ORIGINAL ARTICLE

WILEY

Identification of hub genes associated with neutrophils infiltration in colorectal cancer

Hao Su[1] [iD] | Tianyi Cai[2] | Sen Zhang[1] | Xialin Yan[3] | Leqi Zhou[4] | Zirui He[1] | Pei Xue[1] | Jianwen Li[1] | Minhua Zheng[1] | Xiao Yang[1] | Bo Feng[1]

[1]Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Department of General Surgery, Zhongshan Hospital, Fudan University, Shanghai, China

[3]Department of General Surgery, Tongji University Tenth People's Hospital, Shanghai, China

[4]Department of General Surgery, Changhai Hospital, Shanghai, China

Correspondence
Bo Feng and Xiao Yang, Department of General Surgery, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai, China.
Emails: fengbo2022@163.com; yangxiao19890125@163.com

Abstract

Colorectal cancer (CRC) is the leading cause of cancer-related mortality in the world. Accumulating evidence indicate that tumour infiltrating immune cells participated in cancer progression. Among them, tumour infiltrating neutrophils (TINs) are reported to play crucial role in various cancers. In this study, we used CIBERSORTx, a digital cytometry tool to evaluate the neutrophils infiltration in CRC based on gene expression data of CRC tissues from GSE39582 data set and The Cancer Genome Atlas data set (TCGA-COAD and TCGA-READ). Weighted gene co-expression network analysis (WGCNA) was conducted in GSE39582 data set to identify hub genes associated with neutrophil infiltration. The association of hub gene and neutrophils was then validated in TCGA cohorts and an independent RJ cohort. Functional analysis was performed to investigate the molecular mechanisms of the interested hub gene. We found that neutrophil infiltration is elevated in CRC tissues, and it is related to a poorer prognosis. A total of 18 gene modules are identified by WGCNA in GSE39582 data set, among which lightcyan module is significantly correlated with neutrophils infiltration. Furthermore, Superoxide Dismutase 2 (SOD2) in lightcyan module was proved to correlated with neutrophils infiltration in various cancer types. In addition, SOD2 expression is highly associated with several chemokines, including CXCL8, a neutrophils-related attractant, and functional analysis revealed that SOD2 is involved in neutrophils recruitment biological process. These results indicate that an 'SOD2-CXCL8-neutrophil recruitment' axis plays a potential role in colorectal cancer progression.

KEYWORDS
CIBERSORTx, colorectal cancer, immune cells, neutrophils, tumour microenvironment, WGCNA

- -

1 | INTRODUCTION

Accounting for 1 in 10 people suffering from cancer, colorectal cancer (CRC) ranks as the third in terms of cancer incidence and second of in terms of mortality in the world. In 2018, Over 1.8 million new cases are identified, and approximately 881,000 people died of CRC.[1] Despite the improvements of treatment methods, mainly surgical resection and chemotherapy, patients with advanced CRC are still haunted by the recurrence after surgery, which compromise life quality and expectance.[2] In this context, a better understanding of CRC mechanism is needed to develop novel cures.

In recent years, the success of checkpoint-blockade therapy in clinical trials shed light on the critical role of tumour infiltrating immune cells, especially T cells.[3] As a historical milestone of cancer treatment, checkpoint blockade immunotherapies proved that our immune system can be harnessed to antagonize malignant tumour. However, despite these substantial advancements in clinical care, the majority of patients do not respond to immune checkpoint inhibitor.[4]

Increasing evidence suggests that tumorigenesis is often triggered in a tumour microenvironment (TME), which is composed of extracellular matrix, blood or lymphatic vessels, fibroblasts, immune cells and inflammatory cells. Importantly, the role of the crosstalk between tumour cells and the stromal cells in the process of tumour progression, drug resistance and clinical outcomes has been validated by multiple studies.[5] As shown, various types of myeloid cells, such as tumour-associated macrophages (TAMs), tumour-infiltrated neutrophils (TINs) and T cells, play important roles in the above biological processes. Infiltration of cytotoxic CD8 + T cells has been associated with a favourable prognosis of CRC patients, mainly because of the anti-tumour activity of CD8 + T cells. Gastric cancer patients with higher TINs were prone to overall survival benefit from postoperative adjuvant chemotherapy. A number of studies have shown that high levels of circulating neutrophils and neutrophil-to-lymphocyte ratio in peripheral blood are associated with poor prognosis in several types of cancers. However, the clinical significance of TINs in tumour stromal tissues is controversial.[6]

Activated neutrophils can release multiple toxic components, including MPO, H2O2, cationic proteins defensins and NET to kill the invaders. However, in the context of tumour, neutrophils can be manipulated by tumour cells and switch into pro-tumour status.[7] Emerging evidence indicates that neutrophil plays a contradictory role in TME. Other than the cytotoxicity mentioned above, it can degrade extracellular matrix by secretion of MMPs and elastase, facilitating tumour cells dissemination. It also promotes angiogenesis mediated by releasing VEGF, HGF and MMP9. Moreover, neutrophil mediates immune suppression via the secretion of ROS and Arginase 1 to limit T cell–dependent anti-tumour immunity.[7,8] These pro-tumour effects might be the potential therapeutic target in the treatment of cancer.

Plenty of studies described the phenomenon of neutrophils recruitment to inflammation or invasion spot by robust chemotactic molecules, most importantly CXCL1, CXCL2, CXCL5 and CXCL8.[9] Tumours orchestrate neutrophils recruitment by releasing neutrophil-related chemokines. For example, TME hypoxia induced CXCL8 overexpression through hypoxia inducible factor 1(HIF1) signalling pathway.[10] However, different tumours may allure neutrophils in various ways, and little research work has been done to investigate the molecular mechanisms of interaction between colorectal cancer and neutrophils. In this research, we use bioinformatic tools to identify the genes potentially associated with neutrophil infiltration in colorectal cancer. Moreover, the interested hub gene was validated in CRC cases from TCGA cohort and real-world validation cohort.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

Gene expression data of 9164 sample of 32 types of cancer (including 603 colorectal cancer patients, COAD and READ) from TCGA were obtained from Xena (https://xenabrowser.net/datapages/). Transcriptome data of 566 colorectal cancer (CRC) samples with clinical information from GSE39582 data set were downloaded from Gene Expression Omnibus (GEO).[11] RNA sequencing data of 17 382 samples of 30 different types of tissue from 6426 donators were obtained from Genotype-Tissue Expression (GTEx) project,[12] and transcriptome of 917 cancer cell line derived from tumour in 24 different sites from Cancer Cell Line Encyclopedia (CCLE)[13] was obtained through GSE36133.

Tumor Immune Estimation Resource (TIMER) was used for validation of our testing results.[6,14,15] Additionally, formalin-fixed, paraffin-embedded specimen arrays of 79 paired CRC tumour and normal tissues from our centre were used in this study (RJ cohort). Written informed consent was obtained from every enrolled patient. The study protocol was approved by the Ethical Committee of Ruijin Hospital, Shanghai Jiao Tong University Medical School.

## 2.2 | Evaluation of neutrophil infiltration

In this study, we use CIBERSORTx[16] to evaluate the amount of infiltrated neutrophils in tumour tissue. CIBERSORTx is a digital cytometry tool. The algorithm enables researchers to estimate the faction of many kinds of immune cells inside tumour tissue, just by utilizing the gene expression data from bulk tissue, without running an actual flow cytometry. In this research, we evaluated the neutrophils infiltration in CRC tissue using data from TCGA and GSE39582. The calculation results come with a p-value for each sample. Samples with p-values < 0.01 were selected for further analysis.

2.3 | Tumour purity estimation

Tumour purity was evaluated using R package Estimate.[17] ESTIMATE is a R software for evaluation of tumour purity, stromal

FIGURE 1 Main design of the study. GSE39682 data set was used to identify hub genes associated with TINs. TCGA COREAD cohort were used for validation, DEGs calculation and GSEA analysis. GTEx, CCLE and TIMER database were used for further validation. RJ cohort were utilized for immunohistochemistry
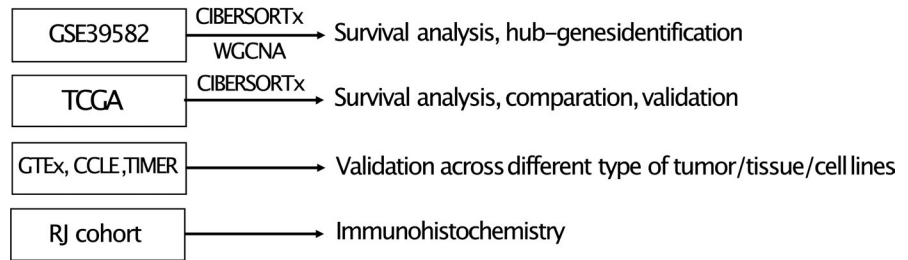
cells proportion and the infiltration level of immune cells in tumour tissues based on expression data. In this research, we use ESTIMATE to evaluate the tumour purity of samples in GSE39582 data set.

2.4 | Co-expression network analysis and visualization

The WGCNA R software package was used to construct the co-expression network of the genes in the GSE39582 data set. We selected the top 6000 genes according to the rank of variance (from large to small). A total of 459 tumour samples in the GSE39582 data set were included in the following analysis. After proper samples and genes were chosen, adjacency matrix was constructed by calculating the Pearson correlation between all pairs of genes in all selected samples using the WGCNA R package. Then, the power of power of $\beta = 5$ (scale-free R2 = 0.8970) was used as a soft threshold to ensure a scale-free network. To further identify functional modules in the co-expression network with these 6000 genes, the adjacency matrix was used to calculate the topological overlap measurement (TOM) representing the overlap in the shared neighbours.

The module eigengenes (MEs) were considered to be a representation of the gene expression profile in the module. Correlation and P-value between the module and neutrophils scores were calculated using MEs and neutrophils score.[18,19] The gene module highly correlated with the neutrophils was selected for visualization using Cytoscape.[20]

2.5 | Differentially expressed genes (DEGs) and Gene Set Enrichment Analysis (GSEA)

The expression data of TCGA colorectal cancer patient were downloaded and processed using R package DESeq2[21] to calculate differentially expressed genes. The genes were then ordered according to log2(Foldchange), and the gene list was subjected to gseGO function from clusterProfiler[22] package for GSEA analysis.

2.6 | Immunohistochemical staining and histochemistry scoring

CD66b served as specific biomarker for neutrophils in TME. Tissue Microarray sections were deparaffinized and dehydrated, and then treated with 3% H2O2 at room temperature for 10 min to block endogenous peroxidase activity. Next, the tissue sections were incubated with citrate buffer for the retrieval of the antigen. Then, the tissues were blocked with 3% BSA at room temperature for 30 min, followed by incubating with anti-CD66b(ab197678, rabbit; 1:200, Abcam, Cambridge, UK), anti-SOD2 antibody (ab13533, Rabbit, 1:100, Abcam, Cambridge, UK) and anti-CXCL8 antibody(ab18672, Mouse, Abcam, Cambridge, UK) at room temperature for 50min. Finally, the tissue slides were counterstained with haematoxylin. In validation phase, slides were then scanned (Pannoramic MIDI, 3D HISTECH), and intensity of three proteins was evaluated using Quant centre software. The expression level of each genes was quantified as 1, 2 and 3, representing weak, moderate and strong expression, respectively, and then, the histochemistry scores(H-Score) of SOD2 and CXCL8 of calculated using the method mentioned previously.[23] Intratumoral neutrophils are identified as CD66b-positive cells with segmented nuclear. When calculating the correlation coefficient of neutrophil infiltration and genes, the amounts of infiltrated neutrophils were evaluated by the mean density of strong-positive pixels of CD66b.

2.7 | Statistics and data visualization

All statistical analyses were performed with R software (version 3.6.2). Wilcoxon test was used to compare the neutrophil scores/counts in different sample groups. Fisher's exact test was used in the categorical clinical traits comparison between the high or low neutrophil group. The Kaplan-Meier method and log-rank were used to present and evaluate the differences of survival between the high or low-neutrophil groups. The cutpoint to determine TINs high/low groups was calculated using 'surv_cutpoint' function from R package, survminer.[24] Data visualization was performed based on ggplot2[25] and ggpubr [26] packages.

3 | RESULTS

3.1 | Study design

The design of the study was shown in Figure 1. Gene expression data from GSE39582 (n = 566) and TCGA colorectal patients (n = 603) were subjected to CIBERSORTx to get neutrophils infiltration scores. After identifying hub genes associated with using WGCNA in GSE39582 data set, the correlation between hub genes and neutrophils infiltration was then validated using database including TIMER, GTEx and CCLE. Samples of RJ Cohort from our hospital were used for immunohistochemical validation. In RJ cohort, a total of 79 patients were enrolled, about 59% of the patients were aged over 60 years, 56% of the patients were male (Table S1).

3.2 | Higher TIN levels indicating poorer prognosis of CRC patients

As shown, neutrophils score calculated by CIBERSORTx, representing amount of neutrophil infiltration, is significantly higher in tumour tissue compared with normal tissue, indicating an increasing neutrophil recruitment in CRC tumour tissue (Figure 2A). Moreover, higher neutrophil score indicates poorer prognosis in both GSE39582 data set

(Figure 2B-2C) and TCGA data set(Figure 2D-2F). These findings highlight the clinical relevance of TINs and the necessity to investigate detailed mechanism of neutrophil-tumour interaction in colorectal cancer.

3.3 | Neutrophil associated Hub gene identification

Gene expression data from GSE39582 data set were subjected to CIBERSORTx and ESTIMATE algorithm to evaluate neutrophil infiltration and tumour purity. We discovered that tumour purity varies dramatically across samples, arranging from 0.238 to 0.972. As tumour purity might have a great influence on the neutrophil score calculated by CIBERSORTx, the heterogeneity of tumour samples might compromise the robustness of WGCNA. Thus, we filtered out the samples with tumour purity less than 0.5. Also, we discarded samples with p-values (derived from CIBERSORTx calculation) greater than 0.01. Finally, 459 samples were subjected to WGCNA analysis.

In this study, the power of $\beta = 5$ (scale-free $R^2 = 0.8970$) was selected (Figure 3A) to ensure a scale-free network. A total of 18 gene modules represented by 18 different colours were successfully identified(Figure 3B), in which the lightcyan module was significantly correlated with neutrophil score (Figure 3C). Genes in lightcyan module were exported and visualized in Cytoscape (Figure 3D). As expected, genes highly expressed in neutrophils
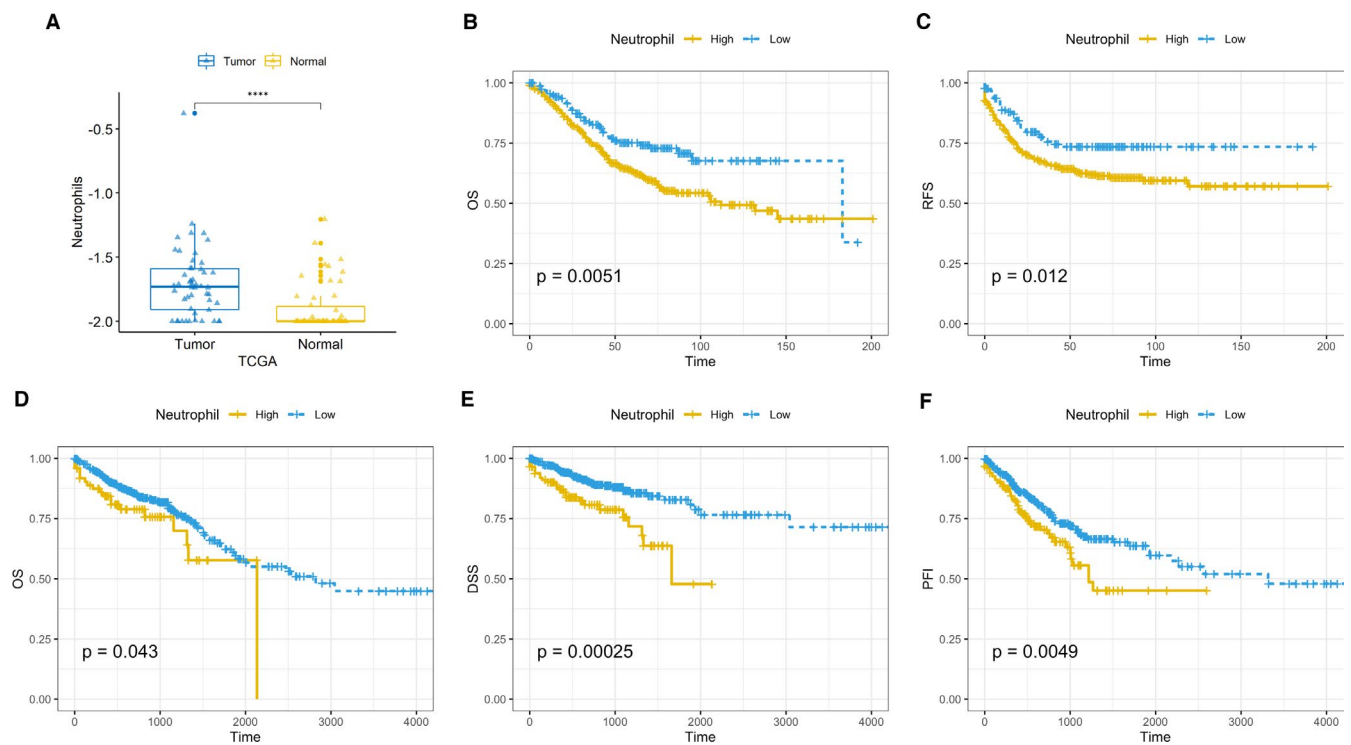

FIGURE 2 Neutrophils infiltration is higher in tumour tissue and associated with poorer prognosis. A, Neutrophil infiltration score counted by CIBERSORTx in 50 paired tumour-normal tissue in TCGA COAD and READ cohort. Y-axis represents log10(neutrophil scores + 0.01). B, Overall survival (OS) in GSE39582 data set. C, Relapse-free survival (RFS) in GSE39582 data set. D, OS in TCGA COREAD cohort. E, Disease-specific survival (DSS) in TCGA COREAD cohort. F, Progression-free interval event (PFI) in TCGA COREAD cohort. *Kaplan-Meier, log-rank P-values were calculated using Wilcoxon tests
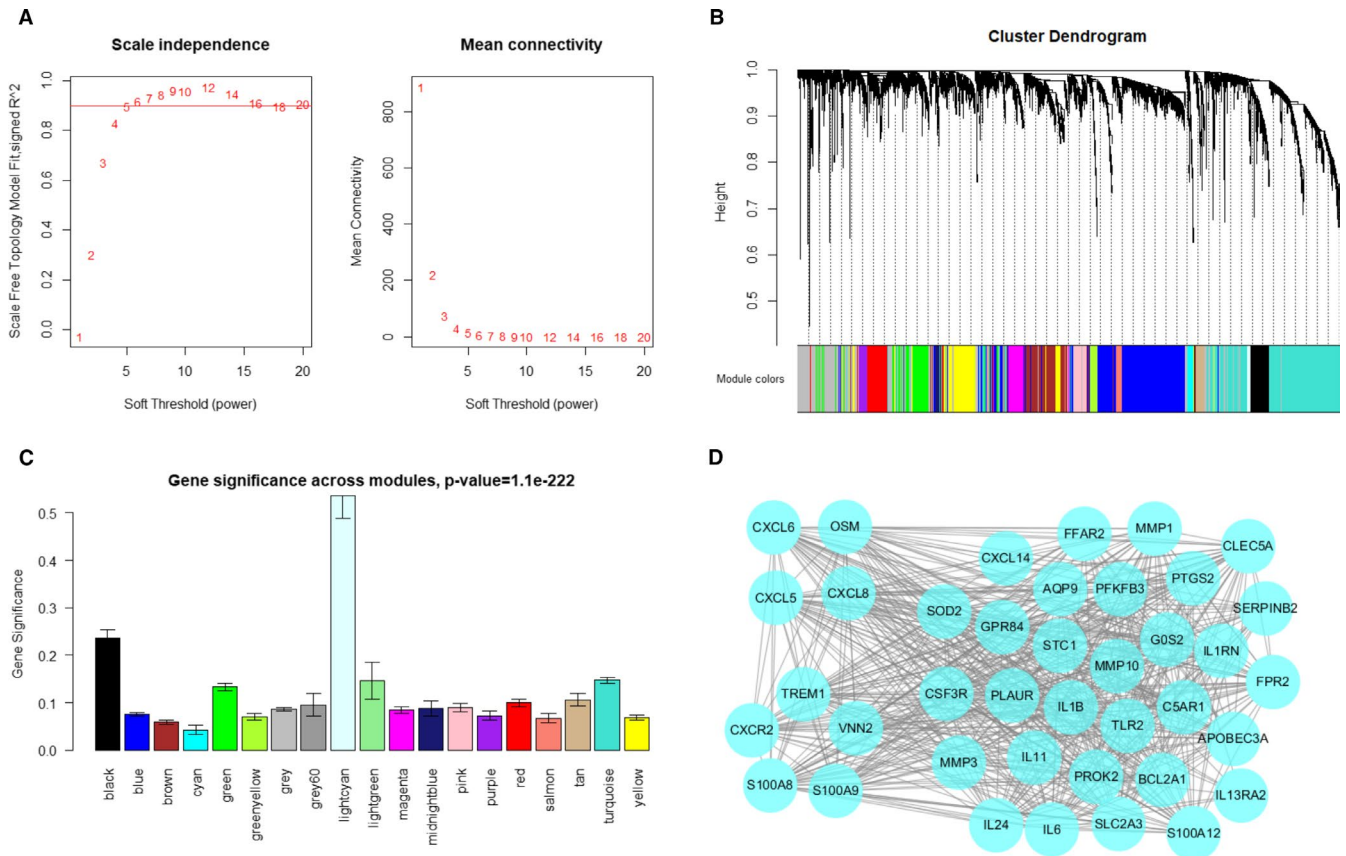
FIGURE 3 Identifying hub genes associated with neutrophil infiltration in GSE39582 data set. A, Analysis of the scale-free fit index and the mean connectivity for various soft-thresholding powers. B, Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure. Dynamic tree cutting was applied to identify modules by dividing the dendrogram at significant branch points. Modules are displayed with different colours in the horizontal bar immediately below the dendrogram. C, Distribution of average gene significance in the modules based on neutrophil score derived from CIBERSORTx. D, Visualization of gene network in lightcyan module

FIGURE 4 Association of SOD2 and neutrophils in TCGA cohort. A, Correlation of SOD2 and TINs in TCGA COAD cohort. B, Correlation of SOD2 and TINs across 32 different type of cancer. The vertical line corresponds to a p-value of 0.1
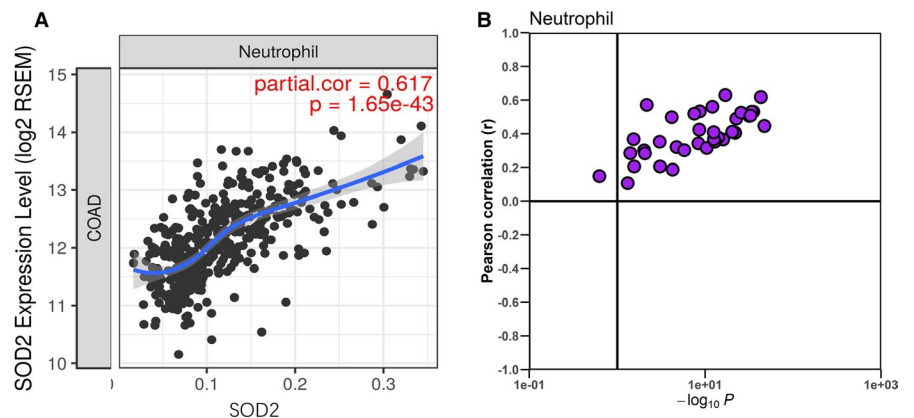
including S100A8, S100A9, VNN2 and TERM1 are included in the module. Besides, three C-X-C Motif Chemokine Ligand members, CXCL5, CXCL6, CXCL8, were also included. These results reflected the robustness of WCGNA.

Theoretically, most of the genes in lightcyan module should be correlated with neutrophil infiltration and neutrophil-attract chemokines, regardless of the expression level. In order to facilitate the

immunohistochemical analysis, we decided to choose a relatively highly expressed gene from this module, SOD2 (Figure S1), for further analysis. Intriguingly, the other members of SOD family, SOD1 and SOD3, are not correlated with neutrophils infiltration. This is an indicator that the correlation between SOD2 and TINs might be irrelevant to oxidative stress induced by neutrophils. Thus, we chose SOD2 for further analysis.

3.4 | Validation of SOD2 associated TME neutrophil infiltration in TCGA and RJ cohort

In order to validate the association of SOD2 and neutrophil infiltration, we firstly tested our findings in TCGA cohort using TIMER. As shown, the expression of SOD2 in CRC tumour tissue is significantly correlated with neutrophils infiltration in CRC tissue (Figure 4A). We further explored the association in 32 different kinds of solid tumours. Intriguingly, SOD2 is related to TINs in most of the cancer types(Figure 4B). This finding indicated a universal role SOD2 in neutrophils infiltration in TME. Intriguingly, SOD1 and SOD3 are not substantially correlated with TINs (Figure S2).

For further verifying, we conducted immunohistochemistry analysis in RJ cohort, which contains 79 cases of CRC patients. As shown

(Figure S3), cells positive for CD66b antibody and with segmented nuclear are identified as neutrophils. As expected, neutrophil count in tumour tissue are significantly higher than that in normal colon tissue(Figure 5A). Expression levels of SOD2 are positively correlated with CXCL8 (Figure 5B, Pearson correlation, R = 0.51, $P <.001$) and neutrophil infiltration (Figure 5C, Pearson correlation, R = 0.44, $P <.001$), as shown in representative images (Figure 5D-5I).

3.5 | Functional investigation of downstream signalling associated with SOD2 expression

TCGA colorectal cohort was used to further investigate the SOD2 related biological activity. Based on the normalized expression level
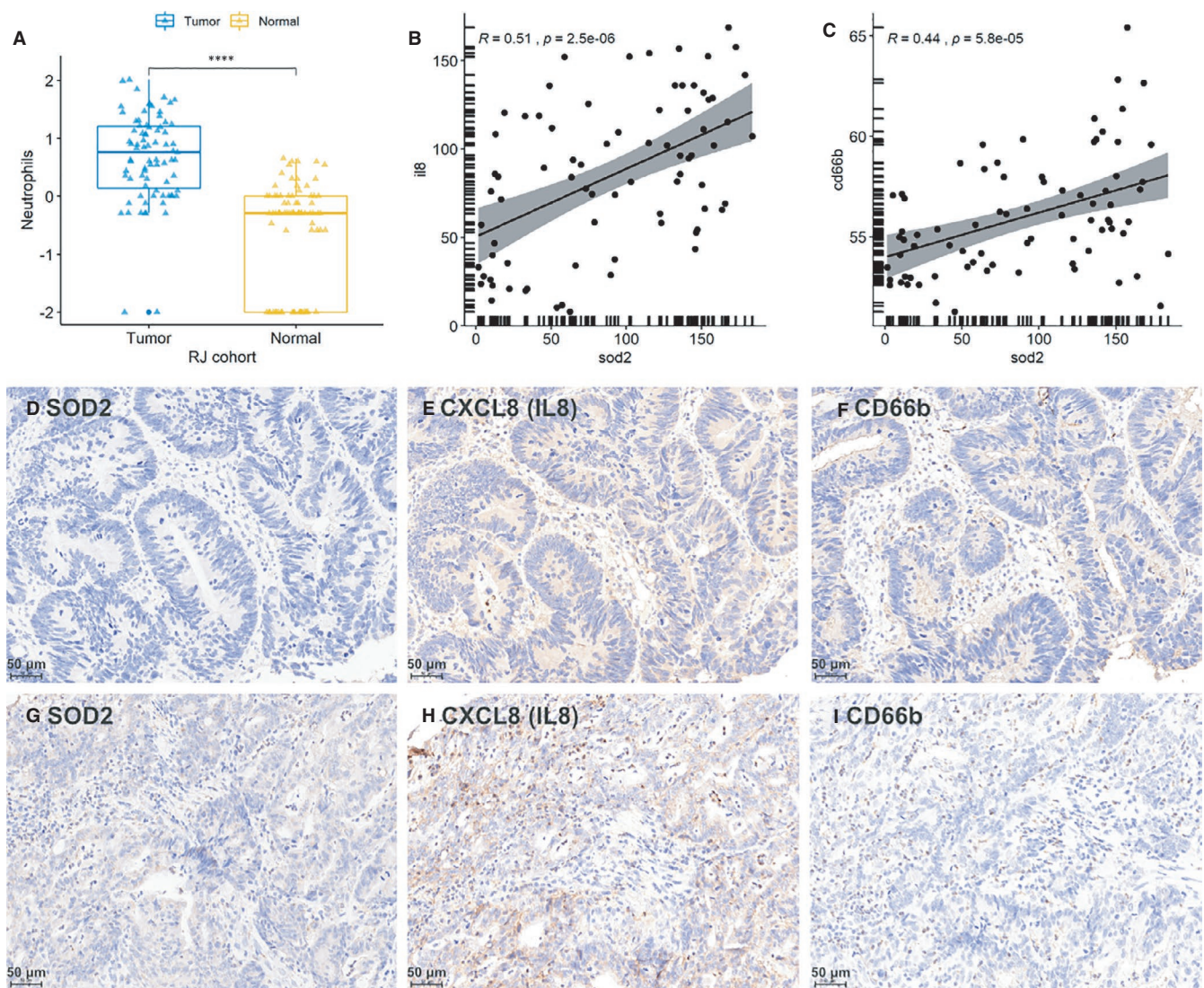

FIGURE 5 Validation of SOD2-CXCL8-Neutrophil infiltration association in RJ cohort. A, Neutrophils counts in tumour tissue and normal tissue in RJ cohort. Y-axis represents log10(neutrophil count + 0.01). B, Pearson correlation of SOD2-CXCL8(IL8) association in RJ cohort (expression levels are quantified by H-score). C, Pearson correlation of SOD2-CD66b association in RJ cohort (expression levels are quantified by H-score). D-F, Representative images of low SOD2 expression, low CXCL8 expression and low neutrophil infiltration in the same tissue region, respectively. G-I, Representative images of high SOD2 expression, high CXCL8 expression and high neutrophil infiltration in the same tissue region, respectively
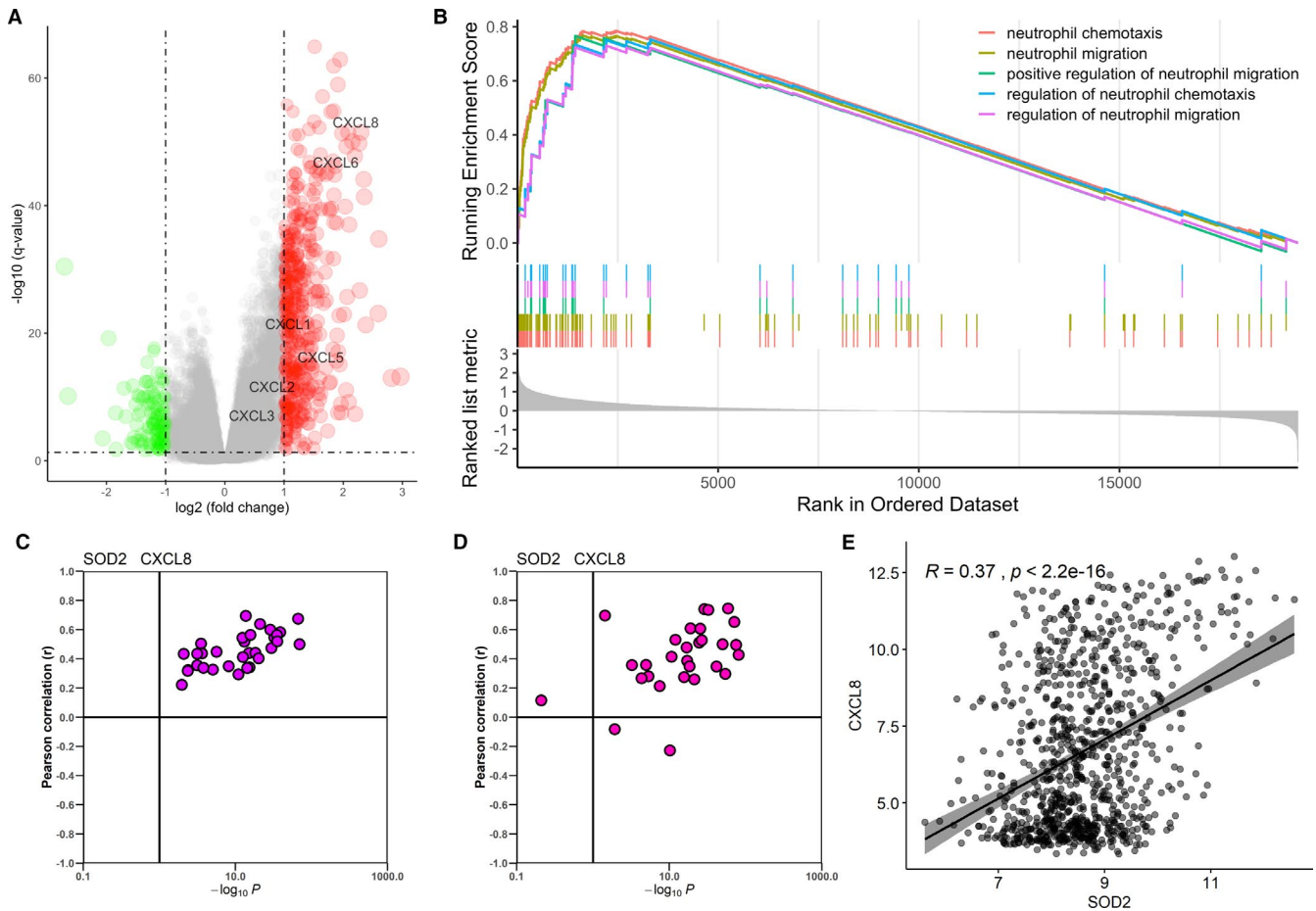
FIGURE 6 Functional exploration of SOD2. A, Volcano plot of Differentially expressed genes (DEGs) of high/low SOD2 group in TCGA COARD cohort. B,GSEA analysis, genes are ordered by logFC value derived from DEGs calculation. C, Correlation of SOD2 and CXCL8 in 32 types of cancer from TCGA data. Vertical line corresponding with a p-value of 0.1. D, Correlation of SOD2 and CXCL8 in 30 types of tissue from GTEx data. Vertical line corresponding with a p-value of 0.1. E, Correlation of SOD2 and CXCL8 in 917 cancer cell lines from CCLE

of SOD2, we divided the samples into 2 groups, with 312 and 313 samples in SOD2 high and low groups, respectively. Differentially expressed genes were calculated using DESeq2 package. As shown in the volcano plot (Figure 6A), we found that CXCL6 and CXCL8 expression levels are significantly higher in SOD2 high group comparing to low group. Gene Set Enrichment Analysis (GSEA) of GO biological process indicates that immunological differences concerning neutrophils recruitment are significantly higher in SOD2 high group (Figure 6B). These findings indicate that a SOD2 involved neutrophil recruitment regulating axis might exist.

As CXCL1, CXCL2, CXCL5, and CXCL8 are frequently reported chemokines to recruit neutrophils, we set our course to explores the expressional correlation between SOD2 and these chemokines. The correlation of SOD2 and the four chemokines were evaluated in CCLE, GTEx and TCGA gene expression data. As shown, the SOD2 are significantly correlated with these chemokines, especially CXCL8, in most of the cancer types and normal tissues (Figure 6C-6E, Figure S4), indicating a universal association between SOD2 and CXCL8.

These results showed that a 'SOD2-CXCL8-Neutrophil infiltration' axis might exists in CRC. However, detailed mechanisms by which SOD2 orchestrating neutrophils infiltration needs further investigation.

4 | DISCUSSION

Due to the success of immune checkpoint inhibitors (ICIs) therapy in clinical trial, CD8 + T cells are put under spotlight in recent years.[3] Although emerging cases showing the efficacy of immunotherapy, a large proportion of patients were not responsive to ICIs. Thus, studies aiming at exploiting alternative therapeutic targets are necessary. Neutrophils, as the first responders of infection and inflammation, have been neglected by tumour biologists for a long time. Recent years, researchers started to pay attention to the role of neutrophils played in cancer progression.[27]

Neutrophils are usually considered as first-line responder against extracellular pathogens. Just as tumour-associated macrophages

(TAMs) being divided into 'M1' anti-tumour phenotype and 'M2' pro-tumour phenotype, emerging evidence exhibit that TINs can polarize into 'N1' anti-tumour phenotype or 'N2' pro-tumour phenotype. In fact, transforming growth factor beta (TGF-β) promote neutrophil differentiation into N2 type, whereas interferon type I (IFN-I) induces N1 type. In a word, the signalling context in the tumour microenvironment can affect the phenotype of neutrophils profoundly.

The increase of neutrophil count in peripheral circulation are found to related to poorer prognosis in CRC patients.[28,29] However, the effect of TINs in CRC patients has been controversial. Rao et al reported that increase in TINs is related to are predictor of poor prognosis in CRC patients,[30] whereas Berry at al. showed that high levels of TINs were associated with improved overall survival.[31]

Several experiments demonstrated that CRC cells recruit neutrophils by secreting chemokines including CCL-9, CCL15, CXCL8.[32,33] Neutrophils also promote the hepatic metastasis of CRC, whereas depletion of neutrophils in murine decreased the vascular density and branching in metastatic site and inhibited the tumour growth.[34] Intriguingly, in CRC patients receiving anti-VEGF (bevacizumab) therapy, the present of CD177 + neutrophils in tumour is related to reduced survival, and depletion of neutrophils restores the sensitivity of anti-VEGF treatment.[35]

Tohme, S et al reported that neutrophil extracellular traps (NETs), complexes of chromosomal DNA, histones and other proteins released by neutrophils, can release High Mobility Box (HMGB)-1 protein and activate of TLR9-dependent pathways in CRC cells to promote their adhesion, proliferation, migration and invasion, thus facilitating liver metastasis.[36]

T cell infiltration in CRC tissue is associated with favourable outcome. As neutrophils frequently colocalize with CD8 + Tcells,[37] the crosstalk between TINs and T cells intrigued researchers in recent years. Due to the plasticity of TINs, the result of neutrophil-T cell interaction in CRC can be controversial. Governa V et al reported that CD8 + T cell activation is enhanced when cocultured with neutrophils in vitro, and combined infiltration of CD66b + neutrophils and CD8 + T lymphocytes in CRC is associated with significantly better survival comparing to CD8 + T cell infiltration alone.[37] However, Neutrophil-derived MMP9 activates TGF-β and consequently induced T cell suppression in CRC.[38] Zhang Y demonstrated that neutrophils selectively deprived of IRAK-M or Tollip significantly activate T cells by enhanced expression of the costimulatory molecule CD80, and reduced expression of the inhibitory molecule PD-L1[39,40] indicating modified neutrophils can be a promising therapeutic strategy in CRC treatment.

Despite the difficulties in TINs investigation, the accumulating evidence is showing the substantial role of neutrophils in tumour initiation and metastasis.[41] This finding raised the hope of inhibiting cancer development by targeting TINs. Thus, it is imperative to elucidate the molecular mechanism regulating the neutrophils and cancer cells interaction. In this study, we found that low neutrophils infiltration in CRC were associated with the better prognosis. With the help of bioinformatic analytical tools, a list of genes highly associated with TINs in colorectal cancer were identified. We then selected SOD2 among all of the genes to conduct a further analysis. The significant correlation of SOD2, CXCL8 and neutrophils infiltration are validated immunohistochemically in both TCGA cohort and RJ cohort.

SOD2 is a member of superoxide dismutase family, located in mitochondria. This gene is known as a scavenger of mitochondrial derived ROS. As ROS are considered to be a crucial regulator of cell signalling network that drive proliferation and malignant progression, the superoxide dismutase plays a putatively crucial role in tumorigenesis. Interestingly, we found that SOD2 is positively correlated with TINs infiltration in most of the cancer type, indicating a universal association between SOD2 and neutrophil in TME. Correspondingly, we found that SOD2 is associated with several neutrophil-attract chemokines. As neutrophils carries cytotoxic materials including H_2O_2, whether the high expression of SOD2 in colorectal cancer is induced by ROS stress derived from neutrophils still needs to be elucidated. Of note, SOD1 and SOD3 have no such association with neither neutrophils nor related chemokines, indicating additional functions of SOD2 in tumour progression except for ROS scavenging.

Altogether, we propose that a 'SOD2-CXCL8-neutrophil recruitment' axis might exist in CRC cells. Our analysis provided insights into the molecular mechanism of crosstalk between neutrophils and tumour, which may facilitate further investigation of TINs in the near future.

5 | CONCLUSION

In summary, based on the comprehensive analysis of transcriptome data using Cibersort and WGCNA algorism, we discovered a list of genes highly associated with TINs. Among which, SOD2 significantly correlated with neutrophil infiltration and neutrophil-related chemokines in CRC, indicating a possible existence of 'SOD2-CXCL8-neutrophil recruitment' in CRC. The findings of our study may facilitate the further research in crosstalk between neutrophil and colorectal cancer cells.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION

Hao Su: Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Writing-original draft (equal). **Tianyi Cai:** Validation (equal); Visualization (equal); Writing-original draft (equal). **Sen Zhang:** Writing-review & editing (equal). **Xialin Yan:** Software (equal). **Leqi Zhou:** Methodology (equal). **Zirui He:** Conceptualization (equal). **Pei Xue:** Resources (equal). **Jianwen Li:** Resources (equal). **MinHua Zheng:** Supervision (equal). **Xiao Yang:** Project administration (equal); Supervision (equal); Visualization (equal). **Bo Feng:** Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing-review & editing (equal).

ETHICAL STATEMENT

Written consent was obtained from all enrolled patients. The study was conducted in accordance with the Declaration of Helsinki, and the Ethic Committee of RuiJin Hospital, Shanghai Jiao Tong University approved the protocol of the study (GCQN-2019-A07).

DATA AVAILABILITY STATEMENT

The RNA sequencing data of CRC patients were downloaded from Xena (http://xena.ucsc.edu/). GSE39582 and GSE36133 were obtained from GEO database (https://www.ncbi.nlm.nih.gov/gds/). Other data analysed during this study are included in this published article and the supplementary material.

ORCID

Hao Su (iD) https://orcid.org/0000-0002-2554-9716

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424. https://doi.org/10.3322/caac.21492
2. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin.* 2017;67(3):177-193. https://doi.org/10.3322/caac.21395
3. Miller BC, Sen DR, Al Abosy R, et al. Subsets of exhausted CD8(+) T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat Immunol.* 2019;20(3):326-336. https://doi.org/10.1038/s41590-019-0312-6
4. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer.* 2019;19(3):133-150. https://doi.org/10.1038/s41568-019-0116-x
5. Ogawa R, Yamamoto T, Hirai H, et al. Loss of SMAD4 Promotes Colorectal Cancer Progression by Recruiting Tumor-Associated Neutrophils via the CXCL1/8–CXCR2 Axis. *Clin Cancer Res.* 2019;25(9):2887-2899. https://doi.org/10.1158/1078-0432.Ccr-18-3684
6. Qin F, Liu X, Chen J, et al. Anti-TGF-β attenuates tumor growth via polarization of tumor associated neutrophils towards an anti-tumor phenotype in colorectal cancer. *J Cancer.* 2020;11(9):2580-2592. https://doi.org/10.7150/jca.38179
7. Granot Z. Neutrophils as a Therapeutic Target in Cancer. *Front Immunol.* 2019;10:1710. https://doi.org/10.3389/fimmu.2019.01710
8. Hurt B, Schulick R, Edil B, El Kasmi KC, Barnett C Jr. Cancer-promoting mechanisms of tumor-associated neutrophils. *Am J Surg.* 2017;214(5):938-944. https://doi.org/10.1016/j.amjsurg.2017.08.003
9. Bonavita O, Massara M, Bonecchi R. Chemokine regulation of neutrophil function in tumors. *Cytokine Growth Factor Rev.* 2016;30:81-86. https://doi.org/10.1016/j.cytogfr.2016.03.012
10. Triner D, Xue X, Schwartz AJ, Jung I, Colacino JA, Shah YM. Epithelial Hypoxia-inducible factor 2alpha facilitates the progression of colon tumors through recruiting neutrophils. *Mol Cell Biol.* 2017;37(5):e00481-16. https://doi.org/10.1128/MCB.00481-16
11. Marisa L, de Reynies A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Medicine.* 2013;10(5):e1001453. https://doi.org/10.1371/journal.pmed.1001453
12. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585. https://doi.org/10.1038/ng.2653

13. Barretina J, Caponigro G, Stransky N, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012;483(7391):603-607. https://doi.org/10.1038/nature11003
14. Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017;77(21):e108-e110. https://doi.org/10.1158/0008-5472.Can-17-0307
15. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016;17(1):174. https://doi.org/10.1186/s13059-016-1028-7
16. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773-782. https://doi.org/10.1038/s41587-019-0114-2
17. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4(1):3612. https://doi.org/10.1038/ncomms3612
18. Langfelder P, Horvath SWGCNA. An R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(1):559. https://doi.org/10.1186/1471-2105-9-559
19. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:4–17. https://doi.org/10.2202/1544-6115.1128
20. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504. https://doi.org/10.1101/gr.1239303
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284-287. https://doi.org/10.1089/omi.2011.0118
23. Yeo W, Chan SL, Mo FK, et al. Phase I/II study of temsirolimus for patients with unresectable Hepatocellular Carcinoma (HCC)- a correlative study to explore potential biomarkers for response. *BMC Cancer.* 2015;15:395. https://doi.org/10.1186/s12885-015-1334-6
24. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.8. 2020. https://CRAN.R-project.org/package=survminer
25. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer; 2016. https://ggplot2.tidyverse.org
26. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. 2020;https://CRAN.R-project.org/package=ggpubr
27. Moses K, Brandau S. Human neutrophils: Their role in cancer and relation to myeloid-derived suppressor cells. *Semin Immunol.* 2016;28(2):187-196. https://doi.org/10.1016/j.smim.2016.03.018
28. Dell'Aquila E, Cremolini C, Zeppola T, et al. Prognostic and predictive role of neutrophil/lymphocytes ratio in metastatic colorectal cancer: a retrospective analysis of the TRIBE study by GONO. *Ann Oncol.* 2018;29(4):924-930. https://doi.org/10.1093/annonc/mdy004
29. Grenader T, Nash S, Adams R, et al. Derived neutrophil lymphocyte ratio is predictive of survival from intermittent therapy in advanced colorectal cancer: a post hoc analysis of the MRC COIN study. *Br J Cancer.* 2016;114(6):612-615. https://doi.org/10.1038/bjc.2016.23
30. Rao HL, Chen JW, Li M, et al. Increased intratumoral neutrophil in colorectal carcinomas correlates closely with malignant phenotype and predicts patients' adverse prognosis. *PLoS One.* 2012;7:e30806.
31. Berry RS, Meng-Jun X, Alissa G, et al. High levels of tumor-associated neutrophils are associated with improved overall survival in patients with stage II colorectal cancer. *PLoS One.* 2017;12(12):e0188799.

32. Ogawa R, Yamamoto T, Hirai H, et al. Loss of SMAD4 promotes colorectal cancer progression by recruiting tumor-associated neutrophils via the CXCL1/8-CXCR2 Axis. *Clin Cancer Res*. 2019;25(9):2887-2899. https://doi.org/10.1158/1078-0432.Ccr-18-3684

33. Yamamoto T, Kawada K, Itatani Y, et al. Loss of SMAD4 promotes lung metastasis of colorectal cancer by accumulation of CCR1+ tumor-associated neutrophils through CCL15-CCR1 Axis. *Clin Cancer Res*. 2017;23(3):833-844. https://doi.org/10.1158/1078-0432.Ccr-16-0520

34. Gordon-Weeks AN, Lim SY, Yuzhalin AE, et al. Neutrophils promote hepatic metastasis growth through fibroblast growth factor 2-dependent angiogenesis in mice. *Hepatology*. 2017;65(6):1920-1935. https://doi.org/10.1002/hep.29088

35. Schiffmann LM, Fritsch M, Gebauer F, et al. Tumour-infiltrating neutrophils counteract anti-VEGF therapy in metastatic colorectal cancer. *Br J Cancer*. 2019;120(1):69-78. https://doi.org/10.1038/s41416-018-0198-3

36. Tohme S, Yazdani HO, Al-Khafaji AB, et al. Neutrophil extracellular traps promote the development and progression of liver metastases after surgical stress. *Cancer Res*. 2016;76(6):1367-1380. https://doi.org/10.1158/0008-5472.Can-15-1591

37. Governa V, Trella E, Mele V, et al. The interplay between neutrophils and CD8(+) T cells improves survival in human colorectal cancer. *Clin Cancer Res*. 2017;23(14):3847-3858. https://doi.org/10.1158/1078-0432.Ccr-16-2047

38. Germann M, Zangger N, Sauvain MO, et al. Neutrophils suppress tumor-infiltrating T cells in colon cancer via matrix metalloproteinase-mediated activation of TGFβ. *EMBO Mol Med*. 2020;12(1):e10681. https://doi.org/10.15252/emmm.201910681

39. Zhang Y, Diao N, Lee CK, Chu HW, Bai L, Li L. Neutrophils deficient in innate suppressor IRAK-M enhances anti-tumor immune responses. *Mol Ther*. 2020;28(1):89-99. https://doi.org/10.1016/j.ymthe.2019.09.019

40. Zhang Y, Lee C, Geng S, Li L. Enhanced tumor immune surveillance through neutrophil reprogramming due to Tollip deficiency. *JCI Insight*. 2019;4(2):e122939. https://doi.org/10.1172/jci.insight.122939

41. Szczerba BM, Castro-Giner F, Vetter M, et al. Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature*. 2019;566(7745):553-557. https://doi.org/10.1038/s41586-019-0915-y

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Su H, Cai T, Zhang S, et al. Identification of hub genes associated with neutrophils infiltration in colorectal cancer. *J Cell Mol Med*. 2021;25:3371–3380. https://doi.org/10.1111/jcmm.16414