12-14-2016

Defining the Inflammation Biomarkers of Inflammatory Bowel Diseases and Colorectal Carcinomas

Jianxu Li

DEFINING THE INFLAMMATION BIOMARKERS OF INFLAMMATORY BOWEL DISEASES AND COLORECTAL CARCINOMAS

by

JIANXU LI

Under the Direction of Yi Jiang, PhD

ABSTRACT

Ulcerative colitis (UC) and Crohn's disease (CD) are the two common forms of inflammatory bowel disease (IBD). They share similar clinical and demographic features as well as harbor key differences in tissue damage and prognosis. Previous studies indicated that they contributed to the increased rick to Colorectal cancer (CRC). However, whether UC and CD share inflammatory signatures still remains controversial. In addition, no inflammatory signatures have been reported on CRC. To answer these questions, a comprehensive study has been conducted on collected microarray datasets. Our analysis suggests that although CD and UC share common inflammatory pathways, they also present difference. Especially, CD patients are likely to have type I response, while UC patients are inclined to undergo type II response. Pathway enrichment analysis on CRC uncovered two potential CRC-specific inflammatory pathways.

INDEX WORDS: Ulcerative colitis (UC), Crohn's disease (CD), Inflammatory bowel disease (IBD), Colorectal cancer (CRC), Microarray, Pathway.

DEFINING THE INFLAMMATION BIOMARKERS OF INFLAMMATORY BOWEL DISEASES AND COLORECTAL CARCINOMAS

by

JIANXU LI

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2016

DEFINING THE INFLAMMATION BIOMARKERS OF INFLAMMATORY BOWEL DISEASES AND COLORECTAL CARCINOMAS

by

JIANXU LI

Committee Chair: Yi Jiang

Committee: Gengsheng Qin

Xin Qi

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2016

DEDICATION

I dedicate this work to my family: my wife, my son Zelin and Noah as they have spiritually supported me throughout the process. A special feeling of gratitude to my mother who has come here from China to help taking care of my son so that I can have spent more time on this thesis. I also want to dedicate this dissertation to my many friends and classmates for their encouragement words.

ACKNOWLEDGEMENTS

I would like to first thank my thesis advisor, Dr. Yi Jiang, without whom this thesis could not have been completed. She is knowledgeable and pleasant to work with, always encouraging and supporting me to explore new ideas.

I also would like to thank Dr. Gengsheng Qin, and Dr. Xin Qi for all their help in many different forms during my graduate study in this department, either academically or materially.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **v**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

LIST OF ABBREVIATIONS ..... **x**

1 INTRODUCTION **1**

1.1 Inflammatory Bowel Disease (IBD) and Colorectal cancer (CRC) **1**

1.2 Expected Results **2**

2 Methods **4**

2.1 Description of Data **4**

2.2 Data processing and differential expression analysis **4**

2.3 Gene ontology over-representation analysis **5**

2.4 Models and statistics **6**

3 RESULTS AND DISCUSSION **8**

3.1 CD and UC Show Different Gene Signatures **8**

3.2 Gene Signatures of CRC **8**

3.3 Pathway Enrichment Analysis on CD, UC, and CRC groups **10**

3.4 Comparative Analysis of Enriched Pathways from CD, UC and CRC **12**

4 CONCLUSIONS **15**

REFERENCES **17**

APPENDICES 20

Appendix A 20

Appendix B 24

Appendix C 31

LIST OF TABLES

Table 1. Data description of the five data set from GEO. 4

Table 2. Differentially expressed genes identified from each datasets. 9

Table 3. Pathways enriched in different groups. 10

Table 4. Different inflammatory features of CD and UC patients. 12

Table 5. Two Candidate GO Pathways for CRC-related Inflammation. 14

Table 6S. Enriched CD- and UC- shared and -specific pathways 24

Table 7S. Enriched pathways for CRC. 31

LIST OF FIGURES

Figure 1. Data processing and analysis pipeline 5

Figure 2. Overlapped gene expression profiles between CD and UC groups 9

Figure 3. Interactive pathways enriched for CD, UC, and CRC diseases. 11

Figure 4. Shared pathways among different diseases. 14

LIST OF ABBREVIATIONS

GEO: Gene Expression Omnibus

IBD : Inflammatory Bowel Disease

CRC: Colorectal cancer

CD: Crohn's disease

UC: Ulcerative colitis

NM: Normal control

FDR: false discovery rate

NCBI ENTRZID: NCBI gene entry Identity

GO ID: Gene Ontology identity

1 INTRODUCTION

1.1 Inflammatory Bowel Disease (IBD) and Colorectal cancer (CRC)

Colorectal cancer (CRC) is the third most common cancer and the fourth most common cancer cause of death globally, accounting for roughly 1.2 million new cases and 600 000 deaths per year [1]. Many known factors such as genetic mutations, inflammatory bowel disease (IBD) can contribute to the occurrence of CRC. As the two common forms of IBD, both Ulcerative colitis (UC) and Crohn's disease (CD) predispose patients to increased risk of CRC. A cohort study from Eaden et al reported that the probability of developing CRC 10 years after diagnosis of UC was 2%, reaching the level of 8% after 20 years and 18% after 30 years [2]. A meta-analysis of 12 published articles by Canavan et al showed that the cumulative risk for patients with Crohn's disease (CD) was 2.9% at 10 years, 5.6% at 20 years, and 8.3% at 30 years after the CD diagnosis [3]. These studies indicated a comparable role of UC and CD in the cumulative risk of developing CRC.

Ulcerative colitis (UC) and Crohn's disease (CD) share clinical and demographic features. However, they also harbor key differences in tissue damage and prognosis. UC is limited to the colon while CD can occur anywhere in the gastrointestinal tract. Inflammation in UC with crypt abscess formation is limited to the mucosa while a deeper, often transmural inflammation with fistula formation is seen in CD. All these different characteristics suggest distinctive etiopathogenic processes. Wu, et al [4] and Lawrance, et al [5] have found different gene signatures between CD and UC diseases through microarray studies. Analysis of the mucosal T cell phenotype revealed that Th1 cells is predominant in CD disease, while Th2 cells is polarized in UC disease [6,7]. However, recent studies from Granlund, et. al showed that no Th preference

between CD and UC diseased [8]. These controversies intrigued us to re-evaluate these conclusions.

It was known that long-standing inflammation secondary to chronic infection or irritation can predispose to cancer [9]. The inverse is also true. Cancer can cause inflammation as smoldering, non-resolving inflammation is one of the consistent features of the tumor microenvironment [10]. Although some general features of cancer-related inflammation have been described, CRC-specific inflammation signature has not been reported. Furthermore, it would be interesting to investigate the similarity and difference between IBD-associated inflammation signatures and CRC-associated ones.

Gene expression based technology has presented a high throughput tool in the field of cancer or diseased-related research. Since the advent of microarray, a large amounts of genome-wide gene expression data has been generated and collected in public archives. To date, the Gene Expression Omnibus (GEO), a repository of array- and sequence-based expression data, stores 1,947,791 samples performed on 16,428 platforms [11]. However, in the case of same disease, analysis on microarrays submitted by different research groups will usually produce different results. Potential causes of this low reproducibility include differences in sample collection methods, processing protocols, and microarray platforms, patient heterogeneity, and analyzing algorithm or criteria [12]. Thus, studies based on a single batch microarray mostly lead to biased results.

1.2 Expected Results

I would like to get the CD- and UC-common and unique inflammatory signatures. Meanwhile, through the CRC samples studies, we would be able to obtain the CRC-associated inflammatory signatures. Comparative analysis on these inflammatory signatures will inform us the similarity

and difference among them. These studies will further our understanding of the underling pathological mechanism of IBD and CRC.

2 METHODS

2.1 Description of Data.

A thorough search of the Gene Expression Omnibus (GEO) results in five datasets that meets the following two criteria: (1), the deposited raw data are in CEL formats; (2), the datasets for IBD should contain both CD and UC samples. The raw data of the five datasets (GEO accession number: GSE6731 [4], GSE36807 [13], GSE4183 [14], GSE15960 [15], GSE37364 [16]) were downloaded from the at (http://www.ncbi.nlm.nih.gov/geo/). Detailed sample information on these microarrays are listed in Table.1, In total, our samples consist of 63 normal control (NM), 24 CD patients, 27 UC patients, and 48 CRC patients.

Table 1. Data description of the five data set from GEO.

GEO Ac.Nm	NM	CD	UC	CRC	Ref
GSE6731	4	6	5		Wu. et al., 2007 [4]
GSE36807	7	12	13		Montero-Meléndez. et al., 2013 [13]
GSE4183	8	6	9	15	Galamb. et al, 2008 [14]
GSE15960	6			6	Galamb. et al., 2010 [15]
GSE37364	38			25	Valcz. et al., 2014 [16]
Total	63	24	27	48	

2.2 Data processing and differential expression analysis

As is illustrated in Figure.1, raw data of these five datasets are separately imported into the R software by using the “affy” package [17], and separately processed by “vsn” package for variance stabilization, background correction, normalization and log transformation [18]. Through a non-specific filtration step using the “genefilter” package, we delete the redundant probes and

the ones lacking NCBI ENTRZID. In addition, this step also eliminates the probes whose variances are less than 0.5 [20]. This filtration process leaves us a total of 4298 genes in GSE6731 and a total of 10267 genes in the other four datasets to start with. By employing the "limma" package, differentially expressed genes were determined, these significantly differentially-expressed genes were further tested through the empirical Bayes method. Over-expressed genes are then selected based on the criteria with false discovery rate (FDR) ≤ 0.05 and fold changer ≥ 1.5 [21].

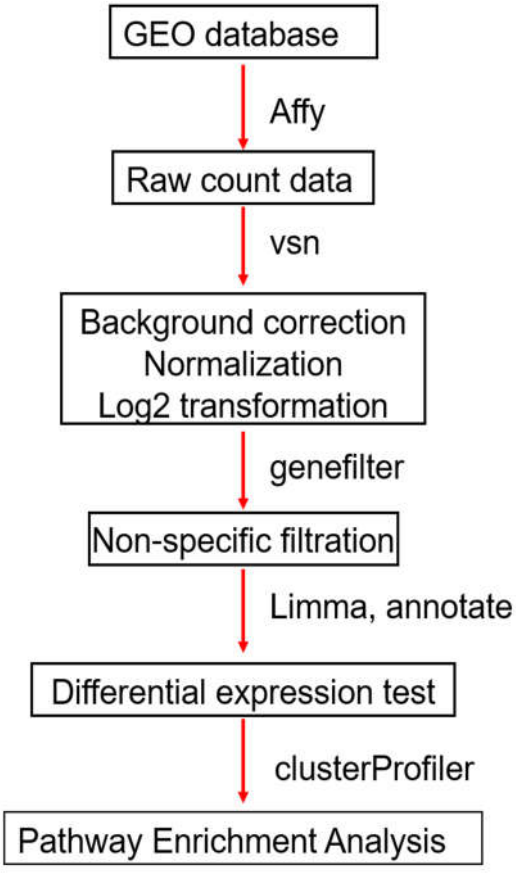

Figure 1. Data processing and analysis pipeline

2.3 Gene ontology over-representation analysis

Pathway over-representation analysis were performed using the "clusterProfiler" package, which uses Hypergeometric Testing to annotates genes to biological processes. To prevent high

false discovery rate (FDR), the Benjamini-Hochberg method ($\alpha = 0.05$) is used in this process [22]. To decrease redundant GO terms, the annotated terms were further filtered at level four. Enriched GO terms are visualized using REVIGO at http://revigo.irb.hr/revigo.jsp [23]. Corresponding R code used for the above analysis was shown in appendix. A.

2.4 Models and statistics

The "vsn" package introduces an additive-multiplicative error model for microarray gene expression data that comprises data calibration, quantification of differential expression, and quantification of measurement error. The package derives a transformation *h* for intensity measurements, in which the variance of the transformed intensities is independent of the mean. The parametric form is $h(x)=arsinh(a+bx)$, where x stands for the probe intensity of each sample, while a and b are derived constant used for determining the mean of each probe intensity. These constants were well defined by Tibshiran R [18] and Huber et al [19]. In addition, a different statistic Δh, whose variance is approximately constant along the whole intensity range, is used to measure differential expression. For highly expressed genes, *h* coincides with the logarithmic transformation, and Δh with the log-ratio. All the parameters are estimated with a robust variant of maximum-likelihood estimation [19].

In order to assess differentially expressed genes, the "limma" package fits a gene-wise linear models $E[\ y_j\ ] = X\beta_j$ to gene expression data, where j stands for each gene, y_j contains the expression data for the gene j, X is the design matrix and β_j is a vector of coefficients which is interest. Here y_j^T is the jth row of the expression matrix. The contrasts of interest are given by $\alpha j = C^T \beta_j$ where C is the contrasts matrix. The linear modelling is performed in a row-wise fashion, with regression coefficients and standard errors either directly estimating the comparisons of interest or via contrasts. Test-statistics are obtained for gene ranking that can be further

summarized at the gene set level to perform gene signature/pathway-level ranking. Additionally, the eBaye function of “limma” borrow information between genes when estimating the variances. It uses a robustified shrinkage strategy to estimate the gene-wise shrinkage factors, which offers the benefits of shrinkage to the majority of the genes, whilst negating the effects of outliers [21].

3 RESULTS AND DISCUSSION

3.1 CD and UC Show Different Gene Signatures

In Table.2, a total of 205 genes were found significantly alternated between the CD and NM groups, and 371 genes between the UC and NM groups in dataset GSE6731 according to the predefined criteria. Among these genes, 125 genes including 60 upregulated genes and 65 downregulated genes were shared by CD and UC groups as was shown in Figure 2. In dataset GSE36807, a total of 186 genes were detected significantly expressed between the CD and NM groups, and 1196 genes between UC and NM groups. Among them, 57 upregulated genes and 99 downregulated genes were shared between the CD and UC groups (Figure 2). More significantly differential expressed genes were discovered in dataset GSE4183. Between the CD and control groups, 1477 genes were detected significantly differentially expressed. This number was 1044 between the UC and control groups. Among them, 580 upregulated and 185 downregulated genes were overlapped between the CD and UC groups (Figure.2). Due to batch effect, platform effect, and sample heterogeneity, the three datasets presented different number of significant genes. For the CD patients, 51.8% -83.8% of their differential expressed genes are shared by the UC patients. Clearly CD and UC patients displayed differential inflammatory signatures as well as similar ones at transcriptome level.

3.2 Gene Signatures of CRC

Through the same procedure, we also obtained the gene signatures of CRC from the three datasets containing CRC patients. The three datasets gave various numbers of differentially expressed genes. As was shown in Table 2, Dataset GSE4183 gave the least number of differentially expressed genes (1180 genes) between CRC and normal groups, while dataset GSE15960 showed more differentially expressed genes (4412 genes) between the CRC and normal

groups. These difference are mainly caused by the high heterogeneity of the CRC patients, which has already indicated by Wu, et. al [4].

Table 2. Differentially expressed genes identified from each datasets.

GEO Ac.Nm	CD Vs NM	UC Vs NM	CRC Vs NM
GSE6731	205	371	
GSE36807	186	1196	
GSE4183	1477	1003	1180
GSE15960			4412
GSE37364			2883

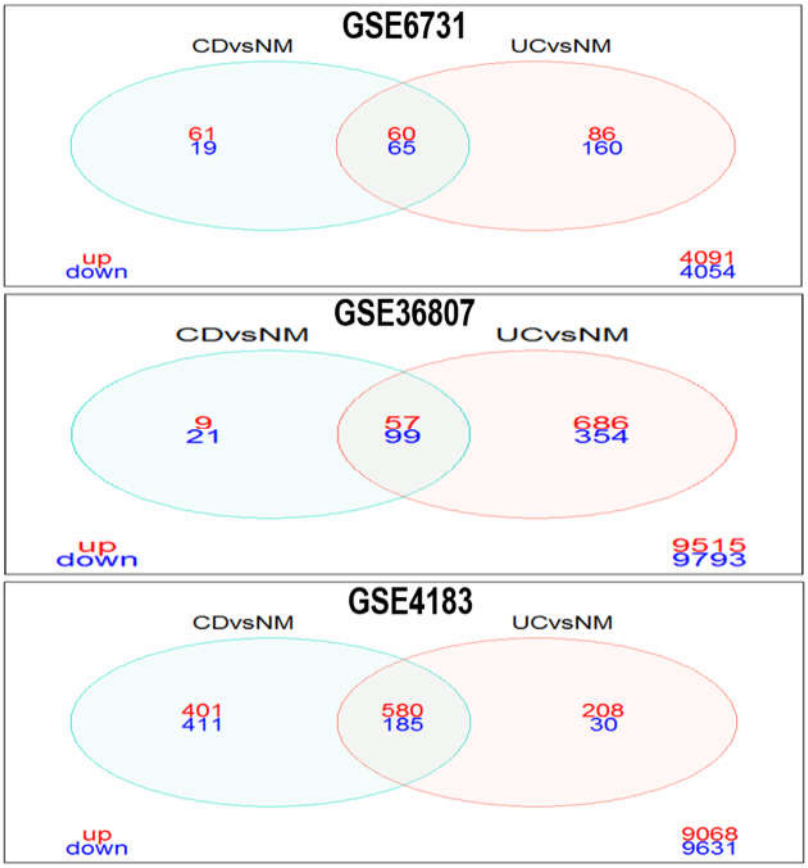

Figure 2. Overlapped gene expression profiles between CD and UC groups

3.3 Pathway Enrichment Analysis on CD, UC, and CRC groups

The alteration of one gene can be attributed to different pathways due to the complexity of the gene regulation. Thus, analysis focused at the gene level usually failed to uncover the underlying biological mechanism. To investigate whether the CD and UC groups show different inflammatory signatures at higher level, we performed pathway over-representation analysis based on these differentially expressed genes by using the "clusterProfiler" package. Enriched pathways from each dataset are shown in Table 3. In the CD and the CRC groups, we could not find any conserved pathways from the three datasets. In the UC group, 2 pathways (GO:0019725, GO:0050921) are found conserved across the three datasets. A combination of all the pathways enriched from different datasets, we achieved a total of 195 CD-associated pathways and 196 UC-associated pathways (Table 3). Similarly, pathway enrichment based on the differentially expressed genes from CRC datasets produced 149 CRC-associated pathways. An overview of these enriched pathways for different disease groups are shown in Figure 3.

Table 3. Pathways enriched in different groups.

GEO Ac.Nm	CD	UC	CRC
GSE6731	39	4	
GSE36807	3	138	
GSE4183	178	157	109
GSE15960			2
GSE15960			107
Conserved	0	2	0
Total	195	196	149

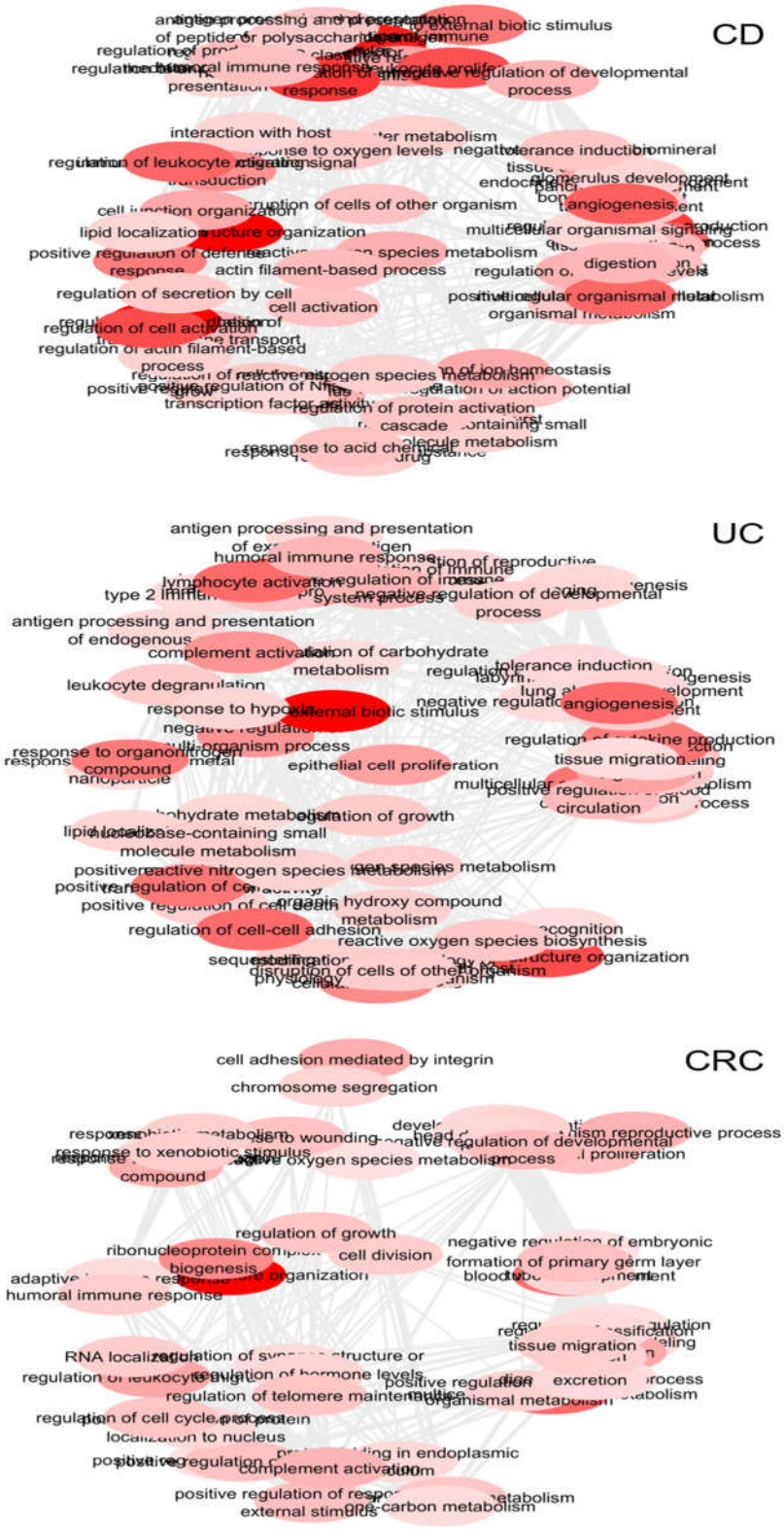

Figure 3. Interactive pathways enriched for CD, UC, and CRC diseases.

3.4 Comparative Analysis of Enriched Pathways from CD, UC and CRC

Comparative analysis of these CD- and UC-associated pathways indicated that 141 pathways were shared between CD and UC groups (listed in Table 6S). As we can see, some of these pathways such as GO:0002285, GO:0072676, GO:0019884, GO:0048002, GO:0002286, GO:0070489, GO:0070489, and so on, are directly associated with adaptive immune response, while complement activation (GO:0006956), and Toll-like receptor mediated signaling pathways (GO:0034122 and GO:0051092) signify the involvement of innate immune system. In addition, several enriched pathways (GO:0043299, GO:0097529, GO:0002274) are clearly indicative of the participation of myeloid cells in the inflammation response of IBD. These observations on the shared inflammatory signatures between CD and UC groups are consistent with those from Granlund, et. al [8].

It was known that Th1 cells secrete interleukin (IL)-2, interferon-γ, and lymphotoxin-α and stimulate type 1 immunity, which is characterized by intense phagocytic activity. Conversely, Th2 cells secrete IL-4, IL-5, IL-9, IL-10, and IL-13 and stimulate type 2 immunity, which is accompanied by high antibody titers [24]. Interestingly, through analyzing the enriched CD-unique and UC-unique pathways, we also found that the CD and UC groups present different inflammatory features. As was seen in Table 4, the CD-unique immune-related pathways are characteristics of type I immune response as cell killing and oxidative burst are the features of phagocytic activity. On the contrary, type II immune response predominates in UC group.

Table 4. Different inflammatory features of CD and UC patients.

CD-unique immune-related pathways	
GO ID	Description
GO:001912	positive regulation of leukocyte mediated cytotoxicity

GO:0031341	regulation of cell killing
GO:0001910	regulation of leukocyte mediated cytotoxicity
GO:0002367	cytokine production involved in immune response
GO:0045730	respiratory burst
GO:0050776	regulation of immune response
GO:0002577	regulation of antigen processing and presentation
UC-unique immune-related pathways	
GO:0042092	type 2 immune response
GO:0002279	mast cell activation involved in immune response
GO:0002377	immunoglobulin production
GO:0002532	production of molecular mediator involved in inflammatory response
GO:0002702	positive regulation of production of molecular mediator of immune response
GO:0019883	antigen processing and presentation of endogenous antigen

Through a similar procedure, in total, 149 CRC-associated pathways (Table 7S) are obtained. As are expected, most of these pathways are the contributors of cancer growth and metastasis. For example, multiple pathways from CRC directly regulate the extracellular matrix such as GO:0033628, GO:0045785, GO:0031589, GO:0030155, GO:0033627, and so on. Various pathways such as GO:0051321, GO:0051781, GO:0051445, GO:0051302, GO:1903046, GO:0090068, GO:0007126, GO:0051301, GO:0010564 are closely related to the cell proliferation processes. Additionally, several pathways such as GO:2000146, GO:2000147, GO:0050921, GO:0010632 clearly regulate cell migration.

Among the 149 CRC-associated pathways, 68 of them were shared with the CD group and 77 of them were shared with the UC group (Figure 3 and Table 7S). Interestingly, 60 pathways are found across the three groups. Among these pathways, 10 of them are inflammation-related such as GO:0002443, GO:0002250, GO:0071674, GO:0009617, GO:0006959, GO:0009611, GO:0002687,

GO:0097529, GO:0006956, GO:0030595, which indicates that CRC shares inflammatory features with CD and UC diseases. In addition, pathways, for example, GO:1901343, GO:1904018, GO:0008285, GO:0045785, GO:0050673, GO:0030155, and GO:0043062, are extracellular matrix-related or cell proliferation-related, suggesting that CRC and IBD also employ similar mechanism to remodel local tissues and regulate the cells which are involved in these biological process.

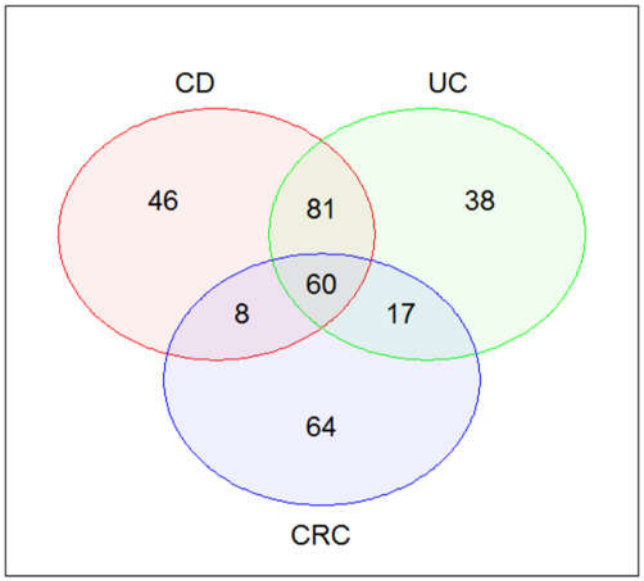

Figure 4. Shared pathways among different diseases.

Analysis on the 64 CRC-unique pathways showed that 31 of these pathways are cell-cycle-regulation- or cell metabolism-related, implying that that CRC are distinct from IBD in their cell regulating mechanism. Further analysis also uncovered two GO pathways for potential CRC-specific inflammation-related pathways (Table.5).

Table 5. Two Candidate GO Pathways for CRC-related Inflammation.

GO ID	Description
GO:0098869	cellular oxidant detoxification
GO:0002686	negative regulation of leukocyte migration

4 CONCLUSIONS

In this study, a comprehensive analysis was conducted on CD, UC and CRC at gene and pathway level. According to our knowledge, this is the most comprehensive studies on these three forms of diseases. Our studies indicated that CD and UC present different gene signatures at transcriptome level, which is consistent with the studies from Wu, et al [4] and Lawrance, et al [5]. Furthermore, our over-representation pathway analysis on these differentially expressed genes revealed that 72% (141/195) of these pathways are shared between CD and UC diseases, including various immune-related pathways. However, they also show different inflammatory features. Especially, type I immune response predominate in CD disease, while type II immune response are more favored in UC diseases. On this point, our results are different from Granlund, et. al [8]. In their studies, they could not detect any signs of Th2 differentiation, while our pathway analysis clearly shows Th2 response in UC patients. Two reasons might cause this difference. On one side, Granlund, et. al's analysis are mainly focused at gene level, the characteristic Th2 gene signatures might not significantly show up in their sample or be detected by their analyzing criteria. On the other side, although they have conducted gene set enrichment analysis, their pathway analysis is enriched at more general level. These reasons possibly mask the difference between CD and UC groups.

Pathway enrichment on the panel of selected genes from CRC uncovered 149 potential pathways. Analyzing these pathways found that CRC shares inflammatory features with IBD diseases and remodels local tissues and regulate involved cells through a similar mechanism. A comparison of CRC-associated pathways with CD- and UC-associated pathways revealed 64 CRC-specific pathways. Further investigation of these CRC-specific pathways indicated that CRC are different from IBD in their cell regulating mechanism. In addition to these, two pathways are

uncovered to be potential candidates for the CRC-associated inflammation. It would be interesting to apply these finding on CRC patients in the TCGA database.

In summary, by using publicly available data sets, the present studies for the first time point out the Th preference of CD and UC patients, which are consistent with previous experiment data. Furthermore, our studies also unmask the similarity and distinction between CRC and IBD diseases, which furthered our understanding of the pathogenesis of these disease.

REFERENCES

[1] Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2014;383(9927):1490-1502.

[2] Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001; 48(4):526-535.

[3] Canavan C, Abrams KR, Mayberry J.Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease. *Aliment Pharmacol Ther* 2006; 23(8):1097-1104.

[4] Wu F, Dassopoulos T, Cope L, Maitra A, et al. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflamm Bowel Dis* 2007;13(7):807-821.

[5] Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 2001,10(5):445-456.

[6] Neurath MF, Finotto S, Glimcher LH.The role of Th1/Th2 polarization in mucosal immunity. *Nat Med* 2002,8(6):567-573.

[7] Shale M, Ghosh S. Beyond TNF, Th1 and Th2 in inflammatory bowel disease. *Gut* 2008, 57(10):1349-1351.

[8] Granlund Av, Flatberg A, Østvik AE, Drozdov I, et al. Whole genome gene expression meta-analysis of inflammatory bowel disease colon mucosa demonstrates lack of major differences between Crohn's disease and ulcerative colitis. *PLoS One* 2013;8(2):e56818.

[9] Seth RN, Why Cancer and Inflammation? *Yale J Biol Med* 2006;79(3-4): 123–130.

[10] Mantovani A, Allavena P, Sica A, et al. Cancer-related inflammation. *Nature* 2008;454(7203):436-444.

[11] Gene Expression Omnibus. Available online: http://www.ncbi.nlm.nih.gov/geo/ (accessed on 5 OCT 2016).

[12] Walsh CJ, Hu P, Batt J, Santos CC. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays* (Basel) 2015;4(3):389-406.

[13] Montero-Meléndez T, Llor X, García-Planella E, Perretti M, et al. Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. *PLoS One* 2013;8(10):e76235.

[14] Galamb O, Györffy B, Sipos F, Spisák S, et al. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. *Dis Markers* 2008;25(1):1-16.

[15] Galamb O, Spisák S, Sipos F, Tóth K, et al. Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *Br J Cancer* 2010;102(4):765-73.

[16] Valcz G, Patai AV, Kalmár A, Péterfia B, et al. Myofibroblast-derived SFRP1 as potential inhibitor of colorectal carcinoma field effect. *PLoS One* 2014;9(11):e106143.

[17] Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004;20(3):307-315

[18] Tibshiran R. Estimating Transformations for Regression Via Additivity and Variance Stabilization. *J Amer Stat Assoc* 1988; 83:394-405.

[19] Huber W, von Heydebreck A, Sueltmann H, Poustka A, et al. "Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression." *Bioinformatics* 2002; 18(Suppl 1):S96-S104.

[20] Gentleman R, Carey V, Huber W, et al. genefilter: genefilter: methods for filtering genes from high-throughput experiments. 2016; R package version 1.54.2.

[21] Ritchie ME, Phipson B, Wu D, Hu Y, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.

[22] Yu G, Wang L, Han Y, He Q. "clusterProfiler: an R package for comparing biological themes among gene clusters." *OMICS* 2012;16(5):284-287.

[23] Supek F1, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011;6(7):e21800.

[24] Spellberg B1, Edwards JE.Type 1/Type 2 immunity in infectious diseases. *Jr.Clin Infect Dis* 2001;32(1):76-102.

APPENDICES

Appendix A

```
# load the required R pacakages;
library("Biobase")
library("annotate")
library("hgu133plus2.db")
library("hgu133plus2.db")
library("hgu95av2.db")
library(limma)
library(genefilter)
library("affy")
library("vsn")
library("clusterProfiler")
#setup data directory and read data into R;
setwd("C:/Users/little-Z/Desktop/thesis data/GSExxxxxx_CEL")
getwd()
Data=ReadAffy()
#do background calculation, normalization, log2 transformation, rename sample names,
checking fitting;
geset=vsnrma(Data)
meanSdPlot(geset)
sampleNames(geset) = substr(sampleNames(geset),1,9)
```

```
sampleNames(geset)
#check the normality of the samples;
boxplot(exprs(geset))
#non-specific filtration to remove control probes and probes which have no entrez ID, are
redundant;
qrange <- function(x) diff(quantile(x, c(0.1, 0.9)))
filt_geset = nsFilter(geset, require.entrez=TRUE, remove.dupEntrez=TRUE, var.func=qrange,
var.cutoff=0.5, feature.exclude="^AFFX")
#construct matrix for model fitting and linear model fitting
group=as.factor(c(rep("NM",a), rep("CD",b), rep("UC",c), rep("CRC",d))); # a, b, c, d, are the
number of samples of NM, CD, UC, CRC;
v = filt_geset$eset
design=model.matrix(~0+group)
colnames(design)=gsub("group", "", colnames(design))
design
contr.matrix=makeContrasts(CDvsNM=CD-NM, UCvsNM=UC-NM, CRCvsNM=CRC-NM,
levels = colnames(design))
contr.matrix
vft=lmFit(v, design)
vft=contrasts.fit(vft, contrasts=contr.matrix)
# empirical bayes moderation;
eft=eBayes(vft)
plotSA(eft)
```

```
#select differentially expressed genes;
dt=decideTests(eft,adjust.method="BH",p.value=0.05,lfc=0.585)
summary(dt)
# vennDiagram shows the overlapping genes;
vennDiagram(dt[,1:2], include = c("up", "down"),counts.col=c("red",
"blue"),circle.col=c("turquoise", "salmon"))
#extract differentially expressed genes for UC, CD, and CRC group.
Syms = unlist(mget(featureNames(v), env = hgu133plus2ENTREZID))
top1=topTable(eft, coef = 1, adjust.method = "fdr",sort.by = "p", number=Inf, genelist = Syms)
top11=top1[abs(top1$logFC)>=0.585 & top1$adj.P.Val<=0.05,]
dim(top11)
top2=topTable(eft, coef = 2, adjust.method = "fdr",sort.by = "p", number=Inf, genelist = Syms)
top22=top2[abs(top2$logFC)>=0.585 & top2$adj.P.Val<=0.05,]
dim(top22)
top3= topTable(eft, coef = 3, adjust.method = "fdr",sort.by = "p", number=Inf, genelist = Syms)
top33=top3[abs(top3$logFC)>=0.585 & top3$adj.P.Val<=0.05,]
dim(top33)
# pathway enrichment using differentially expressed genes from CD, UC, and CRC group;
Gene=top11$ID
Pathway= enrichGO(gene= GSE_UC_4183$ID,
            universe      = TOT,
            OrgDb         = org.Hs.eg.db,
            ont           = "BP",
```

```
            pAdjustMethod = "BH",
            pvalueCutoff  = 0.05,
            qvalueCutoff  = 0.05,
            readable      = TRUE)
summary(x)
L4_path=gofilter(Pathway,level=4)
#produce the Ven diagram of overlapped pathways;
set1=as.character(CD_total_path_order$ID)
set2=as.character(UC_total_path_order$ID)
set3=as.character(CRC_total_path_order$ID)
universe =sort(unique(c(set1, set2, set3)))
Counts =matrix(0, nrow=length(universe), ncol=3)
for (i in 1:length(universe)) {
   Counts[i,1] <- universe[i] %in% set1
   Counts[i,2] <- universe[i] %in% set2
   Counts[i,3] <- universe[i] %in% set3
}
colnames(Counts) <- c("CD","UC","CRC")
cols<-c("Red", "Green", "Blue")
 vennDiagram(vennCounts(Counts), circle.col=cols)
```

Appendix B

Table 6S. Enriched CD- and UC- shared and -specific pathways

Enriched CD and UC shared pathways

ID	Description	pvalue	p.adjust	GeneRatio
GO:0051283	negative regulation of sequestering of calcium ion	0.007174	0.045861	17/1408
GO:0034122	negative regulation of toll-like receptor signaling pathway	0.006952	0.044649	7/1408
GO:0051092	positive regulation of NF-kappaB transcription factor activity	0.006207	0.041039	22/1408
GO:0090130	tissue migration	0.005905	0.039598	39/1408
GO:0033627	cell adhesion mediated by integrin	0.00547	0.037243	14/1408
GO:0010876	lipid localization	0.004976	0.035086	48/1408
GO:0048729	tissue morphogenesis	0.004962	0.035051	85/1408
GO:0051701	interaction with host	0.004905	0.034876	30/1408
GO:0032102	negative regulation of response to external stimulus	0.004545	0.032753	36/1408
GO:2000242	negative regulation of reproductive process	0.00399	0.029702	11/1408
GO:0051282	regulation of sequestering of calcium ion	0.003592	0.027322	18/1408
GO:0072676	lymphocyte migration	0.003259	0.025336	17/1408
GO:0032409	regulation of transporter activity	0.003197	0.025252	27/1408
GO:0043299	leukocyte degranulation	0.003148	0.024975	15/1408
GO:0010524	positive regulation of calcium ion transport into cytosol	0.002867	0.023313	10/1408
GO:0055086	nucleobase-containing small molecule metabolic process	0.002747	0.023047	91/1408
GO:0002701	negative regulation of production of molecular mediator of immune response	0.002652	0.022335	9/1408
GO:0001666	response to hypoxia	0.002359	0.020433	46/1408
GO:0032845	negative regulation of homeostatic process	0.002115	0.018901	29/1408
GO:0050920	regulation of chemotaxis	0.002102	0.047495	11/201
GO:2001057	reactive nitrogen species metabolic process	0.002074	0.047495	7/201
GO:0051235	maintenance of location	0.002012	0.018168	43/1408
GO:0050673	epithelial cell proliferation	0.001962	0.017834	53/1408
GO:0006935	chemotaxis	0.001748	0.043283	20/201
GO:0042330	taxis	0.001748	0.043283	20/201
GO:0042493	response to drug	0.001744	0.016163	60/1408
GO:0048525	negative regulation of viral process	0.00173	0.016102	21/1408
GO:0043900	regulation of multi-organism process	0.001697	0.043001	17/201
GO:1903708	positive regulation of hemopoiesis	0.001595	0.01515	30/1408
GO:0001816	cytokine production	0.001548	0.039984	24/201

GO:0009620	response to fungus	0.001492	0.014463	11/1408
GO:0042267	natural killer cell mediated cytotoxicity	0.001492	0.014463	11/1408
GO:0050778	positive regulation of immune response	0.001486	0.038748	27/201
GO:0010522	regulation of calcium ion transport into cytosol	0.001396	0.013875	16/1408
GO:1903530	regulation of secretion by cell	0.001243	0.012627	77/1408
GO:0002443	leukocyte mediated immunity	0.00116	0.034581	15/201
GO:0001763	morphogenesis of a branching structure	0.001123	0.011652	36/1408
GO:0010035	response to inorganic substance	0.001112	0.011625	68/1408
GO:0007584	response to nutrient	0.001056	0.011125	32/1408
GO:0035295	tube development	0.000989	0.01058	83/1408
GO:1904951	positive regulation of establishment of protein localization	0.000987	0.01058	74/1408
GO:0070482	response to oxygen levels	0.000879	0.00982	51/1408
GO:0070489	T cell aggregation	0.00087	0.02809	20/201
GO:0071559	response to transforming growth factor beta	0.000862	0.009682	38/1408
GO:0044253	positive regulation of multicellular organismal metabolic process	0.000855	0.009625	7/1408
GO:0070168	negative regulation of biomineral tissue development	0.000855	0.009625	7/1408
GO:1903524	positive regulation of blood circulation	0.000817	0.009422	16/1408
GO:0019725*	cellular homeostasis	0.000762	0.025514	27/201
GO:0010243	response to organonitrogen compound	0.000755	0.025514	28/201
GO:2000257	regulation of protein activation cascade	0.000745	0.025514	5/201
GO:0002507	tolerance induction	0.000683	0.00817	9/1408
GO:0008285	negative regulation of cell proliferation	0.000677	0.024266	25/201
GO:0019884	antigen processing and presentation of exogenous antigen	0.000616	0.023428	13/201
GO:0071674	mononuclear cell migration	0.000571	0.007086	18/1408
GO:0032844	regulation of homeostatic process	0.000462	0.005946	65/1408
GO:0003254	regulation of membrane depolarization	0.000449	0.005803	11/1408
GO:0001101	response to acid chemical	0.000449	0.005803	45/1408
GO:0044364	disruption of cells of other organism	0.000429	0.005751	8/1408
GO:0001503	ossification	0.000392	0.005413	58/1408
GO:0048534	hematopoietic or lymphoid organ development	0.00034	0.004803	106/1408
GO:0006959	humoral immune response	0.000332	0.015284	11/201
GO:1901343	negative regulation of vasculature development	0.00033	0.004726	18/1408
GO:0002250	adaptive immune response	0.000253	0.013101	17/201
GO:0043270	positive regulation of ion transport	0.000249	0.003741	33/1408
GO:0060761	negative regulation of response to cytokine stimulus	0.000228	0.003513	12/1408
GO:1904018	positive regulation of vasculature development	0.000222	0.003454	29/1408
GO:0002700	regulation of production of molecular mediator of immune response	0.000198	0.003188	23/1408

GO:0048002	antigen processing and presentation of peptide antigen	0.000193	0.011356	14/201
GO:2000146	negative regulation of cell motility	0.000189	0.003083	43/1408
GO:0002285	lymphocyte activation involved in immune response	0.000164	0.00277	30/1408
GO:0006956	complement activation	0.00012	0.002112	16/1408
GO:0050817	coagulation	0.00012	0.002109	56/1408
GO:1903706	regulation of hemopoiesis	0.000105	0.001864	54/1408
GO:0007586	digestion	0.0001	0.047432	8/165
GO:0040013	negative regulation of locomotion	9.72E-05	0.001748	47/1408
GO:0016337	single organismal cell-cell adhesion	7.19E-05	0.00542	30/201
GO:0008360	regulation of cell shape	6.47E-05	0.001272	31/1408
GO:0009615	response to virus	6.41E-05	0.005212	18/201
GO:0051093	negative regulation of developmental process	5.96E-05	0.001186	108/1408
GO:0050878	regulation of body fluid levels	5.54E-05	0.001125	72/1408
GO:0051271	negative regulation of cellular component movement	5.42E-05	0.001109	48/1408
GO:0060759	regulation of response to cytokine stimulus	4.92E-05	0.001023	29/1408
GO:0072593	reactive oxygen species metabolic process	4.15E-05	0.000875	45/1408
GO:0002218	activation of innate immune response	4.08E-05	0.000867	49/1408
GO:0002286	T cell activation involved in immune response	3.95E-05	0.000842	21/1408
GO:0002523	leukocyte migration involved in inflammatory response	3.67E-05	0.000791	8/1408
GO:0032103	positive regulation of response to external stimulus	2.74E-05	0.002757	18/201
GO:2000021	regulation of ion homeostasis	2.07E-05	0.000483	34/1408
GO:0051607	defense response to virus	1.98E-05	0.002097	15/201
GO:0050921*	positive regulation of chemotaxis	1.75E-05	0.00041	30/1408
GO:0032101	regulation of response to external stimulus	1.72E-05	0.001948	30/201
GO:0007565	female pregnancy	1.66E-05	0.000395	34/1408
GO:0048771	tissue remodeling	1.53E-05	0.000368	32/1408
GO:0001818	negative regulation of cytokine production	1.43E-05	0.000347	40/1408
GO:0097529	myeloid leukocyte migration	1.32E-05	0.001632	14/201
GO:0030595	leukocyte chemotaxis	1.20E-05	0.00163	16/201
GO:0044057	regulation of system process	1.06E-05	0.000269	63/1408
GO:0002699	positive regulation of immune effector process	1.05E-05	0.000269	34/1408
GO:0043901	negative regulation of multi-organism process	9.98E-06	0.00026	31/1408
GO:0009617	response to bacterium	7.94E-06	0.001197	25/201
GO:0071216	cellular response to biotic stimulus	2.24E-06	6.55E-05	39/1408
GO:0045123	cellular extravasation	1.25E-06	3.83E-05	19/1408
GO:0007162	negative regulation of cell adhesion	1.03E-06	3.20E-05	48/1408
GO:0098542	defense response to other organism	8.04E-07	0.000241	23/201

GO:0002757	immune response-activating signal transduction	7.62E-07	2.43E-05	82/1408
GO:0009611	response to wounding	4.88E-07	1.62E-05	103/1408
GO:0002698	negative regulation of immune effector process	4.53E-07	1.51E-05	27/1408
GO:0002274	myeloid leukocyte activation	4.18E-07	1.41E-05	37/1408
GO:0002366	leukocyte activation involved in immune response	2.40E-07	8.48E-06	47/1408
GO:0002696	positive regulation of leukocyte activation	2.00E-07	7.24E-06	60/1408
GO:0001819	positive regulation of cytokine production	1.95E-07	7.15E-06	71/1408
GO:0002263	cell activation involved in immune response	1.10E-07	4.24E-06	48/1408
GO:0050866	negative regulation of cell activation	9.79E-08	3.81E-06	38/1408
GO:0045087	innate immune response	9.12E-08	4.00E-05	40/201
GO:0002687	positive regulation of leukocyte migration	9.09E-08	3.57E-06	34/1408
GO:0031349	positive regulation of defense response	7.88E-08	3.16E-06	77/1408
GO:0043207	response to external biotic stimulus	4.40E-08	2.39E-05	39/201
GO:0050867	positive regulation of cell activation	2.69E-08	1.16E-06	64/1408
GO:0002695	negative regulation of leukocyte activation	1.86E-08	8.23E-07	36/1408
GO:2000147	positive regulation of cell motility	8.98E-09	4.38E-07	85/1408
GO:0040017	positive regulation of locomotion	7.94E-09	3.97E-07	86/1408
GO:0048514	blood vessel morphogenesis	7.28E-09	3.68E-07	97/1408
GO:0003013	circulatory system process	5.90E-09	3.17E-07	78/1408
GO:0002685	regulation of leukocyte migration	4.86E-09	2.65E-07	44/1408
GO:0044236	multicellular organism metabolic process	4.85E-09	2.65E-07	37/1408
GO:0022409	positive regulation of cell-cell adhesion	4.45E-09	2.52E-07	56/1408
GO:0051272	positive regulation of cellular component movement	2.61E-09	1.56E-07	88/1408
GO:0001525	angiogenesis	1.83E-09	1.19E-07	88/1408
GO:0001568	blood vessel development	1.57E-09	1.06E-07	111/1408
GO:0002694	regulation of leukocyte activation	8.25E-10	6.14E-08	87/1408
GO:0046649	lymphocyte activation	5.58E-10	4.74E-08	113/1408
GO:0045785	positive regulation of cell adhesion	3.01E-10	2.77E-08	80/1408
GO:0070661	leukocyte proliferation	2.51E-10	2.42E-08	62/1408
GO:0050865	regulation of cell activation	1.44E-10	1.42E-08	94/1408
GO:0002697	regulation of immune effector process	6.15E-11	7.25E-09	69/1408
GO:0022407	regulation of cell-cell adhesion	3.61E-11	4.81E-09	82/1408
GO:0001817	regulation of cytokine production	2.55E-11	3.74E-09	107/1408
GO:0050777	negative regulation of immune response	8.55E-12	1.34E-09	36/1408
GO:0002683	negative regulation of immune system process	3.42E-13	8.07E-11	84/1408
GO:0030155	regulation of cell adhesion	1.41E-14	4.61E-12	132/1408
GO:0043062	extracellular structure organization	1.02E-14	3.59E-12	85/1408

Enriched CD-specific pathways

GO:0060135	maternal process involved in female pregnancy	0.007553	0.04715	13/1408
GO:0001912	positive regulation of leukocyte mediated cytotoxicity	0.007514	0.04715	8/1408
GO:0051047	positive regulation of secretion	0.007435	0.047035	47/1408
GO:0035637	multicellular organismal signaling	0.00709	0.045466	21/1408
GO:0002577	regulation of antigen processing and presentation	0.006952	0.044649	7/1408
GO:0060711	labyrinthine layer development	0.006516	0.042293	12/1408
GO:0048732	gland development	0.006342	0.041548	58/1408
GO:0048565	digestive tract development	0.006323	0.041548	23/1408
GO:0031589	cell-substrate adhesion	0.00599	0.039962	48/1408
GO:0060348	bone development	0.004905	0.034876	30/1408
GO:0035239	tube morphogenesis	0.004597	0.033078	53/1408
GO:0032835	glomerulus development	0.00434	0.03155	14/1408
GO:0045596	negative regulation of cell differentiation	0.003991	0.029702	77/1408
GO:0032846	positive regulation of homeostatic process	0.00372	0.028099	33/1408
GO:0032411	positive regulation of transporter activity	0.003407	0.026347	14/1408
GO:0032970	regulation of actin filament-based process	0.003281	0.02546	49/1408
GO:0090288	negative regulation of cellular response to growth factor stimulus	0.002842	0.023313	23/1408
GO:0034104	negative regulation of tissue remodeling	0.002777	0.023069	7/1408
GO:0098901	regulation of cardiac muscle cell action potential	0.002777	0.023069	7/1408
GO:0035383	thioester metabolic process	0.002632	0.02232	19/1408
GO:0030278	regulation of ossification	0.002517	0.021719	32/1408
GO:0045648	positive regulation of erythrocyte differentiation	0.002241	0.019538	8/1408
GO:0090066	regulation of anatomical structure size	0.002185	0.01928	64/1408
GO:0008217	regulation of blood pressure	0.001946	0.017725	26/1408
GO:0003012	muscle system process	0.001593	0.01515	49/1408
GO:0051781	positive regulation of cell division	0.001579	0.040389	6/201
GO:0031341	regulation of cell killing	0.001568	0.014988	13/1408
GO:0045646	regulation of erythrocyte differentiation	0.001492	0.014463	11/1408
GO:0034103	regulation of tissue remodeling	0.001475	0.014463	15/1408
GO:0051270	regulation of cellular component movement	0.00122	0.035737	29/201
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0.001096	0.033383	8/201
GO:0042391	regulation of membrane potential	0.000992	0.010586	40/1408
GO:0051222	positive regulation of protein transport	0.000987	0.01058	66/1408
GO:0098900	regulation of action potential	0.000887	0.009827	10/1408
GO:0090287	regulation of cellular response to growth factor stimulus	0.000657	0.007971	39/1408
GO:0098609	cell-cell adhesion	0.000631	0.023428	39/201
GO:0001910	regulation of leukocyte mediated cytotoxicity	0.000582	0.007184	13/1408

GO:0001775	cell activation	0.000501	0.019403	33/201
GO:1901698	response to nitrogen compound	0.000477	0.018752	31/201
GO:0002367	cytokine production involved in immune response	0.000417	0.005654	19/1408
GO:0016477	cell migration	0.000368	0.016371	41/201
GO:0045730	respiratory burst	0.000342	0.004803	10/1408
GO:0002027	regulation of heart rate	0.000247	0.003728	17/1408
GO:0030029	actin filament-based process	0.000235	0.003603	92/1408
GO:0045124	regulation of bone resorption	0.000143	0.002442	12/1408
GO:0002684	positive regulation of immune system process	0.000131	0.009036	38/201
GO:0031348	negative regulation of defense response	0.000106	0.001868	28/1408
GO:0034764	positive regulation of transmembrane transport	9.17E-05	0.001686	23/1408
GO:0098602	single organism cell adhesion	6.73E-05	0.005212	32/201
GO:0008284	positive regulation of cell proliferation	6.56E-05	0.005212	31/201
GO:0050776	regulation of immune response	6.50E-05	0.005212	36/201
GO:0031016	pancreas development	2.37E-05	0.020337	7/165
GO:0034330	cell junction organization	2.15E-05	0.000496	49/1408
GO:0031018	endocrine pancreas development	1.41E-05	0.020337	6/165

Enriched UC-specific pathways

GO:1900076	regulation of cellular response to insulin stimulus	0.006513	0.049867	11/1093
GO:0008037	cell recognition	0.006261	0.048376	16/1093
GO:0042092	type 2 immune response	0.006257	0.043929	7/917
GO:0044246	regulation of multicellular organismal metabolic process	0.006257	0.043929	7/917
GO:0045912	negative regulation of carbohydrate metabolic process	0.006257	0.043929	7/917
GO:0044703	multi-organism reproductive process	0.006133	0.047508	70/1093
GO:0007588	excretion	0.005713	0.045648	9/1093
GO:1990267	response to transition metal nanoparticle	0.005545	0.045046	19/1093
GO:0030324	lung development	0.004891	0.041272	25/1093
GO:0061383	trabecula morphogenesis	0.004601	0.035421	10/917
GO:0035966	response to topologically incorrect protein	0.004556	0.038829	27/1093
GO:0007566	embryo implantation	0.004543	0.035393	9/917
GO:0010632	regulation of epithelial cell migration	0.004375	0.034477	22/917
GO:0060760	positive regulation of response to cytokine stimulus	0.004165	0.036683	8/1093
GO:0030279	negative regulation of ossification	0.00401	0.036146	12/1093
GO:0001704	formation of primary germ layer	0.003652	0.033938	19/1093
GO:0009268	response to pH	0.003572	0.033417	7/1093
GO:0060713	labyrinthine layer morphogenesis	0.003572	0.033417	7/1093
GO:0005975	carbohydrate metabolic process	0.003102	0.030649	59/1093

GO:0002702	positive regulation of production of molecular mediator of immune response	0.003072	0.026092	11/917
GO:0048871	multicellular organismal homeostasis	0.003037	0.030078	36/1093
GO:0009612	response to mechanical stimulus	0.003004	0.030032	26/1093
GO:0019883	antigen processing and presentation of endogenous antigen	0.002942	0.025093	5/917
GO:0045454	cell redox homeostasis	0.002871	0.024852	13/917
GO:0007568	aging	0.002637	0.027203	35/1093
GO:0002643	regulation of tolerance induction	0.002518	0.026232	6/1093
GO:0030195	negative regulation of blood coagulation	0.002492	0.026146	9/1093
GO:1900047	negative regulation of hemostasis	0.002492	0.026146	9/1093
GO:0051048	negative regulation of secretion	0.002405	0.025693	25/1093
GO:0002279	mast cell activation involved in immune response	0.002226	0.020262	9/917
GO:0002532	production of molecular mediator involved in inflammatory response	0.002226	0.020262	9/917
GO:0048286	lung alveolus development	0.002006	0.022386	10/1093
GO:2000241	regulation of reproductive process	0.001894	0.02166	18/1093
GO:0051051	negative regulation of transport	0.00181	0.021065	50/1093
GO:0006949	syncytium formation	0.001349	0.013527	8/917
GO:0050819	negative regulation of coagulation	0.001099	0.014717	10/1093
GO:0051238	sequestering of metal ion	0.000981	0.010371	17/917
GO:0009991	response to extracellular stimulus	0.000959	0.013862	51/1093
GO:1901615	organic hydroxy compound metabolic process	0.000922	0.013486	45/1093
GO:0010942	positive regulation of cell death	0.000899	0.013416	72/1093
GO:0035821	modification of morphology or physiology of other organism	0.000708	0.011104	19/1093
GO:0040008	regulation of growth	0.000532	0.00912	69/1093
GO:0010634	positive regulation of epithelial cell migration	0.000476	0.005884	18/917
GO:0050792	regulation of viral process	0.000379	0.00698	39/1093
GO:0022600	digestive system process	0.000282	0.00558	15/1093
GO:1903409	reactive oxygen species biosynthetic process	0.000231	0.003343	16/917
GO:0043903	regulation of symbiosis, encompassing mutualism through parasitism	0.000156	0.003607	42/1093
GO:0002377	immunoglobulin production	0.000102	0.001677	17/917
GO:1903036	positive regulation of response to wounding	6.27E-05	0.001674	13/1093
GO:0050820	positive regulation of coagulation	5.35E-05	0.001485	9/1093
GO:0030194	positive regulation of blood coagulation	2.82E-05	0.000918	9/1093
GO:1900048	positive regulation of hemostasis	2.82E-05	0.000918	9/1093
GO:0030193	regulation of blood coagulation	9.33E-06	0.000428	19/1093
GO:1900046	regulation of hemostasis	9.33E-06	0.000428	19/1093
GO:0050818	regulation of coagulation	5.72E-06	0.000306	20/1093

*: indicating the two conserved pathways in UC group.

Appendix C

Table 7S. Enriched pathways for CRC.

ID	Description	pvalue	p.adjust	GeneRatio
GO:0007062	sister chromatid cohesion	0.004055	0.048549	22/1114
GO:0021549	cerebellum development	0.003799	0.047023	12/1114
GO:2000257*#	regulation of protein activation cascade	0.003777	0.04797	Dec-18
GO:0051781*#	positive regulation of cell division	0.003599	0.045346	13/1114
GO:0048732*	gland development	0.003583	0.045346	49/1114
GO:0072593*#	reactive oxygen species metabolic process	0.003499	0.045865	63/2618
GO:0045992	negative regulation of embryonic development	0.003424	0.044295	8/1114
GO:0051983	regulation of chromosome segregation	0.003387	0.044084	16/1114
GO:0030902	hindbrain development	0.00324	0.042554	18/1114
GO:0051321	meiotic cell cycle	0.003115	0.041651	25/1114
GO:0051445	regulation of meiotic cell cycle	0.003114	0.041915	13/2618
GO:0050818#	regulation of coagulation	0.003106	0.041651	15/1114
GO:0030324#	lung development	0.003083	0.041651	26/1114
GO:0001822	kidney development	0.003065	0.041685	71/2618
GO:0030850	prostate gland development	0.003049	0.041555	10/1114
GO:0022600#	digestive system process	0.003045	0.041685	23/2618
GO:1990267#	response to transition metal nanoparticle	0.002975	0.040923	37/2618
GO:0007507	heart development	0.002837	0.039279	56/1114
GO:0030728*	ovulation	0.00278	0.038624	6/1114
GO:0001942	hair follicle development	0.00278	0.039079	25/2618
GO:0022404	molting cycle process	0.00278	0.039079	25/2618
GO:0060348*	bone development	0.002733	0.038462	26/1114
GO:0045926	negative regulation of growth	0.002328	0.034689	58/2618
GO:0043207*#	response to external biotic stimulus	0.002283	0.033765	82/1114
GO:1901343*#	negative regulation of vasculature development	0.002256	0.033606	14/1114
GO:0033002	muscle cell proliferation	0.00222	0.033985	47/2618
GO:0022037	metencephalon development	0.002134	0.033306	23/2618
GO:0051302	regulation of cell division	0.002117	0.032518	19/1114
GO:0002686	negative regulation of leukocyte migration	0.002086	0.032173	9/1114
GO:1903046	meiotic cell cycle process	0.002048	0.031814	22/1114
GO:0090068	positive regulation of cell cycle process	0.002047	0.032335	62/2618
GO:0098869	cellular oxidant detoxification	0.002032	0.032217	28/2618
GO:0007126	meiotic nuclear division	0.001986	0.031066	21/1114
GO:0002443*#	leukocyte mediated immunity	0.001885	0.030633	72/2618
GO:0010810	regulation of cell-substrate adhesion	0.001865	0.029492	28/1114
GO:0045596*	negative regulation of cell differentiation	0.001843	0.029256	65/1114
GO:0009566	fertilization	0.001739	0.028548	15/1114

GO:0006730	one-carbon metabolic process	0.001731	0.029202	13/2618
GO:0035036	sperm-egg recognition	0.001671	0.028173	6/1114
GO:0007584*#	response to nutrient	0.001583	0.027975	50/2618
GO:0009994	oocyte differentiation	0.001497	0.025543	9/1114
GO:0008608	attachment of spindle microtubules to kinetochore	0.001452	0.026483	14/2618
GO:0002250*#	adaptive immune response	0.001424	0.026412	84/2618
GO:0007566#	embryo implantation	0.001416	0.024548	11/1114
GO:0050803	regulation of synapse structure or activity	0.001389	0.02606	25/2618
GO:0048771*#	tissue remodeling	0.001361	0.023826	23/1114
GO:0051222*	positive regulation of protein transport	0.001324	0.025293	109/2618
GO:0010632#	regulation of epithelial cell migration	0.001277	0.022584	27/1114
GO:0001890	placenta development	0.001241	0.024771	44/2618
GO:0000075	cell cycle checkpoint	0.001226	0.024573	65/2618
GO:0050878*#	regulation of body fluid levels	0.001202	0.021794	55/1114
GO:0048871*#	multicellular organismal homeostasis	0.001145	0.02338	74/2618
GO:0060688	regulation of morphogenesis of a branching structure	0.001142	0.021153	12/1114
GO:0030278	regulation of ossification	0.001126	0.021135	28/1114
GO:0021700	developmental maturation	0.001126	0.021135	30/1114
GO:0007420	brain development	0.001023	0.019793	68/1114
GO:0034975	protein folding in endoplasmic reticulum	0.000932	0.018367	6/1114
GO:0045787	positive regulation of cell cycle	0.000877	0.019799	84/2618
GO:0007588#	excretion	0.000831	0.01932	17/2618
GO:0060322	head development	0.000711	0.014806	72/1114
GO:0000910	cytokinesis	0.000688	0.01461	22/1114
GO:0009991#	response to extracellular stimulus	0.000663	0.016943	104/2618
GO:1900182	positive regulation of protein localization to nucleus	0.000624	0.01373	23/1114
GO:0007051	spindle organization	0.000573	0.012792	20/1114
GO:0035295*#	tube development	0.000472	0.011122	70/1114
GO:0010035*#	response to inorganic substance	0.000451	0.01071	58/1114
GO:0010817	regulation of hormone levels	0.000417	0.012929	101/2618
GO:0007565*#	female pregnancy	0.000392	0.009534	26/1114
GO:0009612#	response to mechanical stimulus	0.000381	0.009329	29/1114
GO:0007059	chromosome segregation	0.000374	0.009253	46/1114
GO:0071674*#	mononuclear cell migration	0.000356	0.008929	16/1114
GO:0009617*#	response to bacterium	0.000347	0.008841	58/1114
GO:0051965	positive regulation of synapse assembly	0.000338	0.01115	15/2618
GO:0031099	regeneration	0.000334	0.008614	26/1114
GO:0051963	regulation of synapse assembly	0.000329	0.011127	19/2618
GO:2000241#	regulation of reproductive process	0.00032	0.008348	20/1114

GO:0048608	reproductive structure development	0.000304	0.010693	103/2618
GO:1904951*#	positive regulation of establishment of protein localization	0.000261	0.009705	127/2618
GO:0044770	cell cycle phase transition	0.000251	0.006939	65/1114
GO:0033628	regulation of cell adhesion mediated by integrin	0.000251	0.006939	12/1114
GO:0007568#	aging	0.000246	0.006875	39/1114
GO:0019953	sexual reproduction	0.000203	0.005876	64/1114
GO:0044246#	regulation of multicellular organismal metabolic process	0.000199	0.005779	10/1114
GO:0044253*#	positive regulation of multicellular organismal metabolic process	0.000194	0.005722	7/1114
GO:0009636	response to toxic substance	0.000191	0.007633	60/2618
GO:1904018*#	positive regulation of vasculature development	0.000186	0.005722	25/1114
GO:0032206	positive regulation of telomere maintenance	0.000156	0.005189	13/1114
GO:0001556	oocyte maturation	0.000153	0.005127	8/1114
GO:0006805	xenobiotic metabolic process	0.000147	0.006524	28/2618
GO:0007586*#	digestion	0.000146	0.006524	39/2618
GO:0042493*#	response to drug	0.000143	0.004907	54/1114
GO:0003006	developmental process involved in reproduction	0.000136	0.004732	69/1114
GO:0006403	RNA localization	0.000133	0.049701	77/3344
GO:0090130*#	tissue migration	0.00013	0.004593	38/1114
GO:0051301	cell division	0.000126	0.004554	74/1114
GO:2000242*#	negative regulation of reproductive process	0.000122	0.004503	12/1114
GO:0009410	response to xenobiotic stimulus	0.000121	0.005947	31/2618
GO:0008285*#	negative regulation of cell proliferation	0.00011	0.004097	80/1114
GO:0019725*#	cellular homeostasis	0.000109	0.004097	86/1114
GO:0098813	nuclear chromosome segregation	0.000102	0.003929	42/1114
GO:0006959*#	humoral immune response	9.94E-05	0.005421	44/2618
GO:0032844*#	regulation of homeostatic process	7.79E-05	0.003201	57/1114
GO:0032101*#	regulation of response to external stimulus	7.78E-05	0.003201	82/1114
GO:0045785*#	positive regulation of cell adhesion	6.15E-05	0.002718	54/1114
GO:0001763*#	morphogenesis of a branching structure	5.68E-05	0.002591	34/1114
GO:0032846*	positive regulation of homeostatic process	5.37E-05	0.002475	33/1114
GO:0001666*#	response to hypoxia	4.79E-05	0.002282	44/1114
GO:1904872	regulation of telomerase RNA localization to Cajal body	3.84E-05	0.00189	8/1114
GO:1904874	positive regulation of telomerase RNA localization to Cajal body	3.84E-05	0.00189	8/1114
GO:0048729*#	tissue morphogenesis	3.12E-05	0.001746	80/1114
GO:0006520	cellular amino acid metabolic process	2.90E-05	0.001654	47/1114
GO:0040008#	regulation of growth	2.67E-05	0.002058	149/2618
GO:0032204	regulation of telomere maintenance	2.33E-05	0.001382	17/1114
GO:0010564	regulation of cell cycle process	2.25E-05	0.001376	75/1114

GO:0070482*#	response to oxygen levels	1.98E-05	0.001242	48/1114
GO:0042330*#	taxis	1.79E-05	0.001198	71/1114
GO:0006935*#	chemotaxis	1.62E-05	0.001153	71/1114
GO:0009611*#	response to wounding	1.46E-05	0.001069	81/1114
GO:0010942#	positive regulation of cell death	1.33E-05	0.000992	81/1114
GO:0001704#	formation of primary germ layer	1.04E-05	0.000805	25/1114
GO:0002687*#	positive regulation of leukocyte migration	1.04E-05	0.000805	26/1114
GO:0032103*#	positive regulation of response to external stimulus	1.01E-05	0.000805	43/1114
GO:0050673*#	epithelial cell proliferation	9.29E-06	0.000775	52/1114
GO:0097529*#	myeloid leukocyte migration	8.82E-06	0.000749	31/1114
GO:0050920*#	regulation of chemotaxis	7.32E-06	0.000634	34/1114
GO:0050921*#	positive regulation of chemotaxis	5.16E-06	0.000497	27/1114
GO:0031589*	cell-substrate adhesion	2.66E-06	0.000311	51/1114
GO:0051093*#	negative regulation of developmental process	2.61E-06	0.000311	95/1114
GO:0030155*#	regulation of cell adhesion	1.90E-06	0.000245	89/1114
GO:0006956*#	complement activation	1.19E-06	0.000232	26/2618
GO:0040013*#	negative regulation of locomotion	1.03E-06	0.00014	45/1114
GO:0044703#	multi-organism reproductive process	8.78E-07	0.000123	88/1114
GO:0030595*#	leukocyte chemotaxis	6.79E-07	9.80E-05	38/1114
GO:2000146*#	negative regulation of cell motility	5.01E-07	7.48E-05	43/1114
GO:0001101*#	response to acid chemical	4.75E-07	7.48E-05	46/1114
GO:0051271*#	negative regulation of cellular component movement	4.75E-07	7.48E-05	46/1114
GO:0033627*#	cell adhesion mediated by integrin	3.19E-07	5.54E-05	19/1114
GO:0010243*#	response to organonitrogen compound	3.14E-07	5.54E-05	97/1114
GO:0040017*#	positive regulation of locomotion	2.33E-07	4.40E-05	69/1114
GO:2000147*#	positive regulation of cell motility	1.36E-07	2.81E-05	69/1114
GO:0001503#	ossification	1.35E-07	2.81E-05	59/1114
GO:0002685#	regulation of leukocyte migration	9.66E-08	2.20E-05	36/1114
GO:0051272#	positive regulation of cellular component movement	7.07E-08	1.80E-05	71/1114
GO:0001525*#	angiogenesis	1.72E-09	5.34E-07	75/1114
GO:0048514*#	blood vessel morphogenesis	1.54E-09	5.13E-07	84/1114
GO:0022613	ribonucleoprotein complex biogenesis	6.57E-10	2.95E-06	181/3344
GO:0001568*#	blood vessel development	3.62E-12	1.96E-09	101/1114
GO:0044236*#	multicellular organism metabolic process	9.69E-13	7.00E-10	38/1114
GO:0043062*#	extracellular structure organization	3.57E-21	7.75E-18	85/1114

Note: * indicates shared pathways with CD; # indicates shared pathways with UC;