820–42.
23. Kasikova L, Rakova J, Hensler M, et al. Tertiary lymphoid structures and B cells determine clinically relevant T cell phenotypes in ovarian cancer. Nat Commun. 2024;15(1):2528.
24. Nelson BH, Hamilton P, Phung MT et al. Immunological and molecular features of the tumor microenvironment of long-term survivors of ovarian cancer. J Clin Invest. 2024;134(24).
25. Wouters MCA, Nelson BH. Prognostic Significance of Tumor-Infiltrating B Cells and Plasma Cells in Human Cancer. Clin Cancer Res. 2018;24(24):6125–35.
26. Kroeger DR, Milne K, Nelson BH. Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer. Clin Cancer Res. 2016;22(12):3005–15.
27. Cabrita R, Lauss M, Sanna A, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature. 2020;577(7791):561–5.
28. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. Nat Rev Immunol. 2019;19(6):369–82.
29. Chen JJ, Lin YC, Yao PL, et al. Tumor-associated macrophages: the double-edged sword in cancer progression. J Clin Oncol. 2005;23(5):953–64.
30. Jaillon S, Ponzetta A, Di Mitri D, et al. Neutrophil diversity and plasticity in tumour progression and therapy. Nat Rev Cancer. 2020;20(9):485–503.
31. Coffelt SB, Wellenstein MD, de Visser KE. Neutrophils in cancer: neutral no more. Nat Rev Cancer. 2016;16(7):431–46.
32. Fridman WH, Pagès F, Sautès-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12(4):298–306.
33. Sato E, Olson SH, Ahn J, et al. Intraepithelial CD8 + tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. Proc Natl Acad Sci U S A. 2005;102(51):18538–43.

34. Petitprez F, de Reyniès A, Keung EZ, et al. B cells are associated with survival and immunotherapy response in sarcoma. Nature. 2020;577(7791):556–60.
35. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21(8):938–45.
36. Han Y, Li Y. Comprehensive Exploration of M2 Macrophages and Its Related Genes for Predicting Clinical Outcomes and Drug Sensitivity in Lung Squamous Cell Carcinoma. J Oncol. 2022;2022:1163924.
37. Srinivasan L, Sasaki Y, Calado DP, et al. PI3 kinase signals BCR-dependent mature B cell survival. Cell. 2009;139(3):573–86.
38. Fruman DA, Chiu H, Hopkins BD, et al. The PI3K Pathway in Human Disease. Cell. 2017;170(4):605–35.
39. Gunderson AJ, Kaneda MM, Tsujikawa T, et al. Bruton Tyrosine Kinase-Dependent Immune Cell Cross-talk Drives Pancreas Cancer. Cancer Discov. 2016;6(3):270–85.
40. Ammirante M, Luo JL, Grivennikov S, et al. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. Nature. 2010;464(7286):302–5.
41. Choi J, Phelan JD, Wright GW, et al. Regulation of B cell receptor-dependent NF-κB signaling by the tumor suppressor KLHL14. Proc Natl Acad Sci U S A. 2020;117(11):6092–102.
42. Katkere B, Rosa S, Drake JR. The Syk-binding ubiquitin ligase c-Cbl mediates signaling-dependent B cell receptor ubiquitination and B cell receptor-mediated antigen processing and presentation. J Biol Chem. 2012;287(20):16636–44.
43. Zhang M, Veselits M, O'Neill S, et al. Ubiquitinylation of Ig beta dictates the endocytic fate of the B cell antigen receptor. J Immunol. 2007;179(7):4435–43.
44. Drake L, McGovern-Brindisi EM, Drake JR. BCR ubiquitination controls BCR-mediated antigen processing and presentation. Blood. 2006;108(13):4086–93.
45. Sohn HW, Gu H, Pierce SK. Cbl-b negatively regulates B cell antigen receptor signaling in mature B cells through ubiquitination of the tyrosine kinase Syk. J Exp Med. 2003;197(11):1511–24.
46. Drake JR. The immunobiology of ubiquitin-dependent B cell receptor functions. Mol Immunol. 2018;101:146–54.
47. Yu M, Pan Q, Li W, et al. Isoliquiritigenin inhibits gastric cancer growth through suppressing GLUT4 mediated glucose uptake and inducing PDHK1/PGC-1α mediated energy metabolic collapse. Phytomedicine. 2023;121:155045.
48. Zhang Z, Chen WQ, Zhang SQ, et al. Isoliquiritigenin inhibits pancreatic cancer progression through blockade of p38 MAPK-regulated autophagy. Phytomedicine. 2022;106:154406.
49. Luo F, Tang Y, Zheng L et al. Isoliquiritigenin Inhibits the Growth of Colorectal Cancer Cells through the ESR2/PI3K/AKT Signalling Pathway. Pharmaceuticals (Basel). 2023;17(1).
50. Huang Y, Liu C, Zeng WC et al. Isoliquiritigenin inhibits the proliferation, migration and metastasis of Hep3B cells via suppressing cyclin D1 and PI3K/AKT pathway. Biosci Rep. 2020;40(1).

Figures

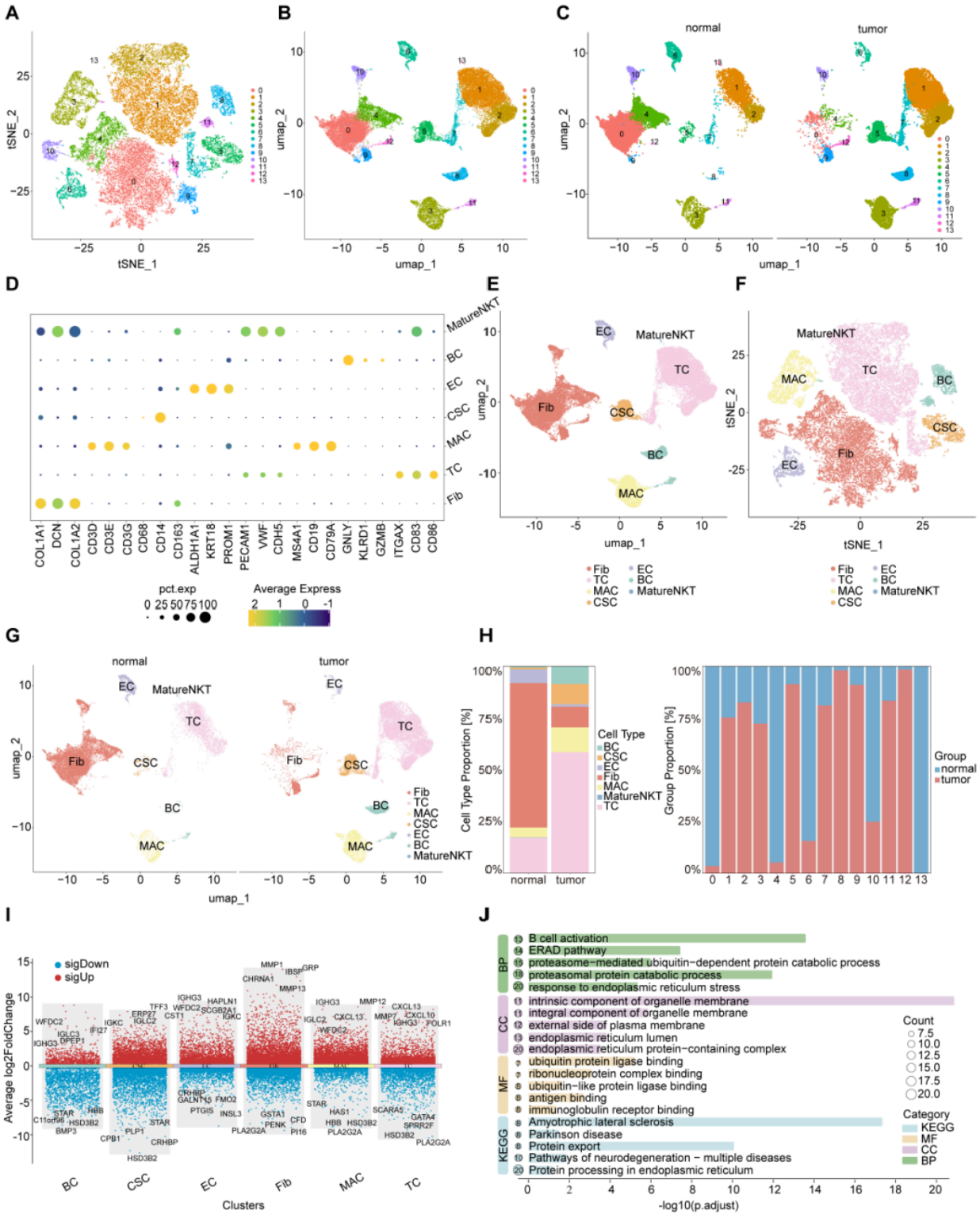


Figure 1

scRNA-seq clustering analysis, differential gene analysis, and enrichment analysis for OC and normal groups in the GSE184880 cohort. (A, B)t-SNE and UMAP visualization of cell cluster distributions after clustering GSE184880 single-cell data. **(C)** UMAP plot comparing cell cluster distributions between the

normal and tumor groups. **(D)** Bubble plot of marker gene expression levels across cell types. **(E, F)** Distribution of GSE184880 single-cell data in t-SNE and UMAP spaces after cell type annotation. **(G)** Comparative display of UMAP distribution differences of various cell types between normal and tumor groups. **(H)** Cell composition proportion plots for normal and tumor groups, with analysis of normal and tumor cell proportions within each cluster. **(I)** DEG analysis for each cell type. **(J)**KEGG and GO enrichment analysis of DEGs in BC.

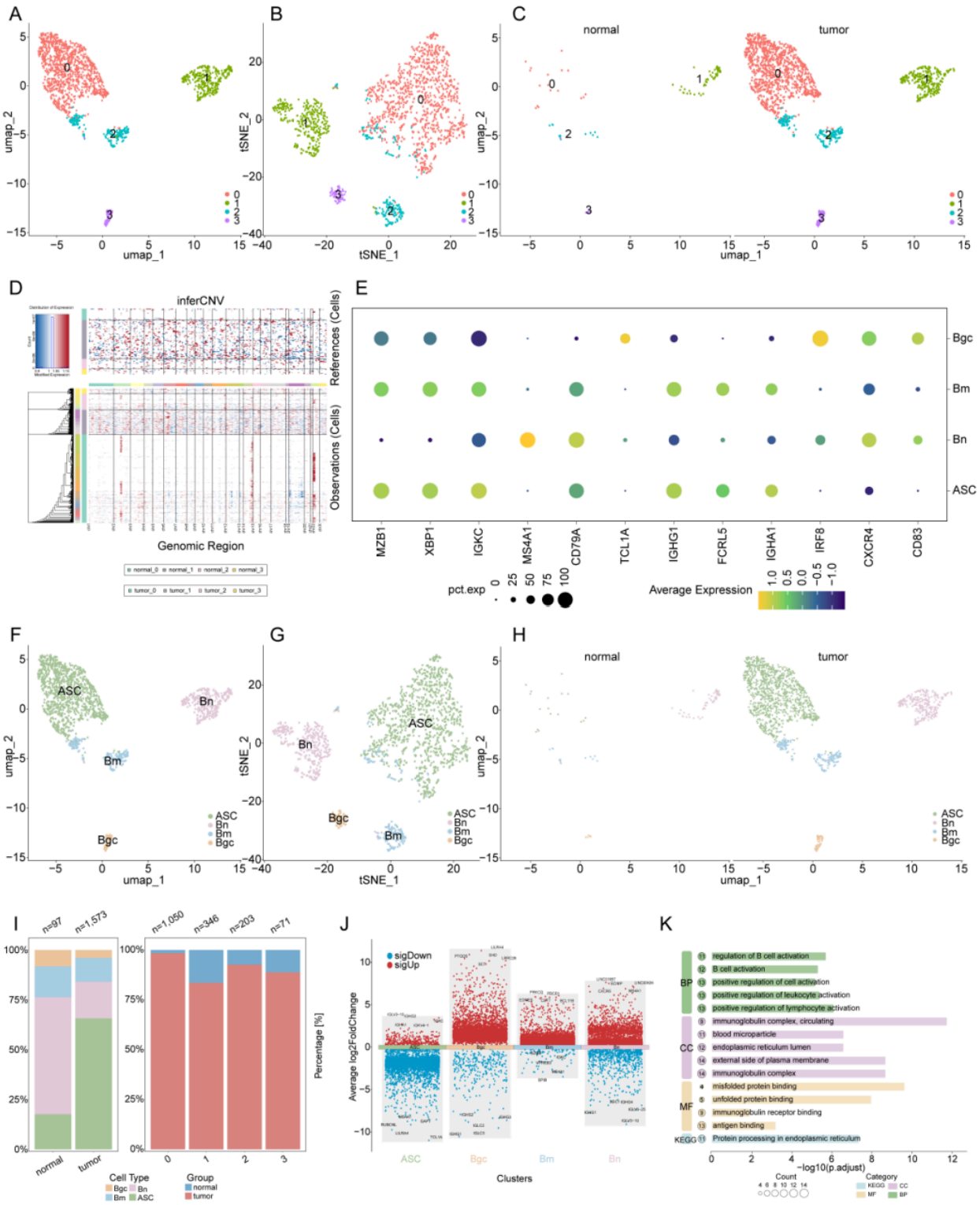


Figure 2

Integrated analysis of B cell subsets subpopulation identification, CNV, and differential expression. (A, B) Dimensionality reduction visualization of B cell subsets using UMAP and t-SNE. **(C)** Comparative UMAP plots showing the distribution of B cell subset clusters between the normal and tumor groups. **(D)** Heatmap of CNV landscapes across B cell subsets in normal and tumor samples. **(E)** Bubble plot of marker gene expression levels in B cell subsets. **(F, G)** Spatial distribution of annotated B cell subsets in t-SNE and UMAP embeddings. **(H)** UMAP plots comparing the distribution of B cell subsets between normal and tumor groups. **(I)** Proportional plots showing the distribution of B cell subsets between normal and tumor groups, with the normal and tumor cell composition quantified for each subset. **(J)**DEG analysis of B cell subpopulations. **(K)** KEGG and GO enrichment analysis of DEGs in ASC.

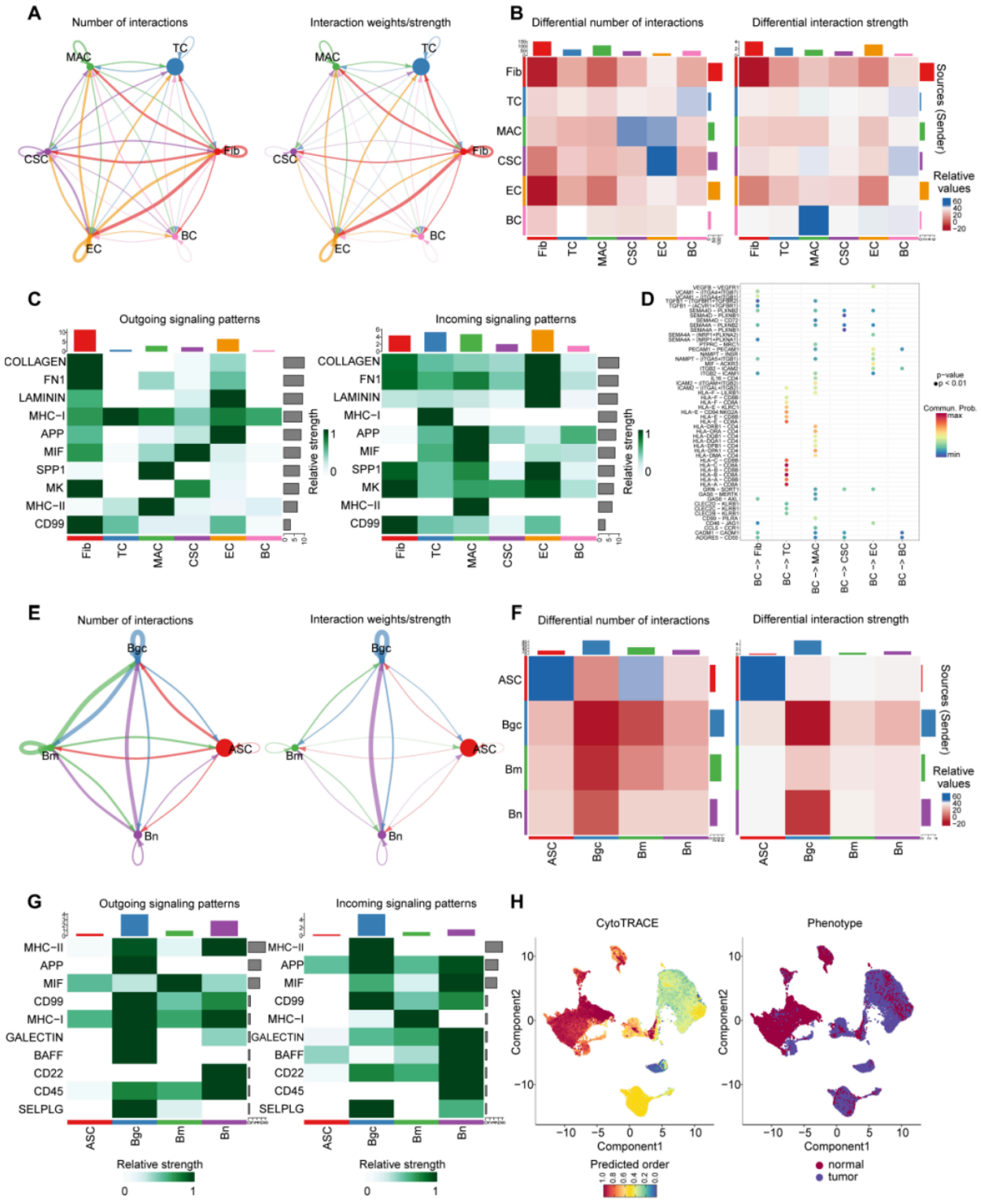


Figure 3

Cellular communication and CytoTRACE cellular pluripotency analysis. (A) Chord diagram of cell interaction numbers and strengths among six cell types. **(B)**Heatmap of cell interaction numbers and strengths among six cell types. **(C, D)** Outgoing and incoming signaling patterns analysis of six cell types. **(E)**Heatmap of ligand-receptor interactions between BC and other cell types. **(F)**Chord diagram of the number and strength of cell interactions among B cell subtypes. **(G)** Heatmap of the number and

strength of cell interactions among B cell subtypes. **(H, I)** Analysis of outgoing and incoming signaling patterns among B cell subtypes. **(G)** Characterization of cellular differentiation levels based on the CytoTRACE algorithm and presentation of phenotypic distributions in normal and tumor cells.

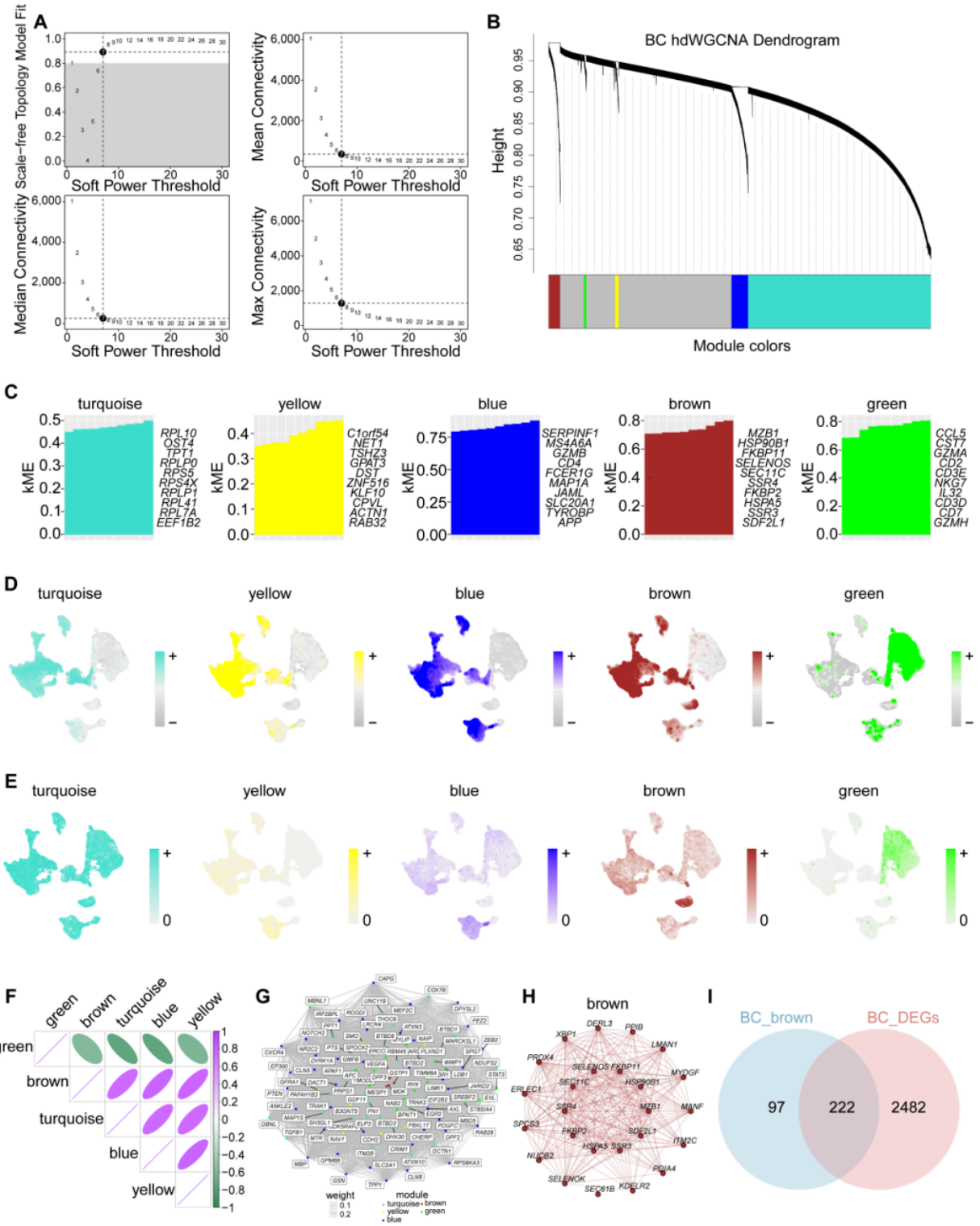


Figure 4

hdWGCNA module analysis and core gene identification in BC. (A) Plots of scale-free topology model fit index and mean connectivity against the soft-thresholding power. **(B)** Dendrogram construction and module assignment from hdWGCNA of BC. **(C)** Connectivity scores of characteristic genes within each gene module. **(D, E)** UMAP visualization of feature gene expression for each co-expression module across single cells in the GSE184880 cohort. **(F)** Correlation matrix between each gene module. **(G)**Interaction network of feature genes derived from gene co-expression modules. **(H)**Interaction network diagram of feature genes in the brown module. **(I)**Overlap of brown module feature genes and B cell DEGs visualized by Venn diagram.

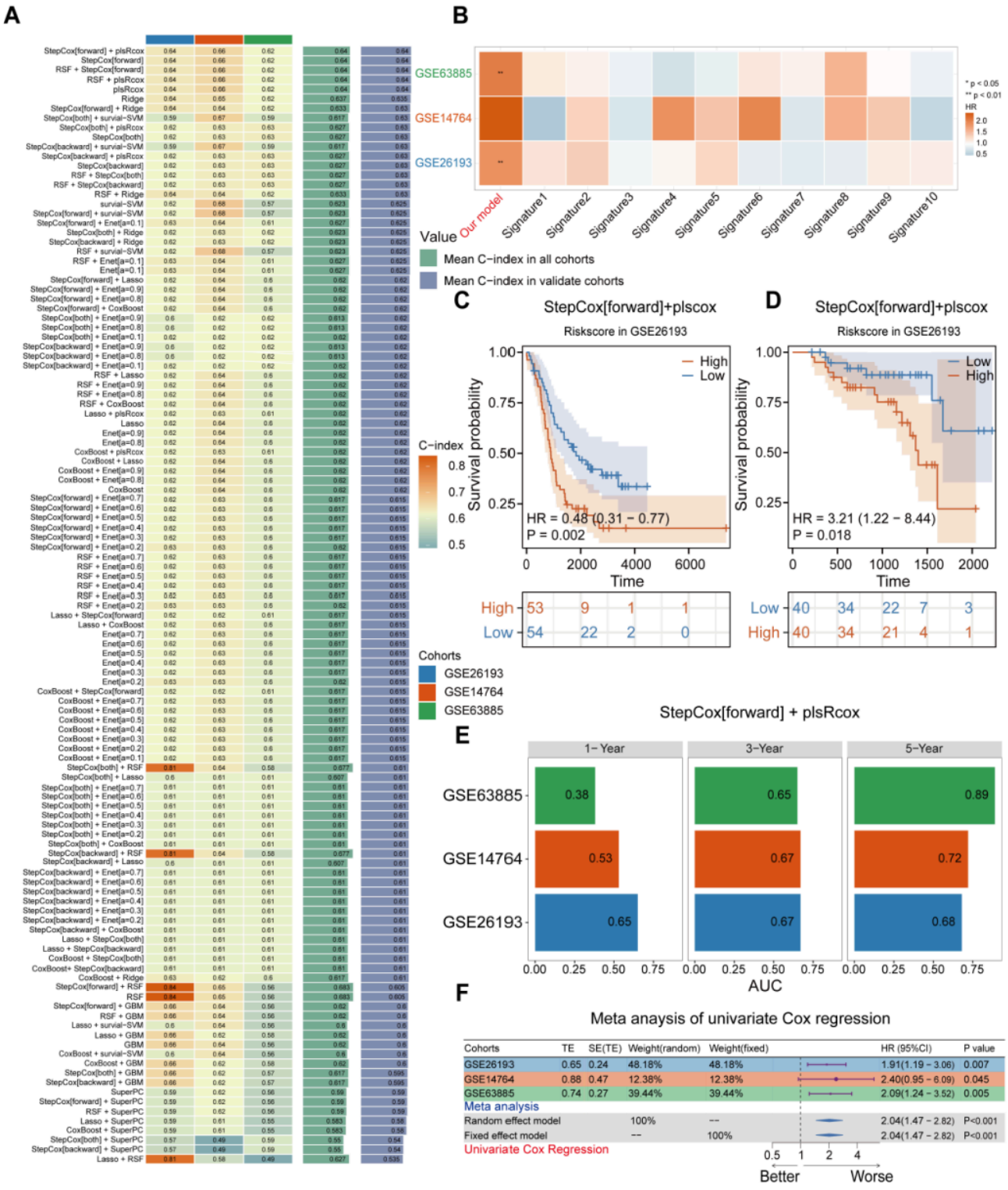


Figure 5

Construction and validation of a scoring model based on combinations of 101 machine learning algorithms. (A) Comparison of predictive performance across different machine learning algorithm combinations in the validation cohort. **(B)** Comparison of prognostic efficacy between this model and previously published prognostic features in an independent validation cohort. **(C, D)** Survival analysis in the GSE26193 and GSE14764 cohort. **(E)**AUC values for 1-, 3-, and 5-year survival prediction in the

validation cohort using the StepCox[forward] + plsRcox model. **(F)** Forest plot of univariate Cox regression meta-analysis for the risk score of the prognostic model. (*p < 0.05, **p < 0.01, ***p < 0.001).

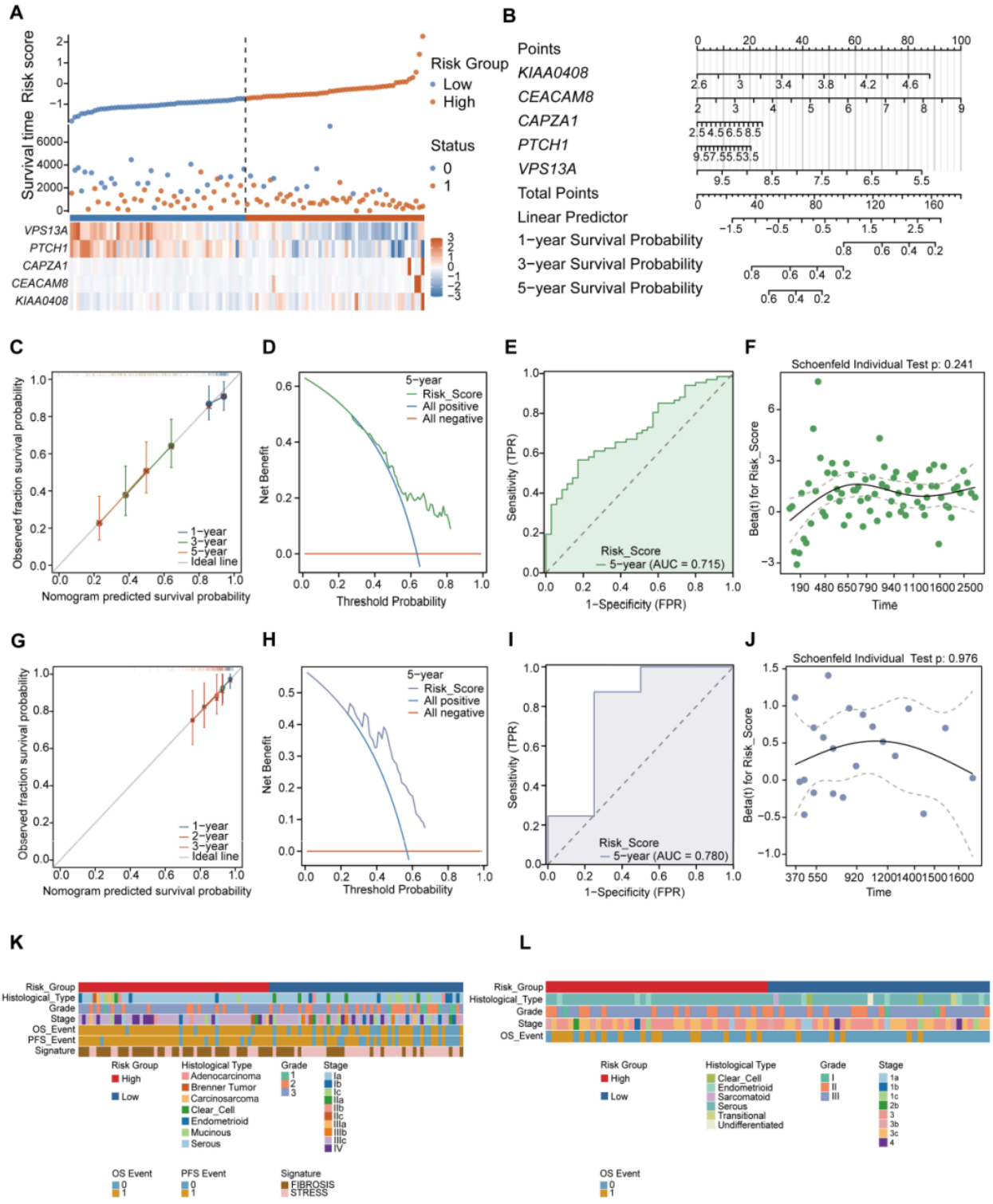


Figure 6

Machine learning model validation and prognostic analysis. (A) Risk score plot of the GSE26193 cohort. **(B)** Prognostic nomogram based on the model genes. **(C, G)** Calibration curve of the GSE26193 and

GSE14764 cohort scatter plot. **(D, H)** DCAfor 5-year survival prediction in the GSE26193 and GSE14764 cohort. **(E, I)**ROC curve for risk score prediction of 5-year survival in the GSE26193 and GSE14764 cohort. **(F, J)** Schoenfeld residual plot for the GSE26193 and GSE14764 cohort. **(K)** Heatmap of clinical characteristics in the GSE26193 and GSE14764 cohort.

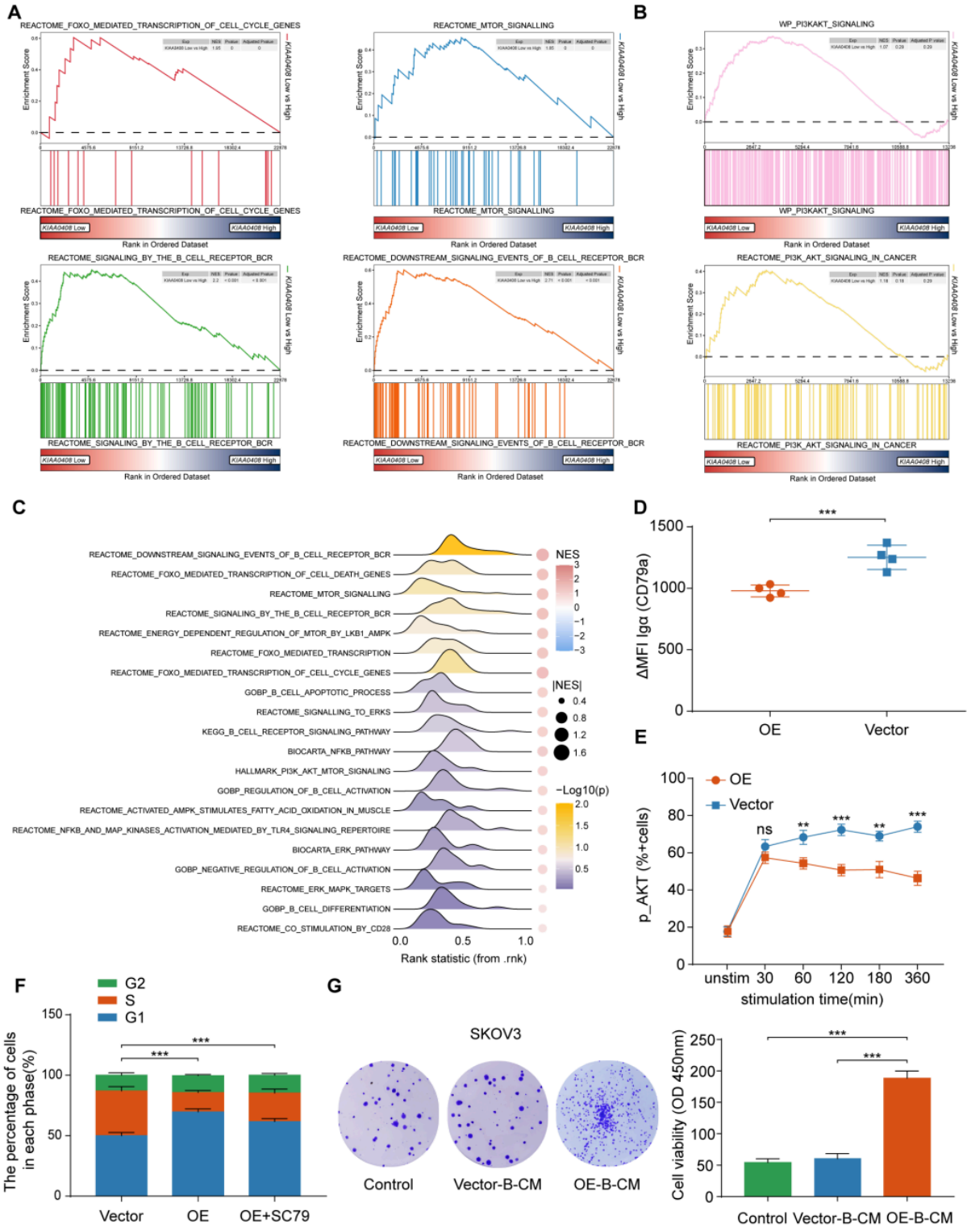


Figure 7

GSEA pathway enrichment analysis based on the *KIAA0408* gene and experimental validation of the *BCR*-PI3K/AKT signaling axis. (A, B) GSEA enrichment analysis of the *KIAA0408* high-expression and low-expression groups in the GSE54388 and GSE14764 cohort. **(C)** Display of significant pathways from *KIAA0408*-related pathway enrichment analysis in the GSE54388 cohort.

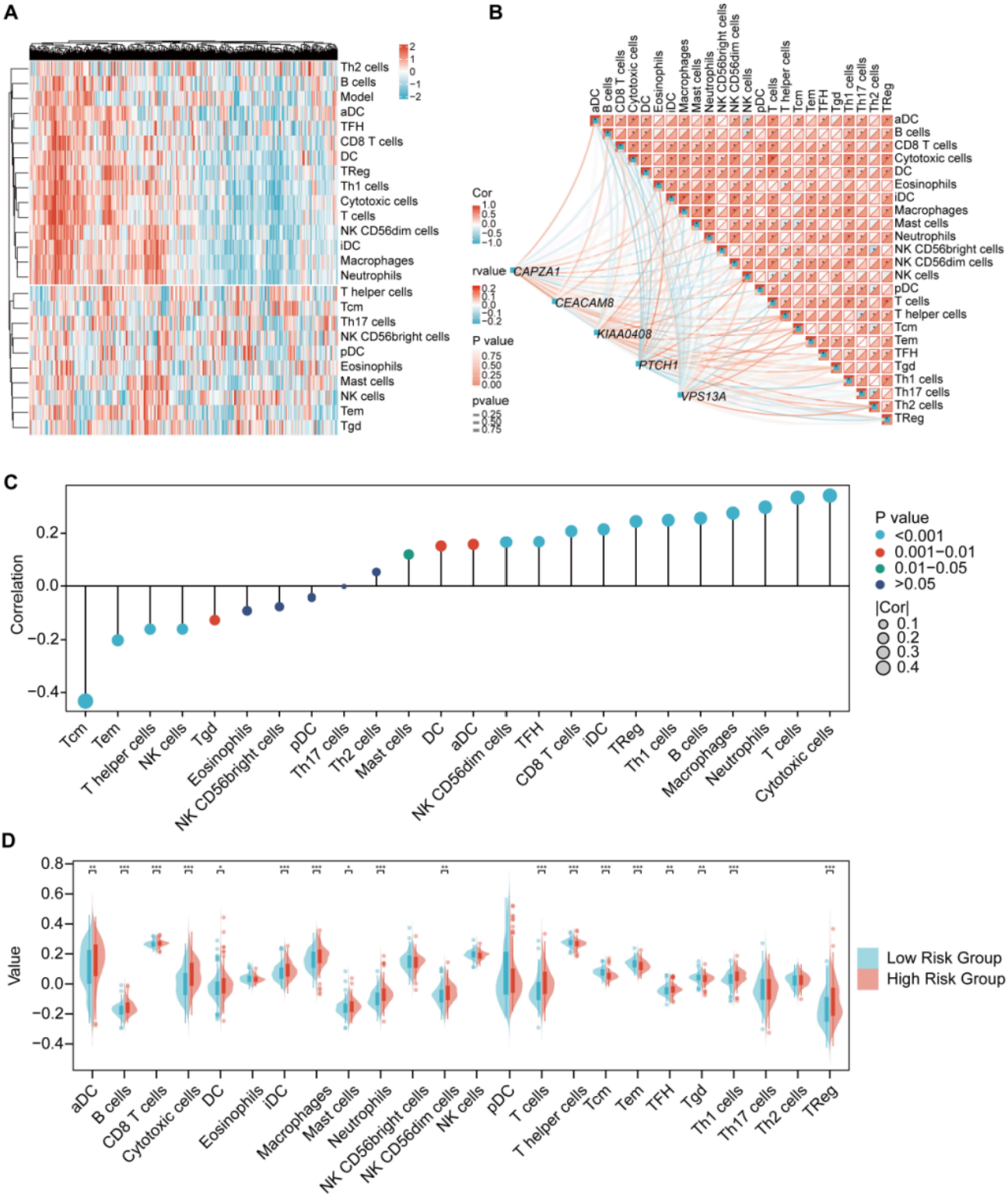


Figure 8

Immune infiltration landscape of the tumor microenvironment across risk groups. (A) Heatmap of immune cell infiltration abundance in the tumor microenvironment. **(B)** Correlation analysis between model genes and immune cell infiltration. **(C)** Lollipop plot showing the correlation between risk scores and immune cell infiltration. **(D)** Violin plot demonstrating differences in immune cell infiltration levels between high- and low-risk groups.

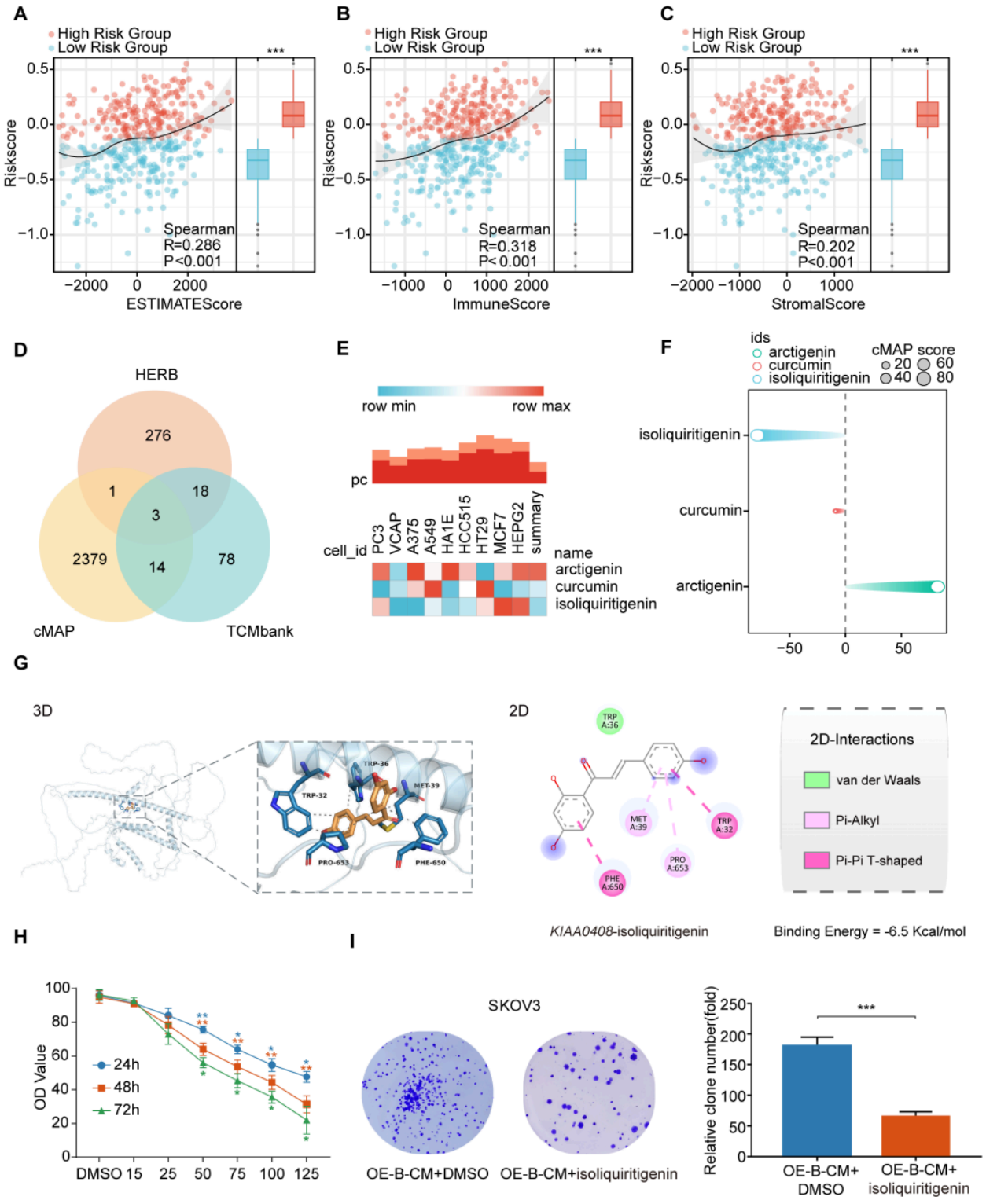


Figure 9

Identifying risk-immunity score correlations and evaluating candidate drug sensitivity. Correlation of risk score with ESTIMATE-derived scores: **(A)** ESTIMATE, **(B)** Immune, and **(C)**Stromal scores. **(D)** Venn diagram of predicted drug intersections. **(E)**Heatmap displaying differential patterns of candidate traditional Chinese medicine small molecules' cMAP scores across different cell lines. **(F)**Comparison of cMAP scores for candidate components. **(G)** Molecular docking results between *KIAA0408* protein and ISL.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx
- floatimage1.png