

Supplementary Materials

Large-scale transcriptomic meta-analysis identifies
constitutive and inflammation-dependent layers in ulcerative colitis

Piernik et al.

Contents

1	Supplementary Figures	2
2	Supplementary Tables	5

1 Supplementary Figures

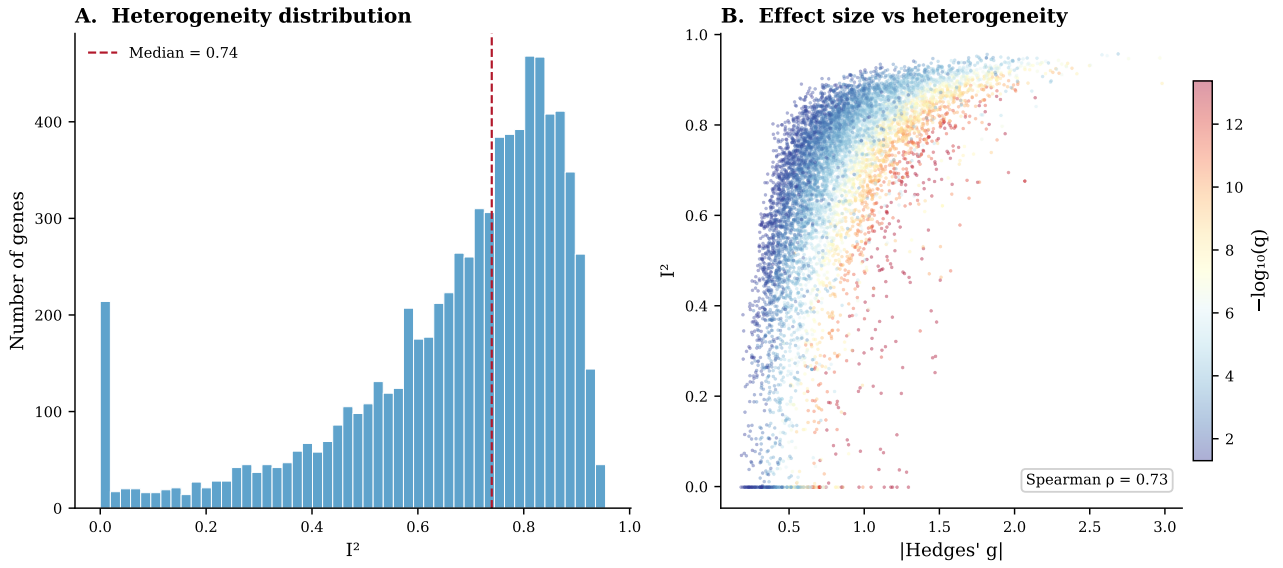


Figure S1: **Heterogeneity diagnostics for the inflamed UC vs control gene-level meta-analysis.** (A) Distribution of I^2 across 7,594 significantly dysregulated genes. The median I^2 of 0.74 reflects variation in effect-size magnitude across the 15 datasets and 9 microarray platforms, rather than inconsistency in direction (controlled by the $\geq 80\%$ direction consistency filter). (B) Relationship between absolute effect size ($|g|$) and heterogeneity (I^2). Colour encodes statistical significance ($-\log_{10} q$). Genes with large effects tend to have high I^2 (Spearman $\rho = 0.73$), while genes ranking highest by q -value have low I^2 , reflecting complementary properties of the two ranking criteria.

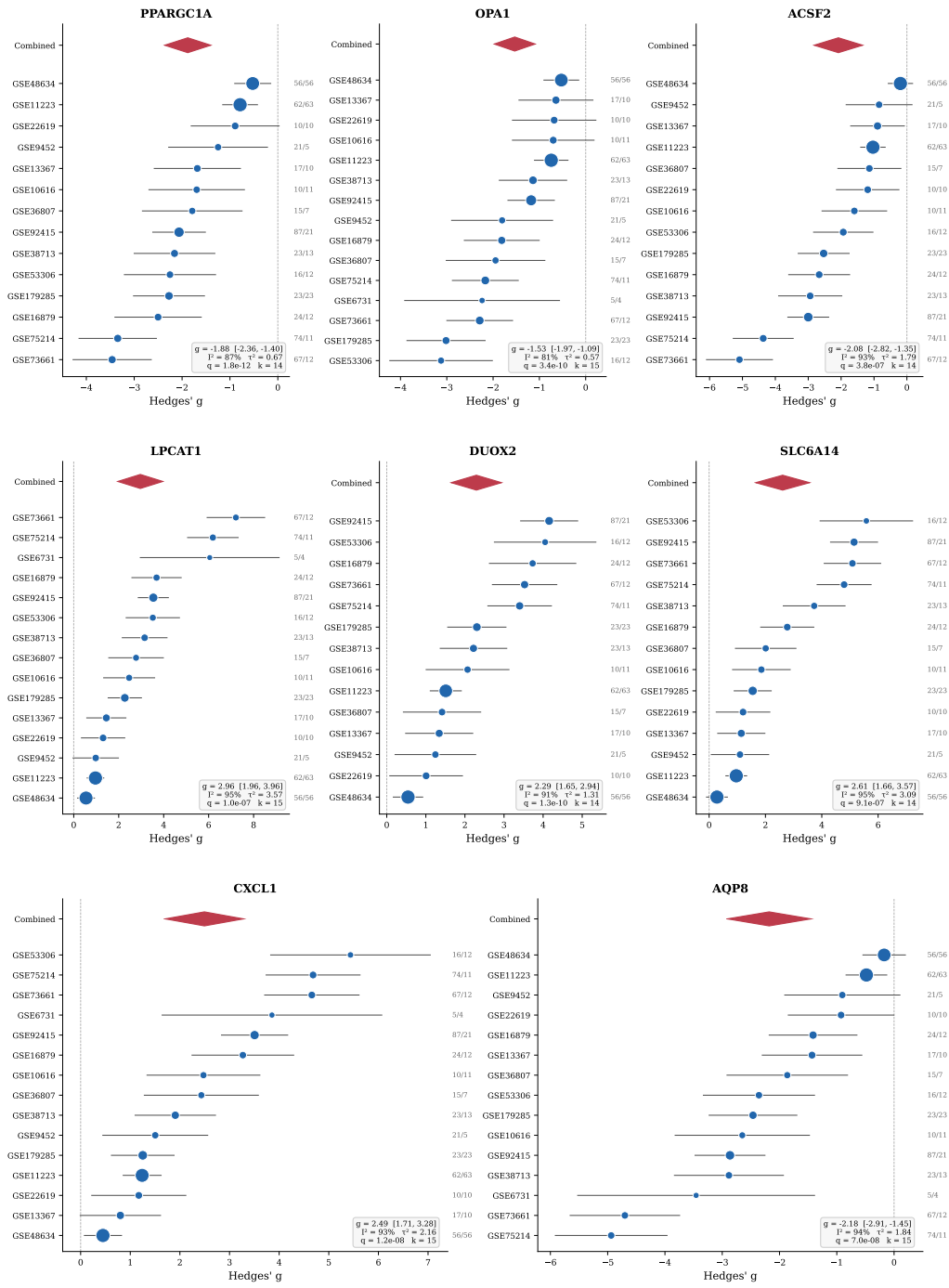


Figure S2: Forest plots for representative genes from the inflamed UC vs control meta-analysis. Each panel shows per-dataset effect sizes (Hedges' g , blue circles scaled by inverse variance) with 95% confidence intervals, and the random-effects combined estimate (red diamond). Numbers at right indicate sample sizes (n_{UC}/n_{Ctrl}). **Top row:** PPARGC1A, OPA1, and ACSF2 (downregulated metabolic genes). **Middle row:** LPCAT1, DUOX2, and SLC6A14 (top upregulated genes). **Bottom row:** CXCL1 (top upregulated immune) and AQP8 (downregulated barrier gene).

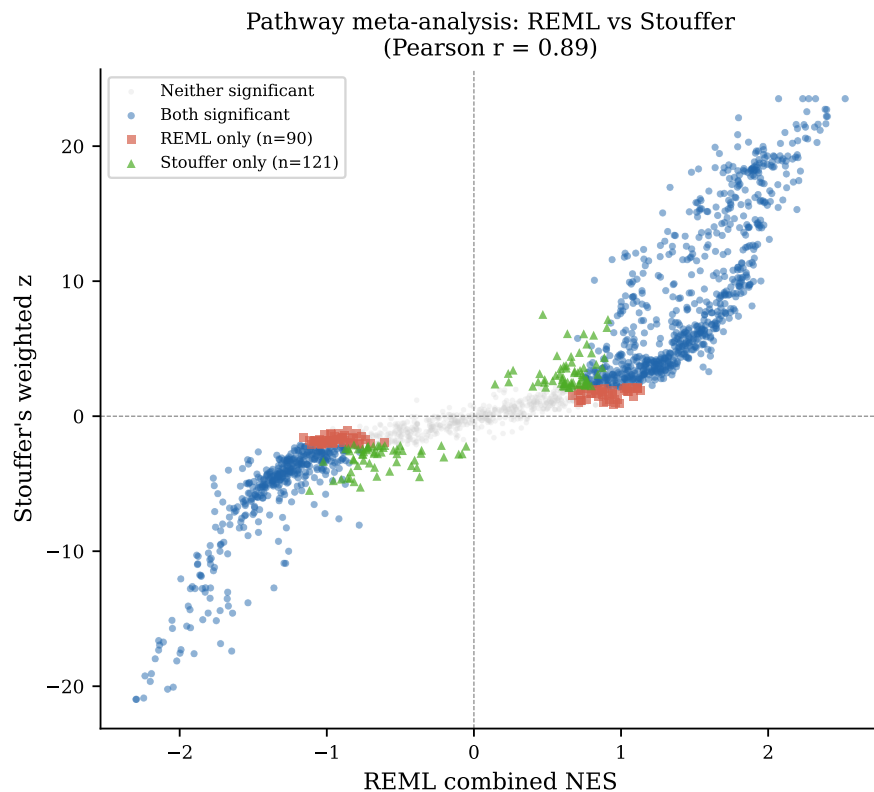


Figure S3: **Sensitivity analysis: REML random-effects vs Stouffer's weighted z -method for pathway-level meta-analysis.** Each point represents one Reactome pathway. The x -axis shows the REML combined normalized enrichment score (NES); the y -axis shows Stouffer's weighted z -statistic (weights = $\sqrt{n}$). Blue points are significant by both methods ($q < 0.05$); red squares are REML-only; green triangles are Stouffer-only; grey points are non-significant by both. 92.5% of REML-significant pathways were confirmed by the Stouffer method, with 100% direction concordance and Pearson $r = 0.89$, supporting the validity of the NES variance approximation used in the REML framework.

2 Supplementary Tables

Tables S1–S5: Gene-level meta-analysis results

Full gene-level meta-analysis results for all five comparisons are provided as separate CSV files:

- **Table S1:** UC inflamed vs healthy controls (7,594 significant genes)
- **Table S2:** UC uninflamed vs healthy controls (93 significant genes)
- **Table S3:** UC inflamed vs UC uninflamed (5,455 significant genes)
- **Table S4:** CD inflamed vs healthy controls (2,861 significant genes)
- **Table S5:** UC vs CD direct comparison (2,882 significant genes)

Each file contains: gene symbol, number of datasets, Hedges' g , standard error, 95% confidence interval, p -value, q -value (Benjamini–Hochberg), τ^2 , I^2 , direction of effect, and direction consistency ratio.

Tables S6–S10: Pathway-level meta-analysis results

Significant pathway-level meta-analysis results ($q < 0.05$) for all five comparisons are provided as separate CSV files:

- **Table S6:** UC inflamed vs healthy controls (978 significant pathways)
- **Table S7:** UC uninflamed vs healthy controls (115 significant pathways)
- **Table S8:** UC inflamed vs UC uninflamed (808 significant pathways)
- **Table S9:** CD inflamed vs healthy controls (382 significant pathways)
- **Table S10:** UC vs CD direct comparison (529 significant pathways)

Each file contains: Reactome pathway term, number of datasets, combined NES, standard error, 95% CI, p -value, q -value, τ^2 , I^2 , direction, direction ratio, and core leading-edge genes.

Pathway clusters — UC inflamed vs healthy controls

Cluster representative	Pathways	Mean NES
Complement Cascade	3	+2.21
IRE1alpha Activates Chaperones	3	+2.17
Assembly Of Collagen Fibrils And Other Multimeric Structures	14	+2.15
Interleukin-10 Signaling	29	+2.14
Antigen processing-Cross Presentation	23	+2.09
Chondroitin Sulfate Biosynthesis	1	+2.07
Regulation Of IGF Transport And Uptake By IGFBPs	1	+1.97
Peroxisomal Protein Import	3	-2.14
Mitochondrial Fatty Acid Beta-Oxidation	1	-2.14
Respiratory Electron Transport, ATP Synthesis By Chemiosmotic Coupling, Heat Production By Uncoupling Proteins	11	-2.12
Glucuronidation	3	-2.11
Response To Metal Ions	2	-2.05

Table S11: Pathway clusters from the inflamed UC vs healthy controls analysis. Each row is a community of co-enriched Reactome pathways identified by Louvain clustering on shared leading-edge genes. Positive NES = upregulated in inflamed UC relative to controls; negative NES = downregulated.

Pathway clusters — uninflamed UC vs healthy controls

Cluster representative	Pathways	Mean NES
SRP-dependent Cotranslational Protein Targeting To Membrane	15	+2.25
Formation Of A Pool Of Free 40S Subunits	7	+2.22
Interferon Gamma Signaling	3	+1.97
Dectin-2 Family	1	+1.86
Platelet Degranulation	2	+1.85
Glucuronidation	1	−2.30
Unblocking Of NMDA Receptors, Glutamate Binding And Activation	1	−1.88
Triglyceride Catabolism	1	−1.82
Synthesis Of PIPs At Late Endosome Membrane	1	−1.72

Table S12: Pathway clusters from the uninflamed UC vs healthy controls analysis. Each row is a community of co-enriched Reactome pathways identified by Louvain clustering on shared leading-edge genes. Positive NES = upregulated in uninflamed UC relative to controls; negative NES = downregulated.

Pathway clusters — UC inflamed vs UC uninfamed

Cluster representative	Pathways	Mean NES
Interleukin-10 Signaling	12	+2.21
Assembly Of Collagen Fibrils And Other Multimeric Structures	16	+2.21
Complement Cascade	3	+2.17
IRE1alpha Activates Chaperones	3	+2.13
Cross-presentation Of Soluble Exogenous Antigens (Endosomes)	26	+2.07
Post-translational Protein Phosphorylation	1	+2.06
Respiratory Electron Transport	12	−2.34
Protein Localization	3	−2.31
Aspirin ADME	2	−2.17
Mitochondrial Fatty Acid Beta-Oxidation	2	−2.17
Response To Metal Ions	1	−2.13

Table S13: Pathway clusters from the direct inflamed versus uninfamed UC analysis. Each row is a community of co-enriched Reactome pathways identified by Louvain clustering on shared leading-edge genes. Positive NES = higher in inflamed than uninfamed UC; negative NES = lower.

Pathway clusters — CD inflamed vs healthy controls

Cluster representative	Pathways	Mean NES
Interferon Alpha/Beta Signaling	8	+2.21
Interleukin-10 Signaling	10	+2.16
Integrin Cell Surface Interactions	9	+2.11
Regulation Of Complement Cascade	6	+2.08
Chondroitin Sulfate Biosynthesis	1	+2.02
Respiratory Electron Transport, ATP Synthesis By Chemiosmotic Coupling, Heat Production By Uncoupling Proteins	6	-2.43
Selenoamino Acid Metabolism	2	-2.37
Protein Localization	2	-2.05
Response To Metal Ions	1	-2.04

Table S14: Pathway clusters from the CD inflamed vs healthy controls analysis. Each row is a community of co-enriched Reactome pathways identified by Louvain clustering on shared leading-edge genes. Positive NES = upregulated in CD relative to controls; negative NES = downregulated.

Pathway clusters — UC vs CD direct comparison

Cluster representative	Pathways	Mean NES
SRP-dependent Cotranslational Protein Targeting To Membrane	1	+2.22
Prefoldin Mediated Transfer Of Substrate To CCT/TriC	5	+2.11
Unfolded Protein Response (UPR)	6	+2.02
Regulation Of Complement Cascade	2	+2.02
Assembly Of Collagen Fibrils And Other Multimeric Structures	10	+1.99
Activation Of ATR In Response To Replication Stress	3	+1.98
POLB-Dependent Long Patch Base Excision Repair	1	+1.91
Tryptophan Catabolism	1	+1.89
Eukaryotic Translation Elongation	1	+1.89
RAF-independent MAPK1/3 Activation	1	+1.88
Respiratory Electron Transport	17	-2.14
Glucuronidation	2	-2.05
Peroxisomal Lipid Metabolism	4	-2.01
Mitochondrial Fatty Acid Beta-Oxidation	8	-1.99
Budding And Maturation Of HIV Virion	1	-1.96
Digestion	1	-1.87
Abacavir ADME	1	-1.87
Degradation Of Cysteine And Homocysteine	1	-1.87

Table S15: Pathway clusters from the UC vs CD direct comparison analysis. Each row is a community of co-enriched Reactome pathways identified by Louvain clustering on shared leading-edge genes. Positive NES = upregulated in UC relative to CD; negative NES = downregulated in UC relative to CD.

Constitutive genes

Genes altered against controls both when inflamed and when uninflamed, in the same direction: the change is already present before inflammation is. g_{Δ} is the effect size of the direct inflamed versus uninflamed contrast, shown for reference. The RNA-seq column grades how well the independent cohorts support the uninflamed effect, the claim on which this layer rests, across both contrasts that define the layer: ●● both replicated, ● one replicated, ○ same direction but neither significant, × not supported, — not tested. The final column marks the genes that are additionally inflammation-dependent; being altered before inflammation and being changed further by it are independent properties, so a gene may carry both.

Table S16: Constitutive genes (56 genes: 40 upregulated, 16 downregulated; 32 are also inflammation-dependent).

Gene	g_{infl}	g_{uninfl}	g_{Δ}	RNA-seq	Infl-dep
SERPINA3	+1.40	+0.92	+0.96	•	•
SHANK3	+1.01	+0.91	+0.69	•	
CYP4X1	+1.41	+0.89	+0.87	••	•
SHC2	+0.99	+0.88	+0.73	•	
TTC9	+1.68	+0.75	+1.10	•	•
CCNO	+1.12	+0.71	+0.60	•	•
AMIGO3	+1.21	+0.69	+0.66	◦	
AMPD3	+1.01	+0.69	+0.70	•	
GPX8	+1.88	+0.68	+1.38	••	•
CADPS	+0.60	+0.65	+0.07	◦	
SPARCL1	+1.07	+0.63	+0.68	×	
ATP6V1F	+1.44	+0.62	+0.96	•	•
GPX7	+1.27	+0.60	+1.05	•	•
WDR54	+1.59	+0.58	+1.37	×	•
ICAM2	+1.24	+0.56	+0.88	•	•
HBA2	+0.47	+0.54	+0.19	◦	
GIMAP2	+1.05	+0.52	+0.60	•	
FGD5	+1.04	+0.52	+0.64	•	
PTAFR	+1.47	+0.52	+0.98	•	•
B3GNT9	+0.92	+0.51	+0.65	••	
FBLN5	+0.97	+0.50	+0.51	•	•
CD93	+1.26	+0.49	+0.86	•	•
HBB	+0.43	+0.49	−0.00	×	
PSAT1	+1.39	+0.48	+1.16	•	•
CHRD2	+1.19	+0.47	+1.05	•	•
TMEM64	+0.75	+0.47	+0.46	•	
GIMAP4	+1.14	+0.47	+0.75	•	•
HPSE	+0.92	+0.47	+0.49	•	
EBF3	+0.78	+0.46	+0.70	•	•
PTGDS	+1.20	+0.46	+0.99	•	•
RNF183	+1.15	+0.46	+0.76	•	•
PDPN	+1.53	+0.46	+1.33	•	•
KYNU	+2.09	+0.46	+1.69	•	•
RGCC	+1.38	+0.46	+1.19	•	
GIMAP7	+1.11	+0.46	+0.79	•	•
HELZ2	+0.92	+0.45	+0.70	•	•
CTSK	+1.43	+0.45	+1.13	•	•
PRUNE2	+0.70	+0.45	+0.26	•	
TGFBI	+1.76	+0.45	+1.50	•	•
SMO	+0.65	+0.45	+0.54	•	•

Gene	g_{infl}	g_{uninfl}	g_{Δ}	RNA-seq	Infl-dep
C1QTNF3-AMACR	−2.24	−1.36	−1.98	◦	•
FTH1P5	−1.21	−1.06	−0.98	—	•
NMRAL2P	−0.86	−1.02	−0.44	•	
TMEM63C	−2.18	−1.01	−1.35	•	
MIR101-1	−0.59	−0.94	−0.05	—	
MAP3K13	−1.00	−0.70	−0.72	•	•
ZNF823	−1.23	−0.68	−1.20	•	•
TRDMT1	−0.92	−0.64	−0.41	•	
OPN4	−0.45	−0.61	+0.02	—	
TMEM65	−1.30	−0.52	−0.87	•	•
ZNF334	−0.62	−0.51	−0.18	•	
PDP2	−0.82	−0.50	−0.75	•	•
ERBB4	−0.55	−0.49	+0.03	—	
PEX1	−0.92	−0.45	−0.72	•	•
TMTC4	−1.09	−0.45	−0.81	•	
CEP152	−0.41	−0.45	+0.12	◦	

Inflammation-dependent genes

Genes altered when inflamed and differing between inflamed and uninflamed mucosa, in the same direction: inflammation measurably changes them. There are 4,192 such genes, too many to print, so they are provided as **Table S17** together with their RNA-seq support.

Leave-one-dataset-out sensitivity

Each meta-analysis was repeated with every contributing cohort omitted in turn. Only the pooling step is repeated; per-dataset statistics and enrichment scores are unchanged. Significance criteria are reapplied within each fold, with the presence threshold set to 50% of the remaining cohorts.

Retained is the percentage of the genes significant in the full analysis that remain significant when a cohort is omitted, taken across all folds. r is the smallest correlation between fold and full effect sizes, taken over every gene the fold estimates rather than only those it still calls significant. The last column gives the largest share of the pooled weight carried by any single cohort.

Table S18: Leave-one-dataset-out results for each meta-analysis.

Analysis	Cohorts	Significant genes	Retained	min r	Largest cohort weight
UC inflamed vs control	15	6,765–7,760	88–94%	0.989	16%
UC uninflamed vs control	6	26–461	18–77%	0.853	34%
UC inflamed vs uninflamed	6	3,380–6,219	60–91%	0.941	30%
CD inflamed vs control	9	1,674–4,262	55–98%	0.970	26%
UC vs CD	7	1,715–3,883	51–93%	0.937	32%

Table S19 is the fold-level companion to this summary: it gives all 43 folds individually (the cohort omitted, its share of the pooled weight, the genes still significant, the fraction of the full list retained, and the fold-to-full effect-size correlation), as a separate CSV file.

Selection of the RNA-seq validation cohorts

The GEO search of the discovery analysis was repeated restricted to expression profiling by high-throughput sequencing, and the same inclusion criteria were applied with only the platform restriction lifted, so that the validation set is every qualifying dataset rather than a selection from among them.

Of 258 candidate series, 9 qualified. The remainder were excluded as follows: 83 single-cell or single-nucleus; 61 no uc vs control colonic comparison; 44 not primary tissue; 23 not intestinal mucosa; 14 no ulcerative colitis samples; 7 sorted or isolated cells; 4 fewer than 6 samples; 3 non-coding rna profiling.

Table S20: RNA-seq datasets meeting the inclusion criteria of the discovery analysis. Samples is the size of the full series; the counts entering each comparison are given in the Methods.

Accession	Samples	Basis for inclusion
GSE193677	2490	Adult registry, colonic biopsies, inflamed and non-inflamed UC
GSE109142	226	Treatment-naïve paediatric UC, rectal biopsies
GSE117993	190	Treatment-naïve paediatric IBD, rectal biopsies
GSE165512	170	Colonic biopsies, UC and healthy
GSE128682	44	Colonic biopsies, active UC, UC in remission and controls
GSE164918	36	Quiescent UC and controls, sigmoid colon; baseline biopsies only
GSE266325	23	Treatment-naïve paediatric IBD, rectal biopsies
GSE224758	22	Colonic biopsies, active UC and healthy volunteers
GSE334803	12	Colonic biopsies, UC and healthy controls

Cross-platform concordance

Tables S22–S25 give the gene- and pathway-level concordance between the microarray discovery meta-analysis and the independent RNA-seq validation meta-analysis, for the inflamed and uninfamed comparisons separately. Each file reports the effect size and q -value on both platforms, whether the gene or pathway was significant on each, and whether the two agree in direction. Replicated means significant on the microarray side and significant in the same direction on the RNA-seq side.

Table S21: Cross-platform concordance summary. r is the Pearson correlation of effect sizes (Hedges' g for genes, NES for pathways) over all shared features.

Comparison	Genes			Pathways		
	Shared	r	Replicated	Shared	r	Replicated
Inflamed UC vs controls	19,781	0.84	66%	1,781	0.89	72%
Uninflamed UC vs controls	16,635	0.42	6%	1,732	0.33	3%
Inflamed vs uninflamed UC	16,310	0.73	45%	1,727	0.79	59%