Original Research

Identification of IMPA2 as the hub gene associated with colorectal cancer and liver metastasis by integrated bioinformatics analysis

Liuli Wang [a,b,c], Deming Liu [d], Shuo Liu [a], Tianyi Liao [a], Yajun Jiao [d], Xianglai Jiang [d], Yongfeng Wang [e], Yaqiong Chen [f], Haizhong Ma [g], Hui Cai [a,b,c,*]

[a] The First Clinical Medical College of Lanzhou University, No. 199, Donggang WestRoad, Chengguan District, Lanzhou, Gansu 730000, China
[b] Department of General Surgery, Gansu Provincial Hospital, Lanzhou 730000, China
[c] Key Laboratory of Molecular Diagnosis and Precision Treatment of Surgical tumor, Gansu Provincial Hospital, Lanzhou 730000, China
[d] Ningxia Medical University, Ningxia 750004, China
[e] Department of Clinical Medicine, Gansu University of Traditional Chinese Medicine, Lanzhou 730000, China
[f] Medical Department of Gansu Provincial Hospital, Lanzhou 730000, China
[g] Department of Quality Control, Gansu Provincial Hospital, Lanzhou 730000, China

ARTICLE INFO

ABSTRACT

Background and Objectives: Colorectal cancer (CRC) is one of the most common malignant tumors worldwide with high incidence and mortality rate, while colorectal liver metastasis (CRLM) is one of the major causes of cancer-related deaths. Therefore, the present study aims to identify the hub gene associated with CRC carcinogenesis and liver metastasis, and then explore its diagnostic and prognostic value as well as the potential regulation mechanism.

Methods: The overlapping differential co-expression genes among CRC, CRLM, and normal tissues were explored on the GSE49355 and GSE81582 datasets from the Gene Expression Omnibus (GEO) database by integrated bioinformatics analysis. Then, the hub prognostic genes were selected from the overlapping genes by univariate Cox proportional hazard analysis and online database Gene Expression Profiling Interactive Analysis 2 (GEPIA2). Subsequently, the clinical value of the hub genes was evaluated in the TCGA and GSE39582 cohorts. Finally, the underlying mechanisms of the hub gene regulating CRC carcinogenesis and metastasis were explored by Gene function annotation and DNA methylation analysis.

Results: Inositol mono-phosphatase 2 (IMPA2) was identified as the hub gene associated with CRC carcinogenesis and liver metastasis. IMPA2 had an excellent diagnostic efficiency, and its expression was significantly decreased in CRC and liver metastasis samples, being positively correlated with poor prognosis. Moreover, its low expression was associated with AJCC stage III+IV, T4, N1+2, and M1. In addition, our results revealed that the potential mechanisms used by IMPA2 to mediate CRC carcinogenesis and metastasis could be associated with lipid metabolism and epithelial mesenchymal transition (EMT). Finally, IMPA2 expression could be regulated by DNA methylation.

Conclusions: IMPA2 was identified and reported for the first time as a hub gene biomarker in the diagnosis and prognosis of CRC, which could regulate CRC carcinogenesis and liver metastasis through the regulation of lipid metabolism, EMT, and DNA methylation.

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors worldwide, with high incidence of morbidity and mortality. Although the diagnostic methods and comprehensive treatments have been improved over the past decades, the mortality rate of CRC is still ranked second and the morbidity rate is ranked third worldwide [1]. As the Global Cancer Incidence, Mortality and Prevalence (GLOBOCAN 2020) statistics reported more than 1.9 million new CRC cases and 935, 000 deaths in 2020 [1]. One of the major causes of cancer-related death
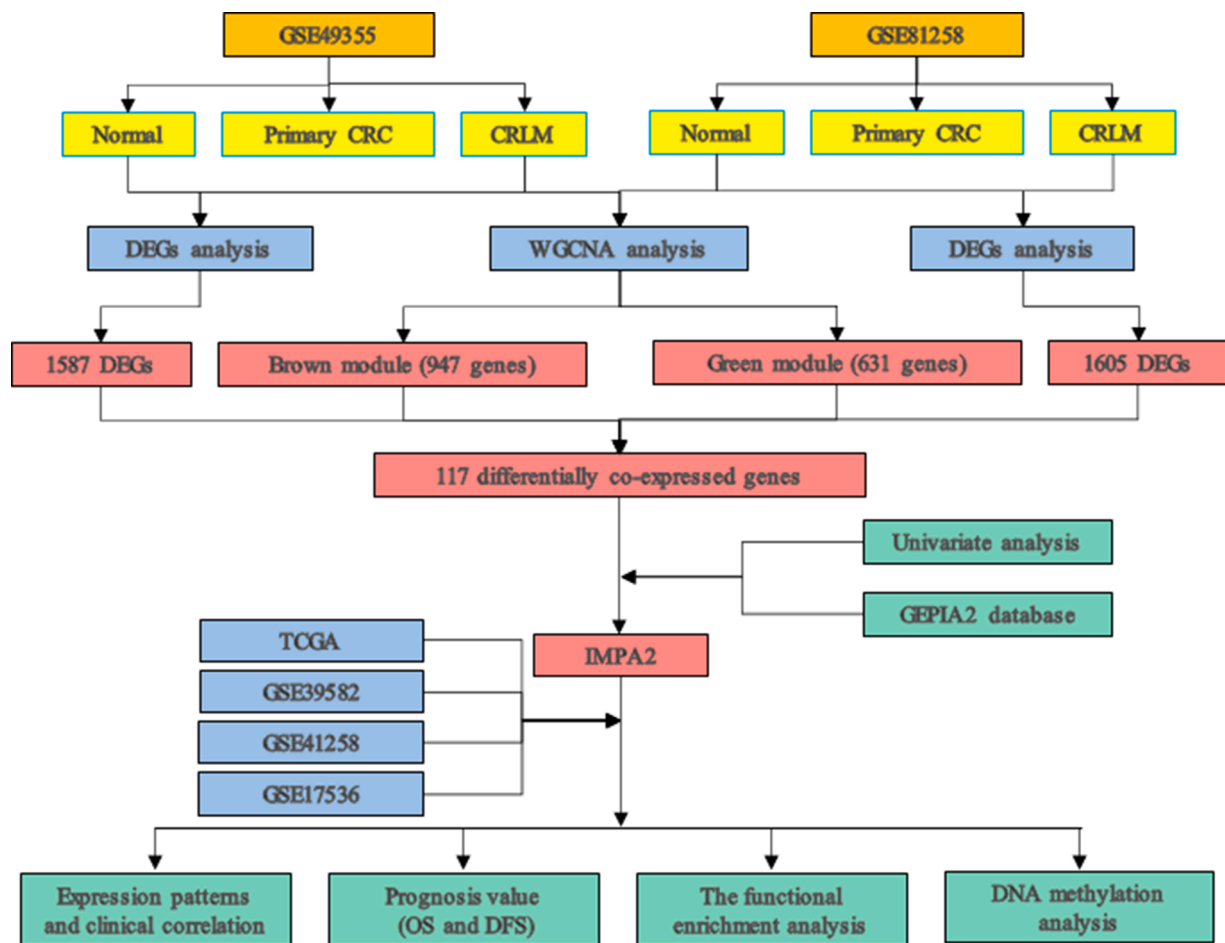

Fig. 1. Flow-chart of the present study. Abbreviations: CRC, colorectal cancer; CRLM, colorectal liver metastasis; WGCNA, weighted gene co-expression network analysis; DEG, differentially expressed genes; IMPA2, inositol Mono-phosphatase 2; GEPIA2, Gene Expression Profiling Interactive Analysis 2; TCGA, The Cancer Genome Atlas; OS, overall survival; DFS, disease-free survival.

in patients with CRC is colorectal liver metastasis (CRLM) [2]. Approximately 25–30% of patients with CRC already have CRLM at the time of the initial diagnosis and up to 60% develop CRLM during the course of treatment [3]. CRLM leads to the death of more than half of CRC patients and contributes to a low 5-year overall survival rate of less than 20% in patients with liver metastasis, which is significant lower compared with that of patients without liver metastasis [4,5]. Therefore, it is crucial to find relevant genes working as biomarkers of CRC carcinogenesis and liver metastasis, as well as discover their potential molecular mechanism, providing an additional approach for an effective prevention and treatment to improve the prognosis and quality of life of patients with CRC.

Gene expression profile analysis based on bioinformatics has been used to identify the key genes associated with progression and prognosis of tumors thanks to the popularization and gradual development of gene microarray and RNA sequencing [6]. One important analysis to identify appropriate key genes is the evaluation of differentially expressed genes (DEGs), which provides the possibility to study the molecular mechanisms underlying genome regulation and identify quantitative changes in their expression between tumors and normal tissues [7,8]. However, previous studies focused more on the role of a single DEG than on the complex connections between DEGs [9–12]. Weighted gene co-expression network analysis (WGCNA) is another novel and effective method to identify the significant modules of highly correlated genes and interested modules associated with clinical traits [13,14]. In the WGCNA network, highly correlated genes are clustered into a module, and the relationships between gene modules and clinical traits are determined to identify significant modules. Therefore, the integration analysis of WGCNA and DEGs allows the identification of the differentially co-expressed genes as the novel candidate biomarkers.

In the present study, the differentially co-expressed genes related to CRC and CRLM were identified on the two datasets of the Gene Expression Omnibus (GEO) database using the integrated analysis of WGCNA and DEGs. Then, the key prognostic genes were identified by univariate Cox proportional hazard analysis and the public database Gene Expression Profiling Interactive Analysis 2 (GEPIA2). The expression pattern, prognostic value, and correlation with clinical characteristics were verified in The Cancer Genome Atlas (TCGA) database and other GEO datasets. In addition, the potential molecular mechanism of CRC progression and metastasis were further explored through the functional enrichment analysis and DNA methylation. Therefore, the present study contributed to the identification of novel biomarkers associated with the diagnosis and prognosis of CRC in patients with liver metastases, as well as to the discovery of their molecular mechanism of action.

Materials and methods

GEO data set and data processing

Gene expression raw microarray profiles of CRC were evaluated in five independent datasets (accession nos. GSE49355 [15], GSE81582 [16], GSE39582 [17], GSE41258 [18], and GSE17536 [19]) from the GEO database (www.ncbi.nlm.nih.gov/geo). GSE49355 included 57
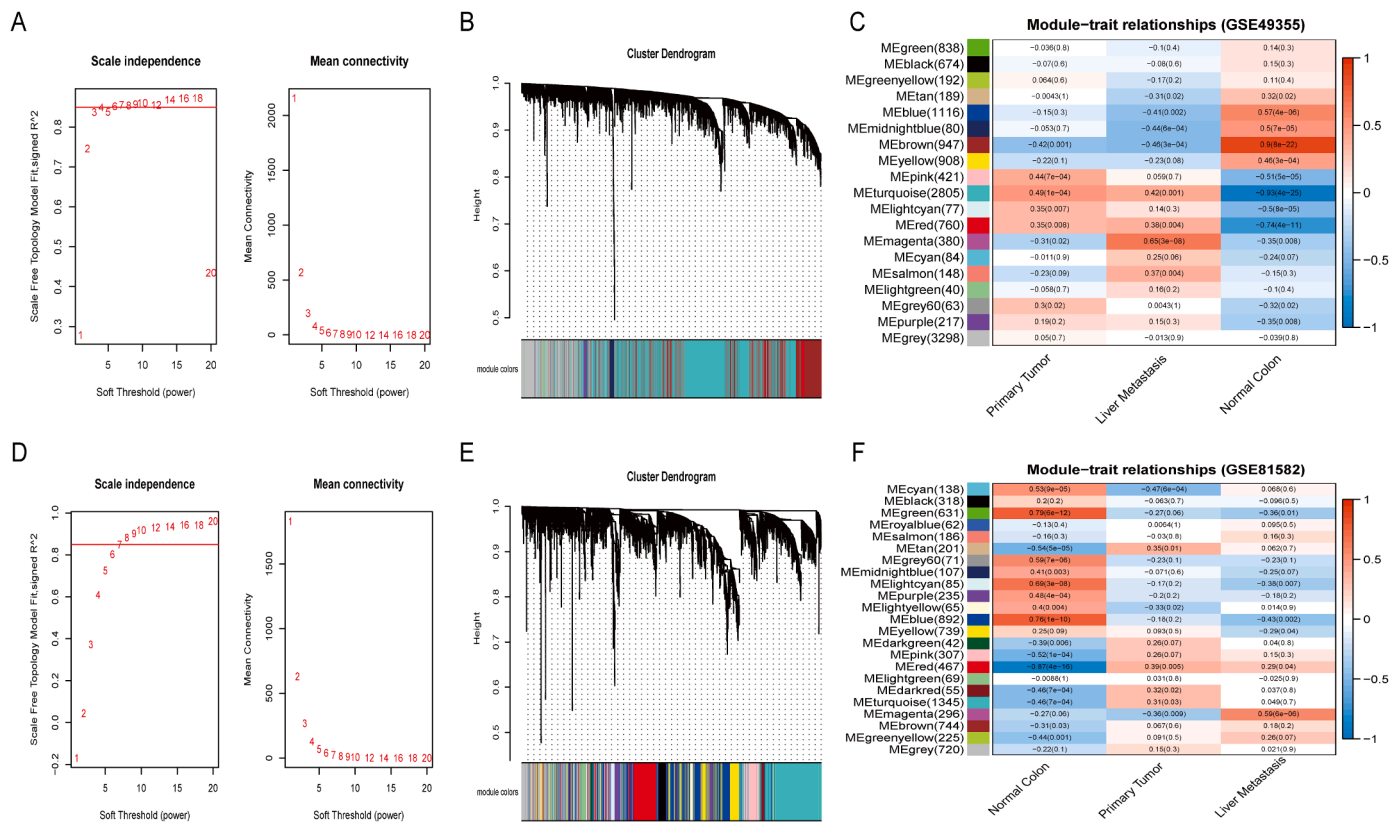
Fig. 2. WGCNA for construction and identification of the interested modules. **(A)** Construction of scale-free network in the GSE49355 dataset; **(B)** Hierarchical clustering tree of genes based on topological overlap different color branches of the cluster tree represent different modules; **(C)** Module-trait relationships. Each row corresponds to a color module and each column correlates to a clinical trait (CRLM, primary CRC and normal tissues). Each cell contains the corresponding correlation and *P*-value; **(D)** Construction of scale-free network in the GSE81582 dataset; **(E)** Hierarchical clustering tree of genes based on topological overlap different color branches of the cluster tree represent different modules; **(F)** Module-trait relationships. Each row corresponds to a color module and each column correlates to a clinical trait (CRLM, primary CRC and normal tissues). Each cell contains the corresponding correlation and *p*-value. Abbreviations: WGCNA, weighted gene co-expression network analysis; CRC, colorectal cancer; CRLM, colorectal liver metastasis.

samples (19 CRLM, 20 primary CRC and 18 normal tissues); GSE81582 included 51 samples (19 CRLM, 23 primary CRC and 9 normal tissues); GSE39582 included 585 tumor samples; GSE41258 included 287 samples (47 CRLM, 186 primary CRC and 54 normal tissues); and GSE17536 included 177 tumor samples. GSE49355 and GSE81582 datasets were used to construct WGCNA and DEGs analyses for the present study. GSE39582, GSE41258, and GSE17536 datasets were used as the training cohort for the selection of hub genes and the validation cohort for the expression, clinical correlation, diagnostic performance, and prognostic value of the hub genes. The expression profiles of the above three datasets were obtained using the "GEOquery" package [20] and the normalization for the microarray data was completed using the "limma" package [21]. Probes were converted to the gene symbols based on a manufacture-provided annotation file, and duplicated probes for the same gene were removed by determining the median expression of all its corresponding probes.

TCGA data set

Transcriptome data of RNA-seq and paired clinical data were downloaded from the TCGA-COAD dataset. The exclusion criteria were the following: 1) patients with unknown survival status, survival time, American Joint Committee on Cancer (JACC) stage, and Tumor Node Metastasis (TNM) stage; 2) patients who died within a follow-up period of 30-day. Finally, a total of 434 samples (including 393 colon tumor tissues and 41 normal tissues) from the TCGA dataset were used as another validation cohort.

This study did not require ethical approval since all data were downloaded from the GEO and TCGA, which are publicly available databases. The flowchart of the present research is shown in **Fig. 1**.

Weighted gene co-expression network construction

The genes in the GSE49355 and GSE81582 datasets were identified for the construction of the gene co-expression networks. A weighted gene co-expression network was constructed using the "WGCNA" package by choosing 6 and 8 as the soft thresholds for GSE49355 and GSE81582, respectively [22]. This package ensured that the network satisfied a scale-free topology ($R^2 > 0.85$) on the basis of the linear regression model fitting index obtained from the function 'pickSoftThreshold' operation. The weighted adjacency matrix was transformed into a topological overlap matrix (TOM) to estimate network connectivity, and the hierarchical clustering method was used to construct the cluster tree structure of the TOM matrix. Different branches of the cluster tree represent different gene modules, and different colors represent different modules. A correlation analysis was performed based on each module eigengene to identify modules that were significantly associated with the sample traits. Then, the modules with high correlation coefficient, which were considered as the candidates relevant to clinical traits, were selected for subsequent analysis.

Identification of DEGs and interaction with the modules of interest

The "limma" package was used to evaluate the GSE49355 and GSE81582 datasets and identify the DEGs among CRLM, primary CRC, and normal tissues. Genes with the cut-off criteria of |logFC| $\geq$ 1 and *p*
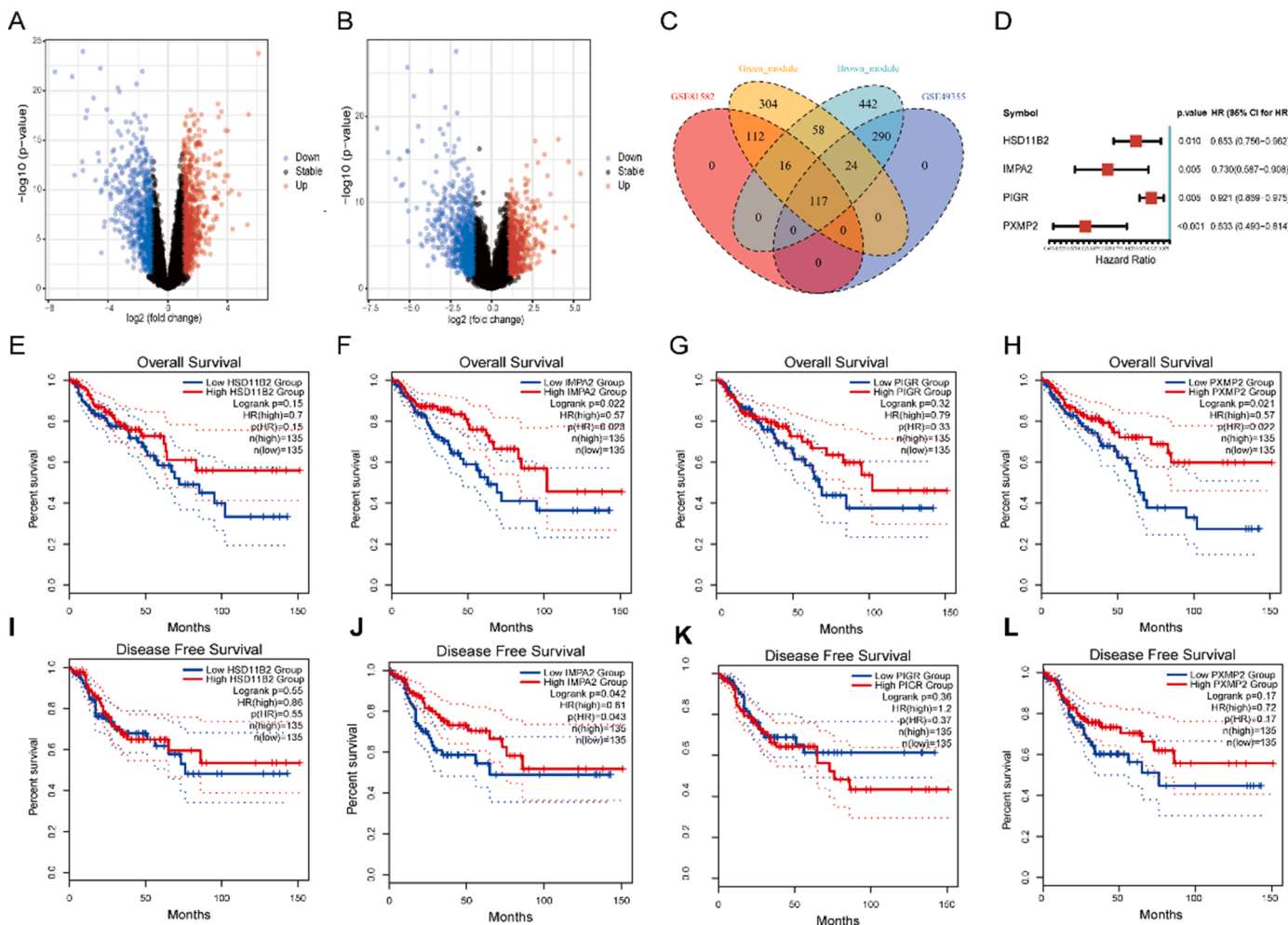
Fig. 3. DEGs analysis and the key gene selection **(A)** Volcano plot of DEGs in the GSE49355; **(B)** Volcano plot of DEGs in the GSE81582; **(C)** The Venn diagram of genes among DEGs and co-expression modules; **(D)** The forest plot of univariate Cox proportional hazard analysis and 4 genes were considered as candidate genes associated with prognosis; **(E-H)** OS analysis for HSD11B2, IMPA2, PIGR, and PXMP2 in the GEPIA2 database. **(I-L)** DFS analysis for HSD11B2, IMPA2, PIGR, and PXMP2 in the GEPIA2 database. Abbreviations: DEG, differentially expressed genes; OS, overall survival; DFS, diseases-free survival; GEPIA2, Gene Expression Profiling Interactive Analysis 2; HSD11B2, hydroxysteroid 11-beta dehydrogenase 2; IMPA2, inositol Mono-phosphatase 2; PIGR, polymeric immunoglobulin receptor; PXMP2, peroxisomal membrane protein 2; HR, hazard ratio; CI, confidence interval.

value < 0.05 were considered as the DEGs. The DEGs of the GSE49355 and GSE81582 datasets were visualized as a volcano plot using the "ggplot" package. Subsequently, the overlapping genes between DEGs and co-expression genes that were extracted from the co-expression networks were used to identify potential prognostic genes, which were presented as a Venn diagram using the "VennDiagram" package.

Selection and validation of key genes

The 117 overlapping genes obtained above were screened in the GSE39582 cohort using the univariate Cox proportional hazard analysis to select the key genes with a prognostic value. The 4 genes with a *p* value < 0.01 were selected as the candidate genes. Then, the public database GEPIA2 (http://GEPIA2.cancer-pku.cn/) [23] was used to further examine the reliability of the selected candidate genes in CRC and inositol mono-phosphatase 2 (IMPA2) was finally identified as the key gene with a prognostic value. Furthermore, the expression patterns, clinical correlation, and survival rates associated to IMPA2 in CRC were validated in the TCGA and GEO datasets. The receiver operating characteristic (ROC) curve was performed to verify the diagnostic performance of IMPA2 using the "pROC" package. Differences in overall survival (OS) and disease-free survival (DFS) were evaluated by the Kaplan-Meier survival curve using the "survival" package.

Functional enrichment and GSEA analysis

The genes positively and negatively correlated with IMPA2 were analyzed in the GSE39582 dataset using the "cor.test" function (stat. ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html) in R [24]. In this correlation analysis, the genes with the absolute value of a Pearson's correlation coefficient ≥ 0.3 and *p* value < 0.05 were included. Then, the Gene Ontology Term (GO), pathway enrichment, and Gene Set Enrichment Analysis (GSEA) were performed to elucidate the potential biological functions of IMPA2. The GO and pathway analysis of the genes correlated to IMPA2 were performed using Metascape (http://metascape.org/), and GSEA analysis was performed using the "clusterProfiler" package with a cut-off criterion of *p* < 0.05 [25] and was visualized using the "ggplot2" package [26].

Correlation between IMPA2 expression and DNA methylation

The correlation analysis was performed to further investigate the correlation between IMPA2 expression and DNA methylation using the MEXPRESS database (http://mexpress.be/), an online tool for the integration and visualization of gene expression, DNA methylation and clinical data in the TCGA dataset [27,28]. Moreover, the online tool cBioPortal (http://www.cbioportal.org/) [29] analyzing the copy
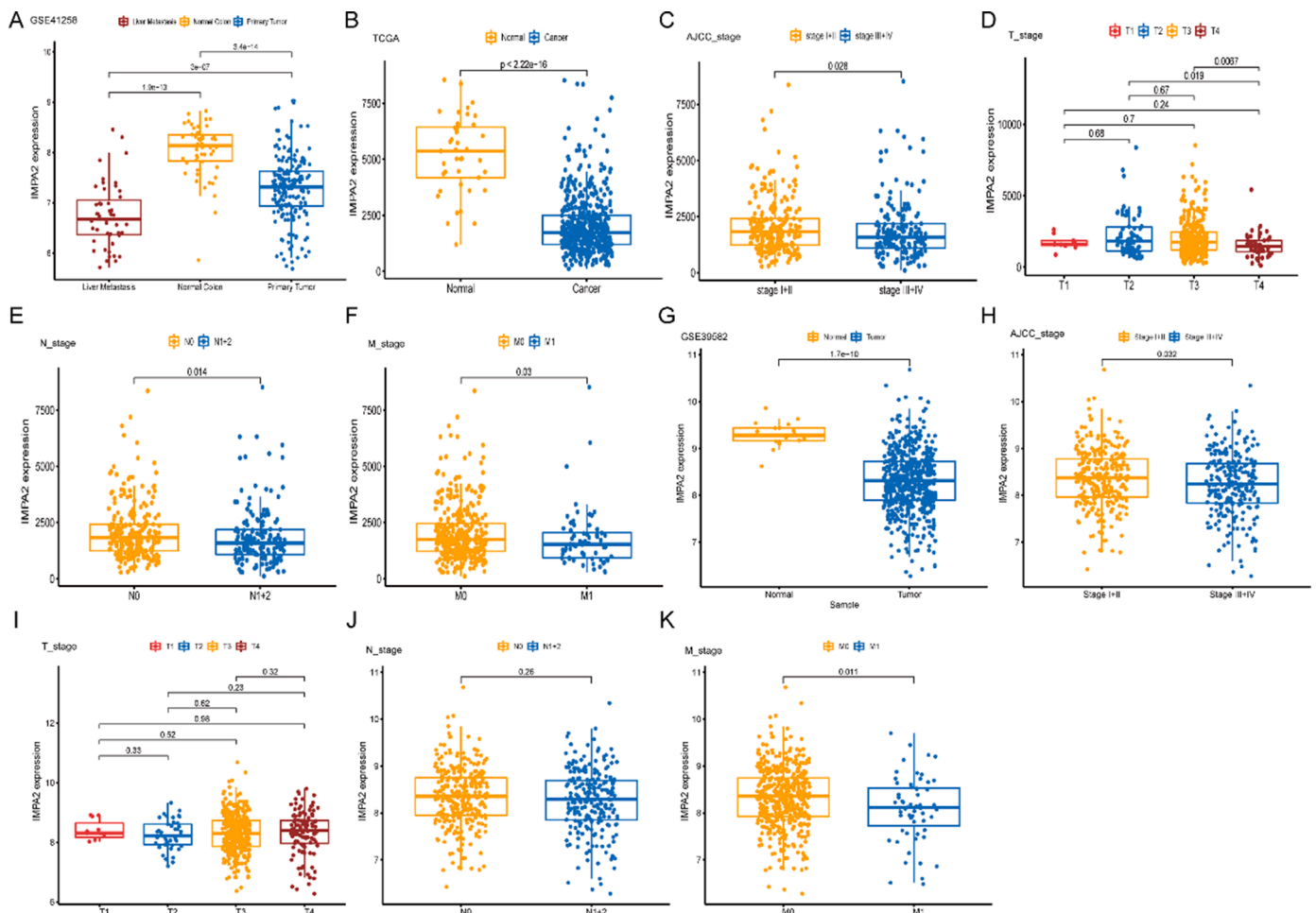
Fig. 4. Expression patterns of IMPA2 in CRC. **(A)** IMPA2 expression among liver metastases, primary tumors and normal tissues in the GSE41258 dataset; **(B)** IMPA2 expression between tumors and normal tissues in the TCGA cohort; **(C-F)** Correlation between IMPA2 expression and clinical characteristics, which showed that IMPA2 expression were significantly associated with AJCC Stage **(C)**, T stage **(D)**, N stage **(E)**, and M stage **(F)**; **(G-K)** Similar results were observed from the GSE39582 cohort. Abbreviations: CRC, colorectal cancer; TCGA, The Cancer Genome Atlas; AJCC, American Joint Committee on Cancer; IMPA2, inositol Mono-phosphatase 2.

number alterations (CNA) and gene mutations of IMPA2 in the colorectal adenocarcinoma (TCGA, PanCancer Atlas, 526 samples) was used to investigate whether the correlation between IMPA2 expression and DNA methylation contributes to the differences in genetic alterations.

Results

Identification of the WGCNA modules

A gene co-expression network was constructed using WGCNA to identify significant gene modules. First, a hierarchical clustering tree was generated in the GSE49355 and GSE81582 cohorts (**Supplementary Fig. S1**). Then, the scale independent value exceeded 0.85 with a low average connectivity when the soft threshold power was 6 and 8 in the above datasets, respectively (**Fig. 2A** and **Fig. 2D**). Finally, a total of 18 modules in the GSE49355 cohort (**Fig. 2B**) and 21 modules in the GSE81582 cohort (**Fig. 2E**) were identified, in which a gray module was excluded because it was not assigned to any cluster. In addition, the heatmap of module-trait relationships was plotted to evaluate the association between each module and the three clinical traits (CRLM, primary CRC, and normal tissues). The brown module (including 947 genes) in the GSE49355 dataset and green module (including 631 genes) in the GSE81582 dataset were closely correlated with CRLM (brown module: $r = 0.90$, $p = 8e - 22$; green module: $r = 0.79$, $p = 6e - 12$) (**Fig. 2C** and **F**).

Identification of overlapping genes between the DEGs and co-expression modules

A total of 1587 DEGs screened from the GSE49355 dataset and 1605 DEGs screened from the GSE81582 dataset were found using $|log2FC| > 1$ and $p < 0.05$ as the threshold. The DEGs of the two datasets were depicted in a volcano plot (**Fig. 3A–B**). Then, 117 overlapping genes from the intersection of the DEGs and co-expression modules were identified and selected for further analysis by Venn diagram analysis (**Fig. 3C**).

Identification of IMPA2 as the key gene

In this step, all the 117 overlapping genes contributed to the construction of the univariate Cox proportional hazards mode. Four candidate genes such as hydroxysteroid 11-beta dehydrogenase 2 (HSD11B2), inositol mono-phosphatase 2 (IMPA2), polymeric immunoglobulin receptor (PIGR), and peroxisomal membrane protein 2 (PXMP2) were identified (**Fig. 3D**, $p < 0.01$), which were further validated using the online GEPIA2 dataset. Among the above 4 candidate genes, only IMPA2 was significantly related to OS and DFS, and was previously rarely reported in the CRC (**Fig. 3E–L**). Therefore, IMPA2 was finally selected as the key gene to comprehensively explore its value in CRC.
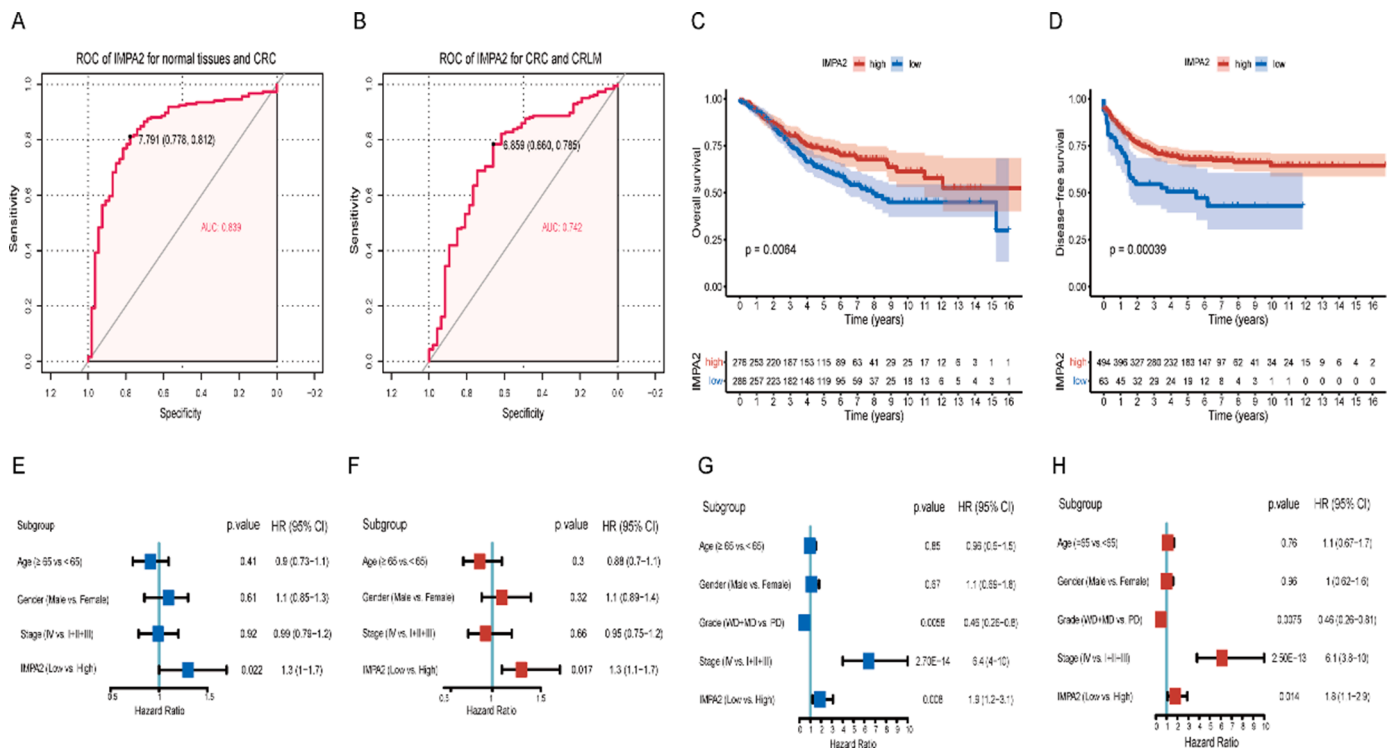
Fig. 5. Verification of the prognostic performance and prognostic value of IMPA2 in CRC. **(A-B)** ROC curves and AUC statistics were used to evaluate the capacity to discriminate CRC from normal controls **(A)**, and CRLM from CRC **(B)** with excellent specificity and sensitivity in the GSE41258 dataset. **(C-D)** Low IMPA2 expression was significantly correlated with OS and DFS in the GSE39582 cohort, which was consistent with survival analysis in the GEPIA2 database; **(E-F)** Low expression IMPA2 was an independent unfavorable prognostic factor in TCGA cohort by univariate **(E)** and multivariate **(F)** Cox proportional hazard analysis; **(G-H)** Low expression IMPA2 was an independent unfavorable prognostic factor in the GSE17358 dataset by univariate **(G)** and multivariate **(H)** Cox proportional hazard analysis. Abbreviations: CRC, colorectal cancer; CRLM, colorectal liver metastasis; IMPA2, inositol Mono-phosphatase 2; ROC, receiver operating characteristic; AUC, area under the curve; OS, overall survival; DFS, disease-free survival; GEPIA2, Gene Expression Profiling Interactive Analysis 2; TCGA, The Cancer Genome Atlas; HR, hazard ratio; CI, CI, confidence interval; WD, Well differentiated; MD, Moderately differentiated; PD, Poorly differentiated.

Verification of IMPA2 expression patterns and clinical correlation

In this part, the expression and clinical correlation of IMPA2 were analyzed among the CRC patients in the TCGA, GSE39582, and GSE41258 datasets. The mRNA expression of IMPA2 in the GSE41258 dataset had a decline trend among normal tissues, primary tumors, and liver metastases. IMPA2 expression was lower in the liver metastases but higher in the normal tissues when compared to the primary tumors (Fig. 4A). Similar results were found in the TCGA (Fig. 4B) and GSE39582 datasets (Fig. 4G). In addition, the correlation between clinical characteristics (including AJCC stage and TNM stage) and the expression of IMPA2 was analyzed. The analysis of the TCGA dataset revealed that IMPA2 expression was remarkably decreased in AJCC stage III+IV compared to stage I+II (Fig. 4C, $p = 0.028$). As regards the TNM stage, patients in T4, N1+2, and M1 stage showed a significantly lower IMPA2 expression than those in T2 and T3 (Fig. 4D, $p < 0.05$), N0 (Fig. 4E, $p = 0.014$), and M0 (Fig. 4F, $p = 0.03$). Similar results were obtained from the GSE39582 dataset (Fig. 4H–K).

Verification of the prognostic performance and prognostic value of IMPA2 in CRC

The ROC curve was generated to verify the diagnostic performance of IMPA2 based on the GSE41258 database. The AUC showed that IMPA2 had an excellent diagnostic efficiency for CRC patients (Fig. 5A, B). The patients were then divided into low-expression group and high-expression group based on the optimal cutoff value of IMPA2 expression in the GSE39582 dataset to further investigate the prognostic value of IMPA2 in CRC. The OS and DFS analyses of IMPA2 were performed by the Kaplan-Meier survival curve using the "survival" package. The

survival analysis indicated that CRC patients in the low-expression group had a significant worse OS and DFS than those in the high-expression group (Fig. 5C, D), which was consistent with the GEPIA2 analysis above.

Moreover, the clinical characteristics and IMPA2 were systematically analyzed by univariate and multivariate Cox proportional hazard analysis to confirm the independent prognostic value of IMPA2 in CRC patients. The results from the TCGA dataset revealed that the low expression IMPA2 was an independent unfavorable prognostic factor (Fig. 5E, F). Similar results were found in the GSE17536 dataset (Fig. 5G, H).

Functional enrichment analysis of IMPA2 co-expressed genes

The genes positively and negatively associated with IMPA2 were analyzed in the GSE39583 dataset to explore the potential role and regulation mechanism of IMPA2 in CRC, and 1013 IMPA2 co-expressed genes were identified. After screening GO and pathway enrichment analysis, up to 100 enriched clusters with IMPA2 co-expressed genes were found (**Supplementary Fig. S2**). The top 20 enriched clusters, originating from three categories (Reactome gene sets, GO biological process and COSUM) are shown in Fig. 6A. The IMPA2-associated genes were mainly enriched in metabolism processes, including metabolism of lipids, carbohydrate metabolic process, small molecule catabolic process, lipid biosynthetic process, organic hydroxy compound metabolic process, amino sugar and nucleotide sugar metabolism, valine, leucine, and isoleucine degradation, multicellular organismal homeostasis, phospholipid metabolism, and regulation of hormone levels. The GSEA analysis was performed for those co-expressed genes to further identify the molecular signaling pathways in CRC. Epithelial mesenchymal
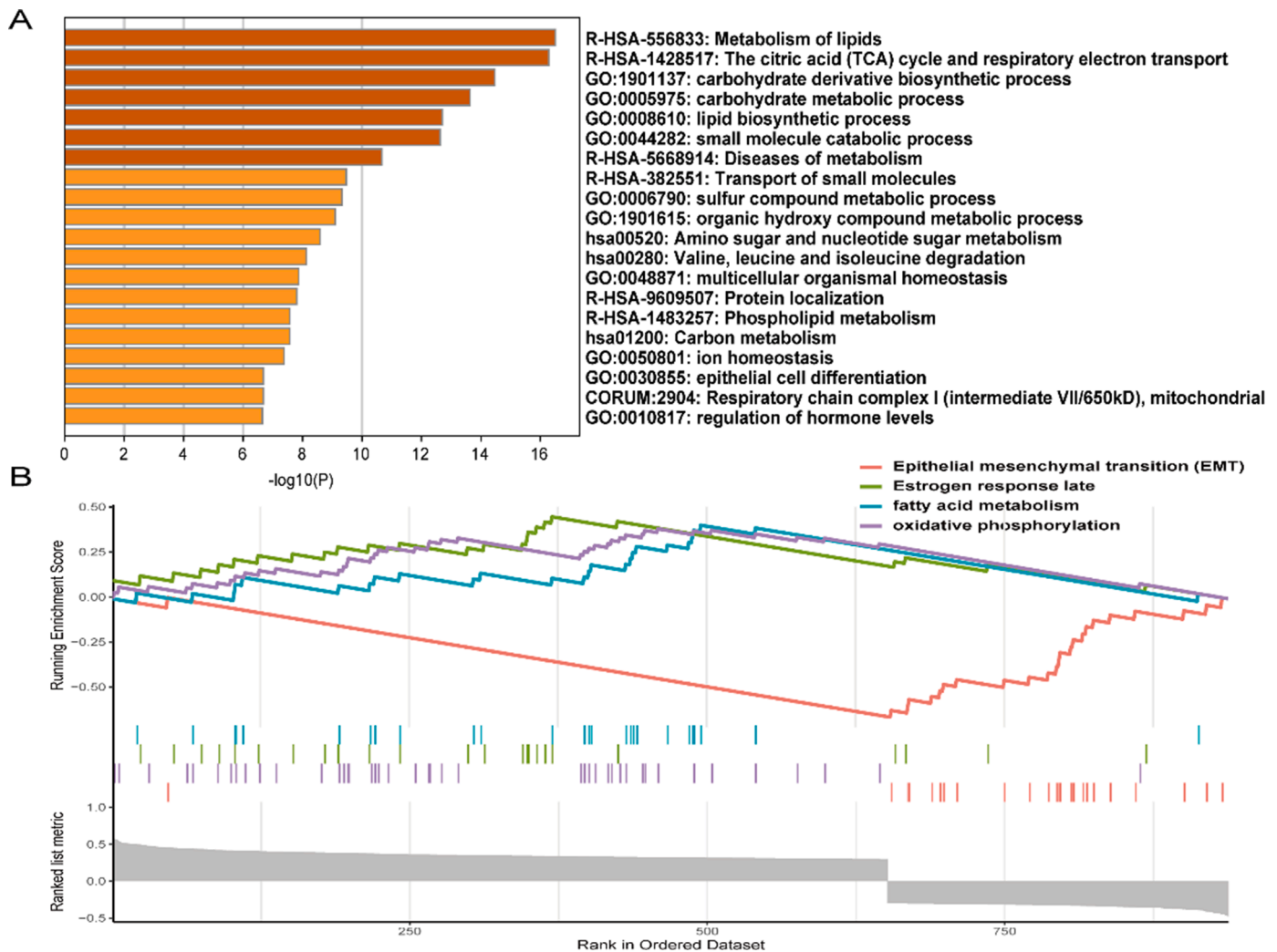
Fig. 6. Functional enrichment of IMPA2-correlated co-expression genes. **(A)** GO and pathway enrichment analysis of IMPA2 co-expressed genes. The top 20 pathway enrichment clusters based on Metascape analysis of IMPA2 co-expressed genes carried out with GO Biological Processes, Reactome Gene Sets and CORUM. Length of bars represent log10 (*p*-value) determined by the best-scoring term within each cluster; **(B)** GSEA for IMPA2 co-expressed genes in the GSE39582 dataset. Abbreviations: GO, Gene Ontology Term; GSEA, Gene Set Enrichment Analysis; IMPA2, inositol Mono-phosphatase 2.

transition (EMT), estrogen response late, fatty acid metabolism, and oxidative phosphorylation were differentially enriched in the gene sets both positively and negatively correlated with IMPA2 (Fig. 6B).

Correlation of IMPA2 expression with DNA methylation

In this part, the correlation between IMPA2 expression and DNA methylation in CRC was explored using the MEXPRESS database. The tumor tissues (primary, recurrent, and metastatic) had lower IMPA2 expression than normal tissues (Wilcoxon rank-sum test, $p = 4.188e\text{-}14$) (Fig. 7A). It was clearly shown that IMPA2 expression was inversely associated with DNA methylation in its four promoter regions located close to the transcription start site (TSS) (Pearson correlation coefficients ranging from -0.247 to -0.231; $p < 0.01$). However, a positive correlation existed between IMPA2 expression and DNA methylation in its another promoter region (Pearson $r = 0.177$; $p < 0.01$). Furthermore, the CNA and gene mutations of IMPA2 were analyzed by the cBioPortal database to investigate whether the lower expression of IMPA2 in CRC was caused by genetic alterations. Only 6 (1.1%, including 2 DNA amplification and 4 missense mutation) of the 526 patients with CRC were identified with existing genetic alterations, suggesting that the lower IMPA2 expression was regulated by DNA methylation rather than CNA and gene mutations (Fig. 7B).

Discussion

CRLM is the major cause of poor prognosis and tumor recurrence. It is well known that liver is the most common metastatic organ of CRC, and only 10% of patients with CRLM can benefit from surgical treatment. However, the recurrence rate after surgical resection is as high as 50–75% [30], and recurrence is the leading cause of cancer-related death [31]. Therefore, it is of utmost importance the identification of novel biomarkers for the diagnosis and prediction of CRC with liver metastasis. In the present study, a total of 117 differentially co-expressed genes associated with the carcinogenesis and liver metastasis of CRC were identified from the GSE49355 and GSE41258 datasets using the integrated analysis of WGCNA and DEGs. Then, 4 candidate genes were identified by constructing the univariate Cox proportional hazards mode. Furthermore, the online database GEPIA2 was further used to estimate the key genes with prognostic value. Finally, IMPA2 was identified as a potential gene to evaluate the risk of liver metastasis in CRC and to predict the prognosis of CRC patients.

The IMPA2 gene is a protein-coding gene for myo-inositol monophosphatase [32], which is one of the key enzymes in inositol phosphate metabolism. It catalyzes the dephosphorylation of inositol monophosphate and plays an important role in phosphatidylinositol signaling [33,34]. Previous studies suggested that IMPA2 is associated with

Fig. 7. Genetic features of IMPA2 in CRC. **(A)** Visualization of the association between IMPA2 expression and DNA methylation in TCGA-COAD cohort using the MEXPRESS tool. The sample are ordered by IMPA2 expression; **(B)** Analyzation of IMPA2 gene alteration in 526 CRC patients using cBioPortal database. Abbreviations: CRC, colorectal cancer; TCGA, The Cancer Genome Atlas; COAD, Colon adenocarcinoma; IMPA2, inositol Mono-phosphatase 2.

neuropsychiatric diseases, such as schizophrenia [35], epilepsy [36], bipolar disorder [37], Huntington's disease [38] and ischemic stroke [39]. Some studies have recently shown that IMPA2 expression is closely related to the progression and metastasis of malignant tumors. IMPA2 expression is predominantly downregulated in clear cell renal cell carcinoma (ccRCC) compared to normal tissues, and the downregulation of IMPA2 is strongly correlated with an unfavorable prognosis in ccRCC patients [40]. Importantly, IMPA2 expression is inversely associated with the metastatic potential of ccRCC cells. Moreover, Zhang et al. [41] indicated that IMPA2 expression is upregulated in cervical cancer. However, the role of IMPA2 in CRC has been rarely studied. In the present study, the role of IMPA2 in CRC progression, metastasis, and prognosis was comprehensively investigated.

The expression, diagnostic and prognostic value of IMPA2 was evaluated by bioinformatics analysis. The GEO and TCGA analyses indicated that the expression of IMPA2 was significantly decreased in CRC and liver metastases compared to that in the normal tissues. This result suggested that IMPA2 may be negatively correlated with CRC progression, however, it is needed to further validated by laboratory experiments. Moreover, IMPA2 expression was inversely correlated with AJCC and TNM stages. The ROC curve indicated that IMPA2 had an excellent diagnostic efficiency for CRC. The prognostic value of IMPA2 in CRC was then validated in the GSE39582 dataset by the Kaplan-Meier survival curve using the "survival" package. The results suggested that the high expression of IMPA2 was positively correlated with favorable OS and DFS in CRC patients. The univariate and multivariate Cox proportional hazard modes were constructed among clinical characteristics and the expression of IMPA2 in the TCGA and GSE17356 datasets to further confirm the independent prognostic value of IMPA2 in CRC

patients. Interestingly, the results indicated that the hazard ratio of the low IMPA2 expression group was significantly higher than that in the high IMPA2 expression group, indicating that the low IMPA2 expression was an independent unfavorable prognostic factor.

The functional and pathway enrichment analyses for IMPA2-associated genes were performed to elucidate the potential role and regulation mechanism of IMPA2 in CRC. Metascape and GSEA analyses suggested that IMPA2-correlated genes were associated with the metabolism of lipids and EMT. Lipid metabolism (including lipid synthesis, lipid utilization, and lipolysis) is associated with the development and progression of various cancers, contributing to tumorigenesis and metastasis by activating specific cell signaling pathways [42,43]. Lipid metabolism is often drastically altered during the transformation of the cells to a malignant phenotype. Tumor cells increase the synthesis of fatty acids for the synthesis of more lipids necessary for the plasma membrane of the newborn cells, as well as energy production and signal transduction to support the rapid proliferation and metastasis of tumor cells [44]. An increasing amount of studies demonstrated that the abnormal lipid metabolism is significantly associated with the occurrence and malignant transformation of CRC [45–48]. The EMT is an important process that occurs prior to tumor metastasis, and is considered as a required condition for tumor invasion and metastasis [49,50]. It is well-documented that lipid metabolism provides the energy and adequate environment for the EMT process of tumor cells and contributes to the metastasis of tumors [51]. Several studies investigated the correlation between EMT induction and lipid metabolism in cancer. For instance, Jiang et al. [52] suggested that PRRX1-induced EMT in salivary adenoid cystic carcinoma activates the metabolic reprogramming of free fatty acids to promote invasion and metastasis. In addition, Wang

et al. [53] indicated that apolipoprotein C-II induced EMT and lipid metabolism to promote gastric cancer peritoneal metastasis through the PI3K/AKT/mTOR pathway. However, the correlation among IMPA2, EMT, and lipid metabolism and the underlying mechanisms regulating their link together to induce the metastasis of CRC have not been investigated to date.

DNA methylation regulates gene expression in cancer and usually induces suppressor gene silencing and oncogene activation through hyper/hypomethylation [54–56]. The present study revealed that DNA methylation in most promoter regions was associated with low IMPA2 expression. Taking into consideration that the low IMPA2 expression could be caused by genetic alterations, the CNA and gene mutations of IMPA2 were analyzed, and the result showed that only 6 (1.1%) of the 526 patients with CRC were identified as carrying genetic alterations, indicating that the low IMPA2 expression was regulated by DNA methylation rather than genetic alterations. These findings suggested that DNA methylation could be a potential mechanism regulating IMPA2 expression.

Our study has some limitations. Firstly, the present study is a retrospective study, since all the data included in this study were obtained from publicly available databases. Therefore, prospective trials should be performed to further evaluate the diagnostic and prognostic value of IMPA2. Secondly, the number of samples in the normal and liver metastasis groups was quite smaller than that in the primary CRC group regarding the evaluation of the expression of IMPA2, which could result in a statistical problem. Thus, a consistent sample size in the CRLM, primary CRC, and normal groups should be considered to further confirm IMPA2 expression in our future research. In addition, further studies including clinical, *in vivo,* and *in vitro* experiments are needed to explore the underlying molecular mechanisms of IMPA2 for clinical applications.

Conclusions

In conclusion, IMPA2 was identified as the hub gene with a diagnostic and prognostic value, and its expression was significantly decreased in CRC and CRLM samples, while that low IMPA2 expression was significantly associated with poor prognosis. In addition, the potential mechanism of IMPA2 regulating tumorigenesis and liver metastasis of CRC could be associated with lipid metabolism, EMT, and DAN methylation.

CRediT authorship contribution statement

Liuli Wang: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Deming Liu:** Methodology. **Shuo Liu:** Validation. **Tianyi Liao:** Validation. **Yajun Jiao:** Data curation. **Xianglai Jiang:** Visualization. **Yongfeng Wang:** Visualization. **Yaqiong Chen:** Data curation. **Haizhong Ma:** Supervision. **Hui Cai:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declarations of Competing interest

None

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101435.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249.

[2] U. Testa, E. Pelosi, G. Castelli, Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells, Med. Sci. (2018) 6 (Basel).

[3] J. Engstrand, H. Nilsson, C. Strömberg, et al., Colorectal cancer liver metastases - a population-based study on incidence, management and survival, BMC Cancer 18 (2018) 78.

[4] P. Landreau, A. Drouillard, G. Launoy, et al., Incidence and survival in late liver metastases of colorectal cancer, J. Gastroenterol. Hepatol. 30 (2015) 82–85.

[5] S. Manfredi, C. Lepage, C. Hatem, et al., Epidemiology and management of liver metastases from colorectal cancer, Ann. Surg. 244 (2006) 254–259.

[6] T. Can, Introduction to bioinformatics, Methods Mol. Biol. 1107 (2014) 51–71.

[7] C.Y. Li, J.H. Cai, J.J.P. Tsai, C.C.N. Wang, Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis, Front Oncol. 10 (2020) 681.

[8] I.S. Segundo-Val, C.S. Sanz-Lozano, Introduction to the gene expression analysis, Methods Mol. Biol. 1434 (2016) 29–43.

[9] P.T. Makondi, P.L. Wei, C.Y. Huang, Y.J. Chang, Development of novel predictive miRNA/target gene pathways for colorectal cancer distance metastasis to the liver using a bioinformatic approach, PLoS One 14 (2019), e0211968.

[10] T. Zhang, J. Guo, J. Gu, et al., Identifying the key genes and microRNAs in colorectal cancer liver metastasis by bioinformatics analysis and *in vitro* experiments, Oncol. Rep. 41 (2019) 279–291.

[11] Z. Qian, G. Zhang, G. Song, et al., Integrated analysis of genes associated with poor prognosis of patients with colorectal cancer liver metastasis, Oncotarget 8 (2017) 25500–25512.

[12] T. Zhang, K. Yuan, Y. Wang, et al., Identification of candidate biomarkers and prognostic analysis in colorectal cancer liver metastases, Front. Oncol. 11 (2021), 652354.

[13] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinform. 9 (2008) 559.

[14] Q. Wan, J. Tang, Y. Han, D. Wang, Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma, Exp. Eye Res. 166 (2018) 13–20.

[15] Del Rio M, Mollevi C, Vezzio-Vie N, Bibeau F, et al., Specific extracellular matrix remodeling signature of colon hepatic metastases, PLoS One (2013) 773–780.

[16] Sayagués JM, Corchete LA, Gutiérrez ML, et al., Genomic characterization of liver metastases from colorectal cancer patients, Oncotarget (2016) 72908–72922.

[17] Marisa L, de Reyniès A, Duval A, et al., Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value, PLoS Med (2013) e1001453.

[18] Sheffer M, Bacolod MD, Zuk O, et al., Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer, Proc Natl Acad Sci U S A (2009) 7131–7136.

[19] Smith JJ, Deane NG, Wu F, et al., Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer, Gastroenterology (2010) 958–968.

[20] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor, Bioinformatics 23 (2007) 1846–1847.

[21] M.E. Ritchie, B. Phipson, D. Wu, et al., limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucl. Acids Res. 43 (2015) e47.

[22] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinform. 9 (2008) 559.

[23] Z. Tang, C. Li, B. Kang, et al., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucl. Acids Res. 45 (2017). W98-w102.

[24] M. Pagano, Simple linear regression and correlation, Nutrition 11 (1995) 179.

[25] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (2012) 284–287.

[26] K. Ito, D. Murphy, Application of ggplot2 to pharmacometric graphics, CPT Pharm. Syst. Pharmacol. 2 (2013) e79.

[27] A. Koch, T. De Meyer, J. Jeschke, W. Van Criekinge, MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data, BMC Genom. 16 (2015) 636.

[28] A. Koch, J. Jeschke, W. Van Criekinge, et al., MEXPRESS update 2019, Nucleic. Acids Res. 47 (2019). W561-w565.

[29] E. Cerami, J. Gao, U. Dogrusoz, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401–404.

[30] S.W. Jung, D.S. Kim, Y.D. Yu, et al., Risk factors for cancer recurrence or death within 6 months after liver resection in patients with colorectal cancer liver metastasis, Ann. Surg. Treat. Res. 90 (2016) 257–264.

[31] J.G. Fortner, Recurrence of colorectal cancer after hepatic resection, Am. J. Surg. 155 (1988) 378–382.

[32] T. Yoshikawa, G. Turner, L.E. Esterling, et al., A novel human myo-inositol monophosphatase gene, IMP.18p, maps to a susceptibility region for bipolar disorder, Mol. Psychiatry 2 (1997) 393–397.

[33] P.J. Bloch, A.E. Weller, G.A. Doyle, et al., Association analysis between polymorphisms in the myo-inositol monophosphatase 2 (IMPA2) gene and bipolar disorder, Prog. Neuropsychopharmacol. Biol. Psychiatry 34 (2010) 1515–1519.

[34] S.A. Jesch, X. Zhao, M.T. Wells, S.A. Henry, Genome-wide analysis reveals inositol, not choline, as the major effector of Ino2p-Ino4p and unfolded protein response target gene expression in yeast, J. Biol. Chem. 280 (2005) 9106–9118.

[35] J. Li, S. Huang, H.R. Dai, et al., A promoter polymorphism rs2075824 within IMPA2 gene affecting the transcription activity: possible relationship with schizophrenia, J. Cell. Mol. Med. 21 (2017) 658–664.

[36] C.A. Gurnett, P. Hedera, New ideas in epilepsy genetics: novel epilepsy genes, copy number alterations, and gene regulation, Arch. Neurol. 64 (2007) 324–328.

[37] Y. Tomioka, E. Jiménez, E. Salagre, et al., Association between genetic variation in the myo-inositol monophosphatase 2 (IMPA2) gene and age at onset of bipolar disorder, J. Affect. Disord. 232 (2018) 229–236.

[38] S. Sarkar, D.C. Rubinsztein, Inositol and IP3 levels regulate autophagy: biology and therapeutic speculations, Autophagy 2 (2006) 132–134.

[39] Q. Ma, Y. Yang, Y. Na, et al., IMPA2 polymorphisms and risk of ischemic stroke in a northwest Han Chinese population, Oncotarget 7 (2016) 75273–75278.

[40] Y.F. Lin, J.L. Chou, J.S. Chang, et al., Dysregulation of the miR-25-IMPA2 axis promotes metastatic progression in clear cell renal cell carcinoma, EBioMed. 45 (2019) 220–230.

[41] K. Zhang, L. Liu, M. Wang, et al., A novel function of IMPA2, plays a tumor-promoting role in cervical cancer, Cell Death. Dis. 11 (2020) 371.

[42] C. Cheng, F. Geng, X. Cheng, D. Guo, Lipid metabolism reprogramming and its potential targets in cancer, Cancer Commun. 38 (2018) 27 (Lond).

[43] R.R. Chen, M.M.H. Yung, Y. Xuan, et al., Targeting of lipid metabolism with a metabolic inhibitor cocktail eradicates peritoneal metastases in ovarian cancer cells, Commun. Biol. 2 (2019) 281.

[44] C. Cheng, Z. Li, X. Zhao, et al., Natural alkaloid and polyphenol compounds targeting lipid metabolism: treatment implications in metabolic diseases, Eur. J. Pharmacol. 870 (2020), 172922.

[45] S. Pan, Y. Guo, F. Hong, et al., Therapeutic potential of melatonin in colorectal cancer: focus on lipid metabolism and gut microbiota, Biochim. Biophys. Acta Mol. Basis Dis. 1868 (2022), 166281.

[46] T. Lu, L. Sun, Z. Wang, et al., Fatty acid synthase enhances colorectal cancer cell proliferation and metastasis via regulating AMPK/mTOR pathway, Onco. Targets Ther. 12 (2019) 3339–3347.

[47] Y. Gao, X. Nan, X. Shi, et al., SREBP1 promotes the invasion of colorectal cancer accompanied upregulation of MMP7 expression and NF-κB pathway activation, BMC Cancer 19 (2019) 685.

[48] H. Ran, Y. Zhu, R. Deng, et al., Stearoyl-CoA desaturase-1 promotes colorectal cancer metastasis in response to glucose by suppressing PTEN, J. Exp. Clin. Cancer Res. 37 (2018) 54.

[49] H. Hugo, M.L. Ackland, T. Blick, et al., Epithelial–mesenchymal and mesenchymal–epithelial transitions in carcinoma progression, J. Cell. Physiol. 213 (2007) 374–383.

[50] M. Yilmaz, G. Christofori, EMT, the cytoskeleton, and cancer cell invasion, Cancer Metastasis Rev. 28 (2009) 15–33.

[51] M. Li, X. Bu, B. Cai, et al., Biological role of metabolic reprogramming of cancer cells during epithelial-mesenchymal transition (Review), Oncol. Rep. 41 (2019) 727–741.

[52] Y.P. Jiang, Y.L. Tang, S.S. Wang, et al., PRRX1-induced epithelial-to-mesenchymal transition in salivary adenoid cystic carcinoma activates the metabolic reprogramming of free fatty acids to promote invasion and metastasis, Cell Prolif. 53 (2020) e12705.

[53] C. Wang, Z. Yang, E. Xu, et al., Apolipoprotein C-II induces EMT to promote gastric cancer peritoneal metastasis via PI3K/AKT/mTOR pathway, Clin. Transl. Med. 11 (2021) e522.

[54] S.A. Belinsky, Gene-promoter hypermethylation as a biomarker in lung cancer, Nat. Rev. Cancer 4 (2004) 707–717.

[55] G.H. Kang, S. Lee, H.J. Lee, K.S. Hwang, Aberrant CpG island hypermethylation of multiple genes in prostate cancer and prostatic intraepithelial neoplasia, J. Pathol. 202 (2004) 233–240.

[56] R. Maruyama, S. Toyooka, K.O. Toyooka, et al., Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features, Clin. Cancer Res. 8 (2002) 514–519.