Computational Analysis of the Colonic Transciptome & In Vitro Biomarker Analysis Using a Novel Microfluidic Quantum Dot Linked Immunoassay

A Thesis

Submitted to the Faculty

of

Drexel University

by

Peter M. Clark

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

March, 2012

DEDICATION

This thesis is dedicated in memory of my co-advisor, mentor and friend Elisabeth Papazoglou. The lessons I have learned under her study extended well beyond the classroom, and will remain with me forever.

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Aydin Tozeren, Dr. Elisabeth Papazoglou and Dr. Moses Noh for their unwavering support and guidance throughout my research, without whom none of this work would have been possible. The knowledge and lessons I have learned extend beyond the classroom and will remain with me forever, helping me realize my full potential as a biomedical engineer. I would also like to thank my colleagues and collaborators throughout my involvement in a number of research projects at Drexel University including Dr. Leonid Zubkov, Rob Hart, Dohyoung Oh, David Hansberry and Chengjie Yu. My collaborations with all of my colleagues and their support of my work has helped shape the scientist I am today and have left a permanent impact on the course of my career within science and engineering. I would also like to thank Dr. Lim and Dr. Seliktar for their counseling, support and guidance throughout my time as a graduate student at Drexel University. Lastly, I would like to acknowledge my parents Michael & Marie Clark for their perpetual support of my academic pursuits and aspirations to pursue a Ph.D within the field of biomedical engineering.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Computational Analysis of the Colonic Transciptome & In Vitro Biomarker Analysis Using a Novel Microfluidic Quantum Dot Linked Immunoassay

Peter M. Clark

Aydin Tozeren, Ph.D. & Elisabeth Papazoglou, Ph.D.

DNA microarray technology facilitates the high throughput analysis of transcriptional disease regulation by measuring the relative expression levels of transcripts present within a tissue. While such computational approaches have been used to study the genetic regulation of a variety of illnesses, such studies often suffer from inadequate patient sample sizes and statistical power resulting in conflicting results and lab-specific bias. In order to overcome these limitations and fully utilize the wealth of publicly available genomic data, an integrated microarray analysis method was used to analyze and interpret microarray data in the context of colonic diseases including colorectal carcinoma (CRC) and inflammatory bowel disease (IBD).

The results of this work indicate widespread genetic perturbations related to IBD in which a variety of cell types are implicated including resident host enterocytes, innate and adaptive immune cells as well as native luminal microflora. Our work has identified subtle genetic differences between IBD phenotypes for the realization of disease specific therapeutic treatments as well as novel diagnostic biomarkers. Furthermore, our analysis has revealed significant overlap in the genetic regulation and predisposition to IBD, lupus, type 1 diabetes, graves disease and rheumatoid arthritis, providing the first genetic link between the enteropathic disease symptoms associated with IBD.

Druggable pathways involved in these diseases as well as known therapeutic drug targets were also analyzed for the potential repositioning of existing therapeutics for the treatment of IBD.

IBD patients are known to be at an elevated risk for developing colorectal carcinoma, with risk increasing with the duration of the disease. In order to better understand the phenotype shift from IBD to cancerous phenotypes, integrated microarray analysis was used to identify gene signatures, implicated pathways and novel discriminatory biomarkers for differentiating between IBD and CRC phenotypes. Our diagnostic panels were shown to accurately differentiate between phenotypes using an independent dataset for validation.

In order to transition the identified biomarkers to the clinic for diagnostic use, a novel microfluidic quantum dot linked immunosorbent assay platform was developed with enhanced surface chemistry and reaction kinetics. The developed prototype has the capability of multiplexed biomarker detection within clinically relevant samples for the stratification of disease phenotypes. In order to validate our design, human samples spiked with the fecal IBD biomarker lactoferrin were analyzed. Results indicate increased sensitivity and signal to noise ratios over our predicate device, with a reduction of the limit of detection. This proof of concept device shows great promise as a portable bedside diagnostic device for multiplexed biomarker analysis within a clinical setting.

CHAPTER 1

BACKGOUND & SIGNIFICANCE

1.1 THESIS ORGANIZATION

The organization of this thesis is article based in which this introductory chapter is succeeded by three freestanding manuscripts. The introductory chapter serves to both frame the thesis and provide the reader with relevant background information related to the topics covered herein. The introduction is followed by a computational, systems biology manuscript entitled, "Bioinformatics analysis reveals transcriptome and microRNA signatures, and drug repositioning targets for IBD and other autoimmune diseases." The manuscript identifies differentially expressed genes and pathways involved in the pathogenesis of IBD as well as therapeutic targets and genes related to the transcriptional regulation of other autoimmune diseases for therapeutic repositioning of existing drugs. The third chapter is a manuscript entitled, "Meta-Analysis of IBD and Colorectal Carcinoma Microarray Data Reveals Transcriptional Regulation, Disease Biomarkers and Regulatory Pathway Associations Between Disease Phenotypes," which identifies unique genes and pathways involved in the transition from IBD to colorectal carcinoma (CRC) as well as unique gene signature biomarker candidates for distinguishing between IBD phenotypes and the four stages of CRC. A biomarker identified within our analysis, was then assessed within spiked human samples using a novel high throughput microfluidic quantum dot linked immunoassay (QLISA) platform (chapter 4). This chapter describes

the design, fabrication and characterization of a second-generation microfluidic QLISA platform with improved surface functionalization to enhance immunoassay kinetics and biosensor performance. This work was presented as a conference paper presented at Micro-Total-Analytical-Systems (uTAS) 2011, one of the foremost international conferences on micro-scale devices. An IBD fecal biomarker, lactoferrin, was then detected within spiked human stool samples in order to characterize biosensor performance in the detection of IBD biomarkers. The concluding chapter summarizes my contributions to the field and addresses future research directions and concerns.

1.2 MOTIVATION & APPROACH

Inflammatory bowel disease is known to affect over 1.5 million people within the United States, although international incidence of the disease are on the rise [1, 2]. Current diagnosis of IBD requires a colonoscopy in combination with endoscopic, radiologic and histological tests. In order to circumvent such invasive procedures, both serological and fecal biomarker assays have been developed for the diagnosis of IBD but require analysis by lab technicians often at a centralized lab facility and often suffer from high false positive rates since many of the markers are inflammatory related proteins that may be elevated due to the presence of other systemic inflammatory conditions [3]. Furthermore, there is currently no biomarker that can accurately distinguish between the two phenotypes of IBD; ulcerative colitis (UC) and crohn's disease (CD).

In order to identify potential IBD biomarkers and better understand the transcriptional regulation of IBD as well as susceptibility genes leading to the progression from IBD to CRC, computational approaches were developed to integrate and analyze large sets of microarray data in a functional context. Biomarkers identified using this approach may then be used to develop clinical biomarker panels for the non-invasive stratification of colonic diseases including differentiating between IBD phenotypes and the stages of CRC.

Lastly, our patented QLISA system was enhanced using novel surface chemistry and translated onto a microfluidic platform to facilitate increased sample thoughput and multiplexed biomarker analysis and

integration with electrothermal flow components to enhance immunoassay kinetics for the realization of a rapid point of care immunosensor platform for the detection of IBD biomarkers.

CHAPTER 2:

Bioinformatics Analysis Reveals Transcriptome And MicroRNA Signatures, and Drug Repositioning Targets for IBD and other Autoimmune Diseases

2.1 ABSTRACT

Inflammatory bowel disease (IBD) is a complex disorder involving pathogen infection, host immune response, and altered enterocyte physiology. Incidences of IBD are increasing at an alarming rate in developed countries, warranting a detailed molecular portrait of IBD.

We used large-scale data, bioinformatics tools, and high throughput computations to obtain gene and microRNA signatures for Crohn's Disease (CD) and Ulcerative Colitis (UC). These signatures were then integrated with systemic literature review to draw a comprehensive portrait of IBD in relation to autoimmune diseases.

The top upregulated genes in IBD are associated with diabetogenesis (REG1A, REG1B), bacterial signals (TLRs, NLRs), innate immunity (DEFA6, IDO1, EXOSC1), inflammation (CXCLs), and matrix degradation (MMPs). The downregulated genes coded tight junction proteins (CLDN8), solute transporters (SLCs), and adhesion proteins. Genes highly expressed in UC compared to CD included anti-inflammatory ANXA1, transporter ABCA12, T cell activator HSH2D, and immunoglobulin IGHV4-34. Compromised metabolisms for processing of drugs, nitrogen, androgen and estrogen, and lipids in IBD correlated with increase in specific microRNA. Highly expressed IBD genes constituted targets of

drugs used in gastrointestinal cancers, viral infections, and autoimmunity disorders such as rheumatoid arthritis, and asthma.

This study presents a clinically relevant gene-level portrayal of IBD phenotypes and genetic co-regualtion of genes involved in other autoimmune diseases. This Study identifies potential therapeutic candidates for the repositioning of existing drugs to manage IBD. Integration of mice and human data point to altered B cell response as a cause for upregulation of genes in IBD involved in other aspects of immune defense such as interferon-inducible responses.

2.2 INTRODUCTION

Inflammatory bowel disease (IBD) is a debilitating illness associated with the altered regulation of the gastrointestinal mucosal immune system leading to intestinal inflammation in the presence of native luminal flora. Nearly 1.5 million people within the United States are affected by IBD, and incidence and prevalence of the disease continue to increase internationally [1, 2]. IBD is comprised of two distinct phenotypes; ulcerative colitis (UC) and Crohn's disease (CD), each of which have unique clinical manifestations despite sharing many genetic and consequently mechanistic features [4-6]. Although, the exact pathophysiology of IBD is not yet fully understood, the etiology of this disease is known to be multifactorial, driven by a number of genetic and environmental factors including loss of regulation of the host's innate immune response, and defects in mucosal barrier function [7].

There appears to be a correlation between IBD and autoimmune disorders. Over a quarter of IBD patients have been known to experience musculoskeletal complications resulting from various subtypes of arthritis [8, 9]. Moreover, microRNA (miRNA) expression changes have high similarity in autoimmune disorders including IBD [10]. The complex host-pathogen interactions within the colon are also known to play a vital role in the pathogenesis of IBD and drive differential gene expression. Recent studies of the intestinal microbiome have shown a shift in the composition of bacterial populations in patients with UC and CD and the involvement of bacterial-host interactions in the pathogenesis of IBD [11-18]. However, it still remains unclear how resident bacteria interact with

the host to activate and sustain chronic activation of the intestinal immune system through genetic regulation.

A system approach has already emerged in the literature for studying IBD within the context of autoimmune disorders. A comprehensive first draft of the human disease network (HDN) was constructed recently using the disease gene information in the literature [19]. This and other system models [20-25] identified disease-disease associations not readily observed in disease-phenotype based Medical Subject Headings (MeSH) disease classification tree. Some of these models utilized text search based data whereas others based the draft network on gene lists identified via literature and genome wide association studies [26, 27]. A number of other disease network drafts utilized protein-protein interaction networks in addition to gene lists to identify gene-disease modules [28, 29]. In all cases, however, the starting point is a set of gene lists specific for each disease. Hu and Agarwal [30] obtained such lists utilizing high throughput microarray analysis of thousands of genomic expression profiles existent in the literature. Their draft of the human disease-drug network brought out new possibilities for further investigation such as the use of malaria drugs in treating Crohn's disease and also showed the potential of known fold changes in transcript copy numbers in guiding the type of drugs (agonists, antagonists) to be targeting druggable proteins and pathways.

The present study benefits from the previous work on the human disease network and studies of global gene and miRNA expression profiling of IBD. It presents gene signature analysis of IBD subtypes UC

and CD in terms of altered pathways, biological processes, disease correlations, transcripts, and possible drug targets, all determined via bioinformatics analysis of integrated microarray and miRNA datasets with open-access datasets. Gene lists derived from microarray data have the advantage of having features such as statistical significance and fold changes to be used for additional annotation. However, previous disease and drug network modeling studies suggest potentially high levels of noise in predictions, extent of which depending on the accuracy of gene lists used to represent the specific diseases [30]. In that regard, our previous meta- and large-scale microarray analyses of cancer tissue datasets from hundreds of different laboratories point to accurate replication of non-microarray literature with microarray data [31]. Our gene lists derived from integrated datasets intersected with the shorter gene lists produced by Noble et al. [32, 33] with highly significant p values in hypergeometric test. Moreover, cellular pathways statistically enriched by Noble et al. gene lists turned out to be a subset of our predicted pathways. Our predictions of IBD associated genes captured non-microarray IBD research literature with significant p values. Taken together, these results illustrate the predictive potential of microarray datasets in identifying genes associated with CD and UC subtypes of IBD. The resulting gene lists come with statistical significance, rank, and fold changes; parameters, which are invaluable in biomarker discovery and drug repositioning and have been included in supplemental files.

2.3 METHODS

Microarray Datasets, Normalization & Probe To Gene Mapping

The microarray datasets used in this study were all in public domain and were generated from RNA isolated from the inflamed colonic mucosa of UC and CD patients and hybridized onto the Affymetrix Human Genome U133 Plus 2.0 GeneChip® Array platform for analysis. A total of 176 colonic mucosal microarray dataset samples were obtained from the NCBI Gene Expression Omnibus (GEO) [34, 35] in raw CEL format. Following the removal of duplicate samples, sample outliers were removed using principal component analysis (PCA) [36]. Two of the principal components accounted for 91% of the variance within the data set. There were a total of 15 outliers detected and removed from further analysis. The final sample population is comprised of 29 healthy control (HC) samples, 48 UC samples and 41 CD samples from six independent studies (GSE9452 [37], GSE9686 [38], GSE10191 [39], GSE10616 [40], GSE13367 [41], and GSE16879 [42]). Gene expression datasets were normalized using robust multi-array averaging (RMA) background adjustment, quantile normalization with median polishing and log2 transformed summarization procedures within MATLAB [43-45] and were mapped to Entrez gene IDs using a custom CDF file during the normalization process [46, 47]. Following pre-processing, batch effect removal was performed using the COMBAT source code in R [48].

Significant Genes

Two distinct microarray data analysis methods – Significance Analysis of Microarrays (SAM) [49] and the rank product (RP) method [50, 51] representing meta analysis – were performed in MATLAB and R (2.10.0), respectively, in order to identify statistically significant upregulated and downregulated genes in UC/HC and CD/HC comparisons. In SAM, statistically significant genes were determined using a fold change (FC) and p-value cutoff of 1.5 and 0.001 respectively over 1000 permutations [48]. For the RP method, only those genes with an enrichment p-value less than 10^{-6} were identified as significant.

Comparison With Gene Lists Obtained Using Other Microarray Platforms

The significant gene lists determined in this study were compared with corresponding lists obtained by Noble et al. comparing UC and CD biopsies to controls on the Agilent platform. In addition to the Agilent annotation file, NCBI, DAVID [52, 53], Clone/Gene ID Converter [54] and Source [55] databases were used to map the Agilent probe IDs to their corresponding Entrez gene IDs. The Agilent platform mapped to 18,989 genes, of which 17,211 were also present on the Affymetrix array. The CD and UC gene lists identified from our analysis were then compared to the two gene lists obtained from Noble et al. A hypergeometric test was used to determine the significance of the intersections in terms of p-value.

Literature Citation Algorithm

Differentially expressed gene lists were used in an automated PubMed text search algorithm [56]. Briefly, a query of the gene symbol and the applicable search term was conducted in PubMed abstracts for all genes available on the Affymetrix HG-U133 Plus 2.0 platform, limiting results to non-microarray literature. The six gene lists were then annotated with these results, identifying those genes that were cited in relation to the particular search terms. Random gene lists from the same platform of equal size to the lists under consideration were obtained and used as a control. The number of related genes in each of the random iterations was determined, and the mean and standard deviation were calculated from these values to obtain the parameters of a normal distribution. The expected value and the standard deviation were then used to compute the p-values for the significant association of each of our cancer gene lists with the known non-microarray literature.

Functional Enrichment

The physiological pathways and biological processes implicated in IBD were determined through the functional enrichment of significant gene lists using DAVID [52, 53]. KEGG and BioCarta pathway enrichment of the differentially expressed gene lists as well as for gene lists identified by Noble et al. were determined using a p-value cutoff of 0.01. KEGG color was used to generate enriched pathway images for a few selected enriched pathways [57]. Gene ontology (GO) biological process enrichment was performed using GO level 3 and a p-value cutoff of $5*10^{-5}$.

Identification Of Known Drug Targets

In order to identify significant genes that encode for known therapeutic targets, differentially expressed genes for both UC and CD (Union, table 2) were first converted to UNIPROT accession numbers using DAVID [52, 53] and then queried against a CHEMBL database [58]. A histogram of the targeted differentially expressed genes is presented in Figure 5 along with their respective fold change in UC and CD and the number of therapeutics targeting each significant gene product.

Gene-miRNA Interactome

Previous research has identified significant differentially expressed colonic miRNAs for both active Crohn's Colitis and active Ulcerative Colitis relative to healthy controls [59-63]. Overall, there were found to be 26 and 33 differentially expressed miRNA extracted from relevant literature for UC and CD respectively. The corresponding validated miRNA targets were found using miRWalk [64]. Targeted mRNAs were then intersected with our list of differentially expressed genes for IBD to find interactions in which miRNAs were upregulated and targeted mRNAs were downregulated and vice versa.

Disease Network Related To IBD

Statistically significant genes identified within this study were queried against the Genetic Association Database (GAD) to find associations of identified genes with other MeSH disease terms as annotated within published, peer reviewed genetic association studies

such as genome wide association studies (GWAS) [27]. The disease list was then further limited to include only those MeSH disease terms with the parent term "Autoimmune Diseases" in order to see linkages between differentially expressed genes within IBD and autoimmune disorders. The node size shown in the resulting network diagram is proportional to the degree of edges associated with each node.

2.4 RESULTS

In this study we integrated microarray data obtained from multiple laboratories (Table 1a) along with other publicly available data and various knowledge bases to better define the transcriptional regulation of UC and CD and discover genetic commonalities with other diseases. Each unique patient sample was hybridized onto the Affymetrix Human Genome U133 Plus 2.0 GeneChip® Array platform, which contains 54,675 oligonucleotide probes, covering over 47,000 human transcripts mapped onto 17,778 NCBI Entrez gene IDs [46]. These samples were merged into one dataset and normalized together prior to subsequent analysis to reduce laboratory specific noise and increase statistical power. Figure 1 shows the sample clustering based upon the principal components, which account for 82% of the variance within the dataset. Significant genes were identified using both SAM and RP (Table 1b) methods. PCA plots were generated using the Matlab code presented in Appendix A. The significant gene lists were then intersected, resulting in 1229 upregulated and 828 downregulated UC/HC genes, 1050 upregulated and 539 downregulated CD/HC genes (Table 1b). From these four lists, significant genes common to both UC and CD (common genes) and exclusive to either UC or CD (exclusive genes) were identified for each phenotype by finding the intersection and unique elements of each list. The resulting six gene lists were then used for subsequent analysis. The top 200 up- and down-regulated gene lists for UC/HC and CD/HC comparisons are provided as appendices E-D. Table 2 shows the top 40 genes present within each of the six comparisons, as ranked by fold

change. Since mucosal tissue is a composite, multiple cell types (enterocytes and inflammatory cells) contribute to the altered regulation of genes in these lists and provide a system level perspective of transcriptional regulation.

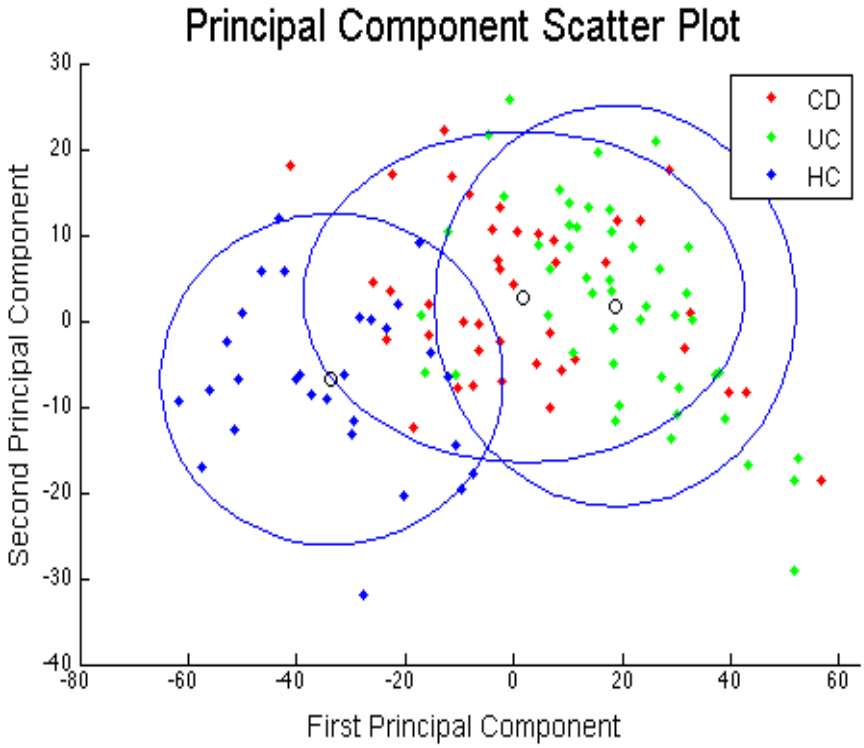


Figure 1. Principal component scatter plot of IBD phenotypes. The ellipse was constructed around the centroid of each data cluster with the semimajor and semiminor axis are equal to two standard deviations of the data in x and y respectively

1a)

GEO Accession #	PMID	HC	UC	CD	Total
GSE9452	19177426	5	8	0	13
GSE9686	18069684	8	5	11	24
GSE10191	18981162	11	8	0	19
GSE10616	18758464	11	10	32	53
GSE13367	19834973	10	8	0	18
GSE16879	19956723	6	24	19	49
Subtotal		51	63	62	176
Duplicates		19	13	11	43
Outliers		3	2	10	15
Total		**29**	**48**	**41**	**118**

1b)

Microarray Data Analysis		
	UC/HC	**CD/HC**
SAM Upregulated	1262	1063
SAM Downregulated	828	539
RP Upregulated	1315	1237
RP Downregulated	1044	867
SAM ∩ RP Upregulated	1229	1050
SAM ∩ RP Downregulated	828	539
Exclusive Up	264	85
Exclusive Down	349	60
Common Up	965	
Common Down	479	
Union	2202	

Table 1. An Overview of the microarray datasets used in the study. Table 1a presents for each dataset the corresponding NCBI GEO accession number and the PubMed citation ID (PMID) as well as the numbers of HC, UC, and CD samples in each dataset. Table 1b is summary of results of the microarray data analysis.

Significant genes in UC and CD

The Table 2A shows nearly 40 fold upregulation of pancreatic cancer associated [65] regenerating islet-derived growth factor genes REG1A and REG1B as the most highly upregulated genes in both phenotypes of IBD. REG1A is a biomarker for celiac disease [66] and was also linked to beta cell generation and damage in type I diabetes [67]. Moreover, REG1A and REG1B are known to induce CXCL6 expression in cancer [68]. Consistent with this observation, CXCL6 is also highly upregulated in our list. Approximately 20 percent of the top 40 commonly upregulated genes in IBD with highest fold rank are CXC chemokines involved in inflammation, including CXCL13, IL8, CXCL6, CXCL11, CXCL9, and CXCL5. The top chemokine in this list, CXCL13 promotes the migration of B lymphocytes to the tissue and is associated in the literature with systemic lupus [69], prostate cancer [70], and rheumatoid arthritis [71]. The second most highly expressed chemokine in IBD is IL8 (CXCL8), one of the major mediators of the inflammatory response. IL8 functions as a chemoattractant, and is also a potent angiogenic factor. The third most highly upregulated gene in our list, S100A8, codes a soluble IBD biomarker protein calprotectin [72-75]. Altered expression of this protein is also linked to cystic fibrosis. [76]. Other genes involved in inflammatory processes such as CHI3L1, ANXA1, TNIP3, BCL2A are also highly upregulated in both phenotypes of IBD. The matrix metalloproteinase (MMP) family proteins are abundant among the top forty most highly

2a)

Significant Genes Common to UC and CD

Upregulated					Downregulated				
Gene	**UC pval**	**CD pval**	**UC FC**	**CD FC**	**Gene**	**UC pval**	**CD pval**	**UC FC**	**CD FC**
REG1A	2.2E-04	3.5E-04	50.7	46.3	CLDN8	4.0E-04	8.0E-07	-22.5	-9.0
REG1B	1.5E-05	9.9E-05	26.6	40.3	OSTALPHA	1.5E-05	4.2E-08	-9.6	-7.5
S100A8	3.1E-04	2.5E-04	35.2	26.5	GUCA2B	7.9E-04	3.3E-04	-7.9	-5.3
MMP3	3.0E-04	2.3E-04	22.6	19.1	SLC16A9	5.0E-04	9.7E-04	-6.5	-6.0
CHI3L1	4.5E-04	6.1E-04	19.1	15.7	MT1M	4.7E-05	7.6E-08	-7.5	-4.9
MMP1	5.7E-05	2.9E-05	17.8	16.6	CA1	2.9E-04	4.7E-05	-8.3	-2.5
IDO1	4.0E-04	3.9E-04	12.1	12.4	GHR	3.2E-04	3.8E-06	-6.6	-3.9
MMP12	3.9E-04	1.5E-04	13.1	10.1	SLC30A10	4.9E-04	2.2E-04	-5.9	-4.5
CXCL13	1.0E-05	1.6E-06	11.4	10.8	ADH1C	5.0E-04	1.3E-07	-6.8	-2.8
FCGR3B	3.2E-04	2.1E-04	11.1	10.2	VLDLR	9.1E-04	9.7E-04	-4.8	-4.5
IL8	1.2E-04	3.4E-05	11.1	9.1	ANPEP	7.6E-06	4.4E-08	-5.1	-4.1
CXCL6	1.0E-04	3.7E-05	9.3	8.5	MEP1B	6.2E-07	2.1E-06	-5.0	-3.8
PI3	5.1E-04	2.8E-05	11.5	5.9	SLC3A1	8.4E-05	1.5E-05	-4.9	-3.8
ANXA1	6.8E-04	3.5E-05	11.2	5.3	HSD17B2	5.6E-04	9.0E-04	-4.2	-4.2
MMP10	6.0E-04	3.6E-04	9.6	6.3	CNTN3	5.9E-04	2.1E-06	-5.6	-2.5
TNIP3	5.0E-04	1.7E-04	10.3	5.7	BEST4	1.6E-04	1.3E-07	-5.1	-2.8
REG4	8.6E-04	4.7E-05	9.9	5.9	PRAP1	5.2E-05	8.2E-05	-4.0	-3.9
DEFA6	2.0E-05	3.3E-04	5.8	9.8	SLC16A1	4.6E-04	3.7E-04	-4.2	-3.7
IGHV1-69	6.1E-04	2.4E-05	9.6	4.4	PNLIPRP2	9.1E-08	7.0E-05	-4.9	-3.0
IGKV4-1	1.5E-04	2.6E-05	8.3	5.7	GBA3	1.7E-04	9.2E-05	-5.4	-2.3
MMP7	9.3E-04	5.6E-05	8.7	4.6	NPY1R	7.5E-05	2.8E-06	-4.8	-2.8
BCL2A1	3.7E-05	3.7E-05	6.2	7.1	UGT1A7	1.2E-04	9.5E-08	-4.6	-3.0
LOC651536	5.3E-05	1.3E-05	6.8	6.2	PTPRR	2.9E-05	6.8E-05	-3.5	-3.8
SERPINB5	2.5E-04	3.9E-08	8.6	4.3	MS4A12	3.7E-06	1.3E-07	-3.9	-3.3
CXCL11	1.3E-05	3.2E-07	6.4	6.3	CD177	7.6E-05	4.2E-08	-4.0	-3.1
IL1B	5.1E-04	5.3E-04	6.1	6.5	MT1G	9.3E-05	7.1E-08	-4.1	-2.9
C4ORF7	1.1E-07	2.7E-07	6.8	5.7	CWH43	3.4E-04	7.5E-06	-4.8	-2.2
SPINK4	4.2E-04	9.7E-05	6.9	5.2	OSTBETA	3.7E-04	1.0E-05	-3.9	-2.9
DMBT1	6.4E-05	1.2E-05	6.4	5.7	PITX2	9.4E-06	1.1E-05	-3.3	-3.4
IGHV3-11	1.1E-05	1.1E-04	5.0	6.6	MT1X	5.2E-05	3.2E-05	-3.4	-3.3
CTHRC1	9.7E-05	1.5E-05	6.1	5.4	TRHDE	1.0E-04	3.0E-07	-4.0	-2.6
LY96	3.1E-04	1.6E-04	6.1	5.3	SLC17A4	3.2E-04	4.5E-05	-3.8	-2.8
MNDA	5.5E-05	5.0E-06	5.8	5.6	PKIB	6.8E-04	6.1E-09	-4.0	-2.5
CXCL9	1.1E-06	3.5E-07	4.7	6.5	C10ORF116	5.3E-04	1.6E-04	-3.7	-2.8
C2CD4A	2.3E-04	1.6E-04	6.3	4.8	CA7	2.4E-04	7.7E-06	-3.7	-2.7
IGLV3-1	1.4E-04	9.2E-07	7.6	3.4	MT1F	1.3E-04	1.2E-06	-3.5	-2.8
SELL	1.9E-05	2.6E-06	5.6	5.2	FLJ32063	8.2E-04	4.1E-08	-4.0	-2.3
TNC	1.5E-04	4.8E-06	6.4	4.4	CDKN2B	7.8E-04	7.6E-04	-3.4	-2.8
CXCL5	2.2E-06	1.0E-06	5.4	5.3	EXPH5	5.9E-04	6.2E-08	-4.0	-2.3
HLA-DQB1	7.2E-04	3.4E-04	5.5	5.2	MEP1A	6.8E-04	1.2E-04	-3.4	-2.8

Table 2: Microarray gene signatures obtained from microarray analysis of IBD patients. Table 2a shows the top 40 significantly genes common in UC/HC and CD/HC comparisons ranked by average fold change (FC) and Table 2b presents the top 40 genes uniquely expressed in either UC or CD.

2b)

Ulcerative Colitis						Crohn's Disease					
Exclusive Upregulated			Exclusive Downregulated			Exclusive Upregulated			Exclusive Downregulated		
Gene	Pval	FC	Gene	Pval	FC	Gene	Pval	FC	Gene	P-Value	FC
DEFA5	2.2E-04	20.8	NR1H4	1.0E-06	-3.2	DUOX2	2.9E-04	12.1	AQP8	7.9E-04	-12.0
LCN2	9.1E-04	15.5	RETSAT	9.2E-04	-2.4	DUOXA2	8.0E-04	6.2	PCK1	1.1E-04	-6.9
UBD	8.1E-04	9.1	ACVR1C	5.6E-05	-2.4	TIMP1	1.0E-03	5.5	UGT2A3	1.4E-05	-5.8
SERPINB7	5.7E-07	4.7	EYA2	5.7E-08	-2.3	CD55	8.7E-04	4.2	LOC10050678	4.4E-04	-5.2
CD74	5.5E-04	4.1	PTPRO	4.9E-07	-2.3	TCN1	6.7E-05	4.1	GUCA2A	5.3E-05	-4.8
ASRGL1	7.8E-04	3.1	HOXA5	1.1E-05	-2.2	VNN1	5.5E-08	3.5	LAMA1	3.8E-04	-3.9
TFF1	1.3E-05	2.5	SGK223	4.5E-04	-2.1	FKBP11	9.9E-04	3.3	SLC26A2	2.3E-08	-3.7
DSG3	6.5E-08	2.5	MST1P9	1.5E-04	-2.1	PDZK1IP1	2.2E-05	3.1	CHP2	3.3E-05	-3.5
SLC6A20	3.1E-05	2.4	ZG16	1.9E-07	-2.1	DUSP4	3.2E-08	2.1	APOBEC3B	5.0E-04	-3.4
PIM3	8.3E-04	2.3	LOC283070	3.8E-05	-2.1	ARNTL2	1.6E-04	2.1	TUBAL3	2.9E-04	-3.2
SERPINB3	1.8E-05	2.3	HRCT1	7.8E-07	-2.0	ABI3BP	6.5E-06	2.0	PHLPP2	8.1E-04	-3.0
POSTN	1.1E-05	2.2	LOC100128822	5.7E-05	-2.0	FPR2	3.1E-07	2.0	RUNDC3B	6.9E-04	-3.0
KLK10	1.1E-05	2.2	LGALS2	3.3E-06	-2.0	IFNG	1.8E-05	1.9	DHRS11	3.4E-05	-2.7
SLCO1B3	1.3E-08	2.2	ZBTB10	2.0E-07	-2.0	P2RY14	4.9E-04	1.8	LRRC19	2.2E-05	-2.3
CTSE	1.3E-07	2.1	CHPT1	9.4E-04	-2.0	GOLGA8B	6.9E-04	1.7	CHRNA1	4.0E-04	-2.1
IGHV3-11	1.3E-05	2.1	MS4A8B	3.5E-07	-1.9	RGL1	1.1E-07	1.7	LOC10028809	3.2E-06	-2.1
IGLV4-60	9.9E-08	2.1	FLJ32810	1.4E-05	-1.9	CLEC4E	6.3E-07	1.7	HIGD1D	4.0E-04	-2.1
SERPINB4	1.3E-04	2.1	NDRG2	2.5E-04	-1.9	PTPN13	1.3E-04	1.7	SEMA5A	2.5E-04	-2.1
TESC	1.0E-04	2.1	NEDD4L	1.1E-05	-1.9	FYB	3.4E-07	1.7	WSCD1	5.4E-04	-2.1
VIP	1.6E-04	2.1	TCEA3	2.9E-05	-1.9	HCP5	3.3E-06	1.7	SNX30	1.9E-04	-2.0
CCL21	2.7E-05	2.0	PRR5L	7.6E-05	-1.9	CYP27B1	1.2E-06	1.7	SLC39A5	2.8E-04	-2.0
HBEGF	2.0E-06	2.0	IL1R2	1.3E-06	-1.9	GSDMB	3.1E-06	1.7	PXMP2	5.4E-04	-2.0
PMEPA1	2.5E-05	2.0	SPHAR	1.2E-04	-1.9	NLRC3	3.2E-04	1.6	SH3BGRL2	1.6E-04	-1.9
SPRR1B	4.9E-05	2.0	RNF125	7.5E-06	-1.9	TFEC	6.9E-07	1.6	CNNM2	1.3E-04	-1.9
ADAM9	2.1E-05	2.0	FMO4	1.6E-04	-1.9	ENTPD3	1.9E-06	1.6	PRDX6	4.6E-04	-1.8
SAA4	5.9E-05	1.9	TP53TG1	4.8E-05	-1.9	SLFN12	1.0E-07	1.6	CIDEC	9.3E-06	-1.8
ANXA3	8.5E-08	1.9	PSD3	6.2E-05	-1.8	DOCK10	6.3E-05	1.6	TNFRSF21	3.2E-06	-1.8
OSMR	4.3E-06	1.9	AMT	2.3E-04	-1.8	MAGEH1	8.5E-06	1.6	NUAK2	2.4E-04	-1.8
NMU	3.4E-07	1.9	KLHL34	2.6E-05	-1.8	IL7	2.1E-04	1.6	USP30	4.4E-04	-1.8
KRT7	5.6E-06	1.9	CFTR	1.1E-05	-1.8	MRPS21	2.4E-04	1.6	CDHR2	2.8E-07	-1.8
TM4SF1	2.3E-04	1.9	CYP2J2	7.0E-06	-1.8	LPAR6	8.8E-08	1.6	HAS3	3.7E-06	-1.7
XDH	2.1E-06	1.9	NRARP	1.1E-05	-1.8	TNF	2.9E-07	1.6	PPP2R3A	6.4E-05	-1.7
RND1	1.3E-07	1.9	RPS6KA6	1.9E-04	-1.8	LOC73142	3.1E-05	1.6	NT5E	1.5E-06	-1.7
SLC40A1	6.4E-06	1.8	FAM162A	6.8E-04	-1.8	IL18RAP	9.5E-05	1.6	PRSS23	1.7E-05	-1.6
PTP4A3	1.4E-05	1.8	METTL7A	1.2E-05	-1.8	BEX5	9.4E-04	1.6	C17ORF97	4.5E-06	-1.6
SLPI	5.0E-06	1.8	TLE2	6.0E-07	-1.8	IL2RA	6.6E-08	1.6	SDCBP2	3.1E-07	-1.6
MIF	3.5E-05	1.8	CFDP1	1.9E-04	-1.8	PVRIG	1.8E-05	1.6	POF1B	3.4E-05	-1.6
GJB3	1.1E-07	1.8	NIPSNAP3A	1.1E-04	-1.8	SLC16A6	4.5E-05	1.6	C20ORF54	9.6E-05	-1.6
ALDH1A2	4.4E-06	1.8	OTC	3.2E-07	-1.8	PGCP	3.7E-08	1.6	PRSS3	4.9E-07	-1.6
LRP8	3.7E-05	1.8	TSHZ1	1.5E-05	-1.8	SPG20	5.9E-08	1.6	MIER1	1.0E-04	-1.6

upregulated genes in IBD. MMP proteins partake in the breakdown of extracellular matrix in biological processes and in disease events, such as arthritis and metastasis [75, 77, 78]. Genes involved in the host defense, antimicrobial enzyme IDO1, exosome component PI3, and defensin DEFA6, were found to be highly expressed in both IBD phenotypes. Thus, top upregulated genes in IBD relate to processes such as diabetogenesis, innate immunity, chemokine-induced inflammation and adapted immunity, and breaking down of the extracellular matrix.

The gene in IBD with the most diminished expression is CLDN8, encoding a claudin tight junction protein. Its expression is downregulated in colorectal and renal carcinoma [79]. Diminished expression of CLDN8 in IBD suggests a breakdown in the physical barrier preventing solutes and water from passing freely between epithelial cell sheets. Several other downregulated genes in the list are also associated with the control of water and ion flux in intestinal epithelia: the solute transporter genes SLC16A9, SLC16A1, SLC30A10 and SLC3A1; and the organic solute transporter OSTα. These genes are all markers of enterocyte differentiation and indicate that IBD is associated with alterations in differentiated cell function.

Differentially expressed genes unique to each phenotype were also identified for both UC and CD (Table 2b). The most upregulated gene exclusive to UC codes for the antimicrobial defensin protein DEFA5. Defensins are abundant in the epithelia of mucosal surfaces of the intestine, respiratory tract, urinary tract, and vagina [80]. Other genes,

which are highly expressed in UC but not in CD include protease and protease inhibitors (SERPINB3, SERPINB4, SERPINB7, ASRGL1, KLK10, CTSE, and ADAM9), transporter genes (SLC6A2, SLCO1B3, ABCA12); anti-inflammatory ANXA1; and the zinc binding glycoprotein AZGP1P1. Several immunoglobulin heavy and light chain variable region encoding genes such as IGHV1-46, IGHV1-69, IGHV2-5, IGLV3-25, IGHV3-73 and IGHV4-34 are also highly expressed in UC but not in CD. Genes with high expression in CD but not in UC include oxireductase DUOX2 and DUOXA2, the protease TIMP1, and the abundant neutrophil granulocyte protein TCN1. Whether such genes can be used as biomarkers differentiating UC from CD will require further investigation using standard assays in stool samples and rectal swaps.

The genes with diminished expression in CD but with typical expression in UC include the water channel protein AQP8, also in the bile secretion pathway. This protein facilitates the diffusion of GPX2, a radical scavenger. The downregulation of Aqp8 in IBD mouse models was related to defending against severe oxidative stress [81]. Other genes downregulated expclusively in UC code for the cytosolic enzyme PCK1 involved in the regulation of gluconeogenesis; the basement membrane protein LAMA1; and the two solute transporters SLC26A2 and SLC39A5. Although downregulated genes are not optimal as biomarkers, nevertheless, when statistically enriched on signaling pathways, these genes can point to the molecular mechanisms of deregulation in IBD. The gene NR1H4 (FXR) is diminished in expression in UC (but not in CD). This gene codes a bile-dependent transcription factor in the bile secretion

pathway. Its downregulation may be responsible for the presence of excessive lipids in the stools of individuals with UC. Other genes poorly expressed in UC include those coding TGFB receptor ACVR1C; tyrosine phosphatase PTPRO, a tumor suppressor; and IL1R2, a cytokine receptor that belongs to the interleukin 1 receptor family.

Validation Of Microarray Predictions Using Independent Data

We have evaluated the statistical significance of our gene lists in comparison to other published gene lists derived from microarray studies not used in our analysis since these datasets were generated on a different platform: Agilent-012391 Whole Human Genome Oligo Microarray G4112A. Comparison of our gene lists to previously published significant gene lists provided by Noble et al. [32, 33] showed a high degree of overlap with significant p-values for each comparison based upon a hypergeometric test (Table 3a). The results shown in Table 3a point to a high degree of match between upregulated gene lists obtained using different platforms. The intersection for downregulated genes is smaller although still statistically significant.

We have also determined the enrichment of our top 200 significantly altered gene lists for both phenotypes (Appendices E-D) within relevant non-microarray literature. Table 3b shows that our predicted genes are highly statistically enriched within cancer literature, establishing a possible link between IBD and colorectal cancer. The match between our IBD upregulated gene list and non-microarray IBD literature is also good, with significant p values. The match between IBD research

literature and our downregulated gene lists lead to higher but still significant p values; a trend also observed when directly comparing our significant gene lists with those obtained by Noble et al

3a)

Cutoffs	HC/UC UP	HC/UC DOWN	HC/CD	HC/CD
top 50	1.30E-11	3.00E-03	6.95E-14	1.89E-04
top 100	< E-50	8.32E-03	< E-50	9.80E-04
top 200	4.44E-16	1.74E-02	< E-50	3.00E-03
top 500	2.00E-14	3.70E-04	< E-50	3.00E-02
All	3.50E-14	1.00E-04	< E-50	3.00E-02

3b)

Search Terms	List Name	P-value	List Articles	Search	Article
cancer NOT microarray	UC_DOWN	2.62E-03	19392	946	78
cancer NOT microarray	UC_UP	2.62E-09	68801	946	279
cancer NOT microarray	CD_UP	2.61E-10	65790	946	298
cancer NOT microarray	CD_DOWN	3.11E-04	15443	946	69
crohn's disease NOT microarray	UC_UP	2.62E-09	68801	946	279
crohn's disease NOT microarray	CD_UP	2.61E-10	65790	946	298
crohn's disease NOT microarray	UC_DOWN	2.62E-03	19392	946	78
crohn's disease NOT microarray	CD_DOWN	3.11E-04	15443	946	69
ulcerative colitis NOT microarray	CD_UP	2.61E-10	65790	946	298
ulcerative colitis NOT microarray	UC_DOWN	2.62E-03	19392	946	78
ulcerative colitis NOT microarray	UC_UP	2.62E-09	68801	946	279
ulcerative colitis NOT microarray	CD_DOWN	3.11E-04	15443	946	69

Table 3: Comparison of significant UC and CD transcriptional enrichment signatures to independent data. a) Statistics for the intersection of the results from our study with Noble et al. [32, 33]. b) Literature enrichment using the top 200 genes from each of the six significant gene lists (up UC, down UC, up CD, down CD, common up, and common down).

Activated Pathways And Processes In IBD

Commonly upregulated pathways involved in the pathogenesis of IBD (CD and UC) include cytokine-cytokine interaction, chemokine signaling, intestinal immune network for IgA production, adhesion and diabedesis of lymphocytes and granulocytes, cell adhesion molecules, complement and coagulation cascades (Table 4). Other upregulated pathways included the NOD-like receptor signaling pathway and the Toll-like Receptor pathway. NOD-like receptor pathway shown in Figure 2a relates bacterial effects on the cell surface (bacterial peptidglycans, pore forming toxins, and other bacterial secretion systems) to expression changes of chemokines and cytokines, which subsequently activate innate and adapted immunity. The nodes upregulated in both UC and CD were shown in yellow, those upregulated in CD, and UC, only shown in red, and brown, respectively. The Toll-like pathway shown in Figure 2b relates the recognition of conserved microbial components by Toll-like receptors to subsequent cellular events, which ultimately lead to inflammation and activation of sentinel immune cells. The genes CASP1, CASP5 and CARD6 are highly upregulated in both UC and CD along the NOD receptor signaling. The genes TLR1 and TLR2 are highly expressed in lipopolysaccharide (LPS) interactions in both phenotypes in the Toll-like pathway. The biological processes associated with these pathways are also significantly enriched within GO biological processes including immune defense response, cellular response to wounding, leukocyte activation, cell proliferation, and response to molecules of bacterial origin amongst others (Figure 3).

4a)

Pathway	UC UP	CD UP	Noble UC UP	Noble CD UP
Natural killer cell mediated cytotoxicity		9.0E-03		3.5E-03
Lck and Fyn tyrosine kinases in TCR Activation	8.2E-03			
The Co-Stimulatory T-cell Activation	6.9E-03			
Prion diseases	4.5E-03	8.4E-03		
Antigen Processing and Presentation	2.1E-03	6.1E-03		1.8E-08
Antigen processing and presentation	2.1E-03	6.1E-03		1.8E-08
NOD-like receptor signaling pathway	4.7E-03	6.0E-04		
IL-10 Anti-inflammatory Signaling Pathway		1.8E-03		
Toll-like receptor signaling pathway	1.7E-03	4.9E-04		
T Cytotoxic Cell Surface Molecules	2.4E-04	1.3E-03		
T Helper Cell Surface Molecules	2.4E-04	1.3E-03		
Autoimmune thyroid disease	5.1E-04	8.2E-04		6.6E-08
Focal adhesion	6.4E-04			
B cell receptor signaling pathway	2.1E-04	6.2E-04		
Asthma	5.5E-04	4.0E-05	5.5E-04	2.7E-04
Primary immunodeficiency	5.3E-05	1.1E-04		
Systemic lupus erythematosus	1.5E-04	2.5E-06		4.8E-03
Viral myocarditis	1.2E-04	8.3E-06	7.3E-03	6.3E-08
Type I diabetes mellitus	2.7E-05	1.4E-06	1.6E-03	4.9E-10
Allograft rejection	2.6E-05	1.3E-06	1.0E-03	1.1E-10
local acute inflammatory response	2.1E-05	7.3E-07		
ECM-receptor interaction	3.3E-07	1.4E-05		
B Lymphocyte Cell Surface Molecules	7.8E-06	3.9E-06		
Chemokine signaling pathway	1.8E-06	8.7E-06	3.7E-03	2.0E-04
Adhesion Molecules on Lymphocyte	3.9E-06	2.1E-06		
Neutrophil and Its Surface Molecules	3.9E-06	2.1E-06		
Adhesion and Diapedesis of Granulocytes	4.1E-06	6.2E-09	3.1E-03	8.4E-04
Hematopoietic cell lineage	3.3E-06	2.7E-08		
Leukocyte transendothelial migration	4.3E-07	2.3E-06		
Graft-versus-host disease	9.4E-07	3.2E-08	1.3E-03	2.4E-10
Complement and coagulation cascades	1.8E-07	5.0E-09		
Intestinal immune network for IgA production	1.1E-07	2.3E-08		
Monocyte and its Surface Molecules	7.4E-08	3.3E-08		
Adhesion and Diapedesis of Lymphocytes	7.8E-09	8.7E-08		
Cytokine-cytokine receptor interaction	6.2E-10	1.1E-10	5.2E-04	5.6E-04
Cell adhesion molecules (CAMs)	6.5E-12	3.0E-11		1.5E-06

4b)

Pathway	UC DOWN	CD DOWN
Arginine and proline metabolism	7.0E-03	
Tryptophan metabolism	5.4E-03	
Nuclear Receptors in Lipid Metabolism and Toxicity	5.1E-03	
Drug metabolism	7.0E-03	8.2E-04
Mitochondrial Carnitine Palmitoyltransferase (CPT) System	3.7E-04	6.0E-03
Starch and sucrose metabolism		2.2E-03
Selenoamino acid metabolism	1.7E-03	
Metabolism of xenobiotics by cytochrome P450		6.6E-04
Valine, leucine and isoleucine degradation	3.3E-07	4.8E-04
Nitrogen metabolism	1.9E-04	7.0E-05
PPAR signaling pathway	4.9E-05	1.1E-05
Propanoate metabolism	4.1E-06	
Butanoate metabolism	4.1E-06	
Fatty acid metabolism	4.3E-09	3.8E-07

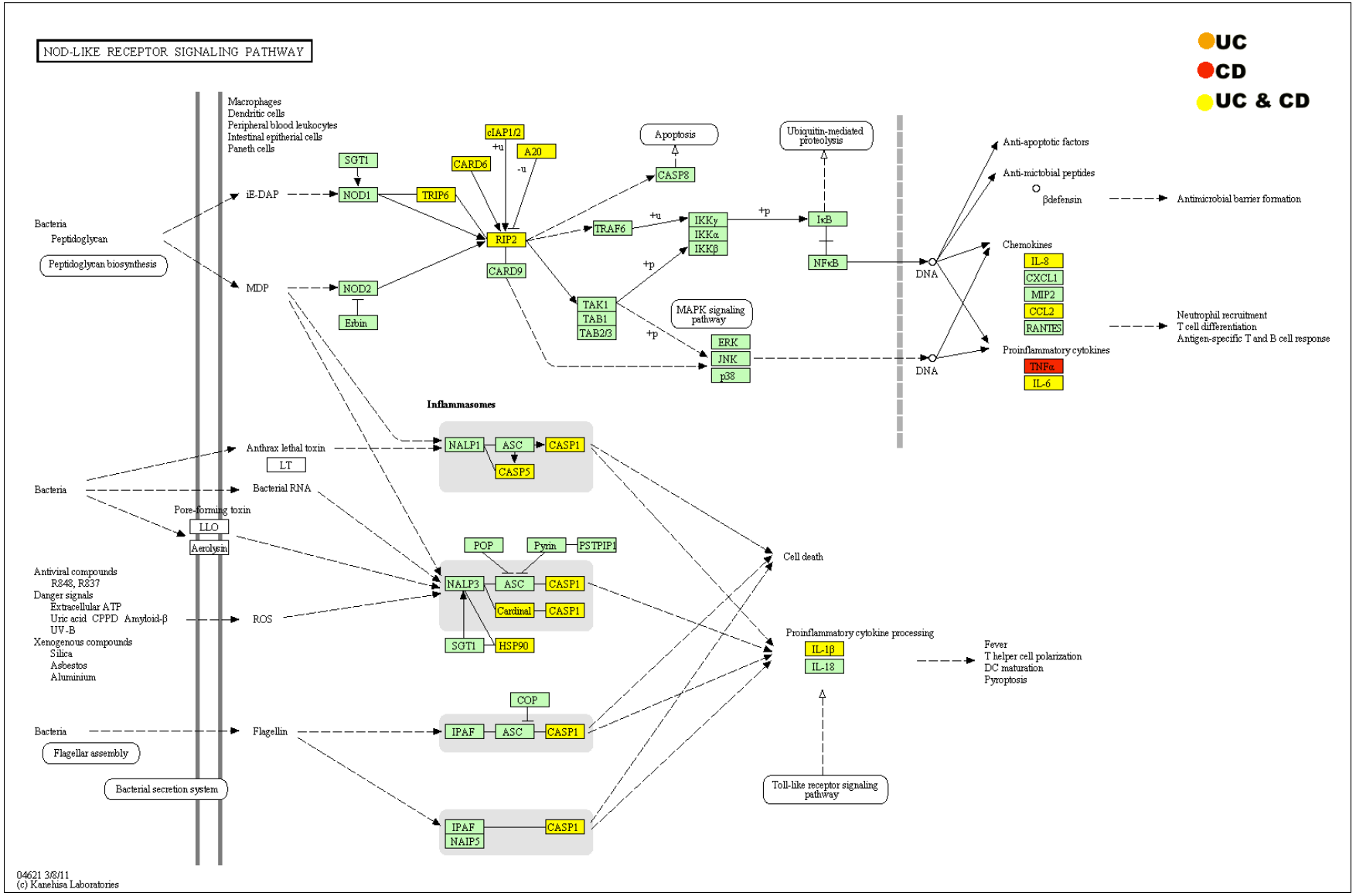
Table 4. KEGG and BioCarta cellular pathways statistically enriched by differentially expressed genes in UC/CD and CD/HC comparisons. Pathways enriched by upregulated and downregulated genes are shown in a) and b), respectively, along with associated p-values. Bolded pathways are involved in the pathogenesis of other diseases.

Drug metabolism pathway is downregulated in both UC and CD (Table 4). Drug metabolism genes ADH1C, ADH6, CYP2B6, FMO5, GSTA1, GSTM4, MAOA, UGT1A7, UGT2A3 are all downregulated in IBD compared to healthy controls. A highly related pathway, the metabolism of xenebiotics by cytochrome p450 is downregulated in CD. The detox enzymes with diminished expression in this pathway include the drug metabolism genes ADH1C, ADH6, UGT1A7, UGT2A3 as well as genes CYP2B6 and CYP2S1 that code enzymes to metabolize anti-cancer drugs, gene EPHX1 for epoxide enzyme, and genes GSTA1 and GSTM4 coding enzymes involved in the detoxification of electrophilic compounds,

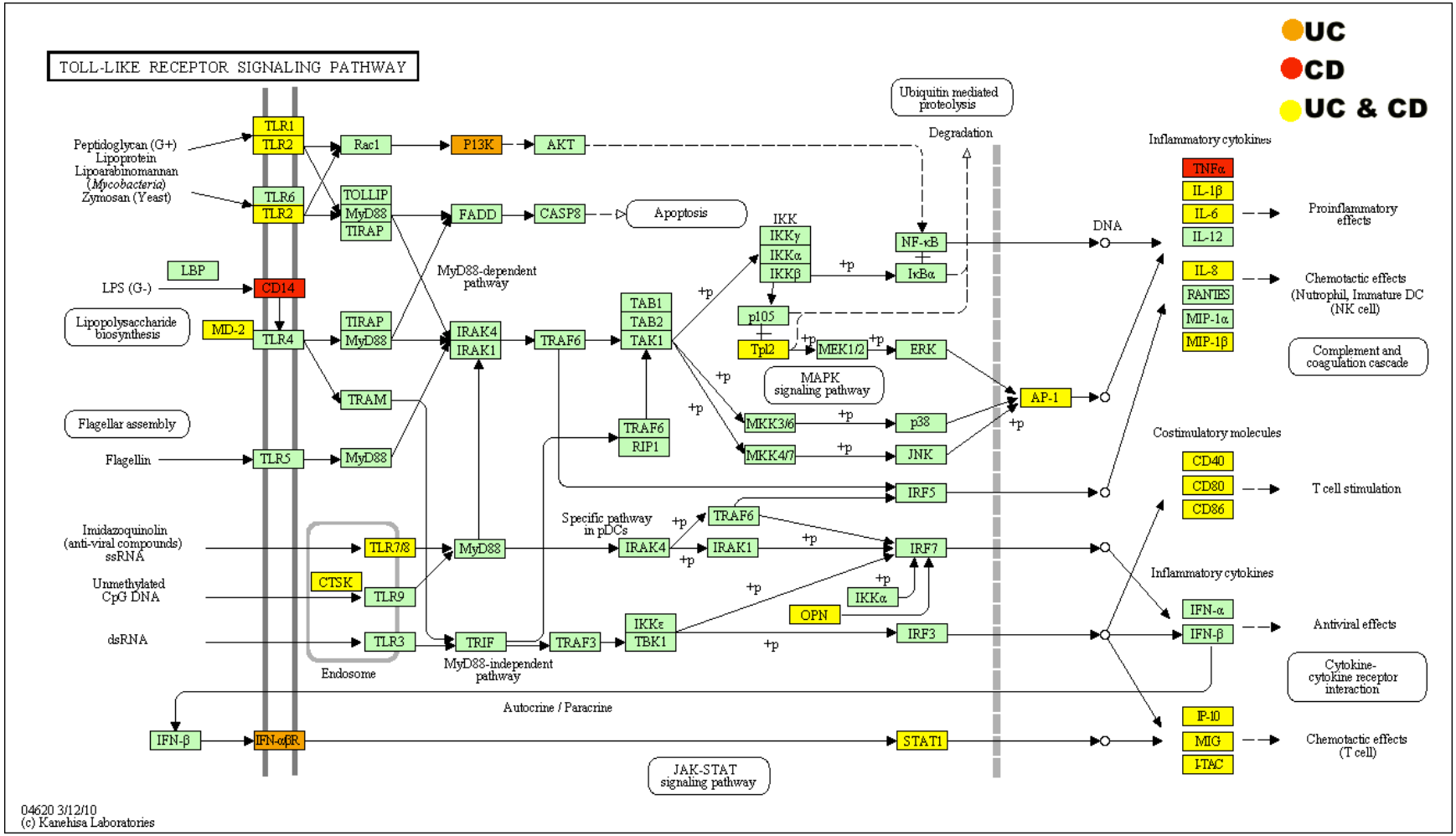
including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.

Table 4 shows that PPAR signaling as well as nitrogen metabolism are diminished in IBD. Genes downregulated in the nitrogen metabilism pathway consist of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide (CA1, CA2, CA4, CA7), cytoplasmic enzyme CTH, and the K-type mitochondrial glutaminase GLS. The PPAR signaling pathway clearly demonstrates the downregulation of the transcription factor PPARG within adipocytes leading to the downregulation of lipid metabolism through a variety of interactions (Figure 2c). The attenuated metabolism of lipids is also represented within the enriched GO biological processes (Figure 3).

2a)


NOD-LIKE RECEPTOR SIGNALING PATHWAY
UC
CD
UC & CD
Macrophages
Dendritic cells
Peripheral blood leukocytes
Intestinal epithelial cells
Paneth cells
Bacteria
Peptidoglycan
Peptidoglycan biosynthesis
iE-DAP
MDP
SGT1
NOD1
NOD2
Erbin
TRIP6
CARD6
RIP2
CARD9
cIAP1/2
A20
+u
-u
Apoptosis
CASP8
TRAF6
TAK1
TAB1
TAB2/3
IKKγ
IKKα
IKKβ
+u
+p
+p
+p
MAPK signaling pathway
ERK
JNK
p38
Ubiquitin-mediated proteolysis
IκB
NFκB
DNA
DNA
Anti-apoptotic factors
Anti-mictobial peptides
βdefensin
Antimicrobial barrier formation
Chemokines
IL-8
CXCL1
MIP2
CCL2
RANTES
Neutrophil recruitment
T cell differentiation
Antigen-specific T and B cell response
Proinflammatory cytokines
TNFα
IL-6
Inflammasomes
Anthrax lethal toxin
LT
NALP1
ASC
CASP1
CASP5
Bacteria
Bacterial RNA
Pore-forming toxin
LLO
Aerolysin
POP
Pyrin
PSTPIP1
NALP3
ASC
CASP1
Cardinal
CASP1
SGT1
HSP90
Antiviral compounds
R848, R837
Danger signals
Extracellular ATP
Uric acid CPPD Amyloid-β
UV-B
Xenogenous compounds
Silica
Asbestos
Aluminium
ROS
Cell death
Proinflammatory cytokine processing
IL-1β
IL-18
Fever
T helper cell polarization
DC maturation
Pyroptosis
COP
Bacteria
Flagellin
IPAF
ASC
CASP1
Flagellar assembly
Bacterial secretion system
IPAF
NAIP5
CASP1
Toll-like receptor signaling pathway
04621 3/8/11
(c) Kanehisa Laboratories

2b)


UC
CD
UC & CD
TOLL-LIKE RECEPTOR SIGNALING PATHWAY
Peptidoglycan (G+)
Lipoprotein
Lipoarabinomannan (Mycobacteria)
Zymosan (Yeast)
LPS (G-)
Lipopolysaccharide biosynthesis
Flagellar assembly
Flagellin
Imidazoquinolin (anti-viral compounds) ssRNA
Unmethylated CpG DNA
dsRNA
TLR1
TLR2
TLR6
TLR2
LBP
CD14
MD-2
TLR4
TLR5
TLR7/8
CTSK
TLR9
TLR3
Endosome
Rac1
P13K
AKT
TOLLIP
MyD88
TIRAP
FADD
CASP8
Apoptosis
TIRAP
MyD88
TRAM
MyD88
MyD88-dependent pathway
IRAK4
IRAK1
TRAF6
MyD88
TRIF
TRAF3
MyD88-independent pathway
TAB1
TAB2
TAK1
TRAF6
RIP1
Specific pathway in pDCs
TRAF6
IRAK4
IRAK1
IKK
IKKγ
IKKα
IKKβ
p105
Tpl2
MEK1/2
ERK
MAPK signaling pathway
MKK3/6
p38
MKK4/7
JNK
IKKε
TBK1
OPN
IKKα
IRF5
IRF7
IRF3
NF-κB
IκBα
Ubiquitin mediated proteolysis
Degradation
DNA
AP-1
Autocrine / Paracrine
IFN-β
IFN-αβR
STAT1
JAK-STAT signaling pathway
Inflammatory cytokines
TNFα
IL-1β
IL-6
IL-12
Proinflammatory effects
IL-8
RANTES
MIP-1α
MIP-1β
Chemotactic effects (Nutrophil, Immature DC (NK cell)
Complement and coagulation cascade
Costimulatory molecules
CD40
CD80
CD86
T cell stimulation
Inflammatory cytokines
IFN-α
IFN-β
Antiviral effects
Cytokine-cytokine receptor interaction
IP-10
MIG
I-TAC
Chemotactic effects (T cell)
04620 3/12/10
(c) Kanehisa Laboratories

2c)

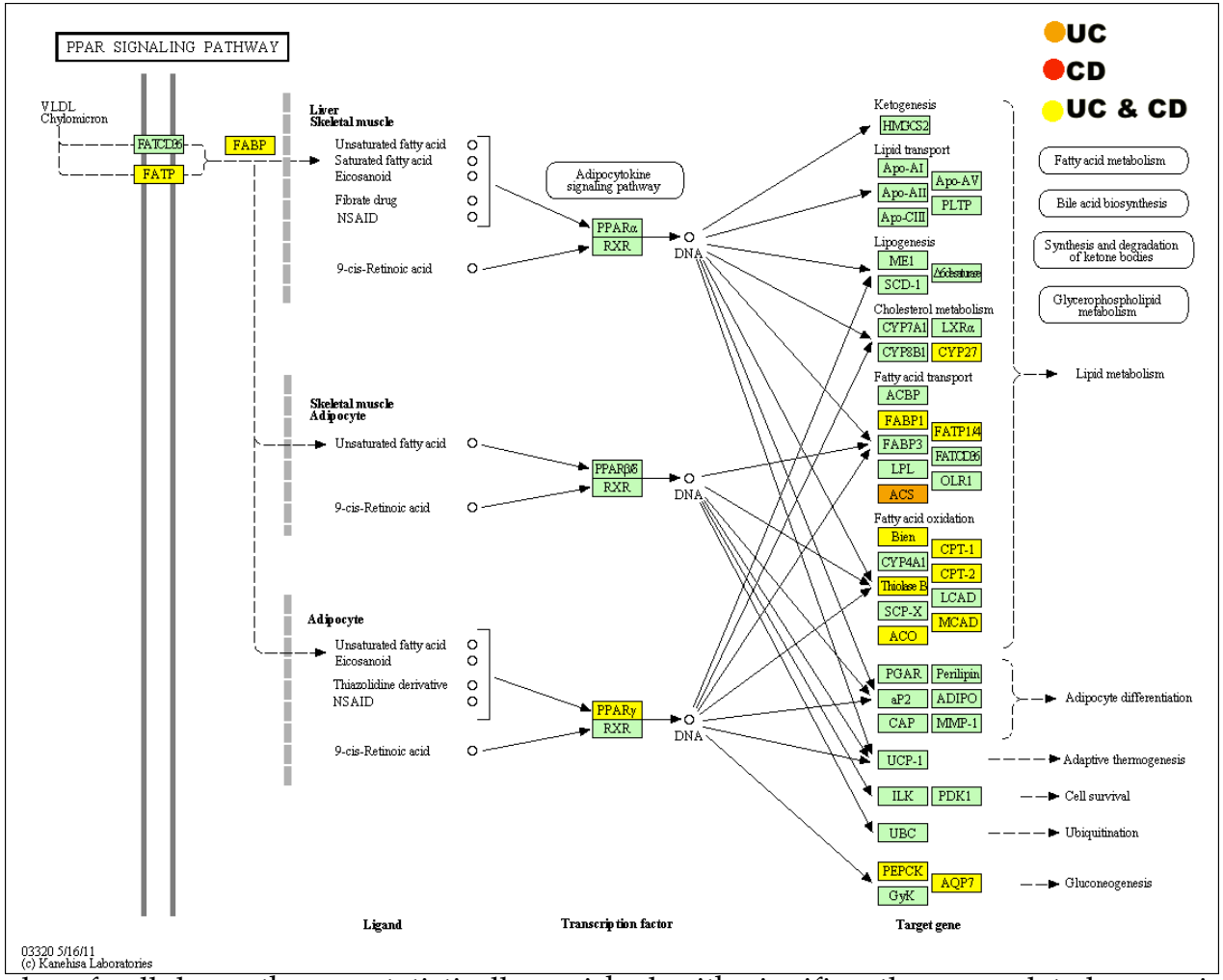


Figure 2. Examples of cellular pathways statistically enriched with significantly upregulated genes in IBD. Figure 2a shows Nod-like receptor signaling pathway, 2b Toll-like receptor signaling pathway, and 2c the PPAR signaling pathway. The nodes shown in brown, red, and yellow indicate significantly upregulated genes in UC/HC, CD/HC and (UC U CD)/HC comparisons, respectively.

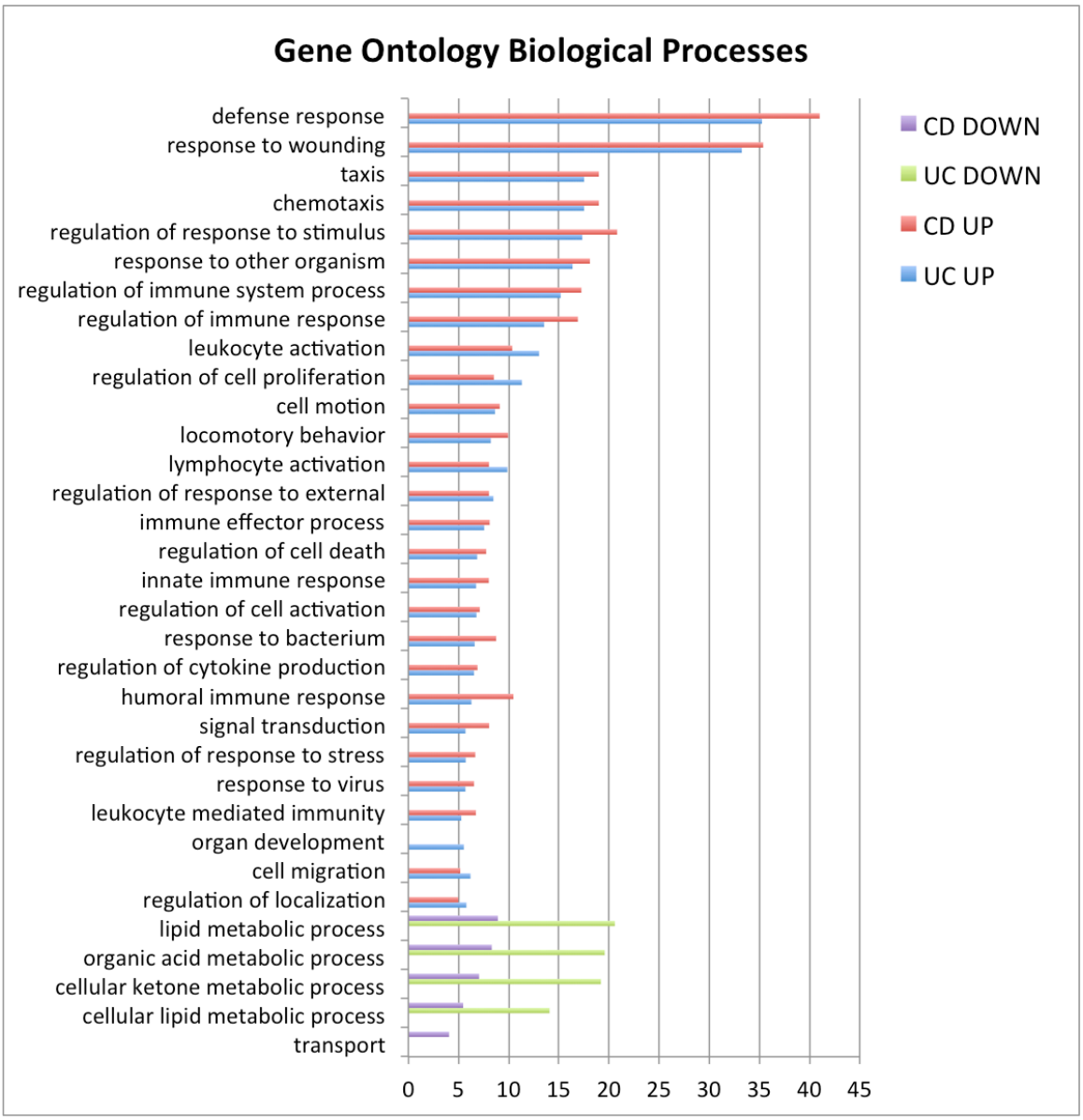


Figure 3. Gene Ontology biological processes (level 3) statistically enriched by significantly upregulated and downregulated genes for UC and CD.

Autoimmune Diseases Related To IBD

A number of KEGG pathways related to other autoimmune diseases are identified in Table 4 as being statistically enriched with IBD genes including type 1 diabetes mellitus, systemic lupus erythematosus, autoimmune thyroid disease, and primary immunodeficiency. For this reason, we investigated the involvement of identified significant genes related to IBD and other autoimmune diseases using GAD. We found 280 differentially expressed genes in IBD in our gene list known to be associated with 35 different autoimmune and IBD related MeSH disease terms through 918 unique associations in (Figure 4). The three largest disease nodes in Figure 4 (nodes with the highest connectivity to IBD genes) are type 1 diabetes mellitus, rheumatoid arthritis, and systemic lupus erythematosus. The Matlab code used for GAD enrichment is presented in Appendix B.

IBD genes found to be associated with autoimmune diseases include the upregulated leukocyte antigen genes HLA-DMB, HLA-DOB, HLA-DPB1, HLA-DQB1, HLA-DRB4, and HLA-F. In addition to the aforementioned genes common with IBD, lupus and IBD share central pattern recognition proteins such as C1QA, C1C1QB, C1QC, C1R, and C1S. Recent research implicated hereditary C1Q deficiency as associated with systemic lupus erythematosus and increased susceptibility to bacterial infections [82]. Other genes common to lupus include the ADA gene coding an enzyme deficiency of which causes immunodeficiency disease, and the genes FCGR2B, FCGR3B code for proteins regulating the antibody production by B cells

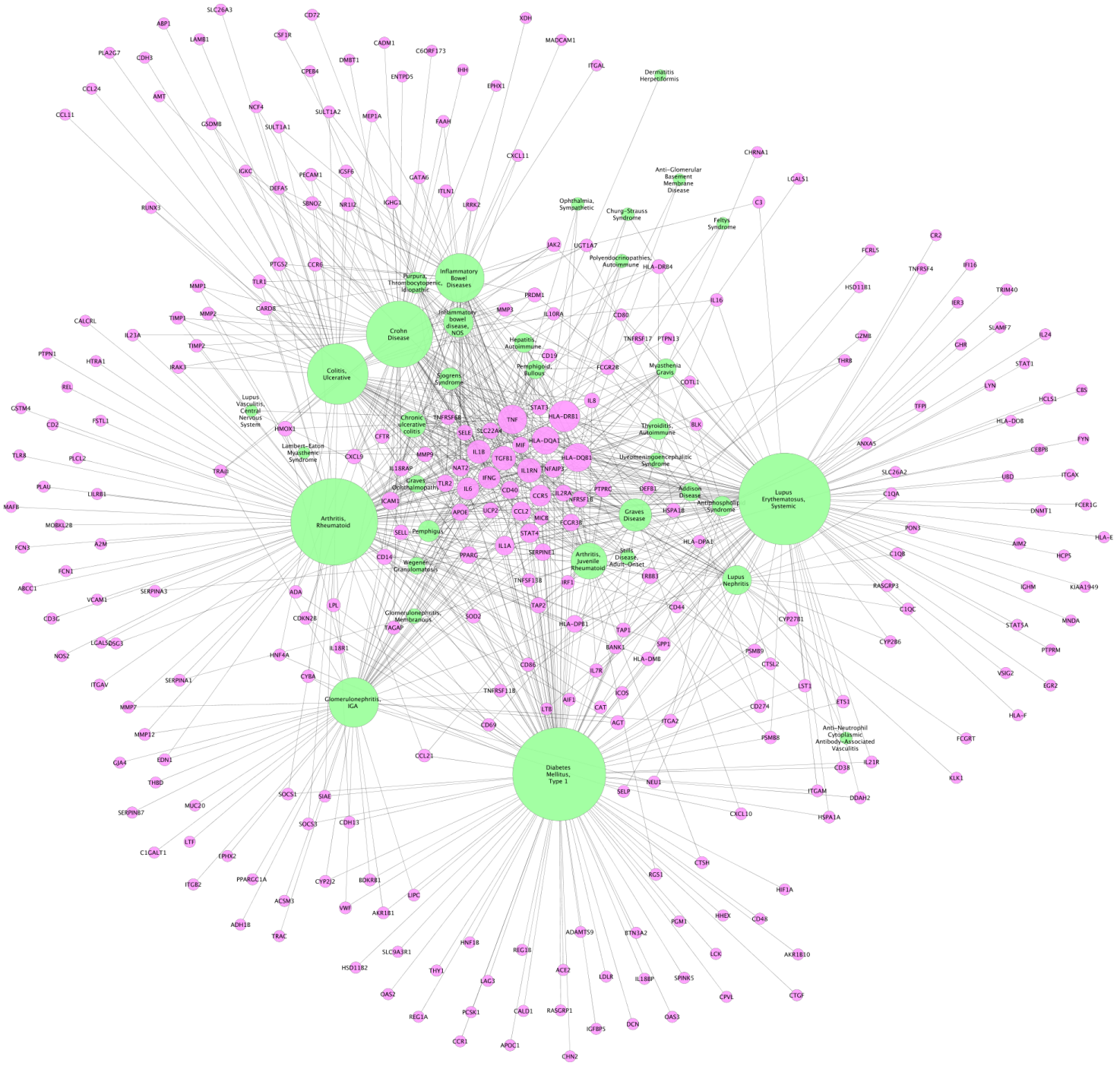


Figure 4. Interactome map of autoimmune diseases. The map was obtained by projecting IBD significant gene list onto gene lists of autoimmune diseases (obtained from the Genetic Association Database). Diseases are colored in green and genes are colored in pink. The size of each node corresponds to the number of interactions with other nodes, with larger nodes having more interactions.

Identification Of Known Drug Targets With Altered Gene Expression In IBD

The significant gene lists were projected onto genes with UNIPROT accessions in ChEMBL, the database for bioactive drug-like small molecules [58]. Those genes with a minimum fold change of 2.0 are presented in Figure 5. The top 200 significantly upregulated and downregualted genes for both UC and CD are presented in appendices E-D. The figure shows gene targets along with their respective fold change in UC and CD relative to HC. The majority of the small-molecule targeted genes encode cell surface receptors, oxidoreductase, G-protein coupled receptors, and kinases. Among the drug targets with the largest fold change expression are MMP3, MMP1, IL-8, PLAU, PTGS2, and CXCR4. Drugs targeting MMP3 include nonspecific MMP inhibitor with poor performance in clinical trials [83]. Doxycycline, a tetracyclin antibiotic, is known to inhibit MMP activity, and is used clinically for the treatment of periodontal disease. IL-8 is an antibody therapeutic target in inflammatory diseases [84] but with mixed results stemming from side effects. The small molecule drugs with activity on IL-8 include Diclofenac, Ibuprofen, Indomethacin, and Tolmetin. The drug target PTGS2 is known to play a role in promoting colon cancer [85]. Nonsteroidal anti-inflammatory drugs targeting PTGS2 appear to inhibit the proliferation of cultured hepatocellular cancer cells [86]. CXCR4 is a receptor playing a role in immune system signaling between cells and is involved in the growth of tumors. It is a drug target in metastatic lung cancer [87]. These observations indicate the challenges of repositioning existing drugs targeting IBD genes with highly elevated expression.

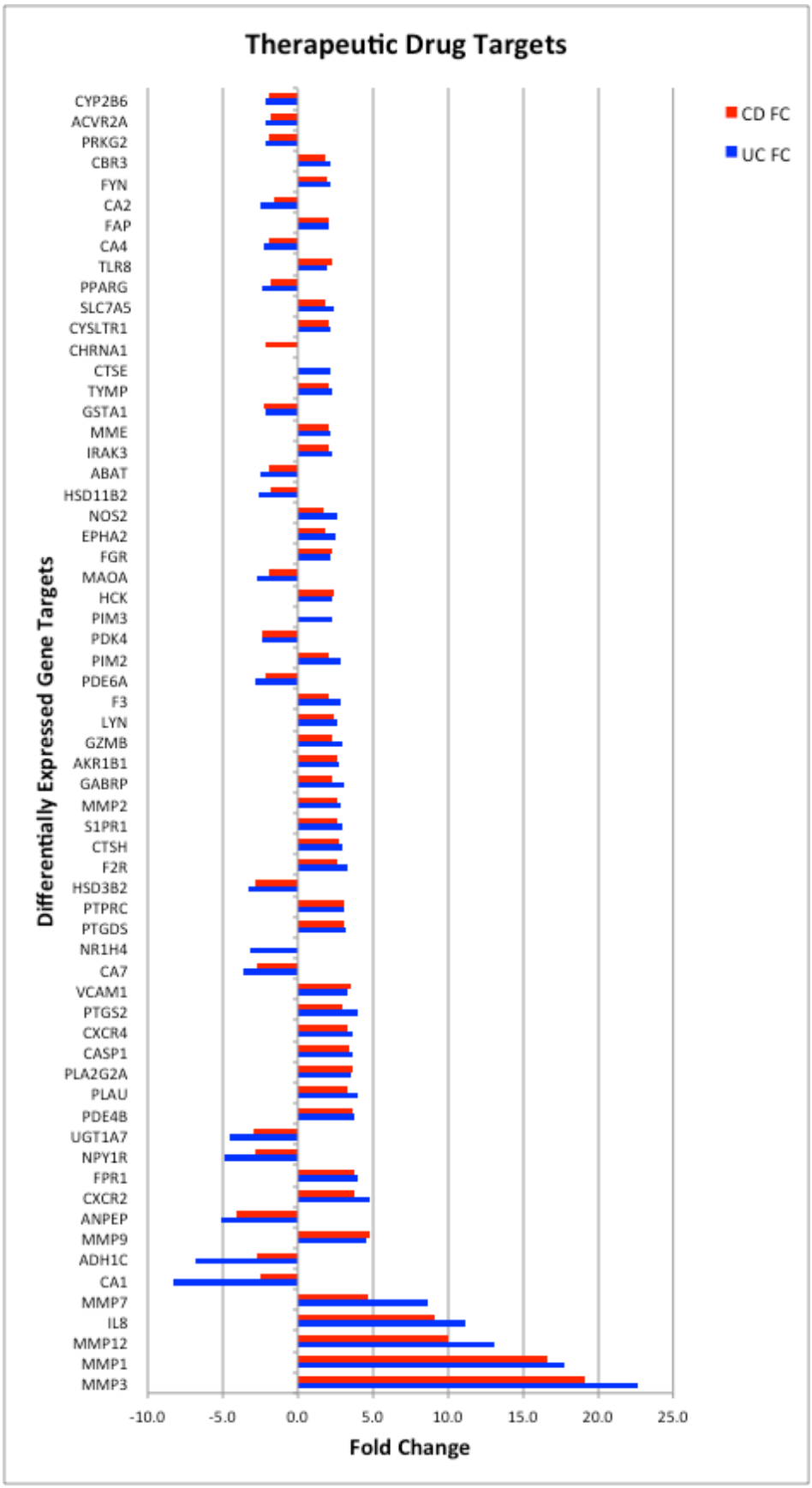

Therapeutic Drug Targets
CD FC
UC FC
Differentially Expressed Gene Targets
CYP2B6
ACVR2A
PRKG2
CBR3
FYN
CA2
FAP
CA4
TLR8
PPARG
SLC7A5
CYSLTR1
CHRNA1
CTSE
TYMP
GSTA1
MME
IRAK3
ABAT
HSD11B2
NOS2
EPHA2
FGR
MAOA
HCK
PIM3
PDK4
PIM2
PDE6A
F3
LYN
GZMB
AKR1B1
GABRP
MMP2
S1PR1
CTSH
F2R
HSD3B2
PTPRC
PTGDS
NR1H4
CA7
VCAM1
PTGS2
CXCR4
CASP1
PLA2G2A
PLAU
PDE4B
UGT1A7
NPY1R
FPR1
CXCR2
ANPEP
MMP9
ADH1C
CA1
MMP7
IL8
MMP12
MMP1
MMP3
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0
25.0
Fold Change

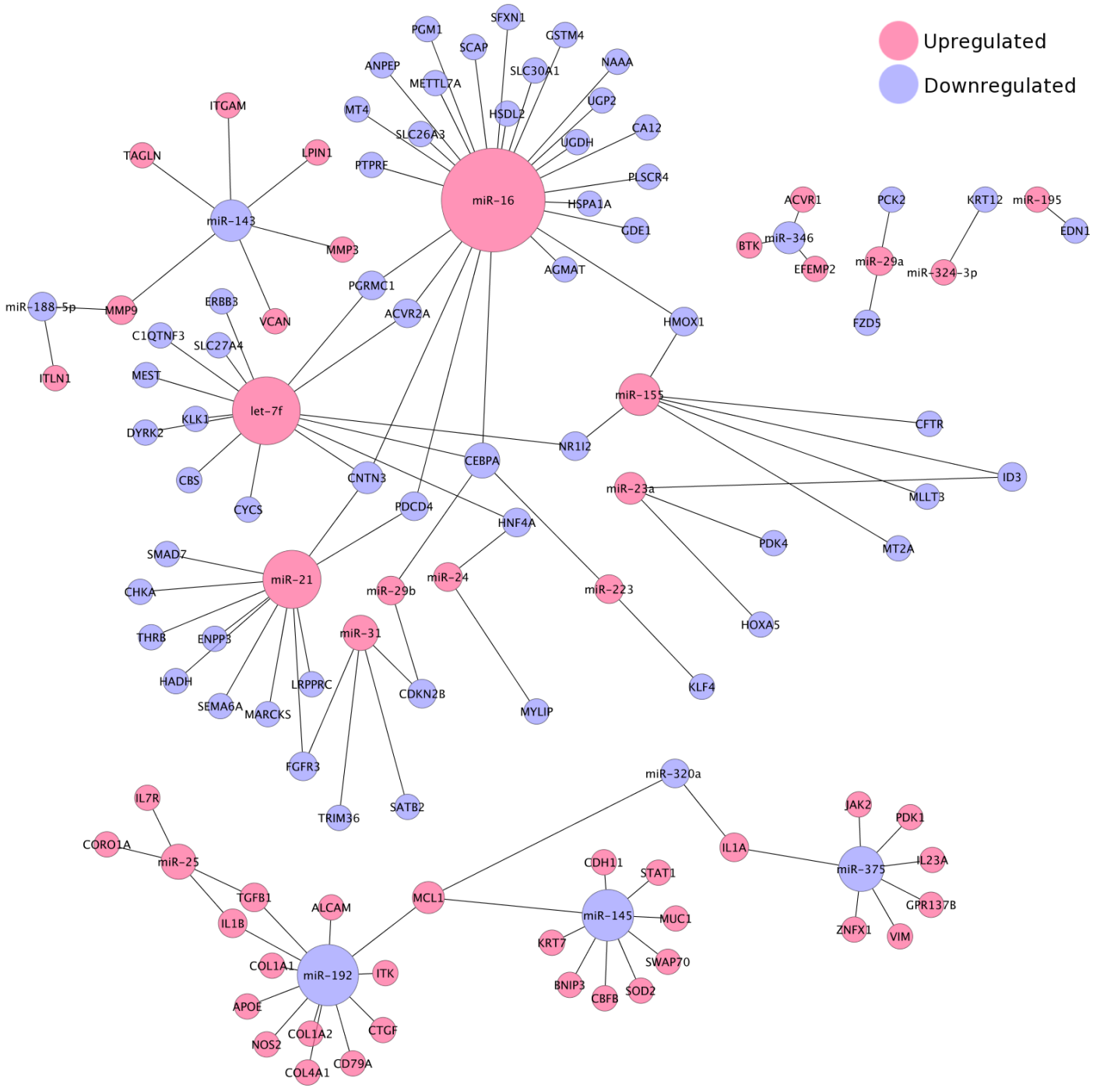
Figure 5. IBD genes targeted by bioactive small molecules. The number of therapeutics targeting each gene is presented in parentheses next to the gene symbol on the y-axis. The fold change of each differentially expressed gene is also shown for both UC and CD.

Drugs with activity on CXCR proteins include Ibuprofen, Naproxen, Indoprofen, and Ketoprofen. Drug targets with negative fold changes include carbonic anhydrases CA1 and CA2, and PPARγ. Existing drugs such as Acetazolamide (Diamox), Methazolamide (Neptazane) lead to the inhibition of carbonic anhydrase and as such appear unsuitable for the treatment of IBD. PPARγ is another drug target with lower expression in IBD. This gene encodes a member of the peroxisome proliferator-activated receptor subfamily of nuclear receptors regulating transcription of genes implicated in the pathology of numerous diseases including obesity, diabetes, atherosclerosis and cancer. Small drugs with activity against PPARγ including Fenofibrate, Pioglitazone, and others have found use in treating Type II diabetes. Drug targeted genes VCAM1, ICAM1, ITGAL, TNF, and IL8 map into the acute inflammatory pathway whereas the leukocyte migration pathway contains drug targets VCAM1, ICAM1, ITK, ITGAL, CXCR4, MMP9, and MMP2. Combination therapies designed for IBD with existing drugs will be limited by the poor drug metabolism associated with this disease.

Mirna - Gene Interactome

A total of 26 and 33 miRNA are known to be differentially expressed within active UC and CD respectively as compared to healthy controls [59-63]. Of these miRNA, 20 were shown to interact with 90 significant genes for UC (Figure 6a) and 19 miRNA were shown to interact with 44 significant genes for CD (Figure 6b). Gene interactions with miRNA in which both were either upregulated or downregulated were not considered for our analysis since these interactions defy the understood paradigm of miRNA-mRNA interactions [88]. We identified 9 miRNA that are coexpressed within UC and CD including miR-126, miR-126*, miR-127-3p, miR-155, miR-21, miR-29b, miR-31, miR-324-3p and miR-375. Each miRNA is upregulated in both phenotypes except miR-375, which is reported to be downregulated in UC [60] and upregulated in CD [62]. There were found to be 17 and 8 genes within the network that are targets for bioactive small molecules for UC and CD respectively. The figure indicates a widespread downregulation of significant genes due to the upregulation of several miRNA.

6a)


Upregulated
Downregulated
miR-16
let-7f
miR-21
miR-143
miR-155
miR-23a
miR-29b
miR-24
miR-223
miR-31
miR-192
miR-145
miR-375
miR-25
miR-320a
miR-346
miR-29a
miR-324-3p
miR-195
miR-188-5p
SFXN1
PGM1
SCAP
GSTM4
ANPEP
METTL7A
SLC30A1
NAAA
MT4
HSDL2
UGP2
SLC26A3
CA12
UGDH
PTPRE
PLSCR4
HSPA1A
GDE1
AGMAT
ITGAM
TAGLN
LPIN1
MMP3
PGRMC1
ACVR2A
HMOX1
ACVR1
PCK2
KRT12
EDN1
BTK
EFEMP2
FZD5
ERBB3
VCAN
C1QTNF3
SLC27A4
MEST
KLK1
DYRK2
CBS
CYCS
CNTN3
CEBPA
NR1I2
CFTR
ID3
MLLT3
PDK4
MT2A
HOXA5
PDCD4
HNF4A
SMAD7
CHKA
THRB
ENPP3
HADH
SEMA6A
MARCKS
LRPPRC
CDKN2B
MYLIP
KLF4
FGFR3
TRIM36
SATB2
IL7R
CORO1A
TGFB1
IL1B
ALCAM
COL1A1
ITK
APOE
NOS2
COL1A2
CTGF
COL4A1
CD79A
MCL1
IL1A
CDH11
STAT1
MUC1
KRT7
SWAP70
BNIP3
CBFB
SOD2
JAK2
PDK1
IL23A
GPR137B
ZNFX1
VIM

6b)

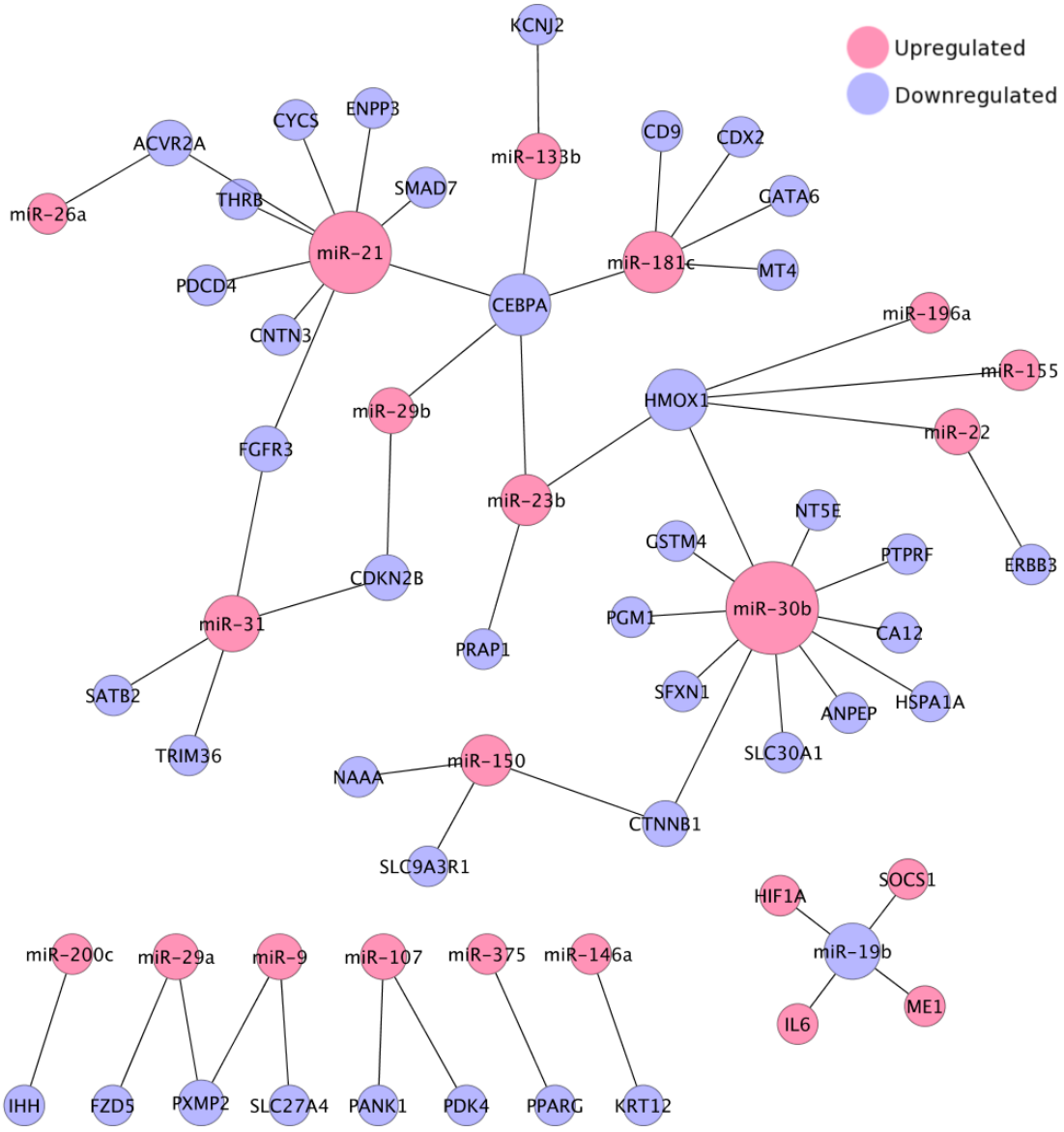


Figure 6. Micro-RNA – Transcript Interactome Map. The figure illustrates the possible regulation patterns of a subset of IBD significant genes by miRNA, significantly increased or decreased in UC (7a) and CD (7b). The colors pink and blue indicate upregulated and downregulated nodes, respectively. The links shown in the figure indicates gene/miRNA pairings reported in the literature (not necessarily for IBD). The miRNA shown in the figure correspond to those significantly upregulated or downregulated in IBD.

2.5 DISCUSSION

Inflammatory bowel disease is a complex disorder involving commensal bacteria, multiple human cell types and a variety of cellular networks including those related to adaptive and innate immunity. Our integrated microarray analysis of inflamed human colonic mucosa captures the widespread genetic perturbations of IBD phenotypes as associated with a variety of cell types, pathways and biological processes; a system level perspective of transcriptional regulation attributable to the heterogeneous nature of colonic biopsy samples. By combining raw microarray data from six independent studies, we were able to increase the statistical power of our analysis while minimizing lab specific bias and experimental noise. A total of 118 unique patient samples were used including 48 UC, 41 CD and 29 healthy controls. Differentially expressed miRNA obtained through literature curation were also incorporated into our analysis in order to investigate possible mechanisms of post transcriptional regulation through validated interactions with differentially expressed genes related to IBD [64]. Bioactive therapeutic small molecules targeting differentially expressed genes as well as gene involvement in other autoimmune diseases were also investigated to facilitate the identification of druggable pathways and therapeutic targets that may be candidates for a wide variety of related illnesses or existing therapeutics that may be repositioned for the treatment of IBD.

Our analysis of microarray data showed as much as forty-fold increase in transcript number for some of the genes in IBD relative to healthy controls. The most upregulated gene in IBD is REG1A, coding a

protein typically secreted by the exocrine pancreas and associated with islet cell regeneration. A survey of the literature shows upregulated REG1A as associated with pancreatic cancer, celiac disease, and type I diabetes [65-67]. The top upregulated genes for IBD in our list were abundant with CXC chemokines involved in inflammation, and were previously linked to such diseases as lupus, prostate cancer, and rheumatoid arthritis [69-71]. The matrix metalloproteinase (MMP) family proteins are also abundant among the top forty most highly upregulated genes in IBD. MMP proteins were previously shown to be involved in arthritis and metastasis [89]. The third most highly upregulated gene in our list, S100A8, codes a soluble IBD biomarker protein calprotectin [72-75]. High levels of expression of this protein is also linked to cystic fibrosis [76]. Commonly upregulated pathways in IBD included the previously recognized Toll-like and NOD-like receptor signaling pathways, cytokine-cytokine interaction, and chemokine signaling. Cellular processes activated by IBD include immune defense response, cellular response to wounding, leukocyte activation, and regulations of immune response and cell proliferation as well as response to molecules of bacterial origin. Some of the pathways activated in IBD appear to be druggable such as the calcium signaling pathway, endocytosis, cytokine-cytokine receptor interaction and leukocyte transendothelial migration pathways. In our gene signatures, the genes UGT1A, HSD11B2, HSD17B2, HSD3B2, LCMT1 are all significantly downregulated and are known to play a role in androgen and estrogen metabolism. The specific role sex hormones play in the pathology of IBD is yet to be explored extensively in the literature

[90] [91]. All these pathways contain at least 8 drug targets each, which are also significant genes in our healthy control to IBD comparison.

KEGG disease pathways enriched by IBD significantly upregulated genes included those with an autoimmune component, namely autoimmune thyroid disease, asthma, primary immunodeficiency, systemic lupus, viral myocardidatis, type I diabetes mellitus, autograft rejection, and the graft-vs-host disease. Patients with IBD were previously associated with a higher prevalence of rheumatoid arthritis, lupus and hypothyroidism, with increased prevalence of asthma, eczema, allergic rhinitis and diabetes [92]. Moreover, literature also contains links with already recognized IBD associated diseases (cancer and malaria) and the diseases statistically enriched by our IBD genes cited above. A polymorphism in FCGR2B, an IBD significant gene, is associated with protection against malaria but susceptibility to systemic lupus erythematosus [93]. Histone deacetylase inhibitors are currently being harnessed as a potential treatment for malaria, systemic lupus erythematosus and a wide range of neurodegenerative conditions, and asthma [94]. In addition, the beneficial effects of malaria drugs have been recognized in the management of systemic lupus erythematosus and rheumatoid arthritis [95]. Add to these other examples listed in the results section, it appears that many of the top upregulated genes in IBD are also upregulated in autoimmune disorders. For this reason, we investigated the upregulated genes shared by IBD and other autoimmune diseases as possible drug targets by intersecting our gene list with the genes associated with specific diseases in the Genetic Association Database [27].

The genes known as drug targets and upregulated in multiple autoimmune diseases provide clues for the multidimensional nature of IBD. For example, the ATP binding transporter gene ABCC1 – significantly upregulated in a number of diseases including, IBD, and rheumatoid arthritis, and is known to bind to HIV Tat protein – is targeted by drugs/supplements such as Abacavir, Tenovir, Nevirapin, and Quercetin, all with antiviral effects, as well as by the anticancer supplement Apigenin and the transplant rejection medicine Cyclosporine. Other significantly upregulated genes shared by rheumatoid arthritis and IBD include TLR8 (a target of the immune response modifier Amiquimod); MMP3 (a target of failed drug Marimastat); ICAM-1 (targeted indirectly by cholesterol lowing Lovastatin); IL8 (targeted by Tolmetin and Ibuprofen); and IL2RA (targeted by Ascomysin and Budesonide, used for treating Crohn's disease and asthma). The supplemental file associated with the present study presents a collection of drug target – drug – disease relationships relevant to IBD and other autoimmune diseases. The intersection of IBD significant genes with known drug targets provides valuable information for possible repositioning of existing drugs in the treatment and management of IBD.

The data on significantly altered miRNA expression in IBD, when integrated with literature and our significant gene lists, indicate a strong role for miRNA in the down regulation of genes in IBD. Increased expression of miR-16 has been shown in the literature to downregulate genes CNTN3, C1QTNF3, ACVR2A, and CYCS, while Let-7f-1 has been shown to downregulate CNTN3, ANPEP, NAAA, PDCD, UGDH, HSDL2,

and ACVR2A. These two microRNAs are highly expressed in UC and the genes shown are highly downregulated. Similarly, increased expression of mir-23b is linked to downregulation of PRAP1 in both CD and UC and increased miR-21 is associated with downregulated CTN3, PDK4, and PDCD genes in CD. The aforementioned proteins play roles in a multitude of events ranging from mitochandrian electron transport (CYCS), anti-auto immune processes (PDCD), regulation of glucose metabolism (PDK4), small intestinal microvillar events (ANPEP), neurite outgrowth-promoting activity (CNTN3), signaling through serine kinases (ACVR2A), and induction of glycosylation (UGDH). It is shown in figure 6 that CEBP is the most connected gene node for both UC and CD miRNA innervation. This gene is shown to be downregulated in the presence of upregulated miRNA. CEBP is known to be upregulated in the presence of platelet activation factors (PAF) and endotoxins such as bacterial lipopolysaccharides (LPS) leading to an inflammatory response though the transcriptional regulation of pro-inflammatory cytokines and proteins [96]. The downregulation of this protein in our analysis indicates

These results suggest the restorative potential of a micro RNA based therapies on IBD specific diminished biological processes.The results of our study can be put in context with mouse studies, emphasizing the effect of B lymphocytes on immunity and metabolism in the gut. A recent study in mice [97] showed that the lack of B cell immunity was associated with upregulation of genes involved in other aspects of immune defense, inflammatory, and interferon-inducible responses. The gene list provided in Supplemental Table 1 of Shulzhenko

et al. [97] contains 14 significant genes from our CD/HC and UC/HC comparisons. Among these genes, DUOX2 and DUOXA2, involved in synthesis of hydrogen peroxide, are highly upregulated in CD. The genes significantly upregulated in both UC and CD with fold changes greater than 2.5 include CFI, which codes complement factor I. This protein proteins is involved in the destruction of foreign invaders (such as bacteria and viruses), trigger inflammation, and remove debris from cells and tissues. Also upregulated in protein deficient mouse and human IBD are interferon-induced proteins IFIT3, which potentiates antiviral signaling, and IFITM1, with antiproliferative effects; and chemokine CXCL9, thought to be involved in T cell trafficking. The genes downregulated in B cell deficient mice and in CD and UC consist of CYP27A1 in the in synthesis of cholesterol, steroids, and other lipids. and EDN3, essential for generation of enteric neurons and was previously linked to Hirschsprung disease and Waardenberg syndrome. NR1H4 is a gene significantly downregulated both in B cell defficient mouse and in UC. It encodes a ligand-activated transcription factor, involved in bile acid synthesis and transport. These observations suggest the role of ineffective B-cell mediated immunity in IBD, possibly causing activation of innate immunity and a reduction in ability to process lipids.

In conclusion, this study presents a portrayal of the molecular rewiring of the human gut mucosa altered by IBD. Our statistical analysis reveals cellular processes altered in IBD, with focus on biomarker selection, and possible repositioning of currently available drugs for the treatment of IBD. Our results show gene signature

commonalities between IBD, and asthma, lupus, and rheumatoid arthritis, among others. Results raise the question of compromised B cell immunity in IBD patients along with over compensation of other components of the immune system, possibly facilitating research on new treatment protocols for this chronic disorder.

CHAPTER 3

Meta-Analysis of IBD and Colorectal Carcinoma Microarray Data Reveals Transcriptional Regulation, Disease Biomarkers and Regulatory Pathway Associations Between Disease Phenotypes

3.1 ABSTRACT

Patients with IBD are known to have an increased risk of developing colorectal carcinoma (CRC), with the likelihood increasing with the duration of IBD. This phenotype shift is however poorly understood as are the genetic factors leading to the development of CRC. In order to better understand this transition, microarray meta-analysis was performed to characterize the gene expression profiles of IBD and CRC patients in a functional context. Disease classifiers were also determined in order to accurately stratify colonic disease phenotypes as well as the various stages of CRC according to the Dukes staging system. Our analysis represents a comprehensive analysis of the colonic transcriptome under various disease conditions and provides for the interpretation of significant genes in the context of physiological pathways and discrete clinical biomarkers.

3.2 INTRODUCTION

Colorectal cancer (CRC) is estimated to affect over 1.2 million people worldwide with over 600,000 attributed deaths annually making it the third most commonly diagnosed cancer in men and the second in women, while being having the fourth and third highest contributor to cancer related mortality among men and women respectively [98]. There are known to be a number of risk factors associated with the onset of CRC including hereditary syndromes of familial adenomatous polyposis, hereditary nonpolyposis colorectal cancer and inflammatory bowel disease (IBD) [99]. The incidence of IBD related CRC varies widely within the literature, although a meta-analysis of 116 studies involving 55,000 UC patients revealed a relative risk of 2%, 8% and 18% in developing CRC after 10, 20 and 30 years respectively [100]. Similarly, a meta-analysis of over 60,000 CD patients revealed a relative risk of 2.9% for CRC and 33.2% for small bowel carcinoma [101].

Previous microarray analysis studies of colonic gene expression in colorectal cancer patients have led to many important findings. Survival outcomes of CRC patients were predicted using a 34-gene classifier set and was found to significantly discriminate patients with recurrence of metastatic tumors from remissive patients [102]. Similarly, a 128 matastasis–associated gene expression profile was shown to accurately predict patient outcomes and may be used to optimize adjuvant treatment strategies for subsets of CRC patients [103]. The response of CRC patients to COX2 inhibitors was also investigated using gene expression measurements and was found to be associated with a shift in gene

expression patterns between normal and ademona cell types, suggesting the efficacy in COX2 inhibitor treatment during early cancer stages [104]. Research performed by the same group previously identified discriminative gene signatures for the differential diagnosis of colorectal disease types and their correlation with gene expression within peripheral blood samples [105].

Diagnosing IBD is also problematic and often requires a colonoscopy or histopathological examination to be performed by a gastroenterologist. Although this practice may allow for a definitive diagnosis, such procedures are costly, time consuming, and invasive. Biomarkers have alternatively been studied as a means to diagnose IBD, monitor disease activity and responsiveness to various treatment modalities. Such diagnostic procedures for IBD may be based on elevated intestinal inflammation, since IBD patients excrete an increased quantity of leukocytes within their stool. The leukocytes in the gastrointestinal tract can be quantified via radioactive labeling in conjunction with an abdominal scintigraphy to identify which regions of the intestine are inflamed. This costly technique, however, requires specialized facilities, and subjects the patient to radiation exposure [106, 107]. An alternative to measuring leukocytes directly in vivo is to measure leukocyte derived proteins. This approach has shown promise in identifying indicators of IBD including fecal lactoferrin [108, 109], calprotectin, lysozyme, polymorphonuclear neutrophil elastase, lipocalin and myeloperoxidase [73, 74, 110]. Among the fecal biomarkers, lactoferrin and calprotectin have been shown to be the most sensitive and specific in discriminating

IBD patients from a population of healthy patients [111], while ASCA and pANCA have shown mixed results in discriminating UC patients from those with CD [112, 113]. Consequently, there is a clear need to identify a set of biomarkers to differentiate patients with IBD from healthy patients as well as identify drug targets for new treatment protocols for IBD.

While previous microarray studies have focused solely on CRC our integrated microarray analysis approach aims elucidate the genetic factors responsible for the transition from IBD to CRC including physiological pathways and unique disease biomarkers that may be used to clinically differentiate between disease phenotypes.

3.3 METHODS

The microarray datasets used in this study were all in public domain and were generated from mRNA isolated from the affected colonic mucosa of UC, CD and CRC patients. Samples were all hybridized onto the Affymetrix Human Genome U133 Plus 2.0 GeneChip® Array platform for analysis. A total of 698 colonic mucosal microarray dataset samples were obtained from the NCBI Gene Expression Omnibus (GEO) [34, 35] in raw CEL format. Following the removal of duplicate samples, sample outliers were removed using principal component analysis (PCA) [36]. There were a total of 15 IBD and 41 CRC outliers detected and removed from further analysis. The final sample population is comprised of 29 healthy control (HC) samples, 48 UC samples, 41 CD and 309 CRC samples from nine independent studies (GSE9452 [37], GSE9686 [38], GSE10191 [39], GSE10616 [40], GSE13367 [41], GSE16879 [42], GSE17536 [102], GSE17537 [102] and GSE14333 [103]). Gene expression datasets for IBD (UC & CD) as well as CRC were independently normalized using robust multi-array averaging (RMA) background adjustment, quantile normalization with median polishing and log2 transformed summarization procedures within MATLAB [43-45] and were mapped to Entrez gene IDs using a custom CDF file during the normalization process [46, 47]. Following pre-processing, batch effect removal was performed using the COMBAT source code in R [48].

Significant Genes

Two distinct microarray data analysis methods – Significance Analysis of Microarrays (SAM) [49] and the rank product (RP) method [50, 51] representing meta analysis – were performed in MATLAB and R (2.10.0), respectively, in order to identify statistically significant upregulated and downregulated genes in UC/HC, CD/HC, Dukes A/HC, Dukes B/HC, Dukes C/HC, Dukes D/HC comparisons. For IBD comparisons, statistically significant genes were determined using a fold change (FC) and p-value cutoff of 1.5 and 0.001 respectively over 1000 permutations [48], while a FC and pvalue cutoff of 2 and 0.001 were used to identify significant genes for CRC/HC comparisons. For the RP method, only those genes with an enrichment p-value less than 10^{-6} were identified as significant in both cases.

Functional Enrichment

Significantly enriched KEGG [114] and BioCarta pathways were determined for UC, CD, Dukes A, Dukes B, Dukes C and Dukes D phenotypes using DAVID [52, 53]. A pvalue cutoff of 0.05 was used to identify significant pathways enriched by significant gene lists for each phenotype. Genes implicated in pathways of interest were colored using KEGG color [57] and MATLAB.

Disease Classifiers

Disease classifier sets were identified by first filtering the normalized data matrix to only contain upregulated genes common to both phenotypes under comparison since downregulated transcripts and related proteins may not exist in sufficient quantity for detection. A second filter was used to identify only those proteins that are expressed within the extracellular region, extracellular space or plasma membrane according to GO cell compartment category. This filter ensures that the identified biomarkers will be present within the stool for use as a fecal biomarker. The data for each sample comparison is then randomly split evenly into training and test classes. Classifiers were then identified on the training set using the ROC class separability metric within Matlab. The resulting classifier sets were used to classify the test samples using a logistic regression model and subsequently calculate the area under the ROC curve (AUC) [115, 116]. This process was performed over 1000 iterations in order to identify the best ten classifiers for each comparison as weighted by their AUC for each comparison. Each optimal classifier set was then used to classify the test samples independently to assess their performance as reported by their AUC.

In Silico Disease Classifier Validation

Classifier performance was evaluated on an independent dataset obtained form GEO, which contained IBD, CRC and healthy control biopsies obtained from the colon of affected individuals [104]. RNA isolated from the samples were hybridized onto the Affymetrix HGU133

plus2 microarray platform for analysis. The dataset did not differentiate between the various cancer stages or IBD phenotypes related to each patient, allowing for the evaluation of only HC/IBD, HC/CRC and IBD/CRC biomarker sets. The dataset contains 8 HC samples, 15 CRC samples and 15 IBD samples.

Data was first normalized using robust multi-array averaging (RMA) background adjustment, quantile normalization with median polishing and log2 transformed summarization procedures within MATLAB [43-45] and were mapped to Entrez gene IDs using a custom CDF file during the normalization process [46, 47]. The top 5 disease classifiers were then used to classify the samples using logistical regression. ROC plots were generated and AUC was calculated to assess classifier performance. The Matlab script used to classify the samples within the validation set using logistic regression and generate ROC curves is provided within appendix C.

Disease-Gene Interactome

Genes with known associations to the disease terms "Colitis, Ulcerative", "Crohn Disease", "Colorectal Neoplasms" and "Colonic Neoplasms" were found using the Genomic Association Database (GAD) [27]. A Gene-disease interaction network map was generated using cytoscape [117]. The node size and node label font size in the resulting network diagram is proportional to the degree of edges associated with each node. Significant genes for each phenotype identified within our study were colored according to phenotype.

Genetic Regulation by Transcription Factors and miRNA

Transcription factor (TF) regulation was inferred from the results of our integrated microarray analysis based upon well-characterized TF target genes using TFactS [118]. Significant TFs were identified using a p-value cutoff of 0.01 for each respective phenotype.

Significant differentially expressed miRNA were obtained from recent reviews of miRNA regulation in both CRC [119] and IBD [120]. Only miRNA identified as being significantly altered in CRC by at least two studies were considered for subsequent analysis. Gene targets of the identified miRNA were found using miRWalk [64]. Only those Gene-miRNA interactions in which an upregulated miRNA interacts with a downregulated gene as identified by our analysis and vise versa were considered for subsequent analysis. MiRNA-Gene-disease networks were then constructed in order to infer the involvement of miRNA and their target genes on regulation of IBD and CRC phenotypes.

3.4 RESULTS

In this study we integrated microarray data obtained from multiple laboratories (Table 5a) along with other publicly available data and various knowledge bases to better define the transcriptional regulation IBD related CRC and identify sets of discrete genetic classifeirs. Each unique patient sample was hybridized onto the Affymetrix Human Genome U133 Plus 2.0 GeneChip® Array platform, which contains 54,675 oligonucleotide probes, covering over 47,000 human transcripts mapped onto 17,778 NCBI Entrez gene IDs [46]. Integrated microarray analysis was performed to reduce laboratory specific noise and increase statistical power. Figure 7 shows the sample clustering based upon the principal components, which account for 82% of the variance within the dataset. Significant gene lists were compared in order to identify commonly regulated genes for IBD (UC & CD) as well as CRC (Dukes A-D) (Table 5b). There were found to be 965 upregulated genes common to both UC and CD, while CRC phenotypes were found to have 315 upregulated genes common to each phenotype. There were found to be 111 genes commonly upregulated within each phenotype (both IBD and CRC). Conversely, IBD phenotypes were found to share 479 downregualted genes, while CRC phenotypes were found to share 712 downregulated genes. There were found to be a total of 227 downregulated genes shared by IBD and CRC phenotypes. The top 200 up and downregulated genes identified within CRC cancer stages as ranked by average fold change across cancer stages is provided in Appendices F and G respectively.

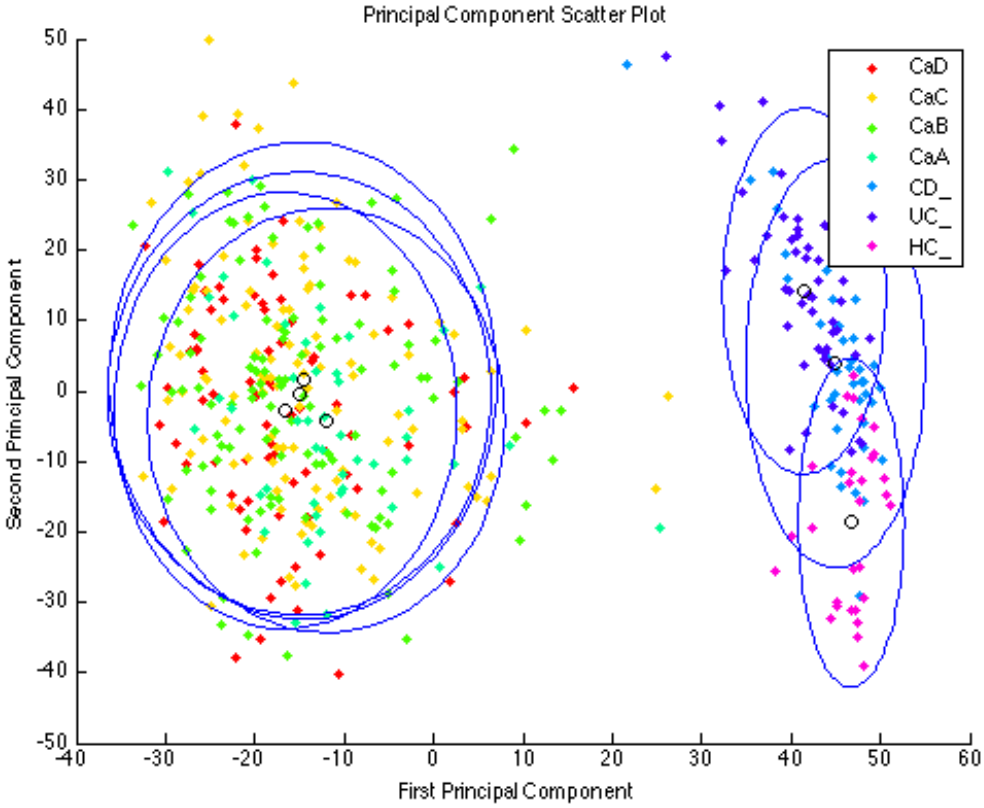


Figure 7. Principal component analysis scatter plot of CRC cancer stages and IBD phenotypes. Colorectal cancer Dukes A (CaA), Dukes B (CaB), Dukes C (CaC), Dukes D (CaD), ulcerative colitis (UC_) , crohns disease (CD_) and healthy control (HC_) samples contained within the meta-analysis are represented.

5a)

GEO Accession #	PMID	HC	UC	CD	Dukes A	Dukes B	Dukes C	Dukes D	Total
GSE17536	19914252	0	0	0	24	57	57	39	177
GSE17537	19914252	0	0	0	4	15	19	17	55
GSE14333	19996206	0	0	0	44	94	92	60	290
GSE9452	19177426	5	8	0	0	0	0	0	13
GSE9686	18069684	8	5	11	0	0	0	0	24
GSE10191	18981162	11	8	0	0	0	0	0	19
GSE10616	18758464	11	10	32	0	0	0	0	53
GSE13367	19834973	10	8	0	0	0	0	0	18
GSE16879	19956723	6	24	19	0	0	0	0	49
Subtotal		51	63	62	72	166	168	116	698
Duplicates		19	13	11	23	53	57	39	215
Outliers		3	2	10	3	14	10	14	56
Total		**29**	**48**	**41**	**46**	**99**	**101**	**63**	**427**

5b)

MICROARRAY DATA ANALYSIS

	UC/HC	CD/HC	Dukes A	Dukes B	Dukes C	Dukes D
Upregulated Genes	1229	1050	424	465	495	486
Downregulated Genes	828	539	917	892	946	957
Common Up	965		315			
				111		
Common Down	479		712			
				227		
Union (All Significant Genes)	2202		1715			
				801		

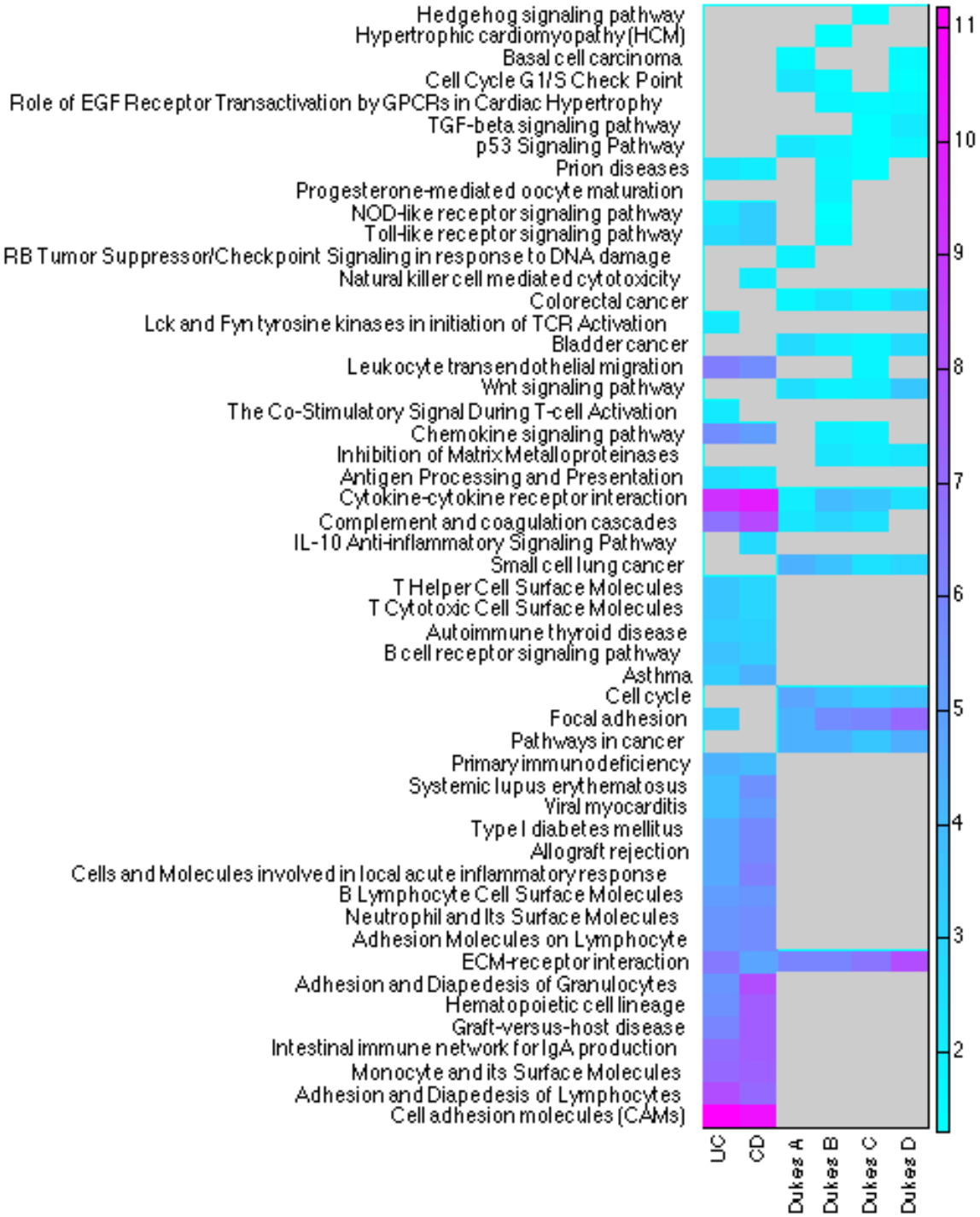
Table 5. Overview of the samples used within the IBD-CRC meta-analysis (a). Identified significant genes for each phenotype including gene list intersections for various sample comparisons (b).

Functional Enrichment

Functional enrichment of individual gene lists for each phenotype was performed in order to visualize significantly altered disease pathways common to both IBD and CRC. Figure 8 demonstrates the widespread pathway alterations associated with IBD and CRC. Figure 8b shows the upregulation of cytokine-cytokine interactions, complement and coagulation cascade and ECM receptor interactions in both IBD and CRC. IBD is shown to be associated with an upregulation in NOD and toll like receptor signaling, leukocyte transendothelial migration, chemokine signaling and numerous other pathways related to an immune response including those related to T and B cell mediated immunity. Significantly upregulated pathways for CRC include the p53 signaling pathway, colorectal cancer, baldder cancer, wnt signaling, small cell lung cancer, cell cycle and pathways in cancer.

Downregulated pathways were found to drastically differ between IBD and CRC phenotypes (Figure 8b). Commonly downregulated pathways include fatty acid metabolism, nitrogen metabolism, PPAR signaling pathway, nitrogen metabolism, valine, leucine and isoleucine degradation, drug metabolism, butanoate metabolism and mitochondrial carnitine palmitoyltransferase (CPT) system. Downregulated pathways solely associated with CRC include pentose and glucuronate metabolism, retinol metabolism, steroid hormone synthesis, amino sugar and nucleotide sugar metabolism and ascorbate and aldarate metabolism.

8a)


Hedgehog signaling pathway
Hypertrophic cardiomyopathy (HCM)
Basal cell carcinoma
Cell Cycle G1/S Check Point
Role of EGF Receptor Transactivation by GPCRs in Cardiac Hypertrophy
TGF-beta signaling pathway
p53 Signaling Pathway
Prion diseases
Progesterone-mediated oocyte maturation
NOD-like receptor signaling pathway
Toll-like receptor signaling pathway
RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage
Natural killer cell mediated cytotoxicity
Colorectal cancer
Lck and Fyn tyrosine kinases in initiation of TCR Activation
Bladder cancer
Leukocyte transendothelial migration
Wnt signaling pathway
The Co-Stimulatory Signal During T-cell Activation
Chemokine signaling pathway
Inhibition of Matrix Metalloproteinases
Antigen Processing and Presentation
Cytokine-cytokine receptor interaction
Complement and coagulation cascades
IL-10 Anti-inflammatory Signaling Pathway
Small cell lung cancer
T Helper Cell Surface Molecules
T Cytotoxic Cell Surface Molecules
Autoimmune thyroid disease
B cell receptor signaling pathway
Asthma
Cell cycle
Focal adhesion
Pathways in cancer
Primary immunodeficiency
Systemic lupus erythematosus
Viral myocarditis
Type I diabetes mellitus
Allograft rejection
Cells and Molecules involved in local acute inflammatory response
B Lymphocyte Cell Surface Molecules
Neutrophil and Its Surface Molecules
Adhesion Molecules on Lymphocyte
ECM-receptor interaction
Adhesion and Diapedesis of Granulocytes
Hematopoietic cell lineage
Graft-versus-host disease
Intestinal immune network for IgA production
Monocyte and its Surface Molecules
Adhesion and Diapedesis of Lymphocytes
Cell adhesion molecules (CAMs)
UC
CD
Dukes A
Dukes B
Dukes C
Dukes D
11
10
9
8
7
6
5
4
3
2

8b)

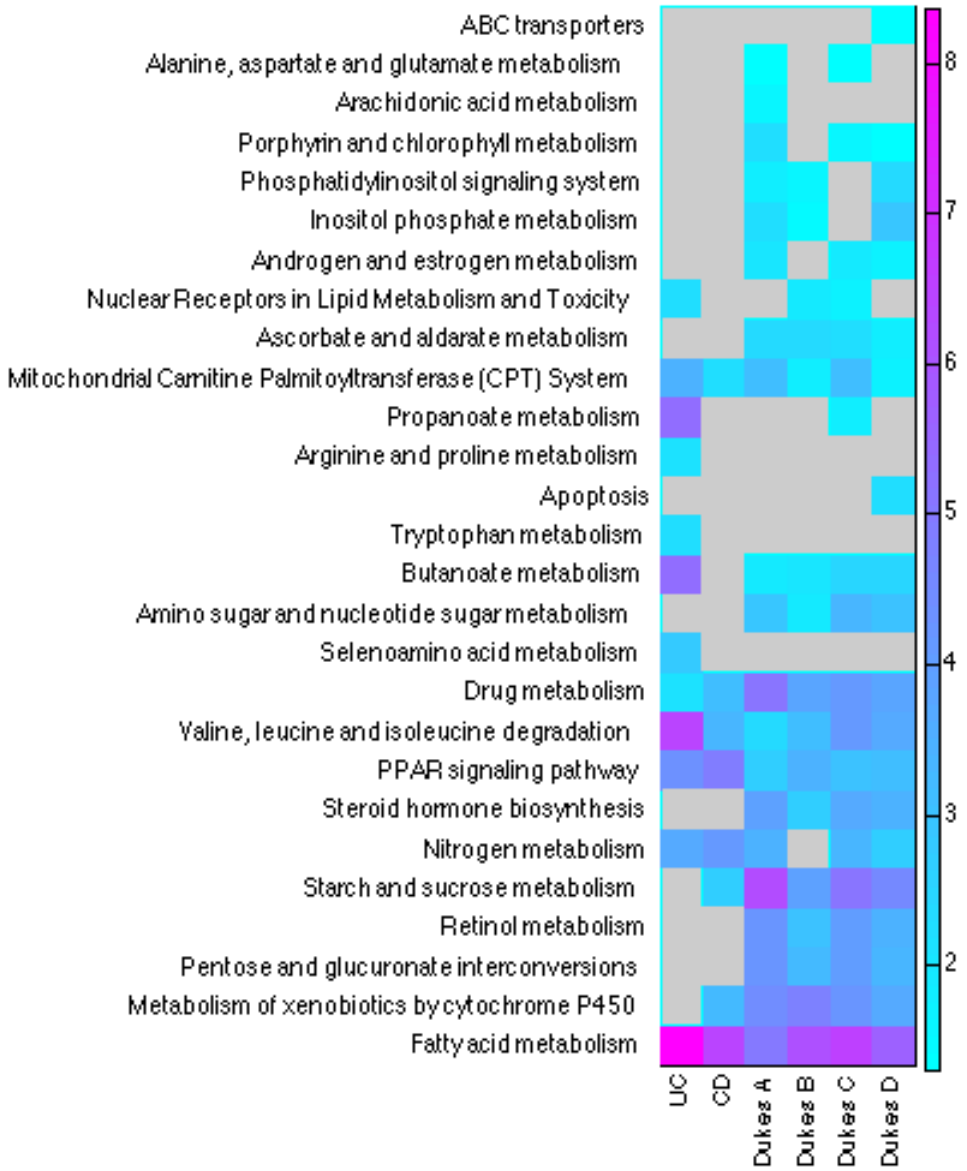


Figure 8. Functional enrichment of identified significant genes for IBD phenotypes and CRC Dukes A-D staging. Statistically enriched upregulated (8a) and Downregulated (8b) KEGG and BioCarta pathways.

Disease Classifiers

Discrete transcriptional disease markers were identified between healthy controls and IBD phenotypes as well as various stages of CRC in order to identify unique classifier sets and assess their performance in differentiating patient populations. The identified classifier lists are provided in Table 6 along with their performance in classifying samples within the test dataset (Figure 9). Classifiers were found to adequately separate healthy controls from IBD and CRC samples regardless of staging, with AUC's approaching unity. There was however great difficulty in differentiating between cancer stages as with AUCs ranging between 0.65 and 0.85 indicating significant overlap in genetic regulation of the various stages of CRC.

In Silico Disease Classifier Validation

Classifier performance was evaluated on an independent dataset obtained form GEO, which contained IBD, CRC and healthy control biopsies obtained from the colon of affected individuals [104]. However since the dataset was not annotated with cancer staging or IBD phenotype only the HC/IBD, HC/CRC and IBD/CRC biomarker sets were validated. It was found that using all 10 disease classifiers perfectly separated disease classes, with an AUC of 1. Consequently, we investigated the minimum number of disease classifiers that could be used to accurately classify samples (AUC>0.9). It was found that using the top five biomarkers for each comparison resulted in an acceptable AUC (Figure 10).

Colorectal Cancer Comparisons									
HC/A	**HC/B**	**HC/C**	**HC/D**	**A/B**	**A/C**	**A/D**	**B/C**	**B/D**	**C/D**
WNT2	TIMP1	WNT2	WNT2	CCL20	COL10A1	SPP1	DEFA6	CXCL3	CXCL3
CALU	MMP14	MMP14	TIMP1	TREM1	SPP1	GRP	ESM1	CXCL1	CXCL1
COL1A1	ADAMTS2	THBS2	TG	SPP1	AGT	CXCL3	CMTM7	CXCL2	CMTM7
TG	TG	COL12A1	MMP14	OLR1	OLR1	CXCL1	C8ORF84	CHI3L1	MMP12
IL28A	SFRP4	MFAP2	IL28A	IGHV@	TREM1	CXCL2	IGFBP3	APOD	AGT
ESM1	ECM1	CALU	THBS2	COL10A1	TIMP1	APOD	SERPINA1	CXCL9	SERPINB5
PGLYRP4	ESM1	TIMP1	LOXL2	AGT	ISLR	LCN2	CCL25	SFTA2	MMP1
TIMP1	COL1A2	SFRP4	PGLYRP4	TNC	CCL20	ASPN	NMU	IL8	LPL
MFAP2	LOXL2	COL1A1	TGFBI	SERPINE1	SERPINE1	CFB	STC1	CCL18	IL8
ADAMTS2	SPON2	IL28A	MFAP2	CFB	FLNA	CCL20	TCN1	MMP1	FOLR1

IBD Comparisons							
HC/UC	**HC/CD**	**UC/CD**	**UC/CRC**	**CD/CRC**	**HC/IBD**	**IBD/CRC**	**HC/CRC**
CALU	CHI3L1	AZGP1P1	COL11A1	COL11A1	CHI3L1	COL11A1	WNT2
REG4	MMP3	GALNT2	IGFL2	IGFL2	TIMP1	IGFL2	MMP14
GALNT2	TIMP1	CKLF	GKN2	GRP	MMP3	GRP	TIMP1
LCN2	SPINK4	LAMB2	GRP	PRB2	SPINK4	PRB2	CALU
LTF	PCSK1	TCN1	PRB2	ESM1	CALU	PNOC	MFAP2
CD44	IGKV4-1	IGLV3-25	PNOC	PNOC	MMP12	PGLYRP4	IL28A
CRELD2	MMP12	MGC29506	CFP	SFRP4	REG4	CCL25	COL1A1
TNFRSF6B	LTF	ST3GAL1	PGLYRP4	GKN2	MMP10	GKN2	TG
SPINK4	CALU	IL23A	NUCB2	COL10A1	APOL1	SFRP4	THBS2
SPON2	LIPG	IGHV3-11	CCL25	CEL	CRELD2	ESM1	PGLYRP4

Table 6. Classifier lists for phenotype comparisons identified by ROC analysis. The classifiers used to classify the exernal validation set samples by logistic regression are highlighted.

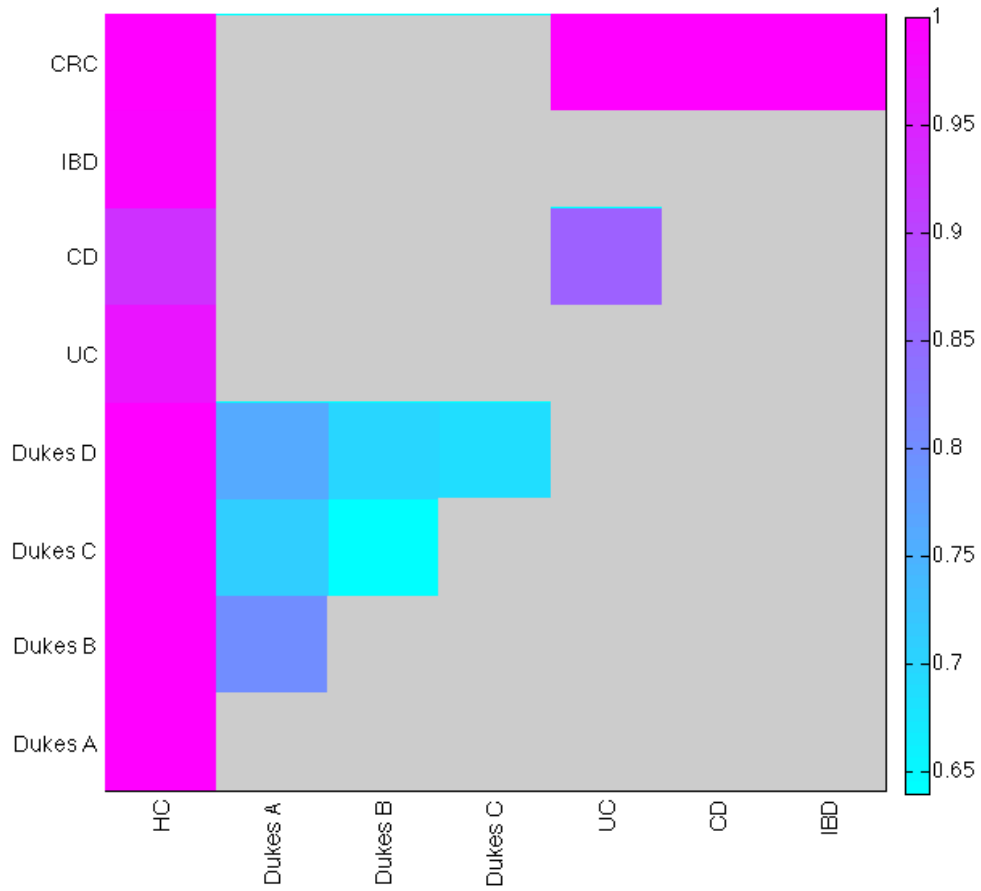


Figure 9. Classifier performance identified for each phenotype comparison on the internal test set. The color bar represents the AUC of each comparison on the respective test dataset.

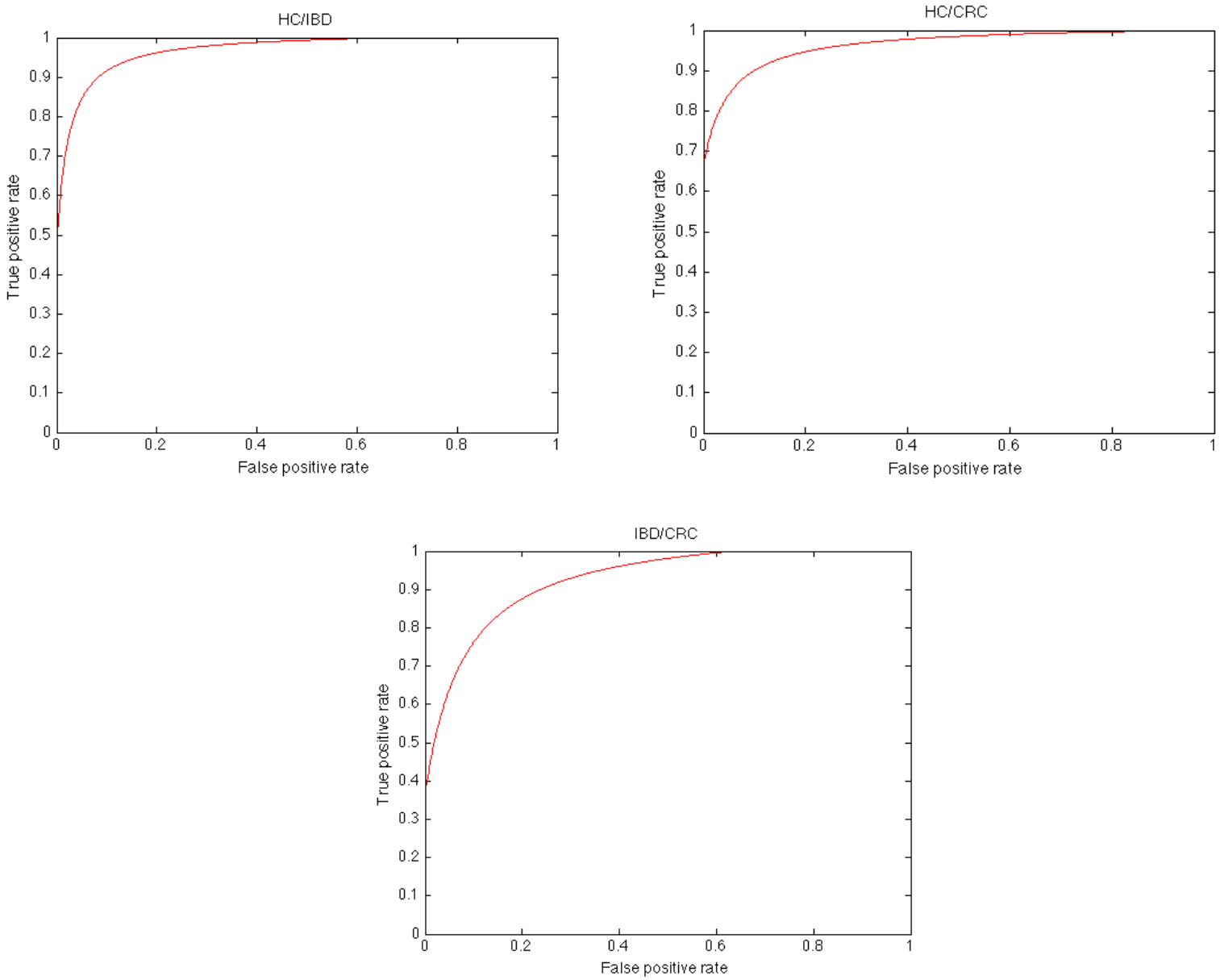


Figure 10. ROC for classification by logistic regression of HC/IBD (AUC = 0.98), HC/CRC (AUC = 0.98) and IBD/CRC (AUC = 0.94) comparisons.

Disease-Gene Interactome

According to GAD, there are 10 genes with known associations to both UC and CD, and 30 genes associated with both IBD and CRC phenotypes (Figure 11). Significant genes identified by our integrated microarray analysis for UC, CD and CRC are colored within Figure 11 in order to portray the overlap with our results and known gene-disease associations. Only one gene, PPARG, is known to be associated with UC, CD and CRC and was also found to be significantly downregulated in HC/UC, HC/CD and HC/CRC comparisons.

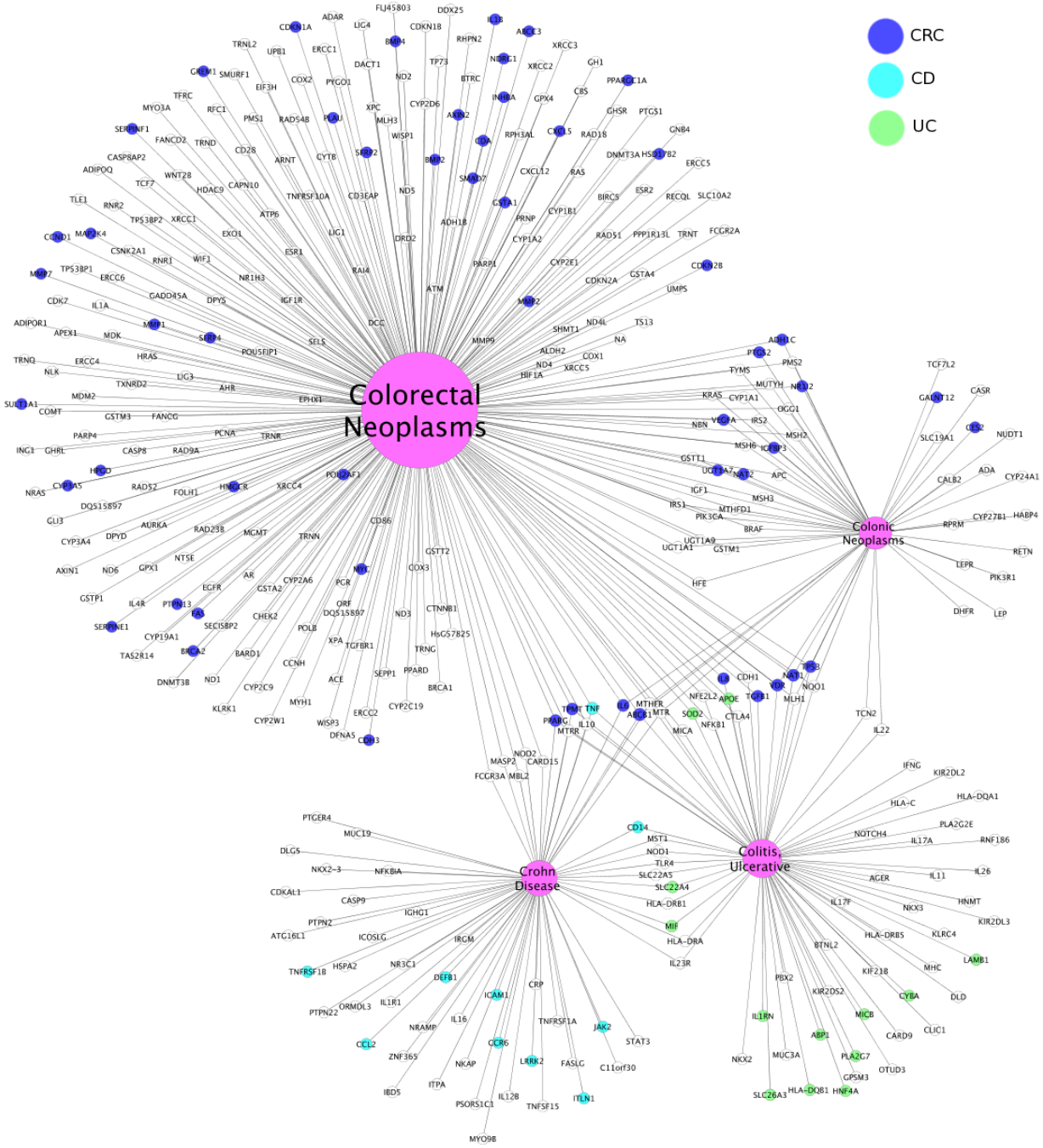


Figure 11. Disease-Gene interactome. Genetic associations for MeSH disease terms related to CRC and IBD. Significant genes identified by this meta-analysis are highlighted.

Genetic Regulation by Transcription Factors and miRNA

Transcription factor enrichment shows differential expression of transcription factors in both IBD and CRC phenotypes. It is shown that CEBPE, CREM, NR2F1, ATF1, HIF1A, STAT3, SP3, COX2, TBP, TP53, RELB and EGR1 are all significantly enriched for CRC, while NRC, REL and CEBPD are all significantly enriched in IBD phenotypes (Figure 12).

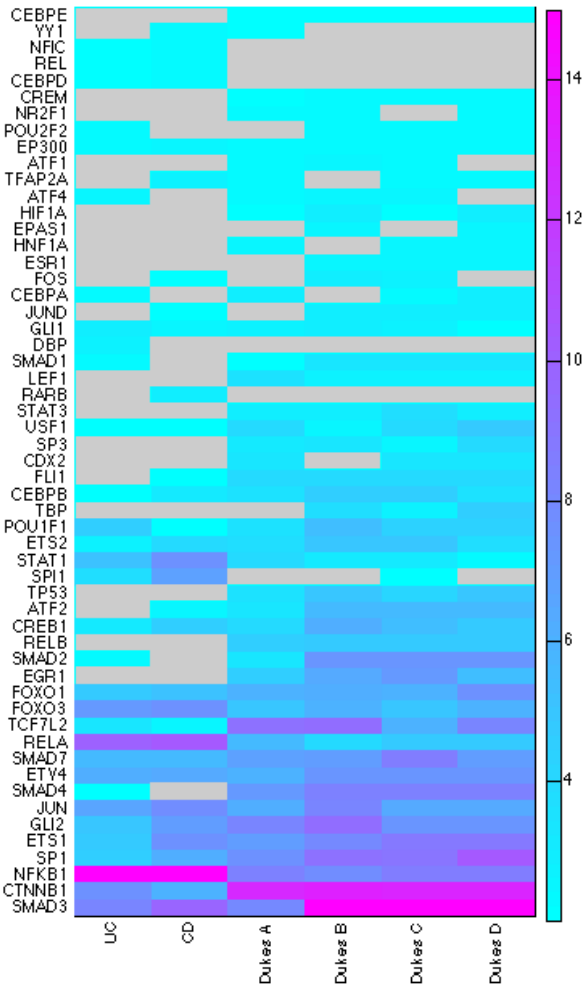


Figure 12. Significantly enriched transcription factors for IBD phenotypes and stages of CRC.

Upregulated miRNA were found to interact with downregulated genes identified within our analysis and overlap between UC, CD and CRC phenotypes. Mir-155 and miR-29b were found to be upregulated in both UC and CD, while miR-21, miR-29a and miR-31 were found to be upregulated in UC, CD and CRC. UC and CRC were found to both have upregulated miR-223, while CD and CRC were found to both have upregulated miR-107 and miR-200c (Figure 13a).

Downregulated miRNA were also found to overlap between UC and CRC phenotypes including miR-143, miR-145 and miR-192 (Figure 13b). The interacting genes of these commonly regulated miRNA were then determined to see the influence of miRNA on the regulation of significant genes for each phenotype. Upregualted miRNA-21 was shown to downregulate PDCD4, THRB, SMAD7, FGFR3, ENPP3 and CNTN3 within all phenotypes. Upregulated miRNA-29a was shown to downregulate FZD5 across all phenotypes. Upregulated miR-31 was shown to downregulate the expression of SATB2, FGFR3, CDKN2B and TRIM36 amongst all phenotypes. FGFR3 was the only gene commonly downregulated across every phenotype by more than one miRNA. The upregulation of miR-223 was shown to downregulate KLF4 in UC and CRC phenotypes, while the upregualtion of miR-107 was associated with the downregulation of PDK4 in UC and CRC. Interestingly, downregulated miRNA were only shown to overlap between UC and CRC phenotypes including miR-143, miR-145 and miR-192. Genes commonly upregulated by the attenuation of miR-192 inlcude COL1A1, COL1A2, COL4A1, CTGF and TGFB1. Genes commonly downregulated by miR-143 include MMP3, TAGLN and VCAN. The only gene commonly downregulated in both phenotypes by miR-145 is CDH11.

Interestingly, P53 and VEGFA were found to be upregulated in CRC by the attenuation of miR-143, miR-145 and miR192.

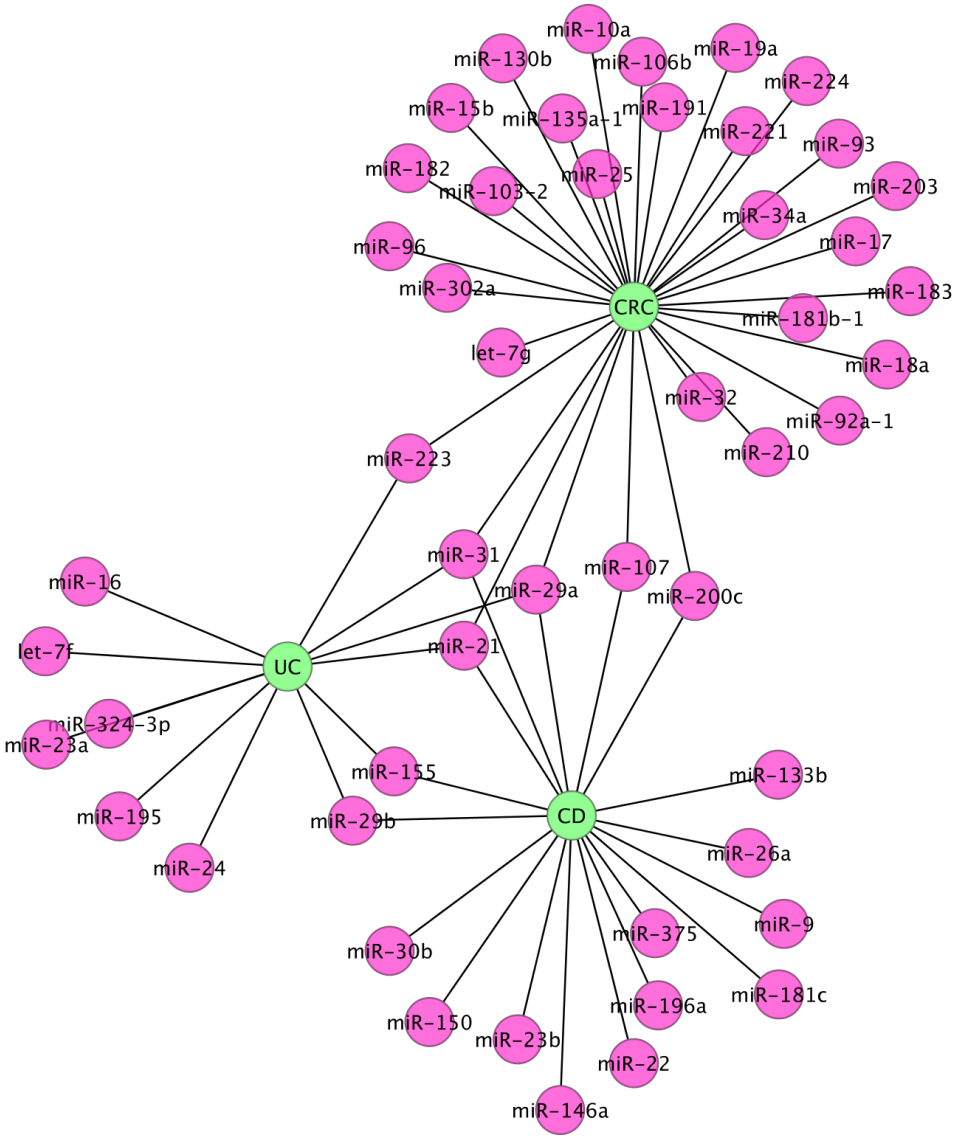


Figure 13a. Upregulated miRNA with validated interactions with significant downregulated genes identified by our integrated microarray analysis for UC, CD and CRC.

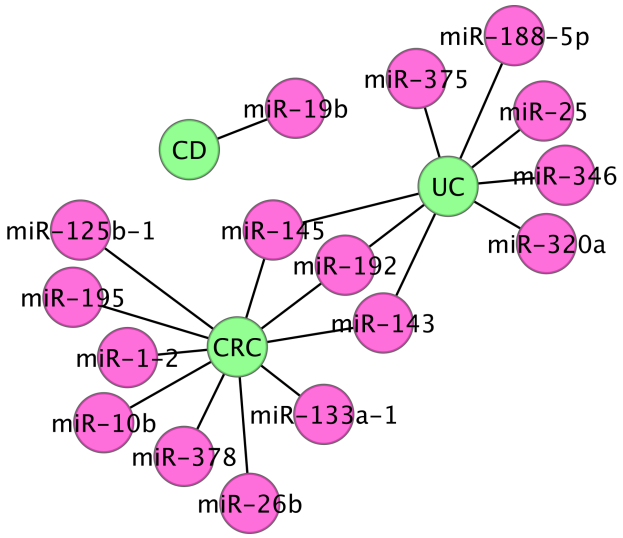


Figure 13b. Downregulated miRNA with validated interactions with significant upregulated genes identified by our integrated microarray analysis for UC, CD and CRC.

3.5 DISCUSSION

Our analysis of 427 unique patient samples obtained from 9 independent microarray studies revealed gene signatures, pathways, biomarkers, transcriptional factors and the role of miRNA regulation in IBD phenotypes and the stages of CRC. Our results indicate transcriptional factors, miRNA and genes involved in the transition from IBD to CRC phenotypes as well as unique genetic regualtors involved in each phenotype respectively. We have furthermore identified biomarker sets that may be used to diagnose and assess the prognosis of colorectal cancer patients.

The identified biomarkers for HC and IBD patient classification that were used to classify the samples within the validation set include CHI3L1, TIMP1, MMP3, SPINK4 and CALU. Chitinase 3-like 1 (CHI3L1) is a glycoprotein member of the glycosyl hydrolase 18 family, secreted by activated macrophages, chondrocytes, neutrophils and synovial fluid. CHI3L1 was found to have a FC value of 19 and 16 for UC and CD respectively. The shear abundance of this transcript make it an excellent biomarker candidate for active inflammation and interventional effectiveness. The second marker, TIMP1 is a natural inhibitor of MMPs and is also known to promote cell proliferation and anti-apoptotic functions. TIMP1 gene expression is known to be induced by various cytokines. Other markers in the list include matrix metallopeptidase 3 (MMP3), serine peptidase inhibitor, kazal type 4 and calumenin (CALU).

Biomarkers identified for the classification of HC and CRC patients include WNT2, MMP14, TIMP1, CALU and MFAP2. TIMP1 and CALU were also found to be excellent classifiers for IBD and HC patients. WNT2 is a member of the wingless-type MMTV integration site (WNT) gene family, which

consists of structurally related genes which encode signaling proteins implicated in oncogenesis. Microfibrillar-associated protein 2 (MFAP2) is major antigen of elastin associated microfibrils and is believed to play a role in the etiology of inherited connective tissue diseases. Biomarkers found to differentiate between IBD and CRC phenotypes include COL1A1, IGFL2, GRP, PRB2 and PNOC.

UC and CRC phenotypes are shown to both have downregulated miR-143, miR-145 and miR-192. The validated interacting mRNA transcripts that were found to be upregulated within this study and known to interact with each miRNA for each disease phenotype were also found. Downregulated miR-143 may result in upregulated ITGAM, LPIN1,and MMP9 in UC, FSCN1, MYC, TP53 and VEGFA in CRC and MMP3, TAGLN and VCAN in both UC and CRC phenotypes. Significant upregualted genes known to interact with miR-145 include BNIP3, CBFB, KRT7, MCL1, MUC1, SOD2, STAT1, SWAP70 and ALCAM for UC, CCND1, CDK4, EIF2C2, FSCN1, MYC, SOX4, TP53 and VEGFA for CRC and CDH11 for both UC and CRC. Downregulation of miR-192 may result in the upregulation of ALCAM, APOE, CD79A, ITK, MCL1 and NOS2 in UC, CCND1, EIF2C2 and TP53 in CRC and COL1A1, COL1A2, COL4A1, CTGF and TGFB1 in UC and CRC.

Upregulated miRNA were also found to have known interactions with a number of significant downregulated genes for UC, CD and CRC. Upregulated miR-21 was shown to downregulate CNTN3, ENPP3, FGFR3, PDCD4, SMAD7, THRB, FZD5, CDKN2B, FGFR3, SATB2 and TRIM36 across UC, CD and CRC patients. Upregulated miR-155 and miR 29b were shown to downregulate HMOX1 in UC and CD, while upregulated miR-29b was found to downregulate CEBPA and CDKN2B in UC and CD. While the coregulation of mRNA has been

shown for shared miRNA, there are a number of miRNA that are known to interact with commonly regulated genes that are not innervated by shared miRNA and these complex redundant interactions warrant further research and discovery in order to map the complex genetic regulation involved in IBD and CRC.

There were found to be a total of 53 significant upregulated transcription factors innervated by downregualted miRNA and 28 significant downregulated transcription factor genes innervated by upregulated miRNA suggesting complex interactions and feedback loops between miRNA and various transcription factors. SMAD7, PPARG, CDKN2B and THRB are all significant transcription factor genes known to be downregulated in UC, CD and CRC. MiR-21 has been shown to be significantly upregulated in UC, CD and CRC and is known to interact with the upregulated transcription factors SMAD7 and THRB, while miR-31 is known to interact with CDKN2B. SMAD7 is a nuclear protein that binds the E3 ubiquitin ligase SMURF2, which interacts with TGF-beta receptor type-1 (TGFBR1). Variations in SMAD7 have been shown to confer susceptibility to colorectal cancer type 3 and several transcriptional variants encoding different isoforms of this protein have been found. Similarily, polymorphisms in PPARG have been shown to confer susceptibility to IBD [121, 122]. The downregulation of these transcription factors in the presence of upregualted miRNA suggest a crucial role in the post-transcriptional regulation of these specific transcription factors by miRNA.

CHAPTER 4

MICROFABRICATED QLISA BIOSENSOR FOR MEASURING FECAL BIOMARKER CONCENTRATIONS IN IBD PATIENTS

4.1 ABSTRACT:

Microfluidic immunosensors have several advantages over standard microtiter based techniques including reduced regent consumption due to their reduced dimensionality, controllable surface properties, integration of electro-optical components and potential for high throughput sample handling and automation. Working towards the realization of a compact microfluidic immunoassay system, our group has developed a platform for the detection of quantum dot conjugated proteins within a capillary based immunoassay device [123]. This first generation device was further improved through two major design improvements including optimized antibody conjugation to PMMA substrates and microfluidic device integration. The device was validated using human samples spiked with lactoferrin, an IBD biomarker. This chapter highlights my contributions to one submitted manuscripts, two patents and one conference paper related to the realization of a second-generation microfluidic QLISA platform.

4.2 INTRODUCTION

Immunoassay Technology

Immunoassays are routinely used to identify and quantitatively evaluate the concentration of an analyte contained within a test sample through the exploitation of selective and reversible antigen-antibody interactions (eq. 1.1) and have been employed in a variety of clinical, pharmaceutical, defense and environmental applications. The specificity of this reaction in combination with characteristically high association constants make them an ideal biosensing modality and are the basis behind the gold standard of in vitro diagnostics; the enzyme linked immunosorbent assay (ELISA).

$$Ag + Ab \Leftrightarrow Ag \bullet Ab \text{ (Complex)} \qquad (1.1)$$

In the ELISA method, antibodies are used to detect specific antigens contained within a biological sample. Traditional heterogeneous solid-phase immunoassays, such as an ELISA are carried out on the solid substrate of a standard microtiter plate, which is typically manufactured from polystyrene and may have 6, 24, 96, 384 or 1536 sample wells. In this format, proteins are passively adsorbed onto the microtiter plate surface through noncovalent hydrophobic interactions between the non-polar or aromatic amino acid residues present within the protein and the hydrophobic polymer chains present on the substrate, resulting in randomly oriented antibodies. Some of the adsorbed antibodies are denatured in this process, leaving as few as 10% of the total adsorbed antibodies available for an affinity reaction [124, 125]. Furthermore, passive adsorption of antibodies onto the substrate results in low, inhomogeneous surface coverage with concentrations of approximately 400ng/cm^2 on PMMA [125]. Alternatively, homogeneous immunoassays may be

carried out in the liquid phase of the wells of a microtiter plate, with no biomolecules adsorbed onto the substrate. Overall, the ELISA method is very sensitive due to the high specificity of antibody binding and the amplification of the binding signal by means of enzymatically catalyzed fluorescence (often accomplished using horseradish peroxidase), which may be detected by measuring the absorbance or transmittance of light using a standard microtiter plate reader.

While microtiter plate based immunoassays offer sufficient automation and throughput, they often require the excessive consumption of expensive reagents and require the oversight of a trained technician as well as a significant amount of analysis time to perform. In order to circumvent these issues microfluidic immunosensors have been investigated as an alternative immunoassay platform, allowing for increased throughput and automation, lower reagent and sample consumption as well as decreased analysis time. The advantages that microfluidic immunosensor platforms have over standard microtiter plate based assays arises from their reduced, micron scale features, which effectively increases their surface area to volume ratio over that of standard microtiter plates while decreasing working volumes and the characteristic diffusion length for biomolecules. As a result, heterogeneous assays are particularly amenable to a microfluidic platform since they may exploit the increased surface area to volume ratio uniquely inherent to such microscale devices, allowing for comparable sensitivity, specificity and limit of detection as microtiter plate based assays in only a fraction of the time and with decreased sample and reagent consumption. Furthermore, microfluidic immunosensors may be integrated with various components including

micromixers, valves, micropumps, electrical and optical components to facilitate the design and development of highly automated, precise and portable micro total analysis system capable of multianalyte detection as a point-of-care diagnostic platform [126-128].

Microfluidic immunosensors may be fabricated from a variety of materials depending upon the final application and the desired chemical, material and optical properties of the substrate. The selection of an appropriate substrate often depends heavily upon the signal transduction mechanism employed by the sensor, the desired surface chemistry and the available fabrication techniques. Most immunosensors that rely on optical excitation and/or detection mechanisms are fabricated from materials with a high optical transmittance within the visible and near UV wavelengths and low autofluorescence. For this reason, glass [129, 130] and various polymeric materials including PMMA [131-136], PDMS [137-139], polycarbonate [140-142] and more recently the cycloolefin Zeonor [143, 144] have all been successfully used as immunosensor substrates in applications where an optical signal transduction and detection mechanism is employed. In addition to their superior optical properties, polymers may also be manufactured on a commercial scale using existing techniques such as hot embossing and injection molding to inexpensively mass-produce disposable and highly reproducible immunosensor platforms. Lastly, the surface of polymeric materials such as PMMA may be easily functionalized and modified to facilitate the covalent immobilization of properly oriented antibodies directly onto the substrate effectively enhancing assay sensitivity and allowing for decreased limits of detection while reducing non-specific binding [131, 145, 146].

Capillary sensors have found extensive application in numerous disciplines of science, ranging from electronics and semiconductors to chemical and biological sampling & detection methods, like chromatography [147]. The versatility of a capillary comes from its geometry, its high surface area to volume ratio, the possibility to create a surface with perfect adhesion and use transparent construction materials [148]. A sturdy minute capillary construction allows the immunosensor device to be compact and portable. The small size ensures easy, cost effective and accurate detection for small sample sizes. A larger surface to volume ratio in a capillary compared to the geometry of wells in well-plates (currently used as standards), can facilitate detection reactions and increase sensitivity, by reducing diffusion[149]. Capillary based biosensors can eliminate the complex process of microfabrication required for microfluidic immunosensors. Capillary tubes can act as both optical waveguides and sampling chambers. The cylindrical shape facilitates easy and uninterrupted flow-in and flow-out of samples and the inner surface can be used for immobilization of probing molecules such as antibodies [150].

Both glass and polymer based capillaries have been used to carry out immunoassays. Specifically glass or fused silica[151-154], polystyrene [155], polymethylpentene [156], polymethylmethacrylate [157], poly-vinyl chloride[158] have been used in fabricating capillary biosensors. Polymeric tubes may pose a manufacturing challenge and glass capillary tubes need additional processing steps of surface activation for successful fixation and immobilization of probes. The inert nature of glass makes it necessary to perform additional chemical modification to form covalent bonds between a glass surface and a biomolecule. Chemicals used for such modification / functionalization comprise

of silinazing agents like 3-mercaptopropyl trimethoxysilane. These silanizing agents bind simultaneously to the hydroxy groups on the glass silicate surface and to various groups on the probe surface ranging from the amino terminal to, thiol groups, based upon the specific silanizing agent [154]. Therefore, the choice of these silinazing agents depends upon the type of immunoassay that they are going to be used in and the probes that they need to attach to. On the other hand, polymeric capillaries, such as the ones we are proposing to use, are of particular interest due to readily available functional groups on their surface offering an appropriate substrate for immobilizing antibodies or antigens, without special modification steps. Furthermore, recent developments in photochemical methods of functionalizing polymeric materials provide compelling reasons to choose polymeric capillaries over fused silica or glass capillaries [159]. Several strategies have been reported in the literature to detect low concentrations of antigens by solid phase immunoassay in capillaries, primarily focusing on excitation of fluorophores followed by collection of the emitted photons. One such approach takes advantage of the evanescent field at the interface of the polymer/liquid interface for collecting the emitted signal; however this method requires the material of the capillary to function as a waveguide [147], restricting the choice of material among those having very low scattering coefficient. Fused silica or glass, though meet the requirements, are fragile and require additional cladding layers to protect the capillaries from mechanical damage. This method is further challenged by issues related to optical alignment and coupling efficiency often resulting in reduced performance.

The first capillary waveguide application for a sensor was reported by Weigl and Wolfbeis in 1994 [160], for carbon dioxide measurement. In this case optical fibers perpendicular to the capillary were used for both excitation and collection. Ligler et al in 2002 provided a summary on four different potential excitation and collection methods for integrating waveguide capillaries [161]. Several biosensors utilize the integrating waveguide capillary biosensor design with data collection from the ends of the capillary. Zhu et al reported development of a capillary based waveguide sensor for detection of E. coli O157 [162]. The method described utilized a capillary based sandwich immunoassay for detecting the presence of E. coli O157. Recently, Stringer et al reported a capillary biosensor based on Forster Resonance Energy Transfer (FRET) [163]. The sensor utilized the superior properties of Qdot fluorescent particles and a Teflon AF capillary as a Liquid Core Waveguide to detect the presence of Troponin I in solution as well as in biological samples. The limit of detection of the sensor was 32nM for troponin I in solution and 55nM in spiked human plasma.

Working to improve the current limitations of ELISA technology, our lab's predicate QLISA platform exhibited comparable sensitivity with current methods while using $1/100^{th}$ of the reagent volume and requiring less than half the operating time required of current ELISA methods [123]. Our preliminary data collected from a proof-of-concept device which includes UV LED as the excitation source indicate that it is indeed possible to design and fabricate a prototype capable of detection MPO and Lf with LDL at 10pg/ml. Building upon this technology to further increase assay performance and throughput, a second generation QLISA technology was developed. An overview of the first

and second-generation devices is shown in Figure 12. The developed microfluidic platform allows for more precise fluidic control, site-specific antibody immobilization and integration with other components.

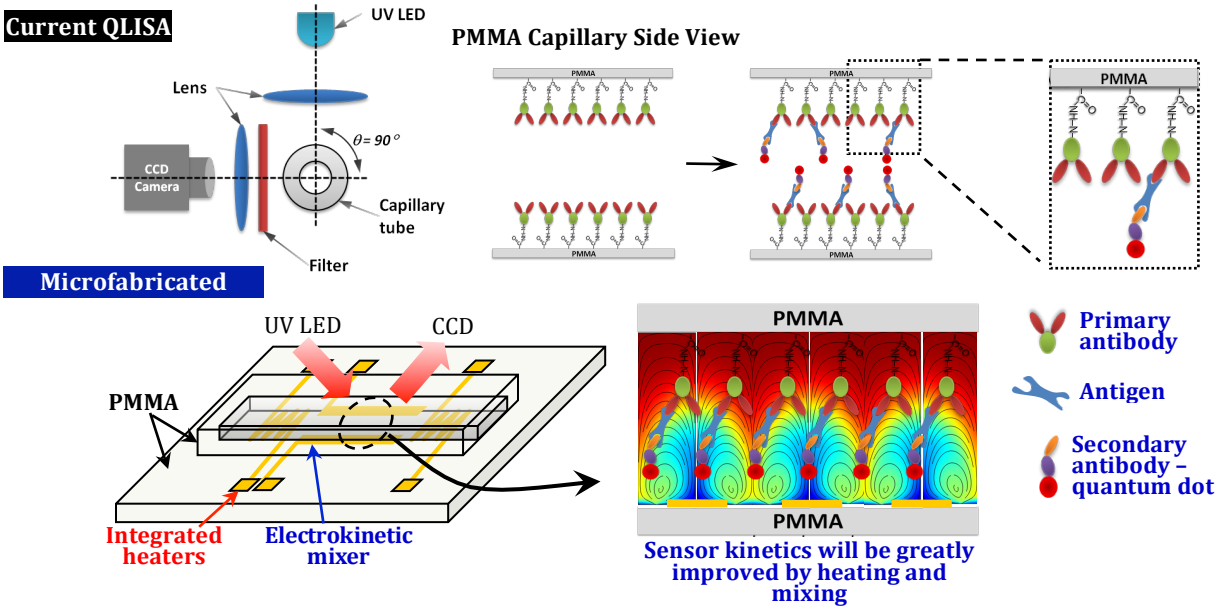


Figure 14. Capillary based first generation QLISA optical setup and surface chemistry using zero-length crosslinking (top) and microfluidic QLISA platform with integrated electrodes for electrothermally driven enhancement of reaction kinetics (bottom).

Biosensor Surface Functionalization

Maintaining the native physiological conformation of an antibody following its immobilization on a substrate is critical in determining the efficiency of any subsequent affinity reaction and thereby influencing important biosensor parameters such as selectivity and sensitivity [164-167]. Therefore, the selection of an antibody immobilization procedure is a key factor in determining

the performance parameters of any immunosensor. Achieving proper antibody orientation following immobilization allows for higher signal to noise ratios, enhanced reaction rates and assay sensitivity while effectively reducing an immunosensor's limit of detection [168, 169].

In order to increase antibody surface coverage for biosensor applications, numerous covalent immobilization strategies have been developed using both zero and non-zero length crosslinkers. Perhaps the most ubiquitous technique used to immobilize antibodies onto a polymeric substrate is through the formation of a covalent amide linkage between any reactive amine present within an antibody and a carboxylate moiety on the substrate using the zero-length crosslinker 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) [170]. EDC facilitates the formation of a permanent covalent amide bond between the antibody and substrate by first activating the carboxylate groups present on the substrate to form an intermediate ester, which then reacts directly with the amines present on the antibody to form a covalent linkage. Sulfo-NHS is often used in conjunction with EDC in order to increase the solubility and stability of the active intermediate, resulting in an increased yield of immobilized antibodies onto the substrate. However, since the covalent bond can be formed between any reactive amine on the antibody surface, this technique results in the random orientation of antibodies on the surface. Although this antibody immobilization strategy has been shown to increase antibody surface coverage when compared to passive adsorption, these randomly oriented antibodies may have a 2-3 fold lower antigen binding capacity when compared to properly oriented antibodies, as a result of their limited mobility and increased steric hindrances at the antigen

binding site [171]. As a result of the reduced antigen binding capacity of antibodies immobilized using zero length crosslinkers, long flexible polymeric linkers such as diamino poly(ethylene glycol) – amino–PEG and poly(ethyleneimine) – PEI – have been investigated as spacers in an attempt to increase the mobility of the immobilized antibodies while maintaining their active binding sites. The use of non-zero length crosslinkers offers several advantages over traditional antibody immobilization methods by allowing a high density of physiologically active antibodies to be immobilized onto a polymeric substrate. Using the amine bearing polymer PEI as a spacer to immobilize an antibody onto a PMMA substrate effectively increases the antigen capture tenfold as compared to passively adsorbed antibodies [145] by limiting steric hindrance and antibody denaturation following immobilization, effectively enhancing assay sensitivity for the detection of an IgG analyte [145, 146]. Amino–PEG spacers have also been used with great success to covalently bind physiologically active antibodies onto PMMA substrates. Amino–PEG spacers have all of the benefits associated with PEI in addition to reducing non-specific interactions with the substrate by creating a hydrophilic polymer brush coating on top of the polymer substrate [172]. Both amino–PEG and PEI also come in a variety of molecular weights allowing for further optimization of chain density and flexibility. Although the effect that the spacer length has on antibody density, non-specific adsorption and the functionality of the immobilized antibodies has not yet been fully elucidated, it has been shown that the length of the spacer is a critical determinant of the spacer coil conformation, which in turn alters the antigen binding capacity of the immobilized antibodies [173]. Protein A, a cell wall component of *Staphylococcus aureus*, has been known for its

specificity in binding to the Fc portion of immunoglobulin G (IgG). This specificity has found several applications in optical [174, 175] and flow injection [176, 177] immunosensor fabrication. It has been postulated that this specificity results in orientation of the antibody [178, 179] rendering the binding sites readily available for the antigen particularly in a sandwich assay

Quantum Dots

Quantum dots are semiconductor nanocrystals that emit photons when excited by light. They do not photobleach and are useful for a variety of imaging applications[180-184]. Commercial QDs are environmentally harmful and toxic to humans due to the organic solvents and heavy metal elements. Many groups have studied the synthesis of water soluble QDs by adding silica coatings using 3-(mercaptoprpyl) trimethoxysilane (MPS) and other polymeric as well as non-toxic surfactants[183].

Quantum dots exhibit a florescence emission based on their physical size (Figure 1). This has to do with the energy gap between the conduction and valence energy bands on these semiconductor metals when arranged as a nanocrystal at sizes smaller than ~10 nm [185, 186]. When a semiconductor gets excited, there is a gap between the excited electron and the hole it leaves behind. The Bohr energy radius is the distance between the excited electron and the gap formed. Therefore, when the size of the semiconductor is on the nanoscale the diameter is close to the Bohr energy radius, which forms a QD with distinct energy levels at these band gaps[187]. The energy spectra is modified at these sizes leading to quantum confinement. As the size decreases the energy gap

increases, which will lead to a blue shift in the emission wavelengths of these nanocrystals[187, 188].

The wide range of biosensor applications using QDs is encouraging. Given their unique and advantageous photophysical properties, biosensing developers attempt to harness and maximize the potential of QDs, which has created a rapidly growing field. Perhaps no potential is greater than the multiplexing capabilities of QDs in biosensor development. Broad overlapping excitation spectra of different sized QDs allow simultaneous excitation with multiple distinct emission spectra correlating to specific QDs. This paves the way for multiple biomarker detection in a single sample and has significant advantages in translational medical applications, like our QLISA system shown in figure 12 and described in further detail in the following sections.

Inflammatory Bowel Disease Diagnostic Biomarkers

Inflammatory bowel disease (IBD) is an autoimmune, chronic inflammatory condition of the proximal most portions of the gastrointestinal (GI) tract, which is characterized by two distinct phenotypes: ulcerative colitis (UC) and Crohn's disease (CD). It is estimated that nearly 780,000 and 630,000 people suffer from UC and CD respectively within the United States alone [2], although these numbers are believed to be an underestimate since IBD is often misdiagnosed and clinical reporting varies significantly due to the presentation of a variety of common symptoms shared by other gastroenterological disorders. Consequently, gastroenterologists mainly rely on colonoscopies and histological examination before prescribing any particular diagnosis. While colonoscopies and subsequent biopsies allow the gastroenterologist to definitively diagnose

IBD from other gastrointestinal disorders, they are highly invasive, time consuming and costly procedures that only provide a qualitative assessment of intestinal inflammation. Alternatively, biomarkers have been studied extensively as a means to non-invasively diagnose and monitor disease activity. While the measurement of serological biomarkers may seem like a feasible measurement of inflammatory related disease indices, these inflammation biomarkers may often be elevated as an artifact of some other systemic inflammatory condition. Consequently, many researchers are actively investigating the presence and concentration of various fecal biomarkers present in patients with IBD.

As a result of elevated intestinal inflammation, IBD patients excrete an elevated quantity of leukocytes within their stool. Traditionally, these leukocytes have been measured via radioactive labeling, which is often coupled to an abdominal scintigraphy procedure to identify inflamed segments of the intestine. ^{111}Indium labeled granulocyte scintigraphy has been validated as a highly sensitive (97% sensitivity) and specific quantitiative assessment of both the extent and severity of intestinal inflammation as compared to traditional endoscopic and histopathological indices [106]. Although this technique is an invaluable measure of the intestinal inflammatory response, the technique requires specialized facilities, subjects the patient to radiation exposure. As a result, this technique is primarily used by researchers and is limited in routine clinical practice. As an alternative to measuring leukocytes directly, leukocyte derived proteins may also be measured and have shown great promise as indicators of disease severity.

Although numerous leukocyte derived biomarkers have been investigated [111, 113, 189, 190], lactoferrin and calprotectin have been shown to be the most

sensitive and specific in discriminating IBD patients from a population of healthy patients [111]. Accordingly, our group has developed a QLISA based assay for lactoferrin and assessed the concentration of lactoferrin within a spiked human sample both within the new microfluidic system and in the first generation capillary based system.

4.3 METHODS

Device Fabrication

Microfluidic devices can be manufactured from thermoplastics using a variety of techniques including direct micromachining, laser ablation, injection molding and hot embossing. In order to develop our prototype polymethylmethacrylate (PMMA) microchannel design, we opted to use a hot embossing technique since a single mold insert may be rapidly manufactured and used to create many pmma microchannels, allowing us to rapidly create microchannels in pmma that are consistently reproducible. In working towards this goal, we first developed a 2.5" square copper mold insert with raised microstructures using the techniques of photolithography and electroplating (apapted from [191]). The raised microstructures are then transferred to the pmma substrate via hot embossing to create microstructures within the pmma substrate. Lastly, the device is sealed using a solvent assisted bonding technique (adapted from [192]). An overview of this process is presented in Figure 13 and a more detailed description of this process is provided in the following sections.

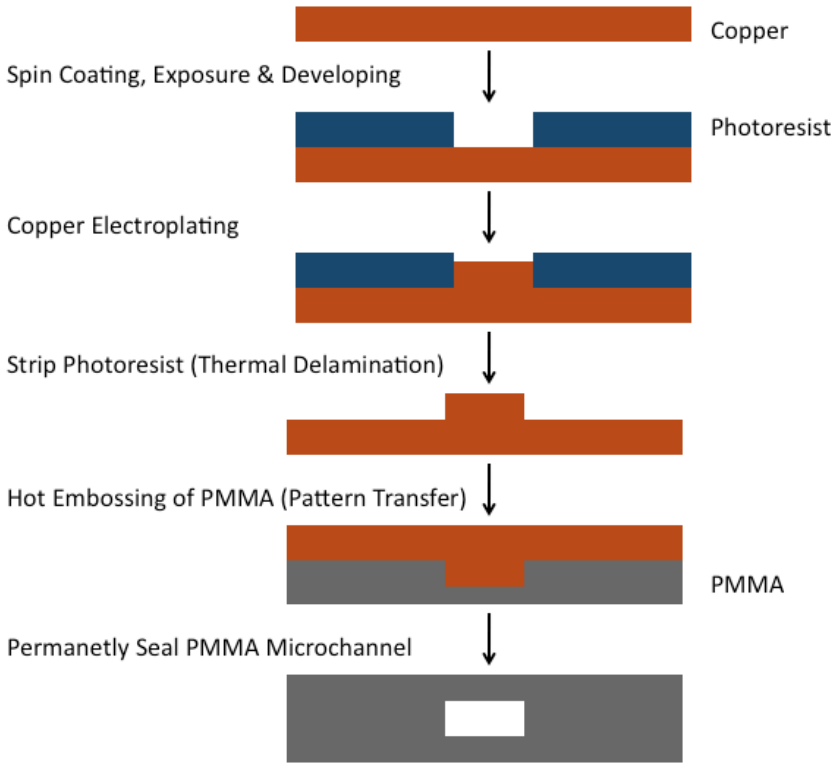


Figure 15. Profile view of the PMMA microchannel fabrication processs. First, the copper is cleaned and patterned with 250μm thick negative photoresist (SU-8 2150), allowing for an active site for copper electroplating in the subsequent step, producing raised microchannel features on a copper substrate (mold insert). The mold insert is then used as a die in the subsequent hot embossing process to produce a microchannel in the desired PMMA substrate. The device is then sealed with a pmma cover, effectively creating an array of sealed microchannels.

The fabrication process begins with a photomask, which in this case is a clear field mask. A simple photomask was first designed in order to optimize the fabrication procedure and demonstrate proof of concept for subsequent immunoassay reactions. The photomask consists of microchannels of various widths (50, 100 and 250 μm) and pitches (5, 7.5 and 10mm) between

microchannels. The photomask was first designed using AutoCAD 2010 and sent to Advanced Reproductions for printing. Upon receipt, the thin film photomask was replicated to create a more durable chrome mask Once the photomask pattern was transferred to a chrome photomask, photolithography was performed on a cleaned (acetone followed by DI water) 1mm thick 2.5" copper substrate (McMaster Carr, catalog# 9821K12). SU-8 2150 photoresist (Microchem) was then spun onto the substrate at 500rpm (100r/s acceleration) for 10 seconds followed by spinning at 17000rpm (300r/s acceleration) for 30 seconds to produce a final thickness of approximately 200um. The copper plate was then placed onto a hot plate and baked at 60°C for 20 minutes followed by a 40 minute bake at 90°C (softbake). The coated substrate is then cooled to room temperature and exposed at 28mW/cm^2 for 17 seconds (476mJ/cm^2). Following exposure, the photoresist patterned substrate was placed onto a hot plate and baked at 65°C for one hour (a latent image of the pattern should not be visible until minutes into the post exposure bake, if a latent image is seen the pattern was likely overexposed). The patterned substrate is then cooled to room temperature and developed for ten minutes in SU-8 developer under manual agitation. The end product (micropatterned SU-8 on copper substrate) is shown below in Figure 14.

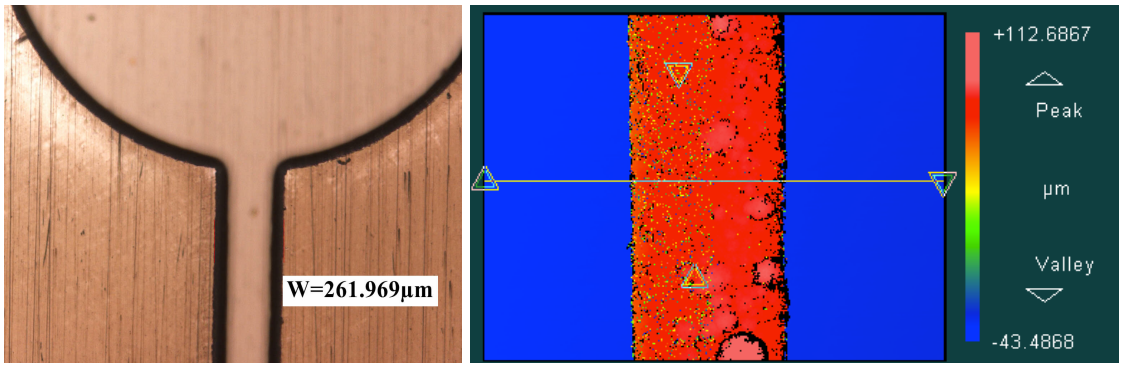


Figure 16. SU-8 patterned copper substrate (left). The raised SU-8 may be seen on top of the copper substrate. The final width of the microchannel is shown to be approximately 260um. The exposed (developed) regions of the SU-8 serve as a substrate for the adhesion of copper during the subsequent electroplating process. Profilometry measurements of the raised copper microchannel structure demonstrated the height of the microchannel is approximately 150µm (right).

Once the SU-8 micropatterned copper substrate has been manufactured copper electroplating is performed to create raised microchannel structures on the copper substrate. First, the patterned substrate is cleaned extensively with acetone followed by a rinse in DI water and air-dried. The patterned copper plate is then placed into a holder 35cm from a bare copper plate (cleaned with acetone & DI water). The apparatus is then submerged in a copper electroplating solution (200g copper (II) sulfate pentahydrate, 25ml of concentrated H_2SO_4 in 1mL of deionized water) with an applied DC current of 0.2A for 1 hour under stir bar agitation. This was found to produce uniform raised microchannel structures of 150µm in height as measured using surface profilometry (figure 16).

Following fabrication of the micropatterned copper mold insert, hot embossing was performed to create microchannel grooves on a PMMA substrate.

In order to accomplish this, 1cm thick PMMA sheets were cut into 2" squares and cleaned with soap and water and allowed to air dry. Once clean a piece of PMMA is sandwiched between a glass plate and the copper mold insert and loaded into a Carver heat press. Both the top and bottom platens were then brought into contact with the copper mold insert and glass plate respectively and set to 140°C. Once the temperature of the platens reached 110°C a pressure of 250psi is applied to the pmma sheet for a period of 5 minutes and then the platens are cooled to room temperature within the press. The molded pmma is then removed from the mold insert. A second 2" wide, 1cm thick pmma sheet is soaked in ethanol for 5 minutes, then brought into contact with the molded pmma piece and loaded into the carver heat press between two pieces of glass and the platens are heated to 75°C under a pressure of 250psi for 5 minutes. The sealed microchannel coupon is then immediately removed from the press and the lumen of the channels is rinsed thoroughly with DI water.

Surface Modification

Surface activation is carried out by circulating 1N NaOH through the microcahnnels for 60 minutes at 60°C using a peristaltic pump at 0.05ml/min. The microchannel is then washed with 2ml of 1X PBS pH 6.0 buffer. A 0.2% diamino PEG (MW 25,000 branched) solution (1X PBS pH 11.5) is then circulated thorugh the microchannels at 0.05ml/min using the peristaltic pump for one hour at room temperature. The microchannel is then rinsed for 5 minutes with deionized water at 3%. A 1% w/v gluteraldehyde solution (1X PBS pH 7.2) is then incubated within the microchannel for 40 minutes at room temperature.

The microchannels are then rinsed for 5 minutes with deionized water at 0.05ml/min. The capture antibody solution (100nM lactoferrin or myeloperoxidase antibodies in 1X PBS pH 7.2) is then circulated through the capillary at 0.02ml/min using a peristaltic pump for one hour at room temperature. The device is then rinsed with 0.05% tween for 2 minutes at 0.05ml/min using the peristaltic pump followed by a rinse with 1X PBS for 2 minutes at 0.05ml/min using the peristaltic pump. The microchannels are then blocked by incubating the channels with 2% FBS for one hour at room temperature followed by a wash with 0.05% tween (1X PBS pH 7.4) for 2 minutes using the peristaltic pump at 0.05ml/min. The microchannel device is then washed with 1X PBS pH 7.4 for 2 minutes using the peristaltic pump at 0.05ml/min and stored at 4°C until needed. An overview of the surface modified biosensor surface is presented in Figure 15.

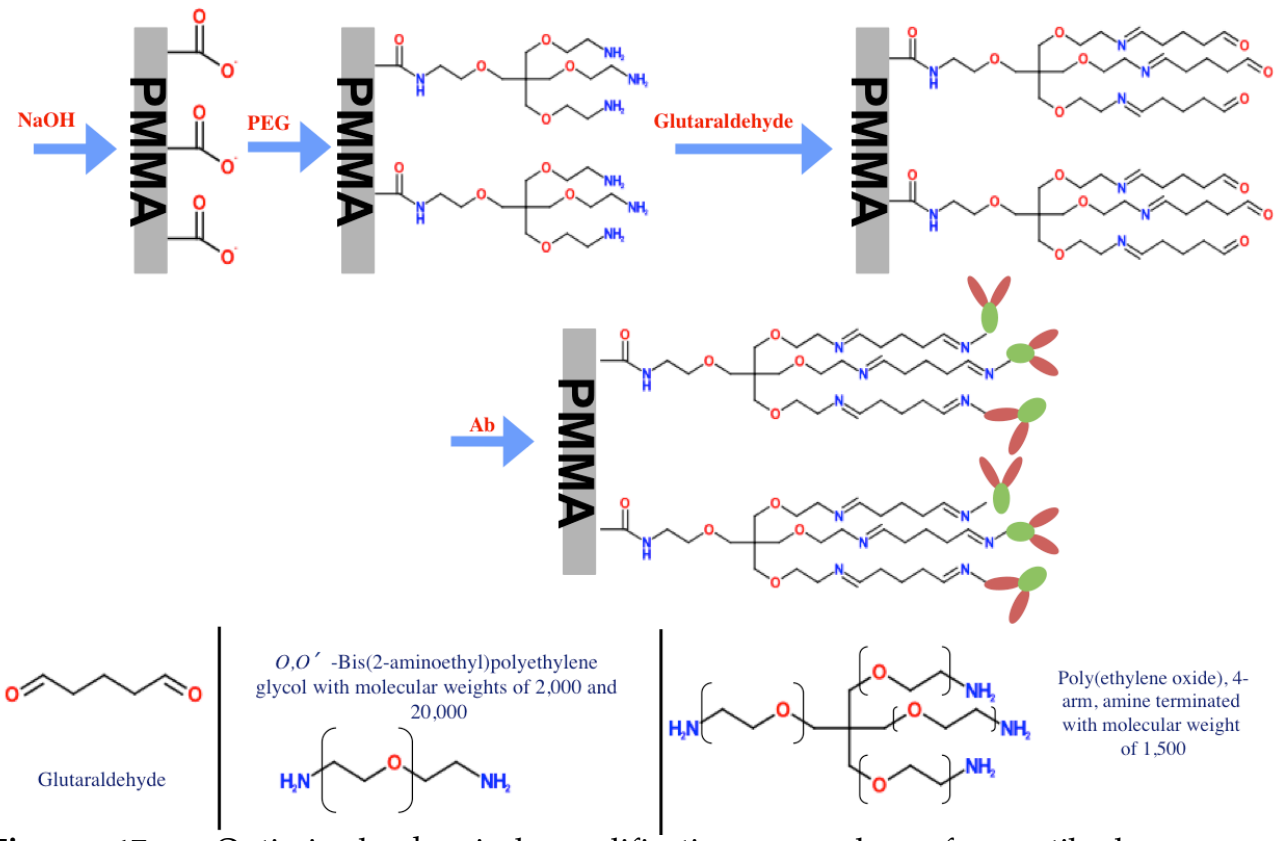


Figure 17. Optimized chemical modification procedure for antibody immobilization within microfluidic QLISA device.

Numerical Simulation

In order to investigate the influence of enhanced surface functionalization on antigen capture and time to equilibrium, a numerical simulation was performed using COMSOL. The geometry for our analysis is presented below in Figure 16 and represents the typical dimensions and volume present within a standard 96-well microtiter plate immunoassay.

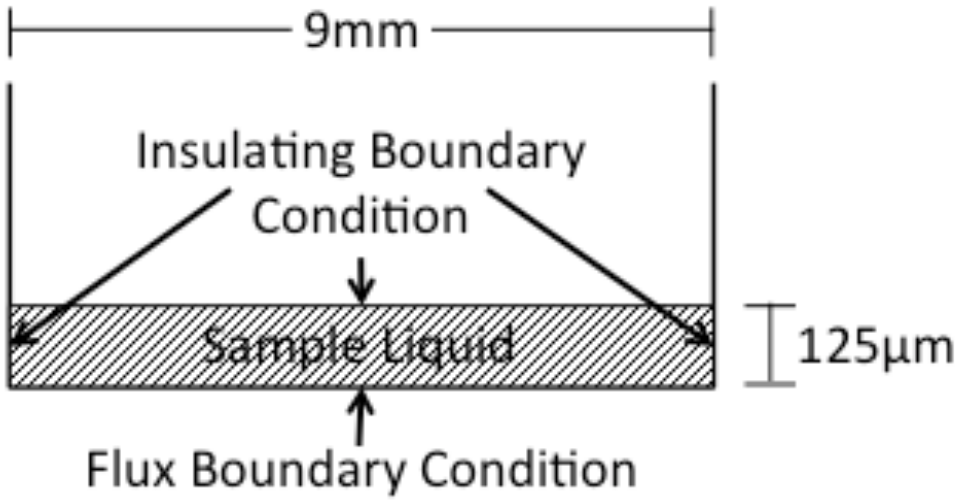


Figure 18. A 96-well microtiter plate geometry. The base of each well is 9mm with a liquid height of 125um within the well. Immobilized antibody is present at the bottom of the plate and represents the reaction surface. The boundary conditions used for the simulation are also shown.

In a typical microtiter plate immunoassay, an antigen solution is introduced into the well and allowed to react with an antibody present on the surface of the bottom of the well under static incubation. This situation results in a diffusion, reaction system in which the antigen species diffuses towards the surface where it reacts with an immobilized antibody. The reaction at the surface proceeds according to Equation 1, in which Ag, Ab and C are the antigen, antibody and antigen-antibody complex, respectively. The association and dissociation constants are represented by k_a and k_d, respectively.

$$[Ag] + [Ab] \underset{k_d}{\overset{k_a}{\rightleftarrows}} [C] \qquad \text{(Equation 1)}$$

From Equation 1, three differential equations may be formulated to represent the rate of change in the concentration of antigen, antibody and antigen-antibody complex. Since we are interested in the rate of change in antigen-antibody complex concentration, the differential equation representing the formation of antigen-antibody complex is provided in Equation 2.

$$\frac{\partial C}{\partial t} = k_a[Ag][Ab] - k_d[C] \quad \text{(Equation 2)}$$

The equilibrium concentration of antigen-antibody complex (C_{eq}) may be found by finding the analytical solution to Equation 3 at equilibrium in which Ab_0 represents the initial concentration of antibodies available for binding.

$$C_{eq} = \frac{k_a(Ab_0)C}{k_a(C)+k_d} \quad \text{(Equation 3)}$$

Since the total number of free antibodies available for binding at any given time (Ab) is equal to the difference between the initial antibody concentration (Ab_0) and the concentration of bound antigen-antibody complex (C), Equation 2 may be simplified to be expressed in terms of antigen-antibody complex and free antigen (Equation 4).

$$\frac{\partial C}{\partial t} = k_a[Ag][Ab_0 - C] - k_d[C] \quad \text{(Equation 4)}$$

$$\frac{\partial Ag}{\partial t} = D\nabla^2 Ag \quad \text{(Equation 5)}$$

The bulk concentration of antigen is given by the general diffusion equation (Equation 5) and may be coupled to concentration of antigen at the surface by establishing appropriate boundary conditions for the system in which the three walls of the liquid interface are treated as insulating boundaries and establishing a flux boundary at the reaction surface at the bottom of the microtiter plate (Figure 2) by imposing a weak form boundary in COMSOL. Constants were chosen to reflect actual immunoassay working conditions with $k_a = 1*10^5\ 1/M*s$, $k_d = 1*10^{-3}\ 1/s$, $C_0 = 100nM$ and $D = 1*10^{-11}\ m/s^2$.

Immunoassay Protocol

Approximately 1g of human stool samples were weighed and diluted 5 times (v/w) with an extraction buffer (8.359mg of ethylenediaminetetraacetic acid, 250μL of FBS, 5mL of glycerol, 12.5μL of tween, 250μL protease inhibitor cocktail and 25mL of 1X PBS pH 7.4). Diluted samples were then incubated at 4ºC for 15 minutes. The solution was then homogenized and incubated again at 4ºC for 15 minutes. To this solution an amount of lactoferrin was added to a final concentration of 100ng/ml. The solution is then centrifuged at 14,000RPM for 30 minutes at 4ºC. All samples are diluted 1:10, 1:100, 1:1000, and 1:10,000 in 1X PBS pH 7.4.

To generate standard curves, standard dilutions of lactoferrin were then introduced into the microchannel using a syringe and allowed to incubate for 60 minutes followed by a wash with 0.05% tween20 in 1X PBS pH 7.4 at a flow rate of 0.05ml/min. 100nM biotinylated secondary antibody to lactoferrin is then incubated within the microchannel for 60 minutes followed by a wash with

0.05% tween20 in 1X PBS pH 7.4 at a flow rate of 0.05ml/min. Next, a 100nM solution of streptavidin conjugated quantum dots is incubated for 60 minutes and then washed with 0.05% tween20 in 1X PBS pH 7.4 at a flow rate of 0.05ml/min.

The microchannel device is then loaded into a holder where quantum dots conjugates are excited with an ultraviolet light emitting diode (385nm wavelength) with an output power of 80mW (Nichia Corporation). Fluorescence is captured with a charge-coupled device (CCD) camera (Stingray, AVT-FS-033B) through the computer software program AVT SmartView. Images are gathered at a variety of integration times and calibration curves Raw images are processed within MATLAB. Image fluorescence intensity is quantified through comparison to a standard calibration curve to determine the lactoferrin concentration. The data acquisition setup is presented in Figure 17. An overview of the entire QLISA workflow is presented in Figure 18. A raw image of the illuminated capillary and the output from the image processing algorithm (Appendix H) is presented in Figure 19.

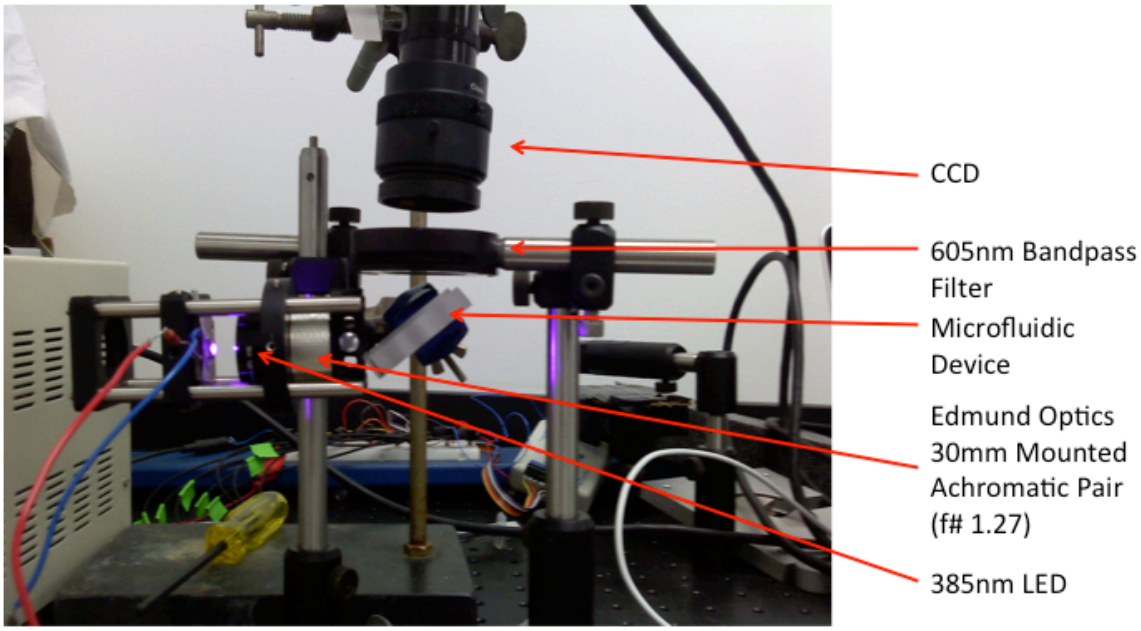


Figure 19. QLISA optical setup used for data acquisition.

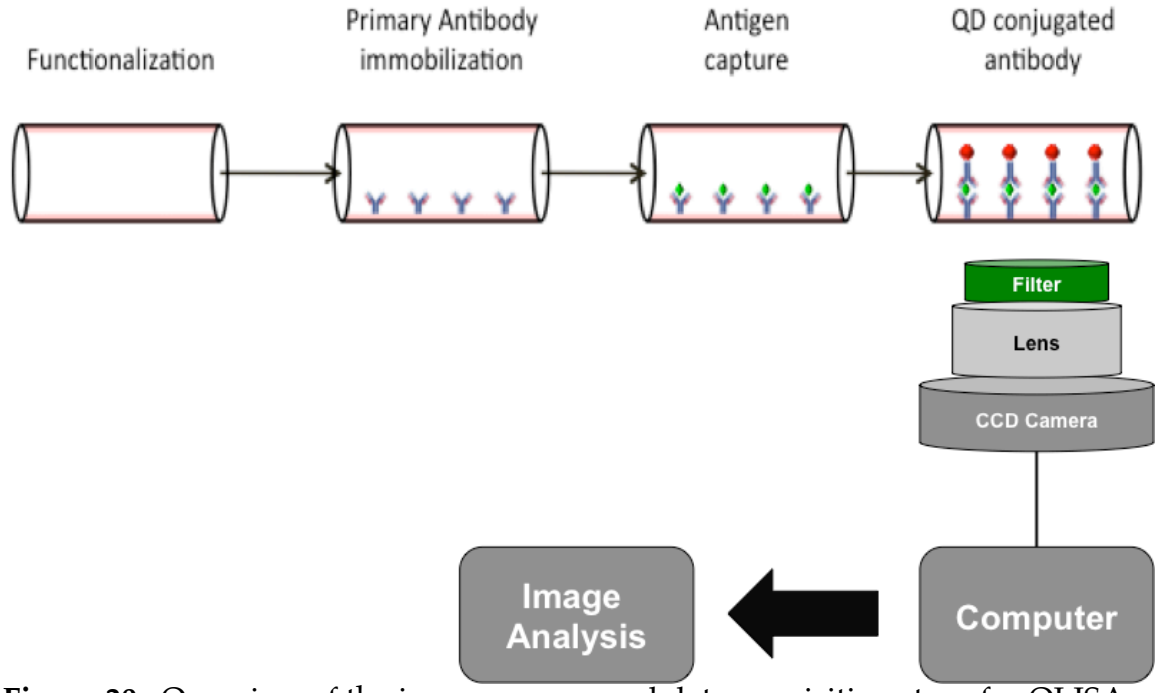


Figure 20. Overview of the immunoassay and data acquisition steps for QLISA.

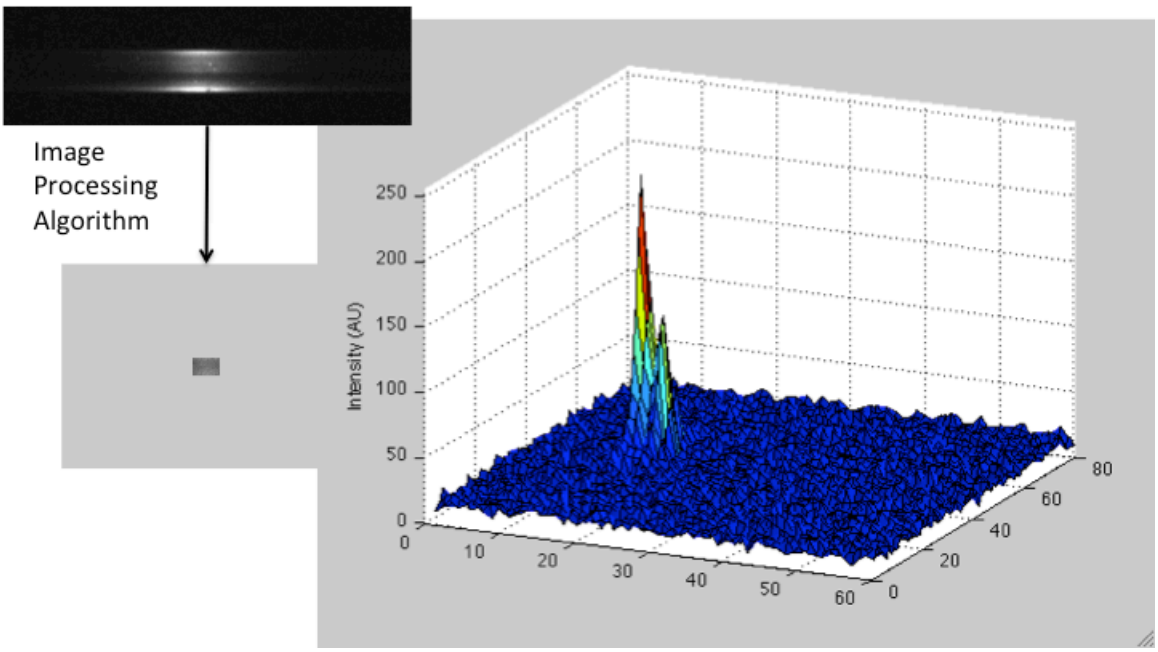


Figure 21. Image analysis performed within Matlab. The function presented in Appendix H was used to determine the intensity of the region of interest (ROI) within the original image at various integration times.

4.4 RESULTS & DISCUSSION

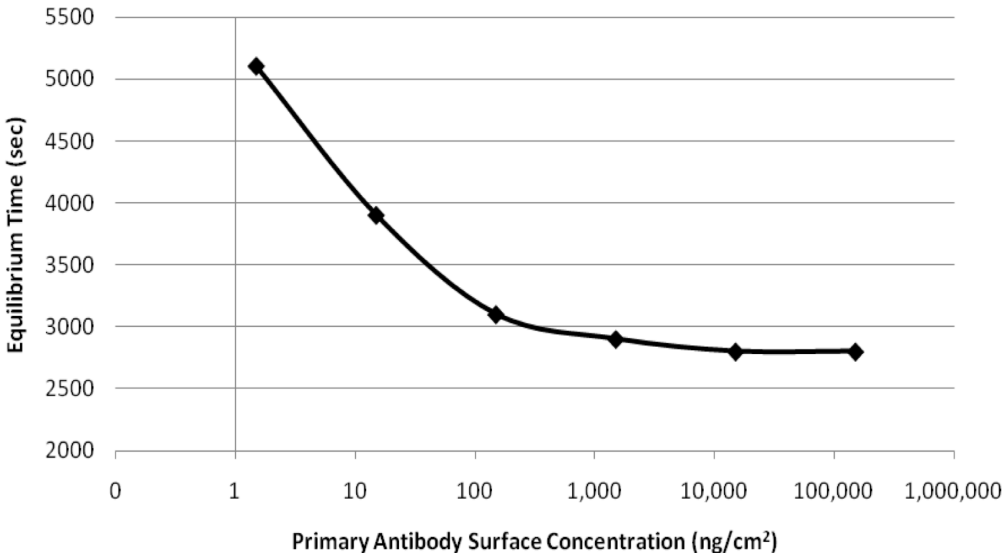


Figure 22: Numerical simulation showing time to equilibration as a function of primary antibody surface concentration.

The results from our simulations indicate that equilibrium time is reduced as the density of primary immobilized antibody is increased. This can be expected, because as the surface concentration of primary antibody is increased, there are increased opportunities for the antigen to bind the antibody and form an antigen-antibody complex at the reaction surface. Our experimental results show a primary antibody concentration of approximately $2ng/cm^2$ and $8ng/cm^2$ for passive adsorption and PEG with glutaraldehyde antibody immobilization strategies, respectively. When comparing these experimental concentrations with our numerical simulations (Figure 20) we see that equilibrium time for the PEG with glutaraldehyde chemistry is over 20% faster than passive adsorption. Further, our simulations support the need for approximately 1 hour of

incubation of primary antibody when using the PEG with glutaraldehyde chemistry. This is demonstrated in Figure 20, where 10ng/cm^2 of primary antibody requires just over 3600 seconds to reach equilibrium with its antigen. It was also observed that as the antibody surface concentration is increased above about 100ng/cm^2, we see limited reduction in equilibrium time. This is due to the saturation of the antibody-antigen reaction given our initial conditions discussed in the Methods section, and thus there is limited benefit to improving equilibrium binding time through increasing primary antibody concentration beyond 100ng/cm^2. Therefore, if possible, increasing primary antibody concentration from 10 to 100ng/cm^2 could further reduce equilibrium binding time by another 20-25%.

Linear or branched PEG has many secondary and primary amino groups available for conjugation with capture antibodies within a heterogeneous immunoassay. It is known that non-zero length crosslinkers like polyethylene glycol or polyethylene imine (PEI) can result in increased sensitivity of an assay when used as the bridge between the primary antibody and the substrate [193]. PEI in particular has been used to covalently bind antibodies onto the surface of PMMA microchannels for ELISA applications [146, 194] and has shown increased antigen binding due to increased mobility of the antibody while maintaining the antigen binding capacity of the epitope [145, 146]. By adapting the length of the antibody crosslinker (di-amino PEG) a nearly 16 fold increase in signal to noise ratio was realized over traditional zero-length crosslinking methods as shown in figure 21. The reduced steric hinderance of the capture antibody in combination with the reduced non-specific protein adsorption was demonstrated with a previously developed capillary based MPO assay [123].

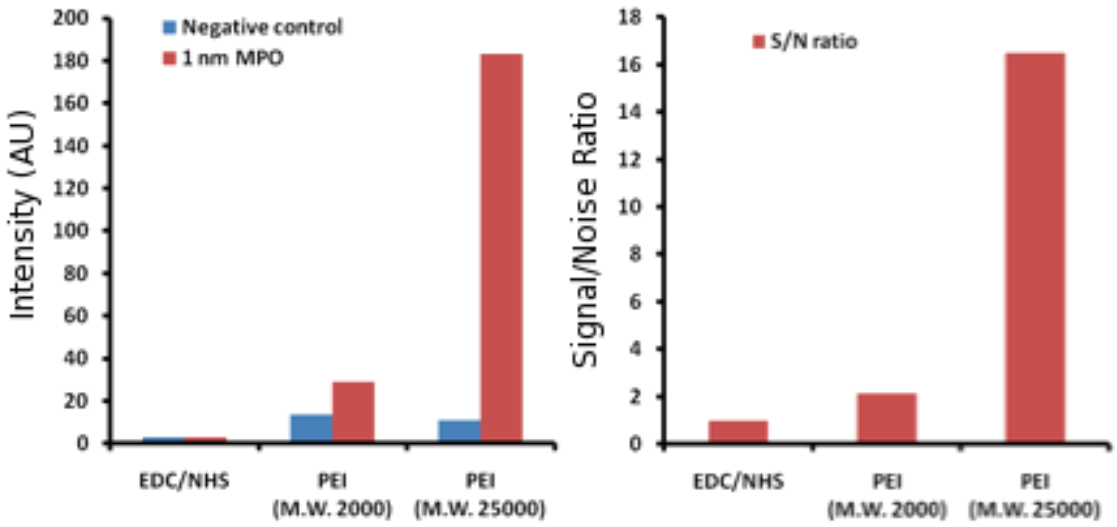


Figure 23 Raw intensity values of negative control sample compared to 1nm MPO on capillary immunoassay with enhanced surface chemistry (a) and signal to noise ratio using enhanced antibody immobilization for 1nm MPO taken at 50us integration time (b).

The increased sensitivity of the microfluidic platform was observed in measuring the IBD biomarker lactoferrin within a spiked human sample (Figure 22). There was a noticeable increase in signal generated from the human samples as a result of non-specific adsorption of extraneous proteins contained within the sample. Limit of detections for the microfluidic lactoferrin assay were calculated as being the concentration of analyte necessary to generate a signal that is equal to the standard deviation plus three times the standard deviation of the background (background noise) as interpolated using the standard curves for each platform (Figure 22) The limit of detection for the microfluidic lactoferrin assay was calculated to be 6.4ng/ml and 39.4ng/ml for the capillary based assay.

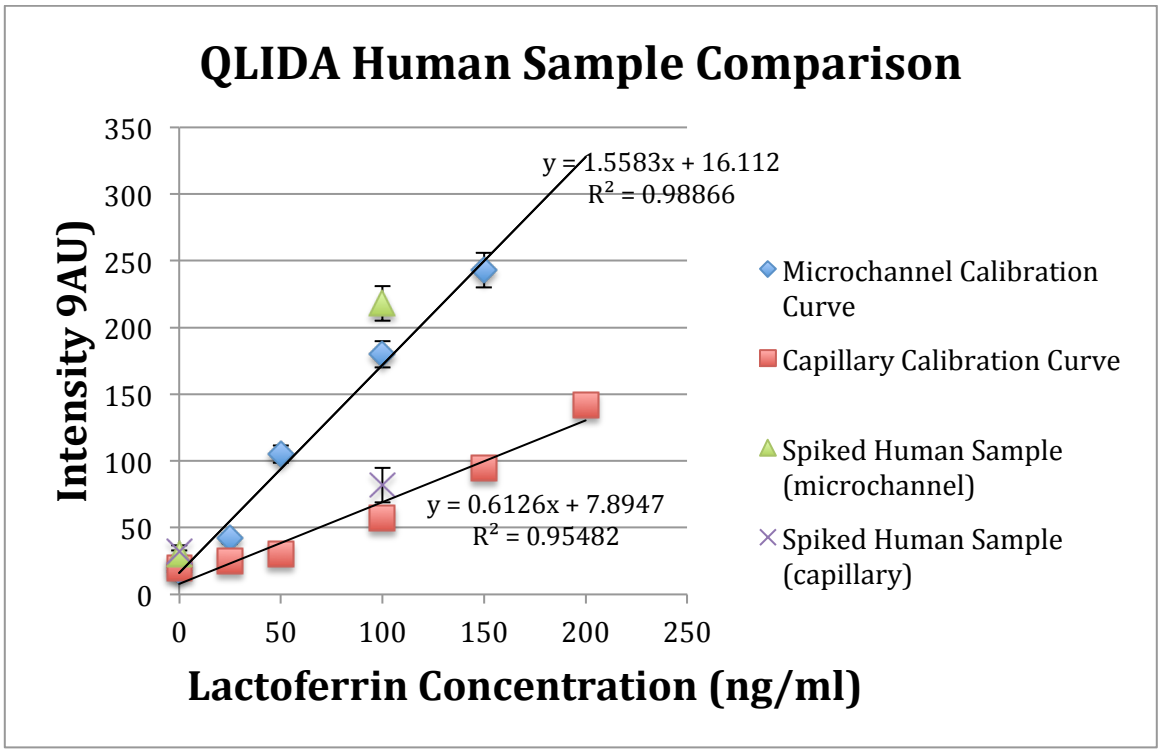


Figure 24. Standard curves for lactoferrin generated using both the microfluidic and capillary immunoassay platforms. Spiked human samples and negative controls were assessed using both platforms for comparison. Images were collected at an integration time of 50us.

CHAPTER 5

CONCLUSION

The pathophysiology of Infammatory bowel disease and the progression to colorectal cancer remains to be fully elucidated. The objective of this work is to integrate computational genomic data including miRNA and mRNA gene expression signatures and various knowledgebases to 1) Better define transcriptional regulation in IBD phenotypes, identify therapeutic targets currently available for other autoimmune diseases for drug repositioning; 2) Identify discrete genetic classifiers that can be used to differentiate between disease phenotypes; 3) Develop a proof of concept highly scalable microfluidic QLISA platform for the detection of the IBD biomarker lactoferrin within human stool samples.

Gene signatures as well as pathway analysis of the IBD and CRC transcriptome reveals significant overlap between the two diseases as well as with IBD and other autoimmune diseases including lupus, type 1 diabetes and rheumatoid arthritis. Our study is the first to map genetic links between enteropathic arthritic symptoms in patients with IBD and gene signatures expressed within the colon. Our integrated analysis of transcriptional regulation maps the involvement of miRNA in IBD as well as other autoimmune diseases and colorectal cancer, while identifying therapeutic targets for the repositioning of existing therapeutics in the treatment of IBD. Unique biomarker sets were found to accurately differentiate between IBD phenotypes and the four Dukes stages of CRC.

The second generation microfluidic QLISA platform with enhanced antibody conjugation strategies using a diamino PEG crosslinker was found to improve signal to noise ratios and increase assay sensitivity while lowering the limit of detection for lactoferrin to 6.4ng/ml. The device was shown to detect lactoferrin within spike dhuman stool samples and demonstrates transitional clinical potential beyond the gastrointestinal biomarker space and is currently being licensed by Net Scientific.

5.1 CONTRIBUTIONS

Throughout my time at Drexel University I have had the opportunity to collaborate on a variety of research projects with many spectacularly talented colleagues and mentors. My work with Dr. Aydin Tozeren has led to the preparation of two manuscripts (chapters 2 & 3) as well as many other research ideas and ongoing collaborations. The manuscript presented in chapter 2 was submitted to the journal of Inflammatory Bowel Disease in February 2012 and is currently under review for publication. The third chapter is a manuscript draft which is intended for submission to a computational or clinical journal related to cancer. The final chapter is composed of work performed under Dr. Elisabth Papazoglou and Dr. Moses Noh has led to the development of two patents (PCT International Patent Application No.PCT/US11/36290 and U.S. Patent Application No.61/334,056) two manuscripts and a book chapter. I co-atuthored one manuscript with David Hansberry regarding the optimization of antibody conjugation methods onto PMMA substrates was submitted to the journal of Biomedical Microdevices and is currently under editorial review. The methods presented within the paper are also currently protected under provisional patent applications. The microfluidic immunoassay platform developed under Dr. Noh's guidance with integrated electrothermal mixing electrodes was prepared in collaboration with Chengjie Yu and was presented at MicroTAS 2011 in Seatlle Washington [195]. I am also a co-author of a book chapter within, "Quantum Dots: Applications, Synthesis and Characterization" entitled, "Quantum Dots for Bioimaging, Cellular Labeling, and Diagnostics," which is currently in press. I am also a coauthor on a paper written with Dr. John Fitzpatrick and Dr. Franco

Capaldi entitled, "Effect of Decellularization Protocol on the Mechanical Behavior of Porcine Descending Aorta," [196].

5.2 FUTURE WORK

Biomarkers in addition to lactoferrin identified though computational analysis for the stratification of colonic disease phenotypes must be validated on a clinically relevant cohort for validation of both the efficacy of patient classification using identified biomarker sets in terms of assay panel sensitivity and specificity and clinical performance of the multiplexed microfluidic immunoassay platform alongside standard ELISA methods. This research requires IBR approval, large sample sizes, extensive assay development and is outside of the scope of this thesis.

Ongoing bioinformatics projects related to this work involve studying the interaction of the intestinal microbiome to model host-pathogen itneractions and their involvement in driving the pathogenesis of IBD and CRC. Microbial species may be identified through 16S rRNA gene sequencing, while the host response may be measured through microarray analysis. Statistically enriched human receptors may be mapped to interacting proteins using the Human Protein Database (HPRD) and homologous proteins or enriched protein domains and conserved eukaryotic binding motifs may then be found within enriched bacterial species to identify unique host-pathogen interactions and better understand this complex phenomena and identify potential therapeutic pathways to limit the mucosal immune response.

The QLISA platform is currently licensed by Net Scientific and being transitioned into the cardiac biomarker market as a bedside diagnostic panel and associated device.. The assay has extended to a micro bead based immunoassay to further increase assay sensitivity and decrease the limit of detection, required

by cardiac immunoassays which need to detect picomolar concentrations of analytes within serum.

There are other means of performing micromixing, particularly related to the manipulation of microparticles within luquids including AC electrokinetic mixing and dielectrophoresis . These methods require extensive characterization within biological systems to determine their efficacy in the enhancement of immunoassay kinetics. Although these methods are not readily transferable to the PMMA substrate, other surface activation techniques may be applied to glass substrates and integrated with electrokinetic mixing components within a particle based assay and microfluidic liquid handling device to develop the next generation microfluidic QLISA platform for clinical use and validation. This work also requires extensive study of the complex interactions between electrical fields and antigen-antibody binding kinetics and affinity.

References

1. Molodecky, N.A., et al., *Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review.* Gastroenterology, 2012. **142**(1): p. 46-54 e42; quiz e30.
2. Loftus, E.V., *Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences.* Gastroenterology, 2004. **126**(6): p. 1504-1517.
3. Assche, G.V., *Fecal biomarkers for the diagnosis and management of inflammatory bowel disease.* Gastroenterol Hepatol (N Y), 2011. **7**(6): p. 396-8.
4. Cho, J.H., *The genetics and immunopathogenesis of inflammatory bowel disease.* Nature Reviews Immunology, 2008. **8**: p. 458-466.
5. Zhang, H., et al., *Genetics of inflammatory bowel disease: clues to pathogenesis.* British Medical Bulletin, 2008. **87**(1): p. 17-30.
6. Xavier, R.J. and D.K. Podolsky, *Unravelling the pathogenesis of inflammatory bowel disease.* Nature, 2007. **448**(7152): p. 427-434.
7. Gassler, N., et al., *Inflammatory bowel disease is associated with changes of enterocytic junctions.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2001. **281**(1): p. G216-G228.
8. Holden, W., T. Orchard, and P. Wordsworth, *Enteropathic arthritis.* Rheum Dis Clin North Am, 2003. **29**(3): p. 513-30, viii.
9. Rodriguez-Reyna, T.S., C. Martinez-Reyes, and J.K. Yamamoto-Furusho, *Rheumatic manifestations of inflammatory bowel disease.* World J Gastroenterol, 2009. **15**(44): p. 5517-24.
10. Ha, T.Y., *MicroRNAs in Human Diseases: From Autoimmune Diseases to Skin, Psychiatric and Neurodegenerative Diseases.* Immune Netw, 2011. **11**(5): p. 227-44.
11. Frank, D.N., et al., *Disease Phenotype and Genotype Are Associated with Shifts in Intestinal-associated Microbiota in Inflammatory Bowel Diseases.* Inflammatory Bowel Diseases, 2011. **17**(1): p. 179-184.
12. Fava, F. and S. Danese, *Intestinal microbiota in inflammatory bowel disease: Friend of foe?* World J Gastroenterol, 2011. **17**(5): p. 557-66.
13. Vanderploeg, R., et al., *Influences of Intestinal Bacteria in Human Inflammatory Bowel Disease.* Infectious Disease Clinics of North America, 2010. **24**(4): p. 977-+.
14. Sartor, R.B., *Genetics and Environmental Interactions Shape the Intestinal Microbiome to Promote Inflammatory Bowel Disease Versus Mucosal Homeostasis.* Gastroenterology, 2010. **139**(6): p. 1816-1819.
15. Friswell, M., B. Campbell, and J. Rhodes, *The Role of Bacteria in the Pathogenesis of Inflammatory Bowel Disease.* Gut and Liver, 2010. **4**(3): p. 295-306.
16. Chichlowski, M. and L.P. Hale, *Bacterial-mucosal interactions in inflammatory bowel disease-an alliance gone bad.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2008. **295**(6): p. G1139-G1149.

17. Gill, S.R., et al., *Metagenomic analysis of the human distal gut microbiome.* Science, 2006. **312**(5778): p. 1355-1359.
18. Mahida, Y.R. and V.E. Rolfe, *Host-bacterial interactions in inflammatory bowel disease.* Clinical Science, 2004. **107**(4): p. 331-341.
19. Goh, K.I., et al., *The human disease network.* Proc Natl Acad Sci U S A, 2007. **104**(21): p. 8685-90.
20. Xu, J. and Y. Li, *Discovering disease-genes by topological features in human protein-protein interaction network.* Bioinformatics, 2006. **22**(22): p. 2800-5.
21. Dudley, J.T. and A.J. Butte, *Identification of discriminating biomarkers for human disease using integrative network biology.* Pac Symp Biocomput, 2009: p. 27-38.
22. Zanzoni, A., M. Soler-Lopez, and P. Aloy, *A network medicine approach to human disease.* FEBS Lett, 2009. **583**(11): p. 1759-65.
23. Linghu, B., et al., *Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network.* Genome Biol, 2009. **10**(9): p. R91.
24. Barrenas, F., et al., *Network properties of complex human disease genes identified through genome-wide association studies.* Plos One, 2009. **4**(11): p. e8090.
25. Suthram, S., et al., *Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets.* PLoS Comput Biol, 2010. **6**(2): p. e1000662.
26. Zhang, S.H., et al., *From phenotype to gene: detecting disease-specific gene functional modules via a text-based human disease phenotype network construction.* FEBS Lett, 2010. **584**(16): p. 3635-43.
27. Zhang, Y., et al., *Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information.* BMC Med Genomics, 2010. **3**: p. 1.
28. Nguyen, T.P. and F. Jordan, *A quantitative approach to study indirect effects among disease proteins in the human protein interaction network.* BMC Syst Biol, 2010. **4**: p. 103.
29. Zhang, X., et al., *The expanded human disease network combining protein-protein interaction information.* Eur J Hum Genet, 2011. **19**(7): p. 783-8.
30. Hu, G. and P. Agarwal, *Human disease-drug network based on genomic expression profiles.* Plos One, 2009. **4**(8): p. e6536.
31. Dawany, N.B., W.N. Dampier, and A. Tozeren, *Large-scale integration of microarray data reveals genes and pathways common to multiple cancer types.* Int J Cancer, 2011. **128**(12): p. 2881-91.
32. Noble, C.L., et al., *Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis.* Gut, 2008. **57**(10): p. 1398-405.
33. Noble, C.L., et al., *Characterization of Intestinal Gene Expression Profiles in Crohn's Disease by Genome-wide Microarray Analysis.* Inflammatory Bowel Diseases, 2010. **16**(10): p. 1717-1728.
34. Barrett, T., et al., *NCBI GEO: mining tens of millions of expression profiles - database and tools update.* Nucleic Acids Research, 2007. **35**: p. D760-D765.

35. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Research, 2002. **30**(1): p. 207-210.
36. Shieh, A.D. and Y.S. Hung, *Detecting outlier samples in microarray data.* Stat Appl Genet Mol Biol, 2009. **8**(1): p. Article 13.
37. Olsen, J., et al., *Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data.* Inflamm Bowel Dis, 2009. **15**(7): p. 1032-8.
38. Carey, R., et al., *Activation of an IL-6:STAT3-dependent transcriptome in pediatric-onset inflammatory bowel disease.* Inflamm Bowel Dis, 2008. **14**(4): p. 446-57.
39. Ahrens, R., et al., *Intestinal macrophage/epithelial cell-derived CCL11/eotaxin-1 mediates eosinophil recruitment and function in pediatric ulcerative colitis.* J Immunol, 2008. **181**(10): p. 7390-9.
40. Kugathasan, S., et al., *Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease.* Nature Genetics, 2008. **40**(10): p. 1211-1215.
41. Bjerrum, J.T., et al., *Genome-wide gene expression analysis of mucosal colonic biopsies and isolated colonocytes suggests a continuous inflammatory state in the lamina propria of patients with quiescent ulcerative colitis.* Inflamm Bowel Dis, 2010. **16**(6): p. 999-1007.
42. Arijs, I., et al., *Mucosal Gene Expression of Antimicrobial Peptides in Inflammatory Bowel Disease Before and After First Infliximab Treatment.* Plos One, 2009. **4**(11).
43. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-264.
44. Irizarry, R.A., et al., *Summaries of affymetrix GeneChip probe level data.* Nucleic Acids Research, 2003. **31**(4).
45. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-193.
46. Dai, M.H., et al., *Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.* Nucleic Acids Research, 2005. **33**(20).
47. Sandberg, R. and O. Larsson, *Improved precision and accuracy for microarrays using updated probe set definitions.* Bmc Bioinformatics, 2007. **8**.
48. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
49. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(9): p. 5116-5121.
50. Hong, F., et al., *RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis.* Bioinformatics, 2006. **22**(22): p. 2825-7.

51. Breitling, R., et al., *Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments.* FEBS Lett, 2004. **573**(1-3): p. 83-92.
52. Dennis, G., et al., *DAVID: Database for annotation, visualization, and integrated discovery.* Genome Biology, 2003. **4**(9).
53. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nature Protocols, 2009. **4**(1): p. 44-57.
54. Alibes, A., et al., *IDconverter and IDClight: Conversion and annotation of gene and protein IDs.* Bmc Bioinformatics, 2007. **8**.
55. Diehn, M., et al., *SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data.* Nucleic Acids Research, 2003. **31**(1): p. 219-223.
56. Dawany, N.B. and A. Tozeren, *Asymmetric microarray data produces gene lists highly predictive of research literature on multiple cancer types.* Bmc Bioinformatics, 2010. **11**.
57. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Research, 2006. **34**: p. D354-D357.
58. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic Acids Res, 2011.
59. Wu, F., et al., *Identification of microRNAs associated with ileal and colonic Crohn's disease.* Inflamm Bowel Dis, 2010. **16**(10): p. 1729-38.
60. Wu, F., et al., *MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha.* Gastroenterology, 2008. **135**(5): p. 1624-1635 e24.
61. Pekow, J.R., et al., *miR-143 and miR-145 are downregulated in ulcerative colitis: putative regulators of inflammation and protooncogenes.* Inflamm Bowel Dis, 2012. **18**(1): p. 94-100.
62. Fasseu, M., et al., *Identification of restricted subsets of mature microRNA abnormally expressed in inactive colonic mucosa of patients with inflammatory bowel disease.* Plos One, 2010. **5**(10).
63. Takagi, T., et al., *Increased expression of microRNA in the inflamed colonic mucosa of patients with active ulcerative colitis.* J Gastroenterol Hepatol, 2010. **25 Suppl 1**: p. S129-33.
64. Dweep, H., et al., *miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes.* J Biomed Inform, 2011. **44**(5): p. 839-47.
65. Park, J.Y., et al., *Proteomic analysis of pancreatic juice for the identification of biomarkers of pancreatic cancer.* J Cancer Res Clin Oncol, 2011. **137**(8): p. 1229-38.
66. Planas, R., et al., *Regenerating gene Ialpha is a biomarker for diagnosis and monitoring of celiac disease: a preliminary study.* Transl Res, 2011. **158**(3): p. 140-5.
67. Astorri, E., et al., *Circulating Reg1alpha proteins and autoantibodies to Reg1alpha proteins as biomarkers of beta-cell regeneration and damage in type 1 diabetes.* Horm Metab Res, 2010. **42**(13): p. 955-60.

68. Usami, S., et al., *Regenerating gene I regulates interleukin-6 production in squamous esophageal cancer cells.* Biochem Biophys Res Commun, 2010. **392**(1): p. 4-8.
69. Ezzat, M., et al., *Elevated production of serum B-cell-attracting chemokine-1 (BCA-1/CXCL13) is correlated with childhood-onset lupus disease activity, severity, and renal involvement.* Lupus, 2011. **20**(8): p. 845-54.
70. El-Haibi, C.P., et al., *CXCL13 mediates prostate cancer cell proliferation through JNK signalling and invasion through ERK activation.* Cell Prolif, 2011. **44**(4): p. 311-9.
71. Meeuwisse, C.M., et al., *Identification of CXCL13 as a marker for rheumatoid arthritis outcome using an in silico model of the rheumatic joint.* Arthritis Rheum, 2011. **63**(5): p. 1265-73.
72. Chung-Faye, G., et al., *Fecal M2-pyruvate kinase (M2-PK): a novel marker of intestinal inflammation.* Inflamm Bowel Dis, 2007. **13**(11): p. 1374-8.
73. Joishy, M., et al., *Fecal Calprotectin and Lactoferrin as Noninvasive Markers of Pediatric Inflammatory Bowel Disease.* Journal of Pediatric Gastroenterology and Nutrition, 2009. **48**(1): p. 48-54.
74. D'Inca, R., et al., *Calprotectin and lactoferrin in the assessment of intestinal inflammation and organic disease.* International Journal of Colorectal Disease, 2007. **22**(4): p. 429-437.
75. Noss, E.H., et al., *Modulation of matrix metalloproteinase production by rheumatoid arthritis synovial fibroblasts after cadherin 11 engagement.* Arthritis Rheum, 2011. **63**(12): p. 3768-78.
76. Werlin, S.L., et al., *Evidence of intestinal inflammation in patients with cystic fibrosis.* J Pediatr Gastroenterol Nutr, 2010. **51**(3): p. 304-8.
77. Zhang, C., et al., *Dieckol from Ecklonia cava Regulates Invasion of Human Fibrosarcoma Cells and Modulates MMP-2 and MMP-9 Expression via NF-kappaB Pathway.* Evid Based Complement Alternat Med, 2011. **2011**: p. 140462.
78. Garamszegi, N., S.P. Garamszegi, and S.P. Scully, *Matrix metalloproteinase-1 contribution to sarcoma cell invasion.* J Cell Mol Med, 2011.
79. Ouban, A. and A.A. Ahmed, *Claudins in human cancer: a review.* Histol Histopathol, 2010. **25**(1): p. 83-90.
80. Salzman, N.H., et al., *Enteric defensins are essential regulators of intestinal microbial ecology.* Nat Immunol, 2010. **11**(1): p. 76-83.
81. Te Velde, A.A., et al., *Glutathione peroxidase 2 and aquaporin 8 as new markers for colonic inflammation in experimental colitis and inflammatory bowel diseases: an important role for H2O2?* Eur J Gastroenterol Hepatol, 2008. **20**(6): p. 555-60.
82. Schejbel, L., et al., *Molecular basis of hereditary C1q deficiency--revisited: identification of several novel disease-causing mutations.* Genes Immun, 2011. **12**(8): p. 626-34.
83. Sparano, J.A., et al., *Randomized phase III trial of marimastat versus placebo in patients with metastatic breast cancer who have responding or stable disease after first-line chemotherapy: Eastern Cooperative Oncology Group trial E2196.* J Clin Oncol, 2004. **22**(23): p. 4683-90.

84. Skov, L., et al., *IL-8 as antibody therapeutic target in inflammatory diseases: reduction of clinical activity in palmoplantar pustulosis.* J Immunol, 2008. **181**(1): p. 669-79.
85. Bocca, C., et al., *Celecoxib inactivates epithelial-mesenchymal transition stimulated by hypoxia and/or epidermal growth factor in colon cancer cells.* Mol Carcinog, 2011.
86. Shiraishi, T., S. Matsuyama, and H. Kitano, *Large-scale analysis of network bistability for human cancers.* PLoS Comput Biol, 2010. **6**(7): p. e1000851.
87. Burger, J.A., et al., *Potential of CXCR4 antagonists for the treatment of metastatic lung cancer.* Expert Rev Anticancer Ther, 2011. **11**(4): p. 621-30.
88. Kong, Y.W., et al., *The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene.* Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8866-71.
89. Wang, J., et al., *Polymorphisms of matrix metalloproteinases in myocardial infarction: a meta-analysis.* Heart, 2011. **97**(19): p. 1542-6.
90. Lennard-Jones, J.E. and J. Powell-Tuck, *Drug treatment of inflammatory bowel disease.* Clin Gastroenterol, 1979. **8**(1): p. 187-217.
91. El-Tawil, A.M., *Oestrogens and Crohn's disease: the missed link.* Andrologia, 2008. **40**(3): p. 141-5.
92. Kappelman, M.D., et al., *Association of paediatric inflammatory bowel disease with other immune-mediated diseases.* Arch Dis Child, 2011. **96**(11): p. 1042-6.
93. Willcocks, L.C., et al., *A defunctioning polymorphism in FCGR2B is associated with protection against malaria but susceptibility to systemic lupus erythematosus.* Proc Natl Acad Sci U S A, 2010. **107**(17): p. 7881-5.
94. Ververis, K. and T.C. Karagiannis, *Potential non-oncological applications of histone deacetylase inhibitors.* Am J Transl Res, 2011. **3**(5): p. 454-67.
95. Ben-Zvi, I., et al., *Hydroxychloroquine: From Malaria to Autoimmunity.* Clin Rev Allergy Immunol, 2011.
96. Wang, H., et al., *Platelet-activating factor and endotoxin activate CCAAT/enhancer binding protein in rat small intestine.* Br J Pharmacol, 2001. **133**(5): p. 713-21.
97. Shulzhenko, N., et al., *Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut.* Nat Med, 2011. **17**(12): p. 1585-93.
98. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
99. Triantafillidis, J.K., G. Nasioulas, and P.A. Kosmidis, *Colorectal Cancer and Inflammatory Bowel Disease: Epidemiology, Risk Factors, Mechanisms of Carcinogenesis and Prevention Strategies.* Anticancer Research, 2009. **29**(7): p. 2727-2737.
100. Eaden, J.A., K.R. Abrams, and J.F. Mayberry, *The risk of colorectal cancer in ulcerative colitis: a meta-analysis.* Gut, 2001. **48**(4): p. 526-35.
101. Canavan, C., K.R. Abrams, and J. Mayberry, *Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease.* Aliment Pharmacol Ther, 2006. **23**(8): p. 1097-104.

102. Smith, J.J., et al., *Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer.* Gastroenterology, 2010. **138**(3): p. 958-68.
103. Jorissen, R.N., et al., *Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer.* Clin Cancer Res, 2009. **15**(24): p. 7642-7651.
104. Galamb, O., et al., *Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor.* Br J Cancer, 2010. **102**(4): p. 765-73.
105. Galamb, O., et al., *Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(10): p. 2835-45.
106. Saverymuttu, S.H., et al., *in-111 autologous leukocyte scanning - comparison with radiology for imaging the colon in inflammatory bowel-disease.* British Medical Journal, 1982. **285**(6337): p. 255-257.
107. Saverymuttu, S.H., et al., *indium-111-granulocyte scanning in the assessment of disease extent and disease-activity in inflammatory bowel-disease - a comparison with colonoscopy, histology, and fecal indium-111-granulocyte excretion.* Gastroenterology, 1986. **90**(5): p. 1121-1128.
108. Kane, S.V., et al., *Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation.* American Journal of Gastroenterology, 2003. **98**(6): p. 1309-1314.
109. Sugi, K., et al., *Fecal lactoferrin as a marker for disease activity in inflammatory bowel disease: Comparison with other neutrophil-derived proteins.* American Journal of Gastroenterology, 1996. **91**(5): p. 927-934.
110. Roseth, A.G., et al., *Assessment Of The Neutrophil Dominating Protein Calprotectin In Feces - A Methodologic Study.* Scandinavian Journal of Gastroenterology, 1992. **27**(9): p. 793-798.
111. Langhorst, J., et al., *Noninvasive markers in the assessment of intestinal inflammation in inflammatory bowel diseases: Performance of fecal lactoferrin, calprotectin, and PMN-Elastase, CRP, and clinical indices.* American Journal of Gastroenterology, 2008. **103**(1): p. 162-169.
112. Quinton, J.F., et al., *Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role.* Gut, 1998. **42**(6): p. 788-791.
113. Schoepfer, A.M., et al. *Discriminating IBD from IBS: Comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies*. 2008. John Wiley & Sons Inc.
114. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
115. Hanley, J.A. and B.J. McNeil, *The Meaning And Use Of The Area Under A Receiver Operating Characteristic (ROC) Curve.* Radiology, 1982. **143**(1): p. 29-36.

116. Lasko, T.A., et al., *The use of receiver operating characteristic curves in biomedical informatics.* Journal of Biomedical Informatics, 2005. **38**(5): p. 404-415.
117. Smoot, M.E., et al., *Cytoscape 2.8: new features for data integration and network visualization.* Bioinformatics, 2011. **27**(3): p. 431-2.
118. Essaghir, A., et al., *Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data.* Nucleic Acids Res, 2010. **38**(11): p. e120.
119. Luo, X., et al., *MicroRNA signatures: novel biomarker for colorectal cancer?* Cancer Epidemiol Biomarkers Prev, 2011. **20**(7): p. 1272-86.
120. Pekow, J.R. and J.H. Kwon, *MicroRNAs in inflammatory bowel disease.* Inflamm Bowel Dis, 2012. **18**(1): p. 187-93.
121. Andersen, V., et al., *Polymorphisms in NF-kappaB, PXR, LXR, PPARgamma and risk of inflammatory bowel disease.* World J Gastroenterol, 2011. **17**(2): p. 197-206.
122. Glas, J., et al., *Role of PPARG gene variants in inflammatory bowel disease.* Inflamm Bowel Dis, 2011. **17**(4): p. 1057-8.
123. Babu, S., et al., *A PMMA microcapillary quantum dot linked immunosorbent assay (QLISA).* Biosensors & Bioelectronics, 2009. **24**(12): p. 3467-3474.
124. Butler, J.E., et al., *The immunochemistry of sandwich elisas--VI. Greater than 90% of monoclonal and 75% of polyclonal anti-fluorescyl capture antibodies (CAbs) are denatured by passive adsorption.* Molecular Immunology, 1993. **30**(13): p. 1165-1175.
125. Butler, J.E., et al., *The Physical and Functional-Behavior of Capture Antibodies Adsorbed on Polystyrene.* Journal of Immunological Methods, 1992. **150**(1-2): p. 77-90.
126. Christodoulides, N., et al., *A microchip-based multianalyte assay system for the assessment of cardiac risk.* Analytical Chemistry, 2002. **74**(13): p. 3030-3036.
127. Henares, T.G., et al., *Integration of multianalyte sensing functions on a capillary-assembled microchip: Simultaneous determination of ion concentrations and enzymatic activities by a "drop-and-sip" technique.* Analytical Chemistry, 2007. **79**(3): p. 908-915.
128. Kricka, L.J., *Prospects for microchips in immunoassays and immunoassay tests at the point-of-care.* Journal of Clinical Ligand Assay, 2002. **25**(4): p. 317-324.
129. Mendoza, L.G., et al., *High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA).* Biotechniques, 1999. **27**(4): p. 778-+.
130. Akkoyun, A. and U. Bilitewski, *Optimisation of glass surfaces for optical immunosensors.* Biosensors & Bioelectronics, 2002. **17**(8): p. 655-664.
131. Yuan, Y., H.Y. He, and J. Lee, *Protein A-Based Antibody Immobilization Onto Polymeric Microdevices for Enhanced Sensitivity of Enzyme-Linked Immunosorbent Assay.* Biotechnology and Bioengineering, 2009. **102**(3): p. 891-901.
132. Qi, S.Z., et al., *Microfluidic devices fabricated in poly(methyl methacrylate) using hot-embossing with integrated sampling capillary and fiber optics for fluorescence detection.* Lab on a Chip, 2002. **2**(2): p. 88-95.

133. Irawan, R., et al., *Integration of optical fiber light guide, fluorescence detection system, and multichannel disposable microfluidic chip.* Biomedical Microdevices, 2007. **9**(3): p. 413-419.
134. Charles, P.T., et al., *TNT displacement immunoassay with integrated microfluidic micromixer components.* Sensor Letters, 2008. **6**(3): p. 417-420.
135. Baldini, F., et al., *Optical PMMA chip suitable for multianalyte detection.* Ieee Sensors Journal, 2008. **8**(7-8): p. 1305-1309.
136. Baldini, F., et al., *A new procalcitonin optical immunosensor for POCT applications.* Analytical and Bioanalytical Chemistry, 2009. **393**(4): p. 1183-1190.
137. Park, H., et al., *Open sandwich FRET immunoassay of estrogen receptor beta in a PDMS microfluidic channel.* Bulletin of the Korean Chemical Society, 2008. **29**(7): p. 1297-1298.
138. Luo, C.X., et al., *PDMS microfludic device for optical detection of protein immunoassay using gold nanoparticles.* Lab on a Chip, 2005. **5**(7): p. 726-729.
139. Eteshola, E. and D. Leckband, *Development and characterization of an ELISA assay in PDMS microfluidic channels.* Sensors and Actuators B-Chemical, 2001. **72**(2): p. 129-133.
140. Morais, S., et al., *Multiplexed Microimmunoassays on a Digital Versatile Disk.* Analytical Chemistry, 2009. **81**(14): p. 5646-5654.
141. Kido, H., A. Maquieira, and B.D. Hammock, *Disc-based immunoassay microarrays.* Analytica Chimica Acta, 2000. **411**(1-2): p. 1-11.
142. Han, S.M., et al., *Plastic enzyme-linked immunosorbent assays (ELISA)-on-a-chip biosensor for botulinum neurotoxin A.* Analytica Chimica Acta, 2007. **587**(1): p. 1-8.
143. Raj, J., et al., *Surface immobilisation of antibody on cyclic olefin copolymer for sandwich immunoassay.* Biosensors & Bioelectronics, 2009. **24**(8): p. 2654-2658.
144. Jonsson, C., et al., *Silane-dextran chemistry on lateral flow polymer chips for immunoassays.* Lab on a Chip, 2008. **8**(7): p. 1191-1197.
145. Bai, Y.L., et al., *Surface modification for enhancing antibody binding on polymer-based microfluidic device for enzyme-linked immunosorbent assay.* Langmuir, 2006. **22**(22): p. 9458-9467.
146. Bai, Y.L., W.C. Huang, and S.T. Yang, *Enzyme-linked Immunosorbent assay of Escherichia coli O157 : H7 in surface enhanced Poly(Methyl methacrylate) microchannels.* Biotechnology and Bioengineering, 2007. **98**(2): p. 328-339.
147. Mastichiadis, C., et al., *Capillary-based immunoassays, immunosensors and DNA sensors - steps towards integration and multi-analysis.* Trac-Trends in Analytical Chemistry, 2008. **27**(9): p. 771-784.
148. Lee, M. and R.A. Durst, *Determination of imazethapyr using capillary column flow injection liposome immunoanalysis.* Journal of Agricultural and Food Chemistry, 1996. **44**(12): p. 4032-4036.
149. Kaneki, N., et al., *Electrochemical Enzyme-Immunoassay Using Sequential Saturation Technique in a 20-Mul Capillary - Digoxin as a Model Analyte.* Analytica Chimica Acta, 1994. **287**(3): p. 253-258.

150. Thomas, T., et al., *Development of a capillary biosensor for Helicobacter hepaticus.* Sensor Letters, 2007. **5**(2): p. 398-404.
151. Torabi, F., et al., *Development of a plasma panel test for detection of human myocardial proteins by capillary immunoassay.* Biosensors & Bioelectronics, 2007. **22**(7): p. 1218-1223.
152. Surugiu, L., et al., *Dextran-modified surface for highly sensitive chemiluminescent ELISA.* Analyst, 2001. **126**(10): p. 1633-1635.
153. Bange, A., H.B. Halsall, and W.R. Heineman, *Microfluidic immunosensor systems.* Biosens Bioelectron, 2005. **20**(12): p. 2488-503.
154. Yacoub-George, E., et al., *A miniaturized ISFET-ELISA system with a pretreated fused silica capillary as reaction cartridge.* Sensors and Actuators B: Chemical, 1996. **34**(1-3): p. 429-434.
155. Misiakos, K. and S.E. Kakabakos, *A multi-band capillary immunosensor.* Biosensors & Bioelectronics, 1998. **13**(7-8): p. 825-830.
156. Mastichiadis, C., et al., *Simultaneous determination of pesticides using a four-band disposable optical capillary immunosensor.* Analytical Chemistry, 2002. **74**(23): p. 6064-6072.
157. Petrou, P.S., et al., *Multi-analyte capillary immunosensor for the determination of hormones in human serum samples.* Biosensors & Bioelectronics, 2002. **17**(4): p. 261-268.
158. Rose, A., et al., *GDH biosensor based off-line capillary immunoassay for alkylphenols and their ethoxylates.* Biosensors & Bioelectronics, 2002. **17**(11-12): p. 1033-1043.
159. McCarley, R.L., et al., *Resist-free patterning of surface architectures in polymer-based microanalytical devices.* Journal of the American Chemical Society, 2005. **127**(3): p. 842-843.
160. Weigl, B.H. and O.S. Wolfbeis, *CAPILLARY OPTICAL SENSORS.* Analytical Chemistry, 1994. **66**(20): p. 3323-3327.
161. Ligler, F.S., et al., *Integrating waveguide biosensor.* Analytical Chemistry, 2002. **74**(3): p. 713-719.
162. Zhu, P.X., et al., *Detection of water-borne E-coli O157 using the integrating waveguide biosensor.* Biosensors & Bioelectronics, 2005. **21**(4): p. 678-683.
163. Stringer, R.C., D. Hoehn, and S.A. Grant, *Quantum dot-based biosensor for detection of human cardiac Troponin I using a liquid-core waveguide.* Ieee Sensors Journal, 2008. **8**(3-4): p. 295-300.
164. Danczyk, R., et al., *Comparison of antibody functionality using different immobilization methods.* Biotechnology and Bioengineering, 2003. **84**(2): p. 215-223.
165. Saphire, E.O., et al., *Contrasting IgG structures reveal extreme asymmetry and flexibility.* Journal of Molecular Biology, 2002. **319**(1): p. 9-18.
166. Kang, J.H., et al., *Improving immunobinding using oriented immobilization of an oxidized antibody.* Journal of Chromatography A, 2007. **1161**(1-2): p. 9-14.
167. Schramm, W., S.-H. Paek, and G. Voss, *Strategies for the Immobilization of Antibodies.* ImmunoMethods, 1993. **3**(2): p. 93-103.

168. Cho, I.H., et al., *Site-directed immobilization of antibody onto solid surfaces for the construction of immunochip.* Biotechnology and Bioprocess Engineering, 2004. **9**(2): p. 112-117.
169. Sok, D., et al., *Novel fluoroimmunoassay for ovarian cancer biomarker CA-125.* Analytical and Bioanalytical Chemistry, 2009. **393**(5): p. 1521-1523.
170. Staros, J.V., R.W. Wright, and D.M. Swingle, *Enhancement by N-Hydroxysulfosuccinimide of Water-Soluble Carbodiimide-Mediated Coupling Reactions.* Analytical Biochemistry, 1986. **156**(1): p. 220-222.
171. Bonroy, K., et al., *Comparison of random and oriented immobilisation of antibody fragments on mixed self-assembled monolayers.* Journal of Immunological Methods, 2006. **312**(1-2): p. 167-181.
172. Patel, S., et al., *Control of cell adhesion on poly(methyl methacrylate).* Biomaterials, 2006. **27**(14): p. 2890-2897.
173. Ting Cao, A.W., Xuemei Liang, Haiying Tang, Gregory W. Auner, Steven O. Salley, K.Y. Simon Ng,, *Investigation of spacer length effect on immobilized <I>Escherichia coli</I> pili-antibody molecular recognition by AFM.* Biotechnology and Bioengineering, 2007. **98**(6): p. 1109-1122.
174. Anderson, G.P., et al., *Effectiveness of protein A for antibody immobilization for a fiber optic biosensor.* Biosensors and Bioelectronics, 1997. **12**(4): p. 329-336.
175. Owaku, K., et al., *Protein A Langmuir-Blodgett Film for Antibody Immobilization and Its Use in Optical Immunosensing.* Analytical Chemistry, 1995. **67**(9): p. 1613-1616.
176. Palmer, D.A., et al., *Rapid Fluorescence Flow-Injection Immunoassay Using a Novel Perfusion Chromatographic Material.* Analyst, 1994. **119**(5): p. 943-947.
177. Ren, X.Z. and J.N. Miller, *Flow-Injection Fluoroimmunoassay for Human Transferrin Using a Protein-a Immunoreactor.* Analytical Letters, 1994. **27**(6): p. 1067-1074.
178. Jagannath, C. and S. Sehgal, *Enhancement of the antigen-binding capacity of incomplete IgG antibodies to Brucella melitensis through Fc region interactions with staphylococcal protein A.* J Immunol Methods, 1989. **124**(2): p. 251-7.
179. Chung, J.W., et al., *Immunosensor with a controlled orientation of antibodies by using NeutrAvidin-protein A complex at immunoaffinity layer.* J Biotechnol, 2006. **126**(3): p. 325-33.
180. Michalet, X., et al., *Quantum dots for live cells, in vivo imaging, and diagnostics.* Science, 2005. **307**(5709): p. 538-44.
181. Chan, W.C. and S. Nie, *Quantum dot bioconjugates for ultrasensitive nonisotopic detection.* Science, 1998. **281**(5385): p. 2016-8.
182. Bruchez, M., Jr., et al., *Semiconductor nanocrystals as fluorescent biological labels.* Science, 1998. **281**(5385): p. 2013-6.
183. Medintz, I.L., et al., *Quantum dot bioconjugates for imaging, labelling and sensing.* Nature Materials, 2005. **4**(6): p. 435-46.
184. Ghasemi, Y., P. Peymani, and S. Afifi, *Quantum dot: magic nanoparticle for imaging, detection and targeting.* Acta bio-medica : Atenei Parmensis, 2009. **80**(2): p. 156-65.

185. Klimov, V.I., et al., *Optical gain and stimulated emission in nanocrystal quantum dots.* Science, 2000. **290**(5490): p. 314-7.
186. Htoon, H., et al., *Highly emissive multiexcitons in steady-state photoluminescence of individual "giant" CdSe/CdS Core/Shell nanocrystals.* Nano letters, 2010. **10**(7): p. 2401-7.
187. Rogach, A.L., *Semiconductor nanocrystal quantum dots: synthesis, assembly, spectroscopy and applications*2008: Springer.
188. Klimov, V.I., *Nanocrystal Quantum Dots, Second Edition*2009: Taylor and Francis.
189. Sutherland, A.D., R.B. Gearry, and F.A. Frizelle, *Review of fecal biomarkers in inflammatory bowel disease.* Diseases of the Colon & Rectum, 2008. **51**(8): p. 1283-1291.
190. Tibbe, J.A. and I. Bjarnason, *Non-invasive investigation of inflammatory bowel disease.* World Journal of Gastroenterology, 2001. **7**(4): p. 460-465.
191. Nugen, S.R., P.J. Asiello, and A.J. Baeumner, *Design and fabrication of a microfluidic device for near-single cell mRNA isolation using a copper hot embossing master.* Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems, 2009. **15**(3): p. 477-483.
192. Hsu, Y.C. and T.Y. Chen, *Applying Taguchi methods for solvent-assisted PMMA bonding technique for static and dynamic mu-TAS devices.* Biomedical Microdevices, 2007. **9**(4): p. 513-522.
193. Hermanson, G.T., *Bioconjugate Techniques*. Second ed2008, New York: Academic Press.
194. Naruishi, N., et al., *Highly efficient dynamic modification of plastic microfluidic devices using proteins in microchip capillary electrophoresis.* Journal of Chromatography A, 2006. **1130**(2): p. 169-174.
195. Clark, P., et al., *MICROFABRICATED QLISA BIOSENSORS WITH AN EMBEDDED MIXING ELEMENT.* uTAS, 2011.
196. Fitzpatrick, J.C., P.M. Clark, and F.M. Capaldi, *Effect of decellularization protocol on the mechanical behavior of porcine descending aorta.* Int J Biomater, 2010. **2010**.

Appendix A: Principal Component Analysis m-file

```
function [cumvar, outliers,z] = PCA(data, k, IDcutoff)
%data is an expression matrix or probestructure as returned by the celintensityread function
%k is an integer multiplier of the standard deviaiton
%IDcutoff is an integer representing the  character length to be associated with each sample as
returned within ProbeStructure.CELNames

if isobject(data)
   CELNames = data.colnames';
   data = data.Matrix;
   n = 0.1*length(data.rownames);
elseif isstruct(data)
   CELNames = data.CELNames;
   data = data.PMIntensities;
   n = 0.1*length(data.ProbeSetIDs);
else CELNames = CELNames;
   data = data;
end
names = cellfun(@(x)(x(1:IDcutoff)), CELNames, 'uniformoutput', false);
[r, ~, J] = unique(names);                                  %J is sample a numeric sample vector

%Variance Filter and PCA
variance = var(data,[],2);
[~, inds] = sort(variance, 'descend');
[~, zscores, pcvars] = princomp(data(inds(1:n),:)');
cumvar = cumsum(pcvars./sum(pcvars) * 100);                 %cumulative sum of the
variances

%find the centroid of each diseae classification
figure
gscatter(zscores(:,1),zscores(:,2),names);
hold on
z = zeros(length(CELNames),1);
clear('n');
for n = 1:1:length(r)                                       %for each phenotype
   points = zscores(:,1:2);
   xmean = mean(zscores(J==n,1));
   ymean = mean(zscores(J==n,2));
   stdev_x = std(zscores(J==n,1));
   stdev_y = std(zscores(J==n,2));
   scatter(xmean,ymean,[],'black')                          %plot centroid
   [x y] = CalculateEllipse(xmean, ymean, k*stdev_x, k*stdev_y, 0,100);
   plot(x,y);
   for c = 1:1:size(points,1)                               %for all pca points
      x2 = points(c,1);
      y2 = points(c,2);
      xc = xmean;
      yc = ymean;
      t = 0;
      a = k*stdev_x;
      b = k*stdev_y;
      TF = strcmpi(names(c,1),r(n,1));
      if TF==1 && ((x2-xc)*cos(t)-(y2-yc)*sin(t)).^2/a^2 + ((x2-xc)*sin(t)+(y2-yc)*cos(t)).^2/b^2 >=
1
         z(c)=1;
      end
```

```
    end
end
outliers = CELNames(z==1,1);
outliers = sort(outliers);
xlabel('First Principal Component');
ylabel('Second Principal Component');
title('Principal Component Scatter Plot');
hold off
```

Appendix B: GAD Gene-Disease Interaction m-file

```
function [output2,output4]=GAD_genes(genelist)
%genelist is a cell array of gene symbols (all letters must be capitalized)
%run in folder containing raw Gene Association Database data

f1 = fopen('genelist.txt');
gad_genes = textscan(f1, '%s','HeaderLines',2,'Delimiter','\t','EndOfLine','\r');
fclose(f1);
gad_genes = gad_genes{1,1};

f2 = fopen('all.txt');
all = textscan(f2, '%*s %*s %*s %*s %*s %s %*s %*s %s
%*[^\r]','HeaderLines',3,'Delimiter','\t','EndOfLine','\r');
fclose(f2);

%Find GAD Annotated Genes within Genelist
tf = ismember(all{1,2},genelist);
output(:,1) = all{1,2}(tf);
output(:,2) = all{1,1}(tf);

for n=1:length(output)
   tf2(n) = isempty(output{n,2});
   tf2 = tf2';
end
output2 = output(tf2==0,:);

%Delimit Diseases
for n=1:size(output2,1)
   str = output2{n,2};
   C = textscan(str,'%s','Delimiter','|');
   out = repmat(output2(n,1),length(C{1,1}),1);
   out(:,2) = C{1,1};
   if n==1
      output3 = out;
   else
   output3 = vertcat(output3,out);
   end
end

%find unique interactions
unique_genes = unique(output3(:,1));
for n=1:length(unique_genes)
   tf = ismember(output3(:,1),unique_genes(n));
   unique_mesh = unique(output3(tf,2));
   out2 = repmat(unique_genes(n),length(unique_mesh),1);
   out2(:,2) = unique_mesh;
   if n==1
      output4 = out2;
   else
      output4 = vertcat(output4,out2);
   end
end
```

Appendix C: ROC Sample classification by logistic regression

```
function [AUC,x,y] = ROC_genes(Expression,probes,classifiers)

tf = ismember(probes,classifiers);
filtered_data = Expression.Matrix(tf,:);

names = cellfun(@(x)(x(1:3)), Expression.colnames, 'uniformoutput', false);
[~, ~, J] = unique(names);

IBD_data = filtered_data(:,J==3);
CRC_data = filtered_data(:,J==1);
HC_data = filtered_data(:,J==2);
data1 = IBD_data;
data2 = CRC_data;
data = [data1, data2];
ids = [zeros(1,size(data1,2)) ones(1,size(data2,2))];

b = glmfit(data', ids'==1, 'binomial');
guesses = glmval(b, data', 'logit');
[x, y, ~, AUC]=perfcurve(ids, guesses, 1);
plot(x,y)
xlabel('False positive rate'); ylabel('True positive rate')
title('ROC for classification by logistic regression')
```

APPENDIX D

Ulcerative Colitis					
Up-regulated			Down-regulated		
Gene Symbol	FC	p-value	Gene Symbol	FC	p-value
REG1A	50.7	2.23E-04	Cldn8	-22.5	3.97E-04
S100A8	35.2	3.14E-04	Ostalpha	-9.6	1.52E-05
REG1B	26.6	1.53E-05	CA1	-8.3	2.88E-04
MMP3	22.6	2.98E-04	GUCA2B	-7.9	7.89E-04
DEFA5	20.8	2.22E-04	MT1M	-7.5	4.69E-05
CHI3L1	19.1	4.52E-04	Adh1c	-6.8	4.95E-04
Mmp1	17.8	5.75E-05	GHR	-6.6	3.16E-04
Lcn2	15.5	9.05E-04	SLC16A9	-6.5	5.01E-04
Mmp12	13.1	3.93E-04	slc30a10	-5.9	4.86E-04
IDO1	12.1	3.96E-04	CNTN3	-5.6	5.91E-04
PI3	11.5	5.10E-04	Gba3	-5.4	1.66E-04
CXCL13	11.4	1.02E-05	ANPEP	-5.1	7.57E-06
ANXA1	11.2	6.82E-04	BEST4	-5.1	1.57E-04
IL8	11.1	1.18E-04	mep1b	-5.0	6.19E-07
FCGR3B	11.1	3.18E-04	Slc3a1	-4.9	8.38E-05
TNIP3	10.3	5.04E-04	pnliprp2	-4.9	9.09E-08
REG4	9.9	8.57E-04	vldlr	-4.8	9.07E-04
MMP10	9.6	5.99E-04	Npy1r	-4.8	7.51E-05
IGHV1-69	9.6	6.10E-04	CWH43	-4.8	3.38E-04
Cxcl6	9.3	1.05E-04	UGT1A7	-4.6	1.19E-04
UBD	9.1	8.10E-04	SLC16A1	-4.2	4.64E-04
mmp7	8.7	9.27E-04	HSD17B2	-4.2	5.57E-04
SERPINB5	8.6	2.47E-04	MT1G	-4.1	9.32E-05
IGKV4-1	8.3	1.49E-04	Prap1	-4.0	5.24E-05
IGLV3-1	7.6	1.41E-04	TRHDE	-4.0	1.01E-04
Spink4	6.9	4.20E-04	Pkib	-4.0	6.82E-04
C4orf7	6.8	1.14E-07	Cd177	-4.0	7.57E-05
LOC651536	6.8	5.33E-05	FLJ32063	-4.0	8.25E-04
ighv1-2	6.8	3.25E-04	EXPH5	-4.0	5.86E-04
TNC	6.4	1.51E-04	Ostbeta	-3.9	3.70E-04
CXCL11	6.4	1.26E-05	Ms4a12	-3.9	3.70E-06
DMBT1	6.4	6.41E-05	SLC17A4	-3.8	3.21E-04
C2CD4A	6.3	2.26E-04	ISX	-3.7	4.04E-05
IGLV3-25	6.3	7.52E-05	SELENBP1	-3.7	4.60E-04
IGHV3-9	6.2	9.45E-05	CA7	-3.7	2.40E-04
BCL2A1	6.2	3.68E-05	C10orf116	-3.7	5.25E-04
LY96	6.1	3.10E-04	PPARGC1A	-3.6	8.12E-04
IL1B	6.1	5.15E-04	MT1F	-3.5	1.28E-04

CTHRC1	6.1	9.70E-05	ANKRD43	-3.5	7.46E-04
GREM1	5.8	1.40E-05	ptprr	-3.5	2.92E-05
C4BPB	5.8	3.72E-04	cdkn2b	-3.4	7.83E-04
IGHV@	5.8	1.81E-04	FAM55D	-3.4	1.04E-04
IGLV2-14	5.8	9.76E-05	MT1X	-3.4	5.17E-05
DEFA6	5.8	1.97E-05	SULT1A2	-3.4	5.76E-04
IGLV3-19	5.8	6.90E-05	TRPM6	-3.4	2.83E-04
MNDA	5.8	5.53E-05	Mep1a	-3.4	6.85E-04
IGLV2-11	5.7	1.72E-04	PADI2	-3.3	9.42E-04
S100A9	5.7	1.77E-04	pitx2	-3.3	9.40E-06
SELL	5.6	1.91E-05	Entpd5	-3.3	8.48E-04
IGLL3	5.6	2.58E-05	DPP10	-3.3	5.84E-04
HLA-DQB1	5.5	7.22E-04	Hsd3b2	-3.3	3.83E-07
IGHV4-34	5.5	1.29E-04	PBLD	-3.3	1.71E-04
CXCL5	5.4	2.25E-06	DEFB1	-3.2	6.37E-05
OLFM4	5.4	1.51E-04	Slc38a4	-3.2	1.25E-06
srgn	5.3	1.73E-04	HEPACAM2	-3.2	2.20E-05
Col6a3	5.3	4.96E-04	NEURL1B	-3.2	3.80E-04
IL13RA2	5.2	6.02E-05	Nr1h4	-3.2	1.01E-06
Cfi	5.2	4.79E-05	BEST2	-3.2	2.29E-05
LOC100133678	5.1	3.46E-06	VIPR1	-3.1	4.28E-04
HLA-DRB4	5.0	8.10E-04	LOC100507594	-3.1	4.47E-04
IGHV3-11	5.0	1.08E-05	C2orf88	-3.1	3.62E-04
c13orf15	5.0	2.05E-04	C1qtnf3	-3.0	9.10E-04
POU2AF1	4.9	7.43E-06	METTL7B	-3.0	3.24E-04
Prok2	4.9	7.08E-05	ACSF2	-3.0	3.88E-04
PCSK1	4.8	3.42E-04	Paqr5	-3.0	6.74E-04
Aldob	4.8	2.93E-05	Mcoln2	-2.9	4.03E-05
CXCR2	4.8	2.39E-04	FAM5C	-2.9	2.01E-04
FOXQ1	4.8	2.18E-04	sult1a1	-2.9	5.03E-04
Ccl11	4.8	2.85E-04	TMEM56	-2.9	3.14E-04
Cxcl9	4.7	1.13E-06	SLC4A4	-2.9	1.95E-04
Col15a1	4.7	7.64E-05	TSPAN7	-2.8	7.08E-04
SPP1	4.7	1.35E-05	Slc23a1	-2.8	1.42E-04
SERPINB7	4.7	5.68E-07	CKB	-2.8	3.97E-06
hla-dpa1	4.7	2.85E-04	Mier3	-2.8	6.37E-04
MGC29506	4.6	2.14E-04	Pde6a	-2.8	5.21E-04
Mmp9	4.6	4.74E-05	ST6GAL2	-2.8	5.24E-08
COL1A2	4.6	2.12E-04	FLJ22763	-2.8	7.00E-05
serpina3	4.6	7.31E-05	C7orf10	-2.8	1.14E-04
S100P	4.5	4.46E-04	Tmed6	-2.8	5.74E-05
Ncf2	4.5	2.80E-04	abca8	-2.8	1.45E-05
C4BPA	4.5	3.82E-05	aifm3	-2.8	1.44E-04

CCL18	4.4	5.63E-05	aqp11	-2.7	2.95E-04
LYZ	4.4	8.38E-05	Ces2	-2.7	8.02E-04
CSF2RB	4.3	1.20E-05	PRR15	-2.7	6.05E-04
fam26f	4.3	3.98E-05	PIGZ	-2.7	3.79E-05
VNN2	4.3	1.01E-04	KCNJ2	-2.7	1.81E-04
PF4	4.2	1.66E-04	Tmem37	-2.7	6.77E-05
GBP1	4.2	8.71E-05	SCNN1B	-2.7	1.66E-07
tgfbi	4.2	1.86E-04	CDHR1	-2.7	1.53E-05
MS4A1	4.2	9.66E-08	MAOA	-2.7	6.42E-04
TRIM22	4.1	1.18E-04	SATB2	-2.7	3.56E-04
CD74	4.1	5.45E-04	ZNF575	-2.6	5.20E-04
FLJ40330	4.1	5.15E-05	HSD11B2	-2.6	1.53E-04
TNFRSF17	4.1	4.22E-07	C5orf35	-2.6	1.48E-04
nnmt	4.1	3.64E-05	SGK2	-2.6	1.62E-04
SAMSN1	4.1	4.92E-05	KIAA1737	-2.6	5.24E-04
CCL20	4.1	7.00E-05	ACOX1	-2.6	2.58E-04
MGP	4.1	1.20E-04	Tex11	-2.6	2.00E-04
CXCL10	4.0	4.38E-08	RAVER2	-2.6	4.95E-04
WNT5A	4.0	2.33E-05	FMO5	-2.6	4.95E-04
SELE	4.0	5.17E-05	Samd13	-2.6	1.29E-04
PLAU	4.0	3.15E-04	Tmem171	-2.5	2.37E-04
EVI2B	4.0	3.72E-05	Slc25a34	-2.5	3.65E-04
FPR1	4.0	1.06E-04	C10orf58	-2.5	1.49E-04
LOC91316	4.0	1.57E-06	IGSF9	-2.5	2.38E-04
psmb9	4.0	7.40E-04	WDR78	-2.5	1.93E-04
Ctsk	3.9	3.99E-04	CA2	-2.5	6.40E-04
CHST15	3.9	1.50E-04	PLCE1	-2.5	3.61E-04
PTGS2	3.9	8.93E-06	gxylt2	-2.5	4.12E-07
Basp1	3.9	2.37E-05	NAAA	-2.5	2.96E-04
CDH3	3.9	3.37E-04	MAGI1	-2.5	3.82E-04
LUM	3.9	2.73E-06	xrcc6bp1	-2.5	3.52E-04
Ifitm2	3.9	4.12E-04	RICH2	-2.5	2.45E-04
HLA-DMB	3.8	4.59E-04	ABAT	-2.5	9.27E-04
MXRA5	3.8	5.81E-05	Ephx2	-2.5	4.46E-04
Igkc	3.8	1.16E-04	HPGD	-2.5	6.00E-05
ROBO1	3.8	1.15E-04	ITPKA	-2.5	6.18E-04
Pde4b	3.8	9.58E-05	tmem38b	-2.4	6.52E-04
LTF	3.8	1.66E-07	retsat	-2.4	9.21E-04
AIM2	3.8	8.89E-05	phyH	-2.4	8.85E-04
HCLS1	3.8	2.94E-05	C7orf31	-2.4	8.95E-05
arhgdib	3.8	2.81E-04	LOC100128893	-2.4	1.18E-04
CR2	3.8	4.33E-07	LOC100505633	-2.4	1.04E-04
PECAM1	3.7	5.31E-04	STBD1	-2.4	1.47E-04

TMEM158	3.7	2.32E-04	MYO1D	-2.4	8.59E-04
CYTIP	3.7	5.10E-06	EFNA1	-2.4	3.46E-05
SERPINA1	3.7	6.88E-04	VAV3	-2.4	7.67E-05
LAMP3	3.7	4.44E-06	GIPC2	-2.4	1.03E-04
cstA	3.7	3.50E-05	ACVR1C	-2.4	5.63E-05
CCL19	3.7	8.76E-08	OSBPL1A	-2.4	3.58E-05
CXCR4	3.7	5.48E-06	PLEKHG6	-2.4	2.90E-04
AG2	3.7	1.78E-05	PPARG	-2.4	2.54E-05
pfkfb3	3.7	2.80E-04	SLC20A2	-2.4	5.48E-04
CASP1	3.7	6.00E-04	PDCD4	-2.4	1.80E-04
SDR16C5	3.7	1.20E-04	MTMR11	-2.3	5.00E-04
Col4a1	3.7	2.47E-04	C1orf210	-2.3	1.92E-04
AZGP1P1	3.6	1.56E-04	ANK3	-2.3	8.36E-05
Col5a2	3.6	6.56E-05	Scin	-2.3	4.67E-05
wars	3.6	1.89E-04	PDK4	-2.3	4.76E-06
cyr61	3.6	2.61E-05	FLJ35024	-2.3	2.68E-05
TNFSF13B	3.6	3.89E-05	Eya2	-2.3	5.66E-08
RGS2	3.5	1.17E-04	LRRC31	-2.3	1.59E-05
PLA2G7	3.5	4.32E-06	UGDH	-2.3	9.58E-04
Pla2g2a	3.5	1.22E-07	rab40b	-2.3	1.61E-04
LAPTM5	3.5	2.96E-05	FGFR2	-2.3	2.64E-05
dapp1	3.5	6.54E-04	fam82a1	-2.3	9.56E-04
IGLV6-57	3.5	5.56E-05	pAG1	-2.3	2.20E-04
cd38	3.5	1.71E-04	CNGA1	-2.3	3.25E-05
Rgs5	3.5	1.60E-04	AMN	-2.3	1.18E-04
GJA1	3.5	1.56E-06	gramd1c	-2.3	2.02E-05
KRT6B	3.4	3.41E-05	CLCN2	-2.3	1.19E-04
rab31	3.4	1.93E-04	SVOPL	-2.3	3.38E-08
LAX1	3.4	2.43E-04	Camk2n1	-2.3	5.27E-04
Rftn1	3.4	2.22E-05	Etfdh	-2.3	1.62E-04
clu	3.4	6.97E-06	scube2	-2.3	3.04E-05
tdo2	3.4	3.19E-07	PTPRO	-2.3	4.93E-07
LOC643733	3.4	3.18E-04	PTGDR	-2.2	4.56E-05
Kynu	3.4	9.27E-04	PANK1	-2.2	1.11E-05
EGR3	3.4	3.50E-05	CYP4F2	-2.2	2.78E-06
Igsf6	3.4	4.59E-05	tmcc3	-2.2	1.90E-04
Abca12	3.4	4.16E-04	Golt1a	-2.2	2.96E-04
Ccl2	3.3	6.32E-07	GLB1L2	-2.2	2.33E-04
TNFRSF6B	3.3	6.47E-04	fxyd3	-2.2	1.78E-04
TRIB2	3.3	2.14E-04	Matn2	-2.2	7.39E-05
Cd53	3.3	6.08E-06	CA4	-2.2	8.77E-08
ELTD1	3.3	1.42E-04	C2orf89	-2.2	8.47E-05
PECAM1	3.3	6.90E-04	PDZD3	-2.2	2.35E-04

COL3A1	3.3	2.51E-05	HOXA5	-2.2	1.13E-05
plek	3.3	5.01E-05	Pla2g12b	-2.2	5.09E-05
Vcam1	3.3	1.45E-07	ZDHHC23	-2.2	1.74E-04
FSTL1	3.3	1.86E-04	Nhej1	-2.2	2.72E-04
Ube2l6	3.3	6.77E-04	LOC100132815	-2.2	1.14E-04
CSRP2	3.3	2.30E-04	hsdl2	-2.2	5.97E-04
APOL1	3.3	2.78E-04	RAPGEFL1	-2.2	1.26E-04
F2R	3.3	7.98E-05	Acads	-2.2	3.13E-04
LOC646057	3.3	5.15E-05	SRI	-2.2	1.91E-04
CD44	3.3	5.56E-04	AGMAT	-2.2	1.13E-04
LOC653879	3.3	1.62E-05	ZNF57	-2.2	4.79E-04
msn	3.3	1.43E-04	ACVR2A	-2.2	5.59E-04
Evi2a	3.2	1.22E-06	fam101a	-2.2	2.39E-04
IL7R	3.2	1.62E-05	ca12	-2.2	1.33E-04
Nptx2	3.2	1.13E-06	bsg	-2.2	1.41E-05
TMEM163	3.2	4.62E-05	KIAA1239	-2.2	7.00E-06
FCER1G	3.2	2.24E-05	ap3s2	-2.2	9.94E-04
CD93	3.2	1.30E-04	LEAP2	-2.2	4.99E-05
ptgds	3.2	6.67E-06	SGK223	-2.1	4.54E-04
GNA15	3.2	8.26E-04	Edn3	-2.1	2.25E-04
Egr2	3.2	2.68E-05	XK	-2.1	1.55E-04
BIRC3	3.2	7.98E-05	CYCS	-2.1	6.35E-05
AQP9	3.2	5.81E-06	thrB	-2.1	2.68E-05
COL1A1	3.2	1.54E-04	Cited2	-2.1	1.65E-04
C1s	3.2	3.67E-05	CDX2	-2.1	5.82E-05
ART3	3.2	2.07E-04	a1cf	-2.1	3.25E-06
TMEM45A	3.2	1.92E-05	Adh1c	-2.1	2.75E-04
HLA-DPB1	3.1	2.87E-06	FABP1	-2.1	4.25E-04
gem	3.1	2.35E-04	Ccnjl	-2.1	5.96E-04
loxl1	3.1	2.17E-04	gsta1	-2.1	1.23E-07
FCRL5	3.1	1.79E-04	fam189a1	-2.1	6.38E-04
Btn3a3	3.1	3.75E-04	P2RY1	-2.1	6.59E-04
Cd52	3.1	4.57E-06	GCNT2	-2.1	4.83E-06

APPENDIX E

Crohn's Disease					
Up-regulated			Down-regulated		
Gene Symbol	FC	p-value	Gene Symbol	FC	p-value
REG1A	46.3	3.52E-04	AQP8	-12.0	7.89E-04
REG1B	40.3	9.94E-05	Cldn8	-9.0	7.99E-07
S100A8	26.5	2.53E-04	Ostalpha	-7.5	4.22E-08
MMP3	19.1	2.25E-04	PCK1	-6.9	1.13E-04
Mmp1	16.6	2.88E-05	SLC16A9	-6.0	9.69E-04
CHI3L1	15.7	6.15E-04	UGT2A3	-5.8	1.41E-05
IDO1	12.4	3.91E-04	GUCA2B	-5.3	3.34E-04
Duox2	12.1	2.88E-04	LOC100506782	-5.2	4.36E-04
CXCL13	10.8	1.61E-06	MT1M	-4.9	7.64E-08
FCGR3B	10.2	2.10E-04	GUCA2A	-4.8	5.27E-05
Mmp12	10.1	1.49E-04	slc30a10	-4.5	2.19E-04
DEFA6	9.8	3.30E-04	vldlr	-4.5	9.68E-04
IL8	9.1	3.36E-05	HSD17B2	-4.2	8.97E-04
Cxcl6	8.5	3.70E-05	ANPEP	-4.1	4.36E-08
BCL2A1	7.1	3.72E-05	LAMA1	-3.9	3.81E-04
IGHV3-11	6.6	1.06E-04	Prap1	-3.9	8.18E-05
IL1B	6.5	5.32E-04	GHR	-3.9	3.79E-06
Cxcl9	6.5	3.50E-07	Slc3a1	-3.8	1.49E-05
CXCL11	6.3	3.16E-07	mep1b	-3.8	2.13E-06
MMP10	6.3	3.64E-04	ptprr	-3.8	6.78E-05
Duoxa2	6.2	8.05E-04	SLC16A1	-3.7	3.72E-04
LOC651536	6.2	1.33E-05	Slc26a2	-3.7	2.29E-08
REG4	5.9	4.74E-05	chp2	-3.5	3.28E-05
PI3	5.9	2.80E-05	APOBEC3B	-3.4	5.01E-04
C4orf7	5.7	2.68E-07	pitx2	-3.4	1.10E-05
DMBT1	5.7	1.17E-05	MT1X	-3.3	3.20E-05
IGKV4-1	5.7	2.62E-05	Ms4a12	-3.3	1.33E-07
TNIP3	5.7	1.70E-04	abca8	-3.2	2.23E-04
MNDA	5.6	4.96E-06	TUBAL3	-3.2	2.86E-04
LOC100133678	5.6	9.99E-06	Cd177	-3.1	4.22E-08
TIMP1	5.5	9.99E-04	pnliprp2	-3.0	7.02E-05
fam26f	5.5	6.52E-05	PHLPP2	-3.0	8.14E-04
CTHRC1	5.4	1.46E-05	UGT1A7	-3.0	9.47E-08
CCL18	5.4	1.18E-04	RUNDC3B	-3.0	6.91E-04
CXCL5	5.3	1.01E-06	MT1G	-2.9	7.10E-08
LY96	5.3	1.56E-04	Ostbeta	-2.9	9.96E-06
ANXA1	5.3	3.51E-05	BEST4	-2.8	1.27E-07
SELL	5.2	2.59E-06	Hsd3b2	-2.8	1.82E-05

HLA-DQB1	5.2	3.37E-04	LOC100507594	-2.8	5.82E-04
Spink4	5.2	9.68E-05	Mep1a	-2.8	1.24E-04
CXCL10	5.1	8.58E-08	SLC17A4	-2.8	4.47E-05
PCSK1	5.0	6.25E-04	C10orf116	-2.8	1.60E-04
CSF2RB	4.9	1.39E-05	cdkn2b	-2.8	7.59E-04
C2CD4A	4.8	1.58E-04	Npy1r	-2.8	2.81E-06
Mmp9	4.8	4.22E-05	MT1F	-2.8	1.21E-06
SPP1	4.7	6.43E-06	Adh1c	-2.8	1.32E-07
TRIM22	4.7	4.97E-04	Slc38a4	-2.7	8.21E-07
hla-dpa1	4.7	3.22E-04	DHRS11	-2.7	3.44E-05
mmp7	4.6	5.56E-05	CA7	-2.7	7.67E-06
srgn	4.6	3.69E-05	TRHDE	-2.6	2.98E-07
LAMP3	4.6	1.48E-05	ST6GAL2	-2.6	2.97E-07
Ncf2	4.6	1.43E-04	TRPM6	-2.6	4.83E-05
HLA-DRB4	4.6	7.23E-04	SLC1A1	-2.6	1.13E-05
OLFM4	4.6	7.26E-05	aqp11	-2.6	4.75E-04
GBP1	4.5	1.23E-04	PADI2	-2.6	6.72E-04
TNC	4.4	4.84E-06	LOC644246	-2.6	2.78E-05
GREM1	4.4	2.92E-07	SULT1A2	-2.6	1.99E-04
IGHV1-69	4.4	2.43E-05	ANKRD43	-2.6	5.73E-05
SERPINB5	4.3	3.92E-08	C1qtnf3	-2.5	6.81E-04
MS4A1	4.3	7.97E-08	DEFB1	-2.5	7.41E-08
Cd55	4.2	8.70E-04	Pkib	-2.5	6.07E-09
TNFSF13B	4.2	1.12E-04	CNTN3	-2.5	2.06E-06
CCL20	4.2	2.35E-04	Entpd5	-2.5	9.71E-05
S100A9	4.2	8.85E-07	CA1	-2.5	4.68E-05
Col15a1	4.2	3.82E-05	DPP10	-2.4	1.11E-05
SAMSN1	4.2	2.60E-05	EFNA1	-2.4	2.53E-05
LYZ	4.1	2.77E-05	PDK4	-2.4	9.69E-06
IGLV2-14	4.1	3.89E-07	sult1a1	-2.4	1.70E-04
HLA-DMB	4.1	9.55E-04	STBD1	-2.4	2.11E-04
TCN1	4.1	6.74E-05	ACSF2	-2.3	8.80E-05
tdo2	4.1	3.32E-07	Gba3	-2.3	9.18E-05
COL1A2	4.1	1.26E-05	Mier3	-2.3	1.62E-04
cstA	4.1	1.03E-04	lrrc19	-2.3	2.18E-05
Col6a3	3.9	3.86E-05	FLJ35024	-2.3	1.97E-05
Cfi	3.9	5.69E-06	Paqr5	-2.3	3.79E-05
S100P	3.9	2.65E-04	VIPR1	-2.3	1.39E-05
IGHV3-23	3.8	1.95E-05	FLJ32063	-2.3	4.13E-08
FPR1	3.8	9.39E-06	FLJ22763	-2.3	1.42E-07
nnmt	3.8	1.30E-06	MAGI1	-2.3	5.14E-04
CXCR2	3.8	1.48E-06	Tmem37	-2.3	1.01E-06
ROBO1	3.8	1.73E-04	Ces2	-2.3	3.45E-04

Ccl11	3.8	5.55E-05	EXPH5	-2.3	6.23E-08
PLA2G7	3.8	3.72E-08	Cyp2s1	-2.2	8.52E-05
Basp1	3.8	5.68E-06	KIAA1737	-2.2	2.36E-04
Igsf6	3.7	3.41E-05	CWH43	-2.2	7.51E-06
c13orf15	3.7	1.54E-05	gsta1	-2.2	1.42E-07
C4BPB	3.7	7.07E-05	PPARGC1A	-2.2	2.23E-08
plek	3.7	2.66E-05	BEST2	-2.2	6.64E-07
Pde4b	3.7	2.60E-05	Mcoln2	-2.2	5.34E-07
Pla2g2a	3.7	2.99E-07	ITPKA	-2.2	3.77E-04
Ifitm2	3.6	2.65E-04	CKB	-2.2	2.93E-06
psmb9	3.6	6.71E-04	PRR15	-2.2	4.83E-05
Ccl2	3.6	2.32E-08	METTL7B	-2.2	1.43E-06
POU2AF1	3.6	2.89E-07	ACOX1	-2.2	4.09E-05
Vcam1	3.6	2.75E-08	ISX	-2.2	3.70E-05
IGLL3	3.6	5.32E-07	NEURL1B	-2.2	3.01E-06
ighv1-2	3.6	7.06E-06	chrna1	-2.1	4.02E-04
Ctsk	3.6	1.33E-04	PIGZ	-2.1	2.89E-07
CCL19	3.6	2.92E-06	bsg	-2.1	8.75E-07
AIM2	3.6	9.07E-05	LOC100288092	-2.1	3.16E-06
VNN2	3.5	1.18E-07	Pde6a	-2.1	2.57E-05
CR2	3.5	2.79E-06	EDN1	-2.1	2.83E-04
WNT5A	3.5	2.85E-08	PDZD3	-2.1	2.90E-04
Vnn1	3.5	5.52E-08	Scin	-2.1	1.13E-07
IL7R	3.5	6.93E-06	Epdr1	-2.1	3.02E-08
Cd53	3.5	9.95E-07	RAVER2	-2.1	2.04E-05
HCLS1	3.4	8.50E-07	Tmem171	-2.1	7.34E-06
CHST15	3.4	2.18E-05	OSBPL1A	-2.1	6.52E-06
AQP9	3.4	8.08E-08	CDHR1	-2.1	3.26E-07
CASP1	3.4	6.41E-04	aifm3	-2.1	4.50E-08
lcp2	3.4	6.84E-07	phyH	-2.1	2.23E-04
FLJ40330	3.4	1.18E-06	HIGD1D	-2.1	3.96E-04
PLAU	3.4	5.39E-05	FMO5	-2.1	1.14E-04
IGLV3-1	3.4	9.21E-07	PBLD	-2.1	6.93E-07
Elovl5	3.3	1.22E-04	LOC100505633	-2.1	3.25E-07
FKBP11	3.3	9.86E-04	SELENBP1	-2.1	1.51E-06
LUM	3.3	5.11E-07	SEMA5A	-2.1	2.48E-04
wars	3.3	5.11E-05	FAM5C	-2.1	4.76E-07
FCER1G	3.3	1.90E-06	tmem38b	-2.1	7.99E-05
TNFAIP6	3.3	3.56E-05	WSCD1	-2.1	5.44E-04
LOC643733	3.3	2.99E-04	SGK2	-2.1	1.25E-06
CXCR4	3.3	9.61E-08	HHLA2	-2.1	3.01E-08
IGLV2-11	3.3	1.71E-08	Matn2	-2.0	1.10E-04
rab31	3.3	1.18E-04	C7orf10	-2.0	2.33E-07

Evi2a	3.2	5.97E-08	KIAA1239	-2.0	3.70E-08
SELE	3.2	2.82E-08	emp1	-2.0	7.80E-06
FOXQ1	3.2	2.38E-06	Snx30	-2.0	1.94E-04
gbp5	3.2	5.08E-08	Slc23a1	-2.0	2.39E-07
dapp1	3.2	9.15E-04	Pla2g12b	-2.0	9.45E-06
ART3	3.2	5.61E-04	C1orf210	-2.0	5.45E-06
IFI44	3.2	6.41E-05	C2orf88	-2.0	5.07E-08
Ifit3	3.2	4.90E-08	IGSF9	-2.0	2.41E-06
Ube2l6	3.2	8.40E-04	TMEM56	-2.0	8.29E-09
Col5a2	3.2	3.08E-08	GLB1L2	-2.0	1.61E-05
tgfbi	3.2	3.49E-06	Slc25a34	-2.0	1.78E-05
Rftn1	3.2	2.52E-06	tmcc3	-2.0	1.18E-04
micB	3.2	3.90E-04	SLC20A2	-2.0	2.12E-04
CYTIP	3.1	9.63E-08	Cited2	-2.0	7.89E-05
Btn3a3	3.1	6.86E-04	C10orf58	-2.0	4.58E-08
PECAM1	3.1	1.33E-04	SLC39A5	-2.0	2.81E-04
Col4a1	3.1	1.74E-05	PDCD4	-2.0	7.11E-05
ptgds	3.1	3.88E-08	Golt1a	-2.0	7.59E-05
PDZK1IP1	3.1	2.16E-05	SVOPL	-2.0	1.41E-05
Prok2	3.1	2.78E-06	PLEKHG6	-2.0	2.55E-05
LAPTM5	3.1	1.04E-08	C5orf35	-2.0	6.36E-08
CD163	3.1	2.49E-07	smpdl3a	-2.0	1.16E-04
TNFRSF17	3.1	2.86E-06	PPP1R3B	-2.0	5.22E-04
LOC646057	3.1	3.09E-05	gxylt2	-2.0	1.54E-05
ELTD1	3.1	3.77E-05	MYO1D	-2.0	9.63E-05
C1s	3.1	1.92E-06	PXMP2	-2.0	5.35E-04
TMEM158	3.1	9.18E-06	CNGA1	-2.0	4.16E-08
MXRA5	3.1	2.71E-08	FABP1	-2.0	2.99E-07
arhgdib	3.1	5.93E-05	Samd13	-2.0	6.57E-08
FSTL1	3.0	2.40E-05	WDR78	-2.0	2.46E-08
DRAM1	3.0	1.23E-04	SLC4A4	-2.0	1.71E-07
ENPP2	3.0	3.95E-05	appl2	-2.0	6.03E-06
Ptprc	3.0	1.15E-07	CA4	-1.9	8.83E-05
HLA-DPB1	3.0	7.93E-08	GIPC2	-1.9	3.00E-07
PTGS2	3.0	9.12E-07	FAM55D	-1.9	2.77E-06
EVI2B	3.0	2.20E-07	PRKG2	-1.9	1.03E-04
CCL4	3.0	6.83E-08	fam189a1	-1.9	4.45E-04
serpina3	3.0	9.23E-07	HPGD	-1.9	1.00E-07
Ifitm1	3.0	4.77E-04	MTMR11	-1.9	6.77E-05
IL13RA2	3.0	8.49E-08	VAV3	-1.9	6.20E-08
Rgs5	2.9	1.97E-05	C1orf115	-1.9	1.04E-04
SERPING1	2.9	4.84E-05	RNF128	-1.9	1.23E-06
LILRB2	2.9	6.44E-08	Etfdh	-1.9	1.02E-05

Egr2	2.9	2.65E-06	UGDH	-1.9	4.99E-05
MGP	2.9	8.24E-08	NAAA	-1.9	6.42E-05
pfkfb3	2.9	6.91E-05	GCOM1	-1.9	5.12E-04
IGHV3-9	2.9	2.28E-06	PLCE1	-1.9	1.07E-07
P2RY13	2.9	1.75E-06	CYP2B6	-1.9	6.67E-07
cd38	2.9	9.00E-05	GATA6	-1.9	5.11E-05
ITGB2	2.9	1.49E-05	Tmed6	-1.9	6.69E-06
Egfl6	2.9	9.92E-07	LOC100133772	-1.9	1.02E-04
PF4	2.9	2.39E-05	ABAT	-1.9	1.14E-04
CD93	2.9	1.58E-05	sh3bgrl2	-1.9	1.65E-04
Cd48	2.9	5.92E-08	ANK3	-1.9	2.64E-08
CLIC2	2.9	3.63E-05	TSPAN7	-1.9	8.11E-06
CPA3	2.9	5.40E-07	MAOA	-1.9	1.68E-06
LOC653879	2.9	4.15E-08	CLCN2	-1.9	5.17E-07
LCP1	2.9	5.38E-08	CYP4F2	-1.9	9.36E-07
AG2	2.9	2.93E-07	fxyd3	-1.9	1.14E-06
Cd52	2.9	7.86E-08	HSPB3	-1.9	2.26E-06
APOL1	2.9	1.53E-04	arrdc4	-1.9	4.96E-05
ifi30	2.9	1.21E-04	CNNM2	-1.9	1.32E-04
COL3A1	2.8	3.10E-07	cobL	-1.9	1.55E-04
COL1A1	2.8	1.19E-06	Cpt1a	-1.9	9.89E-04
Mafb	2.8	2.10E-06	ACVR2A	-1.9	2.61E-04
Rhoh	2.8	4.63E-07	Camk2n1	-1.9	7.26E-05
TMEM45A	2.8	1.21E-06	Ephx2	-1.9	2.53E-06
SNX10	2.8	4.69E-09	CDHR5	-1.9	8.54E-08
LOC91316	2.8	1.54E-05	Ppp1r14a	-1.9	3.22E-06
CSRP2	2.8	2.37E-05	pAG1	-1.9	5.04E-06
LIPG	2.8	2.80E-04	SLC9A3R1	-1.8	3.28E-05
SAMD9L	2.8	2.40E-04	Prdx6	-1.8	4.60E-04
dse	2.8	1.72E-07	LOC100506261	-1.8	6.37E-05
IGLV3-19	2.8	4.61E-06	AHCYL2	-1.8	3.64E-07
MGC29506	2.8	3.56E-08	FCGRT	-1.8	3.29E-04
Rilpl2	2.8	6.07E-04	NR5A2	-1.8	4.59E-05

APPENDIX F

UPREGULATED CRC GENES				
Gene Symbol	**Dukes A FC**	**Dukes B FC**	**Dukes C FC**	**Dukes D FC**
CTHRC1	11.7	17.3	19.8	16.4
FOXQ1	15.0	15.3	15.1	16.1
IL8	13.0	18.1	18.0	12.1
THBS2	9.1		16.0	14.4
SPP1	5.6	13.6	15.1	18.1
MMP7	8.2	10.6	11.7	10.1
CLDN2	9.5	8.3	7.2	8.8
SFRP2	5.1	8.2	9.9	10.2
CDH3	7.2		7.4	7.1
TIMP1	5.7	7.3	7.7	7.7
KRT23	6.8	7.2	5.9	8.5
GDF15	7.4	6.8	6.2	6.8
SULF1	4.8	7.1	7.7	7.2
COL11A1	4.4		8.4	7.1
CLDN1	6.2		6.9	6.7
H19	5.1	5.9	7.4	7.1
COL1A1	5.3		6.9	6.0
COL12A1	5.3		6.7	6.0
CXCL2	8.5	6.2	5.3	3.9
TCN1	6.5	6.9	4.8	5.6
AZGP1P1	5.6	5.9	5.6	6.3
CXCL1	7.5	6.1	5.9	3.8
SNORA71A	5.4	6.2	5.9	5.5
CYR61	4.7	5.7	6.0	6.2
MMP1	4.7	6.7	7.0	4.0
CXCL3	7.6	5.6	5.4	3.5
MFAP2	4.6		6.0	5.9
COL10A1	3.2	5.4	7.3	6.0
VCAN	4.0	5.1	5.7	5.3
FAP	3.8		6.1	5.1
SFRP4	3.9	4.6	6.0	5.6
FLJ44450	5.0	5.3	4.9	4.8
SLC7A5	4.6	5.0	5.2	5.0
COL1A2	3.9	5.1	5.5	4.9
MGP	3.4	4.4	5.2	6.1
TESC	4.3	4.7	5.1	5.0
S100A8	4.0	5.7	5.3	4.0
C2ORF82	4.8	4.8	4.8	4.5
C2CD4A	6.3	4.7	4.0	3.8

COL5A2	3.8	4.9	5.3	4.7
FAM165B	4.5	4.7	4.8	4.3
LCN2	6.5	4.3	4.0	3.3
CHI3L1	4.9	5.0	4.7	3.3
C8ORF62	4.6	4.3	4.5	4.5
DPEP1	4.4	4.5	3.8	4.8
F2R	4.7	4.0	4.4	4.5
RP9P	4.2		4.5	4.4
EPHX4	4.3	4.3	4.4	4.3
NCRNA00164	4.4	4.9	4.1	3.8
TACSTD2	3.3	4.1	4.6	5.2
LY6G6D	4.5	4.0	4.4	4.2
TGFBI	4.0	4.3	4.0	4.7
CXCL11	4.9	4.7	4.3	3.0
MMP3	4.1	4.7	4.8	3.1
PHLDA1	4.0	4.4	3.8	4.2
UBD	4.7	4.1	4.2	3.3
LGR5	4.3	3.6	3.7	4.5
FXYD5	3.8	3.9	4.2	3.7
WNT2	4.0		3.9	3.6
KIAA1199	3.8	4.0	3.4	4.1
ANLN	3.9	3.9	4.0	3.5
FOSL1	3.5	4.0	4.2	3.5
REG1A	4.2	4.3	3.1	3.6
FNDC1	2.7	3.7	4.4	4.3
COL6A3	3.2	3.9	4.2	3.8
COL15A1	3.1	4.0	4.2	3.7
CST1	3.5	3.6	4.2	3.7
GREM1	2.6	3.9	4.7	3.7
LUM	3.4	3.7	4.2	3.6
CRNDE	3.8	4.0	3.4	3.7
CDH11	3.3	3.7	4.0	3.9
OLFML2B	2.9	3.8	4.1	4.0
RPL18AP3	3.8	3.6	3.7	3.7
C13ORF37	3.5	3.7	3.9	3.7
INHBA	2.8	3.6	4.2	3.8
TTYH3	3.7		3.6	3.5
ASPN	2.2	3.6	4.1	4.3
S100A2	3.0	3.7	4.2	3.5
PPBP	3.0	3.5	4.0	3.7
UBE2C	3.6		3.6	3.5
NKAIN3	3.6	3.6	3.5	3.4
ESM1	3.4	3.1	3.7	3.8

VSNL1	3.4	4.2	3.3	3.1
COL4A1	3.2	3.5	3.7	3.6
DUSP4	3.4	3.9	3.8	2.9
FOS	3.3	3.7	3.3	3.5
MAMSTR	3.5	3.4	3.4	3.4
SPON2	2.7	3.6	3.7	3.6
RHOB	3.3	3.5	3.3	3.6
SLC35D3	3.5	2.9	3.7	3.5
ETV4	3.3		3.2	3.6
TPM4	3.3	3.4	3.6	3.1
EREG	2.5	3.4	3.8	3.7
MAD2L1	3.5	3.4	3.3	3.2
LOC652797	3.1	3.5	3.5	3.3
DUOX2	3.3			
GAL	2.9	3.3	3.7	3.3
PLS3	3.1	3.4	3.3	3.4
IFITM1	3.6	3.4	3.3	2.8
SERPINB5	2.4	3.6	4.3	2.8
FAM182B	3.0	3.4	3.2	3.4
LOC100505478	3.2	3.4	3.3	3.1
ASCL2	3.5	2.8	2.9	3.8
TIMP3	2.5	3.3	3.6	3.6
C13ORF33	3.0	3.3	3.4	3.2
S100A11	3.0	3.3	3.3	3.3
ZAK	3.2	3.2	3.0	3.5
CTGF	2.8	3.1	3.5	3.5
TAF1L	3.2	3.3	3.2	3.1
DUSP27	4.1	2.8	2.6	3.3
MMP2	2.8	3.3	3.6	3.2
APCDD1	4.0	3.4	2.7	2.8
THY1	2.7		3.5	3.3
MYC	3.5	3.1	3.1	3.0
MYL9	2.2	3.4	3.6	3.4
SRPX2	2.9	3.1	3.2	3.3
SLC29A1	3.0		3.2	3.2
COL5A1	2.5	3.2	3.5	3.3
S100P	3.2	3.2	3.2	2.9
SOX9	3.3	3.2	3.0	3.0
NID1	3.1	3.0	3.2	3.2
IL33	3.9	2.7	2.9	2.9
PRRX1	2.6	3.2	3.7	2.9
LY6E	2.6	3.2	3.5	2.9
ARNTL2	3.3		3.1	2.8

SERPINE1	2.4	3.2	3.6	3.1
JUB	2.9		3.3	3.0
RGS1	3.0	3.0	3.3	2.9
CXCL5		2.7	3.5	2.9
TRIM29	2.9	3.4	3.1	2.7
CEBPB	2.7	3.1	3.2	3.2
ECT2	3.2	3.1	3.1	2.8
TPX2	3.2	3.0	3.0	3.0
ACTB	3.0	3.0	3.1	3.0
BAG2	2.4	3.2	3.3	3.1
ODAM	2.7	3.1	2.8	3.4
IGHV1-69	3.0			
SLCO4A1	2.8	3.2	3.1	2.9
KRT80	2.5	3.0	3.1	3.3
AGT	3.5	2.7	2.5	3.2
SPOCK1		2.8	3.0	3.1
MSX1	3.0	2.9	2.8	3.1
TAGLN	2.0	3.1	3.4	3.2
OR10A5	2.8	3.2	2.9	3.0
IFI6	2.9	3.2	3.0	2.6
CCL20	2.9			
TG	3.2	2.9	2.8	2.9
CCND1	2.9	3.0	3.0	2.8
EDNRA	2.5	3.0	3.1	3.0
C9ORF16	2.8	3.0	3.1	2.8
SLCO1B3	2.7	3.1	3.1	2.8
SPARC	2.4	2.9	3.3	3.0
BGN	2.3	2.9	3.3	3.1
AMIGO2	2.4	2.7	3.3	3.1
SCD	2.6		2.8	3.3
KLK6	2.3	3.2	2.9	3.2
CALU	2.8		3.0	2.9
WNT5A	3.1	2.8	3.1	2.7
APOLD1	3.1	2.5	2.5	3.4
MACC1	2.8	2.8	2.8	3.1
BHLHE40	2.7	3.1	3.1	2.7
KAL1	2.2	2.9	3.2	3.2
LRRC25	2.8	2.9	2.9	2.8
ITGBL1	2.4	2.5	3.1	3.4
TAGLN2	2.8	3.0	2.9	2.7
KIR2DL3	2.8	2.8	2.7	3.0
C7ORF68	2.8	2.8	2.8	3.0
SERHL2	2.7	3.0	2.7	2.8

PCP4	2.2	3.2	2.8	3.1
KRT6B	2.6	2.8	3.0	2.8
IGHV@	2.8			
CD44	3.0	2.9	2.8	2.5
FAM176A	2.7	2.7	2.8	2.9
EGFL6	2.9	2.9	2.9	2.5
UCA1	2.7	2.6	2.7	3.1
PLAU	2.4	2.8	3.1	2.7
ENC1	3.0	2.7	2.6	2.7
AQP9	2.2	3.1	3.1	2.6
APOD	2.2	2.5	2.9	3.4
PPAPDC1A	2.1	2.8	3.1	2.9
PTMS	2.8	2.7	2.8	2.6
CD74	2.9	2.9	2.8	2.3
MET	2.7	2.8	2.7	2.7
HTRA1	2.3	2.7	3.1	2.8
C14ORF34	2.5	2.9	2.7	2.7
FCGR3B	2.5	3.1	2.9	2.3
CCL25	2.8	2.8	2.5	2.7
TNFAIP6		2.8	3.0	2.2
FABP6	2.7	2.5	2.4	3.0
COL4A2	2.4	2.7	2.8	2.8
IGFBP7	2.4	2.6	2.8	2.9
SNTB1	2.4	2.6	2.6	3.1
ANXA6	2.4	2.7	2.8	2.7
DDIT4	2.3	2.7	2.8	2.8
HOMER1	2.4	2.6	2.7	2.8
NNMT		2.5	2.9	2.5
MMP12	2.9	2.4	2.6	
FCER1G	2.3	2.8	2.9	2.5
TRAV27	2.6	2.6	2.6	2.7
HS6ST2	2.4	2.7	2.8	2.6

APPENDIX G

UPREGULATED CRC GENES				
Gene Symbol	**Dukes A FC**	**Dukes B FC**	**Dukes C FC**	**Dukes D FC**
CLCA4	-53.3		-43.5	-63.0
MS4A12	-52.1		-36.3	-45.0
AQP8	-46.1		-33.1	-43.6
ZG16	-27.6		-27.4	-32.2
GCG	-27.4	-24.0	-23.0	-25.0
UGT2B17	-26.6	-18.8	-25.9	-20.7
ABCG2	-25.8		-20.5	-19.8
ADH1C	-20.5		-19.9	-25.1
CA1	-21.0		-16.4	-21.0
CA2	-19.7		-16.2	-22.0
PKIB	-20.2		-16.7	-19.1
HSD17B2	-20.5		-16.5	-17.8
AKR1B10	-17.8		-16.4	-17.4
SLC26A3	-17.9	-17.8	-13.1	-18.6
DHRS9	-17.8		-12.6	-17.6
SLC30A10	-16.6		-15.0	-15.7
CLDN8	-14.6	-14.1	-12.4	-16.2
GUCA2A	-14.3		-13.3	-15.2
CA4	-12.4		-11.0	-15.1
CD177	-12.9		-10.3	-14.9
MT1M	-12.6	-11.9	-10.0	-12.6
SI	-12.3	-12.1	-10.0	-12.5
CHP2	-11.2		-11.4	-11.2
ANPEP	-12.6	-10.4	-9.6	-12.5
CLCA1	-6.2	-10.1	-14.2	-12.5
TUBAL3	-11.0		-11.1	-9.5
CLDN23	-10.5		-9.6	-11.2
HEPACAM2	-7.2	-9.7	-11.9	-11.3
C6ORF105	-7.1	-10.4	-9.6	-11.5
ABCA8	-10.8	-9.8	-9.2	-8.4
PDE9A	-9.8		-9.8	-8.6
AGPAT9	-10.4		-8.8	-9.0
FAM55D	-9.0		-9.3	-9.6
SLC26A2	-8.5	-10.3	-8.5	-9.4
CEACAM7	-8.1	-9.3	-8.1	-10.2
SLC16A9	-8.4	-9.1	-9.0	-8.1
OSTBETA	-9.0	-8.6	-8.2	-8.9
CWH43	-8.3	-8.9	-8.9	-8.4
ARL14	-8.4		-7.9	-9.6

LGALS2	-6.5	-9.2	-7.5	-10.2
BEST4	-8.5		-8.0	-8.5
NPY1R	-9.4	-8.3	-7.5	-8.0
UGT2A3	-6.1	-9.6	-8.3	-9.1
PCK1	-10.0	-7.9	-7.2	-7.7
SCNN1B	-8.1	-8.0	-7.6	-8.6
LRRC19	-6.6	-9.1	-8.6	-7.8
ITLN1	-5.8	-7.7	-8.5	-10.1
LAMA1	-7.8	-7.4	-7.6	-8.6
OSTALPHA	-8.8	-7.5	-7.4	-7.2
BEST2	-7.5	-7.4	-7.5	-8.0
SST	-8.0	-7.4	-7.7	-7.2
LOC100506782	-7.6	-7.3	-7.5	-6.9
GUCA2B	-7.2	-7.3	-7.0	-7.8
MUC4	-5.6	-6.8	-6.6	-9.7
CNTN3	-6.6	-7.5	-7.6	-6.4
PLCE1	-6.7		-6.9	-7.4
LOC100133944	-4.8	-6.3	-8.1	-8.7
C2ORF88	-6.8		-6.7	-7.3
HHLA2	-6.8	-7.0	-6.7	-7.1
C10ORF99	-4.9	-7.5	-7.5	-7.4
SLC4A4	-6.9	-6.4	-5.9	-7.3
VSIG2	-6.4		-6.2	-7.3
NR3C2	-6.3		-6.7	-6.8
GBA3	-6.8	-6.6	-6.2	-6.6
HPGD	-7.6	-6.2	-5.3	-6.9
C17ORF91	-6.8		-5.6	-6.9
TMCC3	-7.2		-5.3	-6.5
ABCB1	-6.6	-6.3	-7.5	-4.7
ETFDH	-6.6		-6.3	-6.0
DHRS11	-6.3		-6.1	-6.4
UGT2B15	-6.4	-5.6	-6.6	-6.3
SULT1A1	-6.6		-6.3	-5.9
GCNT3	-7.6	-5.5	-5.2	-6.5
PLAC8	-6.1	-6.4	-4.8	-7.3
LRRC66	-6.0	-6.1	-6.2	-6.0
ISX	-5.9	-6.5	-6.2	-5.6
SCIN	-7.2	-5.7	-5.6	-5.7
IL1R2	-5.7	-5.3	-5.5	-7.4
ADAMDEC1	-4.9	-6.7	-5.8	-6.5
FAM55A	-5.4		-5.9	-6.3
TMEM220	-5.7	-6.2	-5.4	-6.1
ADH1C	-6.6	-6.2	-5.8	-4.5

CA12	-5.4		-5.8	-6.1
PHLPP2	-6.0	-5.8	-5.7	-5.5
HSD11B2	-5.4		-6.1	-5.6
PLA2G10	-6.0	-5.4	-5.2	-6.1
RUNDC3B	-5.9	-5.5	-5.8	-5.1
HMGCS2	-5.3	-6.1	-5.4	-5.1
PPARGC1A	-5.4		-5.5	-5.5
SULT1A2	-5.5	-5.5	-5.4	-5.3
TSPAN7	-5.7		-5.5	-5.1
SYTL4	-5.4		-5.0	-5.6
UGT1A7	-5.5	-5.1	-5.2	-5.5
EPB41L3	-5.4	-5.1	-5.1	-5.7
NR1H4	-5.5	-5.4	-5.3	-4.8
MEP1A	-4.0	-5.7	-6.3	-4.6
SPINK5	-5.6	-4.7	-4.0	-6.1
SELENBP1	-3.5	-5.4	-5.9	-5.5
RICH2	-4.6		-5.1	-5.3
ZC3H12C	-3.8	-5.0	-5.8	-5.4
GDPD3	-5.2	-4.8	-4.7	-5.1
SLC3A1	-4.3	-5.1	-6.1	-4.2
SDCBP2	-5.3		-4.4	-5.0
PLCL2	-4.8	-4.7	-4.8	-5.3
GHR	-5.3	-4.8	-5.0	-4.3
PTPRR	-5.7	-4.6	-4.2	-4.8
CEACAM1	-5.2	-4.6	-4.7	-4.8
SLC17A4	-4.9	-4.7	-4.8	-4.9
TMED6	-5.0	-5.1	-4.8	-4.2
BMP2	-4.5		-4.6	-5.2
SLC22A5	-4.7	-4.8	-4.8	-4.7
PAG1	-4.9	-4.9	-4.3	-4.8
STYK1	-4.9	-4.6	-4.6	-4.6
PTPRH	-4.8		-4.3	-4.9
LOC100505633	-5.0	-4.6	-4.4	-4.6
EYA2	-4.7	-4.4	-4.6	-4.8
SMPDL3A	-4.3	-4.6	-4.6	-5.0
EDN3	-4.6		-4.3	-4.8
TMEM171	-4.9	-4.4	-4.2	-4.8
GGT6	-4.0	-4.5	-4.7	-4.9
PNLIPRP2	-4.0	-4.8	-5.2	-4.2
PIGZ	-4.3	-4.6	-4.9	-4.4
APOBEC3B	-5.3	-4.3	-4.3	-4.0
XDH	-4.4	-4.3	-4.4	-4.8
CDKN2B	-5.3	-4.4	-3.9	-4.2

FABP1	-3.9	-5.8	-4.5	-3.4
MIR29C	-4.4	-4.6	-4.3	-4.2
AHCYL2	-4.2	-4.5	-4.1	-4.7
ZNF575	-4.4	-4.3	-4.4	-4.4
TCEA3	-4.0		-4.7	-4.3
HSD3B2	-4.5	-4.3	-4.0	-4.3
SLC41A2	-4.2	-4.2	-4.2	-4.5
KIAA1211	-4.1	-4.3	-3.9	-4.8
NR5A2	-4.2	-4.4	-4.1	-4.3
UPP1	-4.9	-4.1	-3.6	-4.2
PCDH20	-3.9	-4.5	-4.6	-3.9
POU2AF1	-2.8	-4.7	-4.4	-4.9
MATN2	-4.0	-4.2	-4.8	-3.8
NAT2	-3.4	-4.1	-4.7	-4.5
KBTBD11	-3.5	-4.3	-4.2	-4.7
CAPN8	-4.1	-4.1	-3.8	-4.7
ASPA	-4.3	-4.2	-4.0	-4.0
ZFP3	-3.8		-4.4	-4.1
CCDC68	-3.6	-3.8	-3.8	-5.1
FOXP1-IT1	-3.9	-4.1	-4.2	-4.2
NAAA	-4.1		-4.0	-4.2
MAMDC2	-4.6	-3.9	-4.0	-3.8
KRT20	-4.3	-4.7	-4.0	-3.2
AMPD1	-4.0	-4.1	-4.1	-4.0
TP53INP2	-4.5	-4.0	-3.9	-3.7
C4ORF19	-3.5	-3.9	-4.1	-4.5
TEX11	-4.0	-4.0	-3.9	-4.0
EPHX2	-3.3		-4.4	-4.2
TNFRSF17	-3.0	-4.4	-4.3	-4.2
C17ORF73	-3.7	-4.4	-3.8	-3.8
SGK2	-3.7	-4.0	-4.2	-3.8
CD163L1	-3.8	-4.0	-3.9	-4.0
HIST1H1C	-4.1	-3.9	-3.7	-3.9
SECTM1	-4.4	-3.6	-3.4	-4.2
HSBP1L1	-4.0		-3.5	-4.1
HIGD1D	-3.8	-3.9	-3.7	-4.1
MAOA	-4.0	-3.8	-4.2	-3.5
FLJ32810	-4.0	-4.0	-3.6	-3.9
LOC100505989	-4.0	-3.6	-4.1	-3.7
OPN3	-3.7	-4.2	-3.6	-3.8
ZDHHC23	-3.6		-4.0	-3.9
FLJ22763	-4.0	-4.0	-3.6	-3.7
NAT1	-2.9		-4.1	-4.4

IL18	-3.9	-3.7	-3.5	-4.2
AMY2B	-3.5	-3.9	-3.8	-4.0
ELOVL6	-3.9	-3.7	-3.6	-3.9
INSL5	-3.9	-3.7	-3.6	-4.0
CCNYL1	-3.7	-3.9	-3.6	-3.8
CHGA	-3.7	-3.9	-4.0	-3.5
CHST5	-3.7	-3.7	-3.8	-3.9
CPM	-3.9	-3.8	-3.5	-3.8
LOC643008	-3.8	-3.6	-3.8	-3.8
CD1D	-3.7	-3.8	-3.7	-3.9
PTGDR	-3.9	-3.7	-3.7	-3.7
FMO5	-3.8	-4.0	-3.9	-3.3
CES2	-3.9	-3.9	-3.5	-3.6
DNASE1L3	-3.1	-4.0	-4.0	-3.9
PBLD	-3.4	-4.0	-3.9	-3.7
RSAD2	-4.5	-3.5	-3.2	-3.7
SLC46A3	-4.1	-3.7	-3.4	-3.7
EMP1	-4.1	-3.6	-3.3	-3.9
DDX26B	-3.8	-3.7	-3.7	-3.5
MYO1A	-3.4	-3.9	-3.6	-3.8
ARRDC4	-3.6		-3.6	-3.9
ABCA5	-3.6	-3.5	-3.8	-3.8
AIFM3	-3.6	-3.7	-3.9	-3.5
AKR1C2	-4.0	-3.7	-3.5	-3.4
TINAG	-3.7	-3.6	-3.9	-3.3
TLCD2	-3.6		-3.4	-3.9
ST6GALNAC1	-2.8	-3.6	-3.8	-4.4
HRCT1	-3.6	-3.6	-3.5	-3.8
CHGB	-3.9	-3.6	-3.8	-3.2
CDHR5	-3.6	-3.6	-3.5	-3.8
VIPR1	-3.4	-3.6	-3.6	-3.8
SQRDL	-3.4		-3.6	-3.8

APPENDIX H

```
function [output, output2] = QLISA_2_1(directory)
%processess all of the images within a given directory
%output is list of files
%output2 calculates the avg intensity and standard deviation of the ROI
%within each file

%FILE FINDER ALGORITHM
contents=dir(directory);
filenames = {contents.name}';
for k=1:1:size(filenames,1)
    x(k,1) = strfind(filenames(k,1), '.BMP');
    x2 = cell2mat(x(k,1));
    tf(k,1) = isempty(x2);
    if tf(k,1)==0
        z(k,1).value = filenames(k);
    else z(k,1).value = [];
    end
    if ~isempty(z(k,1).value)
        output(k,1) = {z(k,1).value};
    end
end

for k = 1:1:size(output,1)
    G(k) = iscellstr(output{k});
end
G = G';
n = size(output,1)-sum(G);
first = n+1;

%IMAGE PROCESSING
for b = first:1:size(output,1)
    filename = output{b,1}{1,1};
    img=imread(filename);
    info = imfinfo(filename);

    img=imread(filename);
    info = imfinfo(filename);
    %figure, imshow(img);

    %DUST FILTER
    [M N]=size(img);                        %M is rows, N is columns
    threshold = graythresh(img);
    bw = im2bw(img,threshold);
    bw2 = bwareaopen(bw,100);
    dust = imsubtract(bw,bw2);
    img1=img;
    for k=1:1:M
        for c=1:1:N
            if dust(k,c)==1
               img1(k,c)=0;
            else img1(k,c)=img(k,c);
            end
        end
    end
    %figure, imshow(img1);
```

```
    threshold2 = graythresh(img1);
    bw3 = im2bw(img1,threshold2);
    %figure, imshow(bw3);                    %BW ROI

    %CROP IMAGE along X
    x = sum(img)/M;                          %average row intensities
    %figure, plot(x);
    [C xidx] = max(x);                     %C is the value, xidx is the
index
    img2 = img1(:,xidx-5:xidx+5);
    %figure, imshow(img2);

    threshold3 = graythresh(img2);
    bw4 = im2bw(img2,threshold3);

    %BOUNDARY FILTER
    [B,L] = bwboundaries(bw4);
    %imshow(label2rgb(L, @jet, [.5 .5 .5]))
    %hold on
    %for k = 1:length(B)
    %    boundary = B{k};
    %    plot(boundary(:,2), boundary(:,1), 'w', 'LineWidth', 2)
    %end

    k = cell2mat(B);
    idx = mean(k);
    %xidx = round(idx(1,2));
    yidx = round(idx(1,1));

    %CROP IMAGE
    xmin = xidx - 5;
    ymin = yidx - 5;
    width = 10;
    height = 10;

    rect = [xmin ymin width height];
    img2 = imcrop(img,rect);
    figure, imshow(img2);                   %cropped image

    %Compile Data
    avg1 = mean(img2);
    avg_intensity = mean(avg1);
    standard_dev = std(avg1);

    z = cat(1,filename);
    x = cat(1,avg_intensity);
    y = cat(1,standard_dev);
    output2{b,1} = z;
    output2{b,2} = x;
    output2{b,3} = y;
end
```