scientific reports

OPEN

Potential biomarkers and immune characteristics of small bowel adenocarcinoma

Jinggao Feng✉, Xiayu Tang, Liusong Song, Zhipeng Zhou, Yuan Jiang & Yao Huang

Small bowel adenocarcinoma (SBA) is a gastrointestinal malignancy with low incidence but poor prognosis, and its pathogenesis is still unclear. This study aimed to explore potential disease-causing biomarkers of SBA. The gene expression datasets of SBA and normal samples were downloaded from the Gene Expression Omnibus database. First, differential gene expression analysis and weighted gene coexpression network analysis (WGCNA) were performed. Common genes (CGs) were obtained by intersection of differentially expressed genes (DEGs) and optimal modal genes of WGCNA. Subsequently, a protein−protein interaction network was established to screen hub genes, and target genes were obtained by Lasso regression analysis of hub genes. An SBA risk prediction model was established based on target genes. The prediction accuracy of the model was evaluated by the area under the receiver operating characteristic curve (AUC). The levels of immune cell infiltration and activation of immune pathways were compared between SBA and normal samples using the "ggpubr" and "reshape2" packages. A total of 1058 DEGs were identified. WGCNA showed that the signature gene in the brown module was significantly associated with SBA (p = 7E−17), and 469 CGs were obtained. Four target genes (APOA4, APOB, COL1A2, FN1) were identified and showed excellent prediction of SBA risk (AUC = 0.965). In addition, active dendritic cells and macrophages showed higher infiltration levels in SBA. Meanwhile, the APC_co_stimulation pathway and parainflammation pathway were strongly active in SBA. Four target genes (APOA4, APOB, COL1A2, FN1) may be involved in the pathogenesis of small bowel adenocarcinoma.

As a rare tumor, small bowel adenocarcinoma (SBA) accounts for only 3–5% of digestive malignant tumors[1]. In 2021, there were 11,390 new cases of SBA in the United States and 2100 deaths from the disease[2]. Because of the nonspecificity of SBA symptoms, the diagnosis is often due to intestinal perforation, ileus, and uncontrolled gastrointestinal bleeding. Therefore, approximately one-third of patients are diagnosed with distant metastasis[3]. Meanwhile, the prognosis of SBA is poor, especially for patients in an advanced stage[4,5].

Currently, the understanding of the pathogenesis of SBA is still limited. The traditional concept is that the pathogenesis of SBA is similar to that of colon cancer but lacks exact evidence. Several key molecular drivers in the pathogenesis of SBA have been identified by genomic profiling studies, including TP53, SMAD4, KRAS and E-cadherin[6,7]. Fortunately, many studies have reported that the incidence of high microsatellite instability (MSI-H) or dMMR and high tumor mutational burden (TMB) in SBAs is higher than that in gastric cancer and colorectal cancer, suggesting that immunotherapy may be a new therapeutic breakthrough[8–10]. However, the state of immune cell infiltration and immune pathway activation in the immune microenvironment in SBA is still unclear. Hereditary cancer syndromes such as Lynch syndrome, familial adenomatous polyposis and Peutz−Jeghers syndrome are considered to be risk factors for SBA[11,12]. In addition, patients with Crohn's disease and celiac disease are more likely to suffer from SBA[13,14].

To date, few studies on potential biomarkers of SBA have been recorded. In this study, comprehensive bioinformatics methods were utilized to explore the pathogenesis of SBA and screen biomarkers that may have therapeutic value. Meanwhile, the tumor immune microenvironment of SBA was discussed.

Methods

Patient and public involvement. The dataset (GSE61465) downloaded for this study contains 20 normal samples and 25 small bowel adenocarcinoma samples. The GEO database belongs to the public databases. The patients involved in the database obtained ethical approval. Users can download relevant data for free for

Department of Gastrointestinal and Anorectal Surgery, The Central Hospital of Yongzhou, No. 151, Xiaoshui West Road, Lingling District, Yongzhou 425100, Hunan, China. ✉email: 13367358221@163.com

1

research and publish relevant articles. Our study is based on open source data, so there are no ethical issues or other conflicts of interest.

The statistical analysis of this study was completed by R version 4.1.0 (http://www.r-project.org). $P < 0.05$ on both sides was considered statistically significant. The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) was used to download the gene expression dataset for SBA pathological tissues and normal small bowel mucosa samples[15]. Data preprocessing was performed on the dataset using the "limma" and "impute" packages, including conversion of probe names to gene names, missing value filling and data normalization[16]. Deleted data with unrecognized gene names. Genes satisfying FDR < 0.01 and |log2 -fold change (FC)| > 2 were selected through the "limma" package, and these genes were designated differentially expressed genes (DEGs)[17].

Weighted gene coexpression network analysis (WGCNA) was used to explore the interaction between DEGs. The gene coexpression network was constructed by the "WGCNA" package[18]. First, genes with more than 25% variation between samples were introduced into WGCNA. Second, the soft threshold was calculated by the pickSoftThreshold function, and RsquaredCut was set to 0.9[19]. The best soft threshold was chosen, and the adjacency matrix was calculated. Then, the adjacency matrix was converted into a topological overlap matrix (TOM), and the degree of dissimilarity between genes was calculated. Third, gene modules were divided using the dynamic shear tree, the minimum gene module size was set to 50, and then the modules with a dissimilarity coefficient less than 0.2 were merged[20]. Fourth, we selected the module associated with clinical traits, calculated the relationship between genes and traits and modules, and then visualized the characteristic gene network. The intersection genes between DEGs and genes in important modules were defined as common genes (CGs).

CGs were uploaded to the DAVID database (https://david.ncifcrf.gov/tools.jsp) and KOBAS database (http://kobas.cbi.pku.edu.cn/genelist/), and the significantly enriched Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis results were exported, and the results were visualized[21,22]. In addition, CGs were imported into the STRING database (https://www.string-db.org/)[23], protein−protein interaction (PPI) networks were built, the minimum required interaction score was set to 0.9, the results were exported, and Cytoscape Version 3.8.2 was applied to visualize the PPI network[24]. The CytoHubba plug-in in Cytoscape was used to calculate the degree of each CG node and screen out the top 10 hub genes[25].

The "glmnet" package was used for logistic LASSO regression analysis of hub genes, and hub genes with a strong correlation to the risk of SBA were obtained and defined as target genes[26]. The "rms" package was used to draw the nomogram of the model for predicting SBA risk based on the above hub genes[27]. A receiver operating characteristic (ROC) curve was obtained by the "ROCR" package[28]. ROC curves were used to evaluate the prediction accuracy of the model.

The levels of immune cell infiltration and activation of the immune pathway in the SBA and control groups were analyzed by using the "ggpubr" and "reshape2" packages, respectively. In addition, the "corrplot" package was used to explore the relationships between immune cells and between immune pathways.

The miRTarBase, StarBase and TargetScan databases were used to predict the microRNAs (miRNAs) of CGs[29–31]. The miRNAs obtained from the three databases were intersected to obtain the target miRNAs. On the other hand, CGs were imported into the Enrichr database (https://maayanlab.cloud/Enrichr/)[32], and transcription factors (TFs) targeting CGs with p < 0.01 were screened out to obtain the TF-mRNA regulatory network. The above regulatory networks were visualized by Cytoscape.

Ethical approval and consent to participate. The Ethics Committee of The Central Hospital of Yongzhou reviewed the study, and ethics approval was not necessary.

Results

Identification of DEGs. After filtering, one dataset (GSE61465) was downloaded from the GEO database. By analyzing the expression levels of genes in the above dataset, 1058 DEGs were obtained, of which 383 were upregulated and 675 were downregulated.

Identification of gene coexpression networks and modules. First, 25% (5182) of the genes with the largest variance were extracted for subsequent analysis. Second, we defined the threshold to 50 for cluster analysis. Third, R^2 was set to 0.9, and the best soft threshold was 6. Fourth, genes with a dissimilarity coefficient less than 0.2 were combined to obtain 11 modules, and the genes in each module had similar coexpression traits (Fig. 1A). Eleven modules were randomly distinguished by color. The characteristic gene (ME) in the brown module (r = − 0.9; p = 7E−17) showed the highest positive correlation and the most significant correlation with SBA (Fig. 1B). Meanwhile, there was a significant positive relationship between the module members of the genes (MMs, the correlation between specific genes and the characteristic genes of the module) in the brown module and the gene significance (GSs, the correlation between specific genes and clinical variables). A significant correlation (cor = 0.94, P < 1E−200) was observed, as shown in Fig. 1C. Finally, the DEGs and the genes in the brown module were intersected to obtain CGs (Fig. 1D). Furthermore, the brown module was determined to be the key module of SBA, and 653 genes contained in the module were used for the next analysis. The number of genes in each analysis phase is shown in Table 1.

Enrichment analyses of the common genes and hub genes. In biological processes (BP), CGs were mainly enriched in "xenobiotic metabolic process", "retinoid metabolic process", "proteolysis", "fatty acid beta-oxidation" and "retinol metabolic process". CGs are mainly involved in the cell component (CC) ontology, including "extracellular exosome", "apical plasma membrane", "extracellular region", "brush border membrane" and "endoplasmic reticulum lumen". In molecular function (MF), CGs were mostly enriched in "extracellular
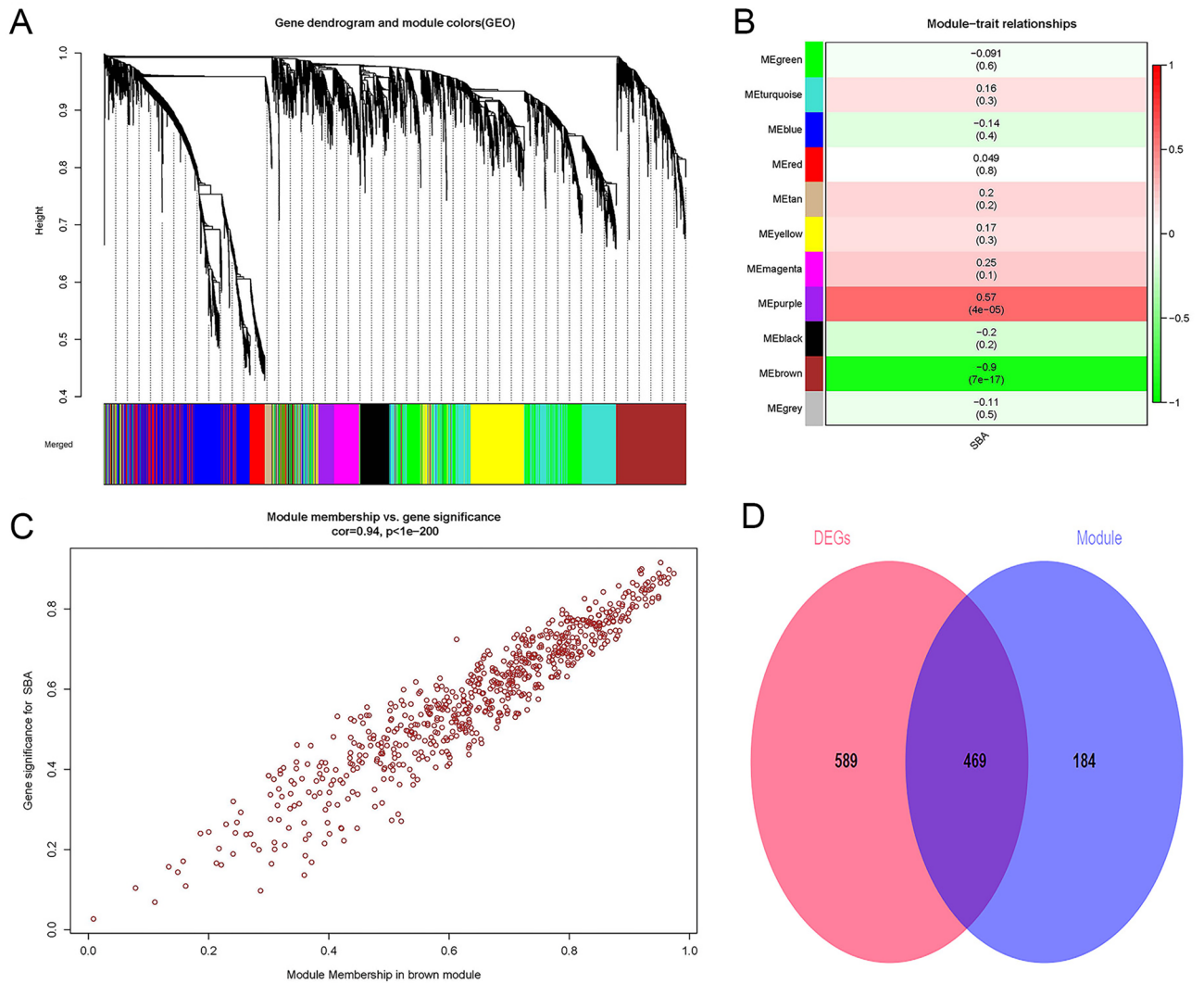
Figure 1. Weighted gene co-expression network analysis and Venn diagram. (**A**) Gene co-expression modules, represented by different colors under the gene tree. (**B**) Heatmap of the association between modules and SBA. The brown module was significantly correlated with SBA. The numbers inside and outside of the brackets represent p-values and correlation coefficients, respectively. (**C**) Correlation plot between MM (X-axis) and (GS) (Y-axis) of genes contained in the blue module. (**D**) Venn diagram showing overlapping genes between the DEGs and the genes in the brown module. *SBA* small bowel adenocarcinoma, *GS* gene significance, *MM* module membership. Color images are available online; DEGs, differentially expressed genes.

Characteristics	Gene number
Original dataset	20,727
DEGs	1058
Input to WGCNA	20,727
Brown gene	653

Table 1. The number of genes in each phase of the analysis. DEGs, differentially expressed genes; WGCNA, weighted gene coexpression network analysis.

matrix structural constituent" and "identical protein binding". In addition, the KEGG enrichment pathways included "Metabolic pathways", "Protein digestion and absorption", "Chemical carcinogenesis", "Retinol metabolism" and "Bile secretion". All enrichment pathways are shown in Fig. 2A. Ten hub genes (APOB, APOC2, APOA4, APOA1, CYP3A4, COL1A2, FN1, DPP4, ACAA2, HADHB) are shown in Fig. 2B.
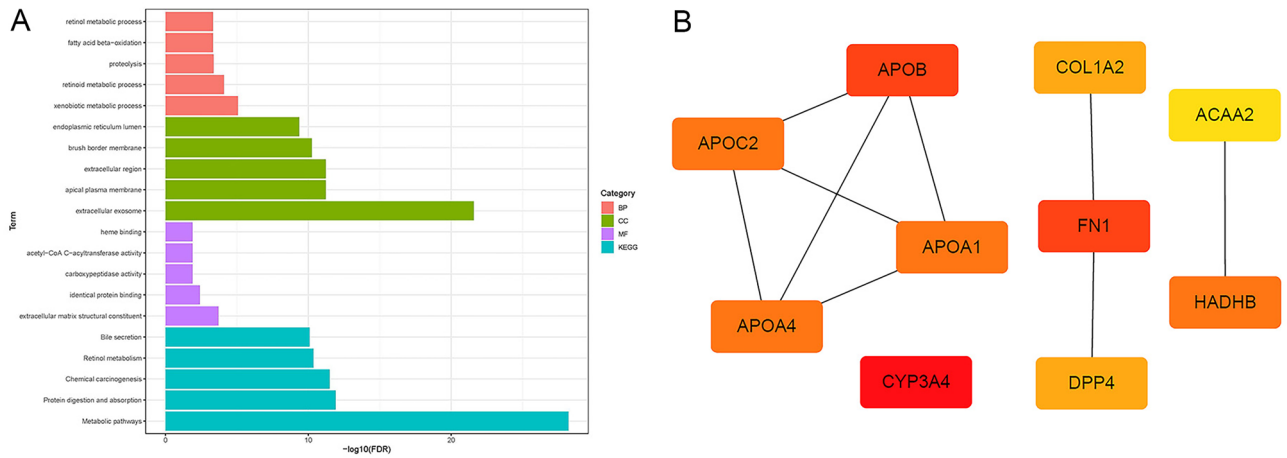
Figure 2. GO analysis, KEGG pathway analysis and hub genes. (**A**) The pink bars represent biological processes, the green bars represent cellular components, the purple lines represent molecular functions, and the blue lines represent KEGG pathways. (**B**) Hub genes. *BP* biological process, *CC* cellular component, *MF* molecular function, *GO* Gene Ontology, *KEGG* Kyoto Encyclopedia of Genes and Genomes.
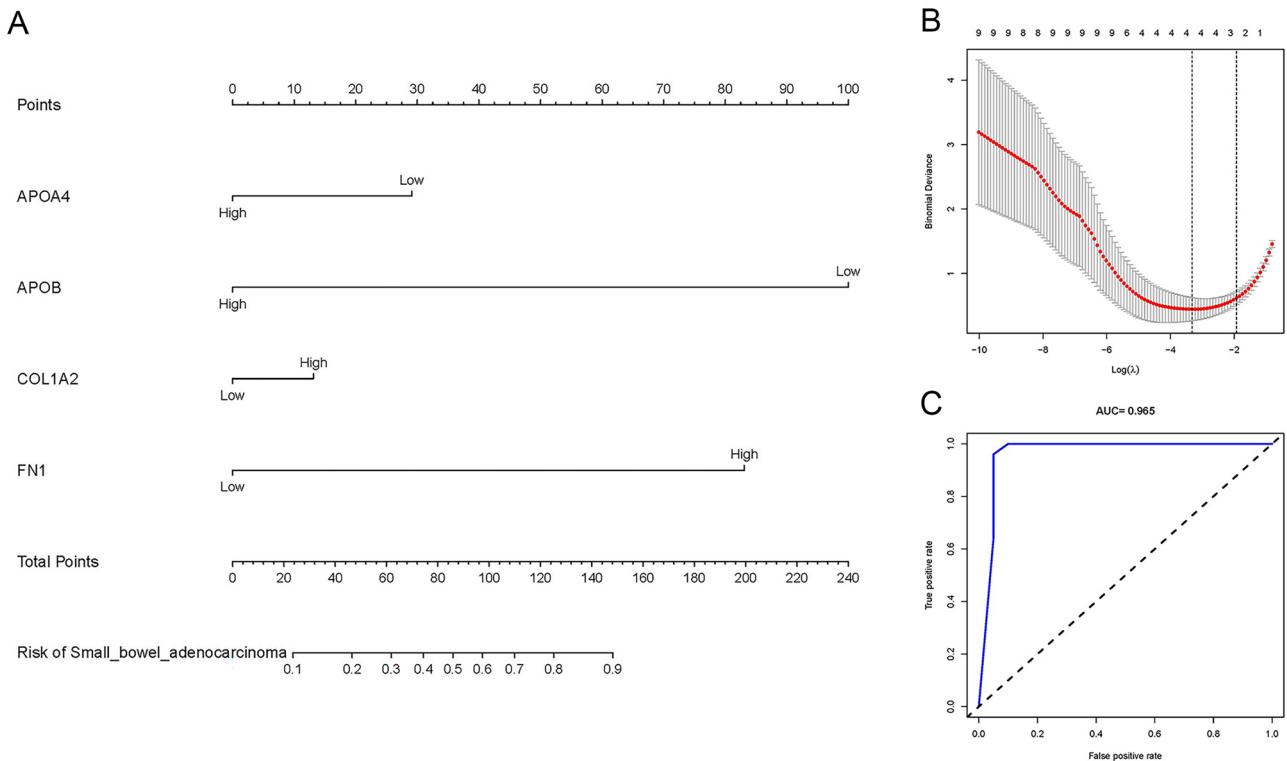

Figure 3. SBA risk prediction model based on 4 target genes. (**A**) SBA risk prediction model based on four target genes. (**B**) Optimal parameter (lambda) selection in the LASSO model used fivefold cross-validation via minimum criteria. The partial likelihood deviance (binomial deviance) curve was plotted versus log(lambda). Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 SE of the minimum criteria (the 1-SE criteria). (**C**) ROC Curves. *SBA* small bowel adenocarcinoma, *LASSO* least absolute shrinkage and selection operator, *SE* standard error, *ROC* receiver operating characteristic.

SBA risk prediction model. Four target genes (APOA4, APOB, COL1A2, FN1) related to the risk of SBA were obtained through LASSO regression analysis, as shown in Fig. 3B. The SBA risk prediction model based on the above four hub genes is shown in Fig. 3A. The ROC curve is shown in Fig. 3C. The area under the curve (AUC) was 0.965, indicating that the model has excellent prediction accuracy.

Immune infiltration analysis. Figure 4A,B show the relationship between immune cells and between immune pathways, respectively. In Fig. 4A, the infiltration of tumor infiltrating lymphocytes (TILs) was posi-
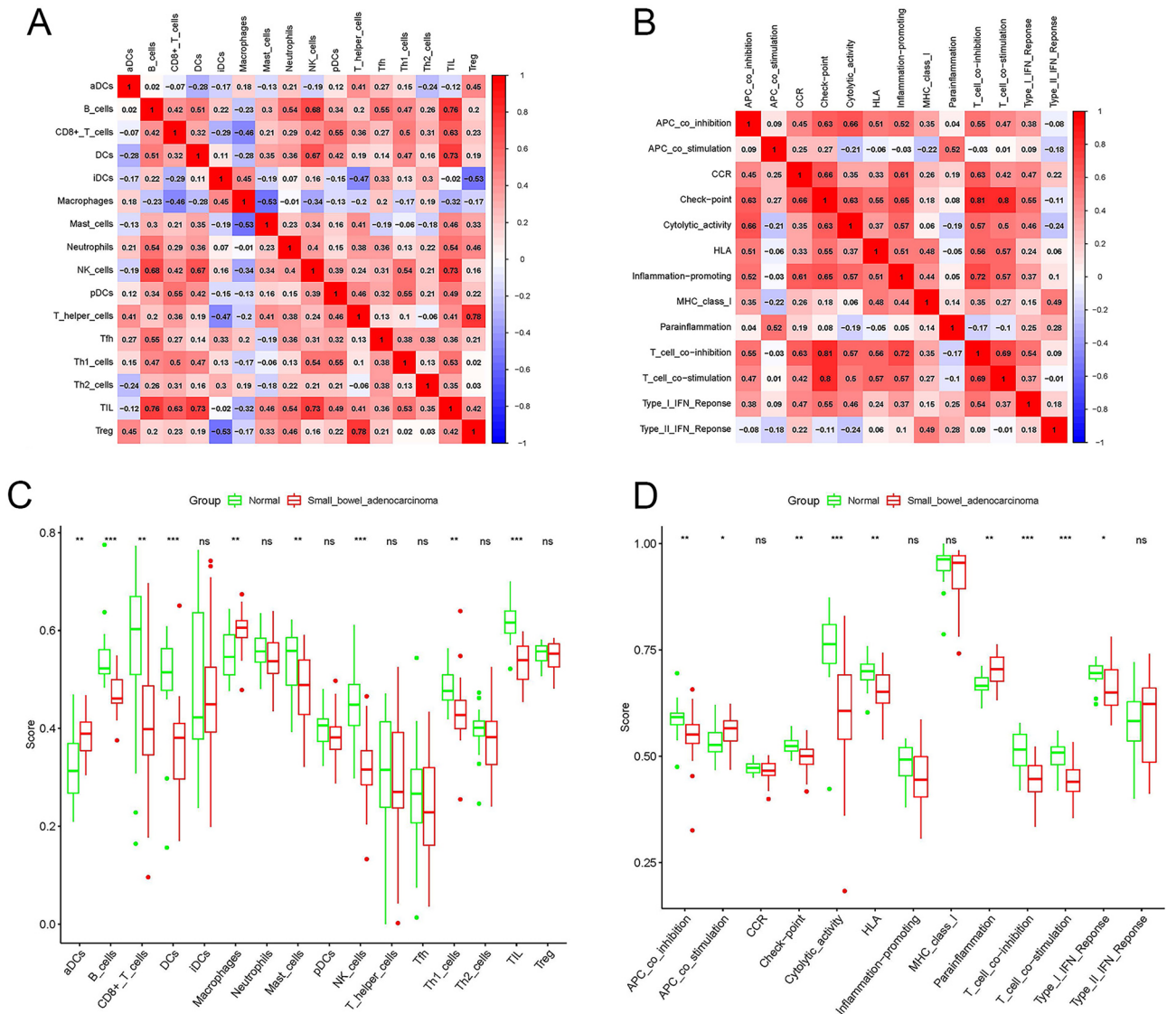
Figure 4. Characteristics of the immune microenvironment in SBA. (**A**) Correlation between immune cells in SBA. Red represents a positive correlation, blue represents a negative correlation. A deeper color indicates a stronger correlation. (**B**) Correlation between immune pathways in SBA. Red represents a positive correlation, blue represents a negative correlation. A deeper color indicates a stronger correlation. (**C**) Comparison of the enrichment scores of 16 types of immune cells between the SBA (red box) and normal group (green box). (**D**) Comparison of the enrichment scores of 13 immune-related pathways between the SBA (red box) and normal group (blue box). *SBA* small bowel adenocarcinoma, P values: *ns* not significant; *P < 0.05; **P < 0.01; ***P < 0.001.

tively correlated with the infiltration of B cells, dendritic cells (DCs), and natural killer (NK) cells. In addition, T regulatory (Treg) cell infiltration was positively correlated with T helper cell infiltration. The check-point pathway was positively correlated with the T_cell_co-inhibition pathway and T_cell_co-stimulation pathway, as shown in Fig. 4B. Figure 4C,D show the level of immune cell infiltration and activation of immune pathways in the SBA and control groups, respectively. Compared with the control group, the infiltration levels of B cells, CD8+-T cells, DCs, mast cells, NK cells, type 1 T helper (Th1) cells and TILs were lower in SBA. However, active dendritic cells (aDCs) and macrophages showed higher infiltration levels in SBA. Meanwhile, in the control group, the activation of the APC_co_inhibition pathway, Check-point pathway, Cytolytic_activity pathway, Human leukocyte antigen (HLA) pathway, T_cell_co-inhibition pathway, T_cell_co-stimulation pathway and Type_I_IFN_Response pathway was stronger than that of SBA. Activation of the APC_co_stimulation pathway and the parainflammation pathway of SBA was stronger than that of the control group.

Target miRNAs and TF-mRNA regulatory network analysis. Figure 5A shows the Venn diagram of predicted miRNAs. This study predicted 435 target miRNAs that may be involved in SBA occurrence (Supple-
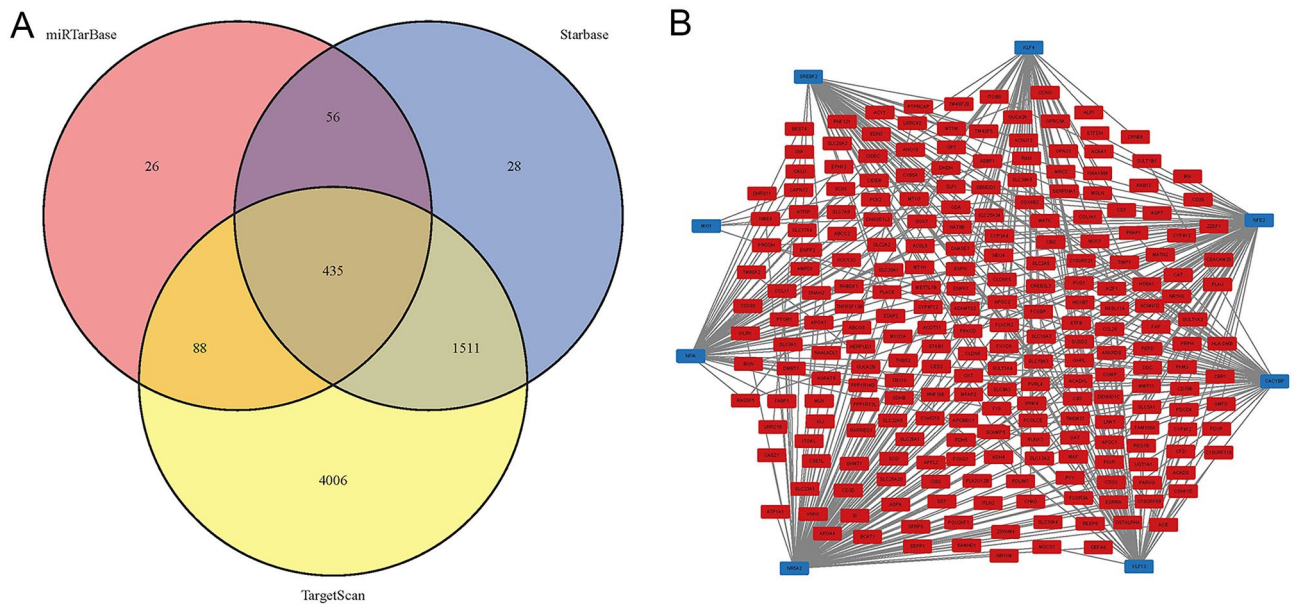
Figure 5. The target miRNA and TF-mRNA regulatory network. (**A**) The Venn diagram of target miRNAs. (**B**) Red represents genes, blue represents TFs. *mRNAs* messenger RNAs, *miRNAs* microRNAs, *TFs* transcription factors.

mentary Material S1). Meanwhile, 8 TFs (NFIA, KLF13, MXI1, CACYBP, NFE2, SREBF2, KLF4, NR5A2) that may be involved in the pathogenesis of SBA are shown in Fig. 5B.

Discussion

SBA is a malignant tumor with low incidence but poor prognosis. In this study, the messenger RNAs (mRNAs), miRNAs and TFs that may be associated with its occurrence were predicted using a comprehensive bioinformatics analysis. Meanwhile, the SBA risk prediction model based on hub genes with good prediction accuracy was established. Finally, analysis of the tumor immune microenvironment suggested that the invasion level of most immune cells in SBA was low, and the activation intensity of the immune pathway was weak, which might be the cause of tumor progression.

Hub genes have attracted extensive attention as potential drug targets. In this study, four hub genes (APOA4, APOB, COL1A2, FN1) were found to be significantly related to the pathogenesis of SBA. Apolipoprotein A4 (APOA4) encodes apolipoprotein A-IV, which is hydrolyzed and glycosylated to produce acidic glycoproteins mainly found in chylomicrons (CMs), very low-density lipoprotein (VLDL), and high-density lipoprotein (HDL)[33–35]. It plays an important role in lipid transport and metabolism, especially in cholesterol reversal[36]. Recent studies showed that APOA4 was significantly overexpressed in Helicobacter pylori-infected atrophic gastritis and intestinal metaplasia tissues, as well as gastric cancer tissues[37]. Furthermore, APOA4 is considered a diagnostic marker for colorectal cancer[38,39]. As a metabolic gene, apolipoprotein B (APOB) is the most important apolipoprotein on chylomicrons and low-density lipoprotein[40,41]. APOB has been confirmed to be associated with the pathogenesis of a variety of gastrointestinal malignancies, including liver cancer, gallbladder cancer, esophageal cancer and pancreatic duct adenocarcinoma[42–45]. Abdominal obesity has been identified as a risk factor for SBA[46]. In this study, APOA4 and APOB were significantly downregulated in SBA, suggesting that lipid metabolism disorders may play an important role in the occurrence of SBA. In addition, functional enrichment analysis showed that CGs were enriched in multiple metabolically related pathways. Unfortunately, no studies have been found on APOA4 or APOB and SBA. Therefore, further research is necessary to clarify the role of lipid metabolism in SBA.

Collagen, Type I, Alpha 2 (COL1A2) is distributed in collagen and cytoplasm and is involved in bone development and the signal transduction pathway of transmembrane receptor protein tyrosine kinase[47,48]. COL1A2 was confirmed to be significantly overexpressed in gastric cancer tissues[49–51]. Similarly, COL1A2 was believed to be significantly overexpressed in colorectal cancer tissues and blood samples, but the specific mechanism remains unclear[52,53]. Increasing evidence shows that COL1A2 is considered to be a diagnostic and prognostic biomarker due to its significant upregulation in many cancers[54–56]. In contrast, COL1A2 was significantly downregulated in bladder cancer, malignant melanoma and head and neck cancer[57–59]. In this study, COL1A2 was significantly overexpressed in SBA. Fibronectin 1 (FN1) is a glycoprotein distributed in the extracellular matrix and plays an important role in carcinogenesis and metastasis[60–62]. FN1 expression was upregulated by the transcription factor CP2, which is involved in the metastasis of hepatocellular carcinoma[63]. FN1 promotes ovarian cancer metastasis by activating the PI3K/Akt pathway[64]. High expression of FN1 has been shown to be carcinogenic in esophageal cancer[65]. In addition, FN1, as an oncogene, is involved in aggressive and poor prognosis of colon cancer[66]. As

expected, high expression of FN1 was significantly associated with poorer prognosis of gastric cancer[67,68]. This study suggests that the highly expressed FN1 plays an important role in the carcinogenesis of SBA.

As innate immune cells, macrophages play an important role in the tumor microenvironment. High macrophage infiltration has been associated with tumor progression or poor prognosis in a variety of solid tumors, including neck squamous cell carcinoma, gliomas, breast cancer, bladder cancer, prostate cancer, and melanoma[69–74]. Conversely, high macrophage infiltration was associated with a better prognosis of gastrointestinal malignancies, such as gastric cancer and colorectal cancer[75]. This study suggested that the level of macrophage infiltration in SBA was low, which may be one of the reasons for the poor prognosis of SBA. In addition, the low infiltration of most immune cells and the weak activation of immune pathways may be important factors for the occurrence and progression of SBA.

Although the present study is novel and rigorous, there are still some shortcomings. First, the data used in this study were obtained from a common public database, and because of the low prevalence of SBA, we were unable to collect sufficient samples for experiments to validate the results of this study. Second, due to the lack of prognostic information on SBA, our study could not be further investigated in the context of prognosis. In view of this, we hope that more studies in the future will further validate our results and investigate them in depth.

In conclusion, four mRNAs (APOA4, APOB, COL1A2, FN1) were predicted to be associated with the occurrence of SBA, and an excellent SBA risk prediction model was established based on these genes. Meanwhile, in SBA, it is speculated that the infiltration level of immune cells was low and the activation state of immune pathways was weak. Finally, TFs and miRNAs that may be involved in the pathogenesis of SBA were predicted.

Discussion
This study predicted that four target mRNAs (APOA4, APOB, COL1A2, FN1) might be involved in the occurrence and progression of SBA. In addition, low infiltration of immune cells and weak activation of immune pathways may be immunological characteristics of SBA.

Data availability
The datasets (GSE61465) generated and analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61465).

References
1. Moati, E., Overman, M. J. & Zaanan, A. Therapeutic strategies for patients with advanced small bowel adenocarcinoma: Current knowledge and perspectives. *Cancers* **14**(5), 1 (2022).
2. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**(1), 7–33 (2021).
3. Aparicio, T. *et al.* Small bowel adenocarcinoma: Results from a nationwide prospective ARCAD-NADEGE cohort study of 347 patients. *Int. J. Cancer* **147**(4), 967–977 (2020).
4. Howe, J. R., Karnell, L. H., Menck, H. R. & Scott-Conner, C. The American College of Surgeons Commission on Cancer and the American Cancer Society: Adenocarcinoma of the small bowel: review of the National Cancer Data Base, 1985–1995. *Cancer* **86**(12), 2693–2706 (1999).
5. Brueckl, W. M. *et al.* Prognostic significance of microsatellite instability in curatively resected adenocarcinoma of the small intestine. *Cancer Lett.* **203**(2), 181–190 (2004).
6. Raghav, K. & Overman, M. J. Small bowel adenocarcinomas–existing evidence and evolving paradigms. *Nat. Rev. Clin. Oncol.* **10**(9), 534–544 (2013).
7. Lee, H. J. *et al.* Combined loss of E-cadherin and aberrant beta-catenin protein expression correlates with a poor prognosis for small intestinal adenocarcinomas. *Am. J. Clin. Pathol.* **139**(2), 167–176 (2013).
8. Schrock, A. B. *et al.* genomic profiling of small-bowel adenocarcinoma. *JAMA Oncol.* **3**(11), 1546–1553 (2017).
9. Laforest, A. *et al.* ERBB2 gene as a potential therapeutic target in small bowel adenocarcinoma. *Eur. J. Cancer* **50**(10), 1740–1746 (2014).
10. Hänninen, U. A. *et al.* Exome-wide somatic mutation characterization of small bowel adenocarcinoma. *PLoS Genet.* **14**(3), e1007200 (2018).
11. Pandya, K., Overman, M. J. & Gulhati, P. molecular landscape of small bowel adenocarcinoma. *Cancers* **14**(5), 1 (2022).
12. Tsuboi, A. *et al.* Genomic analysis for the prediction of prognosis in small-bowel cancer. *PLoS ONE* **16**(5), e0241454 (2021).
13. Axelrad, J. E. *et al.* Inflammatory bowel disease and risk of small bowel cancer: A binational population-based cohort study from Denmark and Sweden. *Gut* **70**(2), 297–308 (2021).
14. Emilsson, L., Semrad, C., Lebwohl, B., Green, P. H. R. & Ludvigsson, J. F. Risk of small bowel adenocarcinoma, adenomas, and carcinoids in a nationwide cohort of individuals with celiac disease. *Gastroenterology* **159**(5), 1686–1694 (2020).
15. Clough, E. & Barrett, T. The gene expression omnibus database. *Methods Mol. Biol.* **1418**, 93–110 (2016).
16. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**(2), 249–264 (2003).
17. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
18. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf.* **9**, 559 (2008).
19. Egger, J. *et al.* Square-cut: A segmentation algorithm on the basis of a rectangle shape. *PLoS ONE* **7**(2), e31064 (2012).
20. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**(5586), 1551–1555 (2002).
21. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**(1), 44–57 (2009).
22. Wu, J., Mao, X., Cai, T., Luo, J. & Wei, L. KOBAS server: A web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* **34**, W720–W724 (2006).
23. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

24. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504 (2003).

25. Chin, C. H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* **8**(Suppl 4), S11 (2014).

26. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenet.* **11**(1), 123 (2019).

27. Nunez, E., Steyerberg, E. W. & Nunez, J. Regression modeling strategies. *Rev. Esp. Cardiol.* **64**(6), 501–507 (2011).

28. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**(20), 3940–3941 (2005).

29. Chou, C. H. *et al.* miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**(D1), D296–D302 (2018).

30. Yang, J. H. *et al.* starBase: A database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* **39**, D202–D209 (2011).

31. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, 1 (2015).

32. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**(W1), W90–W97 (2016).

33. Juneja, R. K., Gahne, B., Lukka, M. & Ehnholm, C. A previously reported polymorphic plasma protein of dogs and horses, identified as apolipoprotein A-IV. *Anim. Genet.* **20**(1), 59–63 (1989).

34. Kamalam, B. S. *et al.* Selection for high muscle fat in rainbow trout induces potentially higher chylomicron synthesis and PUFA biosynthesis in the intestine. *Comp. Biochem. Physiol. A: Mol. Integr. Physiol.* **164**(2), 417–427 (2013).

35. Carmena-Ramón, R., Ascaso, J. F., Real, J. T., Ordovas, J. M. & Carmena, R. Genetic variation at the apoA-IV gene locus and response to diet in familial hypercholesterolemia. *Arterioscler. Thromb. Vasc. Biol.* **18**(8), 1266–1274 (1998).

36. Harney, D. J. *et al.* Proteomic analysis of human plasma during intermittent fasting. *J. Proteome Res.* **18**(5), 2228–2240 (2019).

37. Yin, H., Chu, A., Liu, S., Yuan, Y. & Gong, Y. Identification of DEGs and transcription factors involved in H pylori-associated inflammation and their relevance with gastric cancer. *PeerJ* **8**, e9223 (2020).

38. Voronova, V. *et al.* Diagnostic value of combinatorial markers in colorectal carcinoma. *Front. Oncol.* **10**, 832 (2020).

39. Ahn, S. B. *et al.* Potential early clinical stage colorectal cancer diagnosis using a proteomics blood test panel. *Clin. Proteom.* **16**, 34 (2019).

40. Mahley, R. W., Hui, D. Y., Innerarity, T. L. & Weisgraber, K. H. Two independent lipoprotein receptors on hepatic membranes of dog, swine, and man: Apo-B, E and apo-E receptors. *J. Clin. Investig.* **68**(5), 1197–1206 (1981).

41. Berman, M. *et al.* Metabolsim of apoB and apoC lipoproteins in man: Kinetic studies in normal and hyperlipoproteininemic subjects. *J. Lipid Res.* **19**(1), 38–56 (1978).

42. Chen, H. *et al.* ApoB/ApoA-1 ratio as a novel prognostic predictor in patients with primary small cell carcinoma of the esophagus. *Front. Oncol.* **10**, 610 (2020).

43. Gong, Y., Zhang, L., Bie, P. & Wang, H. Roles of ApoB-100 gene polymorphisms and the risks of gallstones and gallbladder cancer: a meta-analysis. *PLoS ONE* **8**(4), e61456 (2013).

44. Cefalù, A. B. *et al.* A novel APOB mutation identified by exome sequencing cosegregates with steatosis, liver cancer, and hypocholesterolemia. *Arterioscler. Thromb. Vasc. Biol.* **33**(8), 2021–2025 (2013).

45. Pandey, S. N., Srivastava, A., Dixit, M., Choudhuri, G. & Mittal, B. Haplotype analysis of signal peptide (insertion/deletion) and XbaI polymorphisms of the APOB gene in gallbladder cancer. *Liver Int.* **27**(7), 1008–1015 (2007).

46. Lu, Y. *et al.* Comparison of abdominal adiposity and overall obesity in relation to risk of small intestinal cancer in a European Prospective Cohort. *Cancer Causes and Control CCC.* **27**(7), 919–927 (2016).

47. Zhytnik, L. *et al.* Mutational analysis of COL1A1 and COL1A2 genes among Estonian osteogenesis imperfecta patients. *Hum. Genomics* **11**(1), 19 (2017).

48. Warnecke, C. *et al.* Adenovirus-mediated overexpression and stimulation of the human angiotensin II type 2 receptor in porcine cardiac fibroblasts does not modulate proliferation, collagen I mRNA expression and ERK1/ERK2 activity, but inhibits protein tyrosine phosphatases. *J. Mol. Med. (Berl).* **79**(9), 510–521 (2001).

49. Rong, L. *et al.* COL1A2 is a novel biomarker to improve clinical prediction in human gastric cancer: Integrating bioinformatics and meta-analysis. *Pathol. Oncol. Res. POR.* **24**(1), 129–134 (2018).

50. Ao, R., Guan, L., Wang, Y. & Wang, J. N. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. *J. Cell. Biochem.* **119**(6), 4420–4434 (2018).

51. Zhuo, C. *et al.* Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. *Cell. Physiol. Biochem.* **40**(6), 1316–1324 (2016).

52. Zou, X. *et al.* Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis. *J. Proteomics* **94**, 473–485 (2013).

53. Rodia, M. T. *et al.* Systematic large-scale meta-analysis identifies a panel of two mRNAs as blood biomarkers for colorectal cancer detection. *Oncotarget* **7**(21), 30295–30306 (2016).

54. Ji, J. *et al.* Let-7g targets collagen type I alpha2 and inhibits cell migration in hepatocellular carcinoma. *J. Hepatol.* **52**(5), 690–697 (2010).

55. Wu, Y. H., Chang, T. H., Huang, Y. F., Huang, H. D. & Chou, C. Y. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* **33**(26), 3432–3440 (2014).

56. Shintani, Y., Hollingsworth, M. A., Wheelock, M. J. & Johnson, K. R. Collagen I promotes metastasis in pancreatic cancer by activating c-Jun NH(2)-terminal kinase 1 and up-regulating N-cadherin expression. *Can. Res.* **66**(24), 11745–11753 (2006).

57. Misawa, K. *et al.* Hypermethylation of collagen alpha2 (I) gene (COL1A2) is an independent predictor of survival in head and neck cancer. *Cancer Biomark. Sect. A Dis. Mark.* **10**(3–4), 135–144 (2011).

58. Mori, K. *et al.* CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. *Int. J. Oncol.* **34**(6), 1593–1602 (2009).

59. Bonazzi, V. F. *et al.* Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. *PLoS ONE* **6**(10), e26121 (2011).

60. Jerhammar, F. *et al.* Fibronectin 1 is a potential biomarker for radioresistance in head and neck squamous cell carcinoma. *Cancer Biol. Ther.* **10**(12), 1244–1251 (2010).

61. Zhai, J. & Luo, G. GATA6-induced FN1 activation promotes the proliferation, invasion and migration of oral squamous cell carcinoma cells. *Mol. Med. Rep.* **25**(3), 1 (2022).

62. Sun, W. *et al.* The NEAT1_2/miR-491 axis modulates papillary thyroid cancer invasion and metastasis through the TGM2/NFκb/FN1 signaling. *Front. Oncol.* **11**, 610547 (2021).

63. Xu, X. *et al.* Characterization of genome-wide TFCP2 targets in hepatocellular carcinoma: implication of targets FN1 and TJP1 in metastasis. *J. Exp. Clin. Cancer Res. CR.* **34**(1), 6 (2015).

64. Thant, A. A. *et al.* Fibronectin activates matrix metalloproteinase-9 secretion via the MEK1-MAPK and the PI3K-Akt pathways in ovarian cancer cells. *Clin. Exp. Metastasis.* **18**(5), 423–428 (2000).

65. Song, G. *et al.* SATB1 plays an oncogenic role in esophageal cancer by up-regulation of FN1 and PDGFRB. *Oncotarget* **8**(11), 17771–17784 (2017).
66. Wu, J. *et al.* Transcriptional activation of FN1 and IL11 by HMGA2 promotes the malignant behavior of colorectal cancer. *Carcinogenesis* **37**(5), 511–521 (2016).
67. Han, C. *et al.* Identification of the hub genes RUNX2 and FN1 in gastric cancer. *Open Med. (Warsaw, Poland).* **15**(1), 403–412 (2020).
68. Li, L. *et al.* FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci. Rep.* **9**(1), 7827 (2019).
69. Zhang, Y. *et al.* High-infiltration of tumor-associated macrophages predicts unfavorable clinical outcome for node-negative breast cancer. *PLoS ONE* **8**(9), e76147 (2013).
70. Xue, Y. *et al.* Tumorinfiltrating M2 macrophages driven by specific genomic alterations are associated with prognosis in bladder cancer. *Oncol. Rep.* **42**(2), 581–594 (2019).
71. Kumar, A. T. *et al.* Prognostic significance of tumor-associated macrophage content in head and neck squamous cell carcinoma: A meta-analysis. *Front. Oncol.* **9**, 656 (2019).
72. Nishie, A. *et al.* Macrophage infiltration and heme oxygenase-1 expression correlate with angiogenesis in human gliomas. *Clin. Cancer Res.* **5**(5), 1107–1113 (1999).
73. Torisu, H. *et al.* Macrophage infiltration correlates with tumor stage and angiogenesis in human malignant melanoma: Possible involvement of TNFalpha and IL-1alpha. *Int. J. Cancer* **85**(2), 182–188 (2000).
74. Cao, J. *et al.* Prognostic role of tumour-associated macrophages and macrophage scavenger receptor 1 in prostate cancer: A systematic review and meta-analysis. *Oncotarget* **8**(47), 83261–83269 (2017).
75. Cortese, N., Carriero, R., Laghi, L., Mantovani, A. & Marchesi, F. Prognostic significance of tumor-associated macrophages: Past, present and future. *Semin. Immunol.* **48**, 101408 (2020).

Author contributions
J.F., X.T. and L.S. participated in the concept and design, Z.Z. and Y.J. data screening and analysis, and J.F. and Y.H. manuscript drafting. All authors agree to publish.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20599-5.

Correspondence and requests for materials should be addressed to J.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.