Supplementary Material

Table S1. Targets on gene expression array

Gene	Protein Function	Assay ID
IL4	Type 2 cytokine	Hs00174122_m1
IL5	Type 2 cytokine	Hs00174200_m1
IL13	Type 2 cytokine	Hs00174379_m1
IL13RA2	Type 2 cytokine receptor/protein induced by IL-13	Hs00152924_m1
IL33	Cytokine that induces type 2 immune responses	Hs01125943_m1
IL1RL1(s)	IL-33 Receptor (soluble) that regulates type 2 immune responses	Hs01073297_m1
IL1RL1(m)	IL-33 Receptor (membrane) that induces type 2 immune responses	Hs01073295_m1
CCL11	Chemokine associated with eosinophil recruitment and type 2 immune responses	Hs00237013_m1
CHI3L1	Protein induced by type 2 immune responses	Hs00609691_m1
ICOS	T cell co-stimulatory molecular that augments type 2 immune responses	Hs00359999_m1
GATA3	Type 2 T cell transcription factor	Hs00231122_m1
RORA	Group 2 ILC transcription factor	Hs00536545_m1
CLDN2	Tight junction protein induced by IL-13	Hs01549234_m1
IFNG	Type 2 cytokine	Hs00174143_m1
TBX21	Type 1 transcription factor	Hs00203436_m1
IL17A	Type 17 cytokine	Hs00174383_m1
IL23A	Cytokine that augments type 17 immune responses	Hs00372324_m1
RORC	Type 17 transcription factor	Hs01076122_m1
IL22	Type 17 cytokine	Hs01574154_m1
AHR	Group 3 ILC transcription factor	Hs00169233_m1
IL10	T regulatory cell cytokine	Hs00174086_m1
TGFB1	Cytokine that induces T regulatory cell and type 17 T cell differentiation	Hs00171257_m1
S100A8	Component of calprotectin/marker of inflammation	Hs00374264_g1
GAPDH	Reference gene	Hs99999905_m1

Table S2. Comparison between the studied RISK cohort subset and overall RISK cohort

		Non-IBD		CDic		CDc		UC	
	Entire RISK Cohort (n = 1812)	Studied Subset (n = 49)	RISK Cohort (n = 408)	Studied Subset (n = 46)	RISK Cohort (n = 606)	Studied Subset (n = 36)	RISK Cohort (n = 210)	Studied Subset (n = 56)	RISK Cohort (n = 200)
Age, y	12.7 (9.9,14.9)	12.8 (10.8,15.0)	12.7 (9.5,15.0)	12.4 (10.9,13.6)	12.4 (9.9,14.7)	12.7 (10.8,14.5)	12.5 (9.4,15.0)	13.5 (10.8,15.5)	13.1 (10.2,15.0)
A1a: 0 – <10 y	460 (25.4)	9 (18.4)	116 (28.7)	11 (23.9)	152 (25.1)	6 (17.1)	57 (27.1)	13 (23.2)	49 (24.5)
A1b: 10 – <17 y	1348 (74.6)	39 (79.6)	288 (71.3)	35 (76.1)	454 (74.9)	30 (85.7)	153 (72.9)	43 (76.8)	151 (75.5)
Male sex	1055 (58.2)	23 (46.9)	226 (55.4)	25 (54.3)	377 (62.2)	17 (47.2)	120 (57.1)	30 (53.6)	106 (53.0)
Diagnosis									
Non-IBD	408 (22.5)	49 (100)	408 (100)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
CD	1118 (61.7)	0 (0)	0 (0)	46 (100)	606 (100)	36 (100)	210 (100)	0 (0)	0 (0)
UC	200 (11.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	56 (100)	200 (100)
IBD-U	86 (4.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
CD location*									
L1	181 (16.2)			0 (0)	0 (0)	0 (0)	0 (0)		
L2:	210 (18.8)			0 (0)	0 (0)	36 (100)	210 (100)		
L3	606 (54.2)			46 (100)	606 (100)	0 (0)	0 (0)		
L4 only	18 (1.6)			0 (0)	0 (0)	0 (0)	0 (0)		
Insufficient Data	103 (9.2)			0 (0)	0 (0)	0 (0)	0 (0)		
Upper GI									
L4a	531 (47.5)			32 (69.6)	322 (53.1)	17 (47.2)	82 (39.0)		
L4b	147 (13.1)			12 (26.1)	97 (16.0)	0 (0)	20 (9.5)		
UC extent†									
E1	5 (2.5)							2 (3.6)	5 (2.5)
E2	26 (13.0)							7 (12.5)	26 (13.0)
E3	22 (11.0)							7 (12.5)	22 (11.0)
E4	112 (56.0)							36 (64.3)	112 (56.0)
Insufficient data	35 (17.5)							4 (7.1)	35 (17.5)
Macroscopic rectal involvement	846 (60.3)			46 (100)	409 (67.5)	36 (100)	157 (74.8)	56 (100)	177 (88.5)
PGA									

Quiescent	50 (3.6)		0 (0)	13 (2.1)	1 (2.8)	11 (5.2)	1 (1.8)	7 (3.5)
Mild	458 (32.6)		11 (23.9)	168 (27.7)	7 (19.4)	62 (29.5)	20 (35.7)	63 (31.5)
Moderate	684 (48.7)		22 (47.8)	324 (53.5)	19 (52.8)	106 (50.5)	23 (41.1)	101 (50.5)
Severe	209 (14.9)		13 (28.3)	98 (16.2)	9 (25.0)	31 (14.8)	12 (21.4)	29 (14.5)
Insufficient data	3 (0.2)		0 (0)	3 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)
PUCAI	45 (30,60)						45 (35,60)	45 (30,60)
Rectal deep ulcers	158 (11.3)		5 (10.9)	75 (12.4)	12 (33.3)	36 (17.1)	10 (17.9)	28 (14.0)

Quantitative variables expressed as median (quartile 1, quartile 3) and dichotomous variables as n (%).
*L1, terminal ileal ± limited cecal disease; L2, colonic; L3, ileocolonic; L4, upper GI disease; L4a, upper GI disease proximal to ligament of Treitz, L4b: upper GI disease distal to ligament of Treitz
†E1, ulcerative proctitis; E2, left-sided colitis; E3, extensive colitis; E4, pancolitis
CD, Crohn's disease; CDc, colon-only Crohn's disease; CDic, ileocolonic Crohn's disease; IBD, inflammatory bowel disease; IBD-U, inflammatory bowel disease unclassified; PGA, physicians global assessment; PUCAI, Pediatric Ulcerative Colitis Activity Index; UC, ulcerative colitis

Table S3. Characteristics of the Cincinnati Cohort patients

	Non-IBD (n = 17)	CD (n = 20)	UC (n = 14)
Age at biopsy collection, y	15.8 (10.5,16.6)	15.0 (11.2,6.7)	18.0 (14.5,19.0)
Male sex	7 (41.2)	15 (75.0)	25 (54.3)
Age at diagnosis, y		10.6 (9.2,10.6)	14.4 (11.6,16.5)
A1a: 0–<10 y		8 (40.0)	3 (21.4)
A1b: 10–<17 y		11 (55.0)	10 (71.4)
A2: 17–40 y		1 (5.0)	1 (7.1)
Time since diagnosis, y		2.2 (0,4.7)	2.8 (1.0,6.2)
Biopsy collected at diagnostic endoscopy		9 (45.0)	0 (0.0)
CD location			
L1: terminal ileal ± limited cecal disease		0 (0)	0 (0)
L2: colonic		8 (40.0)	0 (0)
L3: ileocolonic		12 (60.0)	46 (100)
L4a: upper disease proximal to ligament of Treitz		11 (55.0)	32 (69.6)
L4b: upper disease distal to ligament of Treitz		2 (10.0)	12 (26.1)
UC extent			
E1: ulcerative proctitis			1 (7.1)
E2: left-sided colitis			3 (21.4)
E3: extensive colitis			3 (21.4)
E4: pancolitis			7 (50.0)
Macroscopic rectal involvement		20 (100)	14 (100)
Clinical Disease Activity			
shPCDAI		25 (12.5,35)	
PUCAI			30 (15,60)
Endoscopic Disease Activity			22 (47.8)

SES-CD	12 (8.5,17)	
Mayo Endoscopic Score		
Mild		4 (28.6)
Moderate		3 (21.4)
Severe		7 (50.0)
Medications		
None	5 (25.0)	0 (0.0)
Oral corticosteroids	2 (10.0)	7 (50.0)
Rectal corticosteroids	1 (5.0)	2 (14.3)
Oral 5-ASA	4 (20.0)	10 (71.4)
Antibiotic	3 (15.0)	0 (0.0)
6-mercaptopurine or azathioprine	5 (25.0)	2 (14.3)
Methotrexate	0 (0.0)	1 (7.1)
Anti-TNF biologic	2 (10.0)	3 (21.4)

Quantitative variables expressed as median (quartile 1, quartile 3) and dichotomous variables as n (%).
5-ASA, 5-aminosalicylic acid; CD, Crohn's disease; IBD, inflammatory bowel disease; PUCAI, Pediatric Ulcerative Colitis Activity Index; SES-CD, Simple Endoscopic Score for Crohn's Disease; shPCDAI, Short Pediatric Crohn's Disease Activity Index; TNF, tumor necrosis factor; UC, ulcerative colitis

Table S4. Univariate logistic regression for discriminating UC from CDc

Gene	OR*	(95% CI)	*P* value
IL5	1.147	(1.050–1.644)	.003
IL13	1.100	(.998–1.212)	.056
IL13RA2	1.311	(1.045–1.644)	.020
IL1RL1(m)	1.798	(1.129–2.865)	.014
ICOS	1.435	(.923–2.229)	.109
IL17A	1.265	(1.032–1.551)	.024
IL23A	1.391	(1.063–1.821)	.016

*Odds of a diagnosis of UC over CDc per unit increase in Cq value for the listed gene.
CI, confidence interval; Cq, quantification cycle; OR, odds ratio

Table S5. Change in effect estimate after bivariate analyses with *S100A8*

	Effect estimate		
Gene	Univariate	Bivariate	Change (%)
IL5	.138	.126	– 8.6
IL13	.095	.078	– 18.1
IL13RA2	.270	.332	+ 22.8
IL1RL1(m)	.587	.561	– 4.4
ICOS	.361	.277	– 23.3
IL17A	.235	.232	– 1.7
IL23A	.330	.328	– .7

Table S6. Inclusion of *S100A8* in multivariate logistic regression model for discriminating UC from CDc

Gene	OR*	95% CI	*P*-value
IL5	1.133	1.032–1.238	.009
IL17A	1.232	.976–1.467	.132
S100A8	.959	.747–1.233	.747

*Odds of a diagnosis of UC over CDc per unit increase in Cq value for the listed gene.
CI, confidence interval; Cq, quantification cycle; OR, odds ratio

Table S7. Univariate logistic regression of gene expression for predicting UC clinical outcomes

	Remission				Response			
	6 month		12 month		6 month		12 month	
Gene	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
IL5	1.119 (.978–1.279)	.104	1.050 (.921–1.198)	.466	1.116 (.981–1.271)	.095	1.126 (.979–1.294)	.094
IL13	1.115 (.973–1.279)	.118	1.126 (.978–1.297)	.099	**1.182 (1.028–1.359)**	**.019**	**1.172 (1.012–1.359)**	**.034**
IL13RA2	1.350 (.939–1.942)	.105	.847 (.595–1.208)	.359	1.206 (.855–1.701)	.286	1.046 (.733–1.493)	.804
IL1RL1(m)	1.241 (.618–2.494)	.543	.716 (.331–1.550)	.397	1.751 (.786–3.891)	.144	.690 (.314–1.517)	.356
ICOS	1.151 (.581–2.283)	.686	1.198 (.631–2.273)	.582	1.032 (.515–2.066)	.930	1.103 (.570–2.137)	.772
IL17A	1.253 (.919–1.706)	.153	1.111 (.854–1.445)	.433	1.239 (.929–1.650)	.170	1.117 (.861–1.449)	.407
IL23A	1.311 (.887–1.934)	.175	1.071 (.746–1.538)	.711	1.199 (.813–1.767)	.360	1.047 (.718–1.527)	.810

*Odds of outcome per unit increase in Cq value for the listed gene.
CI, confidence interval; Cq, quantification cycle; OR, odds ratio

Table S8. Comparison of baseline characteristics and medication exposures between UC patient gene expression clusters

	Clusters 1-3 (n = 33)	Clusters 4-5 (n = 13)	*P* value
Baseline characteristics			
Age, y	13.6 (10.7-15.1)	12.2 (11.5-15.8)	.813
Male sex	19 (57.6)	5 (38.5)	.330
UC extent			
E1: ulcerative proctitis	1 (3.1)	0 (0)	.246*
E2: left-sided colitis	2 (6.3)	2 (15.4)	
E3: extensive colitis	4 (12.5)	4 (30.8)	
E4: pancolitis	25 (78.1)	7 (53.8)	
Data not available	1 (3.1)	0 (0)	
PGA			
Quiescent	1 (3.0)	0 (0)	.546*
Mild	15 (45.5)	4 (30.8)	
Moderate	9 (27.3)	5 (38.5)	
Severe	8 (24.2)	4 (30.8)	
Rectal deep ulcers	5 (15.6)	2 (15.4)	1.00
Medication exposures			
6-month	(n = 31)	(n = 13)	
Corticosteroids	24 (77.4)	12 (92.3)	.401
5-ASA	24 (77.4)	11 (84.6)	.703
Thiopurines	10 (32.3)	6 (46.2)	.496
Methotrexate	0 (0)	0 (0)	–
Anti-TNF biologic	5 (16.1)	4 (30.8)	.414
12-month	(n = 28)	(n = 11)	
Corticosteroids	23 (82.1)	10 (90.9)	.655
5-ASA	22 (78.6)	9 (81.8)	1.00
Thiopurines	12 (46.4)	4 (36.4)	.725
Methotrexate	1 (3.6)	1 (9.1)	.489
Anti-TNF biologic	7 (25.0)	6 (54.5)	.131

Quantitative variables expressed as median (quartile 1, quartile 3) and dichotomous variables as n (%).
5-ASA, 5-aminosalicylic acid; TNF, tumor necrosis factor; UC, ulcerative colitis
*First two groups combined for chi-square test

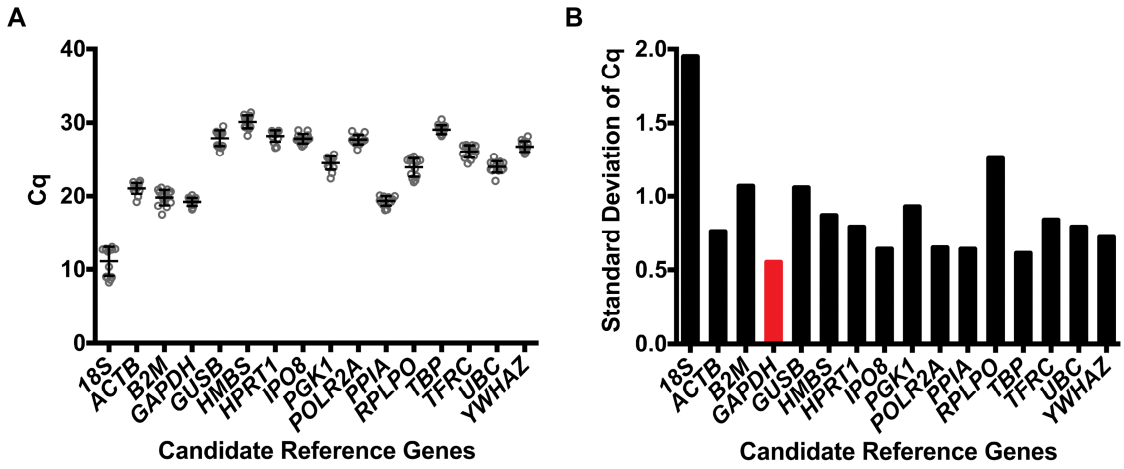

Figure S1. Identification of GAPDH as reference gene with least expression variation across diagnosis groups. (*A*) Dot plot depicting mean Cq and standard errors of 16 candidate reference genes from an endogenous control real-time RT-qPCR microfluidic array (n = 4 each of non-IBD, CDic, CDc, UC). (B) Bar chart depicting standard deviations of the Cq for each candidate reference gene. *GAPDH* exhibited the lowest variability in expression across samples.

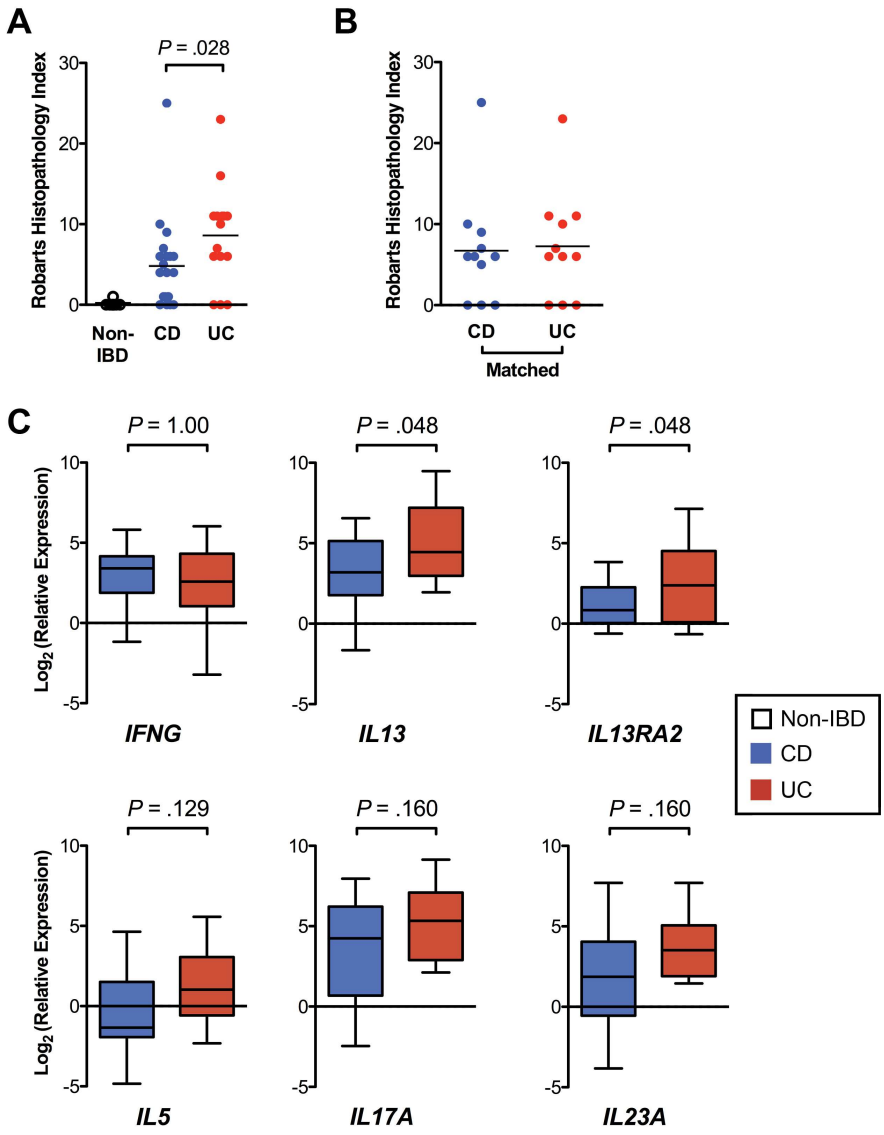

Figure S2. Analysis of Cincinnati cohort rectal mucosa real-time RT-qPCR controlling for histologic disease activity. (A) Dot plot of Robarts Histologic Index scores for CD and UC patients in the Cincinnati cohort. (B) Dot plot of a subset of CD and UC patients in the Cincinnati cohort matched on Robarts Histologic Index scores. (C) Box and whisker chart depicting gene expression normalized to median expression of the non-IBD patient group in CD and UC patients matched on Robarts Histologic Index scores (boxes represent median and interquartile range, whiskers represent the 95% confidence interval).