# Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format.

1. **Grantee Institution:** The Pennsylvania State University

2. **Reporting Period (start and end date of grant award period):** 1/1/2010 - 12/31/2013

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** John Anthony, MPA

4. **Grant Contact Person's Telephone Number:** 814 935 1081

5. **Grant SAP Number:** 4100050904

6. **Project Number and Title of Research Project:** 9. Identification and Function Study of Inflammatory Bowel Disease-Associated DNA Methylation

7. **Start and End Date of Research Project:** 7/1/2010 - 6/30/2011

8. **Name of Principal Investigator for the Research Project:** Zhenwu Lin, PhD

9. **Research Project Expenses**.

   9(A) Please provide the total amount of health research grant funds spent on this project for the entire duration of the grant, including indirect costs and any interest earned that was spent:

   $ 46,530

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

| Last Name, First Name | Position Title | % of Effort on Project | Cost |
|---|---|---|---|
| John, Gerrit | Post-doctoral Fellow | 17% | $7,298 |
| Wang, Yunhua | Research Technician | 19% | $7,838 |

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

| Last Name, First Name | Position Title | % of Effort on Project |
|---|---|---|
| Lin, Zhenwu | Assistant Professor | 5% |

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

| Type of Scientific Equipment | Value Derived | Cost |
|---|---|---|
| None | | |

**10. Co-funding of Research Project during Health Research Grant Award Period.** Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes_____ No____X_____

If yes, please indicate the source and amount of other funds:

**11. Leveraging of Additional Funds**

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes_____ No_____X_____

If yes, please list the applications submitted (column A), the funding agency (National Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert “not funded” in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

| A. Title of research project on grant application | B. Funding agency (check those that apply) | C. Month and Year Submitted | D. Amount of funds requested: | E. Amount of funds to be awarded: |
|---|---|---|---|---|
| None | ☐NIH<br>☐ Other federal (specify:_____)<br>☐ Nonfederal source (specify:_) | | $ | $ |

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes___X _____ No_____

If yes, please describe your plans:

I'll continue working to analyze the data, and hopefully publish 3-4 papers. I also plan to apply for funding in the future to expand the research of DNA methylation in inflammatory bowel disease.

**12. Future of Research Project.** What are the future plans for this research project?

A. Identify common IBD-associated DNA methylation by mega data analysis. Compare our previous DNA methylation results (intestinal tissues from CD-Il, CD-colon, and UC, By Illumina Human methylation 27) (not finally analyzed) with Nimmo et al (whole blood from CD-intestine, CD-colon, and UC, By Illumina Human methylation 27) IBD 2012 18(5)889-, and Cooke et al (Rectal biopsy from CD and UC, By Illumina Human methylation 27) IBD 2012 18(11)2128-

B. Functional study of the IBD-associated DNA methylation identified from this project: using DNA methylation binding protein and target gene expression.

C. Establish an optical in vivo detection method for the target DNA methylation in a single gene. I will use the IBD-associated DNA methylation identified as targets. The method to be developed could be used for in vivo diagnosis and targeted gene therapy.

**13. New Investigator Training and Development**. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

Yes____X_____ No_____

If yes, how many students? Please specify in the tables below:

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| Male | | | | 1 |
| Female | | | | |
| Unknown | | | | |
| **Total** | | | | **1** |

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| Hispanic | | | | |
| Non-Hispanic | | | | 1 |
| Unknown | | | | |
| **Total** | | | | **1** |

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| White | | | | 1 |
| Black | | | | |
| Asian | | | | |
| Other | | | | |
| Unknown | | | | |
| **Total** | | | | **1** |

14. **Recruitment of Out-of–State Researchers**. Did you bring researchers into Pennsylvania to carry out this research project?

    Yes_____ No____X_____

    If yes, please list the name and degree of each researcher and his/her previous affiliation:

15. **Impact on Research Capacity and Quality**. Did the health research project enhance the quality and/or capacity of research at your institution?

    Yes____X_____ No_____

    If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

    *Contributions to Biobank.* Surgical tissues collected from IBD patients improved the sample quality and increased sample quantity. We also provided samples for gene function study and developed a mechanism for investigation of genetic and epigenetic factors in inflammatory bowel disease.

**16. Collaboration, business and community involvement.**

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes____X_____ No_____

If yes, please describe the collaborations:

Chen, Xi, PhD, Assistant Professor, Statistician, Vanderbilt University, for data analysis
Fan, Jian Bing, PhD, Director of Research Dept, Illumina Inc., for using Illumina Human Methylation 27 and disease-associated DNA methylation identification.

16(B) Did the research project result in commercial development of any research products?

Yes_____ No___X_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes_____ No____X_____

If yes, please describe involvement with community groups that resulted from the research project:

**17. Progress in Achieving Research Goals, Objectives and Aims.**

List the project goals, objectives and specific aims (as contained in the grant agreement). Summarize the progress made in achieving these goals, objectives and aims for the period that the project was funded (i.e., from project start date through end date). Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending

publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site. **There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.**

We hypothesize that DNA methylation plays a role in IBD pathogenesis. The objectives of this project are to identify DNA methylation CpG sites in IBD which would further add to the understanding of underlying pathophysiology and to provide a potential clinical application in terms of disease diagnosis.

Specific aim 1. To study DNA methylation in naive B cell and mucosal tissues with the Illumina HumanMethylation 27 BeadChip.

Crohn's disease (CD) and ulcerative colitis (UC) are the two major forms of IBD. CD occurs in both intestine and colon while UC only occurs in the colon. Aberrant DNA methylation has been found in IBD and IBD-associated colon rectal cancer patients. Considering the disease heterogeneity, we aim to identify disease subtype- and affected tissue-specific DNA methylation using terminal ileum (TI) tissues from CD patients and colon tissues from UC patients.

Supported by this grant, we have identified IBD-associated DNA methylation with Illumina Human methylation 27 Beadarray from surgical disease tissues of patients.

## *MATERIALS AND METHOD*

*Table 1. Study samples*

| *Sample* | *Disease* | *Tissue* | *Status* | *Sex* | *Group* | *Sample* | *Disease* | *Tissue* | *Status* | *Sex* | *Group* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **105-03-1** | **CD** | A0 | Nondis | M | W | **118-01-1** | **CD** | C0 | Nondis | F | W |
| | | A1 | Dis | | | | | D1 | Dis | | |
| **117-01-1** | **CD** | A0 | Nondis | F | W | **132-01-1** | **CD** | H0 | Nondis | M | W |
| | | A1 | Dis | | | | | H1 | Dis | | |
| **130-01-1** | **CD** | A0 | Nondis | M | W | **132-01-1** | **CD** | B0 | Nondis | M | W |
| | | A1 | Dis | | | | | B1 | Dis | | |
| **132-01-1** | **CD** | A0 | Nondis | M | W | **777-14-1** | **CD** | H0 | Nondis | F | W |
| | | A1 | Dis | | | | | B1 | Dis | | |
| **777-23-1** | **CD** | A0 | Nondis | F | W | **777-16-1** | **CD** | DO | Nondis | F | W |
| | | A1 | Dis | | | | | H1 | Dis | | |
| **777-25-1** | **CD** | A0 | Nondis | F | W | **777-59-1** | **CD** | J0 | Nondis | M | W |
| | | A1 | Dis | | | | | J1 | Dis | | |
| **777-34-1** | **CD** | A0 | Nondis | F | W | **777-71-1** | **CD** | H0 | Nondis | M | W |
| | | A1 | Dis | | | | | B1 | Dis | | |
| **777-35-1** | **CD** | A0 | Nondis | M | W | **888-129-1** | **CD** | D0 | Nondis | M | W |
| | | A1 | Dis | | | | | D1 | Dis | | |
| **777-39-1** | **CD** | A0 | Nondis | F | W | **777-11-1** | **CD** | C0 | Nondis | | |
| | | A1 | Dis | | | | | C1 | Dis | | |
| **777-57-1** | **CD** | A0 | Nondis | F | W | **120-01-1** | **UC** | B0 | Nondis | F | W |
| | | A1 | Dis | | | | | D1 | Dis | | |
| **777-73-1** | **CD** | A0 | Nondis | F | W | **125-01-1** | **UC** | C0 | Nondis | F | W |
| | | A1 | Dis | | | | | B1 | Dis | | |
| **777-11-1** | **CD** | A0 | Nondis | | | **888-44-1** | **UC** | C0 | Nondis | F | W |
| | | A1 | Dis | | | | | D1 | Dis | | |
| **777-76-1** | **CD** | A0 | Nondis | M | W | **888-46-1** | **UC** | B0 | Nondis | F | W |
| | | A1 | Dis | | | | | C1 | Dis | | |
| 777-85-1 | CD | A0 | Nondis | | | **888-49-1** | **UC** | C0 | Nondis | F | W |
| | | A1 | Dis | | | | | C1 | Dis | | |
| H1 | Health | Intest | Health | | | **888-50-1** | **UC** | C0 | Nondis | F | W |
| H1 | Health | Colon | Health | | | | | D1 | Dis | | |
| H2 | Health | Intest | Health | | | **888-51-1** | **UC** | B0 | Nondis | F | W |
| H2 | Health | Colon | Health | | | | | D1 | Dis | | |
| H3 | Health | Intest | Health | | | **888-56-1** | **UC** | H0 | Nondis | M | W |
| H3 | Health | Colon | Health | | | | | D1 | Dis | | |
| H4 | Health | Intest | Health | | | **888-68-1** | **UC** | D0 | Nondis | M | W |
| H4 | Health | Colon | Health | | | | | D1 | Dis | | |
| H5 | Health | Intest | Health | | | **888-83-1** | **UC** | H0 | Nondis | M | W |
| H5 | Health | Colon | Health | | | | | D1 | Dis | | |
| H6 | Health | Intest | Health | | | **888-85-1** | **UC** | B0 | Nondis | M | W |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| H6 | Health | Colon | Health | | | D1 | Dis | | |
| H7 | Health | Intest | Health | **888-85-1** | **UC** | D0 | Nondis | M | W |
| H7 | Health | Colon | Health | | | D1 | Dis | | |
| H8 | Health | Intest | Health | **888-113-1** | **UC** | *B0* | Nondis | M | W |
| H8 | Health | Colon | Health | | | B1 | Dis | | |
| H9 | Health | Intest | Health | **888-137-1** | **UC** | B0 | Nondis | M | W |
| H9 | Health | Colon | Health | | | B1 | Dis | | |
| H10 | Health | Intest | Health | **888-137-1** | **UC** | C0 | Nondis | M | W |
| H10 | Health | Colon | Health | | | D1sig | Dis | | |

A: TI; B,C,D: right, transverse, left colon, respectively; G: small intestine; H: cecum; J: large bowel;
Intest: small intestine: Health: control unrelated individuals without gastrointestinal diseases

As listed in above (Table 1) 3 groups of IBD patients: 14 pairs of diseased and matched nondiseased TI tissues from 14 CD at TI patients, 8 pairs of diseased and matched nondiseased colon tissues from 8 CD at colon patients, and 15 pairs of diseased and matched nondiseased colon tissues from 13 UC patients. The tissues from patients were collected at the time of surgery and classified as diseased and nondiseased by a pathologist. The IBD patients were from central Pennsylvania. Recruitment and informed written consent of all study participants followed a protocol approved by the Penn State College of Medicine Institutional Review Board. As controls, healthy intestinal and healthy colon tissues were obtained from 10 individuals without gastrointestinal diseases from NDRI (National Disease Research Interchange. Philadelphia, PA).

*1. DNA Isolation*

Frozen tissues were pulverized in liquid nitrogen and the tissue powder was used for genomic DNA isolation. DNA was isolated with QIAamp DNA Mini Kit (Qiagen Inc. Valencia, CA) according to manufacturer's instruction. The concentration of DNA was measured with a Nanodrop ND-1000 spectrophotometer (NanoDrop Technology, Wilmington, DE) and the DNA samples were stored at -70$^{o}$C until use.

*2. Bisulfite Conversion of Genomic DNA*

One microgram of genomic DNA was used for conversion reaction with Methylation-Gold kit (Zymo Research, Orange, CA) according to manufacturer's recommendations. After conversion, the non-methylated Cs in the genome were converted to T and the methylated Cs remained Cs.

*3. Human Methylation 27 BeadArray Assay*

The bisulfite-converted DNA was used for DNA methylation profiling with the Illumina Infinium Human Methylation 27 BeadChip that covers 27,578 highly informative CpG sites from 14,475 consensus coding sequencing in the NBI database (Genome Build 36) and 110 miRNA promoters. The BeadArray assay was performed at the Institute of Molecular Medicine, Cleveland Clinic. Briefly, 500ng bisulfite-converted genomic DNA was used as template for PCR amplification. The single stranded PCR products were prepared by denaturation, and then hybridized to a Sentrix® Array Matrix. The array hybridization was conducted under a temperature gradient program, and arrays were imaged using a BeadArray Reader 1000 scanner.

*4. Bisulfite PCR Based RFLP (restriction fragment length polymorphism) Analysis*

For microarray data validation, we used a PCR-based RFLP method to confirm the methylation status of 4 IBD-associated CpG sites HOXD12 cg03874199 and cg23130254, CPA5 cg02859426, and THYN1 cg09360083. These 4 CpG sites were identified by our microarray analysis as showing distinct methylation profiles in diseased and/or nondiseased tissues. The methodology was described previously (36). PCR primers were designed to flank the corresponding CpG sites, H03sm (5'-GTGGGTTGGTTTTTGACTTTGTAGT) and H04rm (5'-CCAACTAACCTCCATTCTACCTCA) for cg03874199; H03sm and H05rm (CATAACCCCCTCTATAAAAACTAC) for cg23130254; C01sm (GGTTAATGGTTATTATATTGGATAG) and C02rm (CTTATTATAACATCAACACATTA ACC) for cg02859426; and T02smth (GGTAATTTGGGTTATTA AATTT ACG) and T03rm (TTCTTTACAACAAAATAAATCCCTTA) for cg09360083. PCR was performed as described previously(1). PCR products were digested with DpnI (for cg03874199), Hpy99I (for cg23130254), HpyCH4IV (for cg02859426), and BstU1 (for cg09360083) according to manufacturer's instruction (NEB). The digested PCR products were separated by 8% PAGE gel and the bands for methylated or nonmethylated C were quantified using ImageJ. The β-value was calculated as methylated CpGs/methylated +nonmethylated CpGs. For validation, the β-value from RFLP analysis was compared to the β value of the same site and the same sample from the Humanmethylation27 BeadArray assay.

*5. Data Analysis*

DNA methylation assays were performed using Illumina BeadArray technology(2). A β value of

0-1.0 was reported for each CpG site, signifying the methylation level. The β values were calculated by subtracting background using of negative controls on the array and taking the ratio of methylated signal intensity to the sum of both methylated and unmethylated signals.

IBD specific DNA methylation was identified by comparison of DNA methylation profile of B cells from IBD patients vs. non-IBD siblings. To assess the differences between diseased and non-diseased groups, Wilcoxon-signed rank test was used to compare β values due to small size in this experiment and violation of normality assumption.
For disease-associated DNA methylation identification, the threshold is as: FDR (fault discovery rate) $<0.25$, p value (raw data p value) $\leq 0.0001$ for diseased TI tissues from CD patients; FDR=0.25, $p \leq 0.0001$ for diseased colon tissues from CD patients; FDR$<0.03$, $p \leq 0.00005$ for nondiseased tissues of both CD-TI and UC.

Results

DNA methylation varies between tissue and cell types. Considering CD and UC are different subtypes of IBD and diseases can occur at various locations in the intestinal tract, we studied disease-associated methylation identification from CD at TI and UC at colon using tissue- and type-matched diseased and nondiseased tissue pairs (Table 1) in this study. By comparing DNA methylation profile from disease tissues vs. matched nondiseased tissues from the same IBD patients we identify disease-associated DNA methylation for diseased tissues. Such comparison will help to reduce the interference from genetic, life style, and environmental background. For identification of disease-associated DNA methylation from nondiseased tissues we compared DNA methylation profile from nondiseased tissues of IBD patients vs. unrelated healthy tissues.

The rationale is that the nondiseased tissues with normal appearance are adjacent to diseased area, and may be undergoing pathological changes at the level of DNA methylation.

*1. Disease-associated DNA Methylation Identified from Diseased Tissues*

From the analysis of DNA methylation 27 Beadarray, the minimal FDR for diseased tissues is as high as 0.195 for TI tissues of CD patients and 0.372 for colon tissues of UC patients. Considering the comparison is between the tissues from the same patient, we use FDR$<0.250$, $p \leq 0.0001$ as threshold for diseased-associated methylation identification for CD at TI and FDR=0.372, $p \leq 0.0001$ as threshold for UC at colon.

*CD-associated DNA methylation in diseased TI tissue in CD at TI patients*

*Table 2. Diseased associated DNA methylation in diseased tissues from CD-TI patient (FDR<0.25, 9 CpGs)*

| | | Dis/nondis CD-intest | | | Dis/nondis CD-colon | | | Dis/nondis UC-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg00445824 | ISYNA1 | 0.055 | 6.21E-05 | 0.195 | -0.017 | 4.98E-01 | 0.821 | 0.040 | 4.92E-03 | 0.372 |
| cg00947275 | LAS1L | -0.096 | 6.80E-05 | 0.195 | -0.035 | 4.86E-01 | 0.814 | -0.065 | 1.23E-01 | 0.481 |
| cg03578041 | LARP6 | -0.062 | 7.06E-05 | 0.195 | 0.077 | 1.10E-01 | 0.614 | -0.038 | 1.25E-01 | 0.483 |
| cg03874199 | HOXD12 | 0.224 | 1.99E-05 | 0.195 | 0.199 | 7.99E-03 | 0.587 | -0.064 | 2.65E-01 | 0.600 |
| cg05990214 | SPHK2 | -0.022 | 2.73E-05 | 0.195 | -0.017 | 8.57E-03 | 0.587 | 0.004 | 6.55E-01 | 0.845 |
| cg09072120 | PRTFDC1 | 0.044 | 5.10E-05 | 0.195 | -0.035 | 3.27E-02 | 0.592 | -0.013 | 3.72E-01 | 0.673 |
| cg09643313 | NA | 0.023 | 6.90E-05 | 0.195 | -0.043 | 8.86E-04 | 0.543 | -0.005 | 6.70E-01 | 0.852 |
| cg13086586 | PPAT | 0.058 | 3.18E-05 | 0.195 | 0.015 | 5.52E-01 | 0.846 | -0.003 | 8.48E-01 | 0.939 |
| cg14029663 | C14orf138 | 0.028 | 3.99E-05 | 0.195 | 0.004 | 6.98E-01 | 0.904 | 0.004 | 7.24E-01 | 0.880 |
| cg21284636 | PIN4 | -0.116 | 3.48E-05 | 0.195 | 0.011 | 8.68E-01 | 0.963 | -0.053 | 2.94E-01 | 0.622 |
| cg26620157 | PAX9 | 0.059 | 9.09E-05 | 0.228 | 0.018 | 7.06E-01 | 0.908 | 0.014 | 6.19E-01 | 0.824 |
| cg23130254 | HOXD12 | 0.218 | 1.03E-04 | 0.236 | 0.160 | 4.12E-03 | 0.587 | -0.111 | 4.64E-02 | 0.387 |

We identified 12 CD-associated CpGs from diseased TI tissues of 14 CD patients (Table 2). These 12 CpGs are from 11 coding gene regions (2 CpGs cg03874199 and cg23130254 are from the same gene) in which the region for cg09643313 is still undefined. Gene annotation and function for the 10 known genes are listed in Table 3.

*Table 3. Annotation of genes with disease-associated methylation in IBD*

| *Genes* | *Methylation in IBD* | | | *Gene name* | *Gene function* | Other |
|---|---|---|---|---|---|---|
| | *CpGs* | *Indentified from* | *Association with* | | | |
| ZNF22 | cg00899659 | CD-colon dis | CD-colon nondis | zinc finger protein 22 (KOX 15) | DNA binding; zinc ion binding | |
| SCUBE2 | cg01081263 | CD-colon dis | CD-colon nondis, CD-intest dis & nondis | signal peptide, CUB domain, EGF-like 2 | calcium ion binding glycoprotein; regulated by IL1B, LPS, TNF; hedgehog signaling | |
| LDB3 | cg08300860 | CD-colon dis | | LIM domain binding 3 | cytoskeleton | |
| CALHM1 | cg09155001 | CD-colon dis | | calcium homeostasis modulator 1 | calcium ion transport | |
| CENPQ | cg20337106 | CD-colon dis | | centromere protein Q | cell cycle | |
| WNT7B | cg21104946 | CD-colon dis | CD-colon nondis | wingless-type MMTV integration site family, member 7B | Wnt signaling pathway | |
| NA | cg22533573 | CD-colon dis | | | | |
| CLEC4E | cg12762799 | CD-colon dis | | C-type lectin domain family 4, member E; mincle | mycobacterial cord factor receptor; inflammatory response; regulated by LPS, IL13, IL6, | |
| PAPSS2 | cg13926569 | CD-colon dis | CD-colon nondis | 3'-phosphoadenosine 5'-phosphosulfate synthase 2 | ATP sulfurylase and APS kinase | |
| ISYNA1 | cg00445824 | CD-intest dis | CD-colon dis | inositol-3-phosphate synthase 1 | myo-inositol biosynthesis | |
| LAS1L | cg00947275 | CD-intest dis | | LAS1-like | cell proliferation and ribosome biogenesis | |
| LARP6 | cg03578041 | CD-intest dis | | La-related protein 6 | RNA processing; regulation of collagen synthesis | |
| HOXD12 | cg03874199 | CD-intest dis | CD-intest nondis | homeobox D12 transcription factor | transcriptional regulation; development | |
| SPHK2 | cg05990214 | CD-intest dis | | Sphingosine kinase 2 | sphingolipid metabolism | |
| PRTFDC1 | cg09072120 | CD-intest dis | | phosphoribosyltransferase domain-containing protein 1 | Purine Metabolism | |
| NA | cg09643313 | CD-intest dis | | | | |
| PPAT | cg13086586 | CD-intest dis | | amidophosphoribosyltransferase | Purine Metabolism | |
| C14orf138 | cg14029663 | CD-intest dis | | chromosome 14 open reading frame 138; methyltransferase like 21D (METTL21D) | unknown | |
| PIN4 | cg21284636 | CD-intest dis | | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 | Protein folding; rRNA processing | |
| PAX9 | cg26620157 | CD-intest dis | | paired box 9 transcription factor | transcriptional regulation; development | |
| HOXD12 | cg23130254 | CD-intest dis | CD-colon nondis | homeobox D12 transcription factor | transcriptional regulation; development | |
| ID2 | cg13055278 | UC-colon dis | | inhibitor of DNA binding 2, dominant negative helix-loop-helix protein | transcriptional regulation | |
| NEB | cg16753209 | UC-colon dis | UC-colon nondis | nebulin | muscle contraction | |
| LY86 | cg20162076 | UC-colon dis | | lymphocyte antigen 86 | inflammatory response | |
| GLDN | cg07888040 | UC-colon dis | UC-colon nondis | gliomedin | matrix component of peripheral nerves | |
| MYBPH | cg18959422 | UC-colon dis | | myosin binding protein H | cell adhesion; muscle contraction | |

| | | | | | | |
|---|---|---|---|---|---|---|
| ZCCHC14 | cg21865006 | UC-colon dis | UC-colon nondis, CD-colon nondis | zinc finger, CCHC domain containing 14 | cell communication | |
| CCDC91 | cg17489451 | UC-colon dis | UC-colon nondis | coiled-coil domain containing 91 | golgi protein transport | |
| DEFB105A | cg23152755 | UC-colon dis | UC-colon nondis | defensin, beta 105 | antimicrobial response | |
| CSTA | cg26928972 | UC-colon dis | UC-colon nondis | cystatin A | protease inhibitor | |
| ABCC9 | cg20025970 | UC-colon dis | UC-colon nondis | ATP-binding cassette, sub-family C (CFTR/MRP), member 9 | potassium ion transport | |
| IDO1 | cg08465774 | UC-colon dis | UC-colon nondis, CD-colon nondis | indole 2,3-dioxygenase, tryptophan 2,3-dioxygenase | inflammatory response; tryptophan Metabolism | |
| TEAD1 | cg19447966 | UC-colon dis | UC-colon nondis, CD-intes nondis | TEA domain family member 1 (TEF-1) | transcriptional regulation; development | |
| UMOD | cg07456201 | UC-colon dis | | uromodulin | cellular defense response | |
| ASIP | cg13681935 | UC-colon dis | UC-colon nondis | agouti signaling protein | Melanogenesis | |
| ESRRA | cg23073111 | UC-colon dis | UC-colon nondis, CD-colon nondis | estrogen-related receptor alpha | transcriptional regulation | |
| TEX264 | cg03019000 | UC-colon dis | UC-colon nondis | testis expressed 264 | unknown | |
| AQP9 | cg11098259 | UC-colon dis | UC-colon nondis | aquaporin 9 | solute transport | |
| C19orf33 | cg00412772 | CD-colon nondis | UC-colon nondis | chromosome 19 open reading frame 33 | DNA binding | |
| ADAMTSL 5 | cg04601137 | CD-colon nondis | | ADAMTS-like 5 | metalloendopeptidase | |
| C19orf33 | cg07665060 | CD-colon nondis | | chromosome 19 open reading frame 33 | DNA binding | |
| RNF24 | cg10987882 | CD-colon nondis | CD-intest dis | ring finger protein 24 | protein binding | |
| NA | cg14203179 | CD-colon nondis | UC-colon dis&nondis | | | |
| CEBPG | cg15046693 | CD-colon nondis | | CCAAT/enhancer binding protein (C/EBP), gamma | transcriptional regulation | |
| KRT19 | cg16585619 | CD-colon nondis | UC-colon nondis | keratin 19 | Cytoskeleton | |
| SLC22A18 | cg16873863 | CD-colon nondis | | solute carrier family 22, member 18 | Organic cation transport | Imprinted gene |
| GRAP | cg23276115 | CD-colon nondis | | GRB2-related adaptor protein | Ras protein signal transduction; cell-cell signaling | |
| CDHR5 | cg22289115 | CD-colon nondis | UC-colon nondis | cadherin-related family member 5 | calcium-dependent cell adhesion | |
| KCNAB2 | cg19814116 | CD-colon nondis | | potassium voltage-gated channel, shaker-related subfamily, beta member 2 | voltage-gated potassium channel | |
| NA | cg26119796 | CD-colon nondis | | | | |
| PKP3 | cg08314660 | CD-colon nondis | | plakophilin 3 | cell adhesion | |
| CYP7A1 | cg12010995 | CD-colon nondis | UC-colon nondis | Cholesterol 7 alpha hydroxylase | bile acid, cholesterol, xenobiotic metabolism | |
| SGCE | cg01169624 | CD-colon nondis | UC-colon nondis | sarcoglycan, epsilon | ell-matrix adhesion; muscle organ development | |
| NA | cg23051598 | CD-colon nondis | UC-colon nondis | | | |
| CDX1 | cg15452204 | CD-colon nondis | UC-colon nondis | caudal type homeobox 1 | intestinal-specific transcriptional regulation; homeodomain transcription factor | |
| NA | cg16075940 | CD-colon nondis | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| TRPM6 | cg22161874 | CD-colon nondis | UC-colon nondis | transient receptor potential cation channel, subfamily M, member 6 | protein phosphorylation; ion transport | |
| PLAGL1 | cg00702231 | CD-colon nondis | UC-colon nondis | pleiomorphic adenoma gene-like 1; zinc finger protein | cell cycle; transcriptional regulation | Imprinted gene |
| NA | cg01584473 | CD-colon nondis | | | | |
| CSF2 | cg02196805 | CD-colon nondis | | granulocyte-macrophage colony stimulating factor (GN-CSF) | cytokine; immune response | |
| UBE2K | cg02523617 | CD-colon nondis | | ubiquitin-conjugating enzyme E2K | protein ubiquitination | UBE2K **rs652162** (A/G) associated with CD (p-value=6.6188E-4) WTCC. |
| L3MBTL1 | cg02611863 | CD-colon nondis | UC-colon nondis | Lethal(3)malignant brain tumor-like protein 1 | Polycomb group (PcG) protein; transcriptional regulation | |
| CPA5 | cg02855426 | CD-colon nondis | UC-colon nondis | carboxypeptidase A5 | Metallocarboxypeptidase | CPA5 **rs17164838** (A/G) associated with CD (p-value=5.98E-5). WTCC |
| NA | cg03085377 | CD-colon nondis | | | | |
| CTSW | cg03887787 | CD-colon nondis | | cathepsin W | cysteine-type endopeptidase | |
| TMC6 | cg04947157 | CD-colon nondis | | transmembrane channel-like protein 6; epidermodysplasia verruciformis protein 1. | zinc-transporting protein | |
| TRPV3 | cg05790038 | CD-colon nondis | | transient receptor potential cation channel, subfamily V, member 3 | calcium ion transport | TRPV3 **rs7219780** associated with CD (p-value=5.25724E-4). |
| ACY3 | cg06148175 | CD-colon nondis | | aspartoacylase-2 | aminoacylase | |
| HOXD12 | cg03874199 | CD-intest nondis | UC-colon dis | homeobox D12 transcription factor | transcriptional regulation; development | |
| FBLIM1 | cg23002761 | CD-intest nondis | | filamin binding LIM protein 1 | cell adhesion; cytoskeleton | |
| AQP5 | cg23855989 | CD-intest nondis | UC-colon dis | aquaporin 5 | water transport | |
| CYP2A13 | cg07905963 | CD-intest nondis | | cytochrome P450 2A13 | xenobiotic metabolism | |
| VILL | cg04660410 | CD-intest nondis | CD-colon dis | villin-like protein | cytoskeleton | |
| PLD6 | cg05590257 | CD-intest nondis | | phospholipase D family, member 6 | lipid catabolism | |
| GUCA2B | cg19728577 | CD-intest nondis | CD-colon dis | guanylate cyclase C-activating peptide 2; uroguanylin. | activates intestinal guanylate cyclase; cGMP synthesis | |
| EVI5 | cg00519208 | UC-colon nondis | | ecotropic viral integration site 5 | cell cycle | |
| CPA5 | cg02855426 | UC-colon nondis | CD-colon nodis | carboxypeptidase A5 Metallocarboxypeptidase | CPA5 rs17164838 (A/G) associated with CD (p-value=5.98E-5). WTCC | |
| TMPRSS3 | cg25608949 | UC-colon nondis | | transmembrane protease, serine 3 | serine-type endopeptidase; proteolysis | |

| | | | | | | |
|---|---|---|---|---|---|---|
| ABCA9 | cg06213598 | UC-colon nondis | UC-colon dis | ATP-binding cassette, sub-family A (ABC1), member 9 | transmembrane transport | |
| THYN1 | cg09360083 | UC-colon nondis | CD-intest dis, UC-colon dis, CD-colon nondis | thymocyte nuclear protein 1 | apoptosis | |
| EMCN | cg12532667 | UC-colon nondis | | endomucin | endothelial sialomucin | |
| CCDC91 | cg17489451 | UC-colon nondis | UC-colon dis | coiled-coil domain containing 91 | golgi protein transport | |
| GPX1 | cg06613840 | UC-colon nondis | UC-colon dis, CD-colon nondis | | | |
| IDO1 | cg08465774 | UC-colon nondis | UC-colon dis, CD-colon nondis | indole 2,3-dioxygenase, tryptophan 2,3-dioxygenase | inflammatory response; tryptophan Metabolism | |
| ASB5 | cg11698653 | UC-colon nondis | UC-colon dis | ankyrin repeat and SOCS box protein 5 | intracellular signaling | |
| PEBP1 | cg14349467 | UC-colon nondis | UC-colon dis, CD-colon nondis | phosphatidylethanolamine binding protein 1 | Serine protease inhibitor | |
| IAPP | cg15583072 | UC-colon nondis | | islet amyloid polypeptide | hormone activity; apoptosis | |
| PLAC4 | cg20423977 | UC-colon nondis | UC-colon dis | placenta-specific 4 protein | unknown | |
| OTOR | cg21845297 | UC-colon nondis | UC-colon dis | otoraplin | cartilage condensation | |
| NA | cg21663722 | UC-colon nondis | UC-colon dis | | | |
| ABCB5 | cg22066521 | UC-colon nondis | | ATP-binding cassette, sub-family B (MDR/TAP), member 5 | transmembrane transport | |
| TNIP3 | cg20950277 | UC-colon nondis | UC-colon dis, CD-colon nondis | TNFAIP3 interacting protein 3 | regulates NF-kappa-B activation | |
| CACNG3 | cg04721098 | UC-colon nondis | UC-colon dis | voltage-dependent,calcium channel gamma subunit 3 | calcium channel stabilization | |
| TGM4 | cg09111917 | UC-colon nondis | UC-colon dis | glutamine gamma-glutamyltransferase 4 | peptide cross-linking | |
| CP | cg17439694 | UC-colon nondis | UC-colon dis | ceruloplasmin (ferroxidase) | copper ion transport; | |
| DCD | cg13396068 | UC-colon nondis | UC-colon dis | dermcidin | antimicrobial response | |
| KIR3DL1 | cg15627025 | UC-colon nondis | | killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1 | immune response; natural killer cells | |
| FAM50B | cg03202897 | UC-colon nondis | UC-colon dis, CD-colon nondis | family with sequence similarity 50, member B | unknown | imprinted gene |
| FEM1A | cg16118817 | UC-colon nondis | UC-colon dis, CD-colon nondis | fem-1 homolog a | Protein ubiquitination | |
| PTMA | cg13921319 | UC-colon nondis | UC-colon dis, CD-colon nondis | prothymosin, alpha | Immune response; ligand for TLR4; transcriptional regulation | |
| PRSS2 | cg13944141 | UC-colon nondis | UC-colon dis, CD-colon nondis | protease, serine, 2 (trypsin 2) | serine-type endopeptidase; proteolysis | |
| PLAGL1 | cg00702231 | UC-colon nondis | CD-colon nondis | pleiomorphic adenoma gene-like 1; zinc finger protein | cell cycle; transcriptional regulation | imprinted gene |
| TBC1D3C | cg02601403 | UC-colon nondis | UC-colon dis | TBC1 domain family, member 3F | Rab GTPase activation | |
| L3MBTL1 | cg02611863 | UC-colon nondis | CD-colon nondis | Lethal(3)malignant brain tumor-like protein 1 | Polycomb group (PcG) protein; transcriptional regulation | |
| KRTAP13-1 | cg02764897 | UC-colon nondis | UC-colon dis | keratin associated protein 13-1 | cytoskeleton | |

The majority of these genes play regulatory functions, including 3 genes, HOXD12, PAX 9, and LAS1L in transcriptional regulation, 2 genes in cellular signaling, ISYNA1 for myo-inositol signaling and SPHK2 for sphingolipid signaling; and 2 genes LARP6 in RNA processing and PIN4 in protein folding. Other functions include genes PRTFDC and PPAT in purine metabolism, gene C14orf138 likely as a methyltransferase.

Interestingly, the HOXD12 gene has two CD-associated CpGs that were observed in diseased (FDR=0.195, p=1.99E-05 for cg03874199; FDR=0.236, p=1.03E-04 for cg23130254) and nondiseased (FDR=0.006, p=2.35E-07 for cg03874199; FDR=0.054, p=4.54E-05 for cg23130254) TI tissues. The cg03874199 is located at 2 nt upstream of transcription start site, and cg23130254 is located at 131 nt in exon 1. These 2 CpG is within just 132 nt around the transcription start site of the gene. The finding raises very important questions regarding the role of the gene HOXD12 in IBD and how the methylation modification affects IBD pathogenesis. Currently no information is available.

In the 12 CpGs, none of them is associated with UC in diseased nor in non-diseased colon tissues. A weak positive signal from ISYNA1 cg00445824 in diseased colon tissues from UC patients (FDR=0.372, p=4.92E-03) observed here needs further study to confirm.

*CD-associated DNA methylation in diseased colon tissue of CD at colon patients*

*Table 4. Diseased associated DNA methylation in diseased tissues from CD-intestine patients (FDR<0.25, 12 CpGs)*

| | | Dis/nondis CD-colon | | | Dis/nondis CD-intest | | | Dis/nondis UC-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg00899659 | ZNF22 | 0.169 | 4.75E-05 | 0.187 | 0.050 | 1.36E-01 | 0.445 | 0.049 | 1.74E-01 | 0.528 |
| cg01081263 | SCUBE2 | 0.272 | 3.51E-05 | 0.187 | 0.139 | 2.84E-03 | 0.359 | -0.132 | 5.42E-02 | 0.400 |
| cg08300860 | LDB3 | 0.087 | 3.42E-05 | 0.187 | 0.023 | 4.72E-01 | 0.730 | 0.005 | 9.13E-01 | 0.965 |
| cg09155001 | CALHM1 | 0.041 | 3.91E-05 | 0.187 | -0.014 | 1.97E-01 | 0.503 | 0.000 | 9.77E-01 | 0.991 |
| cg20337106 | CENPQ | 0.057 | 1.84E-05 | 0.187 | 0.021 | 2.16E-01 | 0.521 | 0.022 | 8.41E-02 | 0.439 |
| cg21104946 | WNT7B | -0.063 | 4.21E-05 | 0.187 | 0.005 | 6.07E-01 | 0.817 | -0.030 | 3.44E-02 | 0.382 |
| cg22533573 | NA | -0.133 | 2.84E-05 | 0.187 | -0.026 | 5.80E-01 | 0.800 | 0.006 | 8.80E-01 | 0.952 |
| cg12762799 | CLEC4E | 0.085 | 7.91E-05 | 0.242 | 0.020 | 2.24E-01 | 0.531 | -0.045 | 1.23E-01 | 0.481 |
| cg13926569 | PAPSS2 | 0.050 | 7.14E-05 | 0.242 | 0.012 | 3.85E-01 | 0.668 | 0.024 | 3.89E-01 | 0.685 |

*UC-colon associated DNA methylation*

*Table 5. Diseased associated DNA methylation in diseased tissues from UC-colon patients (FDR=0.372, rawp<0.0005, 17 CpGs)*

| | | Dis/nondis UC-colon | | | Dis/nondis CD-intest | | | Dis/nondis CD-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg13055278 | ID2 | 0.199 | 3.35E-05 | 0.372 | 0.042 | 1.92E-01 | 0.499 | 0.017 | 7.54E-01 | 0.372 |
| cg16753209 | NEB | 0.072 | 4.33E-05 | 0.372 | 0.056 | 1.33E-02 | 0.359 | 0.011 | 6.64E-01 | 0.372 |
| cg20162076 | LY86 | 0.079 | 5.45E-05 | 0.372 | 0.006 | 6.74E-01 | 0.856 | 0.019 | 2.65E-01 | 0.372 |
| cg07888040 | GLDN | 0.073 | 8.30E-05 | 0.372 | 0.021 | 3.36E-01 | 0.631 | -0.003 | 8.98E-01 | 0.372 |
| cg18959422 | MYBPH | 0.109 | 1.03E-04 | 0.372 | -0.018 | 4.60E-01 | 0.722 | -0.028 | 6.61E-01 | 0.372 |
| cg21865006 | ZCCHC14 | 0.071 | 1.39E-04 | 0.372 | 0.003 | 7.87E-01 | 0.911 | -0.006 | 7.48E-01 | 0.372 |
| cg17489451 | CCDC91 | 0.123 | 1.40E-04 | 0.372 | 0.070 | 2.90E-02 | 0.359 | -0.021 | 5.58E-01 | 0.372 |
| cg23152755 | DEFB105A | 0.066 | 1.90E-04 | 0.372 | -0.001 | 9.69E-01 | 0.989 | 0.003 | 8.82E-01 | 0.372 |
| cg26928972 | CSTA | 0.090 | 2.76E-04 | 0.372 | 0.024 | 2.55E-01 | 0.560 | 0.032 | 1.42E-01 | 0.372 |
| cg20025970 | ABCC9 | 0.106 | 3.43E-04 | 0.372 | 0.013 | 4.94E-01 | 0.744 | 0.017 | 5.60E-01 | 0.372 |
| cg08465774 | IDO1 | 0.116 | 3.45E-04 | 0.372 | 0.026 | 2.44E-01 | 0.550 | -0.022 | 4.24E-01 | 0.372 |
| cg19447966 | TEAD1 | 0.077 | 4.18E-04 | 0.372 | 0.045 | 8.33E-03 | 0.359 | -0.003 | 9.09E-01 | 0.372 |
| cg07456201 | UMOD | 0.069 | 4.25E-04 | 0.372 | 0.018 | 4.16E-01 | 0.691 | -0.058 | 1.26E-01 | 0.372 |
| cg13681935 | ASIP | 0.056 | 4.57E-04 | 0.372 | -0.002 | 9.01E-01 | 0.962 | -0.021 | 5.01E-01 | 0.372 |
| cg23073111 | ESRRA | 0.132 | 4.76E-04 | 0.372 | 0.051 | 8.18E-02 | 0.391 | -0.017 | 6.65E-01 | 0.372 |
| cg03019000 | TEX264 | 0.089 | 4.89E-04 | 0.372 | -0.003 | 8.40E-01 | 0.935 | -0.026 | 4.91E-01 | 0.372 |
| cg11098259 | AQP9 | 0.102 | 4.98E-04 | 0.372 | 0.017 | 3.40E-01 | 0.634 | -0.001 | 9.63E-01 | 0.372 |

Table 5 lists the identified 5 UC-associated CpGs/genes. These genes function in cell adhesion for GLDN and MYBPH, inflammatory response for LY86, transcriptional regulation for ID2, and muscle contraction for NEB.

All these 5 UC-associated CpGs in diseased colon are also observed in nondiseased colon tissue, FDR<0.1, p<0.01. However, none of them is observed in diseased and nondiseased TI tissues from CD patients.

*2. Disease-associated DNA Methylation in Non-diseased Tissues*

The profile of DNA methylation shows a remarkable difference between nondiseased tissues from IBD patients with healthy control intestinal tissues. The minimal FDR for nondiseased tissues is as low as 0.006 for TI tissues of CD patients and 0.003 for colon tissues of UC patients. Considering the comparison is between the tissues from IBD patients with the same type of tissue from unrelated controls, we use FDR<0.03, p≤0.00005 as threshold for diseased-associated methylation identification for both CD at TI and UC at colon.

*CD-associated DNA methylation in nondiseased TI of CD at TI patients*

*Table 6. Diseased associated DNA methylation in non-diseased tissues from CD-intestine patients (FDR<0.03, 7 CpGs)*

| | | Nondis/health CD-intest | | | Nondis/health CD-colon | | | Nondis/health UC-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg03874199 | HOXD12 | -0.328 | 2.35E-07 | 0.006 | -0.169 | 2.81E-02 | 0.232 | 0.066 | 3.41E-01 | 0.517 |
| cg23002761 | FBLIM1 | -0.260 | 6.31E-07 | 0.009 | -0.022 | 5.63E-01 | 0.843 | 0.115 | 6.88E-02 | 0.225 |
| cg23855989 | AQP5 | -0.229 | 2.01E-06 | 0.018 | 0.003 | 9.71E-01 | 0.993 | 0.095 | 1.89E-01 | 0.373 |
| cg07905963 | CYP2A13 | 0.167 | 3.82E-06 | 0.026 | 0.036 | 2.60E-01 | 0.6502 | -0.048 | 1.60E-01 | 0.340 |
| cg04660410 | VILL | -0.135 | 6.13E-06 | 0.029 | -0.04 | 5.10E-02 | 0.3182 | 0.045 | 1.50E-01 | 0.330 |
| cg05590257 | PLD6 | -0.319 | 7.26E-06 | 0.029 | -0.13 | 6.37E-02 | 0.3566 | -0.111 | 5.28E-02 | 0.200 |
| cg19728577 | GUCA2B | 0.125 | 6.68E-06 | 0.029 | 0.041 | 3.85E-02 | 0.2757 | -0.044 | 4.97E-02 | 0.196 |

From nondiseased TI of CD patients we identified 7 disease associated CpGs (Table 6). As described above, HOXD12 is associated with CD in nondiseased tissues. None of these CpGs is associated with UC in diseased and nondiseased colon tissues and two groups of IBD in nondiseased tissues.

*CD-associated DNA methylation in nondiseased colon of CD at colon patients*

*Table 7. Diseased associated DNA methylation in non-diseased tissues from CD-colon patients (FDR<0.03, top 30 from 365 CpGs)*

| | | Nondis/health CD-colon | | | Nondis/health CD-intest | | | Nondis/health UC-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg00412772 | C19orf33 | -0.171 | 1.13E-06 | 0.008 | -0.066 | 1.52E-01 | 0.674 | -0.098 | 4.99E-04 | 0.045 |
| cg04601137 | ADAMTSL5 | -0.148 | 2.51E-06 | 0.008 | 0.009 | 7.82E-01 | 0.989 | -0.063 | 8.17E-02 | 0.245 |
| cg07665060 | C19orf33 | -0.246 | 2.04E-06 | 0.008 | 0.035 | 4.97E-01 | 0.941 | -0.115 | 9.67E-03 | 0.099 |
| cg10987882 | RNF24 | -0.158 | 2.71E-06 | 0.008 | -0.042 | 7.51E-02 | 0.521 | -0.123 | 2.46E-03 | 0.064 |
| cg14203179 | NA | -0.120 | 1.76E-06 | 0.008 | 0.017 | 4.70E-01 | 0.929 | -0.092 | 8.10E-04 | 0.048 |
| cg15046693 | CEBPG | -0.237 | 3.36E-07 | 0.008 | -0.041 | 3.76E-01 | 0.892 | -0.130 | 8.36E-03 | 0.093 |
| cg16585619 | KRT19 | -0.144 | 1.85E-06 | 0.008 | -0.019 | 5.01E-01 | 0.941 | -0.106 | 1.63E-04 | 0.033 |
| cg16873863 | SLC22A18 | -0.155 | 2.22E-06 | 0.008 | -0.016 | 6.57E-01 | 0.978 | -0.050 | 3.57E-02 | 0.169 |
| cg23276115 | GRAP | -0.213 | 1.28E-06 | 0.008 | 0.013 | 7.30E-01 | 0.986 | -0.076 | 8.79E-03 | 0.095 |
| cg22289115 | CDHR5 | -0.133 | 3.22E-06 | 0.009 | 0.007 | 8.59E-01 | 0.998 | -0.122 | 6.08E-04 | 0.047 |
| cg19814116 | KCNAB2 | -0.270 | 4.62E-06 | 0.011 | -0.054 | 3.90E-01 | 0.897 | -0.097 | 4.84E-03 | 0.078 |
| cg26119796 | NA | -0.103 | 4.84E-06 | 0.011 | -0.041 | 2.00E-01 | 0.737 | -0.043 | 4.33E-02 | 0.184 |
| cg08314660 | PKP3 | -0.136 | 5.32E-06 | 0.011 | 0.013 | 6.67E-01 | 0.980 | -0.083 | 2.36E-03 | 0.064 |
| cg12010995 | CYP7A1 | -0.131 | 5.85E-06 | 0.012 | -0.001 | 9.57E-01 | 0.998 | -0.102 | 4.19E-05 | 0.022 |
| cg01169624 | SGCE | -0.116 | 6.44E-06 | 0.012 | 0.025 | 2.61E-01 | 0.800 | -0.083 | 5.30E-05 | 0.026 |
| cg23051598 | NA | -0.264 | 6.86E-06 | 0.012 | -0.031 | 2.57E-01 | 0.796 | -0.210 | 3.05E-05 | 0.022 |
| cg15452204 | CDX1 | -0.164 | 7.70E-06 | 0.012 | 0.046 | 2.14E-01 | 0.753 | -0.099 | 7.57E-04 | 0.048 |
| cg16075940 | NA | -0.167 | 7.91E-06 | 0.012 | -0.049 | 2.60E-01 | 0.798 | -0.094 | 1.55E-03 | 0.056 |
| cg22161874 | TRPM6 | -0.141 | 8.63E-06 | 0.013 | -0.016 | 5.88E-01 | 0.967 | -0.119 | 2.93E-05 | 0.022 |
| cg00702231 | PLAGL1 | -0.070 | 3.14E-05 | 0.013 | 0.014 | 4.12E-01 | 0.907 | -0.058 | 3.17E-05 | 0.022 |
| cg01584473 | NA | -0.179 | 1.42E-05 | 0.013 | -0.090 | 3.46E-02 | 0.385 | -0.030 | 2.04E-01 | 0.391 |
| cg02196805 | CSF2 | -0.125 | 2.85E-05 | 0.013 | 0.019 | 6.64E-01 | 0.979 | -0.090 | 2.58E-03 | 0.065 |
| cg02523617 | UBE2K | -0.078 | 1.07E-05 | 0.013 | -0.015 | 4.15E-01 | 0.908 | -0.014 | 5.85E-01 | 0.721 |
| cg02611863 | L3MBTL1 | -0.099 | 2.23E-05 | 0.013 | -0.027 | 2.62E-01 | 0.801 | -0.076 | 4.09E-05 | 0.022 |
| cg02855426 | CPA5 | -0.110 | 2.65E-05 | 0.013 | 0.018 | 5.50E-01 | 0.955 | -0.097 | 1.46E-07 | 0.003 |
| cg03085377 | NA | -0.169 | 1.60E-05 | 0.013 | -0.014 | 7.45E-01 | 0.986 | -0.029 | 7.51E-02 | 0.235 |
| cg03887787 | CTSW | -0.138 | 2.45E-05 | 0.013 | -0.026 | 5.71E-01 | 0.962 | -0.093 | 1.37E-02 | 0.112 |
| cg04947157 | TMC6 | -0.266 | 3.01E-05 | 0.013 | 0.035 | 5.45E-01 | 0.954 | -0.106 | 9.54E-03 | 0.099 |
| cg05790038 | TRPV3 | -0.245 | 1.27E-05 | 0.013 | 0.033 | 5.83E-01 | 0.966 | -0.126 | 7.14E-03 | 0.088 |
| cg06148175 | ACY3 | -0.198 | 1.21E-05 | 0.013 | 0.012 | 7.49E-01 | 0.986 | -0.095 | 8.38E-03 | 0.093 |

*UC-colon associated DNA methylation*

*Table 8. Diseased associated DNA methylation in non-diseased tissues from UC-colon patients (FDR<0.03, top30 from 95 CpGs)*

| | | Nondis/health UC-colon | | | Nondis/health CD-intest | | | Nondis/health CD-colon | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CpGs* | *genes* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* | *coef* | *rawp* | *FDR* |
| cg00519208 | EVI5 | -0.134 | 1.91E-07 | 0.003 | -0.002 | 9.46E-01 | 0.998 | -0.055 | 3.51E-02 | 0.262 |
| cg02855426 | CPA5 | -0.097 | 1.46E-07 | 0.003 | 0.018 | 5.50E-01 | 0.955 | -0.110 | 2.65E-05 | 0.013 |
| cg25608949 | TMPRSS3 | -0.091 | 5.97E-07 | 0.005 | 0.009 | 7.40E-01 | 0.986 | -0.055 | 2.37E-02 | 0.210 |
| cg06213598 | ABCA9 | -0.117 | 1.10E-06 | 0.008 | 0.003 | 9.20E-01 | 0.998 | -0.061 | 5.12E-03 | 0.094 |
| cg09360083 | THYN1 | -0.165 | 2.14E-06 | 0.008 | -0.018 | 5.49E-01 | 0.954 | -0.107 | 3.70E-04 | 0.030 |
| cg12532667 | EMCN | -0.122 | 2.10E-06 | 0.008 | 0.040 | 1.32E-01 | 0.642 | -0.081 | 4.59E-03 | 0.089 |
| cg17489451 | CCDC91 | -0.159 | 1.59E-06 | 0.008 | 0.010 | 7.35E-01 | 0.986 | -0.093 | 1.97E-03 | 0.059 |
| cg06613840 | GPX1 | -0.173 | 3.32E-06 | 0.009 | 0.012 | 7.49E-01 | 0.986 | -0.174 | 2.86E-05 | 0.013 |
| cg08465774 | IDO1 | -0.167 | 3.63E-06 | 0.009 | 0.031 | 2.66E-01 | 0.805 | -0.119 | 8.37E-04 | 0.041 |
| cg11698653 | ASB5 | -0.087 | 3.52E-06 | 0.009 | 0.034 | 1.03E-01 | 0.589 | -0.053 | 1.31E-03 | 0.050 |
| cg14349467 | PEBP1 | -0.124 | 4.58E-06 | 0.009 | 0.002 | 9.63E-01 | 0.998 | -0.100 | 1.18E-03 | 0.048 |
| cg15583072 | IAPP | -0.110 | 3.79E-06 | 0.009 | 0.016 | 5.59E-01 | 0.958 | -0.071 | 4.02E-03 | 0.084 |
| cg20423977 | PLAC4 | -0.128 | 4.42E-06 | 0.009 | 0.014 | 6.42E-01 | 0.974 | -0.111 | 1.74E-03 | 0.056 |
| cg21845297 | OTOR | -0.116 | 4.05E-06 | 0.009 | 0.008 | 7.34E-01 | 0.986 | -0.045 | 1.17E-02 | 0.145 |
| cg21663722 | NA | -0.089 | 4.92E-06 | 0.009 | 0.002 | 9.13E-01 | 0.998 | -0.040 | 2.91E-02 | 0.236 |
| cg22066521 | ABCB5 | -0.111 | 5.43E-06 | 0.009 | 0.024 | 3.43E-01 | 0.871 | -0.064 | 4.73E-03 | 0.091 |
| cg20950277 | TNIP3 | -0.250 | 6.39E-06 | 0.010 | 0.018 | 7.24E-01 | 0.986 | -0.134 | 6.34E-04 | 0.036 |
| cg04721098 | CACNG3 | -0.156 | 1.01E-05 | 0.014 | 0.086 | 5.86E-03 | 0.238 | -0.051 | 6.92E-02 | 0.370 |
| cg09111917 | TGM4 | -0.108 | 9.25E-06 | 0.014 | 0.000 | 9.94E-01 | 0.999 | -0.038 | 9.25E-02 | 0.427 |
| cg17439694 | CP | -0.126 | 9.89E-06 | 0.014 | 0.013 | 6.29E-01 | 0.973 | -0.054 | 6.83E-02 | 0.368 |
| cg13396068 | DCD | -0.112 | 1.26E-05 | 0.017 | 0.019 | 5.03E-01 | 0.941 | -0.060 | 1.72E-02 | 0.178 |
| cg15627025 | KIR3DL1 | -0.031 | 1.37E-05 | 0.017 | 0.019 | 9.14E-02 | 0.560 | -0.013 | 1.29E-01 | 0.501 |
| cg03202897 | FAM50B | -0.088 | 1.68E-05 | 0.019 | 0.014 | 4.30E-01 | 0.917 | -0.062 | 3.84E-04 | 0.030 |
| cg16118817 | FEM1A | -0.126 | 1.68E-05 | 0.019 | 0.031 | 1.54E-01 | 0.676 | -0.089 | 3.47E-04 | 0.029 |
| cg13921319 | PTMA | -0.170 | 1.82E-05 | 0.020 | 0.01 | 7.99E-01 | 0.9906 | -0.16 | 3.52E-05 | 0.013 |
| cg13944141 | PRSS2 | -0.152 | 2.01E-05 | 0.021 | -0 | 9.50E-01 | 0.9984 | -0.124 | 7.10E-04 | 0.038 |
| cg00702231 | PLAGL1 | -0.058 | 3.17E-05 | 0.022 | 0.014 | 4.12E-01 | 0.9075 | -0.07 | 3.14E-05 | 0.013 |
| cg02601403 | TBC1D3C | -0.106 | 3.30E-05 | 0.022 | 0.079 | 2.09E-03 | 0.2207 | -0.045 | 7.04E-02 | 0.374 |
| cg02611863 | L3MBTL1 | -0.076 | 4.09E-05 | 0.022 | -0.03 | 2.62E-01 | 0.8009 | -0.099 | 2.23E-05 | 0.013 |
| cg02764897 | KRTAP13-1 | -0.115 | 3.23E-05 | 0.022 | 0.04 | 8.56E-02 | 0.5474 | -0.076 | 1.53E-02 | 0.168 |

From nondiseased colon tissues of UC patients, as many as 55 CpG methylations are associated with UC (Table 8). Compared to the lowest FDR=0.372 from diseased colon tissue, there are 55 CpGs with FDR<0.03 in nondiseased colon tissues from UC patients, reflecting a big difference in DNA methylation between diseased and nondiseased colon.

These aberrant methylations have a broad range of functions that include 1) immune and

inflammation: IDO1, DCD, KIR3DL1, PEM1A, TNIP3; 2) transport: ABCA9, ABCB5, SLC16A7, solute carrier family 16, member 7, for pyruvate transport, CACNG3, and TRPM6, transient receptor potential cation channel, subfamily M, member 6, for ion transport. 3) transcriptional regulation: ESRRA, PTMA, PLAGL1, pleiomorphic adenoma gene-like 1; zinc finger protein, L3MBTL1, Lethal(3)malignant brain tumor-like protein 1; 4) cell adhesion: ENCN, IAPP, OTOR; 5) signaling and activation: TBC1D3C, ASIP, CP, GPX1 and PEBP1; 6) cell cycle, growth, and apoptosis: EVI5, THYN1, and ANXA3; 7) protein processing: CUEDC1, CCDC91, ASB5, TMRPSS3, CAP5, FEM1A, TGM4, PRSS2 (Table 2).

In the top 10 genes list in Table 6, ABC9 and CCDC91 play a role in transmembrane transport and golgi protein transport, respectively; EVI5 and THYN1 are involved in cell cycle and apoptosis, respectively; CPA5, TMPRSS3, and ASB5 function in hydrolysis of proteins or protein processing; and others, EMCN for endothelial slalomucin, GPX1 for ROS signaling, and IDO1 for inflammatory response.

CPA5 is relevant to IBD, not only a top IBD-associated DNA methylated gene, but its SNP rs17164838 (C/G) had been identified to be associated with CD (p=2.8E-3) in a GWAS study from WTCCC. Another top methylated gene is EVI5 that was found to be significantly associated with multiple sclerosis. Whether there is an interaction between genetic variation and DNA methylation, and what the role of these factors is in IBD pathogenesis are currently unknown.

*3. Disease-associated DNA Methylation Across the Disease Subtypes: CD at TI, CD at Colon and UC Patients*

A. No common DNA methylation associated with all the three disease groups: CD at TI, CD at colon and UC patients
B. No common DNA methylation associated with both CD at TI and UC patients
C. One common disease-associated DNA methylation for disease groups CD at colon and CD at TI and 13 common disease-associated DNA methylation for CD at colon and UC

*Table 9 Common disease-associated DNA methylation for CD at colon and CD at TI and for CD at colon and UC*

| *Genes* | *Gene* | *coef* | *rwap* | *FDR* | *coef* | *rwap* | *FDR* |
|---|---|---|---|---|---|---|---|
| | | **CD-colon dis** | | | **CD-intestine nondis** | | |
| cg01081263 | SCUBE2 | 0.272 | 3.51E-05 | 0.187 | -0.148 | 1.12E-03 | 0.047 |
| | | **CD-colon nondis** | | | **UC-colon nondis** | | |
| cg06613840 | GPX1 | -0.174 | 2.86E-05 | 0.013 | -0.173 | 3.32E-06 | 0.009 |
| cg03202897 | FAM50B | -0.062 | 3.84E-04 | 0.030 | -0.088 | 1.68E-05 | 0.019 |
| cg16118817 | FEM1A | -0.089 | 3.47E-04 | 0.029 | -0.126 | 1.68E-05 | 0.019 |
| cg23073111 | ESRRA | -0.197 | 1.44E-05 | 0.013 | -0.161 | 3.12E-05 | 0.022 |
| cg13921319 | PTMA | -0.1604 | 3.52E-05 | 0.013 | -0.170 | 1.82E-05 | 0.020 |
| cg12010995 | CYP7A1 | -0.131 | 5.85E-06 | 0.012 | -0.102 | 4.19E-05 | 0.022 |
| cg01169624 | SGCE | -0.116 | 6.44E-06 | 0.012 | -0.083 | 5.30E-05 | 0.026 |
| cg23051598 | NA | -0.264 | 6.86E-06 | 0.012 | -0.210 | 3.05E-05 | 0.022 |
| cg22161874 | TRPM6 | -0.141 | 8.63E-06 | 0.013 | -0.119 | 2.93E-05 | 0.022 |
| cg00702231 | PLAGL1 | -0.070 | 3.14E-05 | 0.013 | -0.058 | 3.17E-05 | 0.022 |
| cg02611863 | L3MBTL1 | -0.099 | 2.23E-05 | 0.013 | -0.076 | 4.09E-05 | 0.022 |
| cg02855426 | CPA5 | -0.110 | 2.65E-05 | 0.013 | -0.097 | 1.46E-07 | 0.003 |
| cg09360083 | THYN1 | -0.107 | 3.70E-04 | 0.030 | -0.165 | 2.14E-06 | 0.008 |

*4. Annotation and Network of Methylated IBD-associated Genes*

In the 78 CpGs identified to be associated with IBD, 2 CpGs are located in HOXD12 (cg03874199 and cg23130254), 2 CpGs (cg00702231 and cg12757684) are located in PLAGL1. These 2 genes are very important in gene transcriptional regulation.

Two CpGs cg03874199 and cg23130254 are located at the 3'-end of promoter region and exon 1 of the HOXD12 gene, respectively. The two CpGs are within only 131-nucleotides. A role of the aberrant methylation sites is currently unknown.

PLAGL1 (pleiomorphic adenoma gene-like 1; zinc finger protein) plays a role in transcriptional regulation. Two CpGs are associated with UC in nondiseased colon tissues, cg12757684 is located 11 bp 5' upstream and cg00702231 in the exon 1. The PLAGL1 gene is on the anti-sense strand. The sense strand of this region encodes another gene HYMAI (hydatidiform mole associated and imprinted), the cg00702231 is located in exon 2 of the non-coding RNA of the HYMAI. HYMAI exhibits differential DNA methylation between the two parental alleles at an adjacent CpG island, and is expressed only from the paternal allele. It is believed to be one of the causative genes for transient neonatal diabetes mellitus. Interestingly both PLAGL1 and HYMAI are imprinted. Whether the CpG cg00702231 methylation affects both gene expressions is unknown.

In addition to PLAGL1, FAM50B gene is also imprinted. Another important gene with disease-associated CpG methylation is CAP5. The CPA5 gene (rs17164838) was associated with CD (p=2.8E-3, WTCC). CPA5 is a metallocarboxypeptidase.

Among these 78 IBD-associated methylated CpGs, 18 are from CD-TI (diseased 11, nondiseased 7) and 59 from UC-colon (diseased 5, nondiseased 54); and 16 genes from diseased tissues (CD-TI 11, UC-colon 5) and 61 from nondiseased tissues (CD-TI 7, UC-colon 54). The results clearly indicate that the distribution of disease-associated methylation is much higher in UC-colon than CD-TI, in nondiseased tissues next to disease area than diseased tissues. Further looking at the methylation status of the identified methylation sites in other tissues, if considering $p<0.05$ as significant, we found that only 23% (7 out of 18) of CpGs identified from CD-TI and 17% (10 out of 59) of CpGs from UC-colon were observed in UC-colon and CD-TI, respectively. Whether this is due to disease subtypes or tissue types is currently under study by comparing these two groups with CD at colon. However, as high as 50% (8 out of 16) of CpGs identified from diseased tissues, and 82% (50 out of 61) of CpGs from nondiseased tissues were observed in nondiseased and diseased tissues, respectively.

Based on gene function, the 72 genes can be divided into 11 groups: immune and inflammation (8 genes), cell signaling (11 genes), transcriptional regulation (7 genes), cell adhesion (6 genes), cellular transport (6 genes), RNA processing (2 genes), protein processing (12 genes), metabolism (6 genes), cell cycle, differentiation and apoptosis (4 genes), cytoskeleton (4 genes), and others (6 genes). Clearly the aberrant methylated genes are involved in various cellular processes, but the major roles are tissue function related (25%, 18 out of 72, including cellular transport and protein processing), regulatory (25%, 18 out of 72, including cellular signaling and transcriptional regulation), and pathological (19%,14 out of 72, including immune response and cellular adhesion).

*Findings and significance*

*DNA Methylation between CD-TI and UC-colon*

Although two major forms of IBD, CD and UC, have common features at the molecular level and in terms of the clinical conditions, there are differences. In the 99 confirmed IBD risk loci, only about one third are shared between CD and UC (3, 4). Epigenetic distribution between CD and UC has not been estimated. The results from this study indicate that disease associated methylation is highly variable between CD-TI and UC-colon, only 18 of the 78 disease-associated CpGs (23%) are shared for both CD-TI and UC-colon and the levels of the common CpGs are different. The differences between CD-TI and UC-colon may be not only due to disease subtypes, also tissue types and/or both. Further comparison to CD at colon may help to clarify this. However, these belong to the same IBD disease group sharing common phenotypes and they still share 23% of the total aberrant CpGs.

Other supportive evidence is from comparison of this study with our previous study (2). There are several notable differences between the two studies: 1) the study samples are two different sets of IBD patients. For the total sample number 14 CD and 13 UC patients in the present study, and 9 CD and 17 UC patients in the previous study; 2) in the previous study the nondiseased tissue are not perfectly matched diseased tissue in tissue type such as diseased TI vs. nondiseased colon, while all the samples are perfected matched in both disease subtype and tissue type; 3) using different microarray, Goldengate Beadarray (1505 CpGs from 807 genes) for the previous study and DNA methylation 27 Beadarray (27,578 CpGs from 14,475 genes) for this study. Among the 7 IBD-associated genes identified from previous study (5) for IBD, 5 genes are detected by the present study: SERPINA5 for UC-colon; BGN, GABRA5, and AATK for CD-TI; and STAT5 for UC-colon. TNFRSF1A gene is very weakly associated with UC-colon, and no disease association was found for MAPK10. The results from these two microarray assays confirm the common features of IBD-associated DNA methylation.

*DNA Methylation at the Early Stage of Disease*
Disease-associated DNA methylation was identified from surgical nondiseased tissues and the number of these aberrant methylations were much greater than from diseased tissues. One may argue that the identification is by comparing the nondiseased tissues from IBD patients to healthy intestinal tissues from unrelated individuals. The stringent threshold, FDR<0.03 and p<0.00005, was used for identification and majority (82%) of the CpGs identified from nondiseased tissues were observed in matched diseased tissues, strongly support the notion that disease-associated DNA methylation identified here occurs at the early stage of disease development when the tissue appears to be normal at the time of surgery.

*DNA Methylation and IBD Pathophysiology*
DNA methylation becomes an important field with potential impact on human disease and medicine. By regulating gene expression, it can regulate genome-wide gene function and is involved in the process of human health and disease (6). The functions of these genes include immunity and inflammation, transcriptional regulation, cell signaling, cell adhesion, cellular transport, and protein processing. The role of the methylation in these genes in intestinal physiology and IBD pathogenesis is largely unknown.

The potential reversibility of DNA methylation status makes it a viable target for the treatment of human diseases. Evidence indicates that aberrant methylation exists in IBD from both individual gene study and genome-wide search. The results from this study further showed that the epigenetic alteration is very complex even only considering the 2 disease subtypes and 2 tissue locations. As a new player in IBD pathogenesis, most of the disease-associated methylated CpGs/genes identified most of them have not been well studied for their function in human intestine and diseases. Of particular interest, disease-associated methylated gene CPA5 in UC-nondiseased colon has been shown to be CD-associated from a GWAS, in which CPA5 SNP rs17164838 (in intron 5) was associated with CD, suggesting a possible interaction between genetic variation and epigenetic alteration in intestinal biology and IBD etiology. The further characterization and function study of the identified IBD-associated DNA methylation will provide new biomarkers for diagnosis and targeted clinical treatment for IBD.

Published abstract for Aim 1

Lin, Z, Cappel JA, Hegarty JP, Farber PW, Chen X, Wang, Y, Kelly AA, Poritz LS, Koltun WA Inflammatory bowel disease-associated DNA methylation mapping in human intestine and B cells. *Gastroenterology* 2010 May 138(5) Suppl:A-207. Conference proceeding: S1777 Digestive Disease Week, May1-5, 2010, New Orleans, LA.

*Specific aim 2. To study the contribution of DNA methylation to gene expression, and thus to IBD pathogenesis in 4 genes.*

*Specific aim 2A: FUNCTION OF DNA METHYLATION IN IBD SUSCEPTIBILITY GENE NKX2-3*
NKX2-3 (NK2 transcription factor related, locus 3) belongs to the NKX family of vertebrate homeodomain-containing transcription factors that regulate cell-specific expression of genes

involved in maintenance and tissue differentiation [7,8 ]. NKX2-3 is principally expressed in subsets of cells of mesodermal origin in the gastrointestinal tract and spleen and has an important role in spleen and Peyer patch development [9,10 ]. Within the gastrointestinal tract, NKX2-3 is expressed in microvascular endothelial and smooth muscle cells of the lamina propria and submucosa. NKX2-3-deficient mice demonstrate abnormal development of the intestine and secondary lymphoid organs [11] as well as a loss of regional expression of endothelial cell adhesion molecule MAdCAM-1 that acts as co-receptor for lymphocyte homing [12, 13]. Further studies have implicated NKX2-3 as having a central role in the correct localization of lymphocytes in the spleen of mice for both B-cell maturation and T-cell-dependent immune responses [14], which underlines the influence of NKX2-3 on the overall intestinal inflammatory response.

In genome-wide association studies, several genetic variants of NKX2-3 have been shown to be associated with the two main subtypes of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC) [15-17]. The single nucleotide polymorphism (SNP) rs10883365 was the first genetic variant of NKX2-3 to be associated with CD and is located 4900 nt upstream of the NKX2-3 gene [15, 16]. Another IBD-associated genetic variation of NKX2-3, rs11190140, is located in the promoter region of the gene, 1117 nt upstream of the transcription start site (Figure 1A). This variant consists of a C to T polymorphism, with the T allele being over-transmitted in IBD populations, and the risk allele is in complete linkage disequilibrium with the risk allele at SNP rs10883365 ($r^2$ = 1.0 in controls as well as in CD and UC cases) [17]. The C allele of rs11190140 further represents a potential DNA methylation site because it is located in a CG dinucleotide of a CpG island.

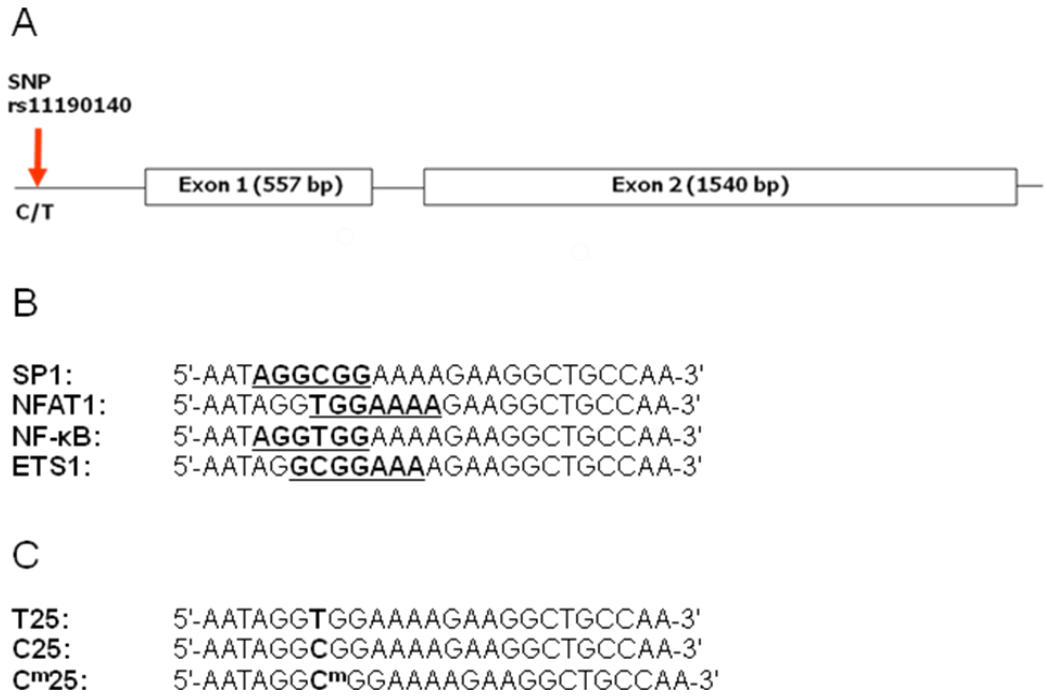


*Figure 1. Gene location of NKX2-3 variant rs11190140 and sequence for primer design and transcription factor selection.* (A) NKX2-3 variant rs11190140 is located in the promoter region of the gene, 1100 nt upstream of the transcription start site. (B) Predicted binding sites of transcription factors SP1, NFAT1, NF-κB, and ETS1. (C) Oligonucleotide sequences of the NKX2-3 promoter region containing the genetic variation of T (T25), non-methylated C (C25) or methylated C ($C^m25$) at rs11190140.

Although genetic predisposition has been demonstrated to be involved in the pathogenesis of IBD, the exact role of NKX2-3 and its genetic variations in this heterogeneous complex disease is still unclear. In our recent study, we replicated the association of SNP rs10883365 with CD in patients from a familial IBD registry from central Pennsylvania. Notably, mRNA and protein expression levels of NKX2-3 in those patients were found to be increased compared with non-CD siblings and healthy controls [18]. Since transcription factors regulate expression of downstream-located genes [19], abnormal levels of NKX2-3 could significantly alter gene expression and function in the intestine. Based upon sequence analysis of NKX2-3 rs11190140, potential transcription factors binding to the region of NKX2-3 rs11190140 were identified; specifically nuclear factor-kappaB (NF-κB), nuclear factor of activated T cells (NFAT1; also known as NFATC2), ETS1 (v-ets erythroblastosis virus E26 oncogene homolog 1, avian), and specificity protein 1 (SP1), (Figure 1B). We first aimed to examine the C allele methylation status, then to study the effect of the genetic variation and epigenetic modification of NKX2-3 rs11190140 on binding of transcription factors SP1, NFAT1, NF-κB, and ETS1.

## *Materials and methods*

### *Study Samples*

IBD patients: Peripheral blood was collected from these study participants and used to derive B cell lines by Epstein Barr virus (EBV) transformation and intestinal tissues harvested at the time of surgery. All human tissues described above were approved by the Human Subjects Protection Offices of The Pennsylvania State University College of Medicine, and were undertaken with the understanding and written consent of each subject.

### *DNA Isolation*

Genomic DNA from B cell lines and peripheral blood was isolated using a QIAamp DNA blood kit (Qiagen Inc., Valencia, CA), and DNA from intestinal tissues was isolated with a QIAamp DNA Mini Kit (Qiagen Inc.) according to the manufacturer's instructions. DNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technology, Wilmington, DE).

### *Genotype Analysis with PCR-based RFLP*

The SNP rs11190140 is a nature Eci I restriction fragment length polymorphism (RFLP). The genotype analysis was carried out with PCR-based RFLP methods [20]. A 225 bp fragment containing SNP rs11190140 was amplified from genomic DNA using primers NKX08s 5'-AGCTCAGAGTTTCCAGAATCTGTGG-3' (forward) and NKX07r 5'-AGTGTGGCACATACTGAGAGCAGGATC-3' (reverse) with the following PCR profile: 95ºC 2 min, 35 cycles of 95°C 30 sec, 58°C 1 min and 72°C 1 min, and 72°C 2 min. PCR products were digested with *Eci I* (New England Biolabs, Ipswich, MA), separated on 8% PAGE and visualized by ethidium bromide staining. The allelic products after digestion were: T allele 225 bp, C allele 153 and 72 bp. Due to its size, the 72 bp product ran out of the gel and could not be detected.

### *Bisulfite Conversion and DNA Methylation Analysis*

For DNA methylation analysis, bisulfite conversion of genomic DNA was performed using an EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's instructions. Briefly, methylated cytosines remain unchanged during bisulfite treatment, whereas non-methylated cytosines are

converted into uracil and can be detected as thymine following PCR [21]. The fragment containing SNP rs11190140 was amplified from bisulfite treated DNA using PCR primers NKX08sM 5'-AGTTTAGAGTTTTTAGAATTTGTGG-3' (forward) and NKX07rM 5'-AATATAACACATACTAAAAACAAAATC-3' (reverse), which are specific for the sense strand of bisulfite converted genomic DNA. The PCR profile was as follows: 95°C 2 min, 35 cycles of 95°C 30 sec, 58°C 1 min and 72°C 1 min, and 72°C 2 min. PCR products were also digested with restriction endonuclease *Eci I*, separated on 8% PAGE and visualized using ethidium bromide staining. The allelic products after digestion were: T and non-methylated C allele 225 bp, methylated C allele 153 and 72 bp. PCR products amplified from the T and non-methylated C allele sequence were not cut by the enzyme, whereas the methylated C allele sequence was cut by the enzyme. The ratio of methylated C to both methylated and non-methylated C was determined by analyzing specific gel bands using densitometry software ImageJ (NIH).

*Nuclear Extraction from B Cell Lines*

Immortalization of B-lymphocytes isolated from peripheral blood was achieved by Epstein-Barr virus (EBV) transformation using standard procedures as previously described [19]. EBV-transformed B cell lines were maintained in RPMI-1640 medium (ATCC, Manassas, VA) supplemented with 12% FBS (Gemini Bio-Products, West Sacramento, CA) and used for purification of nuclear extracts. To maximize transcription factor expression, B cells were activated by treatment with 100 nM PMA (phorbol 12-myristate-13-acetate) and 2 μM ionomycin for 30 minutes at 37°C, and nuclear extracts were isolated using a NE-PER kit (Thermo Scientific, Rockford, IL) according to the manufacturer's instructions. Total protein of nuclear extracts was determined by a BCA Protein Assay (Thermo Scientific) and a NanoDrop ND-1000 spectrophotometer (NanoDrop Technology).

*Transcription Factor Binding Assay, Biotin Pull-down, and Western Blot*

For transcription factor binding, 40 μg protein from B cell nuclear extracts were incubated with [200 ng/μl] biotin-labeled oligonucleotide sequences (IDT, Coralville, IA) of the NKX2-3 promoter region containing the genetic variation of T (T25), or non-methylated C (C25) or methylated C ($C^m25$) at rs11190140 (Figure 1C). The oligonucleotide-protein complexes were isolated using streptavidin-labeled Dynabeads M-280 (Invitrogen, Carlsbad, CA) and a magnetic bead separation system (Invitrogen). After the last wash, samples were mixed with loading buffer, boiled for 5 min, separated by PAGE on 8% Tris-glycine SDS gels (Invitrogen), and transferred onto nitrocellulose membranes. Membranes were blocked with 5% non-fat dry milk in TBS-Tween and probed with antibodies against transcription factors SP1, NFAT1, NF-κB (p65), and ETS1 (Santa Cruz Biotechnology, Santa Cruz, CA) at 4 °C overnight. Specific bands were visualized using enhanced chemiluminescence (Thermo Scientific) and developed onto x-ray films (Denville Scientific, Metuchen, NJ). The experiment was repeated three times.

*ChIP*

B cells (#033-01-1) were cross-linked with formaldehyde for 10 min, and chromatin immunoprecipitation (ChIP) was carried out as described previously [22] with the following modifications. The chromatin fractions from $4 \times 10^7$ cells were sonicated to achieve DNA fragmentation to fragment sizes ranging from 150 to 600 bp. The sheared chromatins were immunoprecipitated with 5 μg specific antibodies for NFAT (sc13033, Santa Cruz

Biotechnology), and normal rabbit IgG (#2729, Cell Signaling), overnight at 4 °C on a rotator. Immunoprecipitated chromatin was reverse protein/histone-DNA cross-linked by heating at 65ºC overnight. The DNA samples were recovered by QIAquick PCR purification kit (Qiagen, Valencia, CA). For PCR, 1 µl from a 20 µl immunoprecipitate was used as a template. The primer pairs used to amplify NKX2-3 were 08S (GCT CAG AGT TTC CAG AAT CTG) and 16r (TGACTGCATTGGTGGTGCGTGGCTG) located at up-stream -1270 to -1249 nt and -1042 to -1018 nt from the transcription site, respectively. The PCR fragment (252 bp) was located in the promoter region from -1270 to -1018. NKX2-3 SNP rs11190140 is located at -1117 nt, where the motif for NFAT binding is located.

*Results*

*Methylation Status of NKX2-3 rs11190140 in IBD Patients*

The methylation status of the C allele was examined in the 5 randomly chosen IBD cases by PCR-based RFLP with bisulfite converted DNA as described above and the results are shown on the right panel of the gel (Figure 2). In the PCR with bisulfite treated DNA, the methylated C will remain as C while unmethylated C will be changed to T. For samples 3 and 4 with the genotype CC, bisulfite treatment of DNA followed by PCR generates a CT indicating that the C allele in these samples is partially methylated.

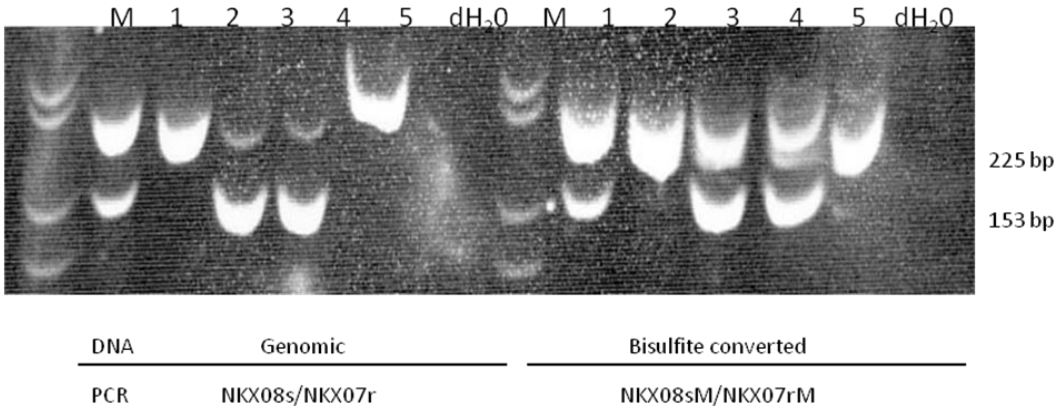


*Figure 2. Genotype and methylation status analysis of SNP rs11190140.* Genomic and bisulfite converted DNA from 5 different and randomly chosen IBD cases after digestion with restriction endonuclease *Eci I*, using two different primer setups specific for either genomic (NKX08s/NKX07r) or bisulfite converted DNA (NKX08sM/NKX07rM). M=marker; $dH_20$=negative control. The allelic products after digestion were: T allele 225 bp, C allele 153 and 72 bp for genomic DNA, and T and non-methylated C allele 225 bp, methylated C allele 153 and 72 bp for bisulfite converted DNA. Due to its size, the 72 bp product ran out of the gel and could not be detected.

To further confirm the above observation and evaluate the degree of the methylated C in IBD patients, we investigated C allele methylation status in 24 IBD cases (12 familial and 12 sporadic IBD patients) and 12 unrelated healthy controls that are homozygous C for rs11190490. We found that the C allele was methylated in all the samples. The methylation level was measured using a methylation ratio (β value, methylated C/ methylated + non-methylated C). An average of approximately 0.5 of β value was obtained in each group (Figure 3). These results indicate that the C allele of NKX2-3 rs11190140 is subject to methylation modification and widely occurs in the study samples from the Milton S. Hershey Medical Center. Although we did not observe a

difference between IBD cases and healthy controls, it may be due to the small number of the samples studied, the individuals heterozygous at rs11190490 have not been tested.

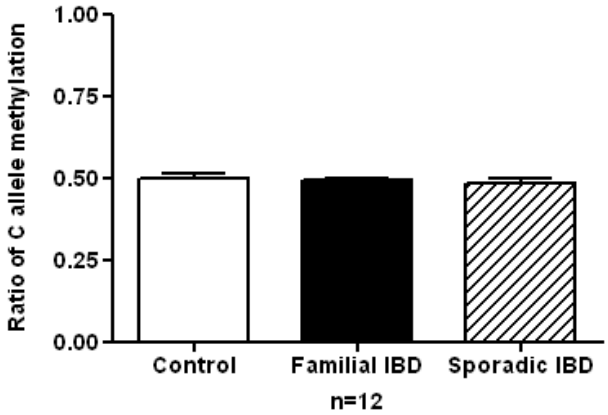


*Figure 3. Ratio of C allele methylation* (= methylated C/ methylated + non-methylated C) of SNP rs11190140 in unrelated healthy controls and IBD cases (familial and sporadic IBD patients) carrying the CC genotype (n=12).

*Effect of Genetic C/T Variation and C Methylation at NKX2-3 rs11190140 on Transcription Factor NFAT1 Binding in vitro*

To assess a functional effect of the genetic C/T variation and C methylation at NKX2-3 rs11190140, a binding assay for transcription factors SP1, NFAT1, NF-κB, and ETS1 was performed. Specific transcription factor binding motifs are present in the region of SNP rs11190140 (Figure 1b). The results of subsequent Western blot analysis are shown in Figure 4. Nuclear protein binding to NKX2-3 oligonucleotides containing the genetic variation of T, non-methylated C or methylated C was not evident for SP1, NF-kB, and ETS1. Only probing for NFAT1 indicated distinct and even differential binding of the different NKX2-3 oligonucleotides. The binding ability was increased for the oligonucleotides containing the genetic variation of C or methylated C.

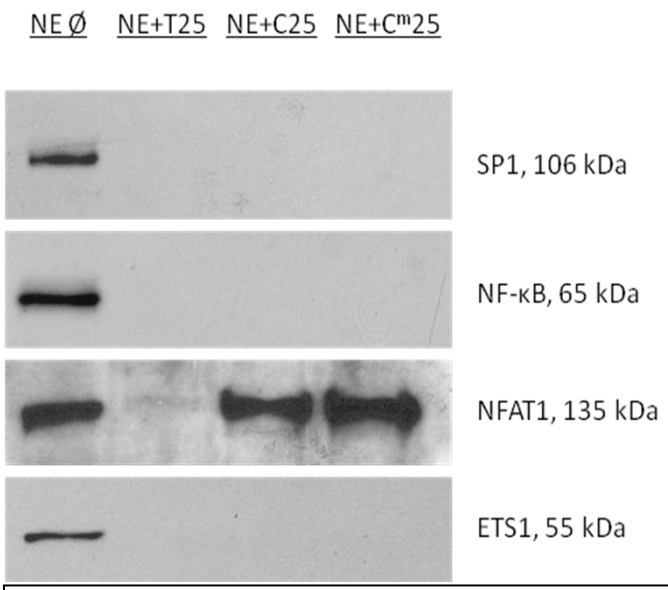


*Figure 4. Western blot analysis of transcription factors SP1, NFAT1, NF-κB, and ETS1 after transcription factor binding assay and biotin pull-down.* Biotin-labeled oligonucleotide sequences of the NKX2-3 promoter region containing the genetic variation of T (T25), non-methylated C (C25) or methylated C (C$^{m}$25) at rs11190140 were incubated with nuclear extracts (NE) from B cells. The oligonucleotide-protein complexes were isolated and subjected to Western blot analysis. Distinct and differential binding to T, non-methylated C or methylated C motifs was observed for transcription factor NFAT1. NEØ=untreated control.

*Transcription Factor NFAT1 Binding to NKX2-3 Promoter Region Flanking rs11190140 in vitro*

To confirm that the endogenous NFAT1 proteins bind to the rs11190140 site of the NKX2-3 promoter region *in vivo*, ChIP assays were carried out. A pair of primers spanning the 252 bp sequences (from -1270 to -1018 nt) of rs11190140 (-1117 nt) region were employed in the PCR to measure the endogenous NFAT1 proteins bound to this region using 033-01-1 EBV-transformed B cell line. From the DNA fractions that were co-immunoprecipitated with NFAT1 antibody, the target sequence was strongly amplified (Figure 5). The results confirmed that transcription factor NFAT1 binds to the region containing rs11190140 and the NFAT1 binding motif.

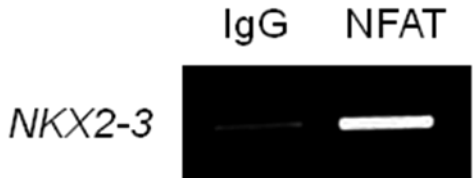


*Figure 5. Binding of NFAT1 protein to the NKX2-3 promoter in vivo.* Chromatin immunoprecipitation (ChIP) assay was performed from 033-01-1 cells. ChIP assay was carried out using anti-NFAT1 antibody and anti-IgG antibody as control. Promoter fragments present in the immunoprecipitates were amplified by PCR using pairs of primers that cover the NKX2-3 promoter region (from -1270 to -1018 nt).

*Findings and significance*

In the present study, we replicate and confirm the genetic association of homeodomain-containing transcription factor NKX2-3 variant rs11190140 with IBD in an IBD population from central Pennsylvania. We found that the C allele of the SNP was subjected to DNA methylation modification and further demonstrate that the genetic variation of C or T, and epigenetic alteration of methylated C and unmethylated C affects binding of transcription factor NFAT1 to the sequence of NKX2-3 containing rs11190140. Our data suggest that genetic variation and epigenetic alteration of rs11190490 play a role in NFAT binding to the gene promoter region and potentially influence NKX2-3 gene transcription.

*NKX2-3 Genetic Association, Methylation and Functional Implication*

Genetic predisposition of multiple genes and environmental factors are believed to be involved in the pathogenesis of IBD [23-25]. Up to now, over 50 IBD-associated genes/loci have been identified [26]. Among those genes, NKX2-3 and its genetic variations have been associated with CD and UC in large study populations using genome-wide association studies [15-18]. Notably, the risk alleles of the two NKX2-3 variants rs10883365 and rs11190140 identified as general IBD susceptibility loci have also been described to be in complete linkage disequilibrium [18]. Recently, we replicated the association of SNP rs10883365 with CD in 75 patients, 137 non-CD family members and 118 unrelated healthy controls from a familial IBD registry from central Pennsylvania [19]. Here, we analyzed the genetic association of NKX2-3 variant rs11190140 with IBD using a study group of 233 IBD patients and 250 unrelated healthy individuals from central Pennsylvania. The results indicated that the T allele of NKX2-3 rs11190140 was significantly associated with IBD in case-control analysis (OR=0.5033, 95% CI=0.3300-0.7636, $p<0.001$).

However, the precise role of IBD-associated genetic variations and their functional implications for IBD pathogenesis are still unknown and could include either direct gene effects or an altered regulation of downstream-located genes. mRNA and protein expression levels of NKX2-3 were found to be increased in CD patients compared with non-CD siblings and healthy controls [19]. Since transcription factors regulate expression of downstream-located genes [20], abnormal levels of NKX2-3 in the intestine could significantly alter expression and function of genes that may be involved in inflammatory processes as seen in IBD. For example the expression of NKX2-3-regulated endothelial cell adhesion molecule MAdCAM-1, which is restricted to mucosal tissues in the adult, has been shown to be up-regulated within focal sites of intestinal inflammation in both animal models of IBD [27] and in human tissue samples from IBD patients [28]. This may alter maturation and migration of antigen-responsive lymphocytes to the gut, affecting the intestinal inflammatory response. Furthermore, recent results from NKX2-3 knockdown human B cell lines identified 138 genes regulated by NKX2-3, including genes directly involved in endothelial cell homeostasis, activation and inflammatory responses, e.g. endothelin-1 (EDN1) and the myocyte enhancer factor 2 (MEF2)/Kruppel-like factor 2 (KLF2) signaling pathway [29], suggesting an implication of NKX2-3-mediated gene regulation in IBD pathogenesis.

The NKX2-3 variant rs11190140 is suitable for a combined genetic association and functional study due to its location in the promoter region of the gene [18], suggesting a potential involvement in transcriptional regulation. This suggests that the genetic variation of C or T, and

epigenetic alteration via C allele methylation at rs11190140 might affect binding of NFAT1 to the promoter region of NKX2-3 and alter its transcriptional regulation. Moreover, we demonstrated identical levels of methylation modification at rs11190140 in homozygous C allele carriers from our IBD-case and control populations. However, we only analyzed a small number of the samples studied, while the individuals heterozygous at rs11190490 have not been tested. Extensive evidence demonstrates that CpG methylation near the transcription start site of the promoter region plays a crucial role in regulating gene expression, for example by affecting the binding of transcription factors and CpG-binding proteins to DNA [30]. Altered genome-wide function due to epigenetic modification is believed to be involved in the development of human diseases [31]. Therefore, we assessed binding of transcription factors SP1, NFAT1, NF-κB, and ETS1 to the promoter region of NKX2-3 at rs11190140. Among the four tested transcription factors potentially regulating NKX2-3 gene expression, only NFAT1 exhibited distinct and differential binding to oligonucleotides containing rs11190140. Interestingly, binding of NFAT1 to oligonucleotides containing the mutation allele T was dramatically decreased compared to the non-methylated C or methylated C allele. Moreover, the binding ability of NFAT1 even slightly increased with C allele methylation compared to non-methylated C, indicating that both genetic variation and epigenetic modification at rs11190140 might affect NKX2-3 gene expression. The above observation of NKX2-3 binding *in vitro* to NKX2-3 promoter has been confirmed *in vivo* by ChIP assay. The further studies on the effects of NFAT1 binding on NKX2-3 transcription will help understanding the role of NKX2-3 in IBD pathogenesis in intestinal microvascular endothelial cells.

*NFAT Transcription Factor in Immune Response and Differentiation*

The immune response is coordinated by an interacting network of transcription factors that dictate gene expression of various cytokines amongst others [20]. Consisting of four distinct classes, NFAT1-4, calcineurin-dependent NFAT proteins were first identified in activated T cells to induce gene transcription of different cytokines during the immune response [32]. They have since been shown to play a role in other classes of immune and non-immune cells, including smooth muscle, endothelial and cardiac cells [33,34]. In this context, NFAT regulates peripheral vascular development during angiogenesis, which involves proliferation, differentiation and migration of endothelial and smooth muscle cells [35,36]. Furthermore, vascular endothelial growth factor (VEGF)-mediated induction of angiogenesis in endothelial cells requires activation of NFAT [37,38]. In lymphatic endothelial cells, NFAT is an important regulator of lymphangiogenesis [33]. During cardiac cell differentiation and development of the cardiovascular system, NFAT is directly involved in regulation of the cardiac-specific transcription factor NKX2-5 [34], the closest family homolog of NKX2-3. To exert its function, NFAT has also been shown to cooperate with several other transcription factors such as GATA4 [34] and AP-1 [35]. Studies in mutant mice carrying a disrupted NFAT1 gene indicate that NFAT1 is a positive regulator of cytokine expression, even though these were conducted in models of allergic inflammation [39, 40]. However, certain primary and secondary immune responses in these mice lacking the NFAT1 gene were enhanced, including increased intrapleural accumulation of eosinophils and increased serum IgE levels [40], and more efficient differentiation toward a Th2 phenotype in stimulated spleen cells [39]. Since mRNA and protein expression levels of NKX2-3 were also found to be increased in IBD patients [19], we speculate that NFAT – probably in coordination with other transcription factors – might exert an inhibitory

effect on NKX2-3 transcription, thereby regulating downstream-located genes involved in the intestinal inflammatory response.

Taken together, we replicated the genetic association of NKX2-3 variant rs11190140 with IBD in an IBD population from central Pennsylvania. We further demonstrated a functional implication of this genetic association as shown by altered binding of transcription factor NFAT1 to oligonucleotides containing the genetic variation of T, C or methylated C at rs11190140, which in turn could affect NKX2-3 gene expression. We speculate that altered genetic and epigenetic regulation of NKX2-3 might affect expression of NKX2-3 regulated genes involved in inflammatory processes and thus contribute to IBD pathogenesis. However, since the NKX2-3 risk allele at rs11190140 is in complete linkage disequilibrium with the risk allele at rs10883365, the question of the etiological variant still needs to be answered.

*Specific aim 2B: function study of DNA methylation in P2RX7 and PPAR-gamma in inflammatory bowel disease*

*PPARγ and P2RX7* were identified as disease-associated methylated genes in our previous study and this study**.** PPAR**γ,** peroxisome proliferator-activated receptor, is a nuclear transcription factor that forms a heterodimer with retinoid X receptor α and binds to DNA at a specific sequence, AGGTCAXAGGTCA, to facilitate gene transcription regulation of genes involved in lipid metabolism, insulin sensitization, inflammation, and cell proliferation. Recently, the PPAR**γ** receptor is thought to contribute to gut homeostasis regulation (41). P2RX7 is a two-transmembrane ionotropic receptor that has a ubiquitous distribution in nearly all tissues and organs of the body and is most highly expressed on immune cells of monocyte-macrophage origin. Activation of the receptor by extracellular ATP opens a cationic channel which gradually dilates to a large pore over tens of seconds. Activation of the P2X7 receptor is associated with a massive K+ efflux that serves as a cofactor for the secretion of pro-inflammatory interleukins (ILs) 1β and 18 from myeloid cells. Prolonged activation of P2X7 leads to apoptotic death of the cell. However, in the absence of ATP, P2X7 has an alternative function in the recognition and phagocytosis of non-opsonized foreign particles, including bacteria and apoptotic cell (42).

Our previous study also indicated that PPARG SNP, rs1801282, is missense mutation that changes amino acid #12 from Pro to Ala, and rs 2960422 is a variation in intron 1 region. The results showed that G allele of rs 2960422 was significantly associated with IBD, but no association for rs1801282 (unpublished data). Recently, several papers report that PPARG is associated with IBD.

The association of genetic variations and DNA methylation with IBD suggests an important role of PPARG and P2RX7 in IBD pathogenesis. In this study we proposed to investigate the expression of these disease-associated methylated genes in the intestinal tissues from IBD patients.

*Study samples and methods*

*Study Samples*

Fifty-two intestinal tissues (diseased and paired normal) from 26 IBD patients (9 CD; 17 UC) were used in this study. All tissues were collected at the time of surgery and classified by a

pathologist. The healthy intestinal tissues were from NRDI (National Disease Research Interchange, Philadelphia, PA).

Intestinal tissue samples were also collected from forty-seven individuals with IBD and ten healthy individuals with no disease. In those patients with IBD, samples were taken from diseased and adjacent non-diseased intestinal tissues. All tissues were collected at the time of surgery and classified by a pathologist.

All the human tissues described above were approved by the Human Subjects Protection Offices of The Pennsylvania State University College of Medicine prior to the beginning of this project, and were undertaken with the understanding and written consent of each subject.

*RNA isolation with QiaGen RNeasy® Mini kit*
RNA was isolated from B cells collected from blood samples of patients with inflammatory bowel disease and their healthy, non-IBD siblings that are subsequently transformed using Epstein-Barr Virus (EBV) in order to immortalize the B cells.

RNA was also isolated from intestinal tissue samples from areas of either Crohn's disease or ulcerative colitis as well as from areas of the GI tract that exhibit normal, non-diseased characteristics.

*Reverse Transcription*
RNA isolated with QiaGen RNeasy® Mini Kit is subjected to reverse transcription in order to synthesize cDNA.

Three micrograms of RNA is collected and combined with primer, 10 mM dNTP mix, and DEPC-treated water and then incubated at 65°C for 5 minutes. The mixture is then cooled on ice for at least 1 minute. A cDNA synthesis mix is made by combining 10x RT buffer, 25 mM $MgCl_2$, 0.1 M DTT, RNaseOUT™, Superscript™ III RT. The mix is added to each RNA and primer mixture and incubated with Oligo$(dT)_{20}$ for 50 minutes at 50°C. Heating the reaction to 85°C for 5 minutes terminates it and then the reaction is cooled again on ice. 1 μL of RNase H is added to each tube and incubated for 20 minutes at 37°C.

*PCR Amplification*
A master mixture is made from adding 1.5 μL Roche Expand Long Template PCR System buffer 1 (17.5 mM $MgCl_2$), 1.5 μL Roche Expand Long Template PCR System buffer 2 (27.5 mM $MgCl_2$), 2.0 μL PCR grade nucleotide mix (10 mM each of dATP, dCTP, dGTP, and dTTP), 0.2 μLTaq DNA polymerase, and 22.6 μL distilled water, and 0.6 μL each of forward and reverse primers amounting to 29 μL per reaction. 1.0 μL of cDNA synthesized through reverse transcription is then added to the master mix.

*Quantification of the Differential Significance using RT-qPCR*
The goal of using RT-qPCR is to quantify differences in gene expression between diseased and non-diseased samples and look for significance of expression differences that are regulated by DNA methylation.

*Gel Electrophoresis and Imaging*
Gel electrophoresis of the PCR reactions is conducted using 1% agarose gels. The gels are run for 35-40 minutes at 120 volts. A DNA ladder (New England BioLabs Quick-Load 100 bp DNA Ladder) is run alongside the samples to indicate the size of the banding. Imaging of the gel is performed using a Ultra-Violet Products Benchtop UV transilluminator.

*Data Analysis*
Images collected of the gel electrophoreses by using a Ultra-Violet Products Benchtop UV transilluminator were processed using Image J imaging processing and analysis software provided by the NIH as a free download on the website http://rsbweb.nih.gov/ij. The luminescence of each gene for each individual tissue sample was determined using Image J and compared to a standard GAPDH luminescence for the same individual tissue sample. A ratio was then established between the gene and GAPDH. The gene to GAPDH ratio of the individual's healthy tissue sample was then compared to that of the diseased tissue sample and another ratio was established between healthy and diseased tissue. Statistical analysis of this ratio was then used to determine associations.

## Results

*Association of DNA Methylation with IBD*
The disease-associated DNA methylation in PPARG and P2RX7 genes has been identified from the Illumina GoldenGate Bead Array (our previous study) and Illumina Human Methylation 27 (this study). The Illumina GoldenGate Bead Array contains 1505 CpG loci selected from 807 genes and the Illumina Human Methylation 27 Bead Array contains 27,578 CpG loci located within the promoter regions of transcription start sites of 14,475 genes.

*P2RX7 Methylation and a Suggestion of an Association with IBD*
At the P2RX7_P119_R CpG site, healthy patients who did not have inflammatory bowel disease had 48% of their sites methylated whereas in those patients with inflammatory bowel disease, only 20% of the sites were methylated. This difference of 28% was statistically significant with a p-value of 0.01.

*PPARy Methylation and a Suggestion of an Association with IBD*
An association of DNA methylation with IBD was also present with PPARy CpG sites from DNA methylation status analysis using the GoldenGate Bead Array (Table 10). At the PPARy_P693_F CpG site (Table 10), healthy patients who did not have inflammatory bowel disease had 76% of their sites methylated whereas in those patients with inflammatory bowel disease, only 53% of the sites were methylated. This difference of 23% was statistically significant with a p-value of 0.03.

*Table 10. GoldenGate Bead Array data illustrating difference in methylated CpG sites among various genes in healthy individuals vs individuals with IBD, PPAR-gamma and P2RX7 methylation as indicated*

| TargetID[a] | Diseased Mean[b] | Diseased SD | Normal Mean[b] | Normal SD | difference[b] | p |
|---|---|---|---|---|---|---|
| PECAM1_P135_F | 0.32 | 0.26 | 0.62 | 0.10 | 0.30 | 0.02 |
| POMC_P400_R | 0.11 | 0.13 | 0.40 | 0.21 | 0.29 | 0.00 |
| IL16_P226_F | 0.26 | 0.23 | 0.54 | 0.25 | 0.29 | 0.03 |
| P2RX7_P119_R | 0.20 | 0.14 | 0.48 | 0.21 | 0.28 | 0.01 |
| MXI1_P1269_F | 0.40 | 0.24 | 0.68 | 0.18 | 0.28 | 0.05 |
| LIG3_P622_R | 0.61 | 0.27 | 0.88 | 0.05 | 0.27 | 0.01 |
| MOS_E60_R | 0.20 | 0.27 | 0.46 | 0.27 | 0.26 | 0.04 |
| ERN1_P809_R | 0.18 | 0.20 | 0.44 | 0.15 | 0.26 | 0.01 |
| ELK3_P514_F | 0.08 | 0.10 | 0.34 | 0.30 | 0.26 | 0.03 |
| HLA_DOB_E432_R | 0.36 | 0.28 | 0.12 | 0.23 | 0.25 | 0.02 |
| PLAT_E158_F | 0.67 | 0.23 | 0.91 | 0.06 | 0.24 | 0.01 |
| MUC1_E18_R | 0.70 | 0.31 | 0.94 | 0.05 | 0.24 | 0.01 |
| ARHGDIB_P148_R | 0.27 | 0.18 | 0.51 | 0.13 | 0.24 | 0.01 |
| CD81_P272_R | 0.54 | 0.26 | 0.77 | 0.16 | 0.23 | 0.04 |
| PPARG_P693_F | 0.53 | 0.21 | 0.76 | 0.11 | 0.23 | 0.03 |
| SMARCB1_P220-R | 0.38 | 0.22 | 0.61 | 0.13 | 0.23 | 0.02 |
| RARA_P176_R | 0.33 | 0.25 | 0.56 | 0.17 | 0.23 | 0.05 |
| CRK_P721_F | 0.15 | 0.07 | 0.37 | 0.24 | 0.22 | 0.05 |
| IL12B_P392_R | 0.59 | 0.28 | 0.80 | 0.16 | 0.22 | 0.02 |
| SEPT9_P58_R | 0.72 | 0.30 | 0.93 | 0.06 | 0.21 | 0.03 |
| BAX_E281_R | 0.46 | 0.23 | 0.66 | 0.12 | 0.21 | 0.04 |
| PPAT_E170_R | 0.30 | 0.13 | 0.49 | 0.14 | 0.20 | 0.01 |
| TEK_E75_F | 0.27 | 0.22 | 0.46 | 0.15 | 0.19 | 0.02 |
| TBX1_P520_F | 0.41 | 0.19 | 0.61 | 0.08 | 0.19 | 0.02 |
| EPS8_P437_F | 0.10 | 0.11 | 0.29 | 0.20 | 0.19 | 0.02 |
| DHCR24_P652_R | 0.29 | 0.14 | 0.48 | 0.18 | 0.18 | 0.03 |
| TAL1_P594_F | 0.09 | 0.11 | 0.27 | 0.25 | 0.18 | 0.05 |
| EYA4_P794_F | 0.10 | 0.11 | 0.28 | 0.19 | 0.17 | 0.03 |
| RAP1A_P285_R | 0.26 | 0.16 | 0.42 | 0.18 | 0.17 | 0.05 |
| EPHB1_E202_R | 0.42 | 0.18 | 0.58 | 0.06 | 0.16 | 0.01 |

[a]ID for target gene and CpG
[b]β value

*Gene Expression in Intestinal Tissues from IBD Patients*

Samples from forty-seven patients with inflammatory bowel disease were taken from affected intestinal tissue and compared to samples of adjacent, non-diseased intestinal tissue from the same individual. In addition, intestinal tissue samples from ten healthy patients were taken from both the small bowel and large bowel.

After performing RT-PCR and gel electrophoresis for P2RX7 and PPARy using two primer pairs previously identified (Figures 6, 7, 8, and 9), the gels were analyzed for gene expression of both genes. Gene expression varied among the same individual between diseased intestinal tissues and adjacent non-diseased intestinal tissue.

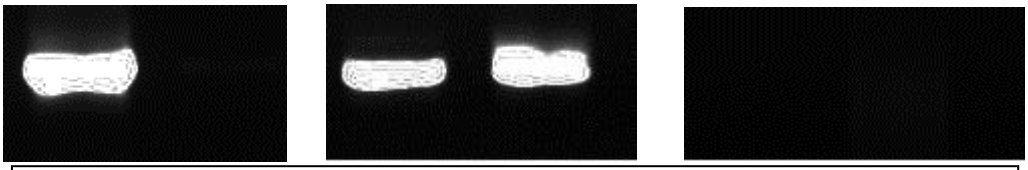

*Figure 6*. P2RX7 expression in intestinal tissues (PCR primers F1/R1).

In the first image, there is decreased P2RX7 gene expression in diseased intestinal tissue as compared to adjacent healthy intestinal tissue. The middle image shows an increased P2RX7 expression in the diseased intestinal tissue as compared to adjacent healthy intestinal tissue. The image on the right displays the same amount of P2RX7 gene expression between both diseased and non-diseased intestinal tissue.

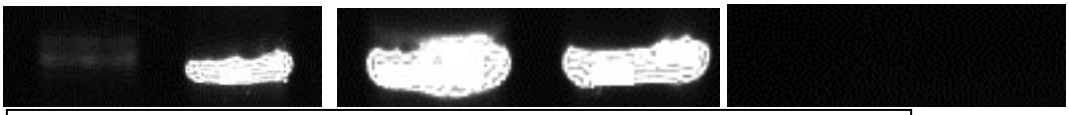

*Figure 7*. P2RX7 expression in intestinal tissues (PCR primers F3/R3).

In the first image, there is increased expression of P2RX7 in the diseased tissue sample as compared to an adjacent non-affected intestinal sample. The middle image is representative of decreased P2RX7 expression in the diseased sample. The image on the right is indicative of having similar gene expression between both samples.

*Figure 8*. PPARG expression in intestinal tissues (PCR primers F1/R1).

In the first image, there is an increased gene expression of PPAR-gamma in the diseased intestinal sample as compared to the adjacent non-affected intestinal tissue. The image on the right is indicative of having similar gene expression between both samples.

*Figure 9*. PPARG expression in intestinal tissues (PCR primers F3/R3).

The first image displays an increased gene expression of PPARy in the diseased intestinal tissue sample. The second image illustrates a decreased PPAR-gamma gene expression in the diseased tissue sample.

*Correlation of Gene Expression of PPARG and P2RX7 with Disease in IBD Patients*

Gene expression was quantified, as previously described, by quantifying luminescence using Image J and comparing the luminescence of the tissue sample to GAPDH expression, establishing a ratio between gene and GAPDH expression. This Gene: GAPDH ratio was then

compared between healthy and diseased tissue samples to look for an association between methylation of PPAR-gamma and/or P2RX7 and inflammatory bowel disease.

The tables 11 and 12 below are grouped by primer pair of each gene and compare healthy tissue against diseased tissue overall, healthy versus diseased tissue from the small bowel in patients with Crohn's disease, healthy versus diseased tissues from the colon in patients with Crohn's disease, healthy vs. diseased tissues from the colon in patients with ulcerative colitis, and tissues from healthy patients without IBD from their small bowel versus tissues from their colon. Table 11 is for the P2RX7 gene in IBD patients, a strong association was found when comparing healthy and diseased small bowel tissue samples in patients with Crohn's disease. In addition, statistically significant correlations were found between healthy and diseased colon samples from patients with Crohn's disease. These correlations were statistically significant regardless of primer pair.

In patients with ulcerative colitis, a statistically significant correlation between P2RX7 methylation and IBD was found when using the F1R1 primer but not the F3R3 primer. In addition, healthy individuals without IBD had a statistically significant correlation between P2RX7 methylation and differing locations of the bowel– small bowel and colon – when using the F1R1 primer but not the F3R3 primer.

*Table 11. Correlation of Gene Expression of P2RX7 with Disease in IBD Patients*

| Gene | Correlation | p-value |
|---|---|---|
| P2RX7 F1R1:GAPDH H vs. D | 0.26 | 0.078 |
| P2RX7 F1R1:GAPDH CD Small bowel H vs. D | 0.996 | <0.0001 |
| P2RX7 F1R1:GAPDH CD Colon H vs. D | 0.8 | 0.002 |
| P2RX7 F1R1:GAPDH UC colon H vs. D | 0.639 | 0.008 |
| P2RX7 F1R1:GAPDH Healthy Small Bowel vs. Healthy Colon | 0.798 | 0.006 |
| P2RX7 F3R3:GAPDH H vs. D | 0.334 | 0.022 |
| P2RX7 F3R3:GAPDH CD Small bowel H vs. D | 0.964 | <0.0001 |
| P2RX7 F3R3:GAPDH CD Colon H vs. D | 0.829 | 0.001 |
| P2RX7 F3R3:GAPDH UC Colon H vs. D | 0.367 | 0.162 |
| P2RX7 F3R3:GAPDH Healthy Small Bowel vs. Healthy Colon | 0.165 | 0.65 |

Table 12 is for the PPAR-gamma gene in IBD patients, a statistically significant correlation was found for PPAR-gamma methylation and IBD when comparing healthy and diseased colon tissue samples from individuals with Crohn's disease, regardless of primer pair. A statistically significant correlation was also found between PPAR-gamma methylation and IBD in samples from the small bowel in individuals with Crohn's disease when using PPAR-gamma F3R3 primer pair. In patients with ulcerative colitis, a statistically significant correlation between PPAR-gamma methylation and IBD was found when using the F1R1 primer but not the F3R3 primer.

In addition, healthy individuals without IBD had a statistically significant correlation between PPAR-gamma methylation and differing locations of the bowel– small bowel and colon – when using the F3R3 primer but not the F1R1 primer.

*Table 12. Correlation of Gene Expression of PPARG with Disease in IBD Patients*

| *Gene* | *Correlation* | *p-value* |
|---|---|---|
| PPARy F1R1:GAPDH H vs. D | -0.004 | 0.98 |
| PPARy F1R1:GAPDH CD Small bowel H vs. D | 0.204 | 0.402 |
| PPARy F1R1:GAPDH CD Colon H vs. D | 0.675 | 0.023 |
| PPARy F1R1:GAPDH UC colon H vs. D | 0.684 | 0.003 |
| PPARy F1R1:GAPDH Healthy Small Bowel vs. Healthy Colon | 0.415 | 0.233 |
| PPARy F3R3:GAPDH H vs. D | 0.225 | 0.133 |
| PPARy F3R3:GAPDH CD Small bowel H vs. D | 0.999 | <.0001 |
| PPARy F3R3:GAPDH CD Colon H vs. D | 0.985 | <.0001 |
| PPARy F3R3:GAPDH UC Colon H vs. D | -0.21 | 0.435 |
| PPARy F3R3:GAPDH Healthy Small Bowel vs. Healthy Colon | 0.749 | 0.013 |

*Findings and Significance*

1) DNA methylation in PPARG and P2RX7 is associated with IBD, indicated by both Illumina GoldenGate and Illumina Human methylation 27 Bead Arrays.
2) Both IBD associated methylated genes PPARG and P2RX7 are expressed in human small intestine and colon.
3) Gene expression of both PPARG and P2RX7 is correlated with IBD

Further study is needed to understand the mechanism how the DNA methylation alteration affects the expression of P2RX7 and PPAR-gamma and what a role of the gene regulation in the pathophysiology of inflammatory bowel disease. Based on the genetic association, DNA methylation association, and gene expression correlation with IBD, PPARG and P2RX7 most likely play an important role in IBD. Further investigation could lead to the discovery of potential targets for disease treatments and may provide a potential cure for the millions affected by this life-altering disease.

Published abstract for Aim 2

John G, Lin Z, Hegarty JP, Berg A, Pastor DM, Kelly AA, Wang Y, Poritz LS, Koltun WA NKX2-3 variant rs11190140 is associated with IBD and alters NFAT transcription factor binding. *Gastroenterology* 2010 May 138(5) Suppl:S-682. Conference proceeding: S1247 Digestive Disease Week, May1-5, 2010, New Orleans, LA.

REFERENCES:


1. Bibikova M, Lin Z, Zhou L, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*. 2006;16:383-393.
2. Lin Z, Hegarty JP, Cappel JA, et al. Identification of disease-associated DNA methylation in intestinal tissues from patients with inflammatory bowel disease. *Clin Genet*.80:59-67.

3. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43:246-252.
4. Thompson AI, Lees CW. Genetics of ulcerative colitis. *Inflamm Bowel Dis*. 2011;17:831-848.
5. Clark SJ, Harrison J, Frommer M. CpNpG methylation in mammalian cells. *Nat Genet*. 1995;10:20-27.
6. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol*. 2008;8:458-466.
7. Tanaka M, Kasahara H, Bartunkova S, Schinke M, Komuro I, Inagaki H, Lee Y, Lyons GE, Izumo S. Vertebrate homologs of tinman and bagpipe: roles of the homeobox genes in cardiovascular development. *Dev Genet*. 1998;22:239-249.
8. Stanfel MN, Moses KA, Schwartz RJ, Zimmer WE. Regulation of organ development by the NKX-homeodomain factors: an NKX code. *Cell Mol Biol (Noisy-le-grand)*. 2005;Suppl 51:OL785-799.
9. Buchberger A, Pabst O, Brand T, Seidl K, Arnold HH. Chick NKx-2.3 represents a novel family member of vertebrate homologues to the Drosophila homeobox gene tinman: differential expression of cNKx-2.3 and cNKx-2.5 during heart and gut development. *Mech Dev*. 1996;56:151-163.
10. Pabst O, Schneider A, Brand T, Arnold HH. The mouse Nkx2-3 homeodomain gene is expressed in gut mesenchyme during pre- and postnatal mouse development. *Dev Dyn*. 1997;209:29-35.
11. Pabst O, Zweigerdt R, Arnold HH. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development*. 1999;126:2215-2225.
12. Pabst O, Forster R, Lipp M, Engel H, Arnold HH. NKX2.3 is required for MAdCAM-1 expression and homing of lymphocytes in spleen and mucosa-associated lymphoid tissue, *EMBO J*. 2000;19:2015-2023.
13. Wang CC, Biben C, Robb L, Nassir F, Barnett L, Davidson NO, Koentgen F, Tarlinton D, Harvey RP. Homeodomain factor Nkx2-3 controls regional expression of leukocyte homing coreceptor MAdCAM-1 in specialized endothelial cells of the viscera. *Dev Biol*. 2000;224:152-167.
14. Tarlinton D, Light S, Metcalf D, Harvey RP, Robb L. Architectural defects in the spleens of Nkx2-3-deficient mice are intrinsic and associated with defects in both B cell maturation and T cell-dependent immune responses. *J Immunol*. 2003;170:4002-4010.
15. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678.
16. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Gene*t. 2008;40:710-712.

17. Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M, Schreiber S. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis.[see comment]. *Nat Genet*. 2008;40:713-715.
18. Yu W, Lin Z, Kelly AA, Hegarty JP, Poritz LS, Wang Y, Li T, Schreiber S, Koltun WA. Association of Nkx2-3 polymorphism with Crohn's disease and expression of Nkx2-3 is up-regulated in B cell lines and intestinal tissues with Crohn's disease. *J Crohns Colitis*. 2009;3:189-195.
19. Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol*. 1997;29:1305-1312.
20. Lin Z, Cui X, Li H. Multiplex genotype determination at a large number of gene loci. *Proc Natl Acad Sci U S A*. 1996;93:2582-2587.
21. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*. 1992;89:1827-1831.
22. Boyd KE, Farnham PJ. Coexamination of site-specific transcription factor binding and promoter activity in living cells. *Mol Cell Biol*. 1999;19:8393-8399.
23. Russell RK, Nimmo ER, Satsangi J. Molecular genetics of Crohn's disease. *Curr Opin Genet Dev*. 2004;14:264-270.
24. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet*. 2005;6:376-388.
25. Elson CO, Cong Y, McCracken VJ, Dimmitt RA, Lorenz RG, Weaver CT. Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota. *Immunol Rev*. 2005;206:260-276.
26. Lees CW, Satsangi J. Genetics of inflammatory bowel disease: implications for disease pathogenesis and natural history. *Expert Rev Gastroenterol Hepatol*. 2009;3:513-534.
27. McDonald SA, Palmen MJ, Van Rees EP, MacDonald TT. Characterization of the mucosal cell-mediated immune response in IL-2 knockout mice before and after the onset of colitis. *Immunology*. 1997;91:73-80.
28. Briskin M, Winsor-Hines D, Shyjan A, Cochran N, Bloom S, Wilson J, McEvoy LM, Butcher EC, Kassam N, Mackay CR, Newman W, Ringler DJ. Human mucosal addressin cell adhesion molecule-1 is preferentially expressed in intestinal tract and associated lymphoid tissue. *Am J Pathol*. 1997;151:97-110.
29. Yu W, Lin Z, Hegarty JP, John G, Chen X, Faber PW, Kelly AA, Wang Y, Poritz LS, Schreiber S, Koltun WA. Genes regulated by Nkx2-3 in siRNA-mediated knockdown B cells: Implication of endothelin-1 in inflammatory bowel disease. *Mol Genet Metab*. 2010;100:88-95.
30. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33 Suppl:245-254.
31. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3:415-428.
32. Crabtree GR, Clipstone NA. Signal transmission between the plasma membrane and nucleus of T lymphocytes. *Annu Rev Biochem*. 1994;63:1045-1083.
33. Kulkarni RM, Greenberg JM, Akeson AL. NFATc1 regulates lymphatic endothelial development. *Mech Dev*. 2009;126:350-365.

34. Chen Y, Cao X. NFAT directly regulates Nkx2-5 transcription during cardiac cell differentiation. *Biol Cell*. 2009;101:335-349.
35. Zaichuk TA, Shroff EH, Emmanuel R, Filleur S, Nelius T, Volpert OV. Nuclear factor of activated T cells balances angiogenesis activation and inhibition. *J Exp Med*. 2004;199:1513-1522.
36. Horsley V, Pavlath GK. NFAT: ubiquitous regulator of cell differentiation and adaptation. *J Cell Biol*. 2002;156:771-774.
37. Armesilla AL, Lorenzo E, Gomez del Arco P, Martinez-Martinez S, Alfranca A, Redondo JM. Vascular endothelial growth factor activates nuclear factor of activated T cells in human endothelial cells: a role for tissue factor gene expression. *Mol Cell Biol*. 1999;19:2032-2043.
38. Rafiee P, Binion D, Li M, Olds C, Adams M, Johnson C. VEGF activates nuclear factor of activated T-cells (NFAT) but not NF-κB in human microvascular endothelial cells. *J Am Coll Surg*. 2000;191:S1-S2.
39. Hodge MR, Ranger AM, Charles de la Brousse F, Hoey T, Grusby MJ, Glimcher LH. Hyperproliferation and dysregulation of IL-4 expression in NF-ATp-deficient mice. *Immunity*. 1996;4:397-405.
40. Xanthoudakis S, Viola JP, Shaw KT, Luo C, Wallace JD, Bozza PT, Luk DC, Curran T, Rao A. An enhanced immune response in mice lacking the transcription factor NFAT1. *Science*. 1996;272:892-895.
41. Somoza M. (2006, October 8). *Peroxisome proliferator-activated receptor*. Retrieved November 12, 2011, from http://en.wikipedia.org/wiki/Peroxisome_proliferator-activated_receptor
42. Wiley JS, Sluyter R, Gu BJ, Stokes L, Fuller SJ. (2011). The human P2X7 receptor and its role in innate immunity. *Tissue Antigens*. 2011;78:321-332.

**18. Extent of Clinical Activities Initiated and Completed**. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__No

**If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed.** (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____Number of subjects originally targeted to be included in the study
_____Number of subjects enrolled in the study

**Note**: Studies that fall dramatically short on recruitment are encouraged to provide the details of their recruitment efforts in Item 17, Progress in Achieving Research Goals, Objectives and Aims. For example, the number of eligible subjects approached, the number that refused to participate and the reasons for refusal. Without this information it is difficult to discern whether eligibility criteria were too restrictive or the study simply did not appeal to subjects.

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics
_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White
_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

**19. Human Embryonic Stem Cell Research.** Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
___X_ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?
____ Yes
____ No

19(C) Please describe how this project involved human embryonic stem cells:

**20. Articles Submitted to Peer-Reviewed Publications**.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication or paper submitted for publication, listed in the table, in a PDF version 5.0.5 (or greater) format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, and an abbreviated title of the publication. For example, if you submit two publications for Smith (PI for Project 01), one publication for Zhang (PI for Project 03), and one publication for Bates (PI for Project 04), the filenames would be:

Project 01 – Smith – Three cases of isolated
Project 01 – Smith – Investigation of NEB1 deletions
Project 03 – Zhang – Molecular profiling of aromatase
Project 04 – Bates – Neonatal intensive care

If the publication is not available electronically, provide 5 paper copies of the publication.

**Note:** The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

| Title of Journal Article: | Authors: | Name of Peer-reviewed Publication: | Month and Year Submitted: | Publication Status (check appropriate box below): |
|---|---|---|---|---|
| 1. None | | | | ☐Submitted<br>☐Accepted<br>☐Published |

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes___X_____ No_____

If yes, please describe your plans:

The results will be organized and published as the following papers:
1) Identification of DNA methylation associated with Crohn's disease.
2) Identification of DNA methylation associated with ulcerative colitis.
3) DNA methylation affects gene expression of IBD susceptibility gene NKX2-3
4) Genetic variation and DNA methylation of genes PPAR-gamma and P2RX7 in inflammatory bowel disease

**21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project.** Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

**22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment**. Describe major discoveries, new drugs, and new approaches for prevention, diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

1) We identified IBD subtype-specific DNA methylation in intestinal tissues.
2) We identified disease-associated DNA methylation in non-flamed intestinal tissue. Whether these occurred at early-stage of disease needs to be further validated.
3) We observed differential methylation associated with tissue location in Crohn's disease
4) DNA methylation affects nuclear transcription factor binding to IBA susceptibility gene NKX2-3 gene

**23. Inventions, Patents and Commercial Development Opportunities.**

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes_____ No____X_____

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention:

b. Name of Inventor(s):

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____
If yes, indicate number of patent, title and date issued:
Patent number:
Title of patent:
Date issued:

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes_____ No____

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes_____ No____X_____

If yes, please describe your plans:

**24. Key Investigator Qualifications.** Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages.

## BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

| NAME | POSITION TITLE |
|---|---|
| Zhenwu Lin | Assistant Professor |
| eRA COMMONS USER NAME<br>zhenwulin | |

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

| INSTITUTION AND LOCATION | DEGREE *(if applicable)* | MM/YY | FIELD OF STUDY |
|---|---|---|---|
| Shandong University of Agriculture, China | B.S. | 07/74 | Genetic/Plant Breeding |
| Institute of Plant Physiology, Academy of Sciences of China, China | M.S. | 07/82 | Biochemistry |
| Institute of Plant Physiology, Academy of Sciences of China, China | Ph.D. | 07/86 | Biochemistry & Molecular Biology |

### A. Personal Statement

As a postdoctoral trained in plant molecular biology, the Principal Investigator was attracted to the progress of the human genome project and shifted his career from plant molecular biology to human genetics in 1993. In his efforts to establish himself in the basic genetic research, he became increasingly interested in the translational study of genetic discovery to clinical application and positioned himself to the research field close to clinical.

In 2006 he moved to the Department of Surgery and started his research in human inflammatory bowel disease with close collaboration with Dr. Koltun, Chief of Division of Colon & Rectal Surgery. To promote the research in IBD genetics, he went to Germany to work in Dr. Schreiber's laboratory, a world leading IBD geneticist, for 4 months in 2007 and learned all the IBD candidate genes identified by GWAS and advanced microarray technology. With his crystallized idea and experimental design, expertise, and excellent technical skill, he and his research team made considerable progress in the rapid expanded IBD research field of disease gene identification, genetic association, IBD-associated DNA methylation marker identification, and regulation of IBD-susceptibility genes. His team has studied inflammatory bowel disease and other digestive diseases from over 10 disease-associated genes, from genetic, epigenetic and gene function, from both nuclear genome and mitochondrial genome. Within 5 years, he has published over 20 peer-reviewed papers. He is interested in IBD susceptibility gene NKX2-3 since 2007 when he worked in Dr. Schreiber's laboratory. The NKX2-3 gene is a homeobox transcription factor, expressed in mucosal endothelial cells, and associated with both CD and UC. His team has studied genetic association of NKX2-3 with IBD, DNA methylation in the gene promoter region, function of the disease-risk variant, gene expression and allele specific expression in IBD patients, and regulation of global gene expression in human intestinal microvascular endothelial cells (HIMECs). To identify the genes regulated by NKX2-3, they have employed siRNA knock down technology and cDNA profiling. They identified multiple genes which function is related to lymphatic system. The recent finding and the progress of lympnatics research lead the P.I. to formalize the hypothesis and propose to generate transgenic mice to study the contribution of the IBD-risk NKX2-3 allele in IBD pathogenic lymphangiogenesis.

### B. Positions and Honors

**Postdoctoral Training**

1988-1990 Visiting Scientist, Botanical Institute, Kiel University, Kiel, Germany (Laboratory of Dr. K. Apel)
1990-1992 Postdoctoral Researcher, Ohio State Biotechnology Center, Ohio State University, Columbus, OH (Laboratory of Dr. D.P.S. Verma)

1992-1993 Postdoctoral Research Associate, Department of Biological Sciences, Purdue University, West Lafayette, IN (Laboratory of Dr. S. Gelvin)
1994-1995 Postdoctoral Research Associate, Coriell Institute for Medical Research (Laboratory of Dr. H. Li)
1996-1997 Postdoctoral Fellow, Department of Cellular and Molecular Physiology, Pennsylvania State University, College of Medicine, Hershey, PA (Laboratory of Dr. J. Floros)

**Positions Held**

1986-1987 Research Associate, Institute of Plant Physiology, Academy of Sciences of China, China
1996-1996 Research Scientist, Coriell Institute for Medical Research, NJ
1997-1999 Instructor, Department of Cellular and Molecular Physiology, Pennsylvania State University, College of Medicine, Hershey, PA
1999-2006 Assistant Professor, Department of Cellular and Molecular Physiology, Pennsylvania State University, College of Medicine, Hershey, PA
2006- Assistant Professor, Department of Surgery, Pennsylvania State University, College of Medicine, Hershey, PA

**Awards and Honors:**

1986 Top Award for Young Scientists, Shanghai Branch, The Academy of Sciences of china, China
1988-1989 Research Fellowship, Alexander von Humboldt Foundation, Germany
1990/1991 Recognized with appreciation, Ohio State University
1998-2000 Initial Investigator Award, American Heart Association
2000-2002 Research Grant Award, American Lung Association
2007-2008 Research Fellowship, Alexander von Humboldt Foundation, Germany

**C. Selected peer-reviewed publications (in chronological order).**

(Publications selected from 47 peer-reviewed publications)

1. **Lin Z**, Xiangfeng C, Li H. Multiplex genotype determination at large number of gene loci. *Proc Natl Acad Sci U S A*. 1996;93:2582-2587. PMCID: PMC39840
2. **Lin Z**, Wang G, deMello DE, Floros J. An alternative spliced surfactant protein B mRNA in normal human lung: disease implications. *Biochem J*. 1999;343:145-149. PMCID: PMC1220535
3. **Lin Z**, Floros J. Heterogeneous allele expression of pulmonary SP-D gene in rat large intestine and other tissues. *Physiol Genomics*. 2002;11:235-243.
4. Bibikova M, **Lin Z**, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, Goldmann T, Seifart C, Jiang W, Barker DL, Chee MS, Floros J, Fan JB. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*. 2006;16:383-393. PMCID: PMC1415217
5. **Lin Z,** Thomas NJ, Bibikova M, Seifart C, Wang Y, Guo X, Wang G, Vollmer E, Goldman T, Garcia EW, Zhou L, Fan B-B, Floros J. DNA methylation markers of surfactant proteins in lung cancer. *Int J Oncol.* 2007;31:181-191.
6. **Lin Z**, Poritz L, Franke A, Li T-Y, Rurther A, Byrnes KA, Wang Y, Gebhard AW, MacNeill C, Thomas NJ, Schreiber S, Koltun WA. Genetic association of nonsynonymous variations of the IL23R with familial and sporadic inflammatory bowel disease in women. *Dig Dis Sci*. 2009;55:739-746.
7. **Lin Z**, Nelson L, Franke A, Poritz L, Li T-Y, Wu R, Wang Y, MacNeill C, Thomas NJ, Schreiber S, Koltun WA. OCTN1 variant L503F is associated with familial and sporadic inflammatory bowel disease. *J Crohns Colitis*. Jun 2010;4(2):132-138.
8. Yu W, **Lin Z,** Pastor DM, Hegarty JP, Chen X, Kelly AA, Wang Y, Poritz LS, Koltun W. Genes regulated by Nkx2-3 in siRNA-mediated knockdown B cells: implication of endothelin-1 in inflammatory bowel disease. *Mol Genet Metab.* 2010;100(1):88-95.

9. Wang Z, Liu T, **Lin Z,** Hegarty J, Koltun WA, Wu R. A general model for multilocus epistatic interactions in case-control studies. *PLoS One*. Aug 18 2010;5(8):pii:e11389. PMCID: PMC2909900
10. **Lin Z,** Hegarty J, Cappel J, Yu W, Chen X, Faber P, Wang Y, Kelly AA, Poritz LS, Peterson BZ, Schreiber S, Fan JB, Koltun WA. Identification of disease-associated DNA methylation in intestinal tissues from inflammatory bowel disease. *Clin Genet.* 2011;80:59-67.
11. **Lin Z,** Hegarty J, Lin T, Ostrov B, Wang Y, Yu W, Kelly AA, Poritz LS, Koltun WA. Failure of anakinra treatment of pyoderma gangrenosum in an IBD patient and relevance to the PSTPIP1 gene. *Inflamm Bowel Dis. Jun* 2011;17(6):E41-2. doi: 10.1002/ibd.21684. Epub 2011 Mar 15.

12 **Lin Z**, John G, Berg A, Hegarty JP, Yu W, Wang Y, Kelly AA, Peterson B, Wu R, Floros J, Koltun WA. Genetic variants and monoallelic expression of surfactant protein D in inflammatory bowel disease. *Annual Hum Genet.* 2011 Sep;75(5):559-68. Epub 2011 Jul 25. PMCID: PMC3155611

13. Yu W, Hegarty JP, Berg A, Chen X, West G, Kelly AA, Wang Y, Poritz LS, Koltun WA, **Lin Z***. NKX2-3 transcriptional regulation of endothelin-1 and VEGF signaling in human intestinal microvascular endothelial cells. PloS One. 2011;6:e20454. Epub 2011 May 26. PMCID: PMC3102722
14. John G, Hegarty JP, Yu W, Berg A, Pastor DM, Kelly AA, Wang Y, Poritz LS, Schreiber S, Koltun WA, **Lin Z***. NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol Genet Metab*. 201;104:174-179. Epub 2011 Jul 6.
15. **Lin Z,** Hegarty J, Wei Yu, Cappel J et al, Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Dig Dis Sci*. 2012 Jul 21. [Epub ahead of print]

**D. Research Support**

**Ongoing Research Support**

None

**Completed Research Support**

No number Lin (PI) 01/012010 - 12/31/2012
Commonwealth of Pennsylvania, Tobacco Settlement Fund
Title: Identification and Function Study of Inflammatory Bowel Disease-associated DNA Methylation.
Project goal: In this grant I proposed to study DNA methylation in naive B cell and mucosal tissues with Illumina HumanMethylation 27 Beadchip and the contribution of DNA methylation to gene expression and thus to IBD pathogenesis in 4 genes that are identified as IBD-associated methylation from our Illumina Goldengate BeadArray data.
Role: PI

No number Lin (PI) 07/01/2007 – 06/30/2010
Surgery Feasibility Grant
Title: Epigenetic Alteration and Inflammatory Bowel Disease (IBD): Identification of DNA Methylation Markers for IBD
Project goal: This study is to identify IBD associated DNA methylation markers and the role of DNA methylation in gene expression.
Role: PI

No number Koltun (PI) 06/01/2007 - 06/30/2012
Philadelphia Health Care Trust Grant
Title: The Genetic and Metabolic Characterization of Inflammatory Bowel Disease.
Project goal: This study includes three goals: to identify and quantify the genetic Variations associated with IBD, to investigate the role of epigenetic alterations in the expression of genes in IBD, and to investigate

other metabolic signature of IBD patients and identify serum markers of disease status that correlates with genetic mutations associated with IBD.
Role: Co-PI

No number Lin (PI) 07/01/2011 - 06/30/2013
Children's Miracle Network
Title: Genetic Association and Function of the IL10 and IL10 Signaling Pathway in Pediatric Bowel Disease
Project goal: to determine the genetic association of genes in the IL 10 signaling pathway with pediatric IBD, to ascertain their genetic contribution to disease risk by epistatic interaction analysis, and to understand the anti-inflammatory function of the IL 10 pathway in human B cells in the context of pediatric IBD.
Role: PI