

Identification of differentially expressed genes between mucinous adenocarcinoma and other adenocarcinoma of colorectal cancer using bioinformatics analysis

Journal of International Medical Research

48(8) 1–20

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0300060520949036

journals.sagepub.com/home/imr



Xue Zhang, Jing Zuo , Long Wang, Jing Han, Li Feng, Yudong Wang and Zhisong Fan

Abstract

Objective: As a unique histological subtype of colorectal cancer (CRC), mucinous adenocarcinoma (MC) has a poor prognosis and responds poorly to treatment. Genes and markers related to MC have not been reported.

Methods: To identify biomarkers involved in development of MC compared with other common adenocarcinoma (AC) subtypes, four datasets were obtained from the Gene Expression Omnibus database. Differentially expressed genes (DEGs) were identified using GEO2R. A protein–protein interaction network was constructed. Functional annotation for DEGs was performed via DAVID, Metascape, and BiNGO. Significant modules and hub genes were identified using Cytoscape, and expression of hub genes and relationships between hub genes and MC were analyzed.

Results: The DEGs were mainly enriched in negative regulation of cell proliferation, bicarbonate transport, response to peptide hormone, cell–cell signaling, cell proliferation, and positive regulation of the canonical Wnt signaling pathway. The Venn diagram revealed eight significant hub genes: *CXCL9*, *IDO1*, *MET*, *SNAI2*, and *ZEB2* were highly expressed in MC compared with AC, whereas *AREG*, *TWIST1*, and *ZEB1* were expressed at a low level. *AREG* and *MET* might be significant biomarkers for MC.

Conclusion: The identified DEGs might help elucidate the pathogenesis of MC, identify potential targets, and improve treatment for CRC.

Department of Medical Oncology, Hebei Medical University Fourth Affiliated Hospital, Shijiazhuang, Hebei, P. R. China

Corresponding author:

Jing Zuo, Department of Medical Oncology, Hebei Medical University Fourth Affiliated Hospital, 12 Jiankang Road, Shijiazhuang, Hebei 050011, P. R. China.
Email: jingzuohb159@163.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative

Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Keywords

Colorectal cancer, mucinous adenocarcinoma, differentially expressed genes, protein–protein interaction network, bioinformatics analysis, biomarker

Date received: 22 March 2020; accepted: 21 July 2020

Introduction

Mucinous adenocarcinoma (MC) is a subtype of colorectal cancer (CRC) that is characterized by the formation of “mucus lakes,” in which the mucinous focus accounts for at least 50% of the tumor. MC accounts for 5% to 15% of total cases of CRC and is associated with proximal locations of the primary tumor.¹ This subtype of CRC tends to be more locally invasive than the other common adenocarcinoma (AC) subtype. Patients with MC have metastatic disease more frequently and have metastases at multiple sites more often than patients with the AC subtype. The common sites of metastasis are also different: MC predominantly metastasizes to the peritoneum, whereas AC more frequently metastasizes to the liver.² Compared with AC, patients with MC generally have a poor response to radiotherapy and chemotherapy and tend to have worse prognoses.^{3–6} At present, there are no specific clinical guidelines for MC patients, and treatment is the standard program used for AC. Because of the lack of confirmed data, the best treatment for MC remains unclear. To further refine the treatment for this subtype, we need an in-depth understanding of its pathogenesis.

Previous studies have suggested that the genetic origin of mucinous colorectal adenocarcinoma is related to the BRAF, microsatellite instability (MSI), and CpG island methylation phenotype (CIMP) pathways.^{7,8} There are differences in the expression of mucin family members in

MC: *MUC2* and *MUC5* are highly expressed, whereas *MUC1* shows low expression. The differential expression of mucin is related to the risk of metastasis and death.⁹ In addition, the expression of p53 and p16 may predict the prognosis of MC patients of CRC.¹⁰ However, the diagnostic or therapeutic value of a single gene remains uncertain. Because of the lack of specific biomarkers and therapeutic targets, the prognosis of MC cannot be effectively improved. Therefore, new signal pathways and molecular targets should be screened to identify potential markers and develop new diagnostic and treatment approaches.^{8–10}

In this study, we used a bioinformatics approach to compare differentially expressed genes (DEGs) between MC and AC based on microarray data to find specific biomarkers and therapeutic targets of MC. Bioinformatics, using extensive data mining at the molecular level, can screen DEGs between MC and AC, construct networks of abnormal gene interactions, and identify key genes. This approach could help clarify the cause of MC and find potential treatment targets. In this study, we analyzed four expression profiling datasets of human samples from the Gene Expression Omnibus (GEO) database and identified the DEGs between MC and AC. Then, we carried out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. A protein–protein interaction (PPI) network analysis was constructed to elucidate the molecular pathogenesis of MC.

Overall, eight hub genes were identified, which could have potential as biomarkers or molecular targets for MC.

Methods

Access to public data

All data used in this study were from publicly accessible databases; therefore, ethical approval was deemed unnecessary.

The Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) is an open functional genomics database of high-throughput resources, including microarrays, gene expression data, and chips. Four expression profiling datasets (GSE101651, GSE103512, GSE101479, and GSE101481) were obtained from GEO that included data from 8 MC and 40 AC cases of colorectal cancer.

DEGs identified by GEO2R

GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) is an interactive online tool used to identify DEGs from GEO series.¹¹ GEO2R can be applied to distinguish DEGs between MC and AC subtypes. Probe sets without corresponding gene symbols or genes with more than one probe set were removed. For the dataset GSE103512, statistical significance was defined at $P \leq 0.05$ and fold change ≥ 1 . For the other datasets, there was no criterion for P -value or fold change. The DEGs were presented in volcano maps drawn using a volcano plotting tool (<https://shengxin.ren>) based on the R language.

Construction of the PPI network

The Search Tool for the Retrieval of Interacting Genes¹² (STRING; <http://string-db.org>), an online database, can trace and predict PPI networks after importing the common DEGs identified in

the GEO datasets. STRING was used to construct the PPI network of DEGs in this study. Cytoscape¹³ (version 2.8; <https://cytoscape.org/>) was used to visualize the PPI network.

Functional annotation of DEGs by KEGG and GO analysis

DAVID (version 6.8; <https://david.ncifcrf.gov/home.jsp>) is an online analysis suite allowing integrated discovery and annotation.¹⁴ Gene Ontology (GO) is widely used in bioinformatics and covers three aspects of biology: biological processes (BP), cellular components (CC), and molecular functions (MF).¹⁵ Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.kegg.jp/>) is one of the most commonly used biological information databases worldwide.¹⁶ To perform GO and KEGG analysis of DEGs, the DAVID online tool was implemented. Significance was defined at $P < 0.05$.

Function and pathway enrichment analysis by Metascape

Metascape (<http://metascape.org/gp/index.html#/main/step1>)¹⁷ is a powerful gene function annotation tool used to apply bioinformatics methods to batch analysis of genes and proteins and thus understand their functions. It can annotate a large number of genes or proteins, perform enrichment analysis, and construct PPI networks. It integrates several authoritative functional databases, such as GO, KEGG, and Uniprot, to analyze not only human data but also that of many other species. It can analyze a single data set or multiple gene sets simultaneously. Metascape was used to conduct the functional and pathway enrichment analysis in the research.

Further analysis for BP and CC Gene Ontology terms

The Biological Networks Gene Oncology tool (BiNGO, version 3.0.3; <http://apps.cytoscape.org/apps/bingo>) was used to analyze and visualize the biological processes and cellular components of identified hub genes.¹⁸

Identification of significant modules and hub genes

The Cytoscape plug-in Molecular Complex Detection (MCODE, version 1.5.1; <http://apps.cytoscape.org/apps/mcode>) identified the most important modules of the network map. The criteria of MCODE analysis were degree cut-off = 2, MCODE score > 5, Max depth = 100, node score cut-off = 0.2, and k-score = 2.¹⁹ The hub genes were identified in the PPI network by using the Cytoscape app cytoHubba (<https://apps.cytoscape.org/apps/cytohubba>). Because a single algorithm sometimes produces false positives, we adopted a four-fold algorithm to jointly identify hub genes and the four algorithms were used in different ways. The algorithms applied to identify hub genes were MCC, DMNC, EPC, and Degree. The common hub genes were analyzed by Venn diagram using FunRich software (<http://www.funrich.org/>).

Screening and analysis of significant hub genes

Significant hub genes were identified by Venn analysis for MCODE and cytoHubba. Then, STRING was used to construct the network of significant hub genes, and DAVID was used to perform BP analysis. Pearson's correlation test was used to assess correlations among the significant hub genes. Heatmaps, presenting correlations among all hub genes, were

prepared using the R language to visualize the expression levels of genes.

Identification of hub genes associated with colorectal cancer

The Comparative Toxicogenomics Database (<http://ctdbase.org/>)²⁰ was searched for integrated chemical–disease, chemical–gene, and gene–disease interactions to predict novel associations and generate expanded networks. The relationships between gene products identified in our datasets and CRC were analyzed using this database.

Quantitative real-time-PCR

Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, MA, Waltham, USA) and reverse-transcribed into cDNA using the Reverse Transcription System (Promega, Fitchburg, WI, USA). Then, the prepared cDNA was subjected to quantitative real-time PCR analysis. The $2^{-\Delta\text{Ct}}$ method was used to analyze relative expression of genes. Three technical replicates per sample were analyzed. The primer sequences are shown in Table 1. Differences between groups were evaluated using the two-tailed Student's *t*-test. All statistical analyses were performed using SPSS version 21.0 software (IBM Corp., Armonk, NY, USA). $P < 0.05$ was considered significant.

Construction and analysis of a target gene-TF-microRNA regulatory network

TFmiR (<http://service.bioinformatik.uni-saarland.de/tfmir/>) can integrate and analyze the interaction among transcription factors (TFs), microRNAs (miRNAs), and target genes, all of which are related to the pathogenesis of disease. The internal operation of cells depends on the correct functioning of an extremely complex activation inhibition system that can be disturbed in

Table 1. Gene primers used in this study.

Gene	Forward (5'→ 3')	Reverse (5'→ 3')
AREG	GCACCTGGAAGCAGTAACATGC	GGCAGCTATGGCTGCTAATGCA
TWIST1	GCCAGGTACATCGACTTCCTCT	TCCATCCTCCAGACCGAGAAGG
ZEB1	GGCATAACCTACTCAACTACGG	TGGGCGGTGTAGAATCAGAGTC
CXCL9	CTGTTCTGCATCAGCACCAAC	TGAACTCCATTCTTCAGTGTAGCA
IDO1	GCCTGATCTCATAGAGTCTGGC	TGCATCCCAGAACTAGACGTGC
MET	TGCACAGTTGGTCCTGCCATGA	CAGCCATAGGACCGTATTTCCGG
SNAI2	ATCTGCGGCAAGGCGTTTTCCA	GAGCCCTCAGATTTGACCTGTC
ZEB2	AATGCACAGAGTGTGGCAAGGC	CTGCTGATGTGCGAACTGTAGG

many ways. TFmiR can help explain cellular mechanisms at the molecular level of the network. TFmiR provides an analysis of topological properties and functions and reveals a disease-specific co-regulatory network between genes, TFs, and miRNAs.

Results

DEGs identified between MC and AC of CRC

Based on analysis of the datasets GSE101651, GSE103512, GSE101479, and GSE101481 with GEO2R, differences between MC and AC samples are presented in volcano plots (Figure 1a, b, c, and d). The complete list of DEGs between MC and AC samples in GEO datasets GSE101651, GSE103512, GSE101479, and GSE101481 is shown in Table 2.

PPI network and DAVID enrichment analysis of DEGs

A PPI network of all DEGs in the four datasets was constructed (Figure 1e). The results of the GO analysis showed that variations in DEGs linked to BP were mainly enriched in maintenance of gastrointestinal epithelium, negative regulation of catenin import into nucleus, Wnt signaling pathway involved in somitogenesis, wound healing, negative regulation of cell adhesion

mediated by integrin, cell differentiation, defense response to virus, negative regulation of canonical Wnt signaling pathway, cellular response to organic cyclic compound, epidermal growth factor receptor signaling pathway, negative regulation of osteoblast differentiation, positive regulation of canonical Wnt signaling pathway, cell proliferation, cell–cell signaling, response to peptide hormone, bicarbonate transport, and negative regulation of cell proliferation (Figure 2a). Variations in DEGs linked to CC were significantly enriched in extracellular space, extracellular region, proteinaceous extracellular matrix, extracellular exosome, basolateral plasma membrane, cell surface, and zymogen granule membrane (Figure 2b). With regard to MF, DEGs were significantly enriched in protein binding, E-box binding, protein homodimerization activity, CXCR3 chemokine receptor binding, and carbonate dehydratase activity (Figure 2c). Analysis of KEGG pathways indicated that the top canonical pathways associated with DEGs were pancreatic secretion, bile secretion, and nitrogen metabolism (Figure 2d).

Metascape enrichment analysis of DEGs

Functional enrichment analysis with Metascape showed that DEGs between MC and AC were significantly enriched in negative regulation of cell proliferation,

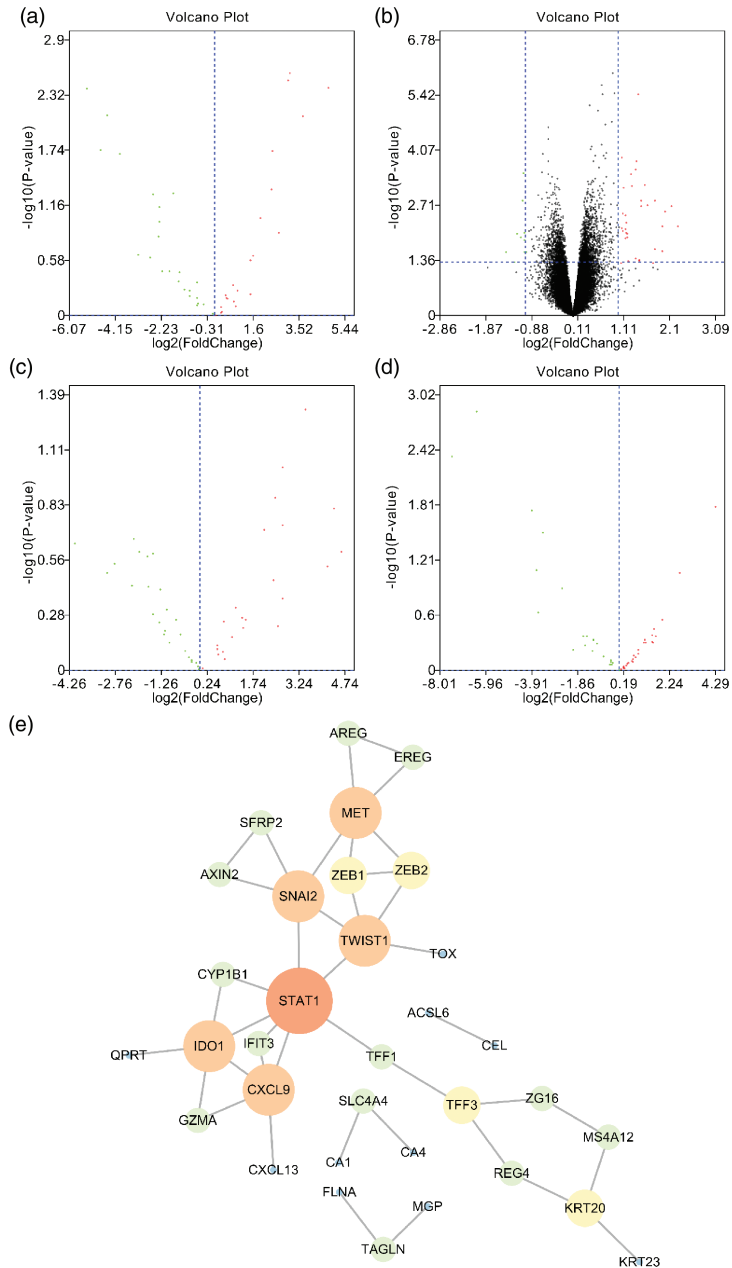


Figure I. Volcano plots of DEGs and the PPI network. The DEGs between mucinous adenocarcinoma and other adenocarcinoma samples in Gene Expression Omnibus datasets (a) GSE101651, (b) GSE103512, (c) GSE101479, and (d) GSE101481. (e) PPI network of all DEGs.

DEG, differentially expressed gene; PPI, protein–protein interaction.

Table 2. Complete list of differentially expressed genes between MC and AC samples in Gene Expression Omnibus datasets GSE101651, GSE103512, GSE101479, and GSE101481.

	Gene symbol
Differentially expressed genes	LY6G6D, GZMA, SLC4A4, CXCL13, KRT23, AXIN2, CXCL9, AREG, STAT1, PCSK1, IFIT3, KRT20, PLEKHB1, MSRB3, MUC2, LINC00261, TWIST1, IDO1, ACSL6, ZG16, ZEB2, CA1, TFF3, RARRES3, EREG, MGP, MS4A12, ZEB1, QPRT, CEL, TFF1, BIRC3, REG4, TOX, CA4, FLNA, CLDN8, SNAI2, CYP1B1, TCN1, TAGLN, COL10A1, SFRP2, MET, SFRP4, CLCA4, AQP8, CFTR, BHLHE41, SPINK4, TACSTD2, REG4, IGF2, AGR3, CLDN2, CYP2B6, MUC2, HSPB6, SERPINA1, LGR5, HMGCS2, FCGBP, DMBT1, LYZ, PIGR, KRT23, GZMA, SLC4A4, CXCL13, STAT1, BIRC3, CXCL9, TFF3, MGP, LY6G6D, SFRP2, RARRES3, PLEKHB1, CA1, SFRP4, MUC2, ZG16, LINC00261, TCN1, MS4A12, TAGLN, COL10A1, IFIT3, CA4, TWIST1, TOX, CLDN8, QPRT, FLNA, AQP8, AREG, IDO1, EREG, MET, KRT20, MSRB3, CLCA4, BHLHE41, CFTR, PCSK1, REG4, CEL, TFF1, AXIN2, CYP1B1, ACSL6, SNAI2, ZEB1, SPINK4, ZEB2, ZG16, MS4A12, CA1, QPRT, CA4, CLCA4, SPINK4, CEL, KRT20, KRT23, IFIT3, GZMA, CXCL13, MGP, PLEKHB1, CXCL9, SLC4A4, EREG, RARRES3, TFF3, MUC2, AQP8, STAT1, PCSK1, REG4, FLNA, TCN1, AREG, CFTR, SNAI2, AXIN2, LY6G6D, LINC00261, MET, COL10A1, ACSL6, MSRB3, CYP1B1, BIRC3, TOX, SFRP2, IDO1, BHLHE41, ZEB2, TAGLN, TWIST1, TFF1, SFRP4, CLDN8, ZEB1

MC, mucinous carcinoma; AC, other adenocarcinoma.

epithelial cell differentiation, maintenance of gastrointestinal epithelium, and regulation of epithelial cell differentiation ($P < 0.05$; Figure 3a, b, and c).

Further analysis for BP and CC using BiNGO

The BP and CC analysis of DEGs by BiNGO verified the results of the GO analysis (Figure 4a, b). The results of BiNGO analysis showed that variations in DEGs linked to BP were mainly enriched in negative regulation of such activities as cell differentiation, fluid transport, and water transport (Figure 4a). Changes in CC could also be enriched in extracellular region, extracellular space, extracellular region part, mucus layer, inner mucus layer, outer mucus layer, apical lamina of hyaline layer, hyaline layer, apical plasma membrane, cytoplasmic vesicle, vesicle, membrane-bound vesicle, stored secretory granule, zymogen granule, cytoplasmic membrane-bound vesicle, zymogen granule

membrane, and cytoplasmic vesicle (Figure 4b).

Significant modules

The three most significant modules were obtained from the PPI network using Cytoscape. The first module consisted of *EREG*, *AREG*, *MET*, *ZEB1*, *ZEB2*, *TWIST1*, *SNAI2*, *SFRP2*, and *AXIN2* (Figure 5a), the second consisted of *ZG16*, *TFF3*, *MS4A12*, *KRT20*, and *REG4* (Figure 5b), and the third consisted of *GZMA*, *CXCL9*, and *IDO1* (Figure 5c).

Hub genes identified from the PPI network

The MCC, DMNC, EPC, and Degree algorithms were used to identify hub genes shown in Figure 6a, b, c, and d, respectively. The common hub genes (*CXCL9*, *MET*, *AREG*, *ZEB1*, *ZEB2*, *TWIST1*, *STAT1*, *SNAI2*, *IDO1*, *IFIT3*) were identified by Venn diagram (Figure 6e).

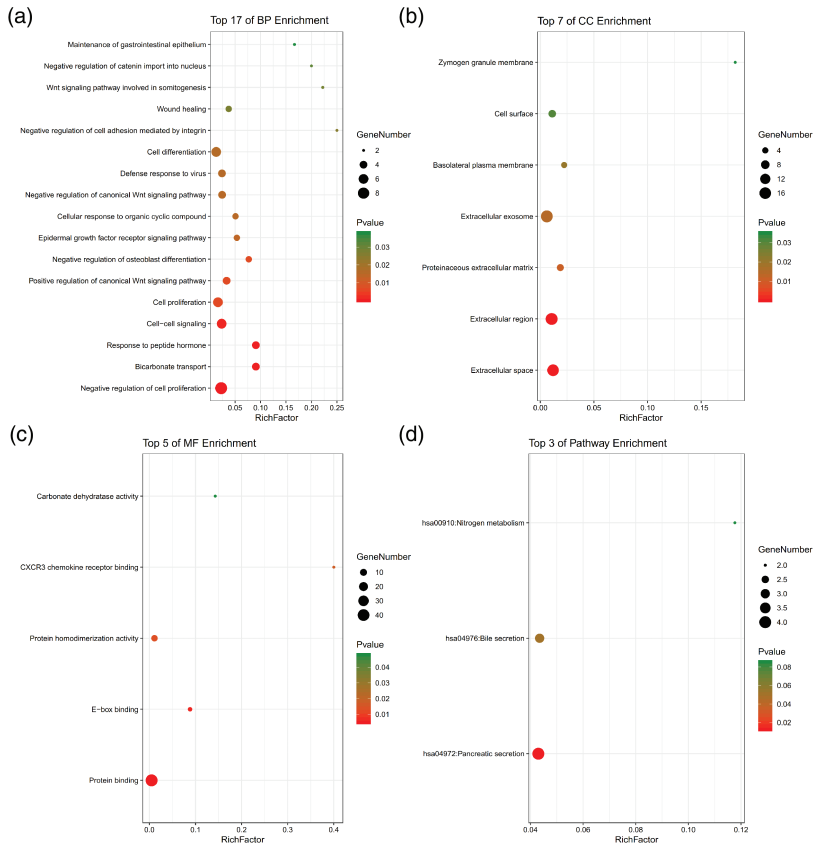


Figure 2. The enrichment analysis of DEGs by DAVID (<https://david.ncicrf.gov/home.jsp>). Detailed information relating to changes in Gene Ontology categories (a) BP, (b) CC, and (c) MF; (d) KEGG pathway analysis for hub genes.

DEG, differentially expressed gene; BP, biological processes; CC, cellular component; MF, molecular functions; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Screening and analysis of significant hub genes

Through the Venn diagram, eight significant hub genes (*CXCL9*, *MET*, *AREG*, *ZEB1*, *ZEB2*, *TWIST1*, *SNAI2*, and *IDO1*) were identified between MCODE and cytoHubba (Figure 7a). A summary of the functions of the eight hub genes is given in Table 3, and the network of significant hub genes is displayed in Figure 7b. Through DAVID analysis, the GO analysis showed that variations in significant hub genes linked to BP were mainly enriched

in such activities as cell proliferation, negative regulation of transcription, negative regulation of gene expression, cell migration, and cell motility (Figure 7c). The correlations among all significant hub genes are presented in Figure 7d.

Identification of hub genes

Analysis by heatmap allowed for simple differentiation of MC from AC tissues CRC based on the expression levels of hub genes. Expression of *AREG*, *TWIST1*, and *ZEB1* was downregulated in MC tissues compared

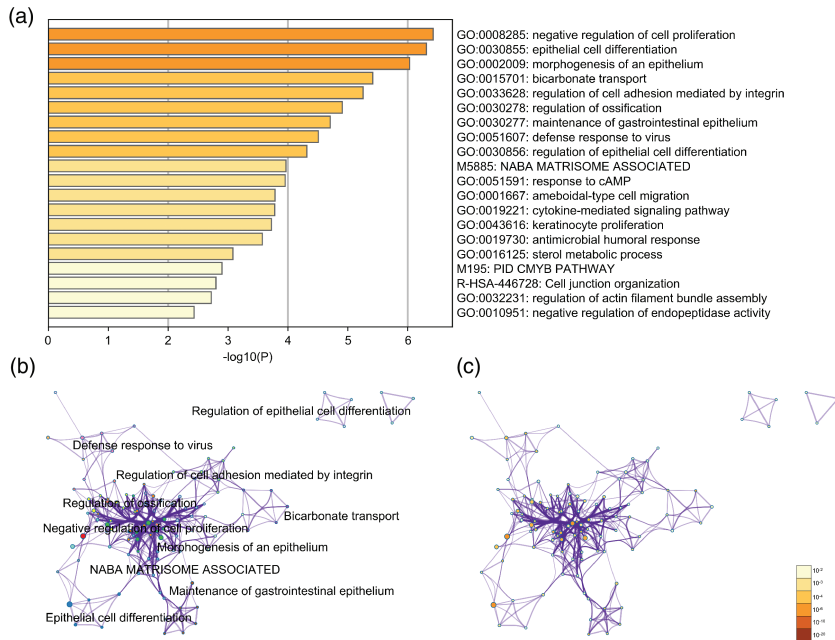


Figure 3. Enrichment analysis of DEGs by Metascape (<http://metascape.org/gp/index.html#/main/step1>). (a) Heatmap of enriched terms across lists of input differentially expressed genes, colored by P -values, via Metascape. (b) Network of enriched terms colored by cluster identity, where nodes that share the same cluster identity are typically close to each other. (c) Network of enriched terms, colored by P -values, where terms containing more genes tend to have more significant P -values. DEG, differentially expressed gene.

with AC tissues. However, expression of the other five significant hub genes was higher in MC (Figure 8). The Comparative Toxicogenomics Database showed that all 8 hub genes were associated with CRC (Figure 9). These results were verified by qRT-PCR in 50 AC tissues and 20 MC tissues (Figure 10). The results showed that compared with AC, expression of *AREG* was significantly downregulated, and expression of *SNAI2*, *IDO1*, and *MET* were significantly upregulated in MC, consistent with bioinformatics results. However, no significant differences were found for the other genes, which may be related to the small sample size. To better understand the mechanism of the *MET* gene, we constructed and analyzed a target genes–TF regulatory network and

target gene–miRNA regulatory network (Figure 11).

Discussion

MC originates more often in the right (proximal) colon and is more common in women.⁶ It is more frequently associated with hereditary non-polyposis colorectal cancer and young-age sporadic CRC.⁹ A meta-analysis of 44 studies and 222,256 patients showed that the prevalence of stage IV disease was similar in patients with AC and MC.⁶

In patients with advanced disease, patients with MC have shown a worse response to fluorouracil and oxaliplatin/irinotecan-based first-line chemotherapy than patients with AC.^{21,22} MC patients

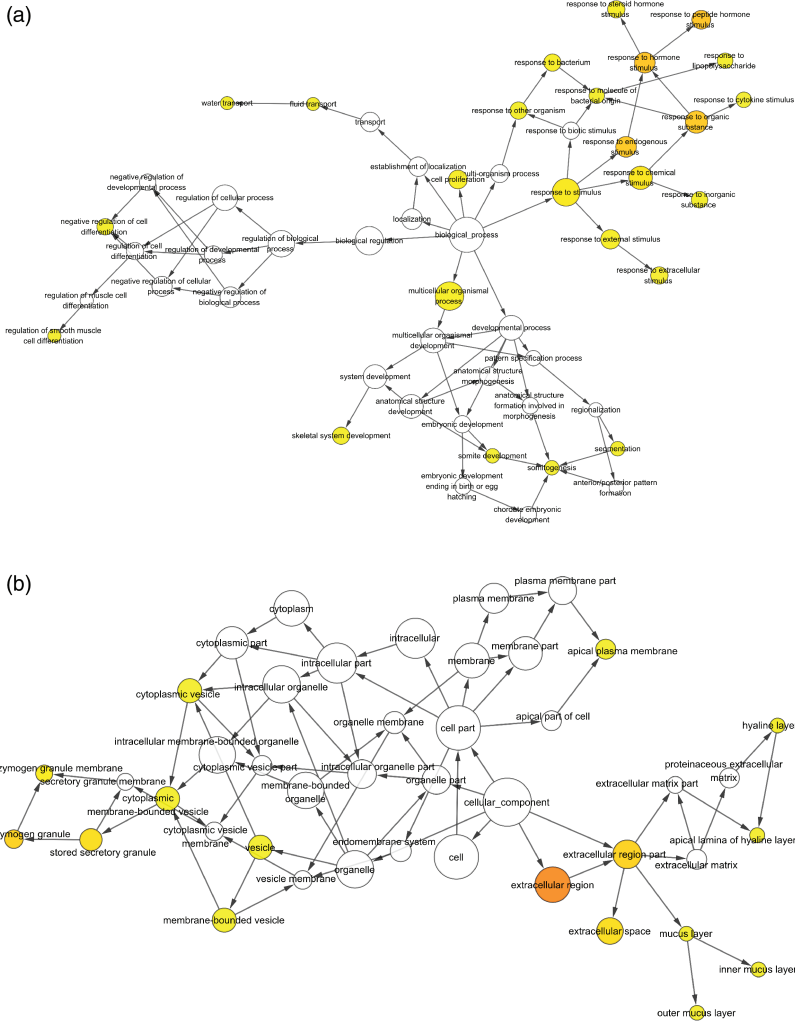


Figure 4. BiNGO (<http://apps.cytoscape.org/apps/bingo>) analysis for all DEGs. (a) BP enrichment for all DEGs; (b) CC enrichment for all DEGs. BiNGO, Biological Networks Gene Oncology tool; DEG, differentially expressed gene; BP, biological processes; CC, cellular component.

tend to have worse overall survival (OS) and quality of life, more metastatic sites, and more peritoneal metastases. For patients with stage II CRC, although American Society of Clinical Oncology (ASCO) and European Society for Medical Oncology (ESMO) guidelines define the prognostic factors to guide adjuvant chemotherapy, including inadequate

sampling of lymph nodes, pT4 primary tumors, obstruction, perforation, lymphovascular and perineural invasion, and poorly differentiated tumors, the effect of histological subtypes remains unclear. A study of 4893 patients with stage II CRC showed that MC patients tended to have a higher T category, a greater percentage of lymph nodes harvested, larger tumor size,

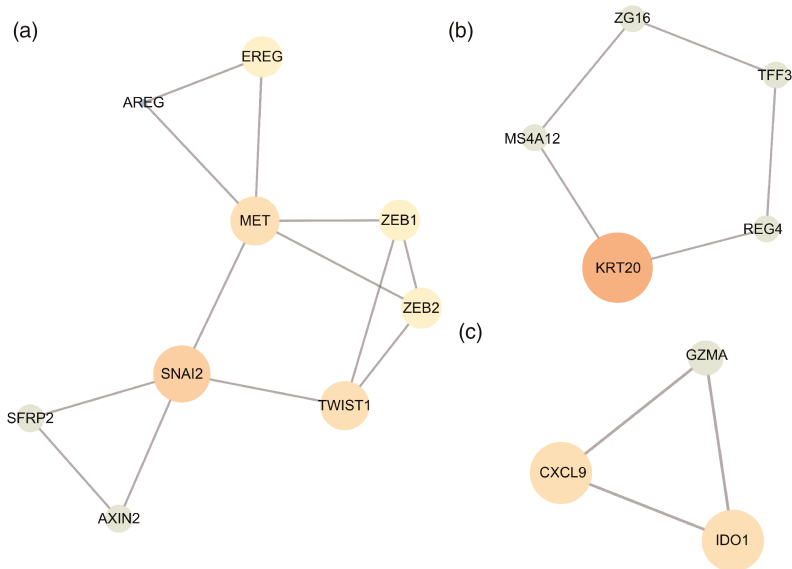


Figure 5. Three significant modules identified from the PPI network. (a) The first module consists of EREG, AREG, MET, ZEB1, ZEB2, TWIST1, SNAI2, SFRP2, and AXIN2. (b) The second module consists of ZG16, TFF3, MS4A12, KRT20, and REG4. (c) The third module consists of GZMA, CXCL9, and IDO1. PPI, protein–protein interaction.

and worse grading. For lymph node-negative CRC patients, the tumor histology of MC is related to poor prognosis. MC is a high-risk factor for stage II CRC patients with good or moderate differentiation and negative vascular or perineural invasion, indicating poor prognosis.²³ This finding is consistent with that of Lee et al.²⁴ For treatment of CRC patients with MC stage II, grade 3, adjuvant chemotherapy with oxaliplatin and fluoropyrimidine for 6 months instead of 3 months is needed because of the worse relapse-free survival and OS in the subgroup of patients with MC.²⁵

Despite the use of standard neoadjuvant and adjuvant regimens, MC remains a negative prognostic factor in young patients with CRC. It is associated with high and early recurrence rates.²⁶ Because of the early recurrence behavior of MC, physicians need to fully evaluate treatment and prognosis of CRC patients, and explore new treatment options.

In the past few decades, bioinformatics technology has been widely used to screen potential genetic targets of diseases; it is a useful approach to identify DEGs and related pathways and it has greatly improved the efficiency of scientific research. In this study, we used previous tissue samples from patients with MC and AC to identify DEGs in the two subtypes. The qRT-PCR results revealed no significant differences in expression of *TWIST1*, *ZEB1*, *CXCL9*, and *ZEB2* between patients with MC and AC, which may be related to the small sample size. However, our results provide a direction for further study of the pathogenesis of MC and thus to optimize individualized treatment of MC.

Amphiregulin (AREG) is one ligand of the epidermal growth factor receptor (EGFR). It is expressed in multiple human tumors, such as liver cancer, prostate cancer, gastric cancer, esophageal cancer, breast cancer, and myeloma. AREG is

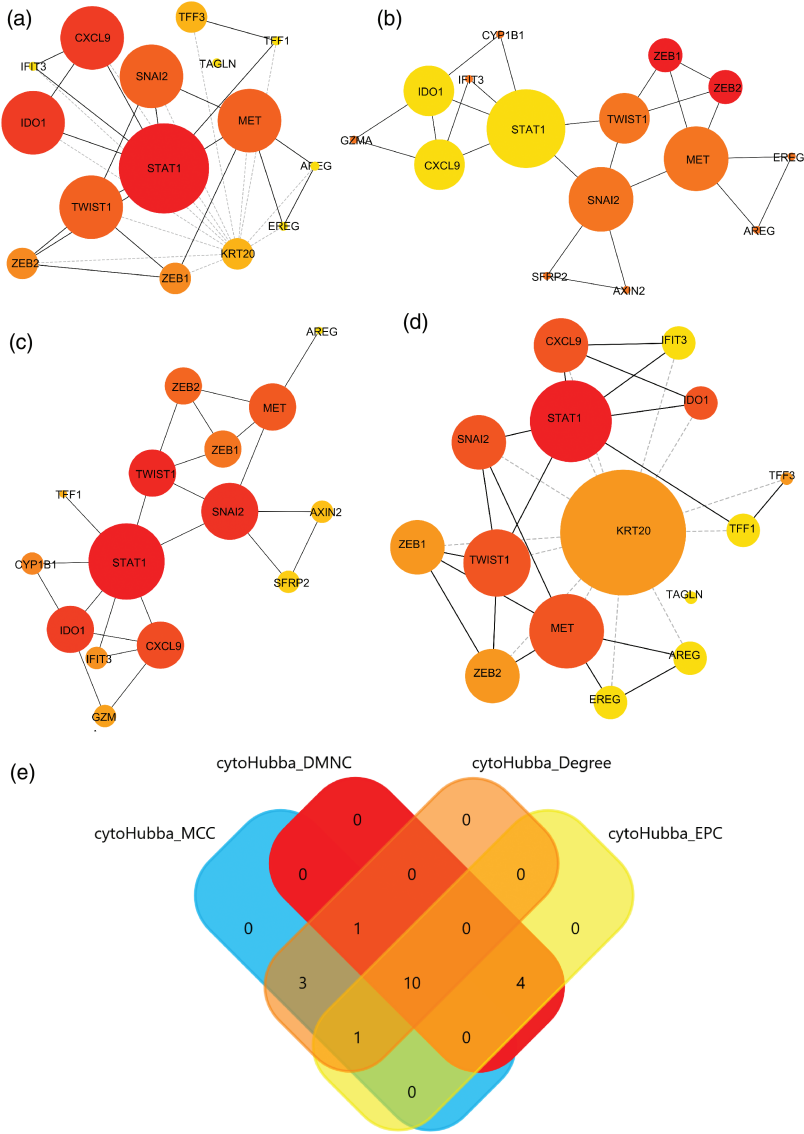


Figure 6. Hub gene networks identified from the protein–protein interaction network using (a) the MCC algorithm; (b) the DMNC algorithm; (b) the EPC algorithm; and (d) the Degree algorithm of the Cytoscape app cytoHubba (<https://apps.cytoscape.org/apps/cytohubba>). (e) The Venn diagram shows the common hub genes (CXCL9, MET, AREG, ZEB1, ZEB2, TWIST1, STAT1, SNAI2, IDO1, IFIT3).

expressed in stromal and epithelial compartments in CRC tissues and might be related to the development and progression of CRC. Researchers have demonstrated the coordinate expression of AREG,

EGFR activity, and cell proliferation markers *in vivo*.²⁷ In addition, AREG may stimulate the growth of tumor cells through autocrine and paracrine pathways *in vitro*. AREG in the extracellular matrix

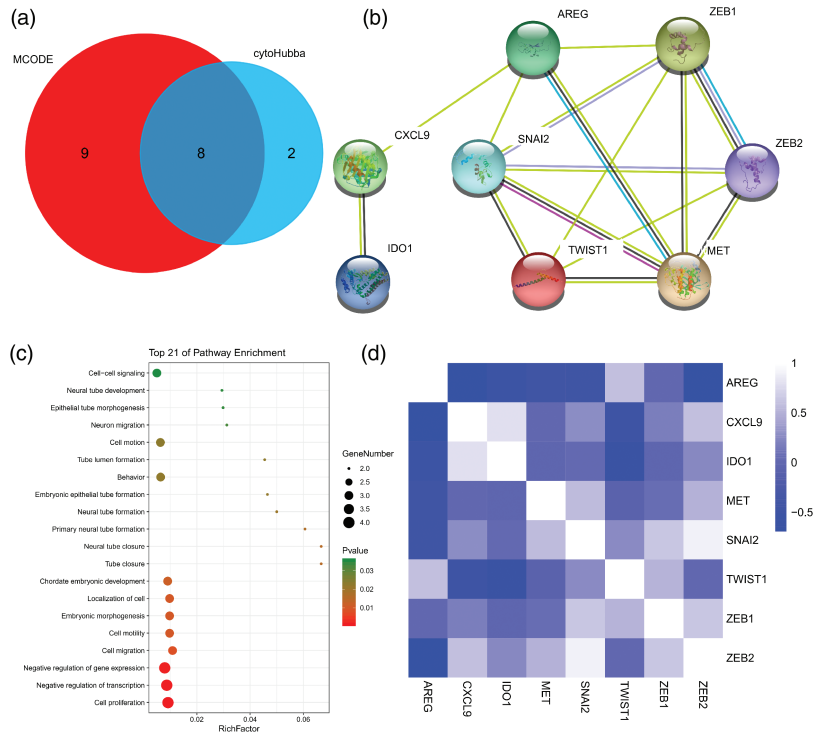


Figure 7. Screening and analysis of significant hub genes. (a) The Venn diagram shows eight significant hub genes (*CXCL9*, *MET*, *AREG*, *ZEB1*, *ZEB2*, *TWIST1*, *SNAI2*, *IDO1*) identified by MCODE and cytoHubba. (b) The network of significant hub genes. (c) The BP enrichment analysis of the significant hub genes by DAVID (<https://david.ncicrf.gov/home.jsp>). (d) The heatmap shows the correlations among all significant hub genes. BP, biological processes.

can stimulate the growth of CRC cells. It has been suggested that *AREG* plays pro-neoplastic roles in CRC.²⁷ The expression of *AREG* is negatively correlated with the *BRAF* mutation and high expression of *CIMP*. In addition, the mRNA expression level of *AREG* was higher in left-sided tumors than that in right-sided tumors.^{28,29} This may further explain the greater response of left-sided tumors to EGFR antibody. Expression of *AREG* is associated with the occurrence of liver metastasis in CRC patients as an important predictive marker.^{30,31} MC predominantly metastasizes to peritoneal sites, whereas patients with AC more frequently have liver metastases, which may be related to

differences in *AREG* expression. Researchers observed a significant downregulation of *AREG* expression in 5-fluorouracil-resistant CRC cells.³² In this study, we found that *AREG* was expressed at a lower level in MC than in AC. This may help explain why MC patients always have a worse response to chemotherapy. It has been shown that high expression of *AREG* in CRC tissues is related to longer OS and progression-free survival, whereas low expression of *AREG* indicates poor survival.^{33–35} However, other researchers have found that high serum expression of *AREG* is associated with poor pathological factors.³⁶ Overall, we believe that low expression of *AREG* affects metastasis,

Table 3. Summary of the functions of the eight hub genes (function descriptions from <https://pmlegacy.ncbi.nlm.nih.gov/>).

	Gene symbol	Full name	Function
1	<i>CXCL9</i>	C-X-C motif chemokine ligand 9	This antimicrobial gene encodes a protein thought to be involved in T cell trafficking. The encoded protein binds to C-X-C motif chemokine 3 and is a chemoattractant for lymphocytes but not for neutrophils.
2	<i>MET</i>	MET proto-oncogene, receptor tyrosine kinase	This gene encodes a member of the receptor tyrosine kinase family of proteins and the product of the proto-oncogene MET. Mutations in this gene are associated with papillary renal cell carcinoma, hepatocellular carcinoma, and various head and neck cancers. Amplification and overexpression of this gene are also associated with multiple human cancers.
3	<i>AREG</i>	Amphiregulin	The protein encoded by this gene is a member of the epidermal growth factor family. It is related to epidermal growth factor (EGF) and transforming growth factor alpha (TGF-alpha). The protein interacts with the EGF/TGF-alpha receptor to promote the growth of normal epithelial cells, and it inhibits the growth of certain aggressive carcinoma cell lines.
4	<i>ZEB1</i>	Zinc finger E-box binding homeobox 1	This gene encodes a zinc finger transcription factor. The encoded protein likely plays a role in transcriptional repression of interleukin 2. Mutations in this gene have been associated with posterior polymorphous corneal dystrophy-3 and late-onset Fuchs endothelial corneal dystrophy. Alternatively spliced transcript variants encoding different isoforms have been described.
5	<i>ZEB2</i>	Zinc finger E-box binding homeobox 2	The protein encoded by this gene is a member of the Zfh1 family of 2-handed zinc finger/homeodomain proteins. It is located in the nucleus and functions as a DNA-binding transcriptional repressor that interacts with activated SMADs. Mutations in this gene are associated with Hirschsprung disease/Mowat-Wilson syndrome. Alternatively spliced transcript variants have been found for this gene.
6	<i>TWIST1</i>	Twist family bHLH transcription factor 1	This gene encodes a basic helix-loop-helix (bHLH) transcription factor that plays an important role in embryonic development. The encoded protein forms both homodimers and heterodimers that bind to DNA E box sequences and regulate the transcription of genes involved in cranial suture closure during skull development. This protein may also regulate neural tube closure, limb development and brown fat metabolism. This gene is hypermethylated and overexpressed in multiple human cancers, and the encoded protein promotes tumor cell invasion and metastasis.

(continued)

Table 3. Continued.

	Gene symbol	Full name	Function
7	<i>SNAI2</i>	Snail family transcriptional repressor 2	This gene encodes a member of the Snail family of C2H2-type zinc finger transcription factors. The encoded protein acts as a transcriptional repressor that binds to E-box motifs and is also likely to repress E-cadherin transcription in breast carcinoma. This protein is involved in epithelial-mesenchymal transitions and has antiapoptotic activity.
8	<i>IDO1</i>	Indoleamine 2,3-dioxygenase 1	This gene encodes indoleamine 2,3-dioxygenase (IDO). This enzyme is thought to play a role in a variety of pathophysiological processes such as antimicrobial and antitumor defense, neuropathology, immunoregulation, and antioxidant activity.

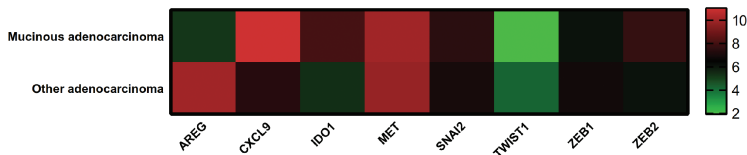


Figure 8. Eight hub genes could differentiate mucinous adenocarcinoma samples from other adenocarcinoma samples. The upregulated genes in MC were *CXCL9*, *IDO1*, *MET*, *SNAI2*, and *ZEB2*, and the downregulated genes were *AREG*, *TWIST1*, and *ZEB1*.

therapeutic effect, and prognosis of MC, but its function and mechanism need further exploration.

In our study, *AREG* was negatively correlated with the mesenchymal–epithelial transition factor (*MET*), a cell-surface receptor tyrosine kinase (RTK) for hepatocyte growth factor (HGF)/scatter factor (SF). *MET* is an important signal in the development of CRC, and its high expression predicts a poor prognosis.^{37,38} A retrospective study of 286 patients with CRC showed that *MET* expression was significantly higher in advanced CRC than in early stages.³⁹ *MET* interacts with other growth factor receptors in CRC, and these interactions or cross-talk could lead to tumor progression and play a key role in drug resistance of targeted therapy.

MET/HGF signaling plays an important role in solid tumors, especially in CRC. Researchers have found that *MET*/HGF signal, together with its downstream PI3K/AKT, MAPK, and STAT3 pathways, is a regulator of vascular endothelial growth factor (VEGF) expression.⁴⁰ Hypoxia-induced c-Met expression by VEGF pathway inhibitors leads to resistance to VEGF inhibition. The *MET* and *RON* tyrosine kinases belong to the scatter factor receptor family of TM RTKs, sharing many similar functional domains, and they are often co-expressed. A study showed that increased immunoreactivity of *RON* or *MET* was associated with shorter survival time and that moderate to strong co-expression of both receptors was associated with a worse prognosis.⁴¹ *MET* could

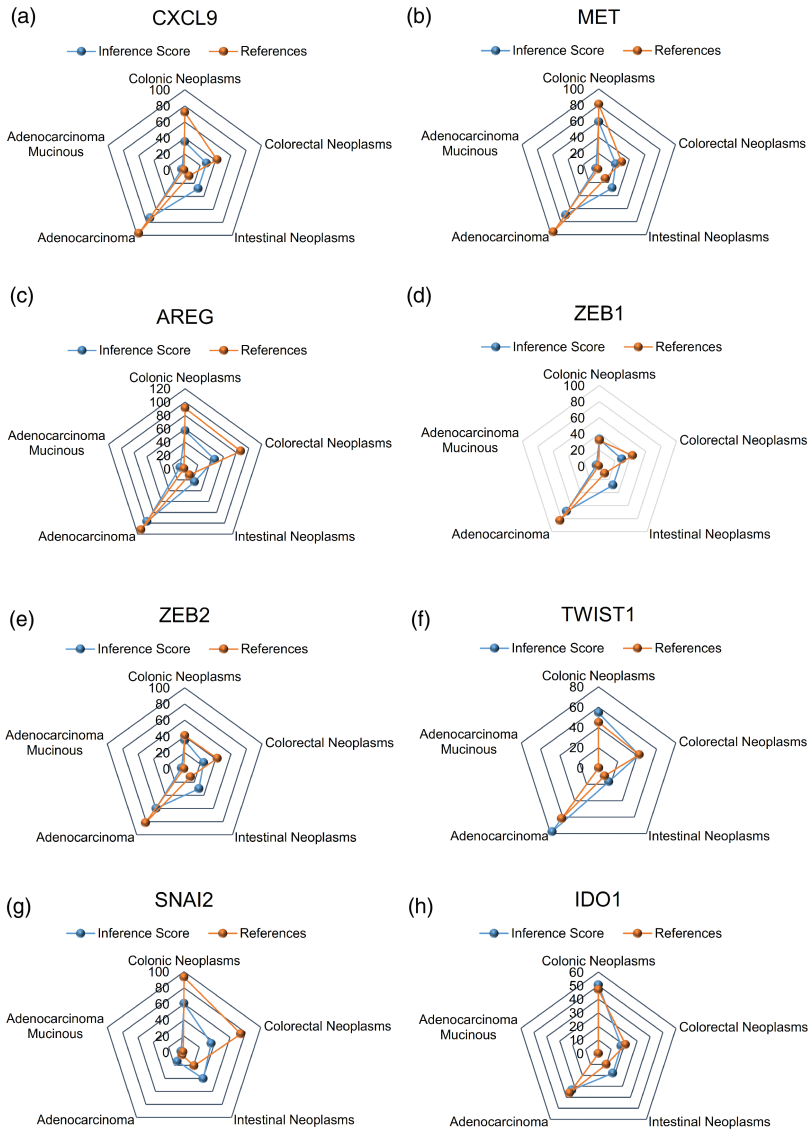


Figure 9. Relationships of colorectal cancer with significant hub genes based on the comparative toxicogenomics database (<http://ctdbase.org/>). (a) *CXCL9*, (b) *MET*, (c) *AREG*, (d) *ZEB1*, (e) *ZEB2*, (f) *TWIST1*, (g) *SNAI2*, and (h) *IDO1*.

compensate for the loss of RON signal.⁴² c-Met and RON signaling pathways rely mainly on the activation of PI3K and MAPK, which leads to the activation of PI3K/Akt, SRC, STAT3, nuclear factor- κ B, focal adhesion kinase (FAK), and

β -catenin, leading to cell responses, including epithelial-mesenchymal transition, cytoskeleton changes, invasion, angiogenesis, proliferation, resistance to apoptosis, among others.⁴³ The crosstalk of RET/c-Met and HER2/c-Met play important

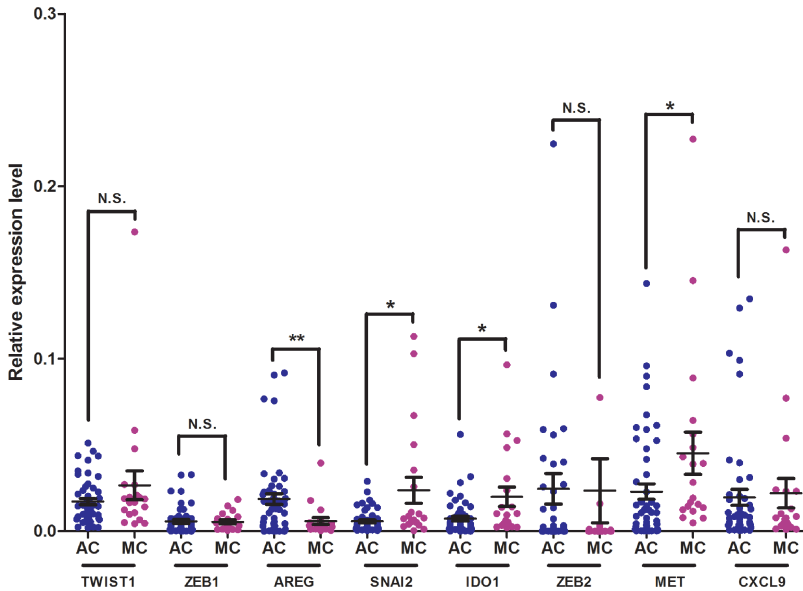


Figure 10. The eight significant hub genes (*TWIST1*, *ZEB1*, *AREG*, *SNAI2*, *IDO1*, *ZEB2*, *MET*, *CXCL9*) were validated by qRT-PCR. * $P < 0.05$, ** $P < 0.01$, N.S., nonsignificant. qRT-PCR, quantitative real-time PCR; MC, mucinous carcinoma; AC, other adenocarcinoma.

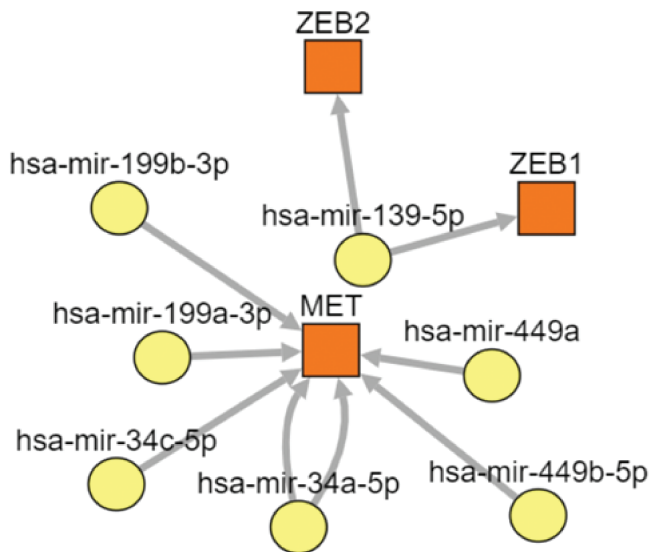


Figure 11. The target gene–transcription factor–microRNA regulatory network.

roles in proliferation, survival, and cell migration. In addition, the heterodimerization of EGFR and HER3 with c-Met could lead to cell proliferation.⁴⁴ In addition,

HGF/c-Met signals have been shown to interact with insulin-like growth factor and its receptor (IGF1/IGF1R)⁴⁵ and MACC1 (MET transcriptional

regulator).^{46,47} In this study, we found that *MET* was highly expressed in MC, which tends to be have a more malignant biological behavior.

The results of this study were based on bioinformatics analysis, without a large number of clinical samples for validation. In the next stage, we will conduct an in vivo or clinical study to verify the conclusions of this study.

In conclusion, the purpose of this study was to determine characteristic gene expression in the MC subtype of CRC. We compared DEGs between MC and AC using microarray data and provided new insights into specific biomarkers and potential therapeutic targets. Data mining and integration are promising tools to predict biomarkers of malignant tumors. However, our results are preliminary and would benefit from further in-depth study.


Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID iD

Jing Zuo  <https://orcid.org/0000-0003-0298-3760>

References

- Lupinacci RM, Mello ES, Coelho FF, et al. Prognostic implication of mucinous histology in resected colorectal cancer liver metastases. *Surgery* 2014; 155: 1062–1068.
- Hugen N, Van De Velde CJ, De Wilt JH, et al. Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. *Ann Oncol* 2014; 25: 651–657.
- Du W, Mah JT, Lee J, et al. Incidence and survival of mucinous adenocarcinoma of the colorectum: a population-based study from an Asian country. *Dis Colon Rectum* 2004; 47: 78–85.
- Glasgow SC, Yu J, Carvalho LP, et al. Unfavourable expression of pharmacologic markers in mucinous colorectal cancer. *Br J Cancer* 2005; 92: 259–264.
- Ooki A, Akagi K, Yatsuoka T, et al. Inverse effect of mucinous component on survival in stage III colorectal cancer. *J Surg Oncol* 2014; 110: 851–857.
- Verhulst J, Ferdinandse L, Demetter P, et al. Mucinous subtype as prognostic factor in colorectal cancer: a systematic review and meta-analysis. *J Clin Pathol* 2012; 65: 381–388.
- Reynolds IS, O'Connell E, Fichtner M, et al. Mucinous adenocarcinoma of the colon and rectum: A genomic analysis. *J Surg Oncol* 2019; 120: 1427–1435.
- Reynolds IS, Furney SJ, Kay EW, et al. Meta-analysis of the molecular associations of mucinous colorectal cancer. *Br J Surg* 2019; 106: 682–691.
- Debunne H and Ceelen W. Mucinous differentiation in colorectal cancer: molecular, histological and clinical aspects. *ActaChirBelg* 2013; 113: 385–390.
- King-Yin Lam A, Ong K and Ho YH. Colorectal mucinous adenocarcinoma: the clinicopathologic features and significance of p16 and p53 expression. *Dis Colon Rectum* 2006; 49: 1275–1283.
- Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013; 41: D991–D995.
- Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; 43: D447–D452.
- Smoot ME, Ono K, Ruscheinski J, et al. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011; 27: 431–432.
- Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007; 8: R183.

15. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29.
16. Kanehisa M and Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000; 28: 27–30.
17. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019; 10: 1523.
18. Maere S, Heymans K and Kuiper M. BiNGO: a Cytoscape plugin to assess over-representation of gene ontology categories in biological networks. *Bioinformatics* 2005; 21: 3448–3449.
19. Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003; 4: 2.
20. Davis AP, Grondin CJ, Johnson RJ, et al. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res* 2017; 45: D972–D978.
21. Catalano V, Loupakis F, Graziano F, et al. Mucinous histology predicts for poor response rate and overall survival of patients with colorectal cancer and treated with first-line oxaliplatin- and/or irinotecan-based chemotherapy. *Br J Cancer* 2009; 100: 881–887.
22. Negri FV, Wotherspoon A, Cunningham D, et al. Mucinous histology predicts for reduced fluorouracil responsiveness and survival in advanced colorectal cancer. *Ann Oncol* 2005; 16: 1305–1310.
23. Hu X, Li YQ, Li QG, et al. Mucinous adenocarcinomas histotype can also be a high-risk factor for stage II colorectal cancer patients. *Cell Physiol Biochem* 2018; 47: 630–640.
24. Lee DW, Han SW, Lee HJ, et al. Prognostic implication of mucinous histology in colorectal cancer patients treated with adjuvant FOLFOX chemotherapy. *Br J Cancer* 2013; 108: 1978–1984.
25. Rosati G, Galli F, Cantore M, et al. Predictive impact of mucinous tumors on the clinical outcome in patients with poorly differentiated, stage II colon cancer: A TOSCA Subgroup Analysis. *Oncologist* 2020; 25: e928–e935.
26. Soliman BG, Karagkounis G, Church JM, et al. Mucinous histology signifies poor oncologic outcome in young patients with colorectal cancer. *Dis Colon Rectum* 2018; 61: 547–553.
27. Guzman MJ, Shao J and Sheng H. Pro-neoplastic effects of amphiregulin in colorectal carcinogenesis. *J Gastrointest Cancer* 2013; 44: 211–221.
28. Kuramochi H, Nakajima GO, Hayashi K, et al. Amphiregulin/epiregulin mRNA expression and primary tumor location in colorectal cancer. *Anticancer Res* 2019; 39: 4729–4736.
29. Lee MS, McGuffey EJ, Morris JS, et al. Association of CpG island methylator phenotype and EREG/AREG methylation and expression in colorectal cancer. *Br J Cancer* 2016; 114: 1352–1361.
30. Watanabe T, Kobunai T, Yamamoto Y, et al. Prediction of liver metastasis after colorectal cancer using reverse transcription-polymerase chain reaction analysis of 10 genes. *Eur J Cancer* 2010; 46: 2119–2126.
31. Yamada M, Ichikawa Y, Yamagishi S, et al. Amphiregulin is a promising prognostic marker for liver metastases of colorectal cancer. *Clin Cancer Res* 2008; 14: 2351–2356.
32. De Angelis PM, Fjell B, Kravik KL, et al. Molecular characterizations of derivatives of HCT116 colorectal cancer cells that are resistant to the chemotherapeutic agent 5-fluorouracil. *Int J Oncol* 2004; 24: 1279–1288.
33. Jing C, Jin YH, You Z, et al. Prognostic value of amphiregulin and epiregulin mRNA expression in metastatic colorectal cancer patients. *Oncotarget* 2016; 7: 55890–55899.
34. Stahler A, Heinemann V, Giessen-Jung C, et al. Influence of mRNA expression of epiregulin and amphiregulin on outcome of patients with metastatic colorectal cancer treated with 5-FU/LV plus irinotecan or irinotecan plus oxaliplatin as first-line treatment (FIRE 1-trial). *Int J Cancer* 2016; 138: 739–746.

35. Sunakawa Y, Yang D, Moran M, et al. Combined assessment of EGFR-related molecules to predict outcome of 1st-line cetuximab-containing chemotherapy for metastatic colorectal cancer. *Cancer Biol Ther* 2016; 17: 751–759.
36. Chayangsu C, Khunsri S, Sriuranpong V, et al. The correlations between serum amphiregulin and other clinicopathological factors in colorectal cancer. *J GastrointestOncol* 2017; 8: 980–984.
37. Gao H, Guan M, Sun Z, et al. High c-Met expression is a negative prognostic marker for colorectal cancer: a meta-analysis. *TumourBiol* 2015; 36: 515–520.
38. Liu Y, Yu XF, Zou J, et al. Prognostic value of c-Met in colorectal cancer: a meta-analysis. *World J Gastroenterol* 2015; 21: 3706–3710.
39. De Oliveira AT, Matos D, Logullo AF, et al. MET is highly expressed in advanced stages of colorectal cancer and indicates worse prognosis and mortality. *Anticancer Res* 2009; 29: 4807–4811.
40. Matsumura A, Kubota T, Taiyoh H, et al. HGF regulates VEGF expression via the c-Met receptor downstream pathways, PI3K/Akt, MAPK and STAT3, in CT26 murine cells. *Int J Oncol* 2013; 42: 535–542.
41. Lee CT, Chow NH, Su PF, et al. The prognostic significance of RON and MET receptor coexpression in patients with colorectal cancer. *Dis Colon Rectum* 2008; 51: 1268–1274.
42. Zhao S, Cao L and Freeman JW. Knockdown of RON receptor kinase delays but does not prevent tumor progression while enhancing HGF/MET signaling in pancreatic cancer cell lines. *Oncogenesis* 2013; 2: e76.
43. Chang K, Karnad A, Zhao S, et al. Roles of c-Met and RON kinases in tumor progression and their potential as therapeutic targets. *Oncotarget* 2015; 6: 3507–3518.
44. Tanizaki J, Okamoto I, Sakai K, et al. Differential roles of trans-phosphorylated EGFR, HER2, HER3, and RET as heterodimerisation partners of MET in lung cancer with MET amplification. *Br J Cancer* 2011; 105: 807–813.
45. Bauer TW, Fan F, Liu W, et al. Insulin-like growth factor-I-mediated migration and invasion of human colon carcinoma cells requires activation of c-Met and urokinase plasminogen activator receptor. *Ann Surg* 2005; 241: 748–756; discussion 756-758.
46. Arlt F and Stein U. Colon cancer metastasis: MACC1 and Met as metastatic pacemakers. *Int J Biochem Cell Biol* 2009; 41: 2356–2359.
47. Ren B, Zakharov V, Yang Q, et al. MACC1 is related to colorectal cancer initiation and early-stage invasive growth. *Am J ClinPathol* 2013; 140: 701–707.