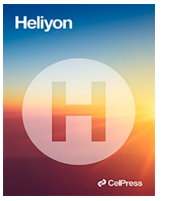
Heliyon 10 (2024) e38078

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Prognostic modeling and Emerging therapeutic targets Unveiled through single-cell sequencing in esophageal squamous Cell carcinoma

Binyang Pan [a,1], Haochun Shi [a,1], Guangyao Shan [a,1], Gujie Wu [a], Kungeng Rao [a], Jiaqi Liang [a], Xing Jin [a], Guoshu Bi [a], Mengnan Zhao [a,*], Weigang Guo [a,b,**]

[a] Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China
[b] Department of Thoracic Surgery and Urology, Shigatse People's Hospital, Shigatse, Tibet Autonomous Region, China

ARTICLE INFO

Keywords:
Esophageal squamous cell carcinoma
Single-cell RNA sequencing
Tumor microenvironment
Immunotherapy

ABSTRACT

ESCC presents a significant global health challenge due to its high mortality rates and varying responses to treatment. This underscores the critical need for novel diagnostic and predictive biomarkers to improve treatment outcomes.

Initially, we conducted single-cell transcriptome sequencing on a total of 128,688 cells obtained from 10 patients as part of our research. Utilizing machine learning and cross-validation techniques, we developed a model incorporating 12 genes that distinguish malignant cells from non-malignant ones. In vitro, we explored the effects of IGFBP2 knockdown on the proliferation, invasion, and migration of ESCC cells. The clinical relevance of IGFBP2 was confirmed through IHC and Kaplan-Meier survival analyses. Furthermore, using bioinformatics tools such as GSVA and xCell on public databases, we discovered that high expression of IGFBP2 is associated with an immunosuppressive tumor microenvironment in ESCC, characterized by reduced $CD8^+$ T cell infiltration. This was validated then through IHC.

In summary, our study integrates single-cell sequencing and sophisticated computational techniques to highlight IGFBP2 as a promising biomarker and therapeutic target in ESCC.

1. Introduction

Esophageal cancer presents a significant global health challenge, with approximately 510,716 new cases reported in 2022, making it the eleventh most prevalent cancer worldwide and accounting for 2.6 % of all cancer diagnoses that year [1]. However, the mortality rate is even more alarming. In 2022, around 445,129 individuals died from esophageal cancer, positioning it as the seventh leading cause of cancer-related deaths and contributing to 4.6 % of all cancer fatalities. Notably, squamous cell carcinoma (SCC) was the predominant subtype, comprising 85 % of all esophageal cancer cases [1,2].

Due to the subtle nature of early-stage esophageal cancer symptoms, a considerable proportion of patients are diagnosed at intermediate to advanced stages [3]. For those ineligible for surgical resection, primary treatment modalities include radiotherapy,

* Corresponding author.
** Corresponding author. Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China.
E-mail addresses: mnzhao19@fudan.edu.cn (M. Zhao), guo.weigang@zs-hospital.sh.cn (W. Guo).
[1] Binyang Pan, Haochun Shi and Guangyao Shan contribute equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38078
Received 5 June 2024; Received in revised form 17 September 2024; Accepted 17 September 2024
Available online 21 September 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

chemotherapy, and immune checkpoint inhibitors [4–7]. To improve the prognosis of esophageal cancer patients and develop new treatment approaches, urgent exploration of novel therapeutic targets and biomarkers is imperative.

In previous decades, most transcriptome analyses of ESCC have relied on conventional bulk RNA sequencing data, significantly advancing our understanding of ESCC occurrence and progression [8]. However, ESCC exhibits a complex tumor immune microenvironment and considerable heterogeneity. Conventional RNA-seq techniques can only analyze tumors at the aggregate sample level, limiting the exploration of cellular diversity and molecular intricacies of tumor cells [9–11]. The emergence of single-cell sequencing technology in recent years has offered a transformative approach [12,13]. This methodology enables the characterization of phenotypic and biological behaviors of distinct cell types at the individual cell level, providing valuable insights into the intricate landscape of ESCC and holding promise for identifying potential therapeutic targets [14,15].

In this study, we identified genes uniquely expressed in cancer cells by analyzing both cancerous and non-cancerous cell types using single-cell RNA sequencing (scRNA-seq) data from ESCC. We aimed to pinpoint genes associated with cancer progression that could be targeted for therapeutic interventions. Additionally, we developed a gene prediction model with a scoring system to improve the accuracy of detecting individual cancerous cells within the scRNA-seq dataset, thereby enhancing our ability to isolate cancer cells for in-depth analysis.

2. Methods

2.1. The construction of a predictive model

A total of 10 samples, including 5 ESCC samples and 5 paired non-cancerous samples, were selected for 10x genomics single-cell RNA sequencing from the Department of Thoracic Surgery, Zhongshan Hospital, Fudan University. Subsequently, we conducted data preprocessing and downstream analysis following the procedures described in previous reports [14]. In brief, the general experimental protocol combined mincing tissue samples, enzymatic digestion, and filtration to obtain single-cell suspensions. Library preparation was done using the 10x Chromium kit, and sequencing was performed on an Illumina HiSeq platform. Data preprocessing included demultiplexing and alignment with Cell Ranger. Analysis in R included cell filtering, doublet detection, batch effect correction, mutual nearest neighbors, and application of Seurat for clustering analysis. Dimensional reduction, clustering, and marker identification were further analyzed, and the most expressed markers for the annotation of cell types were defined.

Next, differentially expressed genes between cancerous and non-cancerous cells were identified using the "FindAllMarkers" function of the Seurat R package, with the following criteria: adjusted P value less than 0.01 and absolute log2 fold change greater than or equal to 0.5. Then, the area under the receiver operating characteristic (ROC) curve (AUC) was calculated using the pROC R package, and we selected all genes with an AUC $>$0.6, as previously reported [16].

We employed a random forest model to select genes aberrantly associated with malignancy in ESCC. The random forest algorithm, known for its robustness, was selected due to its ability to handle high-dimensional data and its relative resistance to overfitting compared to other models [17]. The %IncMSE (percentage increase in mean squared error) metric was used to evaluate variable importance, as it provides a robust measure of how each variable contributes to the predictive power of the model. The number of variables was selected through five replications of 10-fold cross-validation, which ensures the model's stability and helps prevent overfitting. The final equation used in predicting the malignant score was generated by multiple logistic regression.

2.2. Immunohistochemistry (IHC) cohort

Tumor and adjacent esophageal tissues were collected from 100 ESCC patients who underwent radical surgery from January 1 to May 20, 2016. Before surgery, patients were not treated with radiotherapy, chemotherapy, or immunotherapy. The latest follow-up occurred in 2022.IHC was performed using antibodies against IGFBP2 (Rabbit, CY8405, Abways) and CD8 (Rabbit, CY5540, Abways), respectively. In the staining process, Envision two-step method reagent was used, and GTVisionTM III Detection System/Mo&Rb (Including DAB) (Gene Tech, Shanghai, China) was added, where PBS was used as the negative control. For the observation of slides of $CD8^+$ T cells, the mean number of $CD8^+$ cells of three representative fields was counted under the high power of a microscope ($\times$ 200). The staining intensity of IGFBP2 was categorized as follows: 0 (negative), 1 (weak), 2 (moderate), and 3 (strong).

2.3. Cell culture

Human ESCC cell lines TE-1 and KYSE-150 were obtained from the Chinese National Collection of Authenticated Cell Cultures, Shanghai, China. Cells were grown at 37 °C in a humidified atmosphere of 95 % air and 5 % CO2. They were cultured in either high-glucose DMEM (KeyGEN Biotech, Nanjing, China) or RPMI-1640, supplemented with 10 % fetal bovine serum (Every Green, Hangzhou, China), 100 U/mL penicillin, and 0.1 mg/mL streptomycin (Beyotime Biotechnology, Shanghai, China).

2.4. Cell counting Kit-8 assay

Cells were seeded into 96-well microplates and cultured at 37 °C with 5 % CO2 for 5 days. Then, cell counting kit-8 (CCK8) from TOPSCIENCE (Shanghai, China) with 10 μL of reagent in each well to measure cell activity every day [18]. Cell viability was assessed by a SpectraMax iD3 microplate reader (Molecular Devices, USA) with wavelengths at 450 nm.

2.5. Exogenous stimulation experiment

IGFBP2 stimulation experiments were performed by using recombinant IGFBP2 (ab63223; Abcam, Cambridge, MA, USA) under the dose of 10 ng/ml and 100 ng/ml.

2.6. Transwell assay

ESCC cells were seeded into Transwell chambers with 8.0 μm PC membrane at the optimal growth density. The cells were incubated for 48 h. In the incubation process, 200 μL of serum-free medium was added to the upper chamber, while 200 μL of complete medium with 10 % FBS was added to the lower chamber. To investigate their invasiveness, 60 μL of chilled matrix gel was applied to the upper chamber. After incubation, the fix was done using 4 % paraformaldehyde, and then, staining was done using 0.1 % crystal violet.

2.7. qRT-PCR

Quantitative PCR (qPCR) was conducted after RNA was reverse-transcribed into cDNA as previously reported, and specific primers were obtained from Sangon Biotech (Shanghai, China) [19]. Primer sequences are available in the supplementary materials. Gene expression levels were assessed using the $2^{-\Delta\Delta Ct}$ technique and standardized to the ACTB gene.

2.8. Transfection of small interfering RNAs

SiRNAs targeting IGFBP2 (si-IGFBP2) and CALML5 (si-CALML5) were obtained from RIBOBIO (Guangzhou, China). SiRNA-lipid complexes were prepared using Opti-MEM medium and Lipo8000™ Transfection Reagent (Beyotime Biotechnology), then incubated at room temperature for 20 min before being added to the cells at a final concentration of 100 nM siRNAs in each well.

2.9. Western blot assays

Protein concentrations were determined using the BCA Protein Quantification Kit (Yeasen Biotechnology). Equal amounts of protein (20–50 μg) were loaded onto SDS-PAGE and transferred to a PDVF membrane (Millipore, Germany). The membranes were blocked with 5 % bovine serum albumin (BSA) in Tris-buffered saline (TBS) containing 0.1 % Tween-20 from Servicebio (Wuhan, China), followed by incubation with the appropriate primary antibodies against IGFBP2 (1:2000, Rabbit, CY8405, Abways), β-actin (1:2000, Mouse, GB15001, Servicebio). Protein bands were processed by incubating with horseradish peroxidase-conjugated secondary antibodies and BeyoECL Star (Beyotime Biotechnology) and quantified using ImageJ.

2.10. Bioinformatics tools and data analysis

We obtained publicly available RNA sequencing data from the Cancer Genome Atlas (TCGA) database [20]. Utilizing the R packages immunedeconv and GSVA, we evaluated the immune microenvironment, leveraging signatures of various immune cells from previous studies [21–23]. The estimate R package was used for estimating tumor purity, ESTIMATE score, and immune score. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted using the R package cluster profile. The variables were compared between groups using two-way ANOVA and Student's t-test, and survival analysis was carried out through the Kaplan-Meier and log-rank tests. All the statistics were analyzed using GraphPad and the R-4.3.1 program, where p values were two-tailed with $p < 0.05$ considered significant [24].

3. Results

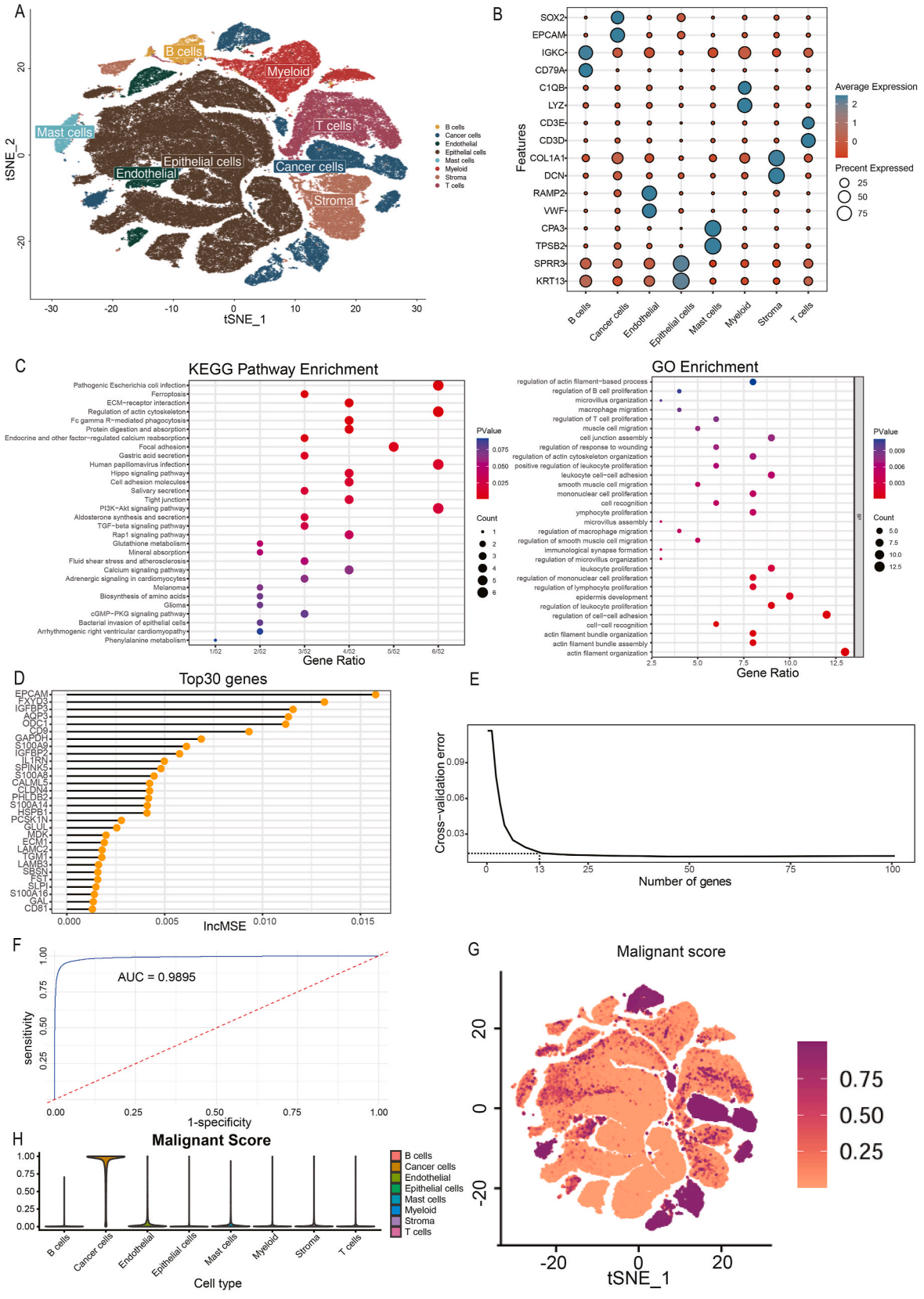
3.1. Construction of the model for malignant cell discrimination with twelve genes

As in our previous study, a total of 10 samples were obtained including 5 samples of the tumor and adjacent normal esophagus samples for the downstream analysis. Following quality control, a sum of 128,688 cells meeting the inclusion criteria were chosen for further analysis [14].

Drawing upon the CellMarker dataset, SingleR package, and previous studies, we employed t-SNE dimensionality reduction clustering to identify 8 distinct cell clusters (Fig. 1A). Fig. 1B illustrates the marker genes for each cell cluster [14]. Two well-known genes that are elevated in ESCC are also correspondingly overexpressed in our clustering results (Fig. S2) [25].

We then found 860 differentially expressed genes (DEGs) among cancer cells and seven other cell types following the criterion of gene expression FC less than 0.70 or more than 1.41 ($|\log2FC| \geq 0.5$), with 687 of which upregulated markedly in tumor cells.

Furthermore, we conducted a ROC curve analysis to evaluate the discriminatory power of each DEG at the single-cell level. Notably, 97 DEGs demonstrated significant efficacy (AUC >0.60) in distinguishing between malignant and non-malignant cells (Supplementary Table 1). Among these, 80 genes were upregulated and 17 downregulated in tumor cells. Functional enrichment analysis revealed that the upregulated genes are primarily associated with processes such as cell adhesion, tight junction formation, and regulation of lymphocyte proliferation. Additionally, these genes were found to be enriched in cancer-promoting signaling pathways including PI3K-AKT and TGF-β (Fig. 1C) [26,27]. These 97 genes were then compiled into a novel predictive gene set.



(caption on next page)

Fig. 1. Construction of the model for malignant cell discrimination with twelve genes. (A) The t-SNE plot to visualize cell-type clusters based on the expression of known marker genes. **(B)** Dotplot illustrating marker gene expression across different cell types. (**C**) GO and KEGG analysis for upregulated genes. (**D**) The lollipop plot displaying the top 30 genes contributing the most to the %IncMSE in the random forest model. (**E**) The line graph depicting the trend in cross-validation error with varying numbers of genes. (**F**) The ROC curve to illustrate the predictive performance of the final model on the test set. (**G**) Feature plot of malignant score across different types of cells. (**H**) Violin plot of malignant score across different types of cells.

We subsequently constructed a regression model based on the expression levels of the 97 genes to predict the cell type of each cell using the random forest method. The number of trees was set to 500, and 70 % of the cells were randomly drawn as the training set, while the rest were used as the test set. The results indicated that the model composed of the expression levels of the 97 genes could explain 89.45 % of the total variance, suggesting a close relationship between the expression levels of these genes and the cell type. Fig. 1D displays the top 30 genes contributing the most to the model.

Additionally, the distribution of 5 trials of 10-fold cross-validation error in a random forest classifier shows that the model could maintain a relatively low error rate while ensuring high simplicity when 13 genes were retained (Fig. 1E). We utilized multivariate logistic regression modeling on these 13 genes to further simplify the model. After excluding the only non-significant gene, S100A8, we ultimately retained 12 genes (EPCAM, FXYD3, IGFBP3, AQP3, ODC1, CD9, GAPDH, S100A9, IGFBP2, IL1R1N, SPINK5, S100A8, and CALML5) to calculate the malignant score for predicting the cell type as followed: Malignant score = −8.1419516 + 1.2387290 × EPCAM+1.2096745 × FXYD3+0.7818696 × IGFBP3−1.4175991 × AQP3+0.9339426 × ODC1+0.8148934 × CD9+0.4946624 × GAPDH+0.1004238 × S100A9+0.3371714 × IGFBP2−0.2899093 × IL1RN−0.8508509 × SPINK5+1.0425721 × CALML5. Fig. S1 showed the flow chart of the mode construction. Fig. 1G highlights the distribution of malignant scores, with a notable overlap observed in the high-score region corresponding to the cancer cell population, as further corroborated by the violin plot in Fig. 1H. Notably, the AUC of the ROC curve reached a remarkable value of 0.986 in the test set, demonstrating the excellent sensitivity and specificity of the model (Fig. 1F).

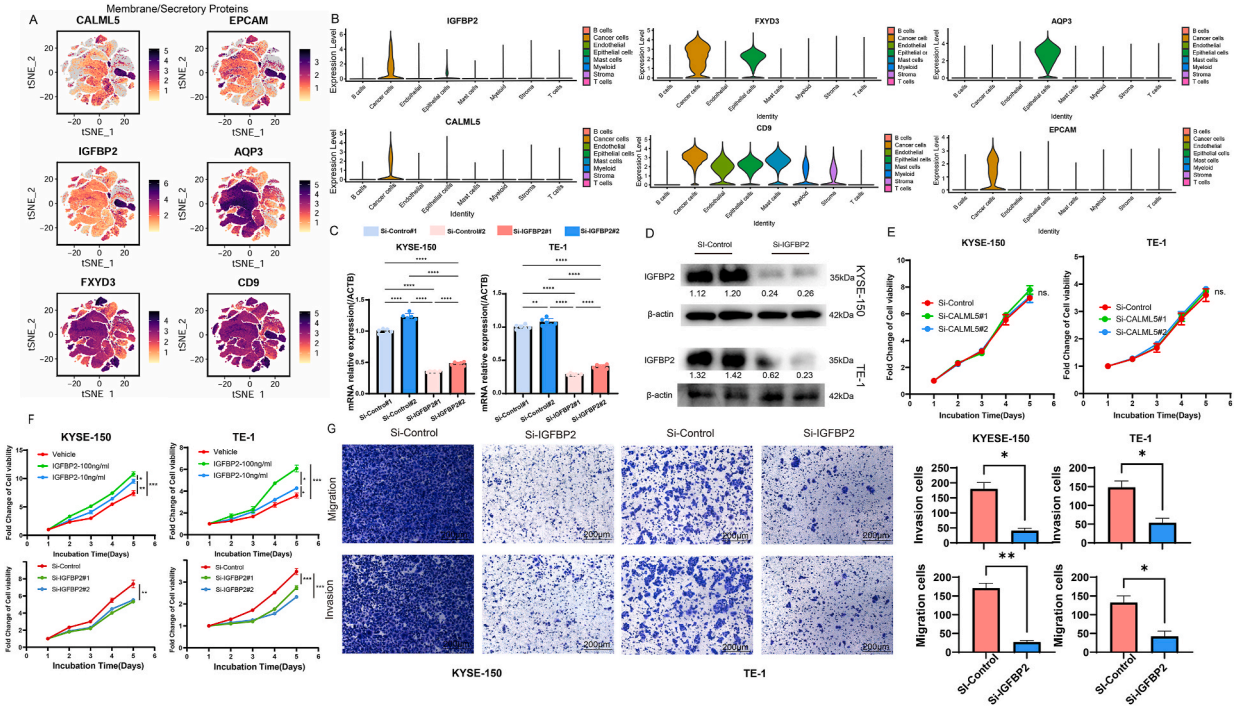


Fig. 2. In vitro validation of IGFBP2. (**A**) Feature plot for mRNA expression of secretory and membrane-bind protein. (**B**) Violin plot for mRNA expression of the secretory and membrane-bind protein in different cell clusters. (**C**) The histogram displaying the suppression of IGFBP2 expression 24 h after siRNA transfection. The presented data (mean ± SD) originate from two separate experiments (one-way ANOVA, n = 3). $^{**}p < 0.01$, $^{****}p < 0.0001$. (**D**) IGFBP2 expression from replicates (n = 2) in KYSE-150 and TE-1 after transfection with the number below showing IGFBP2/β-actin.The uncropped version can be found in supplementary materials. (**E**)The line graph depicting the proliferative ability of KYSE-150 and TE-1 cell lines after the knockdown of CALML5 (two-way ANOVA). n.s non-significant. (**F**) The line graph depicting the proliferative ability of KYSE-150 and TE-1 cell lines after knockdown or exogenous stimulation of IGFBP2 (two-way ANOVA). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. (**G**) Representative images and boxplots illustrating the impact of IGFBP2 knockdown on the migratory and invasive abilities of KYSE-150 and TE-1 cell lines. (one-way ANOVA, n = 2). $^{*}p < 0.05$, $^{**}p < 0.01$.

3.2. In vitro validation of IGFBP2

Among 12 genes in the model, six are membrane or secretory proteins, namely CALML5, EPCAM, IGFBP2, AQP3, FXYD3, and CD9 (Fig. 2A–B). Notably, IGFBP2 and CALML5 were significantly upregulated in cancer cells, which were selected consequently for further validation of their roles in ESCC. We performed transient transfections of TE-1 and KYSE-150 cell lines using two distinct siRNAs targeting IGFBP2 and CALML5 (Fig. 2C and D). The CCK-8 assay indicated that inhibition of IGFBP2 led to an attenuated proliferative capacity of both cell lines, rather than CALML5 (Fig. 2E–F). Conversely, the addition of recombinant human IGFBP2 significantly enhanced the proliferation of both cell lines in a dose-dependent manner (Fig. 2F). Furthermore, the Transwell assay revealed a significant decrease in the invasive and migratory abilities of cancer cells with reduced IGFBP2 expression (Fig. 2G). The above findings suggest that IGFBP2 acts as an oncogene in ESCC.

3.3. Clinical significance of IGFBP2

Next, we analyzed the IGFBP2 expression in a group of 100 paired samples using IHC, obtained from ESCC patients who underwent surgery and were pathologically diagnosed at Zhongshan Hospital (ZSHS) (Supplementary Table 2). As shown in Fig. 3A, IGFBP2 was significantly overexpressed in tumor samples compared with adjacent esophageal tissues, which was consistent with our results of scRNA-seq. To determine the clinical significance of IGFBP2, we conducted a Kaplan-Meier survival analysis. The results revealed that patients exhibiting high IGFBP2 staining intensity are correlated with poorer prognosis compared to those with low expression levels. (Fig. 3B). Moreover, multivariate Cox analysis revealed that the expression of IGFBP2 remained an independent prognostic factor for OS after considering age, gender, and stage (Fig. 3C).

3.4. Immune characteristics associated with IGFBP2

We then analyzed the role of IGFBP2 in shaping the tumor immune microenvironment of ESCC by exploring the interplay between IGFBP2 expression and the infiltration of immune cells, as well as the presence of immune modulators within the TCGA-ESCC dataset. Leveraging computational algorithms such as xCell, ConsensusTME, and GSVA, we uncovered a significant negative correlation between IGFBP2 expression and the infiltration of numerous tumor-associated immune cells, particularly $CD8^+$ T cells (Fig. 4A–C). This finding was further validated in the ZSHS cohort through IHC, which revealed a substantial negative correlation between IGFBP2 expression and the abundance of $CD8^+$ T cells (Fig. 4D).

Furthermore, we observed that elevated IGFBP2 expression is associated with the downregulation of a broad range of immune modulators, encompassing antigen presentation, co-stimulation, T cell effector function, and immune checkpoints (Fig. S3A), which

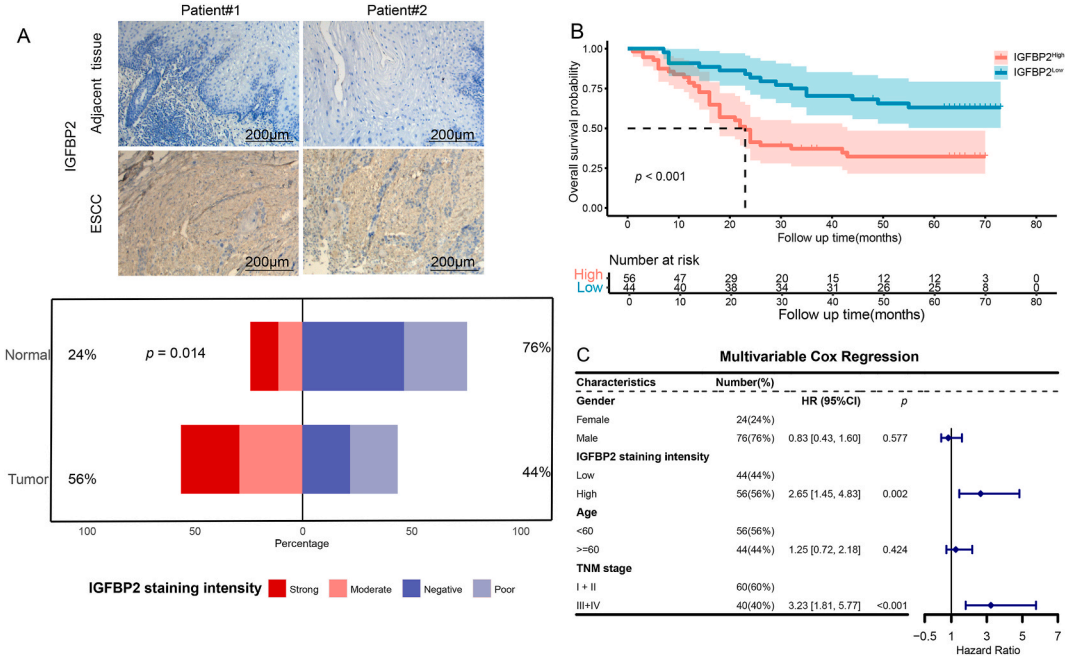
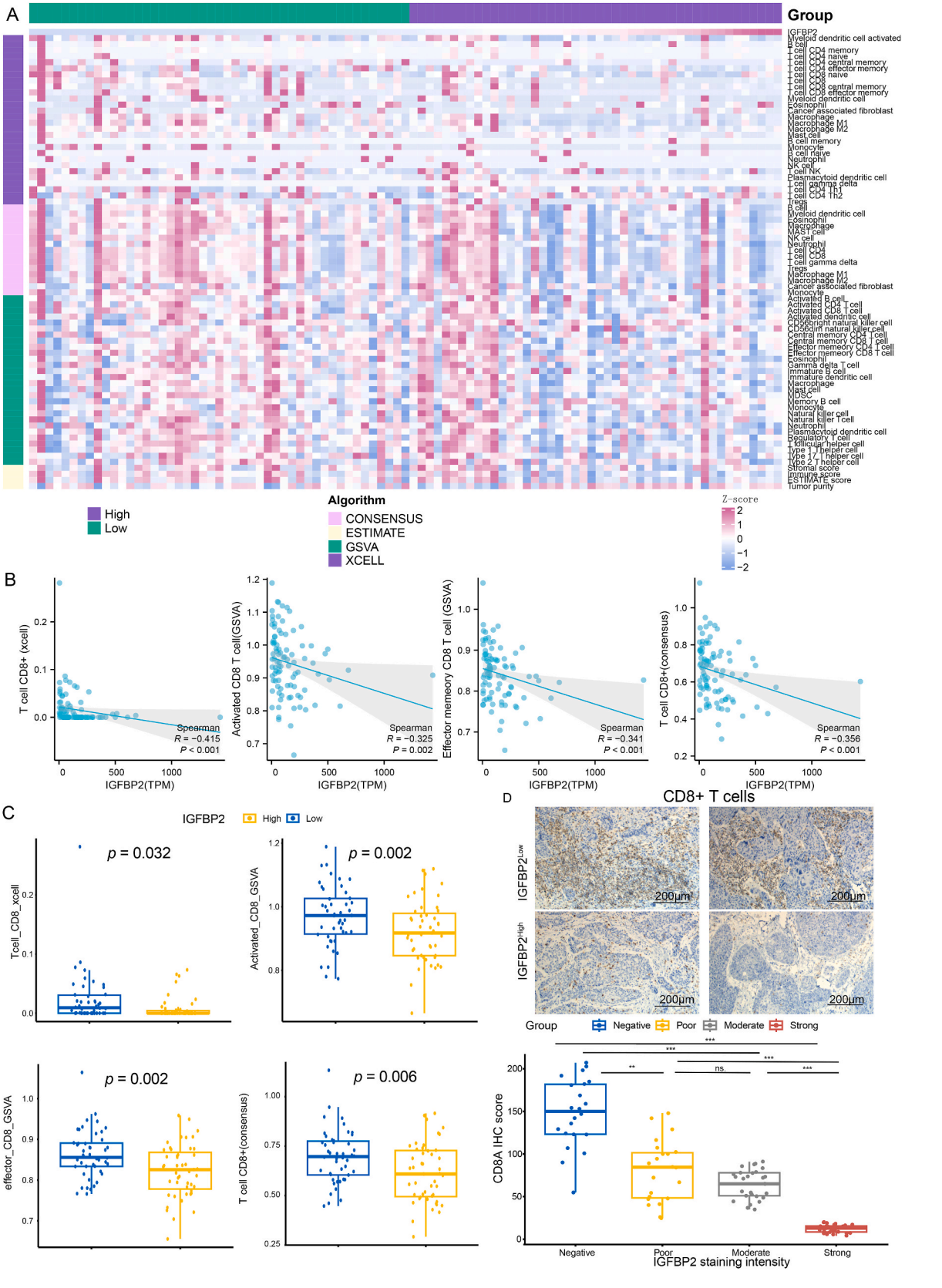


Fig. 3. Clinical significance of IGFBP2. (**A**) Representative images of IGFBP2 IHC staining in ESCC and adjacent esophageal tissues; Likert plot of IGFBP2 IHC staining results in normal esophageal tissues and ESCC. (**B**) Kaplan-Meier survival curves of the ZSHS cohort stratified by IGFBP2 staining intensity, assessed using the log-rank test. (**C**) The forest plot of multivariable Cox regression results in the ZSHS cohort. ZSHS, Zhongshan Hospital.



(caption on next page)

Fig. 4. Immune-related characteristics of IGFBP2. (**A**) Heatmap displaying the correlation between the IGFBP2 and immune infiltrating cells. (**B**) The scatter plot illustrating the correlation between IGFBP2 and $CD8^+$ T cells from different algorithms. (**C**) Box plot displaying the CD8 + T cells levels between two IGFBP2 groups from different algorithms (Student's *t*-test). (**D**) Representative images of $CD8^+$ T cell immunohistochemical (IHC) staining and box plot comparing the distribution of $CD8^+$ T cell IHC scores across different IGFBP2 staining intensities (Student's *t*-test). $^{**}p < 0.01$, $^{***}p < 0.001$.

suggests a potentially immunosuppressive role of IGFBP2 in ESCC. To further support this observation, we conducted the ESTIMATE algorithm to assess the immune score and ESTIMATE score in the TCGA-ESCC dataset, revealing a negative correlation with IGFBP2 expression (Fig. S3B).

4. Discussion

ESCC is reported to be one of the most aggressive and lethal cancers in adults. Despite advancements in therapeutic approaches, the overall prognosis for ESCC remains poor. The high degree of heterogeneity within ESCC further complicates treatment strategies and often leads to therapeutic failure in patients. In this study, we collected 10 paired samples from 5 patients and employed scRNA-seq to characterize the unique transcriptomic landscape of ESCC. Utilizing the "FindAllMarkers" function, ROC analysis, and machine learning, we established the malignant score comprising the expression levels of 12 to delineate the cancer cell-specific transcriptomic features in ESCC. We further selected IGFBP2 as a potential therapeutic target and validated its oncogenic role through cellular experiments as well as bioinformatics tools.

A noteworthy aspect of our research is the discovery of several tumor-related genes that are infrequently detected in traditional bulk data analyses. This achievement is largely attributable to the utilization of scRNA-seq technology. Given that cancer cells typically represent a minor fraction of this mix, tumor-specific markers can be overshadowed by their low abundance when using bulk RNA sequencing. In contrast, single-cell sequencing technologies offer a powerful tool to delve into the expression patterns of individual cell types, which enables us to unravel the intricate cellular heterogeneity and novel therapeutic targets that would otherwise be inaccessible through bulk analysis. Targeting these genes with specific drugs could offer safer and more effective treatment strategies by minimizing cytotoxic effects on non-malignant cells. This underscores the importance of understanding the molecular landscape of ESCC for developing targeted therapies and improving patient outcomes.

The association between IGFBP2 and various cancers remains controversial. It is well known that elevated expression of IGFBP2 is associated with poor prognosis in various cancer types, including pancreatic ductal adenocarcinoma (PDAC), gastric cancer, ovarian cancer, hepatocellular carcinoma, and colorectal cancer [28–30]. When serving as a hub gene of an oncogenic signaling pathway, IGFBP2 primarily mediates downstream cancer-driving signals by promoting IGF-I/II signals and binding to integrins, thus enhancing tumor dissemination and progression [31,32]. In contrast, IGFBP2 seems to play an opposing role in breast cancer. James et al. discovered that IGFBP2 is a potent adipokine secreted by healthy breast adipocytes, which binds and sequesters IGF-II [31]. This finding is consistent with the inhibition of breast cancer xenograft growth by suppressing IGF-II [32]. The results of our study indicate that high IGFBP2 expression in ESCC is implicated in poor prognosis. Exogenous application of recombinant human IGFBP2 further promotes ESCC proliferation, while knockdown of this gene inhibits proliferation, invasion, and migration. These findings, along with previous studies, underscore the importance of investigating the crosstalk between IGFBP2 and the IGF-I/II signaling pathway and distinguishing between intracellular and extracellular pools of IGFBP2 [28,29,31].

IGFBP2 exerts immunosuppressive effects within tumor microenvironment (TME) through multiple mechanisms [33,34]. Zhang et al. reported that the cancer-associated fibroblast (CAF) gene IGFBP2 promotes glioma progression by inducing M2 macrophage polarization [33]. Another study suggests that IGFBP2-mediated upregulation of indoleamine 2,3-dioxygenase (IDO) leads to the accumulation of IDO-derived catabolites, which contribute to regulatory T cell (Treg) differentiation and the establishment of an immunosuppressive TME in PDAC [34]. Given the previous reports on IGFBP2's role in reshaping the TME, we analyzed the RNA sequencing data from the TCGA-ESCA database and found that the expression levels of IGFBP2 were negatively correlated with various immune cells and immune modulators, particularly $CD8^+$ T cells. This conclusion was further validated by the results of IHC staining. These findings suggest that the enrichment of IGFBP2 in ESCC may primarily mediate immune suppression in the TME by influencing the infiltration and function of $CD8^+$ T cells, thereby facilitating immune escape by tumor cells. Additionally, high IGFBP2 expression was associated with low PDCD1 expression, ESTIMATE score, and immune score, indicating that blocking IGFBP2 may potentially transform "cold tumors" into "hot tumors," thereby improving the responsiveness of ESCC patients to immunotherapy [35,36].

However, this study also has the following limitations. Firstly, the potential sources of statistical bias and variability in this study are noteworthy. The sample size of 10 patients is relatively small which restricts the generalizability of our findings. Future studies with larger cohorts are necessary to validate our findings. Additionally, the choice of models and parameters (e.g., the number of trees) can influence the model's performance and its susceptibility to overfitting, although cross-validation and replication were employed to mitigate these effects. Furthermore, selection criteria (such as fold change and AUC) and inherent biological variability among patients are also potential sources of bias. Secondly, while IGFBP2 overexpression is implicated in the growth and invasion of ESCC, it is insufficient on its own to sustain these processes. Inhibition of IGFBP2 function may suppress cancer cell proliferation and invasion, but it is not capable of eradicating the cancer cells. To develop a more comprehensive understanding, it is crucial to conduct functional validation and characterization of other genes upregulated in ESCC, elucidating their synergistic roles in tumor initiation and progression. This approach will enable the design of personalized combination therapy regimens tailored to individual patient profiles. Thirdly, this study employs random forest, which may exhibit potential limitations including susceptibility to overfitting when

parameters are not finely tuned, and a tendency to assign higher importance to variables with greater categorical diversity or broader value range. Additionally, the molecular biological mechanisms underlying the impact of IGFBP2 on ESCC were not deeply investigated in this study. Furthermore, in vivo experiments were not conducted to validate the effects of IGFBP2 inhibitors alone or in combination on prognosis and immune therapy. Nevertheless, we believe that this study has made a certain contribution to exploring potential biomarkers and therapeutic targets for ESCC, as well as revealing tumor heterogeneity and complexity.

Data availability statement

Data used in this study is available upon request. Interested researchers should contact Mengnan Zhao (mnzhao19@fudan.edu.cn) for access.

Ethics declarations

Ethical approval for this study was granted by the Ethics Committee of Zhongshan Hospital, Fudan University, China, with approval number B2021 137R on August 15, 2021. Informed consent was obtained from all patients upon their hospitalization.

CRediT authorship contribution statement

Binyang Pan: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Haochun Shi:** Writing – review & editing, Methodology, Conceptualization. **Guangyao Shan:** Visualization, Data curation. **Gujie Wu:** Data curation. **Kungeng Rao:** Data curation. **Jiaqi Liang:** Data curation. **Xing Jin:** Data curation. **Guoshu Bi:** Data curation. **Mengnan Zhao:** Writing – review & editing, Supervision, Resources, Conceptualization. **Weigang Guo:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Chatgpt to improve and polish the language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

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.

Acknowledgments

This work was funded by the Research Foundation of Shanghai Municipal Health Commission (202140296), the China Postdoctoral Science Foundation (2023M730674), and the Natural Science Foundation of Tibet Autonomous Region (XZ2019ZR-ZY40(Z)).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38078.

References

[1] F. Bray, et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 74 (3) (2024) 229–263.
[2] E. Morgan, et al., The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: new estimates from GLOBOCAN 2020, Gastroenterology 163 (3) (2022) 649–658.e2.
[3] H. Zeng, et al., Disparities in stage at diagnosis for five common cancers in China: a multicentre, hospital-based, observational study, Lancet Public Health 6 (12) (2021) e877–e887.
[4] J.M. Sun, et al., Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study, Lancet 398 (10302) (2021) 759–771.
[5] Y.Y. Janjigian, et al., First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial, Lancet 398 (10294) (2021) 27–40.
[6] Z. Wang, et al., Efficacy and safety of neoadjuvant immunotherapy in surgically resectable esophageal cancer: a systematic review and meta-analysis, Int. J. Surg. 104 (2022) 106767.
[7] X. Zhang, et al., Multifunctional nanoparticles co-loaded with Adriamycin and MDR-targeting siRNAs for treatment of chemotherapy-resistant esophageal cancer, J Nanobiotechnology 20 (1) (2022) 166.
[8] C. Zhan, et al., Landscape of expression profiles in esophageal carcinoma by the Cancer Genome Atlas data, Dis. Esophagus 29 (8) (2016) 920–928.
[9] S. Wang, et al., Low-dose metformin reprograms the tumor immune microenvironment in human esophageal cancer: results of a phase II clinical trial, Clin. Cancer Res. 26 (18) (2020) 4921–4932.

[10] G. Yao, et al., Heterogeneous baseline immune cell infiltration landscape as a predictor of pathological complete response in locally advanced esophageal squamous cell carcinoma (ESCC) following neoadjuvant chemotherapy and immunotherapy: results from a single-arm, phase II clinical trial (SEEK-01), J. Clin. Oncol. 41 (16_suppl) (2023) 4047.
[11] H.Q. Dinh, et al., Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal squamous cell carcinoma microenvironment, Nat. Commun. 12 (1) (2021) 7335.
[12] M. Chen, J. Jiang, J. Hou, Single-cell technologies in multiple myeloma: new insights into disease pathogenesis and translational implications, Biomark. Res. 11 (1) (2023) 55.
[13] F. Li, et al., Single-cell transcriptome analysis reveals the association between histone lactylation and cisplatin resistance in bladder cancer, Drug Resist. Updates 73 (2024) 101059.
[14] Z. Chen, et al., Dissecting the single-cell transcriptome network underlying esophagus non-malignant tissues and esophageal squamous cell carcinoma, EBioMedicine 69 (2021) 103459.
[15] Q. Ren, et al., A fibroblast-associated signature predicts prognosis and immunotherapy in esophageal squamous cell cancer, Front. Immunol. 14 (2023) 1199040.
[16] J. Liang, et al., Signatures of malignant cells and novel therapeutic targets revealed by single-cell sequencing in lung adenocarcinoma, Cancer Med. 11 (11) (2022) 2244–2258.
[17] A. Liaw, M. Wiener, Classification and regression by randomForest, R. News 2 (3) (2002) 18–22.
[18] G. Bi, et al., Polyamine-mediated ferroptosis amplification acts as a targetable vulnerability in cancer, Nat. Commun. 15 (1) (2024) 2461.
[19] J. Liang, et al., MAFF confers vulnerability to cisplatin-based and ionizing radiation treatments by modulating ferroptosis and cell cycle progression in lung adenocarcinoma, Drug Resist. Updates 73 (2024) 101057.
[20] J.N. Weinstein, et al., The cancer Genome Atlas pan-cancer analysis project, Nat. Genet. 45 (10) (2013) 1113–1120.
[21] G. Sturm, F. Finotello, M. List, Immunedeconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data, Methods Mol. Biol. 2120 (2020) 223–232.
[22] P. Charoentong, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (1) (2017) 248–262.
[23] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[24] R Core Team, _R: A Language and Environment for Statistical Computing_, R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.R-project.org/.
[25] L. Chen, et al., RBM4 dictates ESCC cell fate switch from cellular senescence to glutamine-addiction survival through inhibiting LKB1-AMPK-axis, Signal Transduct Target Ther 8 (1) (2023) 159.
[26] M. Martini, et al., PI3K/AKT signaling pathway and cancer: an updated review, Ann. Med. 46 (6) (2014) 372–383.
[27] J. Seoane, R.R. Gomis, TGF-B family signaling in tumor suppression and cancer progression, Cold Spring Harb Perspect Biol 9 (12) (2017).
[28] B. Zhang, et al., Prognostic value of IGFBP2 in various cancers: a systematic review and meta-analysis, Cancer Med. 11 (16) (2022) 3035–3047.
[29] Y. Wang, et al., Spatial transcriptomic analysis of ovarian cancer precursors reveals reactivation of IGFBP2 during pathogenesis, Cancer Res. 82 (24) (2022) 4528–4541.
[30] R. Haschemi, et al., Insulin-like growth factor binding protein-2 (IGFBP2) is a key molecule in the MACC1-mediated platelet communication and metastasis of colorectal cancer cells, Int. J. Mol. Sci. 22 (22) (2021).
[31] J.R.W. Conway, et al., IGFBP2 secretion by mammary adipocytes limits breast cancer invasion, Sci. Adv. 9 (28) (2023) eadg1840.
[32] L.F. Wei, et al., IGFBP2 in cancer: pathological role and clinical significance, Oncol. Rep. 45 (2) (2021) 427–438 (Review).
[33] X. Zhang, et al., Cancer-associated fibroblast-associated gene IGFBP2 promotes glioma progression through induction of M2 macrophage polarization, Am J Physiol Cell Physiol 326 (1) (2024) C252–c268.
[34] L. Sun, et al., IGFBP2 drives regulatory T cell differentiation through STAT3/Ido signaling pathway in pancreatic cancer, J Pers Med 12 (12) (2022).
[35] D. Dyikanov, et al., Comprehensive peripheral blood immunoprofiling reveals five immunotypes with immunotherapy response characteristics in patients with cancer, Cancer Cell 42 (5) (2024) 759–779.e12.
[36] A. Bagaev, et al., Conserved pan-cancer microenvironment subtypes predict response to immunotherapy, Cancer Cell 39 (6) (2021) 845–865.e7.