Research article

Development of a novel prognostic signature for colorectal cancer based on angiogenesis-related genes

Aiqin Chen [a], Kailai Wang [b,c,d], Lina Qi [b,c,d], Wangxiong Hu [b,d,*], Biting Zhou [b,c,d,e,**]

[a] Department of Nursing, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China
[b] Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[c] Department of Medical Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China
[d] Cancer Center, Zhejiang University, Hangzhou, Zhejiang, 310058, China
[e] Department of Radiation Oncology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China

ARTICLE INFO

ABSTRACT

Background: Colorectal cancer (CRC) is the third most common malignant tumor worldwide. Angiogenesis is closely related to tumor metastasis, which is the main cause of cancer death. Although several angiogenesis signatures have been proposed in some cancer types, no angiogenic signature has been developed to predict the prognosis and efficacy of antiangiogenic bevacizumab in CRC patients.

Methods: We developed a novel CRC angiogenic signature by refining seven publicly available angiogenic gene sets using least absolute shrinkage and selection operator (LASSO). Immune and stromal cells within the tumor microenvironment were compared between the high- and low-risk groups in more than 1000 CRC samples classified by calculating the risk score based on the customized angiogenic signature. The correlation of this new gene set with the efficacy of bevacizumab was also compared.

Results: A new prognostic-associated angiogenesis signature gene set was constructed that can divide CRC patients into two high- and low-risk groups. The high-risk angiogenic group was significantly associated with extracellular matrix organization, epithelial-mesenchymal transition (EMT), and myogenesis. In addition, the high-risk group had higher infiltration of stromal and immune cells and was more resistant to bevacizumab than the low-risk group.

Conclusion: Briefly, we constructed a novel angiogenic signature that can predict the prognosis of CRC patients and the efficacy of bevacizumab in treating CRC. Our results provide new insights into the relationships among angiogenesis, metastasis, and medication for CRC.

* Corresponding author. Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.
** Corresponding author. Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.
E-mail addresses: wxhu@zju.edu.cn (W. Hu), zhoubiting@zju.edu.cn (B. Zhou).

1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide, followed by breast and lung cancer, and is the major cause of cancer death [1]. Despite the relatively good prognosis of early-stage CRC, metastasis occurs in ~50 % of CRC patients and the 5-year survival rate decreases to less than 20 %, which leads to the high mortality of CRC patients [2]. Angiogenesis, defined as the formation of new vascular networks from preexisting vessels, putatively plays a crucial role in supporting tumor growth and metastasis by transporting more oxygen and nutrients to malignant tumor cells [3], which offers great clinical potential for tumor therapy. Numerous proangiogenic factors, including vascular endothelial growth factor (VEGF), platelet-derived growth factor, fibroblast growth factor-2, angiopoietins, ephrin, apelin, chemokines, matrix metalloproteases, tissue necrosis factor-α and pleio-trophin, have been reported to promote neovascularization [4]. Therefore, a variety of antiangiogenic therapies have been developed, such as bevacizumab, a humanized monoclonal antibody targeting VEGF-A, and several multitarget tyrosine kinase inhibitors, including sunitinib, sorafenib, pazopanib and vandetanib, which mainly block VEGF receptors [5]. In addition to bevacizumab, aflibercept, ramucirumab, and regorafenib, antiangiogenic drugs have been approved by the FDA/EMA and have demonstrated ef-ficacy in treating metastatic colorectal cancer (mCRC) [6]. However, many tumor patients benefit less from VEGF-targeted therapy due to drug resistance and a lack of effective antiangiogenic biomarkers for efficacy prediction and risk stratification [7].

Several angiogenic signatures have been proposed such as the angiogenic 62 gene expression signature in non-small cell lung cancer [8], 32 gene expression signature in aggressive endometrial cancer [9], nine gene expression signature induced by EGF and hypoxia in Caco-2 CRC cells [10], a core signature of 43 overexpressed genes that were consistently top ranked in different tumors [11], 34 gene expression signature in advanced ovarian carcinoma [12], 36 genes upregulated during the formation of blood vessels collected by MSigDB Team [13], and 100 angiogenic gene expression signature used to predict serous ovarian cancer subtypes [14]; unfortunately, there is no impeccable CRC-derived angiogenic signature available. It is difficult to risk stratify CRC patients based on these publicly available angiogenic signatures. Therefore, we reconstructed a novel angiogenic signature by least absolute shrinkage and selection operator (LASSO) feature selection that could predict the prognosis of more than 1000 CRC patients in the TCGA, GSE39582 and ZUCI (Zhejiang University Cancer Institute) datasets. We found that the high-risk CRC cohort had poorer outcomes than the low-risk cohort and was closely associated with epithelial-mesenchymal transition (EMT) and myogenesis. In addition, higher stromal and immune infiltration was observed in the high-risk angiogenic group. Finally, the high-risk angiogenic group was found to be resistant to bevacizumab but warrants further investigation.

2. Materials and methods

2.1. Somatic mutation data for CRCs

CRC somatic mutational profiles and clinical information were downloaded from The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/, February 06, 2018). TCGA aims to assess the feasibility of a full-scale effort to systematically explore the entire spectrum of genomic changes involved in human cancer. Silent mutations, RNA mutations, and any mutation located within the intron, flanking sequence, 5′ untranslated region (UTR), or 3′UTR were discarded. Then, the clinical information of each patient was added to the mutational information via a unique sample ID. The clinical information is listed in Supplementary Table 1. The oncoplots were constructed by using maftools [15].

2.2. Curation of the angiogenic signature in CRC

To construct a robust CRC angiogenic signature, we collected a list of well-annotated angiogenic gene expression signatures that are highly cited across different cancer types, including the well-known core human primary tumor angiogenesis 43-gene [11], 62-gene angiogenic signature in non-small cell lung cancer [8], 32-gene signature in aggressive endometrial cancer [9], nine angiogenic genes identified in Caco-2 CRC cells [10], 34 gene profile and 100 gene signature in advanced ovarian carcinoma [12,14], and the widely used 36-gene angiogenesis hallmark in MSigDB [13]. Then, batch univariate Cox regression was used to select significant angiogenic variables from these angiogenic union genesets, and subsequently, LASSO was used to develop an angiogenic signature in CRC. The angiogenic score of each CRC sample was calculated by the ssGSEA method implemented in the GSVA package based on the selected gene sets [16].

2.3. Gene expression data processing and normalization

The level 3 mRNA RNASeqV2 datasets for the CRC samples were downloaded from the TCGA (09/15/2018). Genes with expression levels <1 (RSEM-normalized) in more than 50 % of the samples were removed. The raw CEL files of the GSE39582 (Affymetrix HGU133 Plus 2.0 arrays) dataset were downloaded from the public functional genomics data repository Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The MAS5 algorithm was used to determine the gene expression levels as previously described [17]. Analysis of differentially expressed genes (DEGs) was performed using the *DEGSeq* package for R/Bioconductor [18]. Significant DEGs were selected according to a false discovery rate (FDR)-adjusted *P* value < 0.05 and a fold change >2. Heatmaps were generated using the *pheatmap* package in R (64-bit, version 3.0.2).

2.4. Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the *clusterProfiler* package from Bioconductor [19]. Significantly enriched GO terms and KEGG pathways were selected according to an FDR-adjusted *P* value < 0.05. Hallmark gene sets from the Molecular Signatures Database (MSigDB) [20] were used to determine whether any signatures were enriched under highly angiogenic conditions by gene set enrichment analysis (GSEA) [21]. Significantly enriched hallmarks were chosen according to an FDR *q* value < 0.05.

2.5. Deciphering the tumor microenvironment (TME) in CRCs

The stromal and immune scores were calculated by the *estimate* package in R [22]. The infiltration of 64 immune and stromal cells in malignant tumor tissues was assessed by the xCell method [23].

2.6. RNA isolation and microarray in 60 ZUCI CRC samples

mRNA was extracted from 62 CRC freeze fresh tissue samples using TRIzol following a standard protocol. Only the 60 qualified mRNA samples were retained in the microarray. An Agilent Technologies G2600D SG11430001 system was used to scan the fluorescence signal. Background correction, normalization, and summarization of single probes for all probe sets were carried out by GeneSpring.

2.7. Survival analysis

The genes that correlated with CRC patient survival in the multivariate Cox regression analysis were determined using the LASSO method, as described in our previous study [24]. The best λ was determined by 10-fold cross-validation using the *glmnet* package built-in function *cv.glmnet* [25]. Additionally, to confirm the robustness of the newly curated model, a random forest method was used to detect the importance of the variables included in the model with the function *rfsrc* in the R package *randomForestSRC*. Then, we divided the patients into high- and low-risk groups by calculating the prognostic index (PI) using Equation (1) as follows:

$$PIk = \sum_{g=1}^{n} \beta gmgk \tag{1}$$

where n is the number of survival-correlated genes, β_g is the regression coefficient of the Cox proportional hazard model for gene g, and m_{gk} is the expression level of gene g in patient k. Patients were then divided into high- and low-risk groups based on the median PI. Survival differences between the high- and low-risk groups derived from calculating the PI by LASSO were tested by the Kaplan-Meier method and analyzed with the log-rank test with the functions *survfit* and *survdiff* in the *survival* package in R [26]. To test the ability of the signatures to predict CRC prognosis, we independently calculated angiogenic scores using angiogenic gene sets by ssGSEA and evaluated their ability to predict CRC prognosis by Cox proportional hazards regression. Cox univariate analysis was performed with the *coxph* function in the R package *survival*. A *P* value < 0.05 was considered significant.

2.8. Immunohistochemistry (IHC) staining

A tissue microarray of 130 formalin-fixed paraffin-embedded (FFPE) CRC samples was collected from the Second Affiliated Hospital, Zhejiang University School of Medicine. IHC staining was performed as described in our previous work [27], except that the antibody was replaced with an anti-EPHB2 (1:200 dilution, HUABIO, ER1908-25). The staining intensity and distribution of positive cells were evaluated by two independent pathologists. Staining intensity was graded as follows: 0, absent; 1, weak staining; 2, moderate staining; and 3, strong staining. The staining distribution was determined by the percentage of positive cells (0, <5 % positive cells; 1, 5–25 % positive cells; 2, 26–50 % positive cells; and 3, >50 % positive cells). The two scores were summed and divided by 2 to obtain the final score.

2.9. siRNA knockdown

The DLD1 CRC cell line was purchased from ATCC and cultured with RPMI 1640 medium (Gibco) containing 10 % fetal bovine serum (FBS, BI Industry). The cells were incubated at 37 °C with 5 % CO_2. After the cells grew to 50–60 % confluence in 6-well plate, Lipo 3000 (Invitrogen) with specific siRNA (tsingke, Hangzhou) was added to the cells according to the manufacturer's suggestions. The final concentration of siRNA was 100 nM. Cells were incubated with siRNA for 48h and then harvested for RNA extraction. The sequence of *EPHB2* siRNA was as follows: 5′-UCAUAGUCCAGGAUCACGC(dT)(dT)-3'.

2.10. RNA isolation and quantitative real-time reverse-transcriptase-PCR (qRT-PCR)

Total RNA was extracted from DLD1 cells using Trizol following a standard protocol. The Takara PrimeScript TM RT Master Mix Kit (Takara, RR036Q) was used for reverse transcription. The iTaq Universal SYBR Green Supermix (BioRad) and Applied Biosystems 7500
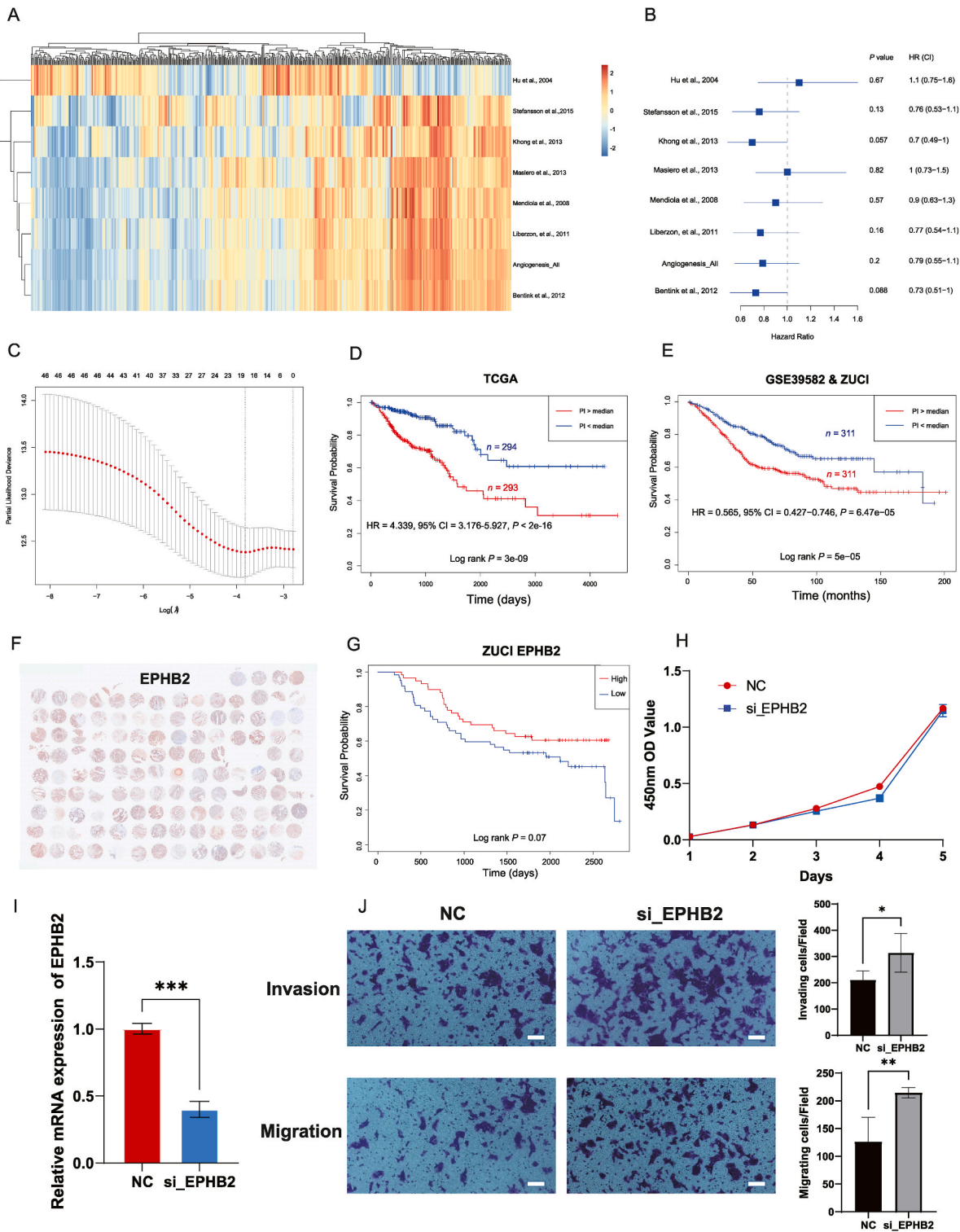
Fig. 1. Construction of an angiogenic gene expression signature that was significantly associated with CRC prognosis. A. Heatmap of angiogenic scores inferred by ssGSEA for each sample from TCGA using eight published angiogenic expression signatures. **B.** Cox regression analysis of the eight angiogenic signature gene sets revealed no significant correlation with prognosis in the TCGA CRC cohort. **C.** Feature selection integrated by LASSO. The left vertical dotted line shows where the CV-error curve hit its minimum (lambda.min) and the right vertical dotted line shows the most regularized model with a CV-error within 1 standard deviation of the minimum (lambda.1se). **D.** The high-risk group determined by our signature was associated with a significantly poorer prognosis than the low-risk group in the TCGA cohort. **E.** Compared with the low-risk group, the

high-risk group had a significantly poorer prognosis in the GSE39582 and ZUCI cohorts. **F.** IHC of EPHB2 in a tissue microarray containing 130 CRC samples from ZUCI. **G.** High expression of EPHB2 was associated with a better prognosis. **H.** Measurement of cell proliferation using cell counting kit (CCK-8) assays after transfection with EPHB2 interfering RNA. **I.** Knockdown of EPHB2 by custom siRNA was determined by qRT-PCR. **J.** Represented images and the number of migrated and invaded DLD1 cells were counted 24 h later. (mean $\pm$ SD, $n = 3$). HR: hazard ratio, CI: confidence interval, TCGA: The Cancer Genome Atlas, ZUCI: Zhejiang University Cancer Institute. Bar: 100 μm.

Fast Real-Time PCR System were applied for qRT-PCR. GAPDH was used as the loading control. Experiments were carried out in triplicate. The results were calculated as follows: $\Delta CT = CT_{Experimental/NC} - CT_{GAPDH}$, $\Delta\Delta CT = \Delta CT_{Experimental/NC} - \Delta CT_{NC,}$ fold change $= 2^{-\Delta\Delta CT}$. The *EPHB2* primers used for qRT-PCR were as follows: forward: 5′-ATGATCCGCAATCCCAACAGC-3′ and reverse: 5′-CAT-CATCTGAGACACGACGTC-3'.

2.11. Cell proliferation and transwell migration assays

Cell proliferation, migration and invasion assays were performed as described in our previous work [28].

2.12. Evaluating the predictive power of the custom angiogenic signature in CRC patients treated with bevacizumab

The tailored angiogenic signature was used to predict the efficacy of bevacizumab in three independent CRC cohorts (TCGA-19 samples, GSE72970-28 samples from Inserm France, and GSE60331-50 samples from the AXEBeam trial) with corresponding expression profiles. The efficacy of bevacizumab in treating CRC was retrieved from accompanying clinical information. The response status of the responders was defined as complete response (CR) and partial response (PR), and nonresponders were defined as stable disease (SD) or progressive disease (PD).

Table 1
Clinicopathological characteristics of the two subtypes based on angingenic signature.

	TCGA			GSE39582		
	High-risk	Low-risk		High-risk	Low-risk	
Variables	$n = 294$	$n = 293$	P-value	$n = 287$	$n = 286$	P-value
Age						
≤ 60	88	97	0.3455	91	69	0.05044
> 60	204	187		195	217	
Unknown	2	9		1		
Gender						
Female	135	129	0.9112	123	133	0.4273
Male	157	155		164	153	
Unknown	2	9				
Clinical stage			**2.39e-6**			**0.000137**
I/II	129	182		130	176	
III/IV	153	94		157	110	
Unknown	12	17				
Location			0.06226			0.1614
Left	175	147		182	164	
Right	110	129		105	122	
Unknown	9	17				
MSI status			0.4269			**2.94e-7**
MSI-H	29	24		17	56	
MSS	145	159		256	198	
Unkown	120	110		14	32	
CIMP						
Positive	27	27	0.9852	37	53	0.2529
Negative/Low	184	193		200	213	
Unknown	83	73		50	20	
BRAF						
WT	105	102	0.159	221	232	0.08434
MUT	24	13		17	32	
Unknown	165	178		49	22	
KRAS						
WT	125	114	0.05701	162	160	0.8875
MUT	56	79		105	108	
Unknown	113	100		20	18	

3. Results

3.1. Reconstruction of the CRC-oriented angiogenic signature

We first independently calculated the angiogenic score using the seven publicly available angiogenic expression gene sets and a gene panel including all genes mentioned in the seven studies referring to "Angiogenesis_All" by ssGSEA and evaluated their ability to predict the prognosis of patients in the TCGA CRC cohort ($n = 587$). The results showed that CRC samples can be divided mainly into high- or low-angiogenesis groups according to known angiogenic signatures, and the distribution trend of the signatures identified by Hu et al. [8] was relatively different from that of the other signatures (Fig. 1A). However, further Cox proportional hazards regression analysis showed that the angiogenic scores deduced from all eight signatures could not significantly predict the risk of CRC in the TCGA cohort (Fig. 1B). Therefore, we reintegrated the gene features of these studies, and 47 significant variables after univariate Cox regression were retained for LASSO regression (Table S2). GO enrichment analysis of the 47 significant genes revealed that blood vessel development (corrected $P = 2.571e-8$), vasculature development (corrected $P = 2.571e-8$), and blood vessel morphogenesis (corrected $P = 3.154e-8$) were the three most enriched terms. Notably, a new 19-gene angiogenic expression prognostic signature panel including *EPHB2* (coefficient: 3.3017e-8), *VEGFA* (5.89738e-5), *DHX15* (−0.00015), *SPP1* (4.288e-6), *PTP4A2* (−0.00016), *VAV2* (0.00015), *PFN2* (0.0001), *TNFAIP2* (1.43497e-5), *NDUFB6* (−2.64119e-5), *GCDH* (−4.07991e-5), *LAMA2* (0.00035), *ANGPTL4* (1.70309e-5), *TNNT1* (0.00091), *PRRX2* (0.00108), HIST1H2BJ (0.00253), NRXN1 (0.00058), *SEMA3E* (0.00123), *FIGF* (0.0087) and *PROK1* (0.00445) was constructed (Fig. 1C). *LAMA2*, *TNNT*1, HIST1H2BJ, and *VAV2* were among the top ten important variables in the random forest validation. GO enrichment analysis of these 19 genes also revealed that blood vessel morphogenesis (corrected $P = 1.949e-3$, represented by *FIGF, PRRX2, TNFAIP2, VEGFA*, and *VAV2*), blood vessel development (corrected $P = 2.218e-3$), vasculature development (corrected $P = 2.218e-3$), and angiogenesis (corrected $P = 2.722e-3$) were the top enriched terms.

Then, we classified the patients into high- and low-risk groups based on the median PI score (−0.073) of the newly constructed angiogenic signature gene set and tested the survival discrimination power in CRC (The pipeline is summarized in Fig. S1). Notably, the low-risk group had significantly better survival than the high-risk group in the TCGA cohort (HR = 0.23, 95 % CI = 0.169–0.315, $P <$
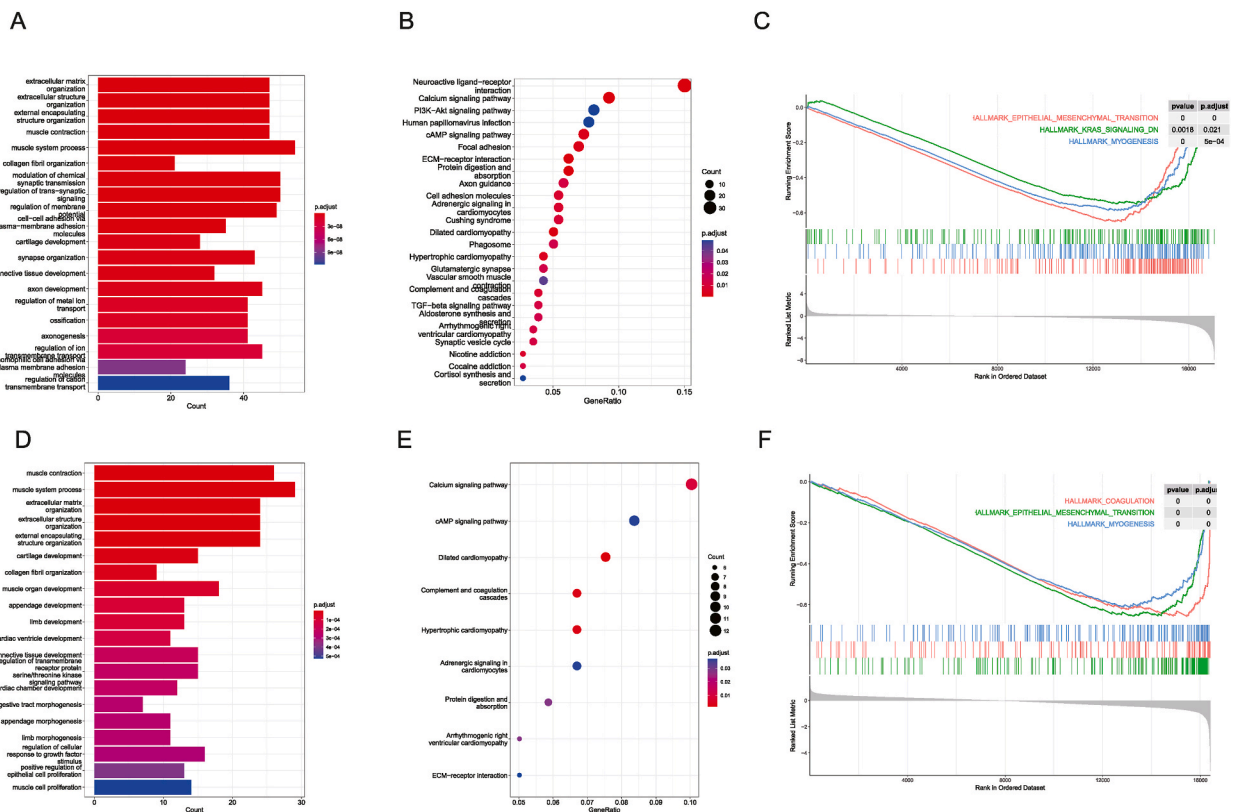

Fig. 2. Functional enrichment analysis of the high- and low-angiogenesis risk groups. A. Enriched GO terms in the TCGA high angiogenic risk group. The red bar indicates a more statistically significant difference. **B.** Enriched KEGG pathways in the TCGA high angiogenic risk group. Larger red solid circles correspond to a larger gene set size and greater statistical significance. **C.** Enriched hallmarks in the TCGA high angiogenic risk group determined by GSEA. **D.** Enriched GO terms in the GSE39582 high angiogenic risk group. **E.** Enriched KEGG pathways in the GSE39582 high angiogenic risk group. **F.** Enriched hallmarks in the GSE39582 high angiogenic risk group determined by GSEA. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
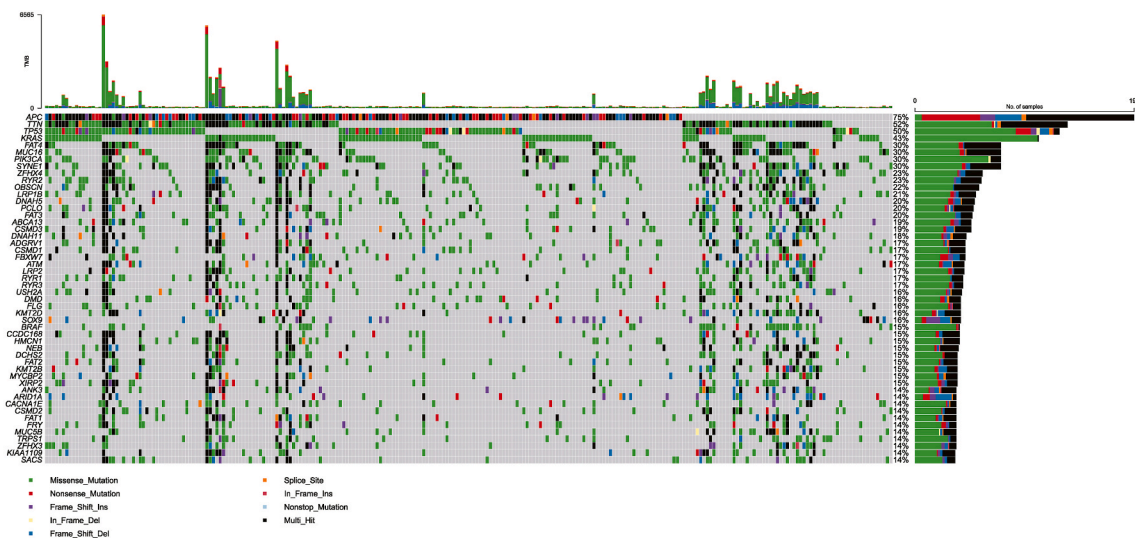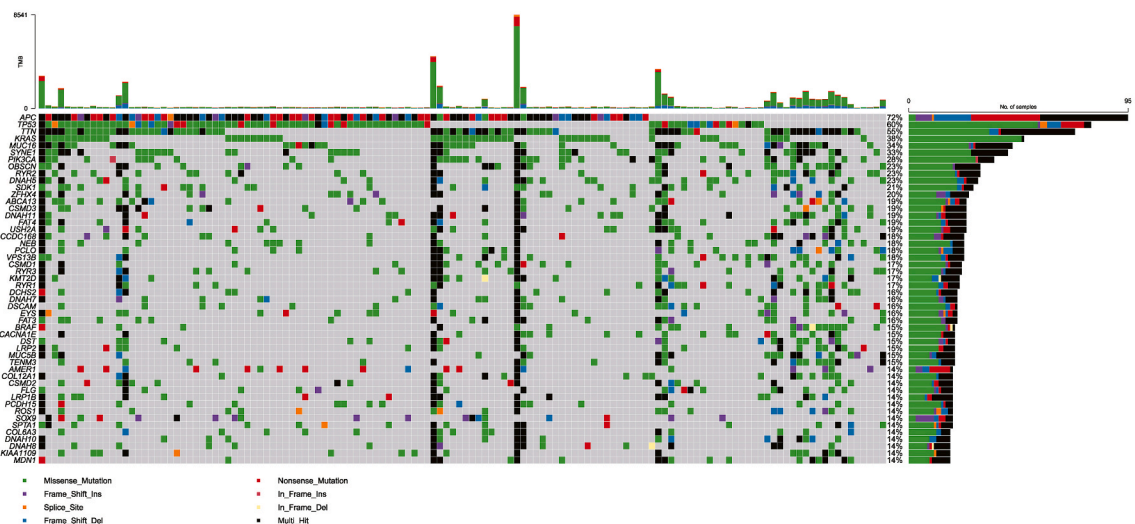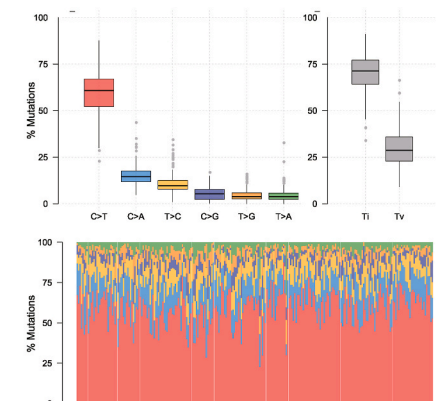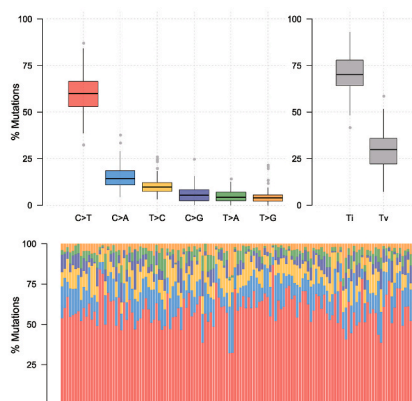
(caption on next page)

Fig. 3. Differences in mutation profiles between the high and low angiogenic risk groups. A. The top 50 mutated genes in the high-risk group were deduced from TCGA. Variants annotated as Multi_Hit are those genes that are mutated more than once in the same sample. **B.** The top 50 mutated genes in the low-risk group were deduced from TCGA. **C.** The proportions of T, C, G, and A mutation frequencies in the TCGA high angiogenic risk group. **D.** The proportions of T, C, G, and A mutation frequencies in the TCGA low angiogenic risk group. Ti: Transitions, Tv: Transversions.

2e-16; log rank P = 3e-9; Fig. 1D). The advantage of our model held by using tumor sample dichotomy according to the PI score when compared with other available signatures (Figs. S2A–C). The five-year survival rates for the high- and low-risk groups were 0.459 (95 % CI, 0.359 to 0.587) and 0.797 (95 % CI, 0.713 to 0.890), respectively. A significant survival advantage for the low-risk group was validated in another independent cohort GSE39582 plus ZUCI (HR = 0.565, 95 % CI = 0.427−0.746, P = 6.47e−05; Fig. 1E). IHC of a tissue microarray containing 130 CRC samples from ZUCI also validated the favorable prognostic role of EPHB2, which was implicated in our model (Fig. 1F and G). To further explore the functional role of EPHB2 in CRC, cell proliferation, cell migration, and invasion assays were performed following EPHB2 knockdown by custom siRNA. Results showed that DLD1 cell migration and invasion were significantly enhanced in EPHB2 knockdown cells, whereas the cell proliferation ability was unaffected after EPHB2 knockdown
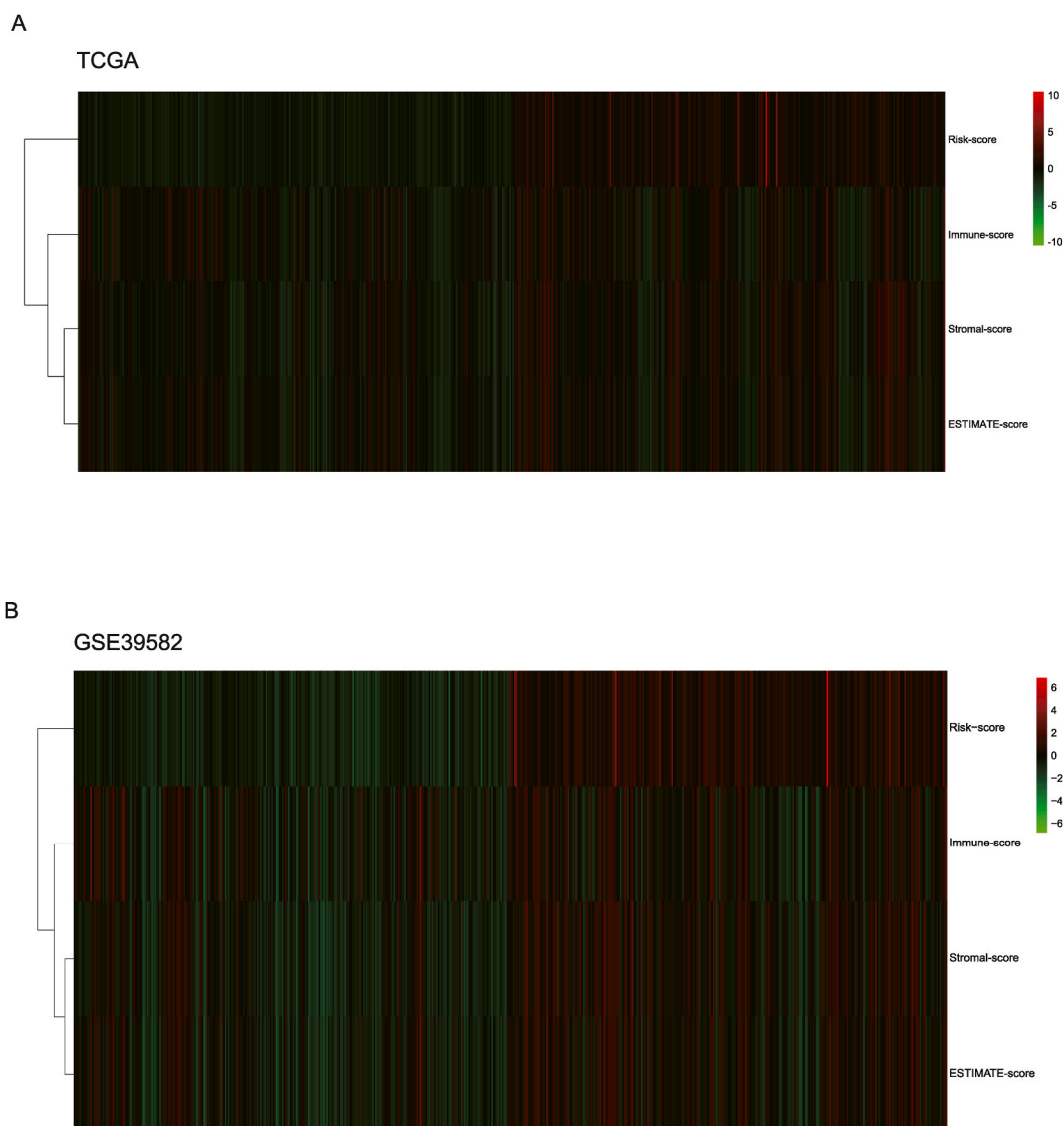

A
TCGA

B
GSE39582

Fig. 4. Differential infiltration of immune and stromal cells between the high- and low-angiogenesis risk groups. A. Heatmap of immune and stromal scores in the high- and low-angiogenic-risk groups in TCGA. Red indicates a high immune or stromal score, and green indicates a low score. **B.** Heatmap of immune and stromal scores in the high- and low-angiogenic-risk groups in GSE39582. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Fig. 1H–J).

Furthermore, analysis of the clinicopathological characteristics of the two subtypes from the TCGA and GSE39582 datasets based on the angiogenic signature revealed that patients with clinical stage III or IV (TNM stage) accounted for a greater proportion of patients in the high-risk group than in the low-risk group, while no other significant differences in clinical features, such as age, sex, left-right location, or *BRAF/KARS* mutation frequency, were detected between these two subtypes (Table 1). A similar trend was observed for the ZUCI, GSE72970, and GSE60331 subtypes, although the difference in stage distribution was not significant due to the relatively small sample size (Table S3).

3.2. Deciphering the molecular properties of the high- and low-risk angiogenic groups

To identify the molecular signatures associated with a particular angiogenic subtype, we searched for subtype-specific enriched genes. A total of 718 DEGs were identified between the high- and low-risk groups in TCGA, most of which were upregulated (97 %, 697/718) in the high-risk group. GO and KEGG enrichment analyses revealed both enhanced extracellular matrix (ECM) organization and ECM-receptor interaction in the high-risk group (Fig. 2A and B). KEGG enrichment analysis also revealed that the tumor-driven PI3K-Akt pathway, TGF-beta pathway, and vascular smooth muscle pathway were the most significantly enriched pathways (Fig. 2B). To further decipher the molecular signatures underlying the two angiogenic subtypes, we performed GSEA by using the well-annotated landscape of signature hallmarks collected in the MSigDB. The results showed that oncogenic features such as EMT (adjusted $P = 0$) and myogenesis (adjusted $P = 5e-4$) were the most significantly enriched hallmarks in the high-risk group compared with the low-risk group (Fig. 2C). These results were consistent with those in another dataset, GSE39582 (Fig. 2D–F).
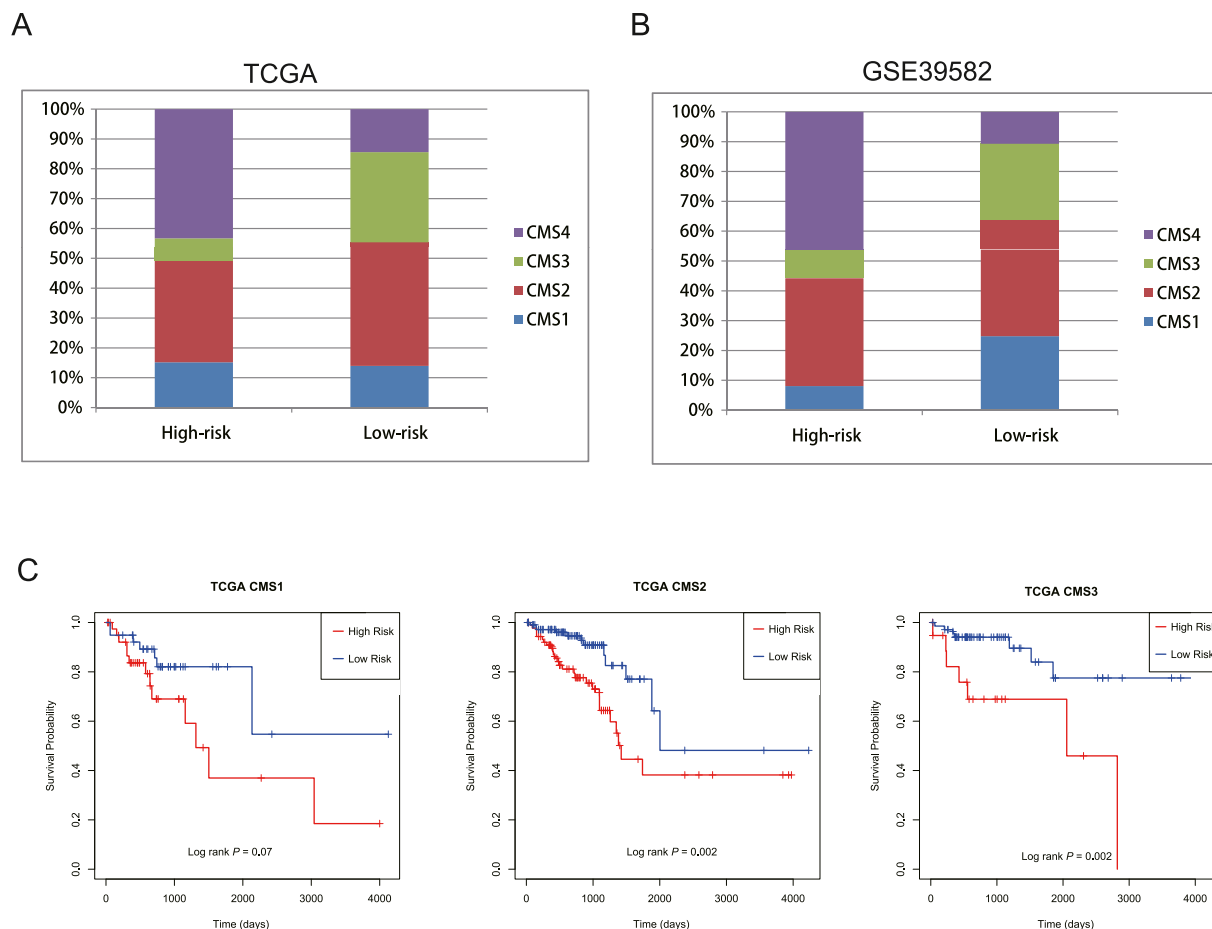

Fig. 5. **Associations between high and low angiogenic risk groups and CMS subtypes. A.** CMS subtype composition in the high- and low-risk groups in TCGA. **B.** CMS subtype composition in the high- and low-risk groups in GSE39582. C. CMS1-3 in the TCGA cohort can be further divided into high- and low-risk subgroups according to our signature, with a significantly poorer prognosis in the high-risk subgroup than the low-risk subgroup.

3.3. Genomic variance between the high- and low-risk angiogenic groups

Given the potential relationship between *KRAS* mutation and enhanced VEGF expression in tumors, we systematically compared the mutation profiles between the two subtypes in TCGA because complete mutation data were available for only the TCGA cohort. First, we summarized the overall somatic mutations in the two groups and found that *APC* (75 % vs. 72 %, high-risk:low-risk), *TP53* (50 % vs. 60 %), and *KRAS* (43 % vs. 38 %) were the most common driver mutation genes shared by both groups (Fig. 3A and B). Additionally, *BRAF* and *ATM* also had similar mutation frequencies in the two groups (~15 %), suggesting that the difference in survival between the high- and low-risk angiogenic groups was not driven by known driver mutations.

Next, we focused our attention on the transitions and transversions of SNPs. Although the proportions of total base transition and transversion were similar between the two subtypes and base substitution C > T was the most common, followed by C > A, T > C and C > G, interestingly, we found that the proportion of base mutation T > G was greater than that of T > A in the high-risk group, while the proportion in the low-risk group was the opposite (Fig. 3C and D).

3.4. Increased infiltration of immune and stromal cells in the high angiogenic risk group

In view of the close relationship between the TME and angiogenesis, we first calculated the stromal score, immune score, and ESTIMATE score (stromal score + immune score) of each sample from TCGA (Fig. 4A). Overall, much higher stromal, immune, and ESTIMATE scores were obtained in the high-risk group. This trend was validated in another dataset, GSE39582 (Fig. 4B). Second, we compared the proportions of the two groups using the well-known consensus molecular subtype (CMS) system [29]. There was no significant difference in the composition of the CMS1-microsatellite instability type or the CMS2-transclassic type, but the proportion of the CMS4-stromal angiogenesis type in the high-risk group was significantly higher (TCGA: 43 % vs. 14 %, GSE39582: 46 % vs. 11 %). In contrast, the proportion of the CMS3-metabolic type was significantly lower than that in the low-risk group (TCGA: 7 % vs. 30 %, GSE39582: 9 % vs. 26 %, Fig. 5A and B). In addition, our angiogenic signature can also classify CMS1-3 into high- and low-risk subgroups separately (Fig. 5C). In addition, our signature was found to be a significant independent prognostic factor in pairwise CMS comparisons, while CMS did not (Fig. S2D).

The great discrepancy in the proportion of CMS4 between the high- and low-risk groups suggested that the two groups differed significantly in terms of stromal infiltration within the TME. We thus further compared the proportions of 64 immune and stromal cells between the high- and low-risk groups in TCGA (Fig. S3). A greater proportion of vascular-related cells (endothelial cells, lymphatic endothelial cells, microvascular endothelial cells, and pericytes) were present in the high-risk group, and stromal cells (preadipocytes, adipocytes, astrocytes, fibroblasts, and mesenchymal stem cells) were also more abundant in the high-risk group. In addition, a strong positive correlation was observed between these vascular-related cells and stromal cells, but a negative correlation was detected between vascular-related cells and epithelial cells (Fig. 6A). This correlation held well in GSE39582 (Table S4). For immune cells, the high-risk subtype had more CD8 T cells, DC cells, basophils, eosinophils, monocytes, macrophages, mast cells, and neutrophils, while CD4 cells, Th1 cells, Th2 cells, natural killer cells, natural killer T cells, gamma delta T cells, and Treg cells were more enriched in the low-risk subtype. Moreover, the high-risk group retained more neurons, which can interact strongly with the immune system [30]. In summary, a high risk of angiogenesis was associated with greater stromal and immune infiltration in the TME.
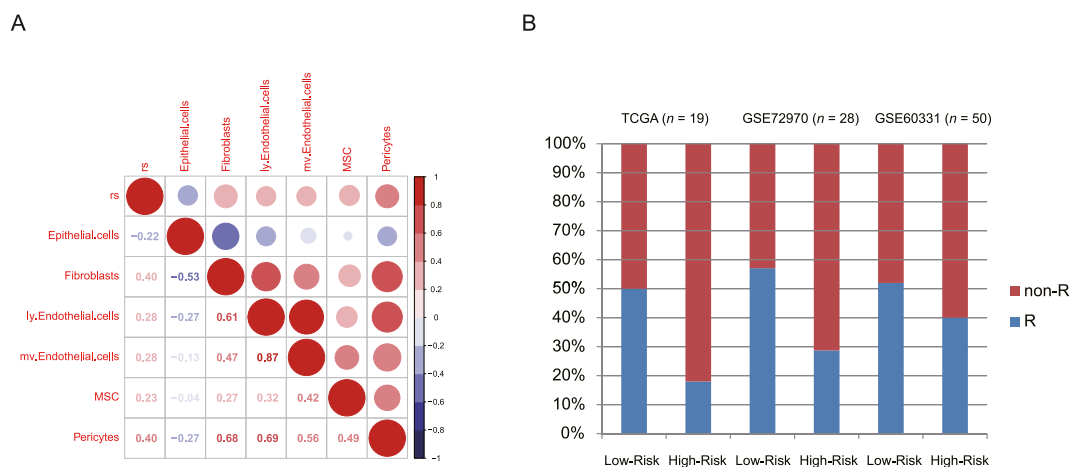

Fig. 6. Comparison of the response rates to bevacizumab between the high- and low-angiogenic-risk groups. A. Pairwise correlation between infiltrated stromal cells according to TCGA CRC samples. **B.** The low-risk group had a greater response rate to bevacizumab than did the high-risk group in the TCGA, GSE72970, and GSE60331 cohorts. R was defined as a complete response or a partial response, and non-R was defined as stable disease or progressive disease. rs: risk score, mv.Endothelial.cells: microvascular endothelial cells, ly.Endothelial.cells: lymphatic endothelial cells, MSC: mesenchymal stem cells.

3.5. The high angiogenic risk group showed bevacizumab resistance

Although bevacizumab has been widely used as an antiangiogenic drug since 2004 [31], early predictive biomarkers to evaluate its clinical efficacy are lacking [32]. Therefore, we further preliminarily explored whether the high- and low-risk groups defined by the tailored angiogenic signature could predict the efficacy of bevacizumab in three independent CRC cohorts. The results showed that the low-risk group had a higher response rate to bevacizumab than did the high-risk group in the TCGA, GSE72970, and GSE60331 datasets (Fig. 6B).

4. Discussion

Angiogenesis is indispensable for tumor progression and metastasis; however, no angiogenic signature can predict CRC prognosis has been established. We first used seven publicly available angiogenic signatures to evaluate their feasibility in CRC survival prediction. Nevertheless, all of these studies failed to distinguish CRC patients with either good or poor prognosis based on model scores. Thus, we constructed a tailored 19-gene angiogenic panel on the basis of seven combined angiogenic signatures identified by LASSO regression that can predict the survival status of patients with CRC. CRC patients can be divided into high- and low-risk groups according to the custom-built gene set, and the high-risk group was associated with a significantly poor outcome. Notably, most members of the 19-gene signature were found to be associated with cancer patient prognosis. High expression of *VAV2* [33], *PFN2* [34], *TNNT1* [35], and *PRRX2* [36] was correlated with poor outcomes in different carcinomas, which was consistent with the risk factors (coefficients >0) in our signature. Although EPHB2 is overexpressed in most tumors and predicts a worse outcome, it is downregulated in CRC and bladder cancer [37], which is in line with the protective factor (coefficient <0) in our signature. Furthermore, EPHB2 IHC in another independent cohort validated the favorable role of EPHB2 in CRC (Fig. 1F). This trend held true for DHX15 [38]. PROK1, a hazardous factor with the second highest weight in the tailored signature, has been demonstrated to be an independent prognostic factor and was significantly more abundant in CRCs with serosal invasion, lymphatic invasion, venous invasion, lymph node metastasis, liver metastasis, and hematogenous metastasis [39].

CRC is highly heterogeneous and is currently classified into four types according to the CMS system. CMS1 (14 %) is characterized by hypermutation, MSI, and strong immune system activation, *BRAF* mutation; CMS2 (37 %) is characterized by mainly epidermal WNT and Myc signaling activation, *APC* gene mutation; CMS3 (13 %) is characterized by mainly metabolic abnormalities, *KRAS* mutation; and CMS4 (23 %) is characterized by EMT and angiogenesis, with high somatic copy number variability. CMS4 has the worst overall survival compared to the other subtypes (HR 1.70–1.77) [29], which was consistent with our finding that the defined high-risk group had threefold more CMS4 samples (Fig. 5). Notably, the CMS system failed to differentiate between pairwise CMS1/2, CMS1/3, and CMS2/3 survival comparisons, but our signature can serve as a significant independent prognostic factor in both single CMS subtypes and any pair of CMS1-3 (Fig. S2D). In this regard, our model outperforms CMS subtyping. The enrichment of EMT and TGF-beta pathway activation may be one of the reasons for stronger metastatic potential and poor outcomes in the high-risk group.

Surprisingly, high angiogenic risk was related to enriched immune cell infiltration, which was previously recognized as a positive marker for good prognosis in tumors [40]. Further inspection revealed enhanced stromal infiltration in the high angiogenic risk group as well. Some stromal cells may aggravate malignant tumor phenotypes [41], particularly cancer-associated fibroblasts (CAFs) [42]. It is tempting to believe that the clinical benefit offered by a strong immune response may be neutralized by malignant stromal cells within the TME.

In mCRC patients with high baseline levels of VEGF-A [43,44], VEGF-D (also known as FIGF) [45], and ANGPTL4 [46], three high-risk factors (coefficient >0) included in our angiogenic prognostic signature, benefited less from bevacizumab therapy. This finding was in line with the lower response rate in the high-risk group of 97 CRC patients who received bevacizumab (Fig. 6B). A plausible explanation for the overexpression of high-risk factors such as VEGF-D could lead to resistance to bevacizumab because VEGF-D has been shown to bind to VEGFR-2, triggering lymphangiogenesis [45]. Interestingly, it was reported that high baseline levels of serum-based collagens type III and VI derived from CAFs significantly predict poor prognosis in mCRC patients treated with bevacizumab and chemotherapy [47], which partially explains why the high-risk patients that identified by our prognostic model are relatively resistant to bevacizumab because of increased infiltration of CAFs. Additionally, high serum levels of cyclophilin A [48], IL6 [49], IL8 [50,51], and proteinase-3 [52] predict poor survival and lower efficacy of bevacizumab in patients with mCRC. To date, no predictive angiogenic transcriptome signatures have been identified in CRC, and the angiogenic signature constructed in this study could significantly predict patient prognosis and the response to bevacizumab. Nevertheless, this promising signature needs further validation in larger scale prospective clinical trials.

5. Conclusions

We summarized and reconstructed a novel angiogenic signature gene set that could predict the prognosis of CRC patients. This detailed inspection will help us better understand the relationships among angiogenesis, the altered TME and the response rate to bevacizumab in CRC patients.

Consent for publication

All authors read and are consent for the publication of this manuscript.

Data availability statement

The datasets used during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Aiqin Chen: Investigation, Data curation. **Kailai Wang:** Visualization, Resources, Investigation. **Lina Qi:** Resources, Methodology. **Wangxiong Hu:** Writing – review & editing, Writing – original draft, Formal analysis. **Biting Zhou:** Writing – review & editing, Project administration, 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33662.

References

[1] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021) 209–249.
[2] L. Carlsen, K.E. Huntington, W.S. El-Deiry, Immunotherapy for colorectal cancer: mechanisms and predictive biomarkers, Cancers 14 (4) (2022).
[3] J. Folkman, Tumor angiogenesis: therapeutic implications 285 (21) (1971) 1182–1186.
[4] R. Lugano, M. Ramachandran, A. Dimberg, Tumor angiogenesis: causes, consequences, challenges and opportunities, Cell. Mol. Life Sci. 77 (9) (2020) 1745–1770.
[5] S. Li, et al., Angiogenesis in pancreatic cancer: current research status and clinical implications, Angiogenesis 22 (1) (2019) 15–36.
[6] H. Hurwitz, et al., Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer, N. Engl. J. Med. 350 (23) (2004) 2335–2342.
[7] P. Carmeliet, R.K. Jain, Molecular mechanisms and clinical applications of angiogenesis, Nature 473 (7347) (2011) 298–307.
[8] J. Hu, et al., Gene expression signature for angiogenic and nonangiogenic non-small-cell lung cancer, Oncogene 24 (7) (2005) 1212–1219.
[9] I.M. Stefansson, et al., Increased angiogenesis is associated with a 32-gene expression signature and 6p21 amplification in aggressive endometrial cancer, Oncotarget 6 (12) (2015) 10634–10645.
[10] T.L. Khong, et al., Identification of the angiogenic gene signature induced by EGF and hypoxia in colorectal cancer, BMC Cancer 13 (2013) 518.
[11] M. Masiero, et al., A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis, Cancer Cell 24 (2) (2013) 229–241.
[12] M. Mendiola, et al., Angiogenesis-related gene expression profile with independent prognostic value in advanced ovarian carcinoma, PLoS One 3 (12) (2008) e4051.
[13] A. Liberzon, et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell Syst 1 (6) (2015) 417–425.
[14] S. Bentink, et al., Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer, PLoS One 7 (2) (2012) e30269.
[15] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (11) (2018) 1747–1756.
[16] S. Hanzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[17] W. Hu, et al., Subtyping of microsatellite instability-high colorectal cancer, Cell Commun. Signal. 17 (1) (2019) 79.
[18] L. Wang, et al., DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (1) (2010) 136–138.
[19] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[20] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc Natl Acad Sci U S A 102 (43) (2005) 15545–15550.
[21] V.K. Mootha, et al., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes, Nat. Genet. 34 (3) (2003) 267–273.
[22] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[23] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, Genome Biol. 18 (1) (2017) 220.
[24] W. Hu, et al., Deciphering molecular properties of hypermutated gastrointestinal cancer, J. Cell Mol. Med. 23 (1) (2019) 370–379.
[25] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J Stat Softw 33 (1) (2010) 1–22.
[26] T.M. Therneau, P.M. Grambsch, Modeling Survival Data: Extending the Cox Model, Springer, New York, 2000.
[27] X. Li, et al., CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway, Cell Commun. Signal. 15 (1) (2017) 38.
[28] L. Qi, et al., HomeoboxC6 promotes metastasis by orchestrating the DKK1/Wnt/β-catenin axis in right-sided colon cancer, Cell Death Dis. 12 (4) (2021).
[29] J. Guinney, et al., The consensus molecular subtypes of colorectal cancer, Nat. Med. 21 (11) (2015) 1350–1356.
[30] J. Sbierski-Kind, N. Mroz, A.B. Molofsky, Perivascular stromal cells: directors of tissue immune niches, Immunol. Rev. 302 (1) (2021) 10–31.
[31] N. Ferrara, et al., Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer, Nat. Rev. Drug Discov. 3 (5) (2004) 391–400.
[32] J. Garcia, et al., Bevacizumab (Avastin®) in cancer treatment: a review of 15 years of clinical experience and future outlook, Cancer Treat Rev. 86 (2020) 102017.

[33] L.F. Lorenzo-Martín, et al., VAV2 signaling promotes regenerative proliferation in both cutaneous and head and neck squamous cell carcinoma, Nat. Commun. 11 (1) (2020) 4788.

[34] X.B. Cui, et al., PFN2, a novel marker of unfavorable prognosis, is a potential therapeutic target involved in esophageal squamous cell carcinoma, J. Transl. Med. 14 (1) (2016) 137.

[35] Y.H. Hao, et al., TNNT1, a prognostic indicator in colon adenocarcinoma, regulates cell behaviors and mediates EMT process, Biosci. Biotechnol. Biochem. 84 (1) (2020) 111–117.

[36] Y.L. Juang, et al., PRRX2 as a novel TGF-β-induced factor enhances invasion and migration in mammary epithelial cell and correlates with poor prognosis in breast cancer, Mol. Carcinog. 55 (12) (2016) 2247–2259.

[37] W. Liu, et al., The roles of EphB2 in cancer, Front. Cell Dev. Biol. 10 (2022) 788587.

[38] S. Ito, et al., RNA helicase DHX15 acts as a tumour suppressor in glioma, Br. J. Cancer 117 (9) (2017) 1349–1359.

[39] T. Nakazawa, et al., Prokineticin 1 protein expression is a useful new prognostic factor for human sporadic colorectal cancer, Ann. Surg Oncol. 22 (5) (2015) 1496–1503.

[40] D. Bruni, H.K. Angell, J. Galon, The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy, Nat. Rev. Cancer 20 (11) (2020) 662–680.

[41] P.G. Loftus, et al., Targeting stromal cell Syndecan-2 reduces breast tumour growth, metastasis and limits immune evasion, Int. J. Cancer 148 (5) (2021) 1245–1259.

[42] T.J. Underwood, et al., Cancer-associated fibroblasts predict poor outcome and promote periostin-dependent invasion in oesophageal adenocarcinoma, J. Pathol. 235 (3) (2015) 466–477.

[43] L. Zhao, et al., High VEGF-A level at baseline predicts poor treatment effect of bevacizumab-based chemotherapy in metastatic colorectal cancer: a meta-analysis, Panminerva Med. 58 (1) (2016) 48–58.

[44] M.A. Bruhn, et al., Proangiogenic tumor proteins as potential predictive or prognostic biomarkers for bevacizumab therapy in metastatic colorectal cancer, Int. J. Cancer 135 (3) (2014) 731–741.

[45] A.J. Weickhardt, et al., Vascular endothelial growth factor D expression is a potential biomarker of bevacizumab benefit in colorectal cancer, Br. J. Cancer 113 (1) (2015) 37–45.

[46] L. Bai, et al., A plasma cytokine and angiogenic factor (CAF) analysis for selection of bevacizumab therapy in patients with metastatic colorectal cancer, Sci. Rep. 5 (2015) 17717.

[47] N.I. Nissen, et al., Prognostic value of blood-based fibrosis biomarkers in patients with metastatic colorectal cancer receiving chemotherapy and bevacizumab, Sci. Rep. 11 (1) (2021) 865.

[48] D.C. Moisuc, , et al.A. Cyclophilin, An independent prognostic factor for survival in patients with metastatic colorectal cancer treated with bevacizumab and chemotherapy, Cancers 16 (2) (2024) 385.

[49] M. Hara, et al., High serum levels of interleukin-6 in patients with advanced or metastatic colorectal cancer: the effect on the outcome and the response to chemotherapy plus bevacizumab, Surg. Today 47 (4) (2016) 483–489.

[50] G. Marisi, et al., IL-8 and thrombospondin-1 as prognostic markers in patients with metastatic colorectal cancer receiving bevacizumab, Cancer Manag. Res. 10 (2018) 5659–5666.

[51] M. Suenaga, et al., Serum IL-8 level as a candidate prognostic marker of response to anti-angiogenic therapy for metastatic colorectal cancer, Int. J. Colorectal Dis. 36 (1) (2020) 131–139.

[52] K. Furuya, et al., High serum proteinase-3 levels predict poor progression-free survival and lower efficacy of bevacizumab in metastatic colorectal cancer, BMC Cancer 24 (1) (2024).