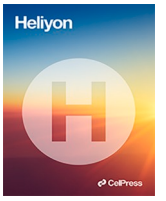
Heliyon 10 (2024) e33367

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on drug response genes to predict prognosis and therapeutic response in ovarian cancer

ZhenWei Zhang, MianMian Chen, XiaoLian Peng *

Jinjiang Municipal Hospital(Shanghai Sixth People's Hospital Fujian Campus), No. 16, Luoshan Section, Jinguang Road, Luoshan Street, Jinjiang City, Quanzhou, Fujian, China

A R T I C L E I N F O

Keywords:
Ovarian cancer
Paclitaxel resistance
scRNAseq

A B S T R A C T

Ovarian cancer represents a severe gynecological malignancy with a dire prognosis, underscoring the imperative need for dependable biomarkers that can accurately predict drug response and guide therapeutic choices. In this study, we harnessed online single-cell RNA sequencing (scRNAseq) and bulk RNA sequencing (RNAseq) datasets, applying the Scissor algorithm to identify cells responsive to paclitaxel. From these cells, we derived a gene signature, subsequently used to construct a prognostic model that demonstrated high sensitivity and specificity in predicting patient outcomes. Moreover, we conducted pathway and functional enrichment analyses to uncover potential molecular mechanisms driving the prognostic gene signature. This study illustrates the critical role of scRNAseq and bulk RNAseq in developing precise prognostic models for ovarian cancer, potentially transforming clinical decision-making.

1. Introduction

Ovarian cancer (OV) ranks as the fifth leading cause of cancer-related mortality among women, with the incidence increasing significantly after the age of 50 [1]. Despite a general five-year survival rate of around 50 %, prognosis dramatically varies depending on the stage at diagnosis [2]. Often termed the 'silent killer,' OV's nondescript symptoms frequently lead to late diagnosis, compounding treatment challenges [3]. Chemotherapy, incorporating agents such as paclitaxel, carboplatin, and doxorubicin, remains the cornerstone of post-surgical treatment for OV [4,5]. However, patient responses to these chemotherapeutic regimens differ markedly due to factors such as cancer stage, tumor cell heterogeneity, and specific drug efficacy [6]. Consequently, there is a critical need to stratify patients into distinct groups based on their chemotherapy responses [7], enabling clinicians to tailor more effective treatment strategies for OV.

Previous studies have leveraged RNAseq to investigate drug responses in OV cells, illuminating the genomic landscape of chemoresistance [8,9]. For instance, one study employed RNAseq to discern a gene set distinctly expressed between drug-sensitive and drug-resistant OV cells [10]. This research revealed that drug-resistant cells exhibit elevated expression of genes related to drug detoxification and cell survival, shedding light on the molecular underpinnings of drug resistance in OV. Another study utilized RNAseq to profile the transcriptomes of OV cell lines, differentiating those sensitive to paclitaxel from those resistant to this commonly

* Corresponding author. Jinjiang Municipal Hospital(Shanghai Sixth People's Hospital Fujian Campus), No. 16, Luoshan Section, Jinguang Road, Luoshan Street, Jinjiang City, Quanzhou, Fujian, China.
E-mail address: pxl94116@163.com (X. Peng).

https://doi.org/10.1016/j.heliyon.2024.e33367
Received 30 March 2024; Received in revised form 19 June 2024; Accepted 20 June 2024
Available online 20 June 2024

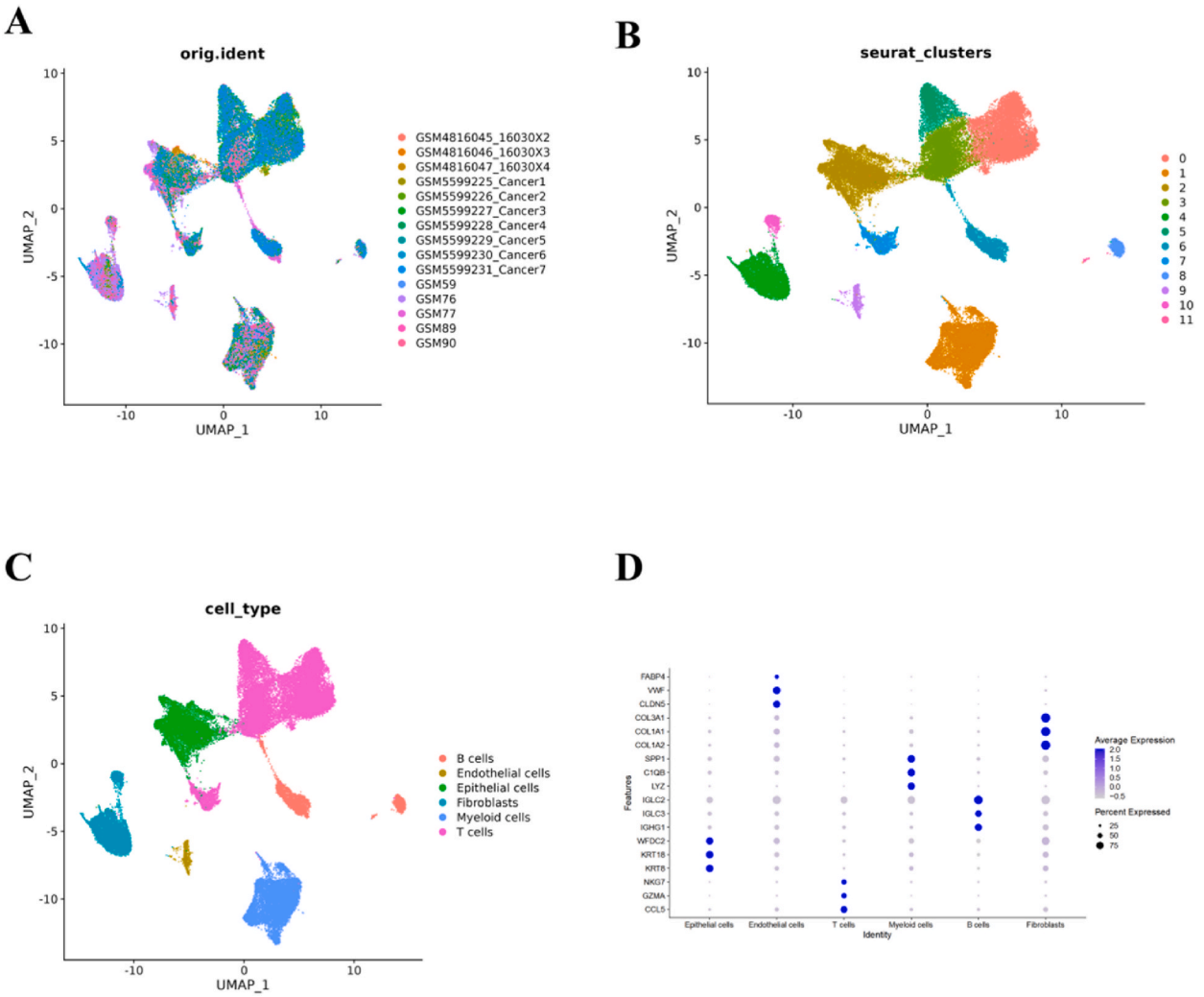


Fig. 1. Characterization of major cell populations of OV patient samples.
(A) UMAP clustering of OV cells integrated from three different scRNAseq datasets including 15 OV samples. Each point indicates one single cell and cells in the same cluster indicates high similarity in transcriptome profile. Cells were color-coded with patient origin. (B) UMAP plot of the integrated dataset depicts 12 transcriptional distinct cell clusters and cells were color-coded with cluster information. (C) UMAP plot of the integrated dataset depicts 6 major cell types and cells were color-coded with cell types. (D) Dot plot of top 3 marker genes upregulated in each cell type.

used chemotherapeutic agent [11]. These investigations underscore the value of RNAseq in elucidating drug responsiveness in OV, offering potential pathways for targeted therapeutic interventions.

Over the past decade, scRNAseq has revolutionized our understanding of cell biology and disease mechanisms [12]. Unlike traditional bulk RNA sequencing, which measures gene expression from a mixed cell population potentially obscuring individual cell differences, scRNAseq assesses gene expression in each individual cell. This approach provides a nuanced view of the cellular diversity and states within tissues. Cancer cells exhibit significant heterogeneity, influenced by their varied tumor microenvironment (TME) and interactions with distinct cell types [13]. This heterogeneity even extends to clonal cancer cells, which can display diverse phenotypic features through clonal evolution. The TME, comprising immune cells, stromal cells, blood vessels, the extracellular matrix, and signaling molecules, plays a crucial role in drug resistance [14]. Drug resistance mechanisms facilitated by the TME include the secretion of signaling molecules such as cytokines and growth factors that activate pathways like PI3K/AKT, promoting cell survival and chemotherapy resistance [15]. Additionally, the TME can physically impede drug delivery to tumor cells through increased interstitial fluid pressure, abnormal blood vessel structures, or extracellular matrix remodeling. Dynamic changes in the TME in response to treatment, such as the induction of growth factors post-chemotherapy, further complicate drug resistance by promoting tumor cell proliferation and survival [16].

To delineate the heterogeneity inherent in OV, recent studies have employed scRNAseq to dissect subpopulations of cancer and stromal cells within the tumor microenvironment (TME), offering novel insights into ovarian cancer progression, metastasis, clonal evolution, and drug resistance [17–23]. Despite its profound impact, single-cell technology remains prohibitively expensive, limiting

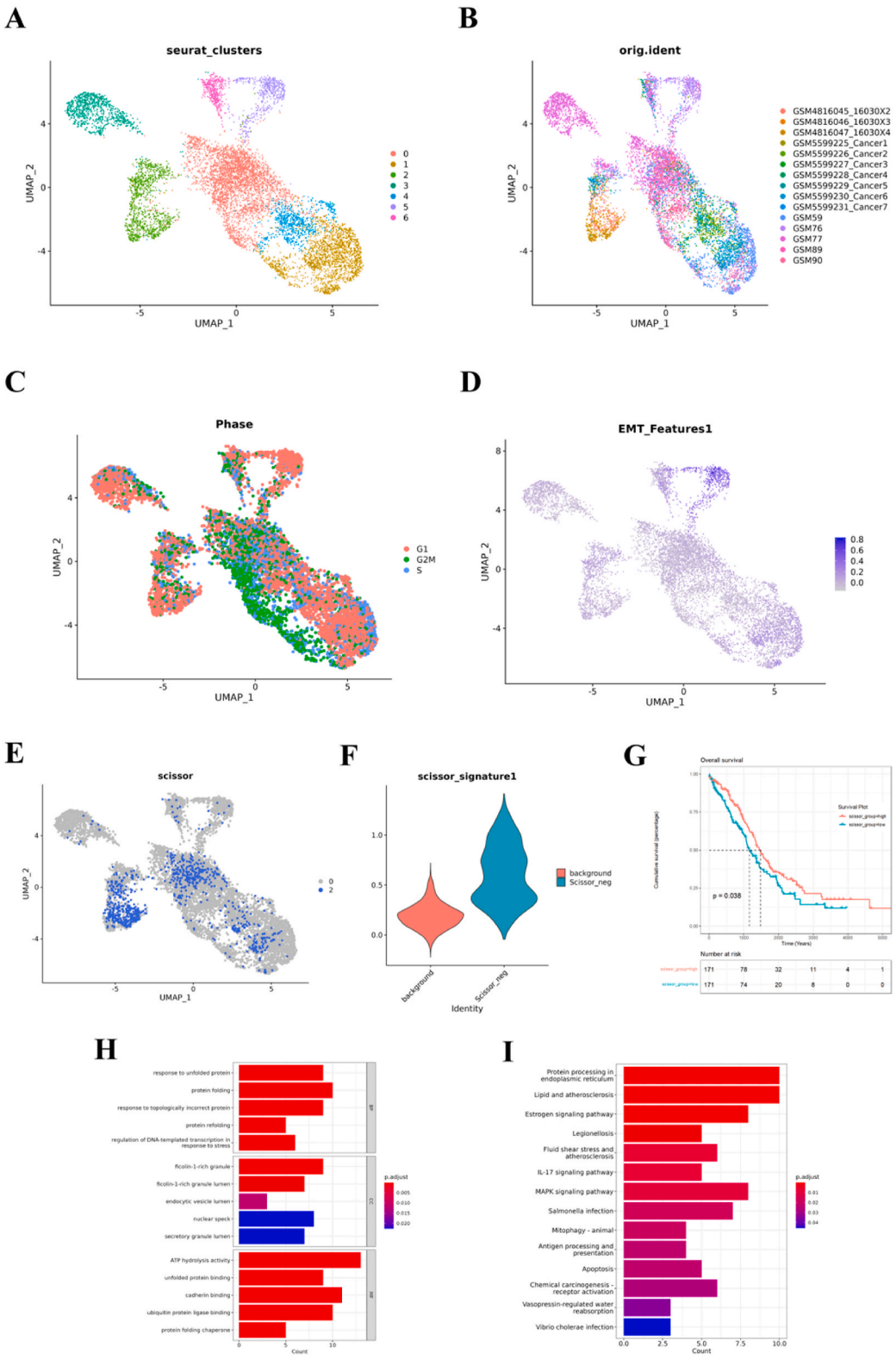


Fig. 2. Identification of tumor cells that are sensitive to by Scissor.
(A) UMAP clustering of epithelial cells. Cells were color-coded with cluster information. (B) UMAP clustering of epithelial cells. Cells were color-coded with patient origin. (C) UMAP clustering of epithelial cells. Cells were color-coded with cell phase. (D) UMAP clustering of epithelial cells. Cells were color-coded with EMT score. (E) UMAP plot of the integrated dataset depicts Scissor status. Scissor-cells were colored in blue. Background

cells were colored in grey. (F) Violin plots of Scissor score of Scissor-cells and background cells. (G) Overall survival analysis of TCGA cohort based on expression of Scissor genes by Kaplan-Meier plotter. The patients were grouped into two groups based on median Scissor score. (H) Bar plot of the GO enrichment analysis for Scissor signature genes. Y-axis represents the enriched GO terms; X-axis represents the amount of the Scissor signature genes GO terms. GO, Gene Ontology. (I) Bar plot of the KEGG pathway enrichment analysis for Scissor signature genes. Y-axis represents pathways; X-axis represents the amount of the Scissor signature genes enriched in KEGG pathways. KEGG, Kyoto Encyclopedia of Genes and Genomes.

the number of samples that can be sequenced. Moreover, the practical challenges of treating patients with varying drug combinations compared to controlled in vitro experiments pose significant hurdles. To bridge these gaps, algorithms have been developed that integrate bulk RNAseq data from large cohorts or cell line assays with scRNAseq datasets. These algorithms aim to identify cell subpopulations driving critical phenotypes such as disease stage, tumor metastasis, treatment response, and survival outcomes. One such computational method, SCISSOR, is designed to pinpoint subpopulations within a single-cell dataset that correlate with specific phenotypes of interest [24]. This method holds promise for advancing the development of cell type-specific therapies and the discovery of prognostic biomarkers.

In this study, we utilized the Scissor analysis to integrate bulk RNAseq data from paclitaxel-resistant OV cells with scRNAseq data from OV patient samples. We examined the association between cellular composition, cell-cell communication, and clinical outcomes. This analysis aims to enhance our understanding of paclitaxel responsiveness across different OV patient groups and to establish a foundation for personalized management strategies for OV. The insights gained could lead to more targeted and effective treatments, tailored to the unique cellular landscapes of individual patients.

2. Materials and methods

2.1. Acquisition of datasets and data preprocessing

Three scRNA-seq datasets GSE154600, GSE158937 and GSE184880 of total 15 OV samples were first downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database. We then used "Seurat" R package (version 4.1.1) to process the datasets [25]. Initially, the percentage of mitochondrial genes, ribosomal genes and erythrocyte genes were determined using the PercentageFeatureSet function. Cells with a gene number <200 and >5,000, mitochondrial gene content >10 %, ribosomal gene content <20 % and erythrocyte gene content >10 % and were excluded. Subsequently, scRNA-seq data were normalized by the NormalizeData function, and the top 2000 genes with highly variable features were identified by FindVariableFeatures function. We used ScaleData function to standardize the count data. The batch effects between different datasets were removed by "harmony" R package (version 0.1.1) [26]. The top 2000 highly variable genes were used for principal component analysis (PCA). Thus, the top 30 principal components were manually selected for cell clustering analysis using the Uniform Manifold Approximation and Projection (UMAP). Canonical marker genes of specific cell types in were used to annotate the clusters of cells. We used FindAllMarker function to identify the genes upregulated in the Scissor-cells. "DoubletFinder" R package (version 2.0.3) was carried out to remove the potential doublet cells in the scRNAseq dataset with default parameters [27].

Moreover, the bulk RNA-seq data of TCGA-OV samples were accessed from TCGA (https://portal.gdc.cancer.gov/) database. The statistical results of somatic mutation in TCGA dataset were visualized with the "maftools" R package software (version 2.12.0). Additional OV samples were obtained from the GSE14764, GSE23554 and GSE26712 datasets. In this study, OV samples from patients without a complete follow-up information were excluded. Bulk RNAseq data of paclitaxel resistance in ovarian cancer cell lines was obtained from the GSE172016.

2.2. Cell-cell communication analysis by CellChat

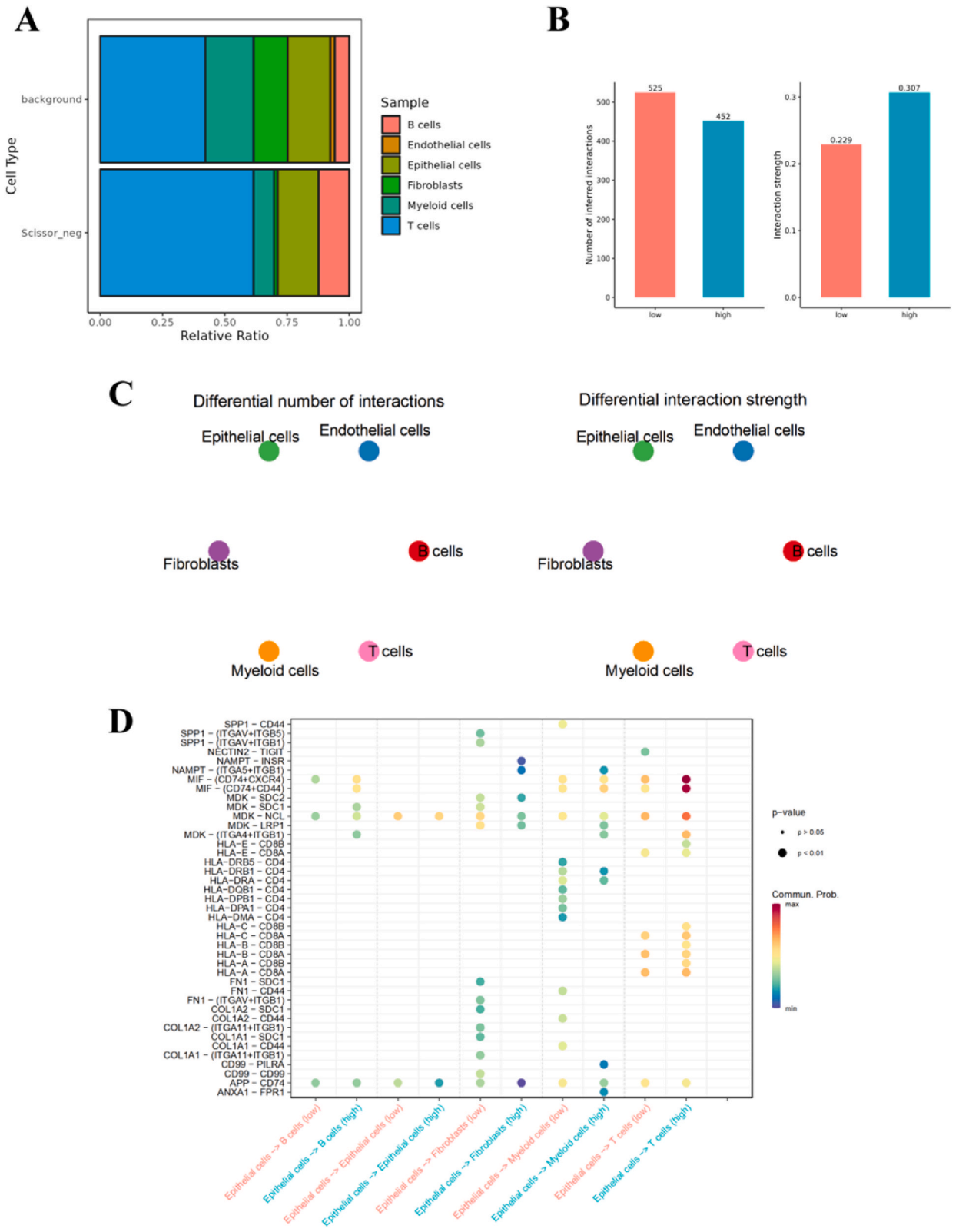
To perform CellChat (version 1.6.1) analysis, we first removed low-quality cells and genes, normalized the expression values, and performed clustering to identify cell types using Seurat R package. We then used "generate_cell_chat" function to generate the cell-cell communication network, which contains information about the ligand-receptor interactions between different cell types. To visualize the cell-cell communication network, the "plot_cell_chat" function was performed.

2.3. Functional and immune cell infiltration analysis

We used ClusterProfiler R package (version 4.4.4) to perform GO and KEGG analysis. To perform GO analysis, we utilized the enrichGO function, which performs enrichment analysis using the GO database. To perform KEGG pathway analysis, we used the enrichKEGG function, which performs enrichment analysis using the KEGG database. To carry out immune infiltration analysis in the TCGA-OV datasets, we uploaded our data to the Timer website (http://timer.cistrome.org/), which allowed us to obtained the immune cell infiltration information of TCGA-OV sample based on different algorithms, including CIBERSORT.

2.4. Identification of prognostic genes and construction of scissor genes prognosis risk model

The LASSO Cox regression analysis was performed to construct a predictive signature. The R package "glmnet" (version 4.1.4) was used to achieve the variable selection and shrinkage of the LASSO algorithm. The risk scores of each patient were calculated by the



(caption on next page)

Fig. 3. Analysis of cell-cell communication between the Scissor high and Scissor low groups using CellChat
(A) The proportion of each cell type in the Scissor high and Scissor low groups. (B) Total interaction number and strength between cells in the Scissor high and Scissor low groups. (C) Overview of selected ligand-receptor interactions of tumor cells and other cells in the Scissor high and Scissor low groups. The line thickness is proportional to the number of ligands when cognate receptors are present in the recipient cell type. The loops indicate autocrine circuits. (D) Bubble plot showing the selected ligand-receptor interactions between tumor cells and other cells in the Scissor high and Scissor low groups. P values are indicated by circle size, with the scale to the right (permutation test).

following formula: Risk score = sum (coeffcients* expression of gene n). The "survminer" (version 0.4.9) and "survival" R packages (version 3.4.0) were used to determine the Kaplan-Meier survival curve. The nomogram of the OV prognosis model was established via univariate and multivariate analyses, combined with clinical characteristics. Then, calibration curves were constructed for the prediction of survival of different time points in OV.

2.5. Statistical analysis

R software (version 4.0.0) were used for statistical analysis. Continuous variables are presented as the mean ± standard deviation. Univariate and multivariate Cox regression analyses were used to demonstrate their prognostic value. Kaplan–Meier curves with a two-sided log-rank test were used to compare the overall survival (OS) of the patients in the high and low expression groups. p-values <0.05 were considered statistically significant.

3. Results

3.1. The cell atlas of OC

To dissect the heterogeneity of tumor cells and cells in tumor microenvironment, we analyzed the transcriptome of 15 OC patients from 3 public scRNA-seq datasets which contains high quality of sequencing data. After a quality screening based on the percentage of the expression of the mitochondrial, ribosomal and erythrocyte genes, we obtained a total of 52,874 high-quality cells with 29,126 total genes for further study (Fig. 1A). We identified twelve clusters (clusters 0–11) in these OC sample through dimension reduction and cluster analysis (Fig. 1B). Canonical markers for different cell type were then used to annotate each cluster of cells and revealed six major cell types, including epithelial cells, fibroblasts, endothelial cells, T cells, B cells and myeloid cells (Fig. 1C). Fig. 1D illustrated typical markers for different cell types.

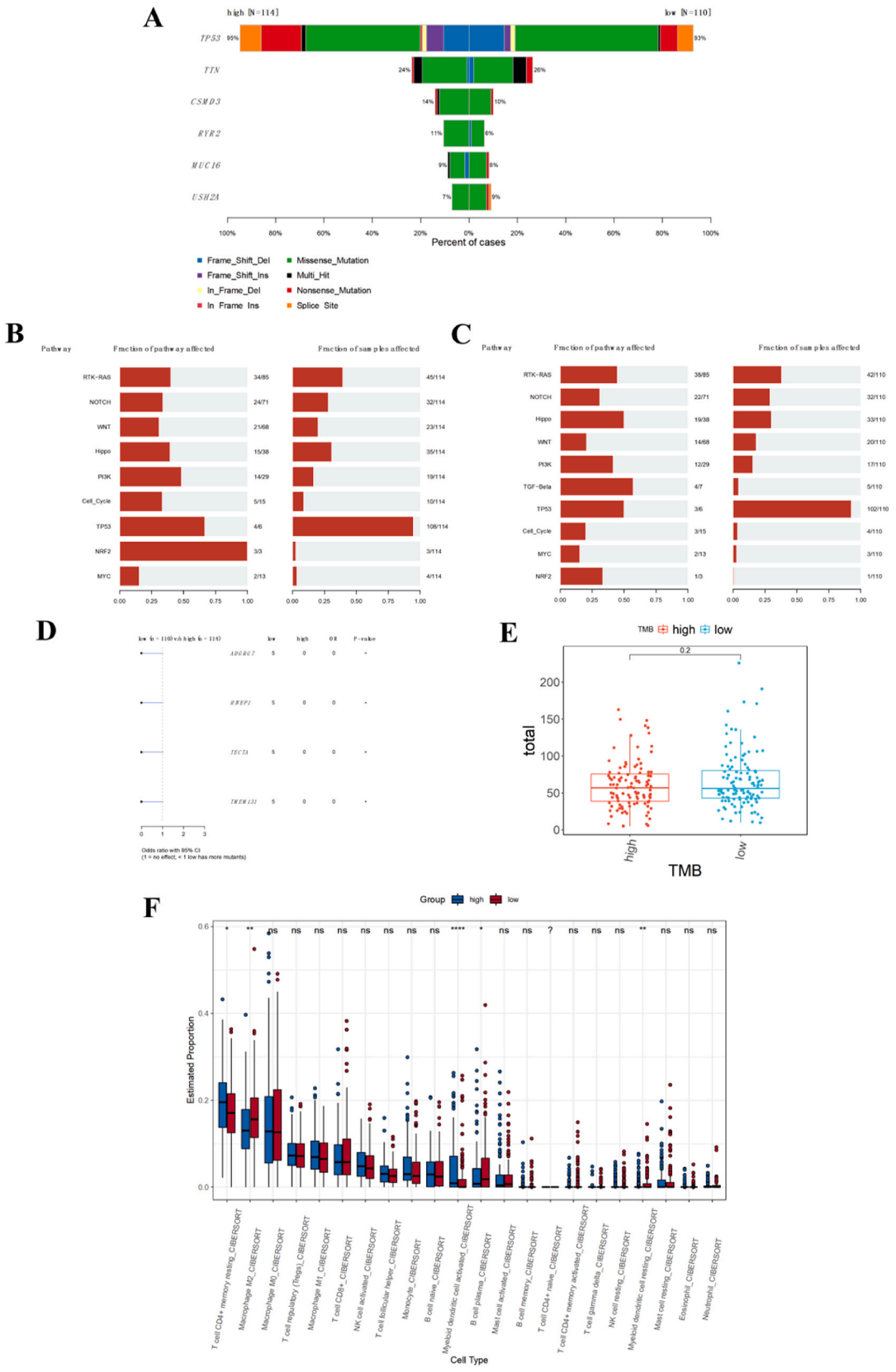
3.2. Identifying the subpopulation of tumor cells that are sensitive to paclitaxel

To further delineate the transcriptional landscape of OC tumor cells, we re-clustered the epithelial cells, identifying seven distinct clusters that illustrate pronounced transcriptional heterogeneity among OC patients (Fig. 2A–B). Subsequently, we conducted cell cycle and epithelial-mesenchymal transition (EMT) score analyses, finding that most high-EMT tumor cells were in the G1 phase (Fig. 2C–D).

To identify cells sensitive to paclitaxel treatment, we applied Scissor analysis to all tumor cells. The results identified 812 paclitaxel-sensitive cells (Scissor-) out of 9161 tumor cells (Fig. 2E). Further investigations revealed 90 significantly upregulated genes in the Scissor-cells, termed paclitaxel sensitive genes (PSG) (Fig. 2F). We then assessed PSG expression in The Cancer Genome Atlas (TCGA-OV) samples, categorizing them into two groups based on median expression levels (Fig. 2G). Notably, there was a significant difference in overall survival (OS) between the high and low Scissor groups ($p = 0.038$, Fig. 2G), suggesting that higher PSG expression correlates with improved survival outcomes. To elucidate the biological pathways and functions associated with PSG, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. The GO analysis showed that PSG were predominantly enriched in protein folding and DNA transcription in response to stress (Fig. 2H). The KEGG analysis highlighted significant enrichment in processes related to protein synthesis in the endoplasmic reticulum, lipid metabolism, atherosclerosis, and estrogen signaling pathways (Fig. 2I).

3.3. Cell-cell interaction profile of OV with different response to paclitaxel

We assessed the PSG score in OC single-cell RNA sequencing samples, dividing them into two groups based on their scores: six samples with positive scores and nine with negative scores. Cell proportion analysis revealed a higher T cell infiltration in the PSG-high group compared to the PSG-low group (Fig. 3A). We then conducted a CellChat analysis to examine the differences in cell-cell communication between the PSG-high and PSG-low groups. This analysis showed a decrease in the number of cell-cell interactions but an increase in their strength in the PSG-high group relative to the PSG-low group (Fig. 3B). Further analysis revealed that the enhanced strength of interactions in the PSG-high group was primarily observed between epithelial cells and T cells/B cells (Fig. 3C). Additionally, ligand-receptor profiles indicated a significant upregulation of MIF/CD74+CXCR4 and MIF/CD74 + CD44 interactions between epithelial cells and T cells in the PSG-high group (Fig. 3D). These findings suggest that the tumor microenvironment (TME) may play a critical role in the response of tumor cells to paclitaxel treatment in OC.



(caption on next page)

Fig. 4. Statistics of mutation information in the Scissor high and Scissor low TCGA OV samples.
(A) Statistical results of the top 5 hyper abrupt genes in the scissor high samples. (B) Bar plot depicts the distributions of 9 pathways affected by mutation in the scissor high samples. (C) Bar plot depicts the distributions of 9 pathways affected by mutation in the scissor low samples. (D) Bar plot depicts the comparison of 4 driver genes in the scissor high and scissor low samples. (E) Comparison of TMB levels between the scissor high and scissor low samples. TMB, tumor mutation burden. (F) Differences of 22 immune cells infiltration between the scissor high and scissor low samples according to CIBERSORT algorithm.

3.4. Mutational landscape in TCGA OV sample with different PSG signature

Given the critical role of genetic mutations in tailoring cancer treatment, we analyzed the somatic mutation profiles in TCGA-OV samples. Our analysis identified genes with high mutation frequencies, including CSMD3, RYR2, and MUC16 (Fig. 4A). Moreover, we observed distinct pathway impacts between risk groups: the NRF2 pathway was more significantly affected in the high-risk group, while the TP53 pathway showed greater alterations in the low-risk group (Fig. 4B–C). We identified five genes—ADGRG7, HIVEP1, TECTA, and TMEM131—that exhibited significant differences in mutation rates between the high and low-risk groups (Fig. 4D). However, there was no observed difference in the overall tumor mutation burden (TMB) between these groups (Fig. 4E).

To further distinguish the immune cell landscape between the high- and low-risk groups, we employed CIBERSORT to analyze 22 types of immune cells. The results showed that the high-risk group had higher proportions of $CD4^+$ memory resting T cells and activated dendritic cells (DCs), whereas the low-risk group was characterized by an enrichment of M2 macrophages, resting DCs, and plasma B cells, which are typically associated with immunosuppression (Fig. 4F). These findings underscore the potential of immune profiles as indicators of risk and therapeutic response in ovarian cancer.

3.5. Identification of a five-gene signature

In our study of the TCGA-OV cohort, all PSGs were evaluated for their association with overall survival (OS) using univariate Cox regression analysis. This analysis revealed that ten genes were significantly correlated with patient prognosis ($P < 0.05$). Subsequently, a paclitaxel-sensitive signature comprising five genes was developed for prognosis prediction using the LASSO regression algorithm (Fig. 5A and B). The formula for the risk score is as follows: Risk score = (−9.842e-07 × expression level of HSP90AA1) + (1.179e-05 × expression level of JUN) + (−1.748e-06 × expression level of CALM1) + (−1.424e-05 × expression level of RBP1) + (−5.665e-05 × expression level of GLRX5). The 342 OC patients were categorized into high- and low-risk groups based on the median risk score, with the high-risk group demonstrating better survival outcomes than the low-risk group (Fig. 5C). The model was also validated across different datasets, including GSE14764, GSE23554, and GSE26712. Notably, in the GSE23554 and GSE26712 datasets, patients with higher signature scores exhibited significantly better survival times compared to those with lower scores (Fig. 5D–F). Further analysis involved both univariate and multivariate Cox regression to evaluate if the risk score could serve as an independent prognostic factor relative to other common clinicopathological parameters for the TCGA-OV patients. The multivariate Cox regression analysis identified both ovarian tumor stage and risk score as significant predictors of OS (Fig. 5G). Similarly, univariate analysis confirmed the significance of the prognostic risk score (Fig. 5H). In the GSE26712 cohort, both multivariate and univariate analyses indicated that the risk score was a prognostic factor for OS (Fig. 5I and J). The risk graph and heatmap were used to detail survival outcomes and display the expression differences of the five genes within the model across the risk groups for both the TCGA and GSE26712 cohorts (Fig. 6A and B). The risk score also showed robust performance in predicting OS for individuals in these cohorts (Fig. 6C and D). Based on these findings, we developed nomograms that incorporate clinical characteristics and the risk score for both the TCGA and GSE26712 cohorts, as illustrated in Fig. 6E and G. These tools offer a quantitative method for predicting survival, enhancing the personalization of treatment strategies in ovarian cancer. The calibration curve effectively depicts the concordance between the predicted and observed survival probabilities at 1-year, 3-year, and 5-year intervals for ovarian cancer patients. The data points closely approach the ideal line, indicating that the prognostic model accurately reflects the survival outcomes based on the gene signature. This precise calibration suggests that the model captures essential biological variables influencing patient survival and highlights its potential reliability for clinical application (Fig. 6F and H). Taken together, the nomogram may serve as a useful quantitative tool for predicting prognoses of OV.

4. Discussion

Ovarian cancer drug resistance is a common problem in the treatment of ovarian cancer, resulting in the cancer continuing to grow and spread, making treatment more difficult [28]. There are several factors that can contribute to drug response in ovarian cancer, including genetic mutations, changes in tumor microenvironment, and the activation of certain cellular pathways [29]. Platinum-based chemotherapeutics remain the principal agents for treating both newly diagnosed ovarian cancer and platinum-sensitive recurrent forms of the disease. The integration of poly (ADP-ribose) polymerase inhibitors (PARP inhibitors) into the ovarian cancer treatment regimen represents a significant advancement, offering substantial benefits to patients exhibiting defects in DNA repair mechanisms. Recent findings underscore the utility of PARP inhibitors in patients with advanced-stage ovarian cancer, extending beyond those harboring BRCA mutations. However, the emergence of resistance to PARP inhibitors among certain patient cohorts presents a critical challenge, tempering the optimism generated by these breakthroughs [30]. To overcome this, effort has been made to develop new drugs or combination therapies that target multiple pathways or mechanisms of resistance [31]. However, it is

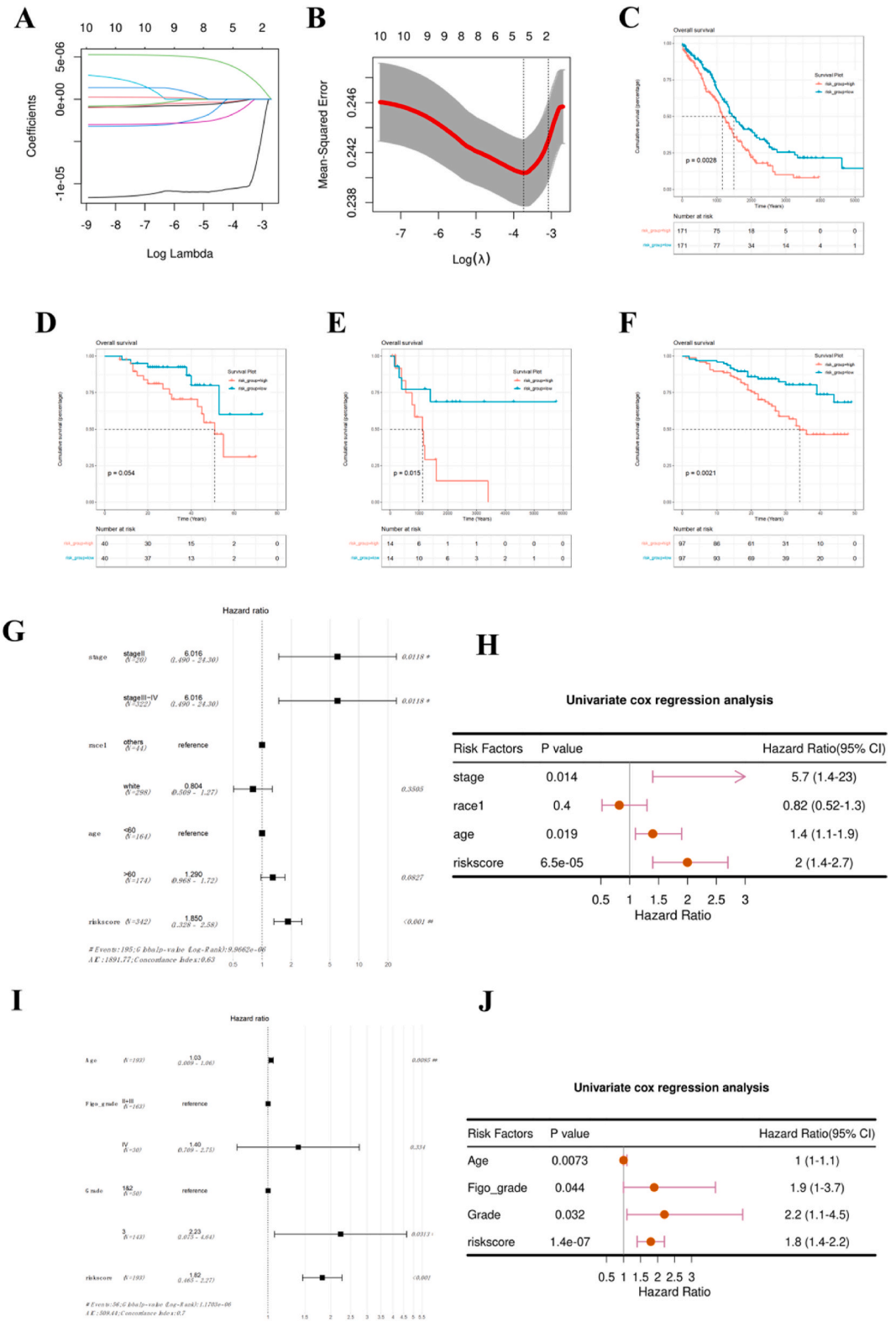


Fig. 5. LASSO Cox regression and overall survival analysis of Scissor signature genes in OV.
(A) LASSO coefficient profile plots of each independent variable. (B) The partial likelihood deviance for the LASSO Cox regression analysis. (C) Analysis of the prognostic value of the RiskScore in TCGA-OV dataset. (D) Analysis of the prognostic value of the RiskScore in GSE14764 dataset. (E) Analysis of the prognostic value of the RiskScore in GSE23554 dataset. (F) Analysis of the prognostic value of the RiskScore in GSE26712 dataset.

(G) Forest plot of the multivariate regression analysis in the TCGA cohort. (H) Forest plot of the univariate regression analysis in the TCGA cohort. (I) Forest plot of the multivariate regression analysis in the GSE26712 cohort. (J) Forest plot of the univariate regression analysis in the GSE26712 cohort.

also important to identify biomarkers that can predict drug response, allowing for the selection of more personalized treatment options for each patient.

Previous research has identified multiple biomarkers capable of predicting drug response in OC. For instance, patients with mutations in the TP53 gene or elevated HE4 protein levels exhibit reduced sensitivity to chemotherapy, suggesting a need for alternative treatment approaches [32,33]. Conversely, mutations in genes associated with the DNA damage response (DDR) pathway, such as ATM, ATR, and CHK1, or low ERCC1 expression levels may enhance responsiveness to platinum-based chemotherapy [34]. However, it is crucial to recognize that biomarkers do not always guarantee accurate drug response predictions; thus, treatment strategies should consider a variety of factors. In this context, RNA sequencing (RNAseq) has been employed to refine drug response prediction accuracy. One notable study utilized TCGA online data and OV GEO datasets to develop a predictive model based on 10 genes [35]. Our findings indicate that the NRF2 pathway is notably active in paclitaxel-sensitive patients, whereas alterations in the TP53 pathway are more prevalent in other patient groups.

Drug treatments, notably chemotherapy, can significantly impact T cell functionality, playing a pivotal role in cultivating an immunosuppressive tumor microenvironment. Our study highlights that in OV, chemotherapy enhances the interaction between epithelial cells and T cells/B cells. Specifically, interactions involving MIF/CD74+CXCR4 and MIF/CD74 + CD44 were found to be substantially upregulated, suggesting their crucial role in the tumor's response to paclitaxel. Chemotherapy is known to prompt the secretion of immunosuppressive cytokines such as TGF-β and IL-10 from both tumor and stromal cells [36]. These cytokines can inhibit T cell activation and proliferation, facilitating the proliferation of regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), which further diminish T cell functionality [37]. Additionally, chemotherapy can alter the expression of major histocompatibility complex (MHC) molecules and components of the antigen-processing machinery in tumor cells, impairing the recognition and activation of T cells [38]. It also influences the expression of co-stimulatory molecules, such as CD80 and CD86, which are essential for robust T cell activation [39]. These effects collectively highlight the complex interplay between chemotherapy and the immune landscape within tumors, underscoring the need for integrated therapeutic strategies that consider both tumor cell killing and the modulation of the immune response.

Our research significantly contributes to advancing ovarian cancer prognostics, yet it is essential to recognize the inherent limitations that could moderate our conclusions. One major challenge is the validation of our prognostic model using publicly available datasets such as TCGA and GSE26712. These datasets often lack comprehensive clinical parameters, especially detailed treatment histories, which are crucial for enhancing the predictive accuracy of the model [40–42]. Therefore, robust validation in a broader range of prospective cohorts, enriched with extensive longitudinal data on treatments and outcomes, is imperative. Such data will not only validate the reliability of the model but also expand its clinical applicability [43].

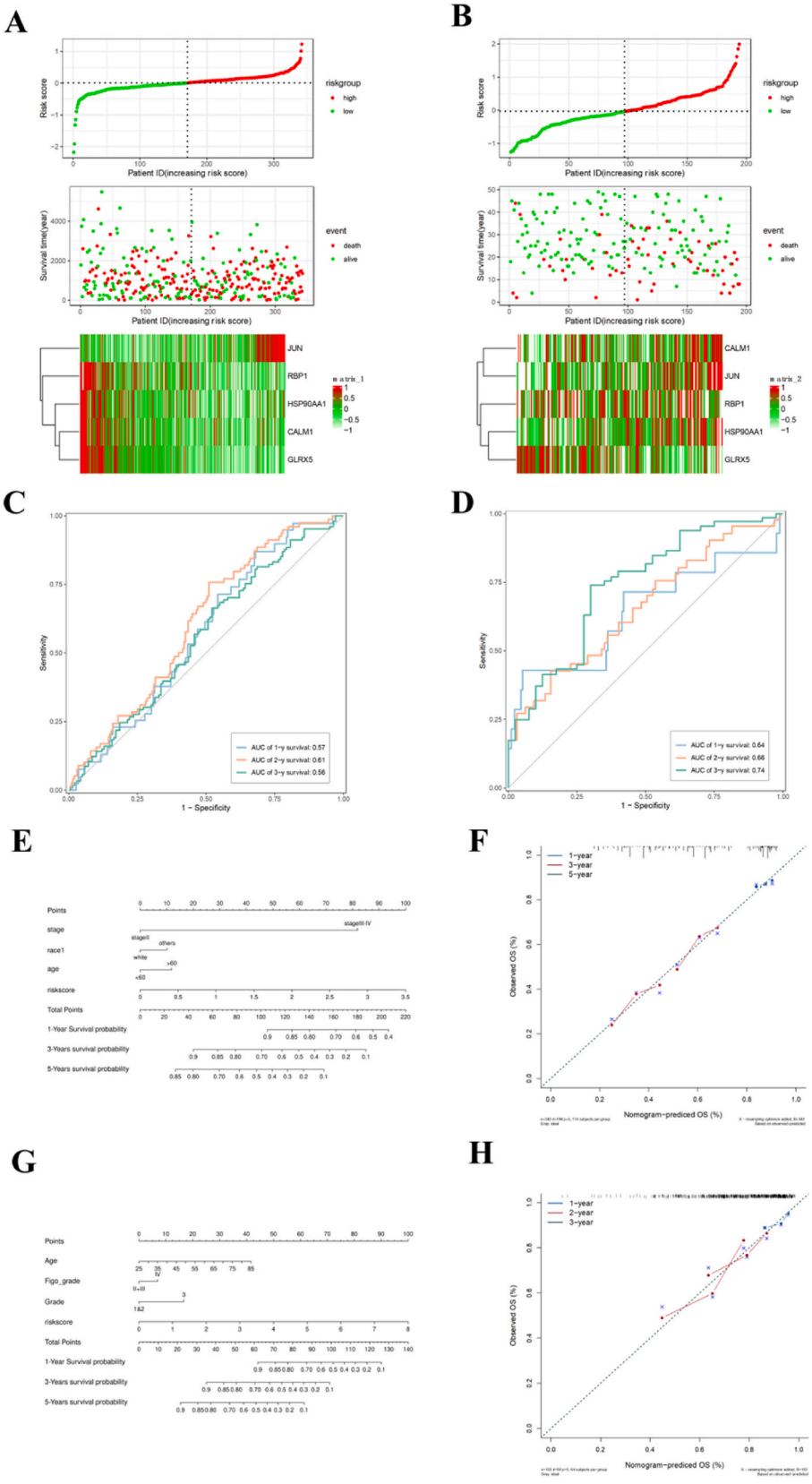
Furthermore, despite the proven robustness of the Scissor algorithm, which has shown superior performance in multiple studies, it does not capture all sources of biological variability [44,45]. The heterogeneity of the tumor microenvironment, a pivotal factor in resistance and prognosis, is still not fully represented in the datasets we analyzed. This limitation may introduce biases in our findings, potentially affecting their applicability across diverse patient cohorts or in different clinical settings. This underscores the need for ongoing refinement of prognostic models to incorporate a more comprehensive view of the biological landscape, ensuring that they remain robust and predictive in a real-world clinical context.

To mitigate the potential limitations identified in our current study, it is imperative for future research efforts to focus on integrating diverse datasets, particularly emphasizing the inclusion of data from multi-institutional and multicentric cohorts. Enriching these datasets with a broad range of clinical variables and comprehensive demographic data is crucial. These enhancements will not only strengthen the external validity of our prognostic models but will also significantly boost their applicability in the field of precision oncology. By ensuring that these models are robust and adaptable, we can better accommodate variations in clinical practices and patient populations across different settings. This approach will enable more personalized and effective treatment strategies, ultimately improving patient outcomes in ovarian cancer care. The strategic integration of varied data sources and detailed clinical information will provide a more holistic understanding of the factors influencing disease progression and treatment response, facilitating the development of universally applicable and highly accurate prognostic tools.

In summary, combining the public scRNAseq dataset and bulk RNAseq databases, we selected the feature DEGs to construct a model to determine the patients' responses to platinum-based chemotherapy. Meanwhile, a prognostic risk model was established to help predict patients' prognoses.

Data availability

GEO data analyzed during this study can be downloaded from https://www.ncbi.nlm.nih.gov/geo website, including GSE154600, GSE158937, GSE184880, GSE14764, GSE23554, GSE26712 and GSE172016 datasets. TCGA data can be downloaded from https://portal.gdc.cancer.gov/database.



(caption on next page)

Fig. 6. Validation of a 6 scissor genes-based OS model in OV.
(A) Risk score and survival status of each patient in the TCGA cohort. Heatmap of 6 scissor genes in the TCGA cohort. (B) Risk score and survival status of each patient in the GSE26712 cohort. Heatmap of 6 scissor genes in the GSE26712 cohort. (C) The ROC analysis of TCGA-OV dataset for prognosis prediction by riskscore. (D) The ROC analysis of GSE26712 dataset for prognosis prediction by riskscore. (E) Columnar plots to predict 1-year, 3-year, and 5-year overall survival of OV patients in TCGA cohort. (F) Calibration curves of the overall survival line plot model in the TCGA group. (G) Columnar plots to predict 1-year, 2-year, and 3-year overall survival of OV patients in GSE26712 cohort. (H) Calibration curves of the overall survival line plot model in the GSE26712 group.

CRediT authorship contribution statement

ZhenWei Zhang: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **MianMian Chen:** Writing – review & editing, Software, Investigation. **XiaoLian Peng:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Yang, et al., Effect of oral contraception on screening tests for primary aldosteronism: a 10-year longitudinal study, J. Clin. Endocrinol. Metab. 108 (7) (2023) 1686–1695, https://doi.org/10.1210/clinem/dgad010.
[2] L.A. Torre, et al., Ovarian cancer statistics, 2018, Ca - Cancer J. Clin. 68 (4) (2018) 284–296, https://doi.org/10.3322/caac.21456.
[3] D.G. Kwolek, et al., Ovarian, uterine, and vulvovaginal cancers: screening, treatment Overview, and prognosis, Med. Clin. 107 (2) (2023) 329–355, https://doi.org/10.1016/j.mcna.2022.10.016.
[4] A. Chandra, et al., Ovarian cancer: current status and strategies for improving therapeutic outcomes, Cancer Med. 8 (16) (2019) 7018–7031, https://doi.org/10.1002/cam4.2560.
[5] M. Cristea, et al., Practical considerations in ovarian cancer chemotherapy, Ther Adv Med Oncol 2 (3) (2010) 175–187, https://doi.org/10.1177/1758834010361333.
[6] S.S. Li, J. Ma, A.S.T. Wong, Chemoresistance in ovarian cancer: exploiting cancer stem cell metabolism, J Gynecol Oncol 29 (2) (2018) e32, https://doi.org/10.3802/jgo.2018.29.e32.
[7] K. O'Leary, A biomarker-driven therapy for ovarian cancer, Nat. Med. (2023), https://doi.org/10.1038/d41591-023-00018-6.
[8] E.L. Christie, D.D.L. Bowtell, Acquired chemotherapy resistance in ovarian cancer, Ann. Oncol. 28 (suppl_8) (2017) viii13–viii15, https://doi.org/10.1093/annonc/mdx446.
[9] R.M. Wadapurkar, A. Sivaram, R. Vyas, RNA-seq analysis of clinical samples from TCGA reveal molecular signatures for ovarian cancer, Cancer Invest. 41 (4) (2023) 394–404, https://doi.org/10.1080/07357907.2023.2182123.
[10] K. Brasseur, N. Gevry, E. Asselin, Chemoresistance and targeted therapies in ovarian and endometrial cancers, Oncotarget 8 (3) (2017) 4008–4042, https://doi.org/10.18632/oncotarget.14021.
[11] J.W. Ricci, et al., Novel ABCG2 antagonists reverse topotecan-mediated chemotherapeutic resistance in ovarian carcinoma xenografts, Mol. Cancer Therapeut. 15 (12) (2016) 2853–2862, https://doi.org/10.1158/1535-7163.MCT-15-0789.
[12] M. Ke, et al., Single cell RNA-sequencing: a powerful yet still challenging technology to study cellular heterogeneity, Bioessays 44 (11) (2022) e2200084, https://doi.org/10.1002/bies.202200084.
[13] I. Dagogo-Jack, A.T. Shaw, Tumour heterogeneity and resistance to cancer therapies, Nat. Rev. Clin. Oncol. 15 (2) (2018) 81–94, https://doi.org/10.1038/nrclinonc.2017.166.
[14] R. Thouenon, G. Verdeil, Tumor microenvironment squeezes out the juice from T cells, Cell Res. (2024), https://doi.org/10.1038/s41422-024-00987-4.
[15] M. Cummings, C. Freer, N.M. Orsi, Targeting the tumour microenvironment in platinum-resistant ovarian cancer, Semin. Cancer Biol. 77 (2021) 3–28, https://doi.org/10.1016/j.semcancer.2021.02.007.
[16] Q. Zhang, et al., Tumor microenvironment manipulates chemoresistance in ovarian cancer, Oncol. Rep. 47 (5) (2022), https://doi.org/10.3892/or.2022.8313 (Review).
[17] T. Kan, et al., Single-cell RNA-seq recognized the initiator of epithelial ovarian cancer recurrence, Oncogene 41 (6) (2022) 895–906, https://doi.org/10.1038/s41388-021-02139-z.
[18] Z. Hu, et al., The repertoire of serous ovarian cancer non-genetic heterogeneity revealed by single-cell sequencing of normal fallopian tube epithelial cells, Cancer Cell 37 (2) (2020) 226–242 e7, https://doi.org/10.1016/j.ccell.2020.01.003.
[19] B. Izar, et al., A single-cell landscape of high-grade serous ovarian cancer, Nat. Med. 26 (8) (2020) 1271–1279, https://doi.org/10.1038/s41591-020-0926-0.
[20] S. Olalekan, et al., Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics, Cell Rep. 35 (8) (2021) 109165, https://doi.org/10.1016/j.celrep.2021.109165.
[21] L. Geistlinger, et al., Multiomic analysis of subtype evolution and heterogeneity in high-grade serous ovarian carcinoma, Cancer Res. 80 (20) (2020) 4335–4345, https://doi.org/10.1158/0008-5472.CAN-20-0521.
[22] L.M. Weber, et al., Genetic demultiplexing of pooled single-cell RNA-sequencing samples in cancer facilitates effective experimental design, GigaScience 10 (9) (2021), https://doi.org/10.1093/gigascience/giab062.
[23] J. Xu, et al., Single-cell RNA sequencing reveals the tissue architecture in human high-grade serous ovarian cancer, Clin. Cancer Res. 28 (16) (2022) 3590–3602, https://doi.org/10.1158/1078-0432.CCR-22-0296.
[24] D. Sun, et al., Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data, Nat. Biotechnol. 40 (4) (2022) 527–538, https://doi.org/10.1038/s41587-021-01091-3.
[25] Y. Hao, et al., Integrated analysis of multimodal single-cell data, Cell 184 (13) (2021) 3573–3587 e29, https://doi.org/10.1016/j.cell.2021.04.048.
[26] I. Korsunsky, et al., Fast, sensitive and accurate integration of single-cell data with Harmony, Nat. Methods 16 (12) (2019) 1289–1296, https://doi.org/10.1038/s41592-019-0619-0.
[27] C.S. McGinnis, L.M. Murrow, Z.J. Gartner, DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors, Cell Syst 8 (4) (2019) 329–337 e4, https://doi.org/10.1016/j.cels.2019.03.003.
[28] H. Ma, T. Tian, Z. Cui, Targeting ovarian cancer stem cells: a new way out, Stem Cell Res. Ther. 14 (1) (2023) 28, https://doi.org/10.1186/s13287-023-03244-4.
[29] P. Yue, B. Han, Y. Zhao, Focus on the molecular mechanisms of cisplatin resistance based on multi-omics approaches, Mol Omics 19 (4) (2023) 297–307, https://doi.org/10.1039/d2mo00220e.

[30] A. Giannini, et al., PARP inhibitors in newly diagnosed and recurrent ovarian cancer, Am. J. Clin. Oncol. 46 (9) (2023) 414–419, https://doi.org/10.1097/COC.0000000000001024.
[31] R.A. Avelar, et al., Small molecule mediated stabilization of PP2A modulates the Homologous Recombination pathway and potentiates DNA damage-induced cell death, Mol. Cancer Therapeut. 22 (5) (2023) 599–615, https://doi.org/10.1158/1535-7163.MCT-21-0880.
[32] F. Delie, P. Petignat, M. Cohen, GRP78 protein expression in ovarian cancer patients and perspectives for a drug-targeting approach, JAMA Oncol. 2012 (2012) 468615, https://doi.org/10.1155/2012/468615.
[33] J. Ferlay, et al., Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012, Eur. J. Cancer 49 (6) (2013) 1374–1403, https://doi.org/10.1016/j.ejca.2012.12.027.
[34] M. Choi, T. Kipps, R. Kurzrock, ATM mutations in cancer: therapeutic implications, Mol. Cancer Therapeut. 15 (8) (2016) 1781–1791, https://doi.org/10.1158/1535-7163.MCT-15-0945.
[35] S. Chen, et al., A risk model of gene signatures for predicting platinum response and survival in ovarian cancer, J. Ovarian Res. 15 (1) (2022) 39.
[36] B. Mirlekar, Tumor promoting roles of IL-10, TGF-beta, IL-4, and IL-35: its implications in cancer immunotherapy, SAGE Open Med 10 (2022) 20503121211069012, https://doi.org/10.1186/s13048-022-00969-3.
[37] N.R. Monu, A.B. Frey, Myeloid-derived suppressor cells and anti-tumor T cells: a complex relationship, Immunol. Invest. 41 (6–7) (2012) 595–613, https://doi.org/10.3109/08820139.2012.673191.
[38] M.L. Axelrod, et al., Biological consequences of MHC-II expression by tumor cells in cancer, Clin. Cancer Res. 25 (8) (2019) 2392–2402, https://doi.org/10.1158/1078-0432.CCR-18-3200.
[39] G. Driessens, J. Kline, T.F. Gajewski, Costimulatory and coinhibitory receptors in anti-tumor immunity, Immunol. Rev. 229 (1) (2009) 126–144, https://doi.org/10.1111/j.1600-065X.2009.00771.x.
[40] C. Maier, et al., A continued learning approach for model-informed precision dosing: updating models in clinical practice, CPT Pharmacometrics Syst. Pharmacol. 11 (2) (2022) 185–198, https://doi.org/10.1002/psp4.12745.
[41] A. Hoesseini, et al., Key aspects of prognostic model development and interpretation from a clinical perspective, JAMA Otolaryngol Head Neck Surg 148 (2) (2022) 180–186, https://doi.org/10.1001/jamaoto.2021.3505.
[42] A.K. Clift, et al., Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study, BMJ 381 (2023) e073800, https://doi.org/10.1136/bmj-2022-073800.
[43] X. Fan, et al., A ferroptosis-related prognostic model with excellent clinical performance based on the exploration of the mechanism of oral squamous cell carcinoma progression, Sci. Rep. 13 (1) (2023) 1461, https://doi.org/10.1038/s41598-023-27676-3.
[44] M. Cheng, et al., Integrating bulk and single-cell sequencing data to construct a Scissor(+) dendritic cells prognostic model for predicting prognosis and immune responses in ESCC, Cancer Immunol. Immunother. 73 (6) (2024) 97, https://doi.org/10.1007/s00262-024-03683-9.
[45] L. Huang, et al., Integrative single-cell sequencing analysis distinguishes survival-associated cells from the breast cancer microenvironment, Cancer Med. 12 (11) (2023) 12896–12911, https://doi.org/10.1002/cam4.5892.