

Functional and protein-protein interaction network analysis of colorectal cancer induced by ulcerative colitis

YONG DAI, JIN-BO JIANG, YAN-LEI WANG, ZU-TAO JIN and SAN-YUAN HU

Department of General Surgery, Qilu Hospital of Shandong University, Jinan, Shandong 250012, P.R. China

Received September 28, 2014; Accepted June 17, 2015

DOI: 10.3892/mmr.2015.4102

Abstract. Colorectal cancer (CRC) is a well-recognized complication of ulcerative colitis (UC), and patients with UC have a higher incidence of CRC, compared with the general population. However, the properties of CRC induced by UC have not been clarified using an interaction network to analyze and compare gene sets. In the present study, six microarray datasets of CRC and UC were extracted from the Array Express database, and gene signatures were identified using the genome-wide relative significance (GWRS) method. Functional analysis was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Prediction of the genes and microRNA were performed using a hypergeometric method. A protein-protein interaction (PPI) network was constructed using the Search Tool for the Retrieval of Interacting Genes/proteins, and clusters were obtained through the Molecular Complex Detection algorithm. Topological centrality and a novel analyzing method, based on the rank value of GWGS, were used to characterize the biological importance of the clusters. A total of 217 differentially expressed (DE) genes of CRC were identified, 341 DE genes were identified in UC, and 62 common genes existed in the two. Several KEGG pathways were the same in CRC and UC. Collagenase, progesterone, heparin, urokinase, nadh and adenosine drugs demonstrated potential for use in treatment of CRC and UC. In the PPI network of CRC, 210 nodes and 752 edges were observed, whereas 314 nodes and 882 edges were identified in UC. Cluster 3 in UC had the highest GWGS, while the topological centrality of Cluster 3 in UC had the lowest degree and betweenness. PPI network analysis provided an effective way to estimate and understand the likelihood of the potential connections between proteins/genes. The results obtained following the use of GWGS to analyze differences between clusters did not agree with the topological degree and betweenness centrality, which indicated that gene fold change

based GWGS was controversial with degree here in CRC and UC.

Introduction

There is convincing evidence from previous studies that patients with ulcerative colitis (UC) have a higher incidence of colorectal cancer (CRC), compared with the general population (1). The increased incidence occurs predominantly in patients with long-standing extensive colitis (2). Although CRC induced by UC only accounts for 1% of all cases of CRC in the general population, it is a serious sequel of the disease and accounts for one sixth of the mortality rate in patients with UC in Asia (3).

Multiple existing genome association approaches have been suggested to account for the mechanism of CRC (4,5), particularly its induction by UC, by identifying the independent effects of individual genes (6). Suzuki *et al* identified a group of genes, which were preferentially hypermethylated in CRC, including *SFRP1* (7). In a genome-scale analysis, 16% of colorectal carcinomas were found to be hypermutated and, excluding the hypermutated types of cancer, colon and rectal types of cancer had considerably similar patterns of genomic alteration (8). However, investigations focussing on the effects of individual gene has omitted genes, which are not only encoded as individual genes or proteins, but also as subnetworks of interacting proteins within a larger human protein-protein interaction (PPI) network in the human genome (9). As a result, several mechanisms of human disease, including CRC remain to be elucidated.

The availability of large protein networks provides one method to, at least partially, address the challenges mentioned above. Since large protein networks are available for humans (10), a number of approaches have been demonstrated for extracting relevant functional pathways, based on the relevant databases (11). Following the measurement of sufficient protein interaction data, a large number of distinct functional pathways can be identified, which enable novel opportunities for elucidating the pathways involved in major diseases and pathologies (10,12). Investigations account for properties in interaction networks, and it has been reported that clustering with overlapping neighborhood expansion can be used as a method for detecting potentially overlapping protein complexes from a PPI network (13).

Network enrichment and topological analysis identifies the target gene set within its interaction environment and identifies

Correspondence to: Mr. San-Yuan Hu, Department of General Surgery, Qilu Hospital of Shandong University, 107 Cultural West Road, Jinan, Shandong 250012, P.R. China
E-mail: sanyuanhu@yeah.net

Key words: colorectal cancer, ulcerative colitis, protein-protein interaction, genome-wide global significance, topological centrality

possible gene cofactors and topologically associated pathways and processes (14). Several groups have suggested a more effective method of combining gene expression measurements in groups of genes that fall within certain pathways. Several approaches have been suggested to score known pathways or sub-networks on the coherency of expression changes among their member genes. For example, Chuang *et al* identified the markers of metastasis within gene expression profiles (15), which involved the identification of gene alterations and prediction of the likelihood of metastasis in unknown samples using a protein-network-based approach. Pržulj *et al* performed a systematic graph theory-based analysis of this PPI network to construct computational models for describing and predicting the properties of life-threatening mutations and proteins involved in genetic interactions, functional groups, protein complexes and signaling pathways (16). However, few investigations combining gene expression and network properties for measurements of groups of genes that fall within pathways and sub-networks have been performed.

The aim of the present study was to determine the formation mechanism of CRC induced by UC, using a combination of methods for the measurement of gene expression (genome-wide global significance; GWGS) and centralities. The analysis pipeline included analysis of differentially expressed (DE) genes, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment, construction of PPI networks, module detection, measurements of topological factors, determination of GWGS values, and predictions of drug genes and miRNA target genes.

Materials and methods

Identification of gene expression datasets. A total of six datasets, including E-GEOD-6731 (17), E-GEOD-36807, E-GEOD-38713 (18), E-GEOD-41258 (19), E-GEOD-4183 (20), and E-MTAB-57 (21), were extracted from the Array Express database (<http://www.ebi.ac.uk/arrayexpress/>). For UC, the E-GEOD-6731 dataset consisted of four normal controls and nine patients; the E-GEOD-36807 dataset consisted of seven normal controls and 15 patients; the E-GEOD-38713 dataset consisted of 13 normal controls and 30 patients; the E-GEOD-41258 CRC dataset consisted of 100 normal controls and 290 patients; the E-GEOD-4183 data consisted of 18 normal controls and 35 patients; and the E-MTAB-57 dataset consisted of 22 normal controls and 25 patients.

Integrated analysis of DE genes. The fold-change (FC) based on the model was used in the present study, as our computational evaluation aimed to identify the changes of gene expression. For each gene in the list of unique genes, a rank number was assigned, in descending order between 1 and *m*, according to their corresponding degree of differential expression. The present study then measured the GWRS of *i*-th gene in the *j*-th dataset, using the following equation (1,22):

$$S_{ij} = -2 \log \left(\frac{r_{ij}}{m} \right)$$

The number of datasets was denoted by *n*, the number of unique genes across *n* datasets was denoted by *m*; where r_{ij} , $i=1-m, j=1-n$, indicate the rank number of the *i*-th gene in the *j*-th study. The range of GWRS values (s_{ij}) was between 0 and $-2 \log (1/m)$.

The GWGS of a gene was estimated based on its corresponding GWRs across the *n* datasets using the following equation (2):

$$S_j^r = \sum_{j=1}^n \omega_j S_{ij}$$

ω_j represents the relative weight of the *j*-th dataset. The value of the weight was assigned based on the data quality of the *j*-th datasets, the value of ω_j is used to reflect the differential importance of biopsy, vs. cell line samples which may be taken into account. The present study assigned equal weights to all data. In addition, the P-values for all genes were recorded following analysis using the Linear Models for Microarray Data (Limma) 3.20.8 package, subsequent to robust multiarray average (RMA) (23) and preprocessing (24). The genes with $\log_2 \text{FC} > 2$ and $P < 0.01$ were selected as DE genes for further investigation. The DE genes were selected if the gene was identified as a DE gene in at least two datasets in each group (UC or CRC).

Pathway enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database is a knowledge base for the systematic analysis of gene functions, linking genomic information with higher order functional information. In the present study, KEGG pathway enrichment analysis was performed for the identified DE genes using the online tool Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.7 (<http://david.abcc.ncifcrf.gov/>) (25). KEGG pathways with $P < 0.01$ were selected, based on the expression analysis systematic explored (EASE) assessment, implemented in DAVID. The principle of EASE was as follows (3):

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}$$

n is the number of background genes; *a*' is the gene number of one gene set in the gene lists; *a*' + *b* is the number of genes in the gene list, which include at least one gene set; *a*' + *c* is the gene number of one gene list in the background genes; *a* is replaced with $a' = a - 1$.

Predictions of drug genes and miRNA targets. The present study performed drug gene and miRNA target prediction using a Web-based gene set analysis toolkit (WebGestalt; <http://bioinfo.vanderbilt.edu/webgestalt/analysis.php>) (26). If there are *n* genes in the CRC gene set of interest (A), *m* genes in the UC reference gene set (B), and there are *k* genes in CRC and *j* genes in UC in a given category (C). Based on the reference gene set, the expected value of *k* is $ke = (n / m) * j$. If *k* exceeds the above expected value, category C is considered to be enriched, with a ratio of enrichment (*r*) determined by $r = k / ke$. If B represents the population from which the genes in A are obtained, WebGestalt uses the hypergeometric assessment to evaluate the significance of enrichment for category C in gene set A (27), as in following equation (4):

$$P = \sum_{i=k}^n \frac{\binom{n}{i} \binom{m}{j+k-i}}{\binom{m+n}{j+k}}$$

The P-values require adjustment for multiple assessments, which was performed using the Benjamini-Hochberg method (28). Genes with quantities > 5 and $P < 0.01$ were considered significant.

Table I. Genes common to colorectal cancer and ulcerative colitis.

Number	Gene
1	AQP8
2	CXCL5
3	MMP3
4	CHI3L1
5	KIAA1199
6	TMEM158
7	CXCL3
8	MMP1
9	CXCL1
10	ABCA8
11	SPP1
12	PSAT1
13	SLC26A2
14	SLC7A11
15	SLC4A4
16	PHLDA1
17	OLFM4
18	MMP7
19	GUCA2B
20	CWH43
21	LCN2
22	REG1A
23	BGN
24	NFE2L3
25	SULF1
26	PRKACB
27	CHP2
28	PTN
29	TRIM29
30	COL1A1
31	CDH3
32	NR5A2
33	HPGD
34	SLCO4A1
35	NXPE4
36	COL1A2
37	PLAU
38	HMGCS2
39	CFB
40	SERPINB5
41	SPINK4
42	CD55
43	MT1M
44	MMP12
45	SGK2
46	SLC17A4
47	PCK1
48	SORD
49	PADI2
50	TNFRSF12A
51	REG1B

Table I. Continued.

Number	Gene
52	ANK3
53	REG3A
54	EPHX2
55	ABCB1
56	OSBPL1A
57	LOXL2
58	WNT5A
59	ENTPD5
60	COL5A2
61	MMP9

Analysis and construction of the PPI network. For protein interaction data, the present study utilized a human PPI dataset from the Search Tool for the Retrieval of Interacting Genes/proteins (STRING) 9.1 (<http://string.embl.de/>) resource. In addition, the PPI network was constructed using Cytoscape 3.1.0 (29), a free software package for visualizing, modeling and analyzing the integration of bimolecular interaction networks with high-throughput expression data and other molecular states.

Molecular complex detection (MCODE) algorithm. The MCODE algorithm (<http://baderlab.org/Software/MCODE>) was used for subnet analysis of the PPI network. The MCODE algorithm predominantly includes three stages: Vertex weighting, complex prediction and optionally post-processing. At the vertex weighting stage, all vertices, based on their local network density, were weighted using the highest k -core of the vertex neighborhood. At the stage of complex prediction, the vertex-weighted graph was taken as input, a complex with the highest-weighted vertex was seeded, and moved outward from the seed vertex recursively. It owned vertices in the complex whose weight was above a specific threshold, a certain percentage away from the weight of the seed vertex. Complexes with a core < 2 (graph of minimum degree 2) were filtered, the 'fluff' option and 'haircut' option were also run. The 'fluff' option was used to increase the size of the complex, according to a given 'fluff' parameter between 0.0 and 1.0. The 'haircut' option removed vertices, which were connected to the core complex alone, resulting in complexes obtained that were 2-core. When both options were performed, 'fluff' runs followed by haircut, the available network properties were as follows: The degree of a node (gene or protein) was the average number of edges (interactions) incident to this node. The degree quantified the local topology of each gene, by combing the number of its adjacent genes (30). This produced a simple count of the number of interactions of a given node.

The node betweenness, $B(v)$, of a node, v , was calculated from the number of shortest paths (σ_{st}) between nodes s and t going through v (5):

$$C_B(v) = \sum_{s \neq v \neq t \in N} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Table II. Kyoto Encyclopedia of Genes and Genome analysis of differentially expressed genes in colorectal cancer.

Term	Genes	P-value
hsa04510:Focal adhesion	CAV1, CCND1, PAK2, VEGFA, COL1A2, COL1A1, FLNC, COL5A2, THBS2, MYLK, SPP1, MYL9	6.82E-04
hsa05219:Bladder cancer	CCND1, IL8, MMP9, VEGFA, MYC, MMP1	7.49E-04
hsa03320:PPAR signaling pathway	SORBS1, HMGCS2, SCD, FABP4, FABP1, MMP1, PCK1	1.20E-03
hsa04270:Vascular smooth muscle contraction	KCNMA1, EDNRA, ACTG2, PPP1R12B, MYH11, PRKACB, MYLK, MYL9	3.25E-03
hsa00910:Nitrogen metabolism	CA12, CA4, CA2, CA1	7.13E-03
hsa00150:Androgen and estrogen metabolism	UGT2B17, HSD17B2, HSD11B2, UGT2B15	2.62E-02
hsa04060:Cytokine-cytokine receptor interaction	CXCL1, INHBA, IL8, CXCL5, CCL20, TNFRSF12A, CXCL3, CXCL2, VEGFA, CXCL12	3.85E-02
hsa04062:Chemokine signaling pathway	CXCL1, IL8, CXCL5, CCL20, CXCL3, CXCL2, PRKACB, CXCL12	4.44E-02
hsa00140:Steroid hormone biosynthesis	UGT2B17, HSD17B2, HSD11B2, UGT2B15	4.58E-02

GWGS values were based on log₂FC, which represented their corresponding degree of differential expression, with genes of a higher degree of differential expression ranked higher.

Results

Identification of DE genes. The value of GWGS was used for integrated analyses of the independent microarray investigations. A gene with a high GWGS value was considered to be globally significant across multiple independent investigations. GWGS can be obtained based on the fold-change, *t*-test and significance analysis microarrays (SAM) (31). In the present study, the fold-change-based algorithm was more suitable for measurement of the significance of differential expression, since the present study aimed to examine the association between gene expression and network properties. By using the intersection of the microarray datasets, 217 DE genes were obtained for CRC and 341 DE genes were obtained for UC. In addition, DE genes present in CRC and UC were identified as common genes, and 62 common genes were identified (Table I).

KEGG analysis. The KEGG pathway database is a collection of manually drawn pathway maps for metabolism, genetic information processing, environmental information processing, including signal transduction, and various other cellular processes and human diseases (32). Pathway enrichment analysis of CRC revealed nine enriched terms (Table II), the most significant term was focal adhesion ($P=6.82E-004$), which contained several genes, including *CAVI*, *CCND1* and *PAK2*. In UC, five enriched terms (Table III) were obtained, the most important of which was ECM-receptor interaction ($P=1.09E-005$), which consisted of genes, including *LAMA1*,

VWF and *COL4A2*. Focal adhesion and the chemokine signaling pathway were presented in CRC and UC.

Drug gene interaction predictions. In the prediction of drug-gene interactions of CRC, the genes were found to be associated with 25 drugs, including collagenase ($P=3.88E-21$), estradiol ($P=6.81E-10$) and progesterone ($P=1.98E-09$; Table IV). A total of 21 drugs were found to be associated with genes in UC, including collagenase ($P=1.02E-20$), heparin ($P=1.45E-14$) and urokinase ($P=1.46E-08$; Table V). The collagenase, progesterone, heparin, urokinase, nadh and adenosine drugs were identified in both UC and CRC.

miRNA target gene prediction. In the prediction of miRNAs in CRC, 27 terms were identified (Table VI), and the most significant three terms were TACTTGA (MIR-26A and MIR-26B), AATGTGA (MIR-23A and MIR-23B) and CAGTATT (MIR-200B, MIR-200C and MIR-429). In UC, miRNA prediction revealed 43 terms of miRNA target genes (Table VII), the most significant three terms were TACTTGA (MIR-26A and MIR-26B), TGGTGCT (MIR-29A, MIR-29B and MIR-29C) and TGCCTTA (MIR-124A).

PPI network construction and analysis. In the present study, PPI networks were constructed for the DE genes in CRC and UC. In the network, nodes represent DE genes and edges between the nodes represent interaction of genes in the network. In the CRC network, there were 210 nodes and 752 edges, which included 217 DE genes (Fig. 1). Among the nodes, *MT2A* was identified with the highest degree at 42, followed by *COL1A1* at 37 and *COL1A2* at 37. In the UC network, there were 314 nodes, 882 edges and 341 DE genes (Fig. 2). *CD44* was identified with the highest degree, at 52, followed by *IL1B* at 50 and *MMP9* at 49.

Table III. Kyoto Encyclopedia of Genes and Genome analysis of differentially expressed genes in ulcerative colitis.

Term	Genes	P-value
hsa04512:ECM-receptor interaction	LAMA1, VWF, COL4A2, COL4A1, CD44, TNC, COL3A1, COL1A2, COL1A1, COL5A2, COL5A1, SPP1	1.09E-05
hsa04610:Complement and coagulation cascades	VWF, CD55, THBD, CFB, C4BPB, C4BPA, CFI, PLAU, PLAUR	4.28E-04
hsa04510:Focal adhesion	COL4A2, VAV3, COL4A1, TNC, COL3A1, COL5A2, COL5A1, VWF, LAMA1, COL1A2, ZYX, COL1A1, PIK3R3, SPP1	2.42E-03
hsa04062:Chemokine signaling pathway	CCL11, CXCL1, VAV3, CXCL5, CXCL3, CXCL9, CXCL6, PRKACB, PIK3R3, CXCL11, STAT1, CXCL10	1.04E-02
hsa04670:Leukocyte transendothelial migration	CLDN8, ICAM1, VAV3, NCF2, MMP9, PECAM1, CLDN1, PIK3R3	3.66E-02

Table IV. Results of drug genes prediction in colorectal cancer.

Drug	C-value	P-value
Collagenase	104	3.88E-21
Estradiol	122	6.81E-10
Progesterone	136	1.98E-09
Cisplatin	135	3.04E-08
Fluorouracil	68	5.37E-08
Acetazolamide	26	1.82E-07
Heparin	188	5.15E-07
Estrone	64	8.66E-07
Gentamicin	71	1.61E-06
Ciprofloxacin	111	1.57E-06
Sodium bicarbonate	42	2.20E-06
Urokinase	80	3.25E-06
Indomethacin	45	3.12E-06
Dexamethasone	89	6.05E-06
Daunorubicin	93	7.81E-06
Netilmicin	97	9.95E-06
Cefacetrile	100	1.19E-05
Cefotaxime	100	1.19E-05
Doxorubicin	103	1.40E-05
Tamoxifen	74	3.66E-05
Etoposide	125	4.21E-05
Hyaluronan	94	1.00E-04
Nadh	243	1.50E-03
Adenosine	477	3.00E-03
Glutathione	341	2.92E-02

Table V. Results of drug genes prediction in ulcerative colitis.

Drug	C-value	P-value
Collagenase	104	1.02E-20
Heparin	188	1.45E-14
Urokinase	80	1.46E-08
Alteplase	86	2.79E-08
Adenine	159	6.18E-08
Amiloride	61	5.28E-07
Nadh	243	4.15E-06
Immune globulin	624	1.13E-05
Cyclosporine	56	8.01E-05
Dinoprostone	61	8.97E-05
Rosuvastatin	134	9.74E-05
Adenosine monophosphate	102	2.48E-04
Glycine	191	7.99E-04
Progesterone	136	8.01E-04
Bupropion	108	1.72E-03
Adenosine triphosphate	299	2.70E-03
Tretinoin	126	3.34E-03
Nitric oxide	131	3.91E-03
Adenosine	477	4.85E-03
Vitamin a	145	5.93E-03
Phosphoric acid	159	8.71E-03

common genes were present in UC and CRC in Cluster 1: *COL1A2*, *MMP3*, *PLAU*, *CXCL5*, *CXCL3* and *CXCL1*. In Cluster 2, *MMP7*, *BGN*, *MMPI1*, *SPPI1* and *COL1A1* were common to UC and CRC. There were four common genes in Cluster 3: *SORD*, *MT1 M*, *MMP9* and *LCN2*.

For UC, three clusters were obtained (Fig. 4). Cluster 1 had the highest score (5.867), numbers of nodes (16 nodes) and number of edges (44 edges). There were five common genes present in UC and CRC in Cluster 1 (*COL1A1*, *SPPI1*,

Clusters. When the Node Score Cut-off=0.2, the Degree Cut-off=4, the *k*-core=4 and the maximum depth was set at 100, for CRC, three clusters were obtained (Fig. 3). Cluster 1 had the highest score (5.8) and number of edges (29 edges), the nodes of the three clusters were identical. A total of six

Table VI. Results of miRNA prediction in colorectal cancer.

miRNA	C-value	P-value
hsa_TACTTGA, MIR-26A, MIR-26B	297	3.64E-07
hsa_AATGTGA, MIR-23A, MIR-23B	417	1.53E-06
hsa_CAGTATT, MIR-200B, MIR-200C, MIR-429	465	4.67E-06
hsa_TATTATA, MIR-374	284	0.99E-05
hsa_TGAATGT, MIR-181A, MIR-181B, MIR-181C, MIR-181D	479	2.01E-04
hsa_CTTGTAT, MIR-381	201	6.11E-04
hsa_TTTTGAG, MIR-373	222	9.23E-04
hsa_ATGAAGG, MIR-205	156	1.21E-03
hsa_TGGTGCT, MIR-29A, MIR-29B, MIR-29C	515	1.24E-03
hsa_AAGCCAT, MIR-135A, MIR-135B	332	1.49E-03
hsa_TGCCTTA, MIR-124A	542	1.82E-03
hsa_ACTGTGA, MIR-27A, MIR-27B	465	2.50E-03
hsa_TGTTTAC, MIR-30A-5P, MIR-30C, MIR-30D, MIR-30B, MIR-30E-5P	572	5.52E-03
hsa_ATTCTTT, MIR-186	270	2.52E-03
hsa_CTACCTC, LET-7A, LET-7B, LET-7C, LET-7D, LET-7E, LET-7F, MIR-98, LET-7G, LET-7I	384	3.41E-03
hsa_TTTGCAC, MIR-19A, MIR-19B	511	4.52E-03
hsa_CACCAGC, MIR-138	223	5.54E-03
hsa_TAATAAT, MIR-126	220	5.20E-03
hsa_AAGCACT, MIR-520F	236	6.90E-03
hsa_TTTGTAG, MIR-520D	335	7.11E-03
hsa_GTTTGTT, MIR-495	252	9.05E-03
hsa_GTGCCTT, MIR-506	714	1.02E-02
hsa_TGCTGCT, MIR-15A, MIR-16, MIR-15B, MIR-195, MIR-424, MIR-497	593	1.05E-02
hsa_ACCAAAG, MIR-9	493	1.26E-02
hsa_CTTTGTA, MIR-524	431	2.21E-02
hsa_TGCTTTG, MIR-330	331	2.61E-02
hsa_AGCACTT, MIR-93, MIR-302A, MIR-302B, MIR-302C, MIR-302D, MIR-372, MIR-373, MIR-520E, MIR-520A, MIR-526B, MIR-520B, MIR-520C, MIR-520D	336	2.76E-02

MIR/miRNA, microRNA.

COLIA2, *BGN* and *MMP9*), four in Cluster 2 (*CXCL5*, *MMPI*, *MMP7* and *PLAU*) and four in Cluster 3 (*LCN2*, *OLFMA*, *PTN* and *REGIB*).

Analysis of network properties. The degree and betweenness centralities for the clusters in CRC and UC were calculated. As shown in Fig. 5, the topological centrality-based degree among the clusters revealed that Cluster 2 of UC had the highest degree at 29), while Cluster 3 of UC had the lowest degree at 13. As shown in Fig. 6, the betweenness of Cluster 3 also had the lowest betweenness (0.02). GWGS is closely associated with log2FC and indicates the corresponding degree of the DE genes, with DE genes of a higher degree exhibiting higher ranking values. As shown in Fig. 7, no significant difference was observed in the rank values between CRC and UC. On comparison of the clusters in CRC, Cluster 1 had the highest rank value (5.02), while cluster 3 of UC had the highest rank value (5.91).

Discussion

In the present study, DE genes with GWGS values in CRC and UC were identified through integrated analysis of multiple high throughput data. Based on the DE genes, PPI networks were constructed using the STRING database, and MCODE algorithm was implemented for sub-network detection. The significance of sub-networks was identified based on the network properties and GWGS values. In addition, functional enrichment analysis, including KEGG enrichment analysis, drug-gene interaction prediction, and miRNA prediction, were performed.

A total of 217 DE genes of CRC, 341 DE genes of UC and 62 common genes were identified. The KEGG pathway analysis revealed nine terms of CRC and five terms of UC, with the focal adhesion and chemokine signaling pathway presented in both. As for the prediction of drug-gene interactions, collage-

Table VII. Results of miRNA prediction in ulcerative colitis.

miRNA	C-value	P-value
hsa_TACTTGA, MIR-26A, MIR-26B	297	1.25E-07
hsa_TGGTGCT, MIR-29A, MIR-29B, MIR-29C	515	8.93E-07
hsa_TGCCTTA, MIR-124A	542	1.78E-06
hsa_CAGTATT, MIR-200B, MIR-200C, MIR-429	465	5.17E-06
hsa_CATTCA, MIR-203	284	1.79E-05
hsa_GTGCCAA, MIR-96	301	3.05E-05
hsa_ACCAAAG, MIR-9	493	2.03E-04
hsa_ACTGTGA, MIR-27A, MIR-27B	465	4.06E-04
hsa_AATGTGA, MIR-23A, MIR-23B	417	5.11E-04
hsa_TTTGCAC, MIR-19A, MIR-19B	511	8.05E-04
hsa_CTACCTC, LET-7A, LET-7B, LET-7C, LET-7D, LET-7E, LET-7F, MIR-98, LET-7G, LET-7I	384	1.00E-03
hsa_TTGGAGA, MIR-515-5P, MIR-519E	145	1.10E-03
hsa_CACCAGC, MIR-138	223	2.00E-03
hsa_AAGCCAT, MIR-135A, MIR-135B	332	1.44E-03
hsa_TGAATGT, MIR-181A, MIR-181B, MIR-181C, MIR-181D	479	1.61E-03
hsa_TAATAAT, MIR-126	220	1.90E-03
hsa_AAGCAAT, MIR-137	217	1.74E-03
hsa_TATTATA, MIR-374	284	1.99E-03
hsa_CTATGCA, MIR-153	214	1.63E-03
hsa_AACTGGA, MIR-145	231	2.51E-03
hsa_ATACCTC, MIR-202	178	3.01E-03
hsa_CAGTGTT, MIR-141, MIR-200A	308	3.20E-03
hsa_GTGCAAT, MIR-25, MIR-32, MIR-92, MIR-363, MIR-367	308	3.22E-03
hsa_AAGCACA, MIR-218	395	4.32E-03
hsa_AAAGACA, MIR-511	199	5.11E-03
hsa_TGTTTAC, MIR-30A-5P, MIR-30C, MIR-30D, MIR-30B, MIR-30E-5P	572	6.02E-03
hsa_TGTATGA, MIR-485-3P	148	6.49E-03
hsa_CAGCAGG, MIR-370	153	7.41E-03
hsa_ATGAAGG, MIR-205	156	8.03E-03
hsa_ACATTCC, MIR-1, MIR-206	293	8.92E-03
hsa_CTGAGCC, MIR-24	229	9.81E-03
hsa_TAGCTTT, MIR-9	234	1.09E-02
hsa_AAGCACT, MIR-520F	236	1.13E-02
hsa_TTGCCAA, MIR-182	324	1.48E-02
hsa_GCAAAAA, MIR-129	183	1.52E-02
hsa_ATACTGT, MIR-144	198	2.06E-02
hsa_ATTCTTT, MIR-186	270	2.05E-02
hsa_CTTGTAT, MIR-381	201	2.16E-02
hsa_CTTTGTA, MIR-524	431	2.18E-02
hsa_TTTGCAG, MIR-518A-2	208	2.48E-02
hsa_ATATGCA, MIR-448	208	2.48E-02
hsa_TGCACTG, MIR-148A, MIR-152, MIR-148B	299	3.16E-02
hsa_ATGTACA, MIR-493	312	3.77E-02

MIR/miRNA, microRNA.

nase was important in the drug associated genes of CRC and UC. The most significant miRNA prediction term in CRC and UC was the same, TACTTGA (MIR-26A and MIR-26B). The

entire PPI network was constructed and subnetwork analyzed, the clusters contained common genes and exhibited similarities between CRC and UC. No significant difference was observed

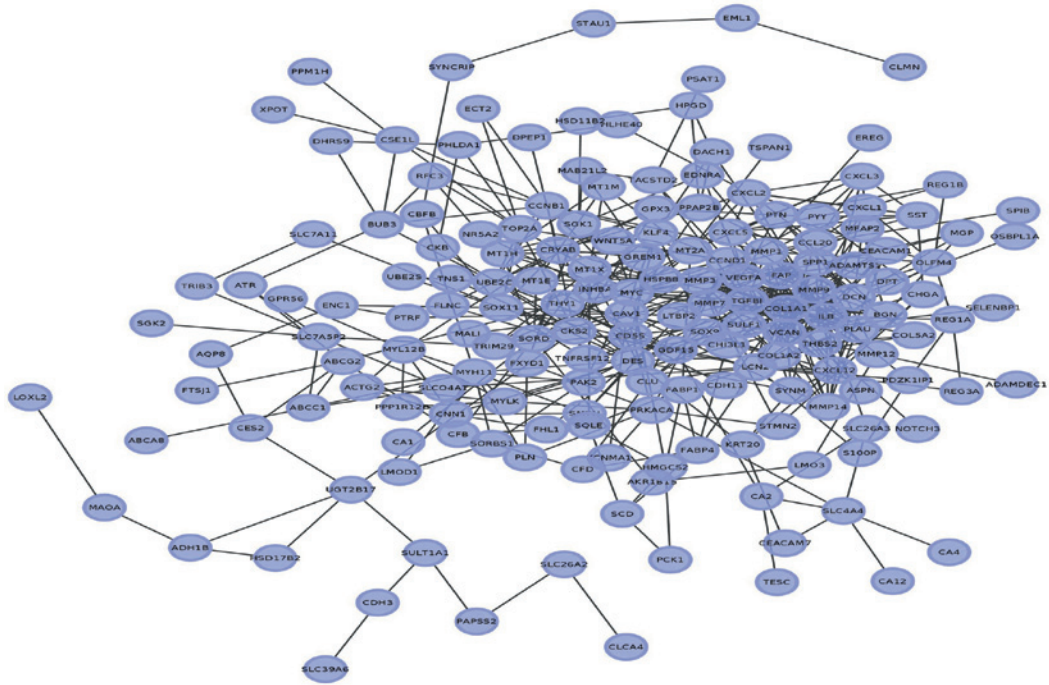


Figure 1. Protein-protein interaction network of colorectal cancer DE genes. A total of 210 node (purple ovals) and 752 edges (lines between) were identified, which included 217 DE genes. Among the nodes, *MT2A* exhibited the highest degree (42), followed by *COL1A1* (37) and *COL1A2* (37). Node sizes correspond to the absolute values of the fold change of the DE gene. Edges were derived from the Search Tool for the Retrieval of Interacting Genes/proteins database. DE, differentially expressed.

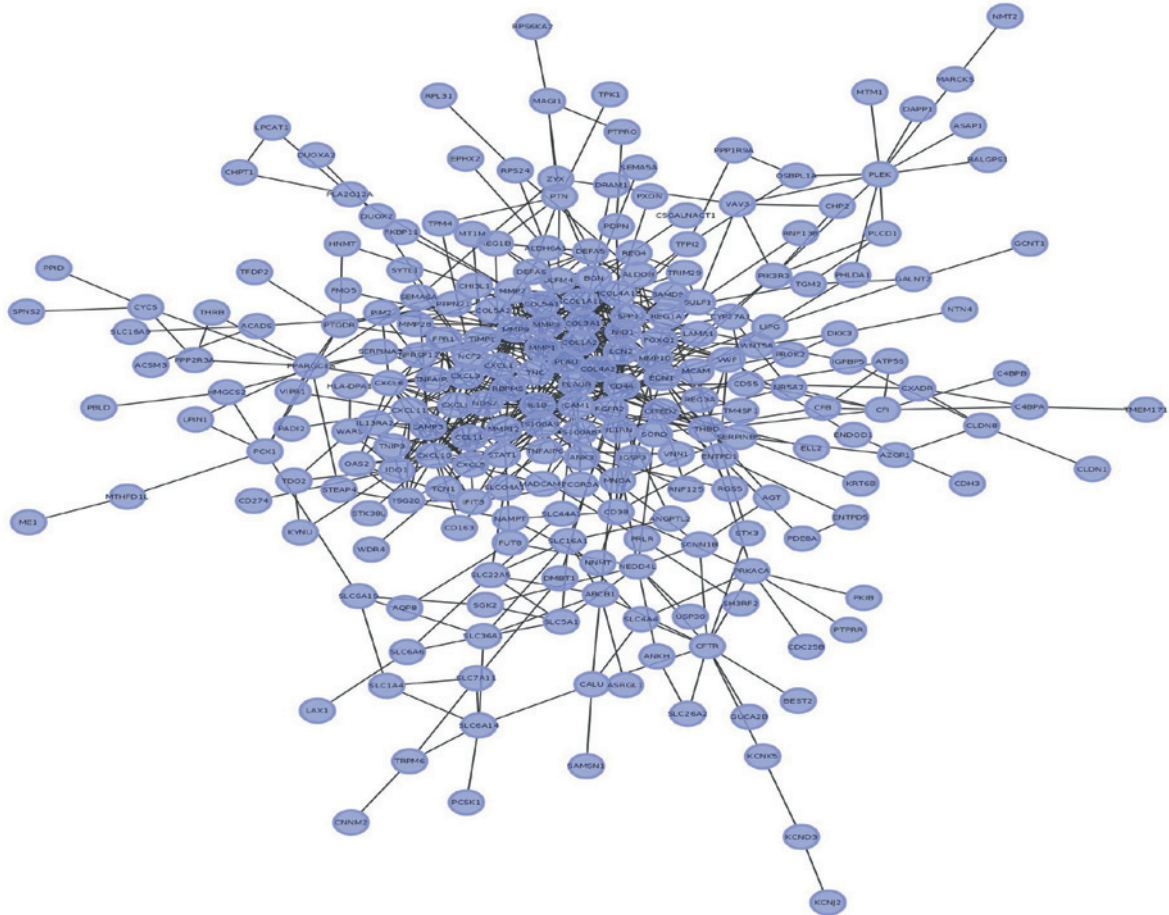


Figure 2. Protein-protein interaction network of ulcerative colitis DE genes. A total of 314 nodes (purple ovals), 882 edges (lines between nodes) and 341 DE genes were identified, where nodes represent gene signatures and edges between nodes represent interaction between genes in the network. Among the nodes, *CD44* exhibited the highest degree (52), followed by *IL1B* (50) and *MMP9* (49). Node sizes correspond to the absolute values of the fold change of the DE genes. Edges were derived from the Search Tool for the Retrieval of Interacting Genes/proteins database. DE, differentially expressed.

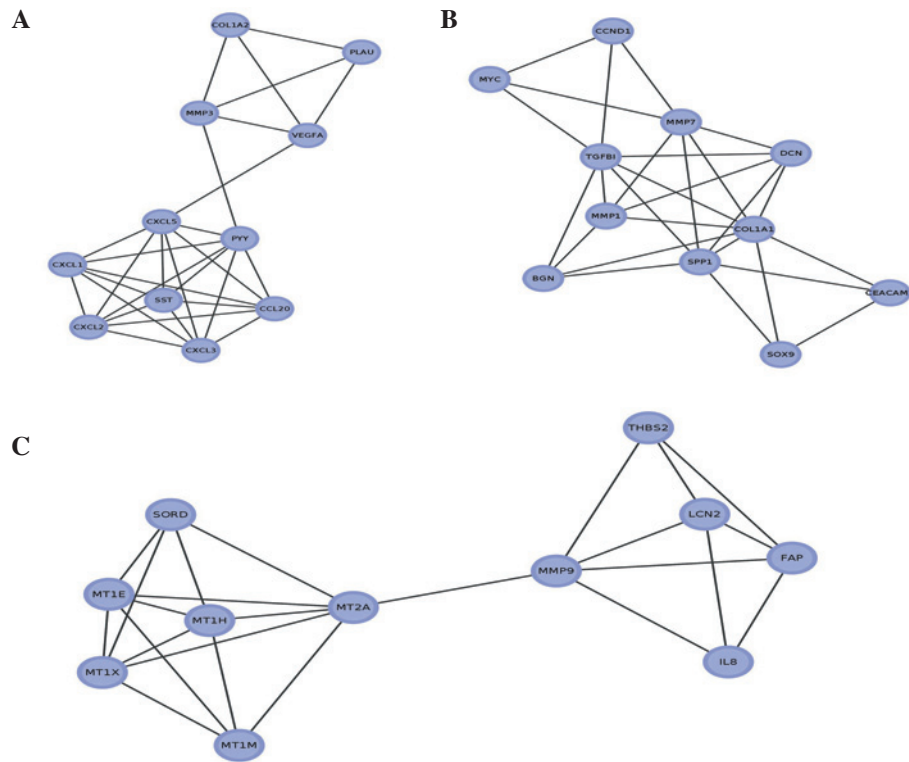


Figure 3. Clusters of the protein-protein interaction network of CRC. (A) Cluster 1, (B) Cluster 2 and (C) Cluster 3. Clusters were identified according to the following cut off-values: Node Score=0.2, degree=4, k -core=4, Maximum depth=100. Cluster 1 had the highest degree (5.8) and number of edges (29), the nodes of the three clusters were same. Common genes to CRC and ulcerative colitis in Cluster 1 were *COL1A2*, *MMP3*, *PLAU*, *CXCL5*, *CXCL3* and *CXCL1*. In Cluster 2, common genes were *MMP7*, *BGN*, *MMP1*, *SPP1*, and *COL1A1*. There were four common genes in Cluster 3 (*SORD*, *MT1 M*, *MMP9* and *LCN2*). Node sizes correspond to the absolute values of the fold change of the differentially expressed genes. Edges were derived from the Search Tool for the Retrieval of Interacting Genes/proteins database; CRC, colorectal cancer.

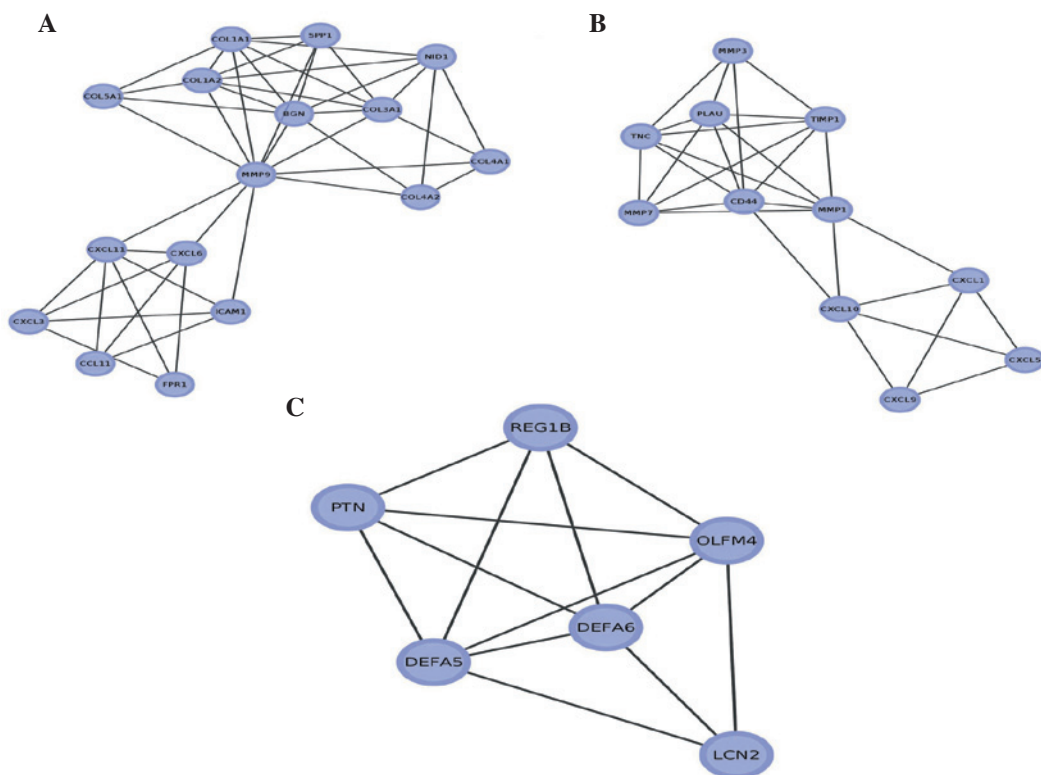


Figure 4. Clusters of the protein-protein interaction network in UC. (A) Cluster 1, (B) cluster 2 and (C) cluster 3. Cluster 1 had the highest degree (5.867), and numbers of nodes (16) and edges (44). There were five common genes to UC and colorectal cancer in cluster 1 (*COL1A1*, *SPP1*, *COL1A2*, *BGN* and *MMP9*), four in cluster 2 (*CXCL5*, *MMP1*, *MMP7* and *PLAU*) and four in cluster 3 (*LCN2*, *OLFMA*, *PTN* and *REG1B*). Node sizes correspond to the absolute values of the fold change of the differentially expressed genes. Edges were derived from the Search Tool for the Retrieval of Interacting Genes/proteins database. UC, ulcerative colitis.

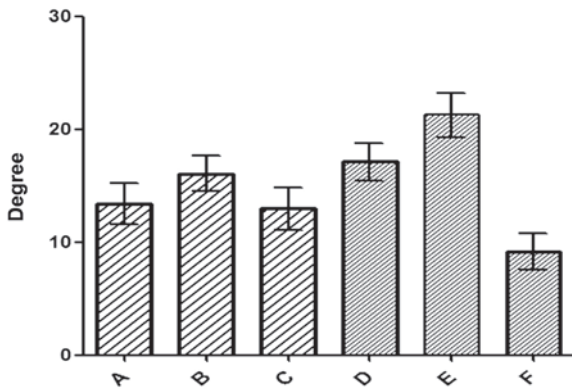


Figure 5. Centralities analysis based on the degree of the clusters. (A) Cluster 1 of CRC; (B) cluster 2 of CRC; (C) cluster 3 of CRC; (D) cluster 1 of UC; (E) cluster 2 of UC; (F) cluster 3 of UC. No significant differences were observed between the degrees of the clusters in CRC. Cluster 2 of UC had the highest degree (29), while Cluster 3 of UC had the lowest degree (13). CRC, colorectal cancer; UC, ulcerative colitis.

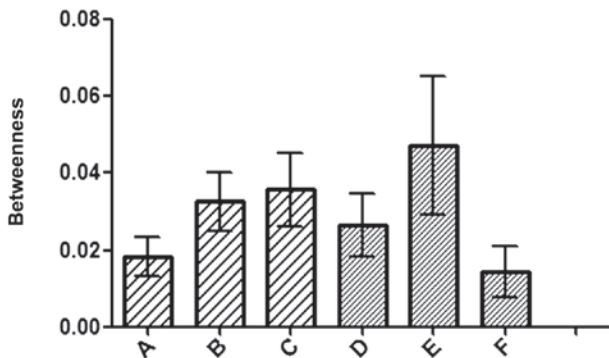


Figure 6. Centralities analysis based on the betweenness of clusters. (A) Cluster 1 of CRC; (B) cluster 2 of CRC; (C) cluster 3 of CRC; (D) cluster 1 of UC; (E) cluster 2 of UC; (F) cluster 3 of UC. Cluster 2 of UC had the highest betweenness, while cluster 3 of UC possessed the lowest betweenness. CRC, colorectal cancer; UC, ulcerative colitis.

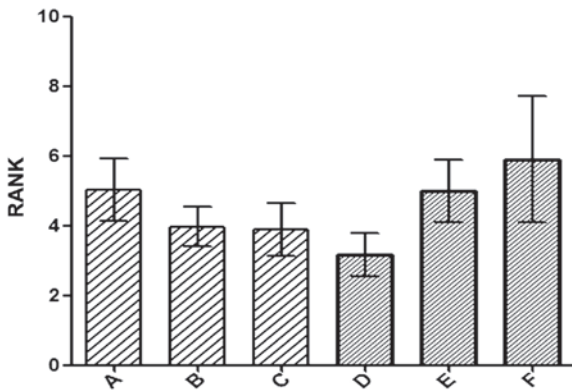


Figure 7. Rank values of the clusters. (A) Cluster 1 of CRC; (B) cluster 2 of CRC; (C) cluster 3 of CRC; (D) cluster 1 of UC; (E) cluster 2 of UC; (F) cluster 3 of UC. No significant differences were observed between the rank values of the clusters in CRC and UC. CRC, colorectal cancer; UC, ulcerative colitis.

between the GWGS values of the clusters in UC and CRC. In UC, cluster 3 had the highest GWGS value, while the topological centrality of this cluster had the lowest degree and betweenness.

Patients with UC have an increased risk of developing CRC, compared with the general population (33), and the increased risk was almost entirely confined to patients with long-standing extensive colitis (3). Important risk factors include primary sclerosing cholangitis (34) and a family history of CRC (35), whereas the role of other factors, including the effect of the age at onset of UC remains to be elucidated. It has been reported that hypermethylation of the promoter region of *CDH1* in CRC is associated with a reduction in UC (36). In the present study, 62 common genes were found between CRC and UC. The most significant two genes were *AQP8* and *CXCL5*. *AQP8* is a water channel protein, and aquaporins are a family of small integral membrane proteins associated with major intrinsic protein, and is closely associated with miRNA in patients with UC patients (37). Thus, it is possible that certain genes expressed in patients with UC are also expressed in patients with CRC, and the inhibition of certain genes in UC may decrease risk of CRC.

The focal adhesion and chemokine signaling pathway were found to be present in CRC and UC. It has been previously revealed that the predominant type of pathway in UC-associated neoplasia is associated with genes and, that genomic instability frequently occurs prior to the development of histologically-defined dysplasia (38). Using genes to construct a predictive model to distinguish patients with and without UC-induced CRC is a useful method to identify the disease (38). Therefore, controlling the pathway of UC to prevent the formation of UC-associated neoplasia may decrease the incidence of CRC.

The present study demonstrated that collagenase, progesterone, heparin, urokinase, nadh and adenosine drugs may be used to treat CRC and UC. Due to this possible mechanism for inflammation-induced cancer, patients taking anti-inflammatory mesalamine drugs may exhibit reduced rates of colorectal neoplasia (39). These results were concordant with the hypothesis that certain drugs may offer potential for use in the treatment of CRC and UC. For example, collagenases, are proteolytic enzymes, which are present within cells in an inactive form and are secreted at sites of inflammation by mononuclear cells and metastatic tumors (40), thus, collagenase not only indicates the potential for preventing UC from inflammation interference, but also a potential effect on tumors, which lead to cancer.

Previous studies have demonstrated that miRNAs are the central regulators of various physiological processes, and that the disruption of miRNA is associated with human diseases (41,42). Therefore, in the present study, miRNA prediction experiments were performed. The most significant miRNA in CRC was the same as that in UC. Therefore, the disruption of miRNA in UC may lead alter miRNA in CRC. The link between miRNA function and cancer pathogenesis was further supported by investigations examining miRNA in clinical samples, with altered miRNA being reported in CRC (43). The present study hypothesized that the altered miRNA in CRC was from UC.

The examination of networks as a tool has attracted significant attention in analyzing several biological and communication systems. Protein interaction network analysis provides an effective method of estimating and understanding the likelihood of potential, yet undetermined,

connections between proteins/genes (44). In PPI networks, the data of large-scale protein interactions has accumulated with the development of high throughput assessment technology, however, a certain number of interactions have not been assessed, which may be important. This type of difficulty had been resolved to a certain extent by the use of clustering methods, which had previously been found to be useful in identifying protein/gene interactions within the same cellular process (45). In the present study, the MCODE algorithm was applied to examine gene-gene connectivity in a more informative way, which revealed three clusters in CRC and UC with highly connected nodes. Several common genes were contained in the clusters, which indicated that the clusters of CRC and UC had certain similarities. Srihari and Ragan performed a straightforward, systematic identification and comparison of modules across pancreatic normal and cancer tissue conditions by integrating PPI, gene-expression and mutation data (46), which provided functional insight into the identified sub-network and thus may be suitable for analysis of CRC.

In several PPI networks, significance is correlated with the topological placement of the proteins/genes in the network, while connectivity provides an indication of the importance of a gene (47). In the present study, the highest ranking gene in degree and betweenness centralities was *PLAU* in both UC and CRC. This gene encodes a serine protease, which is involved in degradation of the extracellular matrix and possibly tumor cell migration and proliferation (48). A specific polymorphism in this gene may be associated with late-onset (49). However, the GWGS value of Cluster 2 in UC was not in accordance with the degree of betweenness. GWGS was a novel method to detect the relevance of genes between clusters, based on the rank value of gene expression. In previous studies, the topological centrality, based on degree, was not consistent with that based on betweenness, even altered rules of the same clusters in degree, betweenness, closeness and other properties (such as cluster coefficient and stress) were different (50,51). Differences among these properties may be explained by the fact that each property has its own target (52); for example GWGS focuses on combining gene expression with the protein network, while the degree concerns the association between genes. Investigation of the worth of rank values in gene signatures and biological analysis is required in the future.

In conclusion, the present study demonstrated the presence of 62 common genes in CRC and UC DE genes and KEGG analysis obtained the same gene terms, therefore, controlling these terms in UC may decrease the risk and rate of CRC formation. Through drug genes prediction, drugs were identified, which may treat UC and CRC simultaneously to cure patients with UC and possibly prevented patients from developing CRC. According to PPI network analysis, a significant PPI network and subnet was produced, with common genes included in clusters. No significant differences were observed in the GWGS values of the clusters in UC and CRC. Cluster 3 in UC had the highest GWGS value, whereas the topological centrality of Cluster 3 in UC had the lowest degree and betweenness. These findings may provide potential biomarkers and reveal information regarding the pathological mechanism of CRC induced by UC.

Acknowledgements

This study was supported by the Shandong Province scientific and technological project (grant. no. 2012G0021830).

References

- Karlén P, Kornfeld D, Broström O, Löfberg R, Persson PG and Ekblom A: Is colonoscopic surveillance reducing colorectal cancer mortality in ulcerative colitis? A population based case control study. *Gut* 42: 711-714, 1998.
- Kisiel JB, Garrity-Park MM, Taylor WR, Smyrk TC and Ahlquist DA: Methylated eyes absent 4 (*EYA4*) gene promoter in non-neoplastic mucosa of ulcerative colitis patients with colorectal cancer: Evidence for a field effect. *Inflamm Bowel Dis* 19: 2079-2083, 2013.
- Watanabe T, Konishi T, Kishimoto J, Kotake K, Muto T and Sugihara K; Japanese Society for Cancer of the Colon and Rectum: Ulcerative colitis-associated colorectal cancer shows a poorer survival than sporadic colorectal cancer: A nationwide Japanese study. *Inflamm Bowel Dis* 17: 802-808, 2011.
- Bardelli A and Siena S: Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. *J Clin Oncol* 28: 1254-1261, 2010.
- Saleh M and Trinchieri G: Innate immune mechanisms of colitis and colitis-associated colorectal cancer. *Nat Rev Immunol* 11: 9-20, 2011.
- Fearon ER: Molecular genetics of colorectal cancer. *Ann Rev Pathol* 6: 479-507, 2011.
- Suzuki H, Gabrielson E, Chen W, Anbazhagan R, van Engeland M, Weijenberg MP, Herman JG and Baylin SB: A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat Genet* 31: 141-149, 2002.
- Cancer Genome Atlas Network: Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487: 330-337, 2012.
- Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumüller RA, Mohr SE and Perrimon N: Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat Methods* 11: 94-99, 2014.
- Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, *et al*: A network-based analysis of systemic inflammation in humans. *Nature* 437: 1032-1037, 2005.
- Goffard N and Weiller G: PathExpress: A web-based tool to identify relevant pathways in gene expression data. *Nucleic Acids Res* 35: W176-W181, 2007.
- Chuang HY, Lee E, Liu YT, Lee D and Ideker T: Network-based classification of breast cancer metastasis. *Mol Syst Biol* 3: 140, 2007.
- Nepusz T, Yu H and Paccanaro A: Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 9: 471-472, 2012.
- Winterhalter C, Widera P and Krasnogor N: JEPETTO: A cytoscape plugin for gene set enrichment and topological analysis based on interaction networks. *Bioinformatics* 30: 1029-1030, 2014.
- Chuang HY, Hofree M and Ideker T: A decade of systems biology. *Annu Rev Cell Dev Biol* 26: 721-744, 2010.
- Pržulj N, Wagle DA and Jurisica I: Functional topology in a network of protein interactions. *Bioinformatics* 20: 340-348, 2004.
- Wu F, Dassopoulos T, Cope L, Maitra A, Brant SR, Harris ML, Bayless TM, Parmigiani G and Chakravarti S: Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflamm Bowel Dis* 13: 807-821, 2007.
- Planell N, Lozano JJ, Mora-Buch R, Masamunt MC, Jimeno M, Ordás I, Esteller M, Ricart E, Piqué JM, Panés J, *et al*: Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. *Gut* 62: 967-976, 2013.
- Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, Paty PB, Gerald WL, Notterman DA and Domany E: Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci USA* 106: 7131-7136, 2009.

20. Galamb O, Spisák S, Sipos F, Tóth K, Solymosi N, Wichmann B, Krenács T, Valcz G, Tulassay Z and Molnár B: Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *Br J Cancer* 102: 765-773, 2010.
21. Ancona N, Maglietta R, Piepoli A, D'Addabbo A, Cotugno R, Savino M, Liuni S, Carella M, Pesole G and Perri F: On the statistical assessment of classifiers using DNA microarray data. *BMC Bioinformatics* 7: 387, 2006.
22. Liu W, Peng Y and Tobin DJ: A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis. *Peer J* 1: e49, 2013.
23. Smyth GK: Limma: Linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*. In: *Statistics for Biology and Health*. Gentleman *et al* (eds). Springer, New York, NY, p397-420, 2005.
24. Diboun I, Wernisch L, Örengo CA and Koltzenburg M: Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics* 7: 252, 2006.
25. Da Wei Huang BTS and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2008.
26. Zhang B, Kirov S and Snoddy J: WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33: W741-W748, 2005.
27. Wang J, Duncan D, Shi Z and Zhang B: WEB-based GENE SeT Analysis toolkit (WebGestalt): Update 2013. *Nucleic Acids Res* 41: W77-W83, 2013.
28. Ferreira J and Zwinderman A: On the Benjamini-Hochberg method. *Ann Statistics* 34: 1827-1849, 2006.
29. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27: 431-432, 2011.
30. Wasserman S and Katherin Faust (eds): *Social Network Analysis: Methods and Applications*. 1st edition. Cambridge University Press, Cambridge, 1994.
31. Altman NS: *Differential expression analysis using LIMMA*. 2013.
32. Kanehisa M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30, 2000.
33. Gillen CD, Walmsley RS, Prior P, Andrews H and Allan RN: Ulcerative colitis and Crohn's disease: A comparison of the colorectal cancer risk in extensive colitis. *Gut* 35: 1590-1592, 1994.
34. Loftus EV Jr, Harewood GC, Loftus CG, Tremaine WJ, Harmsen WS, Zinsmeister AR, Jewell DA and Sandborn WJ: PSC-IBD: A unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. *Gut* 54: 91-96, 2005.
35. Askling J, Dickman PW, Karlén P, Broström O, Lapidus A, Löfberg R and Ekblom A: Family history as a risk factor for colorectal cancer in inflammatory bowel disease. *Gastroenterology* 120: 1356-1362, 2001.
36. Wheeler JM, Kim HC, Efstathiou JA, Ilyas M, Mortensen NJ and Bodmer WF: Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. *Gut* 48: 367-371, 2001.
37. Min M, Peng LH, Sun G, Guo MZ, Qiu ZW and Yang YS: Aquaporin 8 expression is reduced and regulated by microRNAs in patients with ulcerative colitis. *Chin Med J* 126: 1532-1537, 2013.
38. Watanabe T, Kobunai T, Yamamoto Y, Ikeuchi H, Matsuda K, Ishihara S, Nozawa K, Iinuma H, Kanazawa T, Tanaka T, *et al*: Predicting ulcerative colitis-associated colorectal cancer using reverse-transcription polymerase chain reaction analysis. *Clin Colorectal Cancer* 10: 134-141, 2011.
39. Rutter M, Saunders B, Wilkinson K, Rumbles S, Schofield G, Kamm M, Williams C, Price A, Talbot I and Forbes A: Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology* 126: 451-459, 2004.
40. Rosenberg GA, Mun-Bryce S, Wesley M and Kornfeld M: Collagenase-induced intracerebral hemorrhage in rats. *Stroke* 21: 801-807, 1990.
41. Ikeda S, Kong SW, Lu J, Bisping E, Zhang H, Allen PD, Golub TR, Pieske B and Pu WT: Altered microRNA expression in human heart disease. *Physiol Genomics* 31: 367-373, 2007.
42. O'Connell RM, Rao DS and Baltimore D: microRNA regulation of inflammatory responses. *Annu Rev Immunol* 30: 295-312, 2012.
43. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, Barad O, Bentwich Z, Szafrańska AE, Labourier E, *et al*: The colorectal microRNAome. *Proc Natl Acad Sci USA* 103: 3687-3692, 2006.
44. Wen Z, Liu ZP, Liu Z, Zhang Y and Chen L: An integrated approach to identify causal network modules of complex diseases with application to colorectal cancer. *J Am Med Inform Assoc* 20: 659-667, 2013.
45. Jonsson PF, Cavanna T, Zicha D and Bates PA: Cluster analysis of networks generated through homology: Automatic identification of important protein communities involved in cancer metastasis. *BMC Bioinformatics* 7: 2, 2006.
46. Srihari S and Ragan MA: Systematic tracking of dysregulated modules identifies novel genes in cancer. *Bioinformatics* 29: 1553-1561, 2013.
47. Estrada E: Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics* 6: 35-40, 2006.
48. Ploug M, Gårdsvoll H, Jørgensen TJ, Lønborg Hansen L and Danø K: Structural analysis of the interaction between urokinase-type plasminogen activator and its receptor: a potential target for anti-invasive cancer therapy. *Biochem Soc Trans* 30: 177-183, 2002.
49. Finckh U, Van Hadeln K, Müller-Thomsen T, Alberici A, Binetti G, Hock C, Nitsch RM, Stoppe G, Reiss J and Gal A: Association of late-onset Alzheimer disease with a genotype of PLAU, the gene encoding urokinase-type plasminogen activator on chromosome 10q22. 2. *Neurogenetics* 4: 213-217, 2003.
50. Opsahl T, Agneessens F and Skvoretz J: Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc Networks* 32: 245-251, 2010.
51. Ai J, Zhao H, Carley KM, Su Z and Li H: Neighbor vector centrality of complex networks based on neighbors degree distribution. *Euro Phys J B* 86: 1-7, 2013.
52. Kapoor K, Sharma D and Srivastava J: Weighted node degree centrality for hypergraphs. *Proceedings of the 2013 IEEE 2nd International Network Science Workshop*, Westpoint, NY, USA, pp152-155, 2013.