



# LCN2 as a Potential Diagnostic Biomarker for Ulcerative Colitis-Associated Carcinogenesis Related to Disease Duration

Fushun Kou<sup>1,2†</sup>, Yuan Cheng<sup>1,2†</sup>, Lei Shi<sup>2</sup>, Jiajing Liu<sup>3</sup>, Yuyue Liu<sup>2</sup>, Rui Shi<sup>2</sup>, Guiying Peng<sup>3</sup> and Junxiang Li<sup>2\*</sup>

<sup>1</sup> Graduate School, Beijing University of Chinese Medicine, Beijing, China, <sup>2</sup> Gastroenterology Department, Dongfang Hospital, Beijing University of Chinese Medicine, Beijing, China, <sup>3</sup> School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China

## OPEN ACCESS

### Edited by:

Aditi Banerjee,  
University of Maryland, Baltimore,  
United States

### Reviewed by:

Tingting Zhu,  
Zhejiang University, China  
Yali Zhou,  
Guilin Medical University, China

### \*Correspondence:

Junxiang Li  
junxiangli1226@126.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

### Specialty section:

This article was submitted to  
Gastrointestinal Cancers:  
Colorectal Cancer,  
a section of the journal  
Frontiers in Oncology

**Received:** 04 November 2021

**Accepted:** 20 December 2021

**Published:** 17 January 2022

### Citation:

Kou F, Cheng Y, Shi L, Liu J,  
Liu Y, Shi R, Peng G and Li J (2022)  
LCN2 as a Potential Diagnostic  
Biomarker for Ulcerative Colitis-  
Associated Carcinogenesis  
Related to Disease Duration.  
Front. Oncol. 11:793760.  
doi: 10.3389/fonc.2021.793760

**Background:** Patients with long-duration ulcerative colitis (UC) had a higher risk of developing ulcerative colitis-associated carcinogenesis (UCAC) when compared to those with short-duration UC. This study aimed to discover the biomarker for cancer surveillance related to disease duration.

**Methods:** The microarrays were divided into short-duration (<10 years) UC, long-duration (≥10 years) UC, UCAC, and normal groups in the Gene Expression Omnibus (GEO) datasets. Differentially expressed genes (DEGs) of GEO and the hub genes of the selected weighted gene co-expression network analysis (WGCNA) were intersected to obtain the overlapping genes. Among these genes, the key gene was identified by using the protein-protein interaction (PPI) network, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, the cytoHubba of Cytoscape, and the expression levels. Also, immunofluorescence of human colonic mucosa and animal experiment were used to validate the expression trend of the key gene in the progress of UC developing into UCAC.

**Results:** Lipocalin-2 (LCN2) was more relevant with disease duration of UC and significantly negatively correlated with the risk of UCAC. The expression level of LCN2 in short-duration UC was higher than that of long-duration UC ( $P < 0.01$ ), long-duration UC was higher than that of UCAC ( $P = 0.001$ ), and UC and UCAC were all higher than that of the normal ( $P < 0.001$ ). We then discovered that the expression trend of LCN2 in blood and stool samples was consistent with that in colorectal mucosa.

**Conclusion:** The research indicates that LCN2 could be a novel biomarker to evaluate cancer surveillance related to disease duration of developing UC into UCAC. Compared with that of blood samples, stool detection of LCN2 may have more advantages for diagnosis value of early stage of UCAC as a complement to colonoscopy surveillance.

**Keywords:** LCN2, ulcerative colitis, long-duration disease, ulcerative colitis-associated carcinogenesis, colorectal cancer

## INTRODUCTION

Ulcerative colitis (UC) is a group of chronic idiopathic inflammatory diseases of the gastrointestinal tract and characterized by symptoms that evolve in a relapsing and remitting manner. UC has become a more common disease in many parts of the world. The prevalence of UC varies considerably across different countries and is higher in Europe (505/100,000 in Norway) and North America (286/100,000 in the United States). A steady increase has been reported over the last few decades in South America, Africa, and Asia (1). The pathogenesis of UC involves a complex interaction among the host genetic background, microbial shifts, and environmental cues, leading to inappropriate and chronic activation of the mucosal immune system (2, 3). Despite efforts for so many decades, the complex pathogenesis of UC has not yet been fully understood.

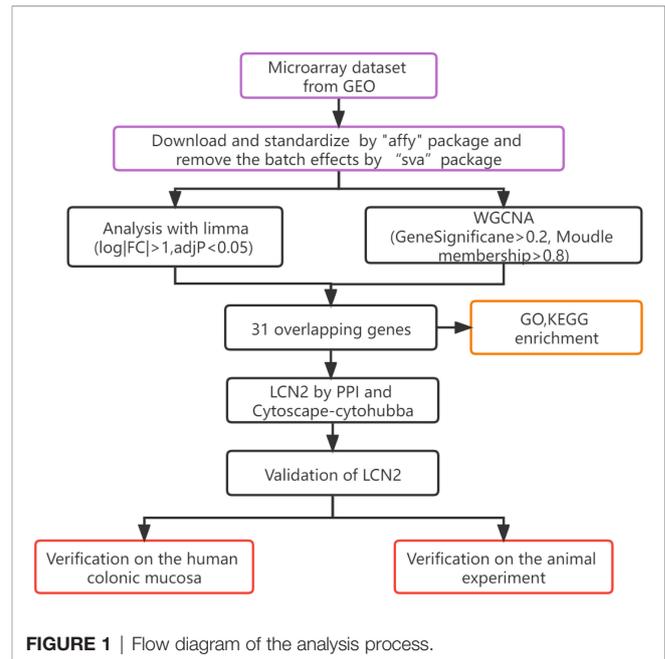
A number of population studies in various parts of the world have shown that patients with UC with longer disease duration ( $\geq 10$  years) have a higher risk of developing ulcerative colitis-associated carcinogenesis (UCAC) when compared to those with shorter duration. Cumulative risk of colorectal cancer (CRC) for patients with UC was found to be 2% at 10 years, 8% at 20 years, and 18% at 30 years (4). Various transcriptomic studies have been carried out in the UC field, including discrimination between disease subtypes (Crohn's disease vs. UC), disease phase (active/mildly active/inactive), colon biopsy locations, genders, etc. (5–7). However, the effects of different disease durations on transcriptome have not yet been explored systematically. Colonoscopy detection of neoplasia has long been employed as the standard procedure, and biomarkers could be a complement to colonoscopy surveillance for detecting UCAC at an early stage (8). Therefore, our study aimed to identify transcriptomic changes in patients with UC of different disease durations developing into UCAC and to provide a foundation study in the development of molecular biomarkers.

## MATERIALS AND METHODS

The workflow of hub gene extraction and creation pipeline were shown in **Figure 1**.

### Datasets From Gene Expression Omnibus Database

The gene expression datasets were obtained through the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>). The microarray studies were searched using the following terms: “Ulcerative Colitis”, “Inflammatory Bowel Disease”, “*Homo sapiens*”, “gene expression” and “microarray”. Statistical group comparisons of UC were divided into two parts, including long-duration UC ( $\geq 10$  years) and short-duration UC ( $< 10$  years). The datasets were included according to the following eligibility criteria: 1) with at least 10 total samples; 2) includes at least five cases and five controls; 3) the age range of



patients undergoing biopsy should be from 18 to 70 years; 4) should contain information with regard to age and disease duration; 5) the raw data or gene expression profiling carried out by array should be available in the GEO; and 6) the samples of baseline data should not receive biologics or any steroid medication.

### Data Preprocessing of Gene Expression Omnibus Datasets

The gene expression matrix and related annotation document for each array dataset were downloaded from the GEO database. After we screened out the dataset that can be used for subsequent analysis, we used the “affy” package to download and standardize the raw “CEL” files of the microarray data (9). RNA sequencing data were transformed by using the “voom” algorithm in order to convert count data to values similar to those resulting from microarrays (10). We used the corresponding annotation document to map the microarray probes to ENTREZ ID. The 4 datasets (5, 11–13) (GSE114603, GSE38713, GSE47908, and GSE37283) were integrated to significantly increase the sample sizes, so as to avoid the single dataset leading to unreliable results and reducing the validity of the bioinformatics analysis. Heterogeneity and potential variables are generally recognized as the main sources of bias and variability in high-throughput experiments. The merged data were preprocessed by the sva package to remove the batch effect and heterogeneity among various datasets (14). If multiple probes were mapped to the same symbol, then the mean value was adopted.

### Identification of the Differentially Expressed Genes

The “LIMMA” (Linear Model of Chip Data) R software package was used to determine the microarrays. The microarrays were then divided into short-duration UC, long-duration UC, UCAC,

and the normal control group in the GEO. The  $|\log_2\text{fold change (FC)}| > 1$  and adjusted  $P < 0.05$  were regarded as the cutoff criteria to determine the differentially expressed genes (DEGs).

## Identification of Co-Expression Modules Based on Weighted Gene Co-Expression Network Analysis

The co-expression networks facilitate network-based gene screening methods that can be used to identify the candidate biomarkers and therapeutic targets. In our research, the gene expression data map of the UC datasets was constructed, and the gene expression network of R (15) was constructed based on the weighted gene co-expression network analysis (WGCNA) package. Hierarchical clustering was performed on the research samples by distance and average methods to detect the outliers and remove the four abnormal samples. WGCNA was used to explore the modules of highly correlated genes among the samples of related modules to that of external sample traits. To build a scale-free network, soft powers of  $\beta = 6$  were selected by using the function `pickSoft Threshold`. Next, the adjacency matrix was established by the following formula:  $a_{ij} = |s_{ij}|^\beta$  ( $a_{ij}$ : adjacency matrix between gene  $i$  and gene  $j$ ;  $s_{ij}$ : similarity matrix composed of Pearson correlations of all gene pairs;  $\beta$ : soft power Value) and was converted to a topological overlap matrix (TOM) and the corresponding degree of dissimilarity (1-TOM). After that, a hierarchical clustering dendrogram of 1-TOM matrix was constructed to divide similar gene expressions into different gene co-expression modules. To further identify the functional modules in a co-expression network, the module-trait associations and clinical trait information were calculated according to previous research studies (16, 17). Therefore, modules with high correlation coefficients were considered as candidates related to clinical features and selected for subsequent analyses.

## Construction of Protein-Protein Interaction and Screening of Hub Genes

In this study, the DEGs of UC datasets and the hub genes of the selected WGCNA modules were intersected to obtain the overlapping genes. The online tool STRING (18) is designed for predicting protein-protein interactions (PPIs) and to construct a PPI network model by selecting genes with a score of  $\geq 0.4$ . To explore the Gene Ontology (GO) of selected genes, the Metascape database (<http://metascape.org>) was used to study the functions among the genes of interest, with a cutoff criterion of  $P < 0.05$  (19). GO annotation contains three sub-ontologies, biological process (BP), cell component (CC), and molecular function (MF), and these can identify the biological properties of genes and genomes of all organisms (20). The Kyoto Encyclopedia of Genes and Genomes (KEGG) Database was applied for pathway enrichment analysis, and significant ( $P < 0.05$ ) genes were selected.

## Elucidation of Hub Gene Expression Pattern

To get the hub genes with the highest risk of carcinogenesis related to disease duration, we confirmed the expression patterns

of the overlapping genes in short-duration UC, long-duration UC, UCAC, the normal control group, and selected hubs by one-way ANOVA (\*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ ). In GEO94648 (21) blood samples and GEO11223 (22) colorectal samples, the expression pattern of the hub gene in the short-duration UC, long-duration UC, and normal tissue was repeatedly verified to observe its expression characteristics. Limited by the current database, we did not find more datasets containing tissue of UCAC in the blood and colorectal samples.

## Verification on the Animal Experiment Animals

All experimental procedures that involved mice in this study were approved by the Animal Ethics Committee of Beijing University of Chinese Medicine (No. BUCM-4-2020122103-4168) in strict accordance with the Guide for the Care and Use of Laboratory Animals published by the Ministry of Science and Technology of China. Male C57BL/6 mice (weight  $20 \pm 2$  g) were purchased from Weitonglihua Laboratory Animal Research Center at Beijing (Number: SCXK-2019-0013). During the adaptation period of 7 days before modeling, mice were housed under standard conditions of temperature and humidity with freedom of access to water and food and were subjected to a 12-h light/dark cycle.

## Modeling

Colitis-associated cancer (CAC): Mice received a single intraperitoneal (i.p.) injection of azoxymethane (AOM) (10 mg/kg). One week later, they were exposed to 2% dextran sulfate sodium salt (DSS) (wt/vol) in drinking water for 7 consecutive days followed by 14 days of regular drinking water. This cycle was then repeated twice. Chronic colitis: Two weeks later, they were exposed to 2% DSS (wt/vol) in drinking water for 7 consecutive days followed by 14 days of regular drinking water. This cycle was then repeated twice. Acute colitis: Nine weeks later, they were exposed to 2% DSS (wt/vol) in drinking water for 7 consecutive days (23).

DSS was purchased from MP Biomedicals (Santa Ana, CA, USA). AOM was purchased from Sigma Aldrich (St. Louis, MO, USA). Body weight was assessed once a week throughout the duration of the experiment, and stool of the four groups was kept in the last 3 days for follow-up testing. All mice were sacrificed by cervical dislocation at the beginning of the 11th week. The colons were excised from the ileocecal junction to anus, and the length and weight of colons were measured, after which the colons were examined macroscopically and tumor numbers were counted.

## Histopathological Analysis

Formalin-preserved colons were dehydrated and embedded in paraffin by standard techniques. Subsequently, paraffin-embedded samples were sectioned at 5  $\mu\text{m}$  and stained with hematoxylin and eosin (H&E).

## Immunofluorescence Staining

The protein expression levels of lipocalin-2 (LCN2) in the normal, acute colitis, chronic colitis, and CAC mice were detected by immunofluorescence staining. The sections were then incubated with an anti-LCN2 antibody (A2092, ABclonal,

Cambridge, MA; 1:50 dilution) at 4°C overnight and then washed with phosphate buffered solution (PBS) three times. The sections were then co-incubated with secondary antibody (GB25303, Servicebio, Wuhan, China; 1:400 dilution). The nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen). The images were taken with confocal microscope (CSIM100, SUNNY, Beijing, China).

### Enzyme-Linked Immunosorbent Assay

The extracted colonic tissues were weighed and homogenized with tissue extraction reagent on ice for 3 min using a homogenizer. Interleukin (IL)-6, tumor necrosis factor (TNF)- $\alpha$ , myeloperoxidase (MPO), calprotectin (CALP), and LCN2 in the colonic mucosa and CALP and LCN2 in the stool were measured using mouse ELISA kits (Cusabio Biotech, Newark, DE, USA). Four samples of each group were randomly chosen for analysis.

### Real-Time PCR and Western Blot

The mRNA expression levels of LCN2 in the normal, acute colitis, chronic colitis, and CAC mice were detected by RT-PCR. The LCN2 primer used for the amplification was as follows: 5'-CACCACGGACTACAACCA-3' and 5'-ACACTCACCACCCATTCA-3' (Sangon Biotech, China). Here,  $\beta$ -actin was regarded as the internal control. The reliability of the PCR results was verified in correspondence with the dissolution curve. The cycle threshold (Ct, which is the inflection point on the amplification power curve) was calculated, and the relative gene expression was calculated using the  $2^{-\Delta\Delta Ct}$  method [ $\Delta Ct = Ct(\text{target gene}) - Ct(\beta\text{-actin})$ ,  $\Delta\Delta Ct = \Delta Ct(\text{target gene}) - \Delta Ct(\beta\text{-actin})$ ]. Five samples of each group were randomly chosen for RT-PCR. The protein expression levels of LCN2 in the normal, acute colitis, chronic colitis, and CAC mice were detected by Western blot. The Anti-LCN2 antibody (ab216462, 1:1,000) and the Anti-GAPDH antibody (ab9485, 1:2,500) were purchased from Abcam. The target protein bands were developed by using enhanced chemiluminescence Western blotting detection reagent, and densities were quantified using Total Lab Quant V11.5 (Newcastle upon Tyne, UK). Three samples of each group were randomly chosen for Western blotting analysis.

### Verification on the Human Colonic Mucosa Immunofluorescence Staining

The intestinal mucosal pathology reports of patients who have undergone colonoscopy in the Department of Pathology,

Dongfang Hospital, Beijing University of Chinese Medicine, including 6 cases of long-duration UC, 6 cases of short-duration UC, 6 cases of normal, and 2 cases of UCAC were obtained. As previous description, the protein expression of LCN2 was then detected by immunofluorescence method.

### Human Protein Atlas Analyses

The Human Protein Atlas (HPA) database (<https://www.proteinatlas.org>) was used for identifying the expression level of LCN2 protein in CRC and normal through immunohistochemistry (IHC) staining. In addition, IHC images were downloaded from HPA, which is an online tool that contains IHC test results, showing the distribution and expression of proteins in various human normal and tumor tissues (24).

### Statistical Analysis

Statistical analyses were performed by RStudio software and the R software (version 3.6.1), SPSS version 25.0 (SPSS Inc., USA), and GraphPad Prism 8.0 (GraphPad Inc., USA). Mean  $\pm$  standard deviation (SD) was used to describe the continuous variables of normal distribution. Median and quartile were used to describe the continuous variable of skewed distribution. One-way analysis of variance (ANOVA) was applied to analyze the data. The least significant difference (LSD) test was performed under the assumption of equal variances; otherwise, Dunnett's T3 test was applied for data with unequal variances. Statistical significance of the difference was set at  $P < 0.05$ .

## RESULTS

### Dataset and Preprocessing

The study finally screened out six datasets, including GSE114603, GSE38713, GSE47908, GSE37283, GSE94648, and GSE11223, and among these, the clinical information of GSE47908 and GSE114603 was provided by the author. The information of each dataset was extracted separately (Table 1). After preprocessing the UC datasets, the data and the eliminating outliers, a total of 17,526 genes were identified as the gene list. In addition, hierarchical cluster analysis (R: hclust function) was used to eliminate outliers and reduced the samples to 39 normal, 40 short-duration UC, 26 long-duration UC, and 15 UCAC samples.

**TABLE 1 |** Characteristics of studies composing the dataset gene expression.

Dataset	Healthy controls	Short-duration UC (<10 years)	Long-duration UC ( $\geq$ 10 years)	UC neoplastic/cancer	Tissues	Platform	Country	Year
GSE38713	13	10	10	NA	Colorectal	GPL570	Spain	2012
GSE114603	6	9	4	NA	Colorectal	GPL14951	Spain	2018
GSE47908	15	21	12	5	Colorectal	GPL570	Denmark	2014
GSE37283	5	NA	NA	10	Colorectal	GPL13158	USA	2012
GSE94648	14	8	14	NA	Blood	GPL19109	Spain	2017
GSE11223	63	28	25	NA	Colorectal	GPL1708	USA	2008

UC, ulcerative colitis; NA, not available.

## Construction of Weighted Gene Co-Expression Modules

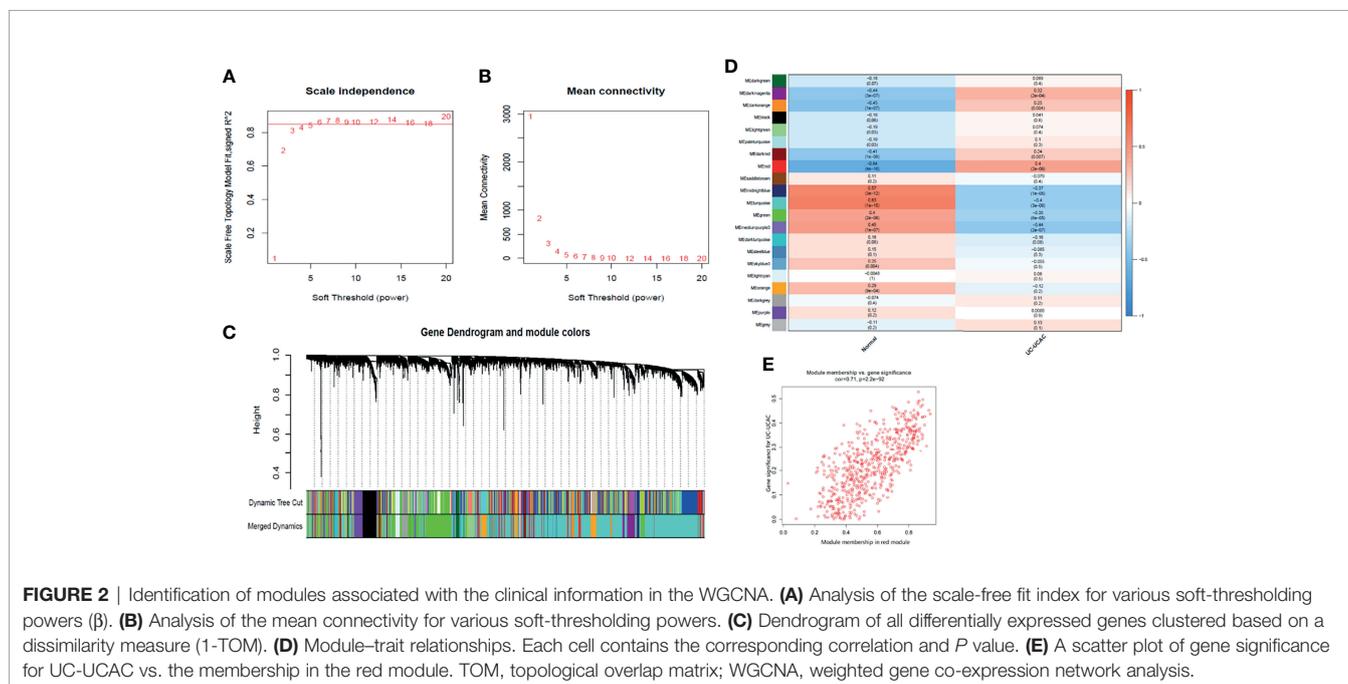
To find the functional clusters related to the carcinogenesis of UC patients, the WGCNA software package was used to construct a gene co-expression network from the processed UC datasets. Before constructing the weighted co-expression network, an appropriate soft threshold  $\beta$  was chosen. After calculation, the soft threshold  $\beta = 6$  was chosen to construct the gene module when the correlation coefficient was close to 0.85 (Figures 2A, B). Following the determination of the soft threshold, the correlation matrix and adjacency matrix of the gene expression profile of UC samples were calculated and subsequently converted into TOM in the light of the WGCNA algorithm. The standard for merging similar modules was to set MEDissThres to 0.3, and 21 modules were generated and assigned the unique color (Figure 2C, excluding gray modules that are not assigned to any cluster). After that, the heat map of the module–trait relationship was drawn to assess the association between each module and the clinical characteristics (normal and UC-UCAC). Among the clinical traits, the trait of UC-UCAC group included different traits such as normal, short duration of UC, long duration of UC, and UCAC, while the trait of the normal group was divided into normal and patients. The results showed that the red module had the highest correlation with the evolution of UC to UCAC ( $r = 0.4$ ,  $P = 2e-06$ ) (Figure 2D). Subsequently, an intramodular analysis of gene significance (GS) and module membership (MM) of the genes in the red module was performed. A high correlation coefficient of GS and MM was found in the red module ( $cor = 0.71$ ,  $P = 2.2e-92$ ) (Figure 2E). The hub genes in the module were screened out according to gene significance of  $>0.2$  and module membership of  $>0.8$ , among which 96 genes are obtained from the red module.

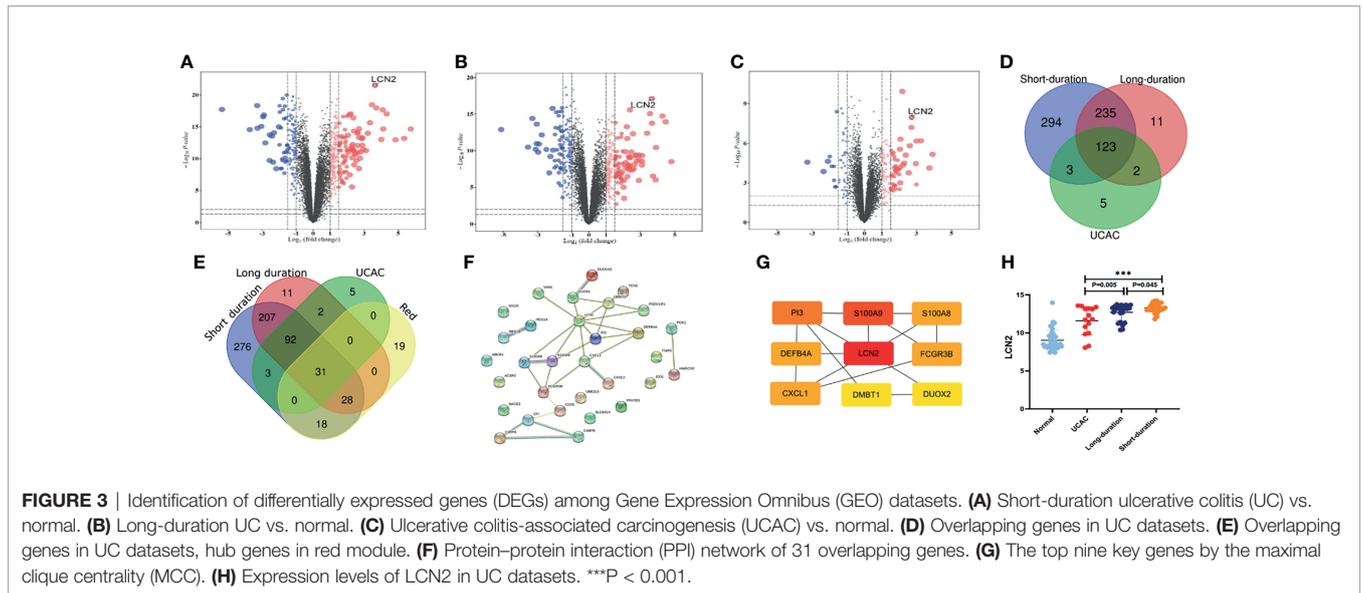
## Identification of Overlapping Genes Between UC Datasets and Co-Expression Module

The genes with the cutoff criteria of  $|\logFC| \geq 1$  and adjusted  $P$  value of  $<0.05$  were regarded as DEGs. Compared with normal tissues, 655 DEGs in short-duration UC (Figure 3A), 371 DEGs in long-duration UC (Figure 3B), and 133 DEGs in UCAC (Figure 3C) were found by using the LIMMA software package. There were 123 overlapping genes among the three parts (Figure 3D). Subsequently, the intersection of the overlapping genes in UC datasets and hub genes in red module was obtained. A total of 31 overlapping genes were extracted for verification (Figure 3E, Supplementary Table S1).

## Protein–Protein Interaction Network and Functional Enrichment Analysis

To further explore the potential functions of the 31 overlapping genes, STRING database was applied to build a PPI network of overlapping genes, with a total of 31 nodes and 29 edges (Figure 3F). The CytoHubba plugin MCC algorithm was used to select the top nine key genes (LCN2, S100A9, P13, S100A8, FCGR3B, CXCL1, DEFB4A, DUOX2, DMBT1) from the PPI network (Figure 3G), and LCN2 ranked first in 9 of 12 methods in cytoHubba plugin (Supplementary Table S2). The expression levels of overlapping genes in different disease durations of UC, UCAC, and normal tissues were proven, and LCN2 was obtained ( $P < 0.01$ ) among the nine key genes with the highest risk of carcinogenesis (Figure 3H). Subsequently, the Metascape database was used for gene enrichment analysis of 31 overlapping genes. After screening out by GO enrichment analysis, several groups of enriched genes were presented in Figure 4. The biological processes (BP) of the 31 genes were mainly from humoral immune response, response to bacterium,



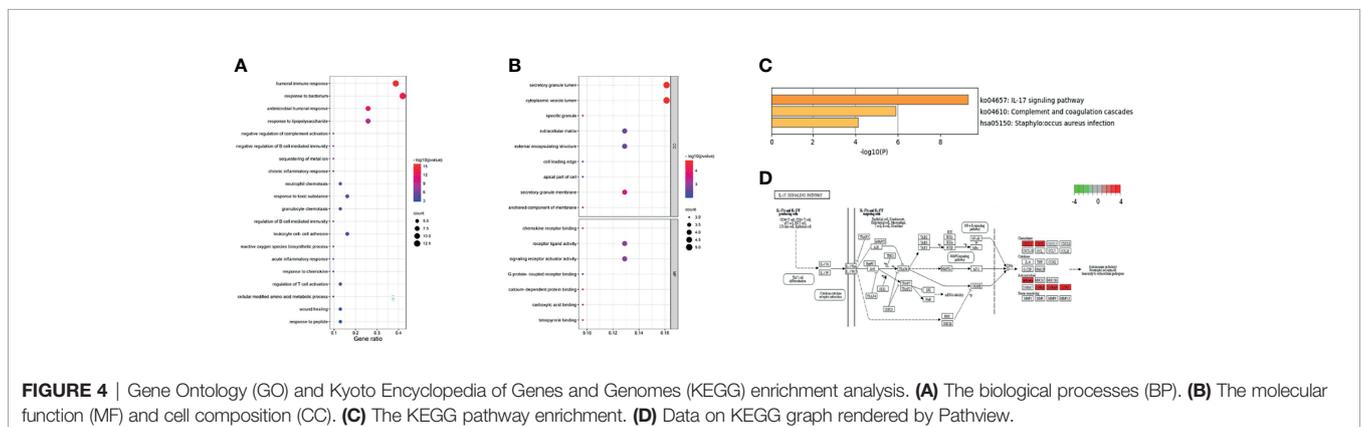


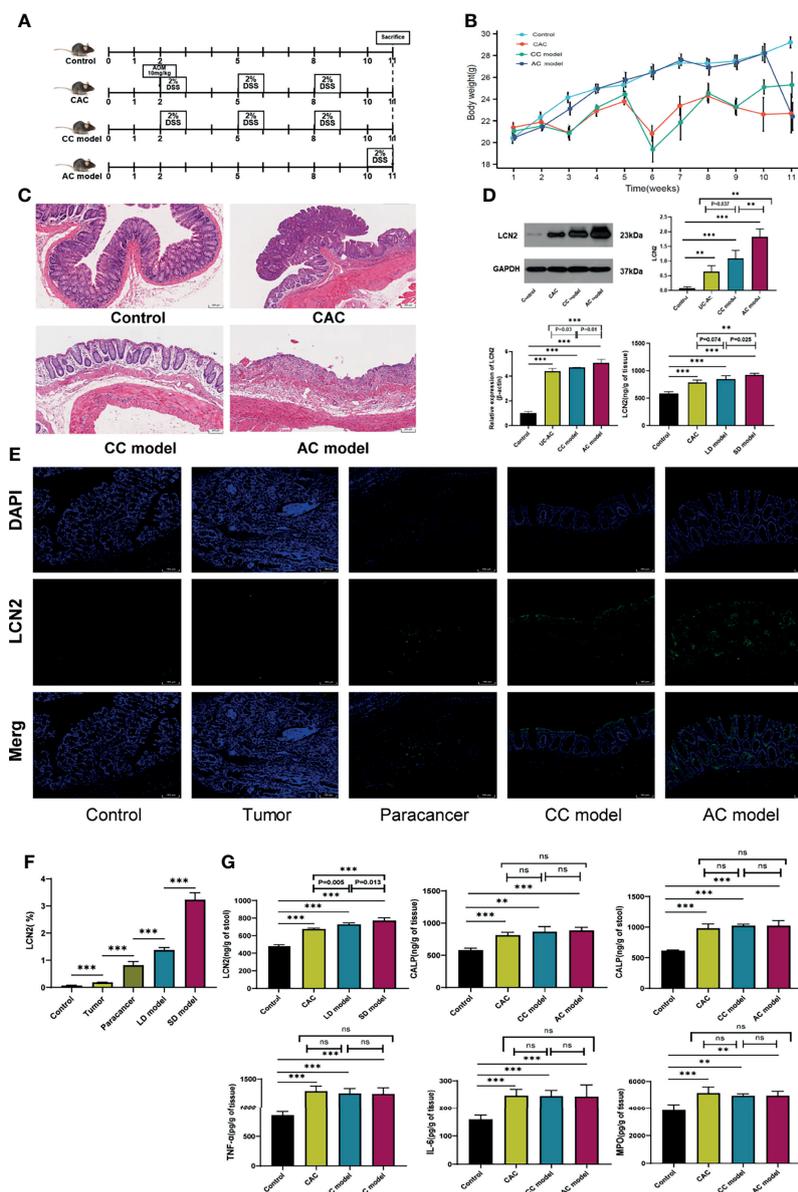
negative regulation of complement activation, and sequestering of metal ion, as well as response to toxic substance and reactive oxygen species biosynthetic process (Figure 4A, Supplementary Table S3). The molecular function (MF) was mainly enriched in receptor ligand activity and signaling receptor activator activity, and the cellular composition (CC) was mainly involved in specific granule lumen, secretory granule lumen, and cytoplasmic vesicle lumen (Figure 4B, Supplementary Table S4). KEGG pathway analysis indicated that the genes might influence the occurrence and progression of UCAC by participating in pathways such as IL-17 signaling pathway and complement and coagulation cascades (Figure 4C), and six genes were enriched in IL-17 signaling pathway (Figure 4D).

### Expression Verification of LCN2 in Animals

To explore the role of LCN2 in the duration of UC disease, the models (acute colitis, chronic colitis, CAC) were built (Figure 5A). As shown in Figure 5B, the significant body weight loss was observed in three modeling groups when compared with the control group. Histologic evaluation showed colon crypts and mucosal damage in acute colitis,

crypt atrophy and abnormal epithelial in chronic colitis, and mucosal structure disorder in CAC (Figure 5C). As demonstrated in RT-PCR and Western blot, the expression levels of LCN2 in acute colitis, chronic colitis, and CAC were all higher than those of normal ( $P < 0.001$ ), acute colitis was higher than that of the chronic colitis ( $P \leq 0.01$ ), and chronic colitis was higher than that of CAC ( $P < 0.05$ ) (Figure 5D). Meanwhile, immunofluorescence staining confirmed a consistent expression trend (Figures 5E, F). Levels of LCN2 and CALP in colonic mucosa and stool were markedly elevated in AOM/DSS- and DSS-induced mice relative to that in the control group ( $P < 0.001$ ). Moreover, we observed that LCN2 expression of chronic colitis could dramatically decrease compared with that of acute colitis ( $P < 0.05$ ). Indeed, we did not make a significant difference of LCN2 expression ( $P = 0.074$ ) between chronic colitis and CAC in the tissue of animal model. However, a downward trend could still be observed between chronic colitis and CAC. To a large extent, chronic colitis and CAC model are consistent in the same modeling time span, it is more likely to result in non-significant differences of LCN2 expression. There was no difference in the expression of





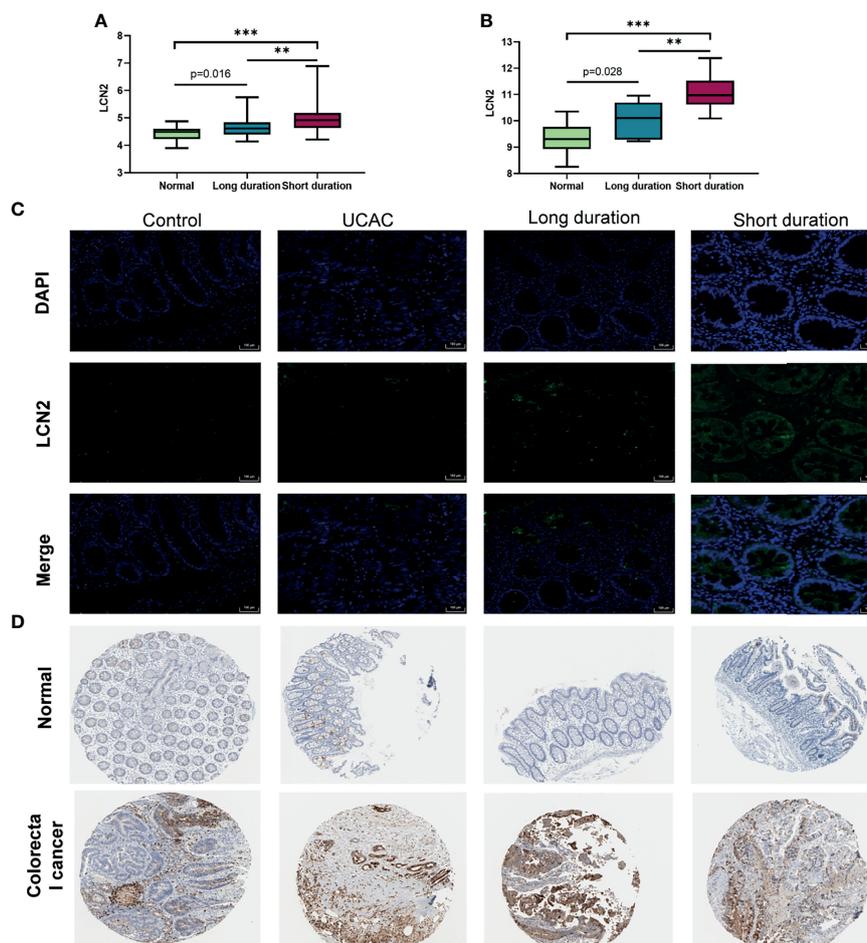
**FIGURE 5** | Protein expression verification in animals. **(A)** Schematic overview of control, acute colitis (AC), chronic colitis (CC) model, and colitis-associated cancer (CAC) model. **(B)** Weight loss in mouse of control and three modeling groups. **(C)** H&E staining of colon sections of mouse at week 11. **(D)** Lipocalin-2 (LCN2) expression level in colorectal mucosa was detected by RT-PCR, Western blot, and ELISA. **(E)** Immunofluorescence of LCN2 in control, tumor, paracancer, long-duration (LD) ulcerative colitis (UC) model, and short-duration (SD) UC model. **(F)** Semiquantitative analysis of immunofluorescence images of LCN2. **(G)** Calprotectin (CALP), myeloperoxidase (MPO), interleukin (IL)-6, and tumor necrosis factor (TNF)- $\alpha$  expression levels in colorectal mucosa and LCN2 in stool were detected by ELISA. \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ; ns, not significant.

IL6, TNF- $\alpha$ , MPO, and CALP in the disease course progresses (Figure 5G).

### Protein Expression Verification of LCN2 in Human

The expression levels of LCN2 among the long-duration UC, short-duration UC, and normal tissues in colorectal samples (GSE11223) and blood samples (GEO94648) were further investigated and shown in Figures 6A, B. The LCN2

immunofluorescence of intestinal mucosa revealed that the LCN2 of short-duration UC was significantly higher than that of the long-duration UC, that of the long-duration UC was higher than that of the UCAC, and those of the short-duration UC and long-duration UC were both higher than that of the normal (Figure 6C). However, there was no difference of LCN2 expression level between UCAC and normal tissues. The cause might be largely explained by the inclusion of a small sample size. At the same time, based on the HPA database, the protein levels



**FIGURE 6** | Protein expression verification in human. **(A)** Expression levels of lipocalin-2 (LCN2) in colorectal mucosa of GSE11223. **(B)** Expression levels of LCN2 in blood sample of GSE94648. **(C)** Immunofluorescence of LCN2 in normal, short-duration ulcerative colitis (UC), long-duration UC, and ulcerative colitis-associated carcinogenesis (UCAC). **(D)** LCN2 in normal colon tissues and colorectal cancer (CRC) from the Human Protein Atlas (HPA) database. \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

of LCN2 gene in tumor tissues were found to be significantly higher than that in normal tissues (**Figure 6D**).

## DISCUSSION

LCN2, also referred to as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin, is a peptide produced by macrophages, neutrophil granulocytes, and other immune and parenchymal cells. It has an antimicrobial function by binding catechol-type siderophores with an extremely high affinity (25). Moreover, it is linked to several physiological roles, including transporting hydrophobic ligands across cell membranes, modulating immune responses, maintaining iron homeostasis, and promoting epithelial cell differentiation (25–28). So, LCN2 is a chemically highly stable protein that makes it as an interesting biomarker candidate (29). Stallhofer et al. (30) have confirmed the excessive *de novo* synthesis of LCN2 is localized in human

colonic epithelial cells, and serum LCN2 has the potential to be a biomarker of active UC. Fecal LCN2 could be used as a biomarker for detecting endoscopic activity in patients with inflammatory bowel diseases. In deep remission, expressions of LCN2 but not CALP remained elevated in the rectum mucosa of UC patients (31). While our study was performed, some reports on the role of LCN2 in other colitis were published except UC, including collagenous colitis and necrotizing enterocolitis. LCN2 was also identified as a potential fecal biomarker (32, 33). Previous review displays that expression of LCN2 in CAC is variable (34). Some *in vitro* and *in vivo* studies in the literature support the role of LCN2 as a tumor promoter (35), but the difference is that in most of the current studies, LCN2 is generally considered to be a protective factor in CRC (36–38). LCN2 expression was significantly lower in metastatic or advanced-stage CRC than that in non-metastatic or early-stage CRC (39). Therefore, there is a clear and urgent need to confirm changes of LCN2 expression in the process of carcinogenesis from UC, so as

to develop diagnostic tools to better monitor the progression of disease.

UC patients with long disease duration have a higher risk of developing CAC compared with patients with short-duration UC. LCN2 has already been proven to be a major downregulated differential gene associated with the duration of UC disease by Low et al. (8), and this was consistent with our study results. On the basis of that, our study compared samples of healthy people vs. samples from patients with the disease in databases, including short-duration UC, long-duration UC, and UCAC. The expression level of LCN2 was significantly higher in the diseased tissues than in the normal tissues. Noteworthy, the expression of LCN2 is significantly decreased in UCAC when compared to long-duration UC, which correlates to the disease duration of UC. We identify that LCN2 is associated with the risk of carcinogenesis based on disease duration among the hub genes. Collectively, from the perspective of BP, CC, and KEGG analyses, LCN2 showed close relation to antibacterial response and immune activation. Based on the previously reported involvement of LCN2 in chemokine induction and in the recruitment of neutrophils at the sites of infection or tissue injury (40). LCN2 could enhance Th17 type responses and cell recruitment (41).

Given the nature of our study, the average ages of healthy people and patients with short-duration UC, long-duration UC, and UCAC were  $43 \pm 13$  years,  $36 \pm 13$  years,  $47 \pm 11$  years, and  $50 \pm 8$  years, respectively. Difference in age might be inevitable. Li et al. (42) have suggested that the molecular pathways in diseased tissue samples are similar between adult and pediatric UC patients. However, there is yet to prove the direct effects of age on transcriptomic studies. Another point to be noted is that verification of proteomics sample obtained in the study was from a single center. A larger population study group is warranted to make any definitive conclusions in this matter. Additionally, a stronger effect of anti-TNF drugs induced a decrease in LCN2 (43), and therefore, the patients who have received treatment were excluded in the study, and only the baseline state was used.

Long-duration UC acts as a risk factor for the development of UCAC, and the transcriptomic at different disease duration intervals can aid in determining the progress, along with colonoscopy. The expression levels of LCN2 in disease groups are increased when compared with the normal group, but the expression levels of LCN2 are gradually reduced in tissues of short-duration UC, long-duration UC, and UCAC. The condition also applies to the serum LCN2, and the expression trend of LCN2 in feces shows consistent results with that of blood and tissue. Because blood and feces are easier to detect than tissues, LCN2 expression is more convenient for disease surveillance. In addition, for the monitoring of colon cancer related with course, fecal LCN2 has more advantages than fecal CALP in our experimental verification. However, there are a few limitations in this study. Firstly, the results are only based on the current database of known gene sequences and transcripts. Because the experiments are done in different centers, this kind of missing information are inevitable in public databases. No research on children with regard to disease duration and

prognosis exists. Then, the chronic UC is similar to that of cancer modeling in the animal experiment; it should be further verified in clinical samples. Finally, the trend of LCN2 expression in disease tissues is clear basically, but corresponding threshold of LCN2 in blood and feces is a significant issue worthy of continued investigation. Noteworthy, other carcinomas, neuropathy, and metabolic diseases (44–46) have a greater impact on the level of LCN2 in the blood, and the detection of stool may be more beneficial for diagnosis value of UCAC in subsequent use.

## CONCLUSIONS

In conclusion, we analyzed transcriptomic changes in patients with UC of different disease durations developing into UCAC to further identify the promising diagnostic biomarkers. We showed that the expression level of LCN2 was significantly negatively correlated with the risk of UCAC. The expression level of LCN2 in short-duration UC was higher than that of long-duration UC, that of long-duration UC was higher than that of UCAC, and those of UC and UCAC were all higher than that of the normal. We then validated this trend in human proteomics and animal experiment. In addition, we also discovered that the expression trend of LCN2 in blood and stool samples was consistent with that in colorectal mucosa. However, LCN2 expression in blood sample was easily affected by a multitude of factors. Stool detection may have more advantages for diagnosis value of early-stage UCAC as a complement to colonoscopy surveillance.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. The animal study was reviewed and approved by the Animal Ethics Committee of Beijing University of Chinese Medicine.

## AUTHOR CONTRIBUTIONS

FSK and YC contributed equally to the study as first authors. YC and FSK: study concept and design. FSK and JLL: acquisition of data. FSK and LS: analysis of data. YC: article writing. YC, JLL, YYL, RS, GYP, and JXL: interpretation of data and critical

revision of the article. JXL consulted fees or honoraria from. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Key R&D Program of China [2018YFC1705403, 2018YFC1705405] and Innovation “One Hundred Million” Talent Project Qihuang Scholar.

## REFERENCES

- Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide Incidence and Prevalence of Inflammatory Bowel Disease in the 21st Century: A Systematic Review of Population-Based Studies. *Lancet* (2018) 390:2769–78. doi: 10.1016/S0140-6736(17)32448-0
- de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-Wide Association Study Implicates Immune Activation of Multiple Integrin Genes in Inflammatory Bowel Disease. *Nat Genet* (2017) 49:256–61. doi: 10.1038/ng.3760
- Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG Clinical Guideline: Ulcerative Colitis in Adults. *Am J Gastroenterol* (2019) 114(3):384–413. doi: 10.14309/ajg.000000000000152
- Kobayashi T, Siegmund B, Le Berre C, Wei SC, Ferrante M, Shen B, et al. Ulcerative Colitis. *Nat Rev Dis Primers* (2020) 6(1):74. doi: 10.1038/s41572-020-0205-x
- Planell N, Lozano JJ, Mora-Buch R, Masamunt MC, Jimeno M, Ordás I, et al. Transcriptional Analysis of the Intestinal Mucosa of Patients With Ulcerative Colitis in Remission Reveals Lasting Epithelial Cell Alterations. *Gut* (2013) 62(7):967–76. doi: 10.1136/gutjnl-2012-303333
- Xu L, Ma L, Lian J, Yang J, Chen S. Gene Expression Alterations in Inflamed and Unaffected Colon Mucosa From Patients With Mild Inflammatory Bowel Disease. *Mol Med Rep* (2016) 13:2729–35. doi: 10.3892/mmr.2016.4880
- Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulssen RH. Transcriptomic Landscape of Treatment-Naïve Ulcerative Colitis. *J Crohns Colitis* (2018) 12:327–36. doi: 10.1093/ecco-jcc/jjx139
- Low END, Mokhtar NM, Wong Z, Raja Ali RA. Colonic Mucosal Transcriptomic Changes in Patients With Long-Duration Ulcerative Colitis Revealed Colitis-Associated Cancer Pathways. *J Crohns Colitis* (2019) 13(6):755–63. doi: 10.1093/ecco-jcc/jjz002
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* (2003) 4(2):249–64. doi: 10.1093/biostatistics/4.2.249
- Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med* (2016) 13(12):e1002194. doi: 10.1371/journal.pmed.1002194
- Lorén V, Garcia-Jaraquemada A, Naves JE, Carmona X, Mañosa M, Aransay AM, et al. ANP32E, a Protein Involved in Steroid-Refractoriness in Ulcerative Colitis, Identified by a Systems Biology Approach. *J Crohns Colitis* (2019) 13(3):351–61. doi: 10.1093/ecco-jcc/jjy171
- Bjerrum JT, Nielsen OH, Riis LB, Pittet V, Mueller C, Rogler G, et al. Transcriptional Analysis of Left-Sided Colitis, Pancolitis, and Ulcerative Colitis-Associated Dysplasia. *Inflamm Bowel Dis* (2014) 20(12):2340–52. doi: 10.1097/MIB.0000000000000235
- Pekow J, Dougherty U, Huang Y, Gometz E, Nathanson J, Cohen G, et al. Gene Signature Distinguishes Patients With Chronic Ulcerative Colitis Harboring Remote Neoplastic Lesions. *Inflamm Bowel Dis* (2013) 19(3):461–70. doi: 10.1097/MIB.0b013e3182802bac
- Leek JT. SvaSeq: Removing Batch Effects and Other Unwanted Noise From Sequencing Data. *Nucleic Acids Res* (2014) 42(21):e161. doi: 10.1093/nar/gku864

## ACKNOWLEDGMENTS

We would like to thank the GEO, HPA, and Metascape databases for providing open access.

## SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.793760/full#supplementary-material>

- Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Yan S, Wang W, Gao G, Cheng M, Wang X, Wang Z, et al. Key Genes and Functional Coexpression Modules Involved in the Pathogenesis of Systemic Lupus Erythematosus. *J Cell Physiol* (2018) 233:8815–25. doi: 10.1002/jcp.26795
- Wang CCN, Li CY, Cai JH, Sheu PC, Tsai JJP, Wu MY, et al. Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *J Clin Med* (2019) 8(8):1160. doi: 10.3390/jcm8081160
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING V11: Protein-Protein Association Networks With Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res* (2019) 47:D607–13. doi: 10.1093/nar/gky1131
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6
- Gene Ontology Consortium. The Gene Ontology (GO) Project in 2006. *Nucleic Acids Res* (2006) 34:D322–6. doi: 10.1093/nar/gkj021
- Planell N, Masamunt MC, Leal RF, Rodríguez L, Esteller M, Lozano JJ, et al. Usefulness of Transcriptional Blood Biomarkers as a Non-Invasive Surrogate Marker of Mucosal Healing and Endoscopic Response in Ulcerative Colitis. *J Crohns Colitis* (2017) 11(11):1335–46. doi: 10.1093/ecco-jcc/jjx091
- Noble CL, Abbas AR, Cornelius J, Lees CW, Ho GT, Toy K, et al. Regional Variation in Gene Expression in the Healthy Colon is Dysregulated in Ulcerative Colitis. *Gut* (2008) 57(10):1398–405. doi: 10.1136/gut.2008.148395
- Wirtz S, Popp V, Kindermann M, Gerlach K, Weigmann B, Fichtner-Feigl S, et al. Chemically Induced Mouse Models of Acute and Chronic Intestinal Inflammation. *Nat Protoc* (2017) 12(7):1295–309. doi: 10.1038/nprot.2017.044
- Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A Pathology Atlas of the Human Cancer Transcriptome. *Science* (2017) 357(6352):eaa2507. doi: 10.1126/science.aan2507
- Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK. The Neutrophil Lipocalin NGAL is a Bacteriostatic Agent That Interferes With Siderophore-Mediated Iron Acquisition. *Mol Cell* (2002) 10(5):1033–43. doi: 10.1016/s1097-2765(02)00708-6
- Kurzrock R, Hickish T, Wyrwicz L, Saunders M, Wu Q, Stecher M, et al. Interleukin-1 Receptor Antagonist Levels Predict Favorable Outcome After Bermekimab, a First-in-Class True Human Interleukin-1 $\alpha$  Antibody, in a Phase III Randomized Study of Advanced Colorectal Cancer. *Oncimmunology* (2018) 8(3):1551651. doi: 10.1080/2162402X.2018.1551651
- Holgerson K, Kutlu B, Fox B, Serikawa K, Lord J, Hansen AK, et al. High-Resolution Gene Expression Profiling Using RNA Sequencing in Patients With Inflammatory Bowel Disease and in Mouse Models of Colitis. *J Crohns Colitis* (2015) 9:492–506. doi: 10.1093/ecco-jcc/jjv050
- Moniruzzaman M, Wang R, Jeet V, McGuckin MA, Hasnain SZ. Interleukin (IL)-22 From IL-20 Subfamily of Cytokines Induces Colonic Epithelial Cell Proliferation Predominantly Through ERK1/2 Pathway. *Int J Mol Sci* (2019) 20(14):3468. doi: 10.3390/ijms20143468
- Thorsvik S, van Beelen Granlund A, Svendsen TD, Bakke I, Røyset ES, Flo TH, et al. Ulcer-Associated Cell Lineage Expresses Genes Involved in Regeneration

- and is Hallmarked by High Neutrophil Gelatinase-Associated Lipocalin (NGAL) Levels. *J Pathol* (2019) 248(3):316–25. doi: 10.1002/path.5258
30. Stallhofer J, Friedrich M, Konrad-Zerna A, Wetzke M, Lohse P, Glas J, et al. Lipocalin-2 Is a Disease Activity Marker in Inflammatory Bowel Disease Regulated by IL-17a, IL-22, and TNF- $\alpha$  and Modulated by IL23R Genotype Status. *Inflamm Bowel Dis* (2015) 21(10):2327–40. doi: 10.1097/MIB.0000000000000515
  31. Zollner A, Schmiderer A, Reider SJ, Oberhuber G, Pfister A, Texler B, et al. Fecal Biomarkers in Inflammatory Bowel Diseases: Calprotectin Versus Lipocalin-2 - a Comparative Study. *J Crohns Colitis* (2020) 15(1):43–54. doi: 10.1093/ecco-jcc/jjaa124
  32. Thibault MP, Tremblay É, Horth C, Fournier-Morin A, Grynspan D, Babakissa C, et al. Lipocalin-2 and Calprotectin as Stool Biomarkers for Predicting Necrotizing Enterocolitis in Premature Neonates. *Pediatr Res* (2021). doi: 10.1038/s41390-021-01680-7
  33. Bakke I, Walaas GA, Bruland T, Royset ES, van Beelen Granlund A, Escudero-Hernández C, et al. Mucosal and Faecal Neutrophil Gelatinase-Associated Lipocalin as Potential Biomarkers for Collagenous Colitis. *J Gastroenterol* (2021) 56(10):914–27. doi: 10.1007/s00535-021-01814-y
  34. Buisson A, Vazeille E, Minet-Quinard R, Goutte M, Bouvier D, Goutorbe F, et al. Fecal Matrix Metalloprotease-9 and Lipocalin-2 as Biomarkers in Detecting Endoscopic Activity in Patients With Inflammatory Bowel Diseases. *J Clin Gastroenterol* (2018) 52:e53–62. doi: 10.1097/MCG.0000000000000837
  35. Zhang W, Pan R, Lu M, Zhang Q, Lin Z, Qin Y, et al. Epigenetic Induction of Lipocalin 2 Expression Drives Acquired Resistance to 5-Fluorouracil in Colorectal Cancer Through Integrin  $\beta$ 3/SRC Pathway. *Oncogene* (2021) 40(45):6369–80. doi: 10.1038/s41388-021-02029-4
  36. Feng M, Feng J, Chen W, Wang W, Wu X, Zhang J, et al. Lipocalin2 Suppresses Metastasis of Colorectal Cancer by Attenuating NF- $\kappa$ B-Dependent Activation of Snail and Epithelial Mesenchymal Transition. *Mol Cancer* (2016) 15(1):77. doi: 10.1186/s12943-016-0564-9
  37. Kim SL, Lee ST, Min IS, Park YR, Lee JH, Kim DG, et al. Lipocalin 2 Negatively Regulates Cell Proliferation and Epithelial to Mesenchymal Transition Through Changing Metabolic Gene Expression in Colorectal Cancer. *Cancer Sci* (2017) 108(11):2176–86. doi: 10.1111/cas.13389
  38. Xiao X, Yeoh BS, Vijay-Kumar M. Lipocalin 2: An Emerging Player in Iron Homeostasis and Inflammation. *Annu Rev Nutr* (2017) 37:103–30. doi: 10.1146/annurev-nutr-071816-064559
  39. Santiago-Sánchez GS, Pita-Grisanti V, Quiñones-Díaz B, Gumpfer K, Cruz-Monserrate Z, Vivas-Mejía PE. Biological Functions and Therapeutic Potential of Lipocalin 2 in Cancer. *Int J Mol Sci* (2020) 21(12):4365. doi: 10.3390/ijms21124365
  40. Pai RK, Hartman DJ, Rivers CR, Regueiro M, Schwartz M, Binion DG, et al. Complete Resolution of Mucosal Neutrophils Associates With Improved Long-Term Clinical Outcomes of Patients With Ulcerative Colitis. *Clin Gastroenterol Hepatol* (2020) 18:2510–7.e5. doi: 10.1016/j.cgh.2019.12.011
  41. Hau CS, Kanda N, Tada Y, Shibata S, Uozaki H, Fukusato T, et al. Lipocalin-2 Exacerbates Psoriasisiform Skin Inflammation by Augmenting T-Helper 17 Response. *J Dermatol* (2016) 43(7):785–94. doi: 10.1111/1346-8138.13227
  42. Li K, Strauss R, Ouahed J, Chan D, Telesco SE, Shouval DS, et al. Molecular Comparison of Adult and Pediatric Ulcerative Colitis Indicates Broad Similarity of Molecular Pathways in Disease Tissue. *J Pediatr Gastroenterol Nutr* (2018) 67(1):45–52. doi: 10.1097/MPG.0000000000001898
  43. Milanesi E, Dobre M, Manuc TE, Becheanu G, Tieranu CG, Ionescu EM, et al. Mucosal Gene Expression Changes Induced by Anti-TNF Treatment in Inflammatory Bowel Disease Patients. *Drug Dev Res* (2019) 80(6):831–6. doi: 10.1002/ddr.21566
  44. Lim D, Jeong JH, Song J. Lipocalin 2 Regulates Iron Homeostasis, Neuroinflammation, and Insulin Resistance in the Brains of Patients With Dementia: Evidence From the Current Literature. *CNS Neurosci Ther* (2021) 27(8):883–94. doi: 10.1111/cns.13653
  45. Bhusal A, Lee WH, Suk K. Lipocalin-2 in Diabetic Complications of the Nervous System: Physiology, Pathology, and Beyond. *Front Physiol* (2021) 12:638112. doi: 10.3389/fphys.2021.638112
  46. Jaber SA, Cohen A, D'Souza C, Abdulrazzaq YM, Ojha S, Bastaki S, et al. Lipocalin-2: Structure, Function, Distribution and Role in Metabolic Disorders. *BioMed Pharmacother* (2021) 142:112002. doi: 10.1016/j.biopha.2021.112002

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kou, Cheng, Shi, Liu, Liu, Shi, Peng and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.