

# Computational Approaches towards Micro-RNA Expression Analysis for Identification of Biomarkers in Inflammatory Bowel Disease Subtypes

Manisha Mandal<sup>1</sup>, Shyamapada Mandal<sup>2,\*</sup>

## ABSTRACT

**Background and Aim:** The potential biomarkers of inflammatory bowel diseases (IBDs) were analyzed from GSE53867 dataset. The current study aimed to identify miRNAs as potential diagnostic and prognostic biomarkers to delineate the IBD subtypes. **Methods:** Differentially expressed microRNAs (DEMs)-genes and protein-protein interaction networks were constructed and hub genes were selected using Cytoscape. Differentially expressed genes were analyzed for GO (Gene Ontology) and Reactome-pathway. **Results:** Seven DEMs were upregulated in Crohn's disease (CD), 4 downregulated in ulcerative colitis (UC), 8 upregulated and 2 downregulated in IBD. A 620, 2377 and 1821 target-genes were in CD, UC and IBD, respectively. SOCS3, upregulated by miR-650, was hub gene in CD, induced by cytokines, through NFkB-signalling pathway to mediate ubiquitin-proteasomal degradation. CIRH1A, downregulated by miR-16, was hub gene of UC, acted by impairing ribosome-biogenesis. SKP2 and ASB1, up- and downregulated, by miR-142 and miR-665, respectively, were hub genes of IBD, induced cytokines through activation of TLR- and TNF-signalling pathways to mediate ubiquitin-proteasomal degradation. **Conclusion:** The SOCS3, CIRH1A, SKP2 and ASB1 genes might serve as valuable biomarkers to differentiate CD, UC and IBD. **Key words:** MicroRNA, Biomarkers, Genes, Inflammatory bowel disease, Crohn's disease, Ulcerative colitis.

## INTRODUCTION

The inflammatory bowel diseases (IBDs), including ulcerative colitis (UC) and Crohn's disease (CD), are chronic, idiopathic inflammation of the gastrointestinal tract (GIT), with 6.8 million cases worldwide, as reported in 2017.<sup>[1]</sup> As reported by Kedia and Ahuja in 2010, it has been 1.64 million cases in the USA and 1.4 million in India.<sup>[2]</sup> The potential causes of IBD are genetic susceptibility, environmental factors and dysregulation of immune system.<sup>[3]</sup> The CD can affect any area of the GIT, including the small intestine and colon, has patchy transmural inflammation; while UC affects only the colon, has continuous superficial inflammation.<sup>[4]</sup> The overlapping symptoms and complications of the disease along with limited treatment options having side effects, highlights the significance of developing specific diagnostic and novel therapeutic strategies in IBD. Recent studies have indicated the role of microRNA (miRNA) in IBD pathogenesis with therapeutic prospects. miRNAs are small (18–24 nucleotides), single stranded, conserved, noncoding RNAs that bind to complementary 3'-UTRs (untranslated regions) of target mRNA (messenger RNA) causing instability and inhibition of translation, rendering miRNAs as important epigenetic

modifiers of gene expression.<sup>[5]</sup> For example, miR-21, miR-31 and miR-141 have been implicated in IBD pathogenesis by targeting PDCD4 and Rho GTPase RhoB, IL-25 and CXCL5 and CXCL12 $\beta$  genes mediated inflammatory responses.<sup>[6]</sup> However, the pivotal molecular interactions of IBD based on differentially expressed genes (DEGs) have not been extensively investigated. Therefore, the current study is an attempt to identify the potential biomarkers that distinguish IBD and its subtypes, investigate underlying molecular interactions based on DEGs profiling, functional enrichment and significant pathways associated with the DEGs, protein-protein interaction (PPI) network of common DEGs. To achieve this, the GSE53867 dataset was downloaded from the Gene Expression Omnibus (GEO) database ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). Differentially expressed miRNAs (DEMs) were identified between and unique to UC and CD patients, compared with controls. The DEMs-gene network and protein-protein interaction (PPI) network were constructed to identify potential biomarkers and major regulated genes. GO (gene ontology) biological process and reactome pathway analysis was performed with modules of PPI network to determine the functional

Manisha Mandal<sup>1</sup>,  
Shyamapada Mandal<sup>2,\*</sup>

<sup>1</sup>Department of Physiology, MGM Medical College, Kishanganj, Bihar, INDIA.

<sup>2</sup>Laboratory of Microbiology and Experimental Medicine, Department of Zoology, University of Gour Banga, Malda, West Bengal, INDIA.

### \*Correspondence

Prof. Shyamapada Mandal

Laboratory of Microbiology and Experimental Medicine, Department of Zoology, University of Gour Banga, Malda-732103, West Bengal, INDIA.

Phone: +91 9831279239

Email: samtropmed@gmail.com

### History

- Submission Date: 31-07-2020;
- Review completed: 07-09-2020;
- Accepted Date: 19-09-2020.

DOI : 10.5530/ijcep.2020.7.3.25

Article Available online

<http://www.ijcep.org>

### Copyright

© 2020 Phcog.Net. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

**Cite this article:** Mandal M, Mandal S. Computational Approaches towards Micro-RNA Expression Analysis for Identification of Biomarkers in Inflammatory Bowel Disease Subtypes. Int J Clin Exp Physiol. 2020;7(3):100-6.

enrichment and significant pathways associated with the DEGs, to gain further insight into the pathogenesis of UC and CD at the molecular level, which may help in the differential diagnosis of UC and CD and determine prospective molecular targets for the IBD subtypes.

## MATERIALS AND METHODS

### Microarray Data Acquisition

The miRNA expression dataset GSE53867 was obtained from the gene expression omnibus database ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)), containing two CD (CD1 and CD2), two UC (UC1 and UC2) compared to a single non-IBD control (C) human colon biopsy samples.

### Identification of DEMs

The DEMs were identified between UC and C (defined as UC group), CD and C (defined as CD group), UC and CD (defined as IBD group) groups, using GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) for performing comparisons on GSE53867 raw data applying GEOquery, limma R and Biobase packages from the Bioconductor project.<sup>[7,8]</sup> The raw microarray data was preprocessed to obtain "NA" filtered expression data and normalized using "normalizeBetweenArrays" in limma, with quantile method. The normalized microarray expression data was fit to a linear model and compared between groups using Empirical Bayes moderated t test, also in limma, to obtain log<sub>2</sub>FoldChange (log<sub>2</sub>FC) differential expression of CD, UC, IBD, Average Expression (AveExpr), F-statistics (F), P value and False Discovery Rate (FDR) adjusted P value (adj.P.Value). The threshold for DEMs were set as p value <0.05 and |log<sub>2</sub>FC| >1.

### Analysis of Target Genes of DEMs

The target genes, possessing miRBase (release 22.1, [www.mirbase.org](http://www.mirbase.org/)),<sup>[9]</sup> ID, of DEMs were analyzed using both validated and predicted targets obtained from the database miRWalk3.0 ([http://mirwalk.umm.uni-heidelberg.de](http://mirwalk.umm.uni-heidelberg.de/)).<sup>[10]</sup> Only those predicted targets were selected which overlapped all five databases including miRWalk, miRDB ([www.mirdb.org](http://www.mirdb.org/)),<sup>[11]</sup> miRanda ([www.microrna.org/microrna](http://www.microrna.org/microrna/)),<sup>[12]</sup> RNA22 (version 2.0, <https://cm.jefferson.edu/~rna22/>),<sup>[13]</sup> and TargetScan (Release 7.2, [www.targetscan.org](http://www.targetscan.org/)).<sup>[14]</sup>

### Construction of DEMs-gene Network

A regulatory network of DEMs associated to genes was constructed using cytoscape (<https://cytoscape.org/>) software, version 3.7.2.<sup>[15]</sup>

### PPI Network Analysis

A PPI network was constructed with the predicted and validated target genes of DEMs using STRING (Search Tool for the Retrieval of Interacting Genes) (<http://string-db.org/cgi/input.pl>; version 1.5.0) App of the Cytoscape software.<sup>[16]</sup> version 3.7.2 (<https://cytoscape.org/>).<sup>[15]</sup>

### Hub Genes Selection and Identification of Significant Modules

The top 10 genes in the PPI network were selected using the cytoscape plugin CytoHubba, version 0.1, ranked by maximal clique centrality (MCC) method.<sup>[17]</sup> Genes possessing higher scores in the MCC analysis were considered hub genes.

The cytoscape plugin MCODE (molecular complex detection) version 1.5.1 was applied to obtain significant modules associated with the PPI network with degree cutoff 2, node score cutoff 0.2, k-core 2 and maximum depth 100.<sup>[18]</sup>

## GO and Reactome Pathway Enrichment Analyses

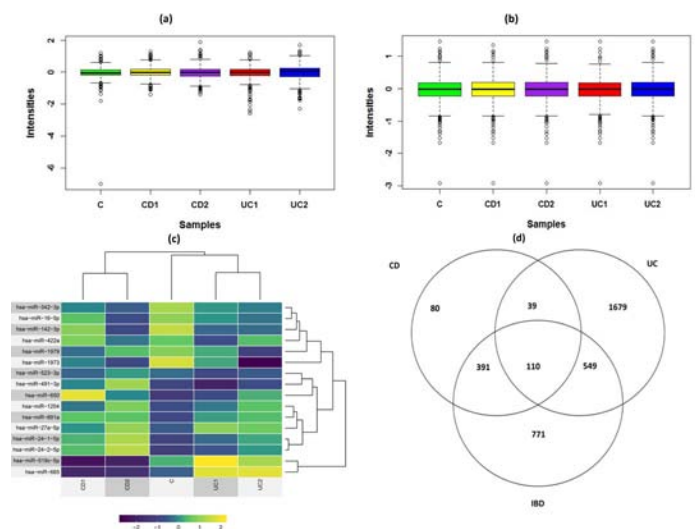
The DEGs were analyzed at the functional level for GO and Reactome pathway enrichment using the STRING Enrichment App of the cytoscape software, with p value <0.05 as the cutoff criterion.

## RESULTS

The miRNA microarray signal intensity downloaded from NCBI GEO dataset GSE53867 before and after normalization have been indicated in Figure 1a and Figure 1b.

Analysis of normalized GSE53867 expression data using GEO2R yielded 16 DEMs within CD types (CD1 and CD2), UC types (UC1 and UC2) and C samples as shown with a heatmap in Figure 1c. There were 7 upregulated DEMs including hsa-miR-24-2-5p, hsa-miR-650, hsa-miR-24-1-5p, hsa-miR-491-3p, hsa-miR-891a-5p, hsa-miR-27a-5p and hsa-miR-1204 in the CD group; 4 downregulated DEMs including hsa-miR-1973, hsa-miR-142-3p, hsa-miR-342-3p and hsa-miR-16-5p in the UC group; 10 DEMs including 8 upregulated (hsa-miR-491-3p, hsa-miR-24-2-5p, hsa-miR-650, hsa-miR-24-1-5p, hsa-miR-523-3p, hsa-miR-422a, hsa-miR-142-3p and hsa-miR-1973) and 2 downregulated (hsa-miR-519c-5p and hsa-miR-665) in the IBD group; miRNAs expressed in order of descending log<sub>2</sub>FoldChange (log<sub>2</sub>FC) (P value <0.05 and |log<sub>2</sub>FC| >1).

The number of target genes obtained from the databases were 725 (478 validated and 247 predicted), 3313 (3003 validated and 310 predicted) and 2201 (1654 validated and 547 predicted), in CD, UC and IBD, group respectively; after removal of duplicate target genes resulted into 620, 2377 and 1821 number of genes in the CD, UC and IBD, group respectively, an amount of which were further used for construction of regulatory network between the miRNAs and their targets. The miRNAs, hsa-miR-886-3p, hsa-miR-886-5p and hsa-miR-1979 were excluded from the DEMs-Gene network, due to non-availability of records in the miRNA database, miRBase and withdrawn status from Hugo Gene Nomenclature Committee (HGNC) (<https://www.genenames.org/>); hsa-miR-519d was excluded from the network due to lack of mature sequence miRNA in the dataset. Figure 1d represents the Venn diagram indicating a total of 620 DEGs in the CD group, 2377 DEGs in the UC group and



**Figure 1:** miRNA and gene expression in UC, CD and IBD.

Note: (a) miRNA microarray signal intensity before normalization; (b) miRNA microarray signal intensity after normalization; (c) Heatmap of normalized expression dataset, d. Venn diagram of differentially expressed genes. UC: Ulcerative colitis; CD: Crohn's disease; IBD: Inflammatory bowel disease.

1821 DEGs in the IBD group, out of which a total of 110 DEGs fitted all the three groups, while 149, 659 and 501 DEGs overlapped CD and UC, UC and IBD, CD and IBD respectively.

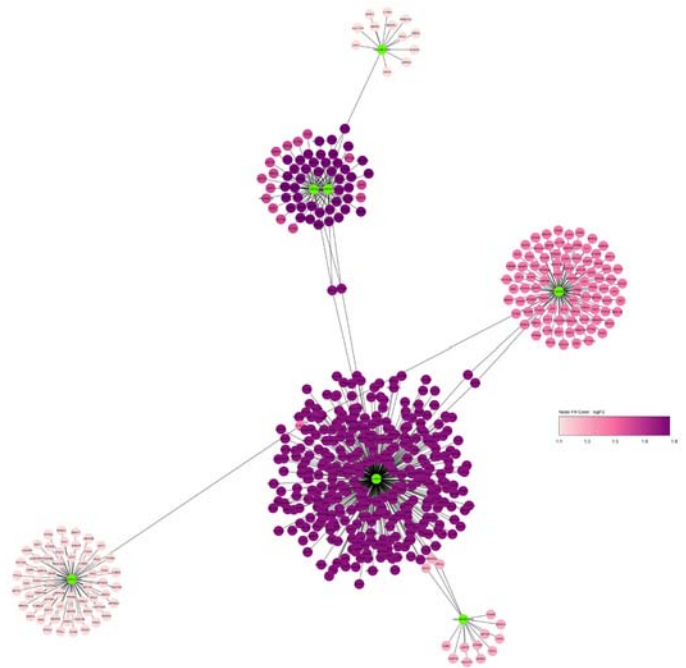
The upregulated and downregulated DEMs among CD, UC and IBD groups have been represented with volcano plots in Figure 2a, Figure 2b and Figure 2c, respectively. The results of the Empirical Bayes moderated *t* test have been represented graphically in terms of  $\log_2FC$ , Average Expression (AveExpr), F-statistics (F), *P* value (*P*.Value) and FDR (False Discovery Rate) adjusted *P* value (adj.*P*.Value) in Figure 2d.

The DEMs-Gene regulatory network of the CD, UC and IBD, group constructed using cytoscape, is depicted in Figure 3-5, respectively (confidence score cutoff 0.7).

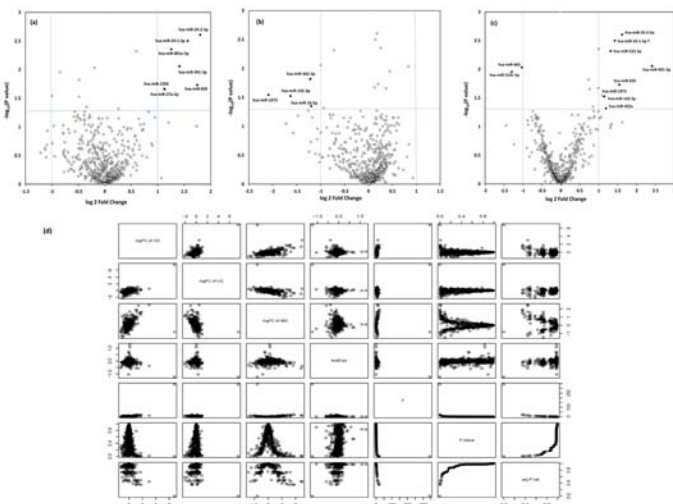
The PPI network visualized using Cytoscape showed 1147, 422, 1482, 1020 protein-protein interactions among 571, 300, 947, 762 proteins (confidence score cutoff 0.7 and  $P < 0.05$  as the cut-off criterion) in the CD, UC, upregulated and downregulated IBD, groups respectively. The top 10 hub genes in the CD and UC groups are represented in Figure 6a and Figure 6b, respectively with *MCC* scores ranging from 363039-362884 and 175-85, respectively. The score of *hub genes* by *MCC* was  $2.09E+13$  and  $1.31E+12$  in the upregulated and downregulated IBD group respectively, as shown in Figure 6c and Figure 6d, respectively. The top 10 upregulated hub genes of the CD group included SOCS3, WSB1, ASB6, UBE2D2, UBE2D3, CDC27, UBE2B, NEDD4L, UBR1, LTN1 and top 10 downregulated hub genes specific to the UC group included CIRH1A, PNO1, PWP2, UTP3, DNMT1, RSL1D1, RPSA, KDM6B, H2AFX. In the IBD group, the top 10 upregulated hub genes were SKP2, CDC27, SOCS3, CUL5, ASB4, WSB1, ASB6, UBE2D3, UBE2D2, UBE2B, while the top 10 downregulated hub genes included ASB1, WSB1, FBXW8, FBXW2, UBE2F, KLHL25, ASB11, UBA52, RPS27A, SMURF2. MCODE was used to establish major modules, which demonstrated 16, 7, 22 and 16 modules containing genes, number ranging from 22-5, 14-3, 44-3 and 33-3 respectively and score ranging from 10-2.5, 6.571-2.6, 17-2.8 and 16-3 in the CD, UC, IBD upregulated and IBD downregulated groups, respectively.

A functional enrichment of the screened DEGs in the CD groups (PPI enrichment value  $1.0E-16$ ), indicated that most of the DEGs were closely associated with protein ubiquitination and class I MHC mediated antigen processing reactome pathway related to ubiquitination and

proteasome degradation (FDR  $9.69E-1$ ) (Figure 7). The other GO terms and pathways included post-translational protein modification, negative regulation of intracellular signal transduction, neddylation, synthesis of active ubiquitin including roles of E1 and E2 enzymes in CD exhibited significant expression. Among the UC genes, GO process and reactome pathway enrichment were significantly related to rRNA processing and modification respectively (PPI enrichment value  $3.28E-10$ ) (Figure 7). In addition, other biological process significantly enriched by these genes were chromatin organization and remodeling, ribosomal small subunit assembly, double-strand break repair via nonhomologous end joining. However, other enriched pathways included oxidative stress induced senescence, DNA methylation, regulation of gene silencing,

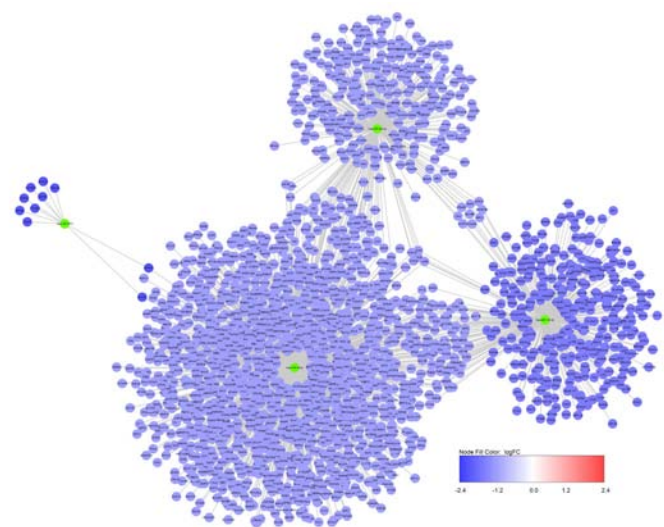


**Figure 3:** Differentially expressed microRNA-gene regulatory network in ulcerative colitis.



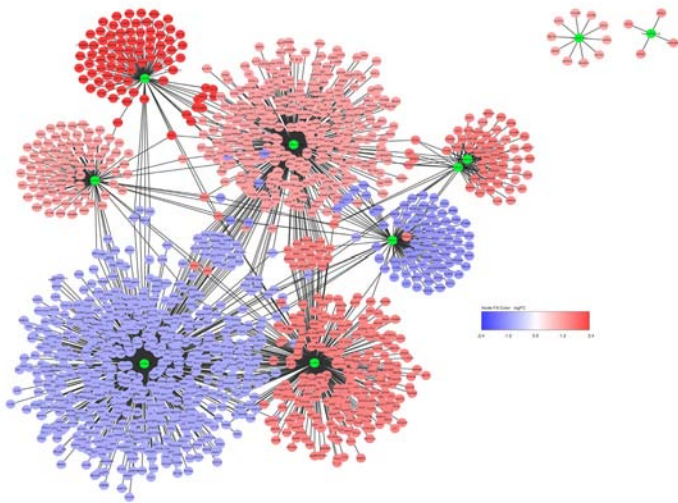
**Figure 2:** Volcano plots of the upregulated and downregulated differentially expressed microRNAs.

Note: (a) Crohn's disease; (b) Ulcerative colitis; (c) Inflammatory bowel disease; (d) Plot of Empirical Bayes moderated *t* test

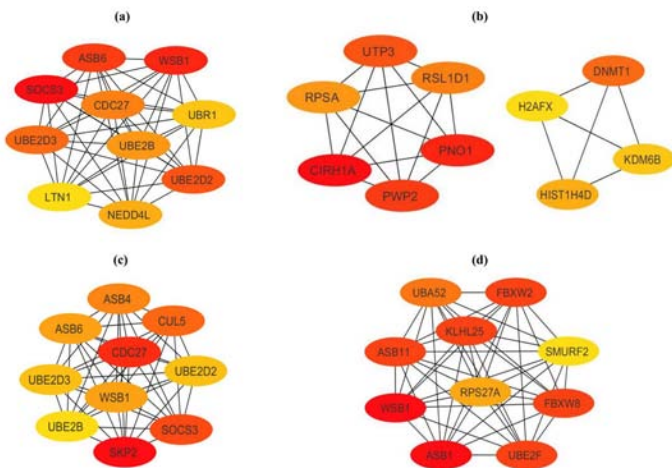


**Figure 4:** Differentially expressed microRNA-gene regulatory network in ulcerative colitis.

pallium development, positive regulation of DNA metabolic process in UC exhibited significant expression (Figure 7). In GO analysis, the top enriched biological processes in upregulation of IBD included protein ubiquitination and mitotic cell cycle phase transition (Figure 7). The top enriched pathways in upregulation of IBD included antigen processing related to ubiquitination and proteasome degradation; synthesis of active ubiquitin including roles of E1 and E2 enzymes. In addition, APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1 exhibited significantly enriched pathways. In GO analysis, the top enriched biological processes in downregulation of IBD included protein ubiquitination, post-translational protein modification, nucleotide-binding oligomerization domain containing signaling pathway, regulation of catabolic process (Figure 7). The top enriched pathways in downregulation of IBD included antigen processing related to ubiquitination and proteasome degradation, downregulation of SMAD2/3:SMAD4 transcriptional activity and Josephin domain deubiquitins. The representative KEGG pathways of statistically enriched DEGs in CD, IBD and UC are represented in Figure 8-10, respectively.



**Figure 5:** Differentially expressed microRNA-gene regulatory network in inflammatory bowel disease.

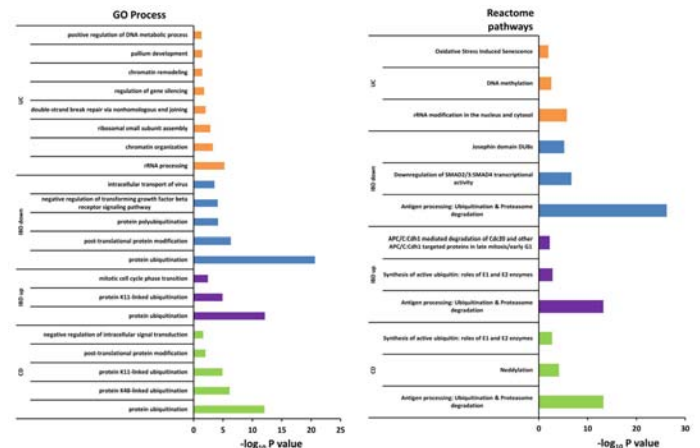


**Figure 6:** Top 10 hub genes. Note: (a) Crohn's disease; (b) Ulcerative colitis; (c) Inflammatory bowel disease upregulated; (d) Inflammatory bowel disease downregulated.

## DISCUSSION

Idiopathic IBD, which includes CD and UC, creates several diagnostic impediments owing to common clinical, radiographic, endoscopic and histologic characteristics. Thus it is important to recognize the subtypes of IBD for developing specific diagnostic and novel therapeutic intervention in IBD. The miRNAs are potential candidates for therapeutic application owing to well-defined mechanism of action during dysregulation in human IBD and simple oligonucleotide design.<sup>[19]</sup> This paper investigated the expression pattern of miRNAs in IBD in comparison to the other IBD subtypes, namely CD and UC, in order to identify miRNAs as potential diagnostic and prognostic biomarkers to better delineate the IBD subtypes.

In the present study, an analysis of the GSE53867 expression dataset revealed 16 DEMs (P value <0.05 and  $|\log_2FC| >1$ ), including 10 upregulated and 6 downregulated DEMs, in IBD and its subtypes. Among the 16 DEMs, 3 miRNAs including miR-891a-5p, miR-27a-5p, miR-1204 were unique to the CD group, 2 miRNAs such as miR-342-3p, miR-16-5p were unique to the UC group, 4 miRNAs including miR-523-3p, miR-422a, miR-665, miR-519c-5p belong exclusively to the IBD group. However, the common miRNAs under the UC and IBD were miR-1973, miR-142-3p and those under CD and IBD were miR-24-2-5p, miR-650, miR-24-1-5p, miR-491-3p; no common miRNAs belonged concomitantly to both UC and CD groups. miR-342-3p, miR-16-5p could be applied as a biomarker to identify UC; miRNAs including miR-523-3p, miR-422a, miR-665, miR-519c-5p could implicate the possibility of IBD diagnosis, while the presence of common miRNAs within CD and IBD, UC and IBD could be suggestive of overlapping epigenetic modifiers targetable by combined therapeutic applications. Occurrence of a similar few signature miRNAs included elevated miR-16 and miR-24, decreased miRNA miR-650, in mucosal tissue of UC; similar miRNAs miR-16 and miR-422a have been found significantly upregulated in the mucosal tissue of CD compared to normal healthy controls.<sup>[20-22]</sup> The panel of microRNAs that have been found differentially expressed between UC and CD in colonic tissue included miR-19b, miR-23b, miR-100a-3p, miR-100b-5p, miR-106a, miR-150, miR-191, miR-196b, miR-223, miR-320a and miR-629.<sup>[23]</sup> The colonic microRNA panels reported by different studies attempting to delineate UC, CD, between UC and CD exhibited little overlap, due to discordance in the activity status of IBD during sampling, location of tissue, inflammation level, previous medication, platforms for micro-RNAs estimation.<sup>[24]</sup>



**Figure 7:** Gene Ontology and pathway enrichment analyses of differentially expressed genes.

In the current study, genes which concurrently fitted 5 databases (miRWalk, miRDB, miRanda, RNA22 and TargetScan) were chosen as predicted target genes and validated target genes from miRWalk were taken, amounting to 1673 upregulated and 3145 downregulated genes. The miRWalk database contains information about predicted interactions based on integration of several algorithms and experimentally validated interactions as well (<http://mirwalk.umm.uni-heidelberg.de>).<sup>[10]</sup> Selecting interactions predicted by multiple tools ensures enhanced recovery of identical miRNA-gene interactions. The cytoscape tool was used to visualize the interaction data on the predicted and validated miRNA and their putative target gene regulatory network structure.<sup>[15]</sup> In the current study, there were 620 DEGs identified in the CD group targeted by 6 upregulated DEMs including miR-24 ( $n=98$ ), miR-650 ( $n=341$ ), miR-491 ( $n=92$ ), miR-891 ( $n=14$ ), miR-27 ( $n=62$ ) and miR-1204 ( $n=12$ ) ( $n$ = DEGs); 2377 DEGs identified in the UC group targeted by 4 downregulated DEMs including miR-1973 ( $n=10$ ), miR-142 ( $n=409$ ), miR-342 ( $n=358$ ) and miR-16 ( $n=1600$ ); 1821 DEGs targeted by 9 DEMs amongst which 7 upregulated DEMs including miR-491 ( $n=92$ ), miR-24 ( $n=98$ ), miR-650 ( $n=341$ ), miR-523 ( $n=4$ ), miR-422 ( $n=99$ ), miR-142 ( $n=409$ ), miR-1973 ( $n=10$ ) and 2 downregulated DEMs including miR-519 ( $n=95$ ), miR-665 ( $n=673$ ) identified in the IBD group. The UC and CD groups had 149 DEGs in common, while 110 DEGs fitted all the three groups, implying overlapping genetic constituent within CD, UC and IBD. The PPI network was visualized using cytoscape among 571, 300, 947, 762 proteins in the CD, UC, upregulated and downregulated IBD, groups respectively, followed by analysis of the most connected nodes, considered as key regulators of pathways and biological functions. Amongst the eleven topological analysis methods in *CytoHubba*, the MCC method, which was based on the degree of connectivity of each protein and the size of the interactions connected to the rest of the network, was proposed to have superior accuracy of identifying hub genes encoding proteins in the network.<sup>[17]</sup> Hub genes, being central elements of the PPI network, serve as prospective biomarkers, therapeutic targets and novel tool for analyzing crucial mechanisms regulating disease processes. Within the top hub genes of each group, in our study, the top 7 upregulated hub genes of the CD group namely, SOCS3, WSB1, ASB6, UBE2D2, UBE2D3, CDC27, UBE2B were also expressed as the upregulated hub genes of the IBD group, while the WSB1 gene from the CD group was concomitantly present in the downregulated hub genes of the IBD group.

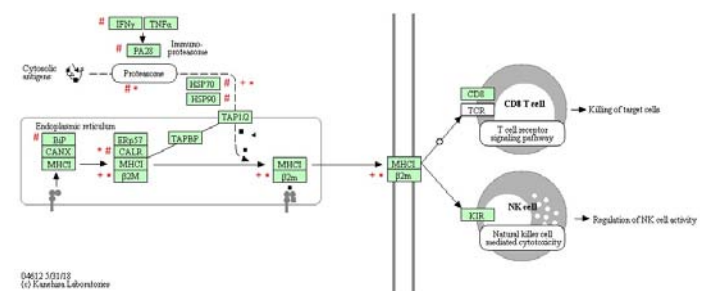
The top ranking hub gene of the CD group was SOCS3 (suppressor of cytokine signaling 3) upregulated by miR-650 ( $\log_2FC = 1.745$ ,  $P$  value = 0.018561). Cheng *et al.*<sup>[25]</sup> reported the role of miR-19b in inhibiting inflammatory response by downregulating SOCS3 to alter chemokine production in CD. The PPI from the cytoscape showed cytoplasmic expression of SOCS3 in most tissues and additionally in the plasma membrane; among blood cell types, it was specific to neutrophils. SOCS3 was expressed in the cells of epithelium and lamina propria in the colon in IBD, UC and CD cases.<sup>[26]</sup> SOCS3 as a member of the SSI (STAT-induced STAT inhibitor) family act to inhibit cytokine signal transduction through the JAK/STAT pathway.<sup>[27]</sup> SOCS3 also acted as a substrate recognition component of a SCF-like ECS (Elongin B/C-Cul2/Cul5-SOCS-box protein) E3 ubiquitin ligase complex to mediate ubiquitination and subsequent proteasomal degradation of cytokine receptors in inflammation.<sup>[28]</sup>

We showed that in the IBD group, the key gene SKP2 (S-Phase Kinase Associated Protein 2) was upregulated by miR-142-3p ( $\log_2FC = 1.145$ ,  $P$  value = 0.029848), while ASB1 (Ankyrin repeat and SOCS box protein 1) was downregulated by miR-665 ( $\log_2FC = -1.03$ ,  $P$  value = 0.009245). Li *et al.*<sup>[29]</sup> demonstrated the role of miR-665 in promoting apoptosis and colitis in IBD by inhibiting ER stress components XBP1

and ORMDL3. SKP2 was tissue enhanced in placenta (<https://www.proteinatlas.org>) with nuclear and cytoplasmic expression and low blood cell type specificity (PPI data); ASB1 was localized in the nucleoplasm (<https://www.proteinatlas.org>), had low tissue specificity and the protein expression was mainly in the CNS, adrenal medulla, parathyroid gland and enteroendocrine cells (<https://www.proteinatlas.org>).

In our study, the DEGs in both CD and IBD groups were significantly enriched (PPI enrichment 1.0E-16) in pathways such as class I MHC mediated antigen processing related to ubiquitination and proteasome degradation. Analysis of biological processes showed that the DEGs in both CD and IBD groups were significantly enriched in protein ubiquitination. Intracellular antigens and cytokines induce proteasome, immunoproteasome containing PA 28 and other accessory particles, for further degradation of the ubiquitinated substrates by an ATP-dependent mechanism. Peptides in association with HSP chaperones are translocated by TAP to the endoplasmic reticulum (ER) lumen, where they are placed on the MHC class I complex, made of heavy chain and  $\beta_2m$  (Figure 8). MHC I fold and assemble in ER, assisted by several ER chaperones. The MHC class I complexes present peptides on the cell surface, for recognition by CD8<sup>+</sup> T cells and NK cells. The SOCS3 and SKP2 genes, in the present study were up-regulated during the course of CD and IBD respectively concomitant with upregulation of pro-inflammatory cytokine TLR6 in both CD and IBD; upregulation of pro-inflammatory cytokines such as IL6, CCL22, IL-1A in IBD only, through activation of TLR- and TNF-signalling pathways via activation of NFKB (nuclear factor kappa B) pathway; JAK signaling pathway and cytokine-cytokine receptor interaction all leading to ubiquitin mediated proteolysis, by K63-linked polyubiquitin chains. The ubiquitin proteasome pathway, is essential for protein degradation carried out by three classes of enzymes E1, E2 and E3; E1 and E2 prepare ubiquitin chains that are then linked to proteins by the E3.<sup>[30]</sup> The SOCS3, SKP2 and ASB1 are components of the SOCSbox in ECS complex of E3 ubiquitin ligase containing CUL5, RNF7/RBX2, Elongin BC complex and SOCSbox.<sup>[28]</sup>

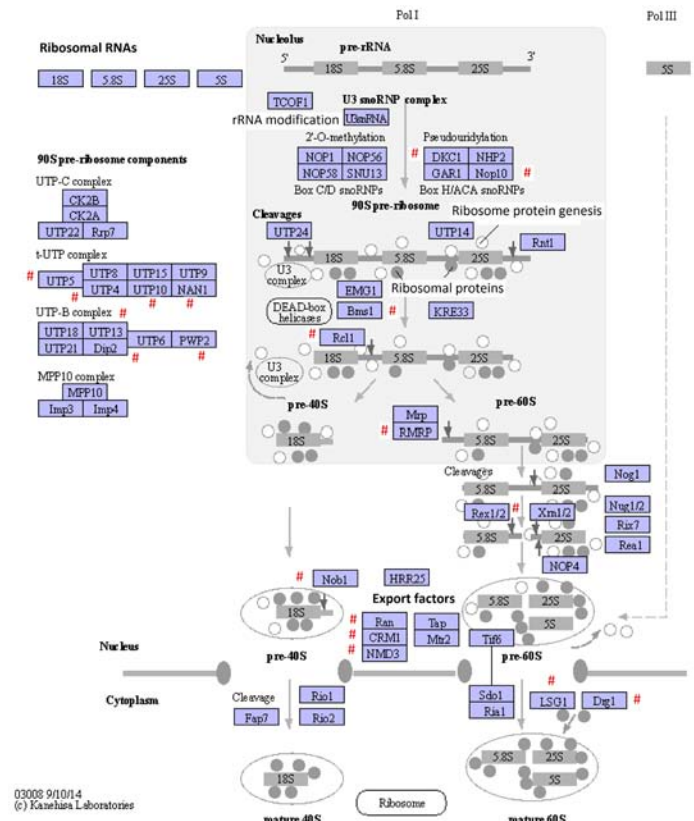
In the current study, in TLR (toll-like receptor) mediated signaling, the microbial components like flagellin, peptidoglycan and lipopolysaccharide are recognized by TLR6 (paralog of TLR2), which in association with TLR2, trigger the recruitment of the adaptor protein MyD88 (Myeloid Differentiation Primary Response 88), the TOLLIP (Toll Interacting Protein) and TIRAP (Toll/Interleukin-1 Receptor Domain-Containing Adapter Protein), through interleukin-1 receptor-associated kinases IRAK1 and IRAK4, in combination with RNF152 (Ring Finger Protein 152/ RING-Type E3 Ubiquitin Transferase RNF152) to stimulate K63-linked polyubiquitination. This complex, in turn, creates a binding platform for the assembly of the signaling genes, the TAK1, TAB1, TAB2, the TGF- $\beta$ -activated kinase 1, TAK1-binding protein 1, 2 respectively, to cause the attachment of linear polyubiquitin



**Figure 8:** KEGG pathways of statistically enriched differentially expressed genes: Antigen processing and presentation-MHC I pathway. Note: Genes expressed in Crohn's disease (+), ulcerative colitis (#) and inflammatory bowel disease (\*).

chains on NFKB (Nuclear Factor Kappa B) essential modifier (NEMO) from the inhibitor of Kappa B kinase (IKBK) complex comprising CHUK (conserved helix-loop-helix ubiquitous kinase), IKBK-β/IKKB (Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta) and NEMO (Figure 9). The NEMO from the IKBK complex mediate K63-linked polyubiquitin binding on RNF152 to facilitate IKBK-β activation and successive phosphorylation of inhibitor of Kappa B alpha (IKBA) for ubiquitination with K63 linkages by the E3 ligase ECS complex to induce IKB degradation. In ubiquitin-mediated JAK signaling, cytokine-induced receptor dimerization results in tyrosine phosphorylation of JAK2 followed by binding of SOCS protein at its SH2 domain leading to ECS complex mediated K63 polyubiquitination and proteasomal degradation of JAK2 and termination of JAK signaling process. In TNF signalling pathway herein, the tumor necrosis factor (TNF) binds to the tumor necrosis factor receptor (TNFR)1, to trigger the recruitment of TRADD (TNFR1-Associated Protein), TRAF2/5 (TNF Receptor Associated Factor 2/5), RIP1 [Receptor (TNFRSF)-Interacting Serine-Threonine Kinase 1]. This, in turn, causes binding of TAK1, TAB1 and TAB2 ultimately leading to NFKB signaling pathway followed by K63-linked ubiquitin mediated proteolysis of IKBA.

Among the genes screened from the differentially expressed genes in UC, CIRH1A (cirrhosis, autosomal recessive 1A), in the present study, were downregulated by miR-16-5p ( $\log_2FC = -1.195$ ,  $P$  value = 0.04479). CIRH1A showed low blood cell type and tissue specificity (PPI data) with nucleolar expression in several tissues (<https://www.proteinatlas.org>), mainly in neuronal cells. CIRH1A, encode WD40-repeat-containing protein, involved in nucleolar processing of small subunit pre-18S rRNA. In the GO and pathway analysis, the top enriched pathways in UC included rRNA processing and modification in the nucleus and cytosol. The CIRH1A gene, specific to the UC group, was down regulated concomitant with downregulation of U3 snoRNP complex, 90S pre-ribosome components, pre-40S, pre-60S and export factors components (Figure 10). Impairment of ribosome biogenesis induce ER stress leading to proinflammatory cytokine production via NFKB activation with IKK (IKBA kinase) activity maintained at basal level via IRE1 (inositol-requiring ER-to-nucleus signal kinase 1), an ER stress sensor.<sup>[31]</sup> The significantly downregulated genes in UC, in

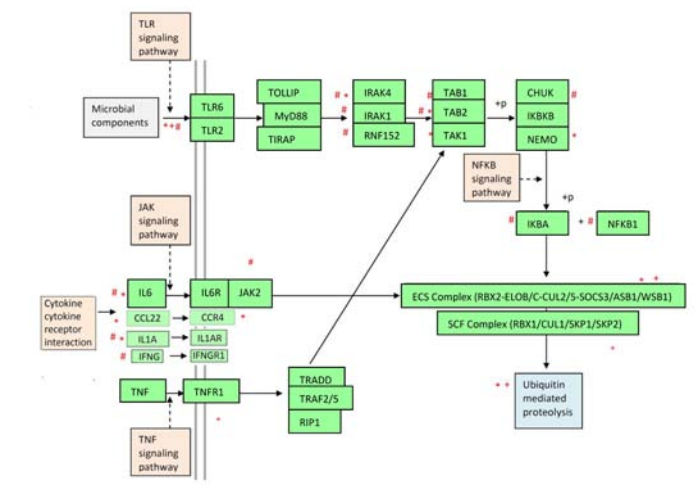


**Figure 10:** KEGG pathways of statistically enriched differentially expressed genes: Ribosome biogenesis. Note: Genes expressed in Crohn's disease (+), ulcerative colitis (#) and inflammatory bowel disease (\*).

the present study, included receptor families such as TLR6 via NFKB signaling pathway; downregulation of pro-inflammatory cytokines such as IL6, IL-1A, IFNG through activation of JAK signaling pathway and cytokine-cytokine receptor interaction. IL-6 enhanced upregulation of rRNA transcription stimulated the MDM2 (mouse double minute 2 homolog)-mediated p53 proteasomal digestion, by reducing the availability of ribosome proteins for MDM2 binding.<sup>[32]</sup>

**CONCLUSION**

It was found that the SOCS3, CIRH1A, SKP2 and ASB1 genes might play predominant role in the pathogenesis of CD, UC, IBD during upregulation and downregulation respectively. The SOCS3 was upregulated by miR-650 in CD, induced by various cytokines, through the NFKB signalling pathway to mediate ubiquitination and subsequent proteasomal degradation of cytokine receptors in inflammation. CIRH1A gene downregulated by miR-16 was specific to the UC group, acted by impairing ribosome biogenesis to induce ER stress via NFKB signaling pathway, JAK signaling pathway and cytokine-cytokine receptor interaction all leading to ubiquitin mediated proteolysis. The SOCS3, CIRH1A, SKP2 and ASB1 genes might serve as valuable biomarkers to differentiate CD, UC and upregulated IBD and downregulated IBD respectively.



**Figure 9:** KEGG pathways of statistically enriched differentially expressed genes: TLR, TNF signalling pathways. Note: Genes expressed in Crohn's disease (+), ulcerative colitis (#), and inflammatory bowel disease (\*). TLR, TNF signalling pathways via activation of NFKB pathway; JAK signaling pathway and cytokine-cytokine receptor interaction, all leading to ubiquitin mediated proteolysis.

## ACKNOWLEDGEMENT

The corresponding author acknowledges Research Square for posting preprint of the article, DOI: 10.21203/rs.3.rs-27227/v2

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## ABBREVIATIONS

**IBD:** Inflammatory Bowel Diseases; **DEM:** Differentially Expressed Micrnas; **GO:** Gene Ontology; **CD:** Crohn's Disease; **UC:** Ulcerative Colitis; **miRNA:** Microna; **mRNA:** Messenger RNA; **UTR:** Untranslated Region; **DEG:** Differentially Expressed Genes; **PPI:** Protein-Protein Interaction; **GEO:** Gene Expression Omnibus; **MCODE:** Molecular Complex Detection; **MCC:** Maximal Clique Centrality; **HGNC:** Hugo Gene Nomenclature Committee; **FDR:** False Discovery Rate; **SSI:** STAT-Induced STAT Inhibitor; **SKP2:** S-Phase Kinase Associated Protein 2; **ASB1:** Ankyrin Repeat And SOCS Box Protein 1; **TLR:** Toll-Like Receptor; **Myd88:** Myeloid Differentiation Primary Response 88; **TOLLIP:** Toll Interacting Protein; **TIRAP:** Toll/Interleukin-1 Receptor Domain-Containing Adapter Protein; **NFKB:** Nuclear Factor Kappa B; **IKBK:** Inhibitor Of Kappa B Kinase; **CHUK:** Conserved Helix-Loop-Helix Ubiquitous Kinase; **IKBK-B:** Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta; **TNF:** Tumor Necrosis Factor; **TNFR1:** Tumor Necrosis Factor Receptor 1, **TRADD:** TNFR1-Associated Protein; **TRAF2/5:** TNF Receptor Associated Factor 2/5.

## REFERENCES

- GBD-Inflammatory Bowel Disease Collaborators. Inflammatory bowel disease collaborators, the global, regional and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol.* 2020;5:17-30.
- Kedia S, Ahuja V. Epidemiology of inflammatory bowel disease in India: The great shift east. *Inflamm Intest Dis.* 2017;2(2):102-15.
- Ramos GP, Papadakis KA. Mechanisms of disease: Inflammatory bowel diseases. *Mayo Clin Proc.* 2019;94(1):155-65.
- Schaefer JS. MicroRNAs: How Many in Inflammatory Bowel Disease?. *Curr Opin Gastroenterol.* 2016;32:258-66.
- Bartel DP. MicroRNAs: Target recognition and regulatory functions. *Cell.* 2009;136:215-33.
- Soroosh A, Koutsoumpa M, Pothoulakis C, Iliopoulos D. Functional role and therapeutic targeting of microRNAs in inflammatory bowel disease. *Am J Physiol Gastrointest Liver Physiol.* 2018;314(2):G256-62.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomshesky M, et al. NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res.* 2013;41:D991-5.
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.* 2015;12(2):115-21.
- Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res.* 2004;32:D109-11.
- Dweep H, Gretz N. MiRWalk 2.0: A comprehensive atlas of microRNA-target interactions. *Nat Methods.* 2015;12(8):697.
- Nathan W, Xiaowei W. MiRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 2015;43(D1):D146-52.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: Targets and expression. *Nucleic Acids Res.* 2008;36(Suppl\_1):D149-53.
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell.* 2006;126(6):1203-17.
- Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* 2015;4:e05005.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
- Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape stringapp: Network analysis and visualization of proteomics data. *J Proteome Res.* 2019;18(2):623-32.
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY, et al. CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8(S4):S11.
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4(1):1-27.
- Ling H. Non-coding RNAs: Therapeutic strategies and delivery systems. *Adv Exp Med Biol.* 2016;937:229-37.
- Wu F, Zikusoka M, Trindade A, Dassopoulos T, Harris ML, Bayless TM, et al. MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. *Gastroenterol.* 2008;135(5):1624-35.
- Iborra M, Bernuzzi F, Correale C, Vetrano S, Fiorino G, Beltran B, et al. Identification of serum and tissue micro-RNA expression profiles in different stages of inflammatory bowel disease. *Clin Exp Immunol.* 2013;173(2):250-8.
- Wu F, Zhang S, Dassopoulos T, Harris ML, Bayless TM, Meltzer SJ, et al. Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm Bowel Dis.* 2010;16(10):1729-38.
- Fasseu M, Treton X, Guichard C, Pedruzzi E, Cazals-Hatem D, Richard C, et al. Identification of restricted subsets of mature microRNA abnormally expressed in inactive colonic mucosa of patients with inflammatory bowel disease. *PLoS One.* 2010;5(10):e13160.
- Kurt F, Jingmei L. MicroRNA in inflammatory bowel disease: Translational research and clinical implication. *World J Gastroenterol.* 2015;21(43):12274-82.
- Cheng X, Zhang X, Su J, Zhang Y, Zhou W, Zhou J, et al. MiR-19b downregulates intestinal SOCS3 to reduce intestinal inflammation in Crohn's disease. *Sci Rep.* 2015;5:10397.
- Suzuki A, Hanada T, Mitsuyama K, Yoshida T, Kamizono S, Hoshino T, et al. CIS3/SOCS3/SSI3 plays a negative regulatory role in STAT3 activation and intestinal inflammation. *J Exp Med.* 2001;193(4):471-81.
- Sasaki A, Yasukawa H, Suzuki A, Kamizono S, Syoda T, Kinjo I, et al. Cytokine-inducible SH2 protein-3 (CIS3/SOCS3) inhibits Janus tyrosine kinase by binding through the N-terminal kinase inhibitory region as well as SH2 domain. *Genes Cells.* 1999;4(6):339-51.
- Kamura T, Maenaka K, Kotoshiba S, Matsumoto M, Kohda D, Conaway RC, et al. VHLbox and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev.* 2004;18(24):3055-65.
- Li M, Zhang S, Qiu Y, He Y, Chen B, Mao R, et al. Upregulation of miR-665 promotes apoptosis and colitis in inflammatory bowel disease by repressing the endoplasmic reticulum stress components XBP1 and ORMDL3. *Cell Death Dis.* 2017;8(3):e2699.
- Stewart HL, Alfred LG, William EM. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J American Soc Nephrol.* 2006;17(7):1807-19.
- Moon Y. Ribosomal alteration-derived signals for cytokine induction in mucosal and systemic inflammation: Noncanonical pathways by ribosomal inactivation. *Mediat Inflamm.* 2014;2014:1-10.
- Su H, Kang Q, Wang H, Yin H, Duan L, Liu Y, et al. Changes in expression of p53 and inflammatory factors in patients with ulcerative colitis. *Exp Therapeutic Med.* 2019;17(4):2451-6.

**Cite this article:** Mandal M, Mandal S. Computational Approaches towards Micro-RNA Expression Analysis for Identification of Biomarkers in Inflammatory Bowel Disease Subtypes. *Int J Clin Exp Physiol.* 2020;7(3):100-6.