

RESEARCH

Open Access



Identification of a novel lymphangiogenesis signature associated with immune cell infiltration in colorectal cancer based on bioinformatics analysis

Hong Liu^{1†}, Huiwen Shi^{2†} and Yinggang Sun^{3*}

Abstract

Background Lymphangiogenesis plays an important role in tumor progression and is significantly associated with tumor immune infiltration. However, the role and mechanisms of lymphangiogenesis in colorectal cancer (CRC) are still unknown. Thus, the objective is to identify the lymphangiogenesis-related genes associated with immune infiltration and investigation of their prognosis value.

Methods mRNA expression profiles and corresponding clinical information of CRC samples were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. The lymphangiogenesis-related genes (LymRGs) were collected from the Molecular Signatures database (MSigDB). Lymphangiogenesis score (LymScore) and immune cell infiltrating levels were quantified using ssGSEA. LymScore and immune cell infiltrating levels-related hub genes were identified using weighted gene co-expression network analysis (WGCNA). Univariate Cox and LASSO regression analyses were performed to identify the prognostic gene signature and construct a risk model. Furthermore, a predictive nomogram was constructed based on the independent risk factor generated from a multivariate Cox model.

Results A total of 1076 LymScore and immune cell infiltrating levels-related hub genes from three key modules were identified by WGCNA. LymScore is positively associated with natural killer cells as well as regulator T cells infiltrating. These modular genes were enriched in extracellular matrix and structure, collagen fibril organization, cell-substrate adhesion, etc. NUMBL, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB were eventually identified as the prognostic gene signature in CRC. And patients were divided into high-risk and low-risk groups based on the median risk score, the patients in the high-risk group indicated poor survival and were predisposed to metastasis and advanced stages. NUMBL and PHF21A were upregulated but PDGFRA was downregulated in tumor samples compared with normal samples in the Human Protein Atlas (HPA) database.

[†]Hong Liu and Huiwen Shi have contributed equally to this study.

*Correspondence:
Yinggang Sun
sunny4367@sohu.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusion Our finding highlights the critical role of lymphangiogenesis in CRC progression and metastasis and provides a novel gene signature for CRC and novel therapeutic strategies for anti-lymphangiogenic therapies in CRC.

Keywords Colorectal cancer, Lymphangiogenesis, Prognosis, Immune infiltration, WGCNA

Introduction

Colorectal cancer (CRC), encompassing both rectal adenocarcinoma (READ) and colon adenocarcinoma (COAD), stands as the most prevalent cancer and the third leading cause of cancer-related deaths, accounting for over 1.9 million new cases and 935,000 deaths in 2020 [1]. It has been estimated that CRC will escalate to 3.2 million new cases and 1.6 billion deaths by 2040 [2]. Notably, CRC often serves as a marker of socioeconomic development, exhibiting increased incidence rates during periods of significant national transition [3, 4]. Several risk factors contribute to the prevalence of CRC, including reduced physical activity, excessive body weight, alcohol consumption, smoking, and the intake of red or processed meat [5, 6]. Over the past few decades, CRC has also emerged as an early-onset cancer, affecting individuals under the age of 50 [7]. Despite advancements in early screening methods and effective treatment strategies that have improved the prognosis for CRC patients, it continues to pose a serious threat to public health, constituting a substantial burden [8–10].

Previous studies have established that lymphangiogenesis is linked to lymphatic metastasis, distant metastasis, and adverse clinical outcomes in various types of tumors [11–13]. Moreover, the presence of functional lymphatic vessels plays a pivotal role in regulating the formation of inflammatory and immune microenvironments within tumors [14, 15]. The lymphatic system is essential for the collection and cycling of tissue-extravasated fluids, macromolecules, and immune cells into the bloodstream, especially mediates tumor metastasis and provides more chances for immunocytes to migrate through lymphatic vessels [16–18]. Lymphangiogenesis, extracellular matrix remodeling, and immunosuppressive cell enlisting in lymph nodes are indispensable for pre-metastatic niche formation [19]. Lymphangiogenesis occurs during the initial stage of metastasis in various types of malignant tumors, and recent evidence suggests that lymphatic vessels are involved in shaping antitumor immunity [20, 21]. Antigens, cytokines, and danger signals are drained from the tumor to sentinel lymph nodes to promote T-cell infiltration [22]. Various tumor-associated immune cells play an important role in lymphangiogenesis, for example, tumor-associated macrophages (TAMs) and tumor-associated neutrophils (TANs) act as the important drivers of lymphangiogenesis [23–25]. Lymphangiogenesis is an intricate process governed by a multitude of factors and activated by numerous genes. Lymphangiogenesis-related genes (LRGs) have the potential to serve

as prognostic indicators in human tumors. Lymph Node Metastasis Associated Transcript 1 (LNMAT1) promotes lymphatic metastasis and acts as a potential therapeutic target for LN-metastatic bladder cancer therapy [26]. Vascular endothelial growth factor C (VEGF-C), interleukin-4 (IL-4), colony-stimulating factor 2 (CSF2), prospero homeobox 1 (PROX1), and TEK receptor Tyrosine Kinase (TEK) is significantly associated with lymph node metastasis in renal cell carcinoma (RCC) [27, 28]. Collagen and calcium-binding EGF domain-1 (CCBE1) and neuropilin-2 (NRP2) promote lymphangiogenesis and lymphatic metastasis in CRC and can be used as therapeutic targets in CRC [29–31]. The above studies have confirmed that lymphangiogenesis is involved in tumor metastasis and poor clinical outcomes, targeting lymphangiogenesis emerges as a potential and effective therapeutic strategy. However, there are still numerous LRGs associated with the prognosis of CRC that remain unknown.

In the present study, a weight gene co-expression network analysis (WGCNA) was performed to identify the lymphangiogenesis and immune-related modules and screened the hub genes based on The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Then, the prognostic values of those hub genes were investigated using univariate Cox and Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses. NUMBL, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB were selected as the lymphangiogenesis and immune-related signature that could be used for prognosis and prediction of therapeutic responses in CRC.

Methods

Data collection and processing

The gene expression data (fragments per kilobase per million mapped reads (FPKM) standardized data) and corresponding clinical information of 583 CRC patients from TCGA-COAD and TCGA-READ were downloaded from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) database. Besides, the gene expression microarray data and corresponding clinical data of 232 CRC patients from the GSE17538 dataset were downloaded from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database. The GSE17538 dataset was generated using the GPL570 platforms. Additionally, the 17 lymphangiogenesis-related genes (LymRGs) were collected from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/>).

[org/gsea/msigdb/](https://www.gsea-msigdb.org/gsea/msigdb/)) and used for subsequent analyses (Table S1-2).

Lymphangiogenesis score (LymScore) calculating and immune cell infiltration analysis

Single sample gene set enrichment analysis (ssGSEA) was performed to quantitatively calculate the LymScore of 17 LymRGs in each CRC sample using the GSVA package in R. Besides, the immune cell infiltration level in each CRC sample also was quantified using the GSVA R package based on the 28 immune genes sets (Table S3).

Weighted gene co-expression network analysis (WGCNA) and identification of hub models and hub genes

WGCNA is a biology method used to illustrate the correlation patterns between gene modules and clinical traits [32]. Here, the WGCNA R package was utilized to construct a co-expression network and identify the hub genes that relate to LymScore and immune cell infiltration. The input genes for network construction were selected based on the median absolute deviation (MAD), and then the soft threshold power (β) was estimated using a nearly scale-free topology to construct a scale-free network. The topological overlap matrix (TOM) similarity was used to determine the distance between each gene pair. Hierarchical clustering analysis with the average method was used to establish the cluster tree, and then the variant set of genes was stratified into different modules with a minimum of 30 genes in each module. The first principal component of each module's expression was summarized as a module eigengene (ME), and Pearson's correlations between MEs and LymScore as well as immune cell infiltration were estimated to identify the significant modules with the greatest absolute module significance (MS) for further analyses. For each module, module membership (MM) was characterized with the correlation coefficient between ME and gene expression, and gene significance (GS) value was used to quantify the correlation between individual genes and clinical traits. Genes with $MM > 0.6$ and $|GS| > 0.4$ were identified as hub genes in the module.

Gene ontology (GO) annotation and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis

The clusterProfiler R package was utilized to explore the biological functions of hub genes [33], which included Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). GO consists of three terms, including biological process (BP), molecular function (MF), and cellular component (CC). P -value < 0.05 was considered as the statistical significance.

Construction and validation of a lym-risk score model

To assess the correlation between the hub genes and survival status of patients, 583 CRC patients from TCGA were divided into a training set and a test set on a 7:3 ratio, then, a univariate Cox analysis was performed using survminer R package to identify the survival-related genes. Genes with P -value < 0.05 were incorporated into a LASSO regression model using the glmnet R package to shrink the number of genes. Then, the Lym-risk score was calculated as follows, Lym-risk score = $(-0.41902565) * TSPAN11$ expression + $(-0.27690515) * PDGFRA$ expression + $0.03754875 * ZNF386A$ expression + $0.3706381 * NUMBL$ expression + $0.4806545 * PHF21A$ expression + $0.48551082 * RIMKLB$ expression. Then, the patients from the training set, test set, and external validation set were classified into high- and low-Lym-risk groups based on the median Lym-risk score. The overall survival (OS) of each patient between high- and low-Lym-risk groups was determined using Kaplan-Meier curves and the log-rank test.

Construction of a predictive nomogram

The clinical characteristics (age, gender, T/N/M stages, grade) and risk score were incorporated into a multivariate Cox model to identify the independent factors in CRC. Then, the rms R package was performed to construct a predictive nomogram. The accuracy of the nomogram was detected using the calibration curve.

Validation of Lym-related signature via human protein atlas (HPA) databases

The protein levels of the Lym-related signature were investigated based on the immunohistochemical (IHC) staining images in CRC tissues that were downloaded from the HPA online database (<https://www.proteinatlas.org/>).

Statistical analysis

Statistical analyses were performed by using R (version 4.0.2) software packages, and $P < 0.05$ was considered statistically significant.

Results

Construction of a co-expression network of LymScore and immune cell infiltration in CRC

The LymScore and the immune cell infiltration level in each CRC sample were quantified (Table S2-3). Then, we constructed a co-expression network to identify the significant modules that were associated with the LymScore and immune cell infiltration in CRC. Firstly, we evaluate the global gene expression patterns using hierarchical clustering to eliminate the outlier samples that would impact the subsequent analyses. Here, no outlier samples were removed according to the sample clustering results

(Figure S1A). Then, a trait heatmap was constructed that showed the distribution of samples according to the corresponding clinical features (Figure S1B). Afterward, seventeen modules were generated from the TCGA cohort based on the differential expression of genes (Figure S2A). The optimal $\beta=10$ was considered the soft threshold ensuring that the network was scale-free (scale-free $R^2=0.85$, Figure S2B). The adjacency matrix transformed to the topological overlap matrix (TOM) and 17 modules were ensured using a cutoff of 0.25 and a minimum module size of 100 (Figure S2C-1D).

Identification of LymScore and immune cell infiltration-related hub genes in CRC

Based on the co-expression network of LymScore and immune cell infiltration, a heatmap was used to describe the relationship between modules and immune cell infiltration (Fig. 1A). As a result, brown (Nature killer cells, $r=0.81, p=2e-159$; LymScore, $r=0.78, p=2e-137$), purple (Nature killer cells, $r=0.75, p=5e-123$; LymScore, $r=0.87, p=1e-205$), and green-yellow (Regulator T cells, $r=0.94, p=1e-302$; LymScore, $r=0.60, p=2e-66$) modules were significantly associated with LymScore and immune cell infiltration in CRC with the

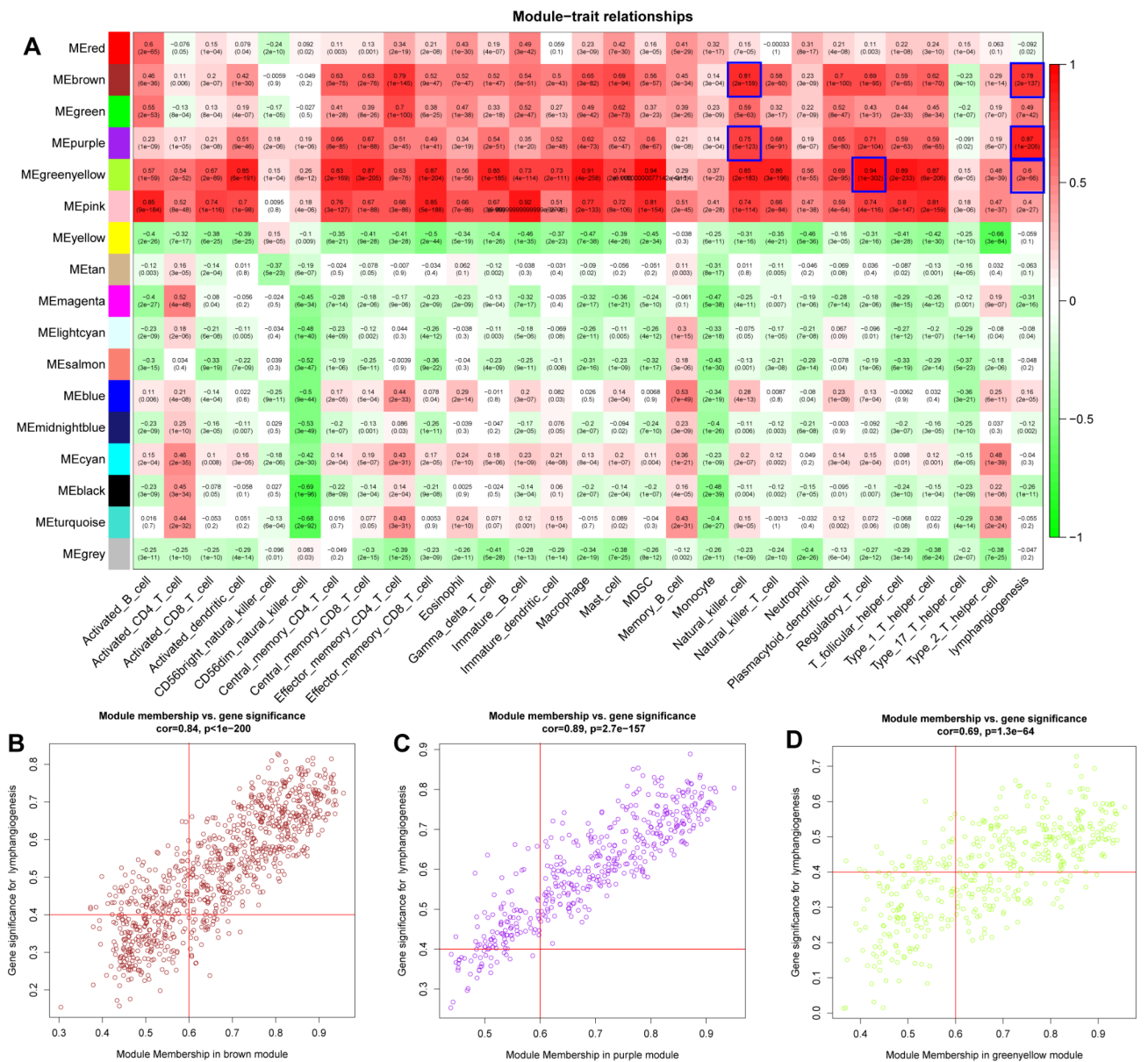


Fig. 1 Identification of LymScore and immune cell infiltration-related hub genes in CRC based on WGCNA. **A**. A heatmap indicating the correlation between modules and LymScore and immune cell infiltration levels. **B-D**. Scatter plot indicating the memberships and the gene significance for lymphangiogenesis in brown, purple, and green-yellow modules

p -value < 0.05 and $|r| > 0.5$ (Table S4). A total of 1076 hub genes with $MM > 0.6$ and $|GS| > 0.4$ were identified as hub genes from three modules (Fig. 1B-D, Table S4).

Functional analysis of hub genes

We further conducted the GO and KEGG enrichment analyses to investigate the potential mechanisms of the above modules. A total of 1,217 GO terms and 43 KEGG pathways were enriched (Table S5). As shown in Fig. 2A, the top 10 BP terms included extracellular matrix and structure, external encapsulating structure, collagen fibril organization, cell-substrate adhesion, skeletal system development, ossification, regulation of angiogenesis and vasculature development, and positive regulation of cell adhesion. For CC analysis, such as the collagen-containing extracellular matrix, collagen trimer, endoplasmic reticulum lumen cell-substrate junction, and focal adhesion were enriched. For MF analysis, several MF were identified, such as extracellular matrix structural constituent, collagen binding, integrin binding, and extracellular matrix structural constituent conferring tensile strength. KEGG pathway enrichment analysis revealed that several pathways were enriched in CRC, including focal adhesion, ECM-receptor interaction, phagosome, complement, and coagulation cascades, protein digestion and absorption, and cell adhesion molecules (Fig. 2B).

Construction of the LymScore and immune cell infiltration-related prognostic signature

We also investigated the prognostic values of above 1076 hub genes by conducting a univariate Cox model. The 583 CRC patients from TCGA were distributed into a training set ($n=409$) and test set ($n=174$) with 7:3.

And the univariate Cox results indicated that NUMBL, PCDH18, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB were closely related to prognosis (Fig. 3A, Table S6). Those genes were incorporated into a LASSO regression model to select the prognostic signature for CRC, resulting in six genes (NUMBL, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB) were finally identified for the LymScore and immune cell infiltration-related risk model construction (Fig. 3B-C). CRC patients in each cohort were distributed into high- and low-risk groups with the median value of risk score (Fig. 3D-F), Kaplan-Meier curves indicated that CRC patients in the high-risk score group showed poor survival compared to those in the low-risk score group (Fig. 3G-I). The AUC values of ROC curves for 1-, 3-, and 5-year OS prediction were more than 0.6 in both three cohorts (Fig. 3J-L). Correlation analysis between the risk score and clinical characteristics (age, gender, T/N/M stages) revealed that the high-risk score is closely related to tumor metastasis and advanced stage (Fig. 3M-O, Table S7).

Validation of the hub gene expression based on the HPA database

To validate the protein expression levels of prognostic hub genes, the IHC images of the tumor and normal tissues, indicated that NUMBL and PHF21A were strongly expressed in tumor tissues compared with normal tissues (Fig. 4A and C), but PDGFRA was deeply stained in normal tissues compared with tumor tissues (Fig. 4B). The results demonstrated that NUMBL and PHF21A might act as risk factors, while PDGFRA might function as the protective factor.

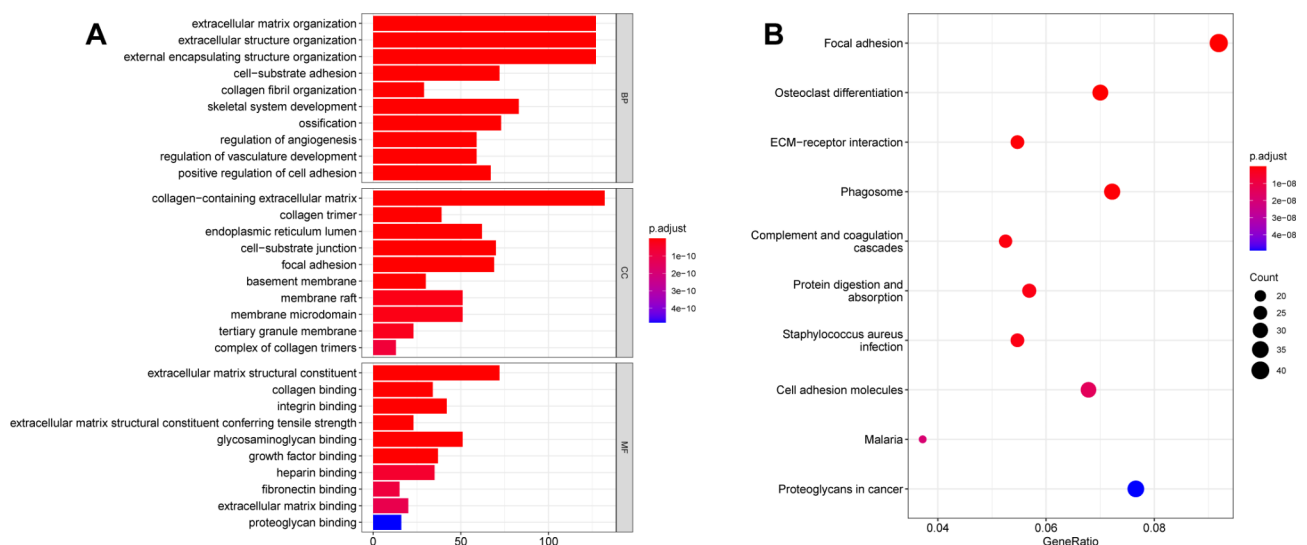


Fig. 2 Functional analysis of hub genes. **A.** GO annotation analysis of hub genes from brown, purple, and green-yellow modules. GO annotation comprises biological process (BP), cellular component (CC), and molecular function (MF) terms. **B.** KEGG enrichment analysis of hub genes from brown, purple, and green-yellow modules

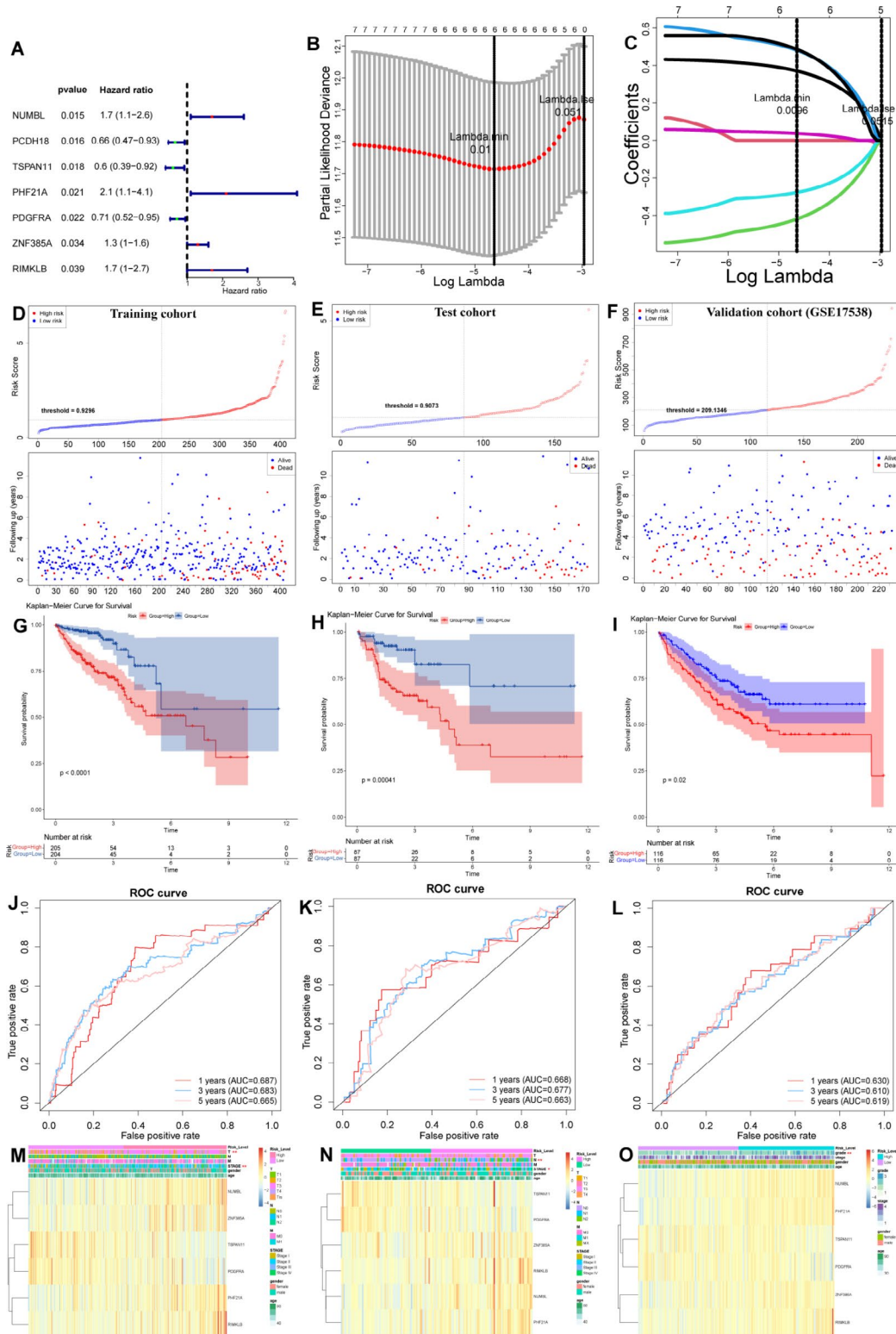


Fig. 3 Construction of the LymScore and immune cell infiltration-related prognostic signature. **A**. Forest plot showing the univariate Cox analysis for identifying survival-associated genes. **B**. LASSO coefficient profiles of the seven survival-associated genes in the TCGA cohort. **C**. Selection of the optimal parameter (lambda) in the LASSO model. **D-F**. CRC patients of training, test, and external validation cohorts were distributed into high-risk and low-risk groups based on risk score, and scatter plots showing the survival status. **G-I**. Kaplan-Meier analysis for OS of high-risk and low-risk groups both in three cohorts. **M-O**. Heatmap showing the correlation between clinical characteristics (age, gender, T/N/M stages) and risk score both in three cohorts

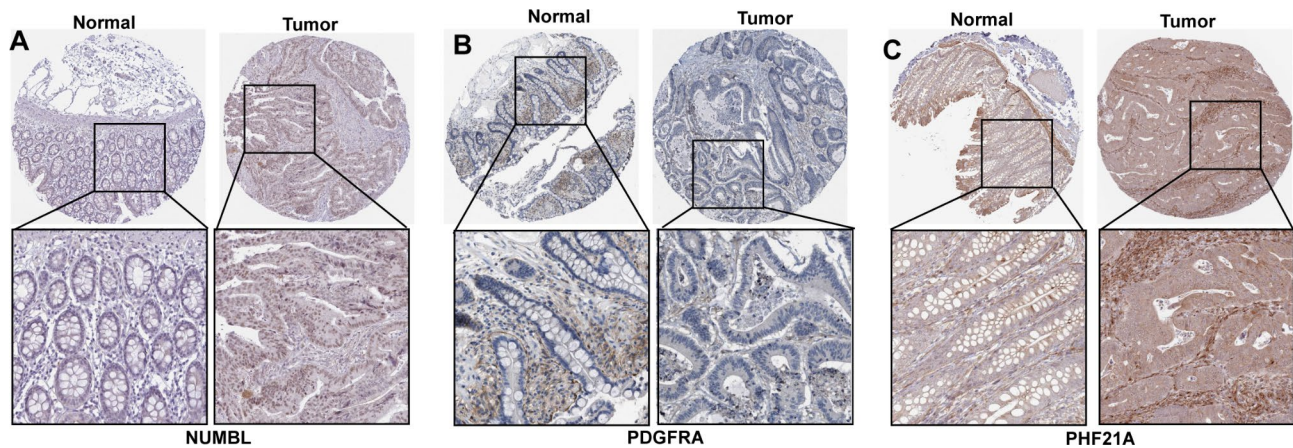


Fig. 4 Validation of the hub gene expression based on the HPA database. **A-C.** protein levels of NUMBL, PDGFRA, and PHF21A in CRC tumor specimens from the Human Protein Atlas database

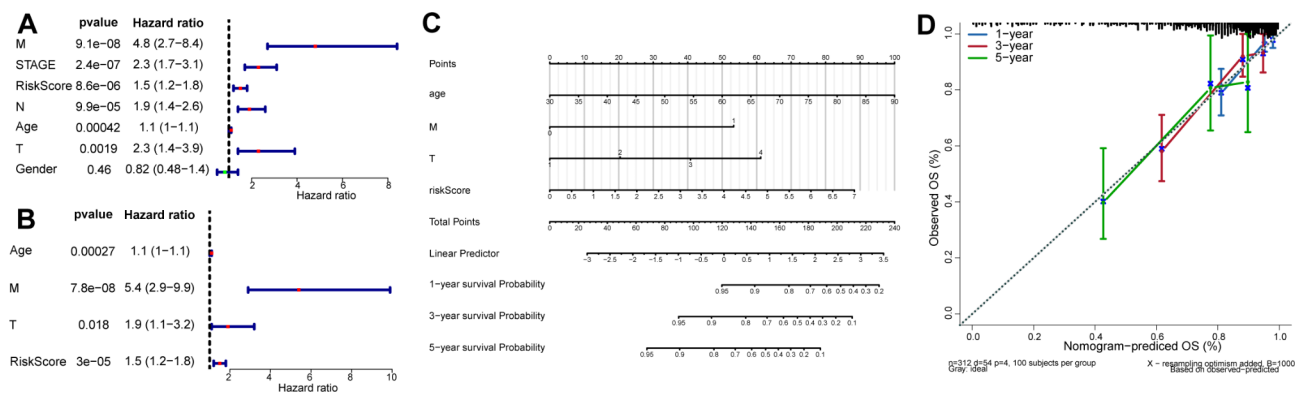


Fig. 5 Development of a predictive nomogram. **A-B.** Forest plots showing the univariate and multivariate Cox analyses for identifying independent risk factors. **C.** A nomogram for predicting 1-, 3-, and 5-year OS probability in CRC. **D.** The calibration curve shows the performance of the nomogram for predicting 1-, 3-, and 5-year OS

Development of a predictive nomogram

Finally, we also constructed a nomogram to predict 1-, 3-, and 5-year OS probability in the training set by including age, gender, and T/N/M stages that were selected as the independent risk factor in CRC based on univariate and multivariate Cox analyses (Fig. 5A-C). The calibration curve indicated the good performance of the nomogram for predicting 1-, 3-, and 5-year OS (Fig. 5D).

Discussion

CRC is one of the most common malignant tumors worldwide among men and women, with the aging population, dietary habits, obesity, lack of physical exercise, and smoking, CRC has been frequently diagnosed, in China with increased incidence and mortality in recent years [34, 35]. Despite the inspiring improvements in diagnosis and therapy of CRC so far, the survival rate in patients with CRC remains low, the overall survival for advanced CRC without metastasis to 3 years [8], and metastatic CRC exhibits approximately 70–75% of patients survive beyond 1 year, 30–35% beyond 3 years,

and fewer than 20% beyond 5 years from diagnosis [36]. We need to identify novel biomarkers and mechanisms for the development and optimization of CRC preventive strategies.

Lymphangiogenesis is an important event during the metastasis in CRC [37], and anti-lymphangiogenic therapies become novel therapeutic strategies [38]. However, the role of lymphangiogenesis contributes to the progression of CRC remains unclear. It has been found that the interaction between the tumor and lymphatic system affects tumor immunity, such as tumor microenvironment and the immune response, and then regulates tumor metastasis [39, 40]. Nevertheless, the biomarkers that are associated with the interaction between the lymphatic system and tumor immune microenvironment are still unknown. In the present study, we construct a WGCNA to identify three modules associated with Lym-Score and the immune cell infiltration level, then 1076 hub genes from the modules. The biological function analysis has indicated that those genes were associated with extracellular matrix and structure, cell-substrate

adhesion, regulation of angiogenesis and vasculature development, positive regulation of cell adhesion, and so on. The extracellular matrix proteins, such as polydome and heparinase, were involved in lymphatic vessel remodeling or lymphangiogenesis [41, 42].

Then, with the survival analysis, the prognostic valuable biomarkers (NUMBL, PCDH18, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB) were selected and then NUMBL, TSPAN11, PHF21A, PDGFRA, ZNF385A, and RIMKLB were identified as the LymScore and immune cell infiltration-related gene signature. We also demonstrated that high expression of NUMBL and PHF21A, but low expression of PDGFRA in tumor tissues compared with normal tissues. NUMBL-like (NUMBL) is a docking protein with a PTB domain that serves as a Notch signaling inhibitor, the aberrant expression of NUMBL can promote carcinogenesis by the dysregulation of the WNT-Notch signaling cycle [43]. Besides, NUMBL plays a crucial role in various processes, such as targeting proteins for ubiquitination and endocytosis, asymmetric cell division, cell migration, and cell adhesion [44–46]. TSPAN11 is one number of the Tetraspanins (TSPANs), which are a class of four transmembrane segmented proteins, TSPANs play dual roles in pan-cancer [47]. A previous study has demonstrated that TSPAN11 was downregulated in pan-cancer and positively associated with CD4+T cells, macrophages, neutrophils, and dendritic cells [48]. PHD finger protein 21 A (PHF21A), also known as BHC80, is a molecule that recognizes unmethylated H3K4 and plays an important role in neuronal and craniofacial development [49, 50]. It has been found that PHF21A serves as a biomarker for the progression and prognosis of hepatocellular carcinoma (HCC) [51]. PDGFRA is a classical proto-oncogene that encodes receptor tyrosine kinases responding to platelet-derived growth factor (PDGF) [52]. Previous study has revealed that high frequency of PDGFRA mutations in high-grade gliomas [53], and overexpression of PDGFRA in oral cancer patients [54]. Moreover, PDGFRA serves as a therapeutic target in young CRC patients [55]. Zinc finger protein 385 A (ZNF385A) is an RNA-binding Cys2 His2 (C2H2) zinc finger protein that is involved in the regulation of cell cycle and apoptosis and acts as a prognostic biomarkers that are associated with immunosuppression in HCC [56]. RimK-like family member B (RIMKLB) is an enzyme that post-translationally modulates ribosomal protein S6 and contributes to the regulation of immune cell development [57]. RIMKLB has been identified as a prognostic biomarker in CRC [58–61]. Our results were consistent with previous findings. Furthermore, based on the LymScore and immune cell infiltration-related gene signature, patients were divided into high-risk and low-risk groups. The patients with high-risk scores showed poor survival, tumor metastasis,

and more severe stages compared to those with low-risk scores. We also constructed and validated a predictive nomogram for predicting 1-, 3-, and 5-year OS probability in CRC. Our finding confirmed that lymphangiogenesis is significantly associated with poor prognosis [62, 63].

Despite our findings elucidating the crucial role of lymphangiogenesis in CRC and identifying the prognostic value of LymScore and immune cell infiltration-related genes, there remain some limitations. Firstly, our analyses were performed based on TCGA and GEO online databases, more explicit clinical data should be collected for validating the expression levels of risk signatures and their prognostic values in the following study. Secondly, our study does not include experimental evidence to demonstrate the expression levels and regulatory mechanisms of risk signatures, we further explore the expression levels and regulatory mechanisms of risk signatures in vitro and in vivo in the following study.

Conclusion

Our finding provides lymphangiogenesis and immune cell infiltration level-related biomarkers for predicting the prognosis of CRC patients and supplies novel therapeutic strategies for anti-lymphangiogenic therapies in CRC.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-023-01781-8>.

Supplementary Material 1: Table S1 Lymphangiogenesis-related genes. **Table S2** lymphangiogenesis score. **Table S3** ssGSEA of immune cell infiltration. **Table S4** Hub modules. **Table S5** Biological functions. **Table S6** Univariate Cox analysis. **Table S7** Correlation risk score and clinical characteristics

Supplementary Material 2: Figure S1. Construction of a WGCNA. A. Sample clustering of genes from the TCGA database to identify outliers. B. Clustering dendrogram of CRC samples and associated clinical traits that included Lymphangiogenesis score (LymScore) and immune cell infiltration levels

Supplementary Material 3: Figure S2. Construction of a co-expression network of LymScore and immune cell infiltration. A. Clustering dendrograms of genes based on the topological overlap and together with assigned module colors. B. The scale-free fit index for soft-thresholding powers. C. Clustering dendrograms of module eigengenes based on the topological overlap. D. Hierarchical clustering dendrograms of genes based on optimal soft-thresholding power

Acknowledgements

Not application.

Author contributions

Yinggang Sun designed the research, and reviewed and revised the manuscript. Hong Liu and Huiwen Shi performed the bioinformatics analyses and wrote the manuscript. Hong Liu and Huiwen Shi downloaded and managed the gene expression, and clinical data from public databases and participated in the discussion. All authors read and approved the final version of the manuscript.

Funding

Not application.

Data Availability

The datasets used in this study are available from the public databases and can be found here: TCGA (<https://portal.gdc.cancer.gov/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/>), and MSigDB, (<https://www.gsea-msigdb.org/gsea/msigdb/>). The supplementary files are stored at the Git-hub (<https://github.com/DrSun2023/Supplementary-Tables.git>).

Declarations**Ethics statement**

Not application.

Competing interests

The authors declare no competing interests.

Author details

¹Department of General Surgery, Wuxi Fifth People's Hospital Affiliated to Jiangnan University, Wuxi, Jiangsu, China

²Department of General Surgery, No.971 Hospital of PLA Navy, Qingdao, China

³Department of General Surgery, The 960th Hospital of Joint Logistics Support Force of Chinese People's Liberation Army, Jinan, China

Received: 7 August 2023 / Accepted: 16 December 2023

Published online: 02 January 2024

References

- Sung H, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
- Morgan E, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut.* 2023;72(2):338–44.
- Bray F et al. *World cancer report 2014* 2014.
- Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index. *Int J Cancer.* 2016;139(11):2436–46.
- Siegel RL, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(3):145–64.
- Clinton SK, Giovannucci EL, and S.D.J.T.J.o.n. Hursting, *The world cancer research fund/American institute for cancer research third expert report on diet, nutrition, physical activity, and cancer: impact and future directions* 2020;150(4):663–671.
- Ugai T, et al. Is early-onset cancer an emerging global epidemic? Current evidence and future implications. *Nat Rev Clin Oncol.* 2022;19(10):656–73.
- Dekker E, et al. Colorectal cancer. *Lancet.* 2019;394(10207):1467–80.
- Doubeni CA, et al. Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut.* 2018;67(2):291–8.
- Levin TR, et al. Effects of organized colorectal cancer screening on cancer incidence and mortality in a large community-based population. *Gastroenterology.* 2018;155(5):1383–1391.e5.
- Miyahara M, et al. Tumor lymphangiogenesis correlates with lymph node metastasis and clinicopathologic parameters in oral squamous cell carcinoma. *Cancer.* 2007;110(6):1287–94.
- Zhang S, et al. High lymphatic vessel density and presence of lymphovascular invasion both predict poor prognosis in breast cancer. *BMC Cancer.* 2017;17(1):335.
- Hou Q et al. Extracellular Hsp90 α promotes tumor lymphangiogenesis and lymph node metastasis in breast cancer. *Int J Mol Sci.* 2021;22(14).
- Lund AW, et al. Lymphatic vessels regulate immune microenvironments in human and murine melanoma. *J Clin Invest.* 2016;126(9):3389–402.
- Wagner M, Steinskog ES, Wiig H. Blockade of lymphangiogenesis shapes tumor-promoting adipose tissue inflammation. *Am J Pathol.* 2019;189(10):2102–14.
- Dieterich LC, et al. Lymphatic vessels in cancer. *Physiol Rev.* 2022;102(4):1837–79.
- Tammela T, Alitalo K. Lymphangiogenesis: molecular mechanisms and future promise. *Cell.* 2010;140(4):460–76.
- Yuan Z, et al. Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. *Mol Cancer.* 2023;22(1):48.
- Gillot L, et al. The pre-metastatic niche in lymph nodes: formation and characteristics. *Cell Mol Life Sci.* 2021;78(16):5987–6002.
- Lund AW et al. Lymphatic vessels, inflammation, and immunity in skin cancer. 2016;6(1):22–35.
- Swartz MA. J.C.i.r. Immunomodulatory roles of lymphatic vessels in cancer progression. 2014;2(8):701–7.
- Lund AW, et al. Lymphatic vessels regulate immune microenvironments in human and murine melanoma. 2016;126(9):3389–402.
- Bieniasz-Krzywiec P, et al. Podoplanin-expressing macrophages promote lymphangiogenesis and lymphoinvasion in breast cancer. *Cell Metab.* 2019;30(5):917–936.e10.
- Hwang I, et al. Tumor-associated macrophage, angiogenesis and lymphangiogenesis markers predict prognosis of non-small cell lung cancer patients. *J Transl Med.* 2020;18(1):443.
- Zhang Q, et al. ETV4 mediated tumor-associated neutrophil infiltration facilitates lymphangiogenesis and lymphatic metastasis of bladder cancer. *Adv Sci (Weinh).* 2023;10(11):e2205613.
- Chen C, et al. LNMAT1 promotes lymphatic metastasis of bladder cancer via CCL2 dependent macrophage recruitment. *Nat Commun.* 2018;9(1):3826.
- Shi L, et al. Expression of vascular endothelial growth factor C in renal cell carcinoma and its correlation with pathological parameters and prognosis. *Transl Androl Urol.* 2020;9(4):1670–7.
- Chen K, et al. A novel lymphangiogenesis-related gene signature can predict prognosis and immunosuppressive microenvironment in patients with clear cell renal cell carcinoma. *Int J Med Sci.* 2023;20(6):754–70.
- Song J, et al. CCBE1 promotes tumor lymphangiogenesis and is negatively regulated by TGF β signaling in colorectal cancer. *Theranostics.* 2020;10(5):2327–41.
- Ou JJ, et al. Neuropilin-2 mediates lymphangiogenesis of colorectal carcinoma via a VEGFC/VEGFR3 independent signaling. *Cancer Lett.* 2015;358(2):200–9.
- Ding W, Tang W, Zhi J. The lymphangiogenic factor CCBE1 promotes angiogenesis and tumor growth in colorectal cancer. *Curr Mol Med.* 2022;22(9):819–25.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
- Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284–7.
- Li N, et al. Incidence, mortality, survival, risk factor and screening of colorectal cancer: a comparison among China, Europe, and northern America. *Cancer Lett.* 2021;522:255–68.
- Cao W, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J (Engl).* 2021;134(7):783–91.
- Billir LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA.* 2021;325(7):669–85.
- Huang C, Chen Y. Lymphangiogenesis and colorectal cancer. *Saudi Med J.* 2017;38(3):237–44.
- Sundlisaeter E, et al. Lymphangiogenesis in colorectal cancer—prognostic and therapeutic aspects. *Int J Cancer.* 2007;121(7):1401–9.
- Lund AW, Swartz MA. Role of lymphatic vessels in Tumor immunity: passive conduits or active participants? *J Mammary Gland Biol Neoplasia.* 2010;15(3):341–52.
- Algars A, et al. Type and location of tumor-infiltrating macrophages and lymphatic vessels predict survival of colorectal cancer patients. *Int J Cancer.* 2012;131(4):864–73.
- Morooka N, et al. Polydom is an extracellular matrix protein involved in lymphatic vessel remodeling. *Circ Res.* 2017;120(8):1276–88.
- Hunter KE, et al. Heparanase promotes lymphangiogenesis and tumor invasion in pancreatic neuroendocrine tumors. *Oncogene.* 2014;33(14):1799–808.
- Katoh M, Katoh M. NUMB is a break of WNT-Notch signaling cycle. *Int J Mol Med.* 2006;18(3):517–21.
- Jafar-Nejad H, Norga K, Bellen H. Numb: Adapting notch for endocytosis dev cell. 2002;3(2):155–6.
- Belle VA, et al. NUMB inhibition of NOTCH signalling as a therapeutic target in prostate cancer. *Nat Rev Urol.* 2014;11(9):499–507.

46. Xin Y, et al. SLC8A1 antisense RNA 1 suppresses papillary thyroid cancer malignant progression via the FUS RNA binding protein (FUS)/NUMB like endocytic adaptor protein (Numb) axis. *Bioengineered*. 2022;13(5):12572–82.
47. Huang R, et al. The role of tetraspanins pan-cancer. *iScience*. 2022;25(8):104777.
48. Liu J, et al. Identification and development of a novel invasion-related gene signature for prognosis prediction in colon adenocarcinoma. *Cancer Cell Int*. 2021;21(1):101.
49. Garay PM, Wallner MA, Iwase S. Yin-Yang actions of histone methylation regulatory complexes in the brain. *Epigenomics*. 2016;8(12):1689–708.
50. Vallianatos CN, Iwase S. Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. *Epigenomics*. 2015;7(3):503–19.
51. Huang JQ, et al. PHF21A expression as a biomarker of hepatocellular carcinoma progression and prognosis. *Neoplasma*. 2022;69(6):1349–58.
52. Guérit E, et al. PDGF receptor mutations in human diseases. *Cell Mol Life Sci*. 2021;78(8):3867–81.
53. Chen CCL, et al. Histone H3.3G34-Mutant Interneuron progenitors co-opt PDGFRA for gliomagenesis. *Cell*. 2020;183(6):1617–1633e22.
54. Ong HS, et al. PDGFRA mRNA is overexpressed in oral cancer patients as compared to normal subjects with a significant trend of overexpression among tobacco users. *J Oral Pathol Med*. 2017;46(8):591–7.
55. Kim TW, et al. The role of PDGFRA as a therapeutic target in young colorectal cancer patients. *J Transl Med*. 2021;19(1):446.
56. Peng Q et al. ZNF385A and ZNF346 serve as prognostic biomarkers associated with an inflamed immunosuppressive tumor microenvironment in hepatocellular carcinoma. *Int J Mol Sci*, 2023;24(4).
57. Kino K, Arai T, Arimura Y. Poly-alpha-glutamic acid synthesis using a novel catalytic activity of RimK from *Escherichia coli* K-12. *Appl Environ Microbiol*. 2011;77(6):2019–25.
58. Li N, et al. Comprehensive analysis of differentially expressed profiles of mRNA N6-Methyladenosine in colorectal cancer. *Front Cell Dev Biol*. 2021;9:760912.
59. Huang C, Zhao J, Zhu Z. Prognostic nomogram of prognosis-related genes and clinicopathological characteristics to predict the 5-Year survival rate of colon cancer patients. *Front Surg*. 2021;8:681721.
60. Ren Y, et al. A prognostic model for Colon adenocarcinoma patients based on ten amino acid metabolism related genes. *Front Public Health*. 2022;10:916364.
61. Cao Y, et al. The prognostic significance of RIMKLB and related immune infiltrates in colorectal cancers. *Front Genet*. 2022;13:818994.
62. Royston D, Jackson DG. Mechanisms of lymphatic metastasis in human colorectal adenocarcinoma. *J Pathol*. 2009;217(5):608–19.
63. Li X, et al. Roles of VEGF-C and Smad4 in the lymphangiogenesis, lymphatic metastasis, and prognosis in colon cancer. *J Gastrointest Surg*. 2011;15(11):2001–10.

Publisher's Note

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