OXFORD

Original Article

Colonic Mucosal Transcriptomic Changes in Patients with Long-Duration Ulcerative Colitis Revealed Colitis-Associated Cancer Pathways

Eden Ngah Den Low,[a] Norfilza Mohd Mokhtar,[a] Zhiqin Wong,[b] Raja Affendi Raja Ali[b]

[a]Department of Physiology, Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur, Malaysia [b]Gastroenterology Unit, Department of Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur, Malaysia

Corresponding author: Professor Dr Raja Affendi Raja Ali, MD, FRCP, Gastroenterology Unit, Department of Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia. Tel: 603-9145-6094; Fax: 603-9145-6679; Email: draffendi@ppukm.ukm.edu.my

Abstract

Background and aims: Patients with ulcerative colitis [UC] with long disease duration have a higher risk of developing colitis-associated cancer [CAC] compared with patients with short-duration UC. The aim of this study was to identify transcriptomic differences associated with the duration of UC disease.

Methods: We conducted transcriptome profiling on 32 colonic biopsies [11 long-duration UC, ≥20 years; and 21 short-duration UC, ≤5 years] using Affymetrix Human Transcriptome Array 2.0. Differentially expressed genes [fold change > 1.5, $p < 0.05$] and alternative splicing events [splicing index > 1.5, $p < 0.05$] were determined using the Transcriptome Analysis Console. KOBAS 3.0 and DAVID 6.8 were used for KEGG and GO analysis. Selected genes from microarray analysis were validated using qPCR.

Results: There were 640 differentially expressed genes between both groups. The top ten upregulated genes were *HMGCS2, UGT2A3* isoforms, *B4GALNT2, MEP1B, GUCA2B, ADH1C, OTOP2, SLC9A3*, and *LYPD8*; the top ten downregulated genes were *PI3, DUOX2, VNN1, SLC6A14, GREM1, MMP1, CXCL1, TNIP3, TFF1*, and *LCN2*. Among the 123 altered KEGG pathways, the most significant were metabolic pathways; fatty acid degradation; valine, leucine, and isoleucine degradation; the peroxisome proliferator–activated receptor signalling pathway; and bile secretion, which were previously linked with CAC. Analysis showed that 3560 genes exhibited differential alternative splicing between long- and short-duration UC. Among them, 374 were differentially expressed, underscoring the intrinsic relationship between altered gene expression and alternative splicing.

Conclusions: Long-duration UC patients have altered gene expressions, pathways, and alternative splicing events as compared with short-duration UC patients, and these could be further validated to improve our understanding of the pathogenesis of CAC.

Key Words: Inflammatory bowel disease; ulcerative colitis; transcriptome; microarray analysis; long duration

1. Introduction

Ulcerative colitis [UC] is a subtype of inflammatory bowel disease [IBD]. IBD is becoming more common in many parts of the world, including Malaysia.[1,2] Malaysia has a multiethnic population comprised of Malays, Chinese, and Indians. An increasing incidence of IBD in Malaysia has been observed over the past two decades [from 0.07 to 0.69 per 100 000 person-years], and the highest prevalence has been observed among Indians at 24.91 as compared with 7.0 and 6.9 per 100 000 persons among the Malays and Chinese, respectively.[3] The pathogenesis of IBD involves a complex interaction between the host genetic background, microbial shifts, and environmental cues, leading to inappropriate chronic activation of the mucosal immune system.[4–6] Despite decades of effort, the complex pathogenesis of IBD is not yet fully understood.

Various transcriptomic studies have been carried out in the IBD field, including several efforts to discriminate between disease subtypes [Crohn's disease vs UC], disease phase [active/mildly active/inactive],[7,8] colon biopsy locations,[9] genders,[10] etc. Some of these studies have also been previously reviewed.[11] However, to date, the effects of different disease duration on transcriptome have yet to be explored. The type of colon cancer that is preceded by clinically detectable IBD is called colitis-associated cancer [CAC]. A number of population studies in various parts of the world have shown that long duration of UC is one of the risk factors for the development of CAC. A recent review discussed in detail the clinicopathological and molecular features of colorectal neoplastic lesions complicating IBD, zooming in on aspects like morphology of dysplasia in IBD, clinicopathological specificities of colorectal cancer [CRC] complicating IBD, molecular specificities of CAC as compared with sporadic CRC, the role of chronic intestinal inflammation in colorectal carcinogenesis, the role of oxidative stress, and chemoprevention and therapeutic perspectives.[12]

Colonoscopy detection of neoplasia has long been employed as the standard procedure. To date, there are no clinically sensitive molecular biomarkers that are widely used to aid the detection of CAC at an early stage. Our study aimed to identify transcriptomic changes, including gene expression and alternative splicing [AS], in inflamed colonic biopsies of patients associated with UC of different disease duration, and to provide a foundation study in the development of molecular biomarkers to detect mucosal neoplastic changes.

2. Materials and Methods

2.1. Patient population

This study has been approved by the institutional ethical committee [UKM PPI/111/8/JEP-2016–357]. All patients that came into the Endoscopy Unit of Hospital Canselor Tuanku Muhriz Universiti Kebangsaan Malaysia from 2015 to 2017 were screened for the inclusion criteria. Patients with an established diagnosis of active UC were recruited [after written consent forms were completed] and stratified into two groups: long-duration UC [≥20 years] and short-duration UC [≤5 years]. Patients who were newly diagnosed were also included in the short-duration category. The disease activity was assessed using the UC disease activity index [UCDAI].[13] Histologically, all these patients were analyzed using Geboes score and endoscopically analyzed using partial Mayo index score. Clinical characteristics and demographic details of the recruited patients were collected.

2.2. Sample collection

We only used inflamed colonic biopsies removed from the descending colon of patients with UC during the routine colonoscopy procedure. The biopsies were stored in RNAlater® and held at –80°C until further analysed.

2.3. RNA extraction and quality assessment of RNA

All biopsies collected were free of evidence of colitis-associated dysplasia or neoplasia based on the histopathological examination. Total RNA extraction was performed using an AllPrep DNA/RNA/miRNA universal kit and QIAshredder columns [Qiagen, Hilden, Germany] according to the manufacturer's instructions. RNA quality and quantity were checked with a DeNovix DS-11 Spectrophotometer [DeNovix, Wilmington, USA], a 2100 Bioanalyzer, and an RNA 6000 Nano kit [Agilent, CA, USA]. Only RNA with an integrity number of >7 was selected for transcriptome profiling.

2.4. Human transcriptome array 2.0

Microarray analysis was conducted using Human Transcriptome Array 2.0 [HTA 2.0; Affymetrix, CA, USA] based on the manufacturer's protocol. Briefly, approximately 1000 ng of total cellular RNA samples was converted to cDNA using a custom-designed random primer [containing the T7 polymerase promoter region] and conventional enzymatic steps. The T7 polymerase was used to produce and amplify antisense cRNA, as a starting material for producing double-stranded labelled cDNA for hybridization. Random primers were then annealed to the cRNA. The first and second strand synthesis reactions were performed using deoxyribonucleotide triphosphates [dNTPs] with both thymine and uracil at a ratio of 4:1. The double-stranded cDNA was fragmented by uracil DNA glycosylase, and the digested fragment was labelled with deoxynucleotidyl transferase and the biotin-conjugated nucleotide analogue. The samples were hybridized to a GeneChip® Human Transcriptome Array 2.0 overnight. The array was washed, stained, and scanned using an Affymetrix Fluidics Station and GeneChip® Scanner.

2.5. Validation of microarray by quantitative PCR

Validation assays were performed using quantitative PCR [qPCR] on additional frozen tissue from UC patients to confirm specific differential expression in short- and long-duration UC for selected genes. We used the subset of samples for the transcriptome array experiment [23 samples], with an additional eight samples that fulfilled the inclusion criteria. From the list of differentially expressed genes from HTA 2.0, nine genes [*CHP2*, *ENTPD5*, *AKR1B10*, *HOXA7*, *ABCG2*, *SGK2*, *SCIN*, *CDKN2B*, *TFF1*] related to cancer pathways were selected for validation by qPCR. Primers of these genes were commercially designed and coated on the 96-well plate by Qiagen [Qiagen, Hilden, Germany]. The sequences were propriety information and were not revealed by the company. Approximately 1 μg of RNA was used for cDNA synthesis, performed using the RT² First Strand Kit [Qiagen, Hilden, Germany] according to the manufacturer's protocol. A total of 12.5 μL of RT² SYBR Green Mastermix [Qiagen, Hilden, Germany] in each well was used. The cDNA was then amplified using the RT² qPCR mastermix with customized RT² PCR Arrays [cat. no. 330171 CLAH26710, Qiagen, Hilden, Germany]. The initial heating was done at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Melting curve analysis was done at 95°C for 1 min, 65°C for 2 min, proceeded with an

increased temperature from 65°C to 95°C in increments of 2°C/min. Peptidylprolyl isomerase A [*PPIA*] was used as the housekeeping gene for results normalization. Genomic DNA control, reverse transcription controls, and positive PCR controls were added in each qPCR run of the samples. All qPCR reactions were performed using the Bio-Rad CFX96 [Bio-Rad Laboratories, CA, USA]. Analysis of relative gene expression was done by the $2^{-\Delta\Delta CT}$ method.[14]

2.6. Statistical analysis for differentially expressed genes

Raw data from Human Transcriptome Array 2.0 were processed with the Expression Console software [Affymetrix, CA, USA] for background correction and normalization as recommended by the manufacturer. Transcriptome Analysis Console [TAC] 3.1 software [Affymetrix, CA, USA] provided by the Affymetrix Corporation was then used to analyse significantly differentially expressed genes [fold change ≥±1.5] using a default algorithm one-way between-subject ANOVA [unpaired] and a filter criteria ANOVA *p*-value [Condition pair] < 0.05.

2.7. Alternative splicing analysis

The design of Human Transcriptome Array 2.0 [Affymetrix, CA, USA], with approximately ten probes per exon and four probes per exon–exon splice junction, enabled an analysis of expression to be performed at the exon level. The AS analysis was carried out using default parameters. The expression of each exon [corresponding to a Probe Selection Region, PSR] in two different conditions [long-duration UC versus short-duration colonic biopsies] was compared by the software. The splicing index [SI] of a PSR is calculated from the ratio of the PSR signal intensity under two sets of conditions after normalization by the gene expression levels, i.e. [Intensity of Exon X in long-duration UC colonic biopsies:Intensity of Gene *Y* in long-duration UC colonic biopsies]:[Intensity of Exon X in short-duration UC colonic biopsies:intensity of Gene *Y* in short-duration UC colonic biopsies]. A positive index describes exons prevalently expressed in long-duration UC colonic biopsies, whereas a negative index shows a prevalent expression in short-duration UC colonic biopsies. Significant results were defined as SI ≥ ±1.5 and *p* ≤ 0.05.

2.8. Gene ontology and pathway analysis

To identify significantly enriched functional pathways, the Kyoto Encyclopedia of Genes and Genomes[15–17] [KEGG, http://www.genome.jp/kegg/] pathways were analysed using the KEGG Orthology Based Annotation System 3.0[18,19] [KOBAS 3.0, http://kobas.cbi.pku.edu.cn/]. To identify significantly enriched gene ontology [GO] terms, bioinformatics resource The Database for Annotation, Visualization and Integrated Discovery version [DAVID] 6.8[20,21] [http://david.abcc.ncifcrf.gov/tools.jsp] was used.

3. Results

A total of 32 patients, with 11 having long-duration UC [≥ 20 years] and 21 having short-duration UC [≤5 years] were included in the study. The median age of patients with short- and long-duration UC was 36 and 65 years, respectively. Of the 32 patients, 31 came with an established UC diagnosis, and a single patient was newly diagnosed. Short-duration UC patients were predominantly men, whereas long-duration UC patients were included more women than

men. In terms of ethnicity, Malays were found more in the short-duration UC group, but Indians were more commonly in the long-duration UC group. Most of the patients in both categories were non-smokers. The mean disease durations for short- and long-duration UC were 3.38 ± 1.36 years and 26.36 ± 5.31 years, respectively. Patients were either diagnosed as having pancolitis or left-sided colitis, and had a UCDAI score of 4–7. In addition, the long-duration UC group had a score of 2–4 based on the partial Mayo index score, but the short-duration UC group had a score of 3–5. In the study group of patients, the disease was of mild to moderate severity. Histologically, both groups of patients had Grade 1 to 2B UC based on Gebeos score. Most patients from both categories were on 5ASA up till the time of the sample collection. None of the patients had a family history of CRC or a personal history of primary sclerosing cholangitis. Patient characteristics are shown in Table 1.

3.1. Global gene expression profiling revealed effects of disease duration

Human Transcriptome Array 2.0 analysis was performed on the biopsies of UC patients with different disease durations. When comparing between long-duration and short-duration UC, a total of 640 transcripts [503 coding and 137 non-coding] were significantly differentially expressed (fold change [linear] ≥±1.5, *p* < 0.05, Figure 1A, B). Among these, 477 of the transcripts were upregulated and 163 were downregulated [Table 2] in long- versus short-duration UC. In Figure 1B, the heatmap shows that Group 1 [orange, long duration] samples were clustered together but not very well separated from Group 2 samples [green, short duration], because this is a comparison between two groups with disease, not a group with disease versus healthy group comparison. The complete list for differentially expressed transcripts can be found at Supplementary Table 1.

The top ten upregulated and top ten downregulated genes in long- versus short-duration UC are listed in Table 3, with fold change ranging from –16.31 to 22.33. These genes were mainly reported to be linked with metabolic and drug metabolism pathways [Supplementary Table 2]. The expression of nine genes from the microarray data was validated using qPCR [Figure 1C]. As for the non-coding transcripts, the top ten transcripts that were upregulated are listed in Table 4, with fold change ranging from –9.82 to 21.42.

3.2. Gene ontology and KEGG analysis of the coding transcripts

A GO term enrichment analysis was performed to functionally characterize the genes, which were found to be differentially expressed in the different disease durations. For each of the three major GO categories, the top three GO terms associated with the filtered genes included antioxidant activity, metalloendopeptidase activity, and calcium ion binding for 'molecular function' [Figure 2A]; fatty acid beta-oxidation, metabolic process, and inflammatory response for 'biological process' [Figure 2B]; extracellular exome, plasma membrane, and basolateral plasma membrane for 'cellular component' [Figure 2C].

Using the online platform KOBAS 3.0, 123 significant KEGG pathways [*p* < 0.05] based on differentially expressed coding transcripts were found. The most significant pathways included metabolic pathways; fatty acid degradation/short-chain fatty acid [propanoate and butanoate] metabolism; valine, leucine, and isoleucine degradation; the peroxisome proliferator-activated receptor

Table 1. Clinical and demographic details of the recruited patients. All data are expressed as *n* [%] except where indicated in the table.

	Long-duration UC	Short-duration UC
Age, median years [range]	65 [52–76]	36 [22–56]
Gender		
Male	5 [45.5]	14 [66.7]
Female	6 [54.5]	7 [33.3]
Race		
Malay	3 [27.3]	17 [80.95]
Chinese	3 [27.3]	3 [14.29]
Indian	5 [45.4]	1 [4.76]
Smoking status		
Yes	1 [9.1]	2 [9.5]
No	10 [90.9]	14 [66.7]
Ex-smoker	0	5 [23.8]
Mean duration of disease [Years ± SD]	26.36 ± 5.31	3.38 ± 1.36
Extent of disease		
Pancolitis	7 [63.6]	9 [42.9]
Left-sided	4 [36.4]	12 [57.1]
UCDAI score [range]	4–7	5–7
Partial Mayo index score	2–4	3–5
Treatment		
5ASA	8 [72.7]	13 [61.9]
Azathioprine	0	1 [4.8]
Steroids	0	1 [4.8]
5ASA + Azathioprine	3 [27.3]	2 [9.5]
5ASA + Steroids	0	1 [4.8]
5ASA + Azathioprine + Steroids	0	3 [14.3]
Family history of CRC	0	0
Primary sclerosing cholangitis	0	0

[PPAR] signalling pathway; and bile secretion [Table 5]. UC with different disease duration showed significant differences in the metabolic pathways, and this pattern was also reflected in the abovementioned GO molecular function terms.

3.3. The impact of disease duration on alternative splicing

To address the impact of disease duration on AS in UC, we performed a genome-wide splicing events profile analysis on long-duration UC vs short-duration UC. We identified 3560 transcripts that underwent significant splicing events [$p < 0.05$] with a splicing index of ≥1.5 or ≤–1.5. We found that the majority of the top ten upregulated and top ten downregulated coding genes [18/20] also showed significant differences in AS [Table 3], while only a quarter of the top ten upregulated and top ten downregulated non-coding genes [5/20] showed this. More than half of these annotated AS events [1228] corresponded to cassette exons [57.1%], followed by the second most common class of disease duration–regulated AS events which was alternative 3′ acceptor site [18.7%]. The rest included alternative 5′ donor site [18.3%], intron retention [5.7%], and mutually exclusive exons [0.2%, Figure 3A]. These results demonstrated that long duration of UC changed not only the transcript level, but also the AS patterns. It was also noted that 392 out of 3560 alternatively spliced transcripts were also significantly differentially expressed. This showed that there is a possibility that an intrinsic association exists between gene expression and AS [Figure 3B].

4. Discussion

This study has identified significant differentially expressed genes and genes with differential splicing between long- and short-duration UC. In addition, significant pathways and GO terms associated with coding genes with expression changes have also been identified. We have also obtained some data on genes for which splicing was affected by disease duration difference.

As none of the UC patients had a family history of CRC, the risk of them developing CAC is unpredictable and should be closely monitored. As the study compared samples from patients with disease [but with different duration of the disease], not samples of patients with disease versus samples from healthy patients, the small number of differentially expressed genes was expected.

Among the top upregulated genes, *HMGCS2* encodes a protein that belongs to the HMG-CoA synthase family. The protein is a mitochondrial enzyme that catalyses the first reaction of ketogenesis, a crucial alternative metabolic pathway that provides lipid-derived energy for various organs during times of carbohydrate deprivation. Ketone bodies may serve as critical fuel in ketogenesis for tumour initiation or metastasis. Chen et al.'s study, in which clinical specimens, cellular experiments, and animal models were used, suggests that *HMGCS2* expression increases cancer cell invasion and metastasis ability.[22] In addition, a study by Lee et al. indicates that high expression of *HMGCS2* predicts poor susceptibility of rectal cancer to preoperative chemoradiotherapy and could act as a promising prognostic factor.[23] Its high expression in long-duration UC as compared with short-duration UC warrants further study into how ketogenesis might play a role during UC progression into CAC. One of the top downregulated genes, peptidase inhibitor 3 [*PI3*], encodes an elastase-specific inhibitor that is also known as elafin. A recent study by Zhang et al. compared the expression of elafin in the colonic mucosa of patient with IBD versus that in healthy controls by immunohistochemistry staining and qRT-PCR.[24] They showed that the expression of elafin decreased in patients with active IBD, and was negatively correlated with disease activity, suggesting that elafin may play a protective role in IBD. Thus, it is not surprising to find major downregulation of *PI3* in the long-duration UC samples in our study.

The expressions of *VNN1* and *SLC6A14* in this study could be further studied for their interesting trends. *VNN1* expression has been shown to be upregulated in colon biopsies of UC patients versus healthy controls.[25] This gene was able to control inflammation-driven carcinogenesis in the colitis-associated vanin-1$^{-/-}$ mouse model.[26] Zhang et al. showed that *GPRC5a* is a potential biomarker for colon cancer and promotes tumourigenesis through stimulation of vanin-1 expression and oxidative stress in CAC.[27] The expression of *VNN1* being downregulated in the long-duration group in our study could mean that its functions in promoting inflammation in terms of ROS production could play an even greater role during inflammation in the early stage of IBD rather than in the later stage. Solute carrier family 6 member 14 [*SLC6A14*] encodes a protein that is a member of the solute carrier family 6. Members of this family are sodium- and chloride-dependent neurotransmitter transporters. *SLC6A14* has been previously shown to be upregulated in UC samples versus normal samples.[28] That the *SLC6A14* transcript had downregulated expression in the long-duration group in our study is therefore interesting, since it is involved in the host's antibacterial response.[29] This finding may indicate that as UC progresses, microflora plays less of a pathogenic role in UC.
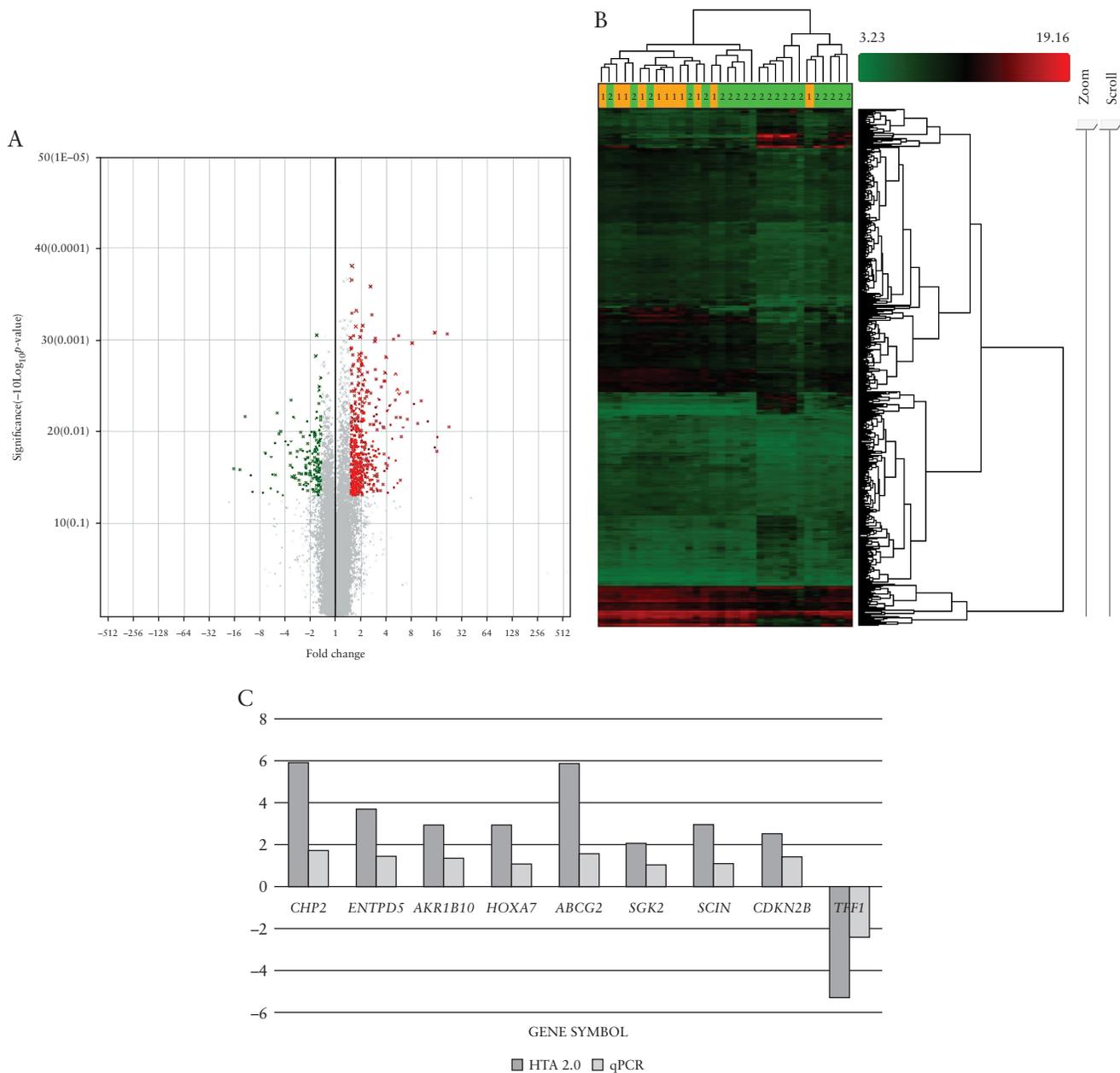
Figure 1. Global gene expression profiling revealed effects of disease duration. [**A**] Volcano plot of the differentially expressed genes using a cut-off value of $p < 0.05$ and fold change <-1.5 [green dots] and >1.5 [red dots]. [**B**] Heatmap presentation of the differentially expressed genes. [**C**] Nine of the ten selected genes were validated using qPCR.

Table 2. Summary of upregulated/downregulated transcripts along with coding/non-coding categories.

	Upregulated	Downregulated	Sum
Coding	380	123	503
Non-coding	97	40	137
Sum	477	163	

In terms of signalling pathways, fatty acid metabolism has been linked with both inflammation and CAC.[30–32] Butyrate and propionate are short-chain fatty acids metabolized by gut bacteria from otherwise indigestible fibre-rich diets. Butyrate has been shown to ameliorate inflammation in dextran sulphate sodium [DSS]-induced colitis.[33] On the other hand, a diet rich in long-chain fatty acids has been shown to promote inflammation and CAC in an animal model as well.[34] The valine, leucine, and isoleucine degradation pathway was previously identified as one of the overrepresented and altered pathways in CRC versus normal.[35,36] In a proteomics study on CAC, the pathway was identified as one of the important pathways, when the differential protein expressions were analysed for both non-dysplastic and high-grade dysplastic rectal tissues from UC progressors [UC patients with dysplasia/cancer] versus non-dysplastic rectal tissues from UC non-progressors [UC patients without dysplasia/cancer].[37] Regarding the PPAR signalling pathway, PPARs are nuclear hormone receptors that are activated by fatty acids and their

Table 3. Top ten upregulated and downregulated [coding] transcripts.

	Transcript ID	Gene symbol	Entrez ID	Fold change	*p*-value
Upregulated	TC01003050.hg.1	*HMGCS2*	3158	22.33	0.008985
	TC04001263.hg.1	*UGT2A3*	79799	16.02	0.016416
	TC4_ctg9_hap1000005.hg.1	*UGT2A3*	79799	15.11	0.014833
	TC17000638.hg.1	*B4GALNT2*	124872	14.89	0.000835
	TC18000135.hg.1	*MEP1B*	4225	12.39	0.00778
	TC01000525.hg.1	*GUCA2B*	2981	9.54	0.008164
	TC04001411.hg.1	*ADH1C*	126	8.52	0.005001
	TC17000836.hg.1	*OTOP2*	92736	8.08	0.001091
	TC05001096.hg.1	*SLC9A3*	6550	7.17	0.003719
	TC01004094.hg.1	*LYPD8*	646627	7.15	0.007311
Downregulated	TC20000341.hg.1	*PI3*	5266	−16.31	0.025312
	TC15001305.hg.1	*DUOX2*	50506	−13.86	0.025794
	TC06002119.hg.1	*VNN1*	8876	−11.91	0.006869
	TC0X000571.hg.1	*SLC6A14*	11254	−10.34	0.029836
	TC15000226.hg.1	*GREM1*	26585	−6.87	0.017174
	TC11002234.hg.1	*MMP1*	4312	−5.93	0.041395
	TC04000411.hg.1	*CXCL1*	2919	−5.85	0.019067
	TC04001511.hg.1	*TNIP3*	79931	−5.69	0.029314
	TC21000486.hg.1	*TFF1*	7031	−5.33	0.013355
	TC09000677.hg.1	*LCN2*	3934	−5.25	0.029787

**Table 4.** Top ten upregulated and downregulated [non-coding] transcripts.

	Transcript ID	Gene symbol	Entrez ID	Fold change	*p*-value
Upregulated	TC17002262.hg.1	*B4GALNT2*	124872	21.42	0.00086
	TC04002580.hg.1	–	–	16.13	0.011561
	TC05002797.hg.1	*SLC9A3*	6550	10.45	0.004611
	TC05002796.hg.1	*PP7080*	25845	5.15	0.002343
	TC14002083.hg.1	*ENTPD5*	957	4.67	0.011941
	TC19002140.hg.1	*CYP2B7P*	1556	4.4	0.005204
	TC16001604.hg.1	*CES2*	8824	3.94	0.022386
	TC20001228.hg.1	*SGK2*	10110	3.39	0.044092
	TC07002600.hg.1	*AKR1B10*	57016	3.15	0.044556
	TC01004608.hg.1	*GSTM4*	2948	3.12	0.009539
Downregulated	TC11003312.hg.1	*MMP1*	4312	−9.82	0.0451
	TC11003311.hg.1	*MMP1*	4312	−7.33	0.04662
	TC08001811.hg.1	*TUSC3*	7991	−3.42	0.00452
	TC12002282.hg.1	*SLCO1B3*	28234	−3.39	0.01284
	TC01005928.hg.1	*SELL*	6402	−3.34	0.03105
	TC11003313.hg.1	*MMP3*	4314	−3.09	0.03007
	TC09002547.hg.1	–	–	−2.69	0.03766
	TC21000717.hg.1	*KCNJ15*	3772	−2.56	0.02875
	TC17002479.hg.1	*MYH10*	4628	−2.3	0.02002
	TC19001977.hg.1	*ICAM1*	3383	−2.29	0.02911

derivatives. The PPARs have sparked much interest as a drug target for cancer prevention, including cancers related to the gastrointestinal system,[38,39] and their roles have been studied in colonic inflammation and colitis-associated tumourigenesis.[40] Bile acids have been found to be involved in apoptotic pathways in the colon.[41] The pathways that were significantly altered and the respective involved genes could be further investigated to validate their places in UC disease progression and their possible links to CAC.

Among molecular function GO terms, oxidative stress, which accompanies chronic inflammation, has long been deemed as one of the leading factors contributing to neoplastic transformation. Thus, antioxidant activity, listed as the most significantly perturbed function, may allow continuous damage throughout disease progression. In addition, metalloendopeptidase activity, which involves matrix

metalloproteinases [MMPs], has also been altered. [The roles of the MMPs involved in inflammation and CAC have been discussed in detail by Lewins et al.][42] Furthermore, fatty acid beta-oxidation was the most significant biological process, corresponding to fatty acid metabolism, in our findings for the KEGG pathway. Regarding the cellular component [CC], most of the most significant CCs were related to plasma membrane, tight junctions, and brush border, indicating changes in epithelial permeability as UC disease progresses.

The complexity of AS was evident in our study because most of the top differentially expressed coding genes underwent significant AS, which was not the case for the top non-coding genes. The alternative spliced transcripts could be further validated using RT-PCR or other platforms to generate more definite results. To date, there is still no specific evidence linking AS to the pathogenesis of UC.
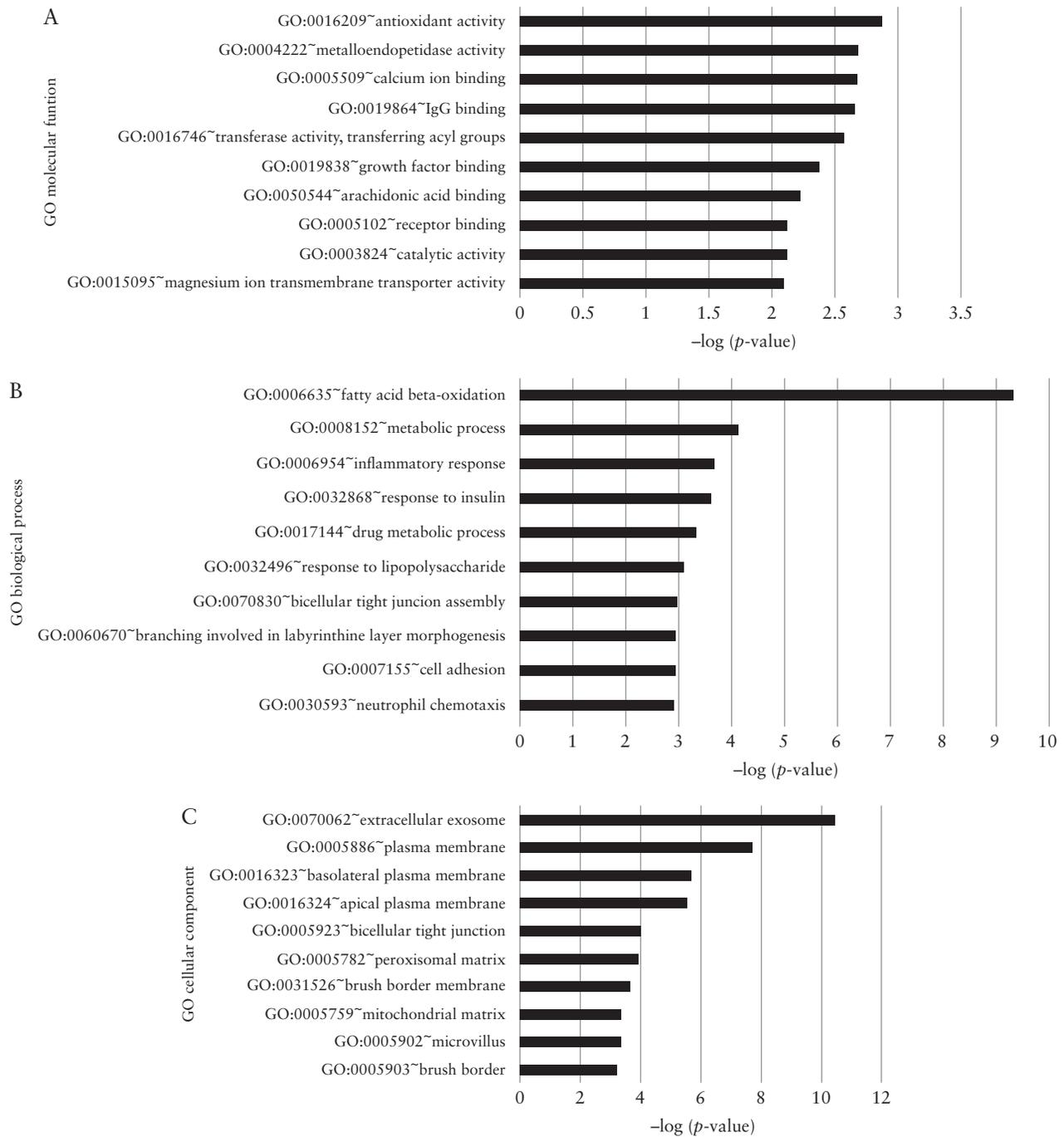
Figure 2. Representation of most significant gene ontology terms from genes with altered expression: [**A**] Molecular function; [**B**] biological process; [**C**] cellular component.

In addition, given the nature of our study, comparing long-duration to short-duration UC, the mean age of our patients was 65 versus 36 years in the two groups, respectively. There are not much data regarding how difference in age might play a role in transcriptomic studies using colonic biopsies. There is a recent study in which colonic microarray data shows that the molecular pathways in diseased tissue samples are similar between adult and paediatric UC patients.[43] But there is yet to be a study to show the direct effect of age on transcriptomic studies that specially compares elderly patients with those who are younger. Another point

to note is that in our population group, there were fewer males [45.5%] in the long-duration group than in the short-duration group [66.7%]. Taman et al. recently showed that their PCA analysis not only stratified treatment-naive UC patients from controls, but could also distinguish transcripts [to a lesser extent] and in a gender-specific manner.[10] However, there is insufficient information about gender-specific differences in UC, and studies to date have shown some contradictory results.[44] In our study, the majority of our long-duration UC patients were of Indian ethnicity, whereas the majority of our short-duration UC patients were of Malay ethnicity.

Table 5. Most significant KEGG pathways with altered expression in terms of disease duration.

Name of pathway	*p*-value
Metabolic pathways	3.81532623535e-18
Fatty acid degradation	2.07125533482e-12
Valine, leucine, and isoleucine degradation	4.97540814314e-12
PPAR signalling pathway	2.32291062327e-11
Bile secretion	2.6993561333e-10
Fatty acid metabolism	1.4175253173e-09
Propanoate metabolism	1.87208729251e-08
Drug metabolism—cytochrome P450	3.14232789678e-08
Tryptophan metabolism	8.46154556076e-08
Butanoate metabolism	1.4814311646e-07

A

B
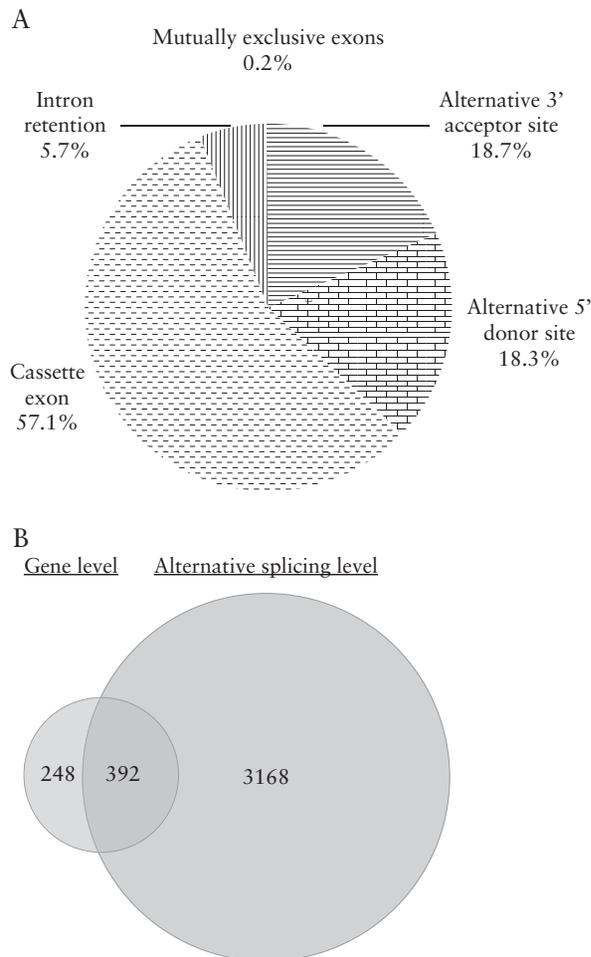
Figure 3. [**A**] Quantification of the different annotated alternative splicing events affected by disease duration. [**B**] Overlapping of transcripts with altered splicing events and transcripts with differential expression.

Currently, no such studies have been conducted in the Asia Pacific region regarding the potential contribution of race and ethnicity in terms of changing the transcriptomic landscape in biopsies of UC patients. A larger population study group is needed to make any definite conclusions in this matter. In addition, 5 out of 21 of the short-duration UC patients were treated with steroids. These patients were on tapering course of corticosteroids, ranging from 5 to 15 mg. These doses should not significantly influence the transcriptomic changes. None of our patients were on biologics, because biologics would most likely alter the medical progression of UC.

There were a few limitations in this study. This is a single-centre study, and thus the number of patients in each arm was limited, based on our strict inclusion and exclusion criteria. As the microarray technology was employed, the results were only based on the current database of known gene sequences and transcripts.

It has long been known that long duration of UC is a risk factor for the development of CAC. However, the transcriptomic changes/ biomarkers at different disease duration intervals [which can aid in determining the prognosis of UC, along with colonoscopy] remain largely unexplored. Having well-studied biomarkers could be a compliment to colonoscopy surveillance, because flat lesions can be overlooked. Also, notable molecular differences between sporadic CRC and CAC have been studied in patients and animal models. Therefore, genes with differential expression identified in this study could serve as the foundation not only for studying specific changes that lead to cancer, but also the development of potential molecular biomarkers that lead to CAC. More studies could be done by recruiting a larger patient population covering different disease durations, e.g. at 5-year intervals, to investigate the changes in gene expression during disease progression.

Funding

Conflict of Interest

The authors declare that there are no conflicts of interest to disclose.

Acknowledgments

Author Contributions

All authors have made a significant contribution to the research described in this manuscript. ENDL: Study concept and design, acquisition of data, analysis and interpretation of data, and manuscript writing. NMM: Study concept and design, interpretation of data, and manuscript writing. ZW: Acquisition of data. RARA: Interpretation of data, and critical revision of the manuscript. All authors approved the final manuscript as well as the authorship list.

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

1. Ng SC, Shi HY, Hamidi N, *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.

2. Prideaux L, Kamm MA, De Cruz PP, Chan FK, Ng SC. Inflammatory bowel disease in Asia: a systematic review. *J Gastroenterol Hepatol* 2012;**27**:1266–80.

3. Hilmi I, Jaya F, Chua A, Heng WC, Singh H, Goh KL. A first study on the incidence and prevalence of IBD in Malaysia—results from the Kinta Valley IBD Epidemiology Study. *J Crohns Colitis* 2015;**9**:404–9.

4. Kim SC, Tonkonogy SL, Karrasch T, Jobin C, Sartor RB. Dual-association of gnotobiotic IL-10⁻/⁻ mice with 2 nonpathogenic commensal bacteria induces aggressive pancolitis. *Inflamm Bowel Dis* 2007;**13**:1457–66.

5. Jostins L, Ripke S, Weersma RK, *et al.*; International IBD Genetics Consortium [IIBDGC]. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119–24.

6. Morgan XC, Tickle TL, Sokol H, *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;**13**:R79.

7. Planell N, Lozano JJ, Mora-Buch R, *et al.* Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. *Gut* 2013;**62**:967–76.

8. 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.

9. Smith PJ, Levine AP, Dunne J, *et al.* Mucosal transcriptomics implicates under expression of *BRINP3* in the pathogenesis of ulcerative colitis. *Inflamm Bowel Dis* 2014;**20**:1802–12.

10. 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.

11. Chan SN, Low END, Raja Ali RA, Mokhtar NM. Delineating inflammatory bowel disease through transcriptomic studies: current review of progress and evidence. *Intest Res* 2018;**16**:374–83.

12. Svrcek M, Borralho Nunes P, Villanacci V. Clinicopathological and molecular specificities of inflammatory bowel disease–related colorectal neoplastic lesions: the role of inflammation. *J Crohn's Colitis* 2018;**12**:1486–98.

13. Sutherland LR, Martin F, Greer S, *et al.* 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis, and proctitis. *Gastroenterology* 1987;**92**:1894–8.

14. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(–Delta Delta C[T]) Method. *Methods* 2001;**25**:402–8.

15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.

16. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;**44**:D457–62.

17. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.

18. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res* 2006;**34**:W720–4.

19. Xie C, Mao X, Huang J, *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;**39**:W316–22.

20. Huang dW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;**37**:1–13.

21. Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57.

22. Chen S-W, Chou C-T, Chang C-C, *et al.* HMGCS2 enhances invasion and metastasis via direct interaction with PPARα to activate Src signaling in colorectal cancer and oral cancer. *Oncotarget* 2017;**8**:22460–76.

23. Lee YE, He HL, Shiue YL, *et al.* The prognostic impact of lipid biosynthesis-associated markers, HSD17B2 and HMGCS2, in rectal cancer treated with neoadjuvant concurrent chemoradiotherapy. *Tumor Biol* 2015;**36**:7675–83.

24. Zhang W, Teng G, Wu T, Tian Y, Wang H. Expression and clinical significance of elafin in inflammatory bowel disease. *Inflamm Bowel Dis* 2017;**23**:2134–41.

25. Gensollen T, Bourges C, Rihet P, *et al.* Functional polymorphisms in the regulatory regions of the *VNN1* gene are associated with susceptibility to inflammatory bowel diseases. *Inflamm Bowel Dis* 2013;**19**:2315–25.

26. Pouyet L, Roisin-Bouffay C, Clément A, *et al.* Epithelial vanin-1 controls inflammation-driven carcinogenesis in the colitis-associated colon cancer model. *Inflamm Bowel Dis* 2010;**16**:96–104.

27. Zhang L, Li L, Gao G, *et al.* Elevation of GPRC5A expression in colorectal cancer promotes tumor progression through VNN-1 induced oxidative stress. *Int J Cancer* 2017;**140**:2734–47.

28. Eriksson A, Flach CF, Lindgren A, Kvifors E, Lange S. Five mucosal transcripts of interest in ulcerative colitis identified by quantitative real-time PCR: a prospective study. *BMC Gastroenterol* 2008;**8**:34.

29. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol* 2006;**3**:390–407.

30. De Preter V, Arijs I, Windey K, *et al.* Decreased mucosal sulfide detoxification is related to an impaired butyrate oxidation in ulcerative colitis. *Inflamm Bowel Dis* 2012;**18**:2371–80.

31. Blouin JM, Penot G, Collinet M, *et al.* Butyrate elicits a metabolic switch in human colon cancer cells by targeting the pyruvate dehydrogenase complex. *Int J Cancer* 2011;**128**:2591–601.

32. Sivaprakasam S, Prasad PD, Singh N. Benefits of short-chain fatty acids and their receptors in inflammation and carcinogenesis. *Pharmacol Ther* 2016;**164**:144–51.

33. Vieira EL, Leonel AJ, Sad AP, *et al.* Oral administration of sodium butyrate attenuates inflammation and mucosal lesion in experimental acute ulcerative colitis. *J Nutr Biochem* 2012;**23**:430–6.

34. Kim IW, Myung SJ, Do MY, *et al.* Western-style diets induce macrophage infiltration and contribute to colitis-associated carcinogenesis. *J Gastroenterol Hepatol* 2010;**25**:1785–94.

35. Wang Y, Zheng T. Screening of hub genes and pathways in colorectal cancer with microarray technology. *Pathol Oncol Res* 2014;**20**:611–8.

36. Vogtmann E, Hua X, Zeller G, *et al.* Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One* 2016;**11**:e0155362.

37. May D, Pan S, Crispin DA, *et al.* Investigating neoplastic progression of ulcerative colitis with label-free comparative proteomics. *J Proteome Res* 2011;**10**:200–9.

38. Pazienza V, Vinciguerra M, Mazzoccoli G. PPARs signaling and cancer in the gastrointestinal system. *PPAR Res* 2012;**2012**:560846.

39. Gou Q, Gong X, Jin J, Shi J, Hou Y. Peroxisome proliferator–activated receptors [PPARs] are potential drug targets for cancer therapy. *Oncotarget* 2017;**8**:60704–9.

40. Wang D, Fu L, Ning W, *et al.* Peroxisome proliferator–activated receptor δ promotes colonic inflammation and tumor growth. *Proc Natl Acad Sci U S A* 2014;**111**:7084–9.

41. Barrasa JI, Olmo N, Lizarbe MA, Turnay J. Bile acids in the colon, from healthy to cytotoxic molecules. *Toxicol In Vitro* 2013;**27**:964–77.

42. Walter L, Harper C, Garg P. Role of matrix metalloproteinases in inflammation/colitis-associated colon cancer. *Immunogastroenterology* 2013;**2**:22–8.

43. Li K, Strauss R, Ouahed J, *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**:45–52.

44. Lesuis N, Befrits R, Nyberg F, van Vollenhoven RF. Gender and the treatment of immune-mediated chronic inflammatory diseases: rheumatoid arthritis, inflammatory bowel disease and psoriasis: an observational study. *BMC Med* 2012;**10**:82.