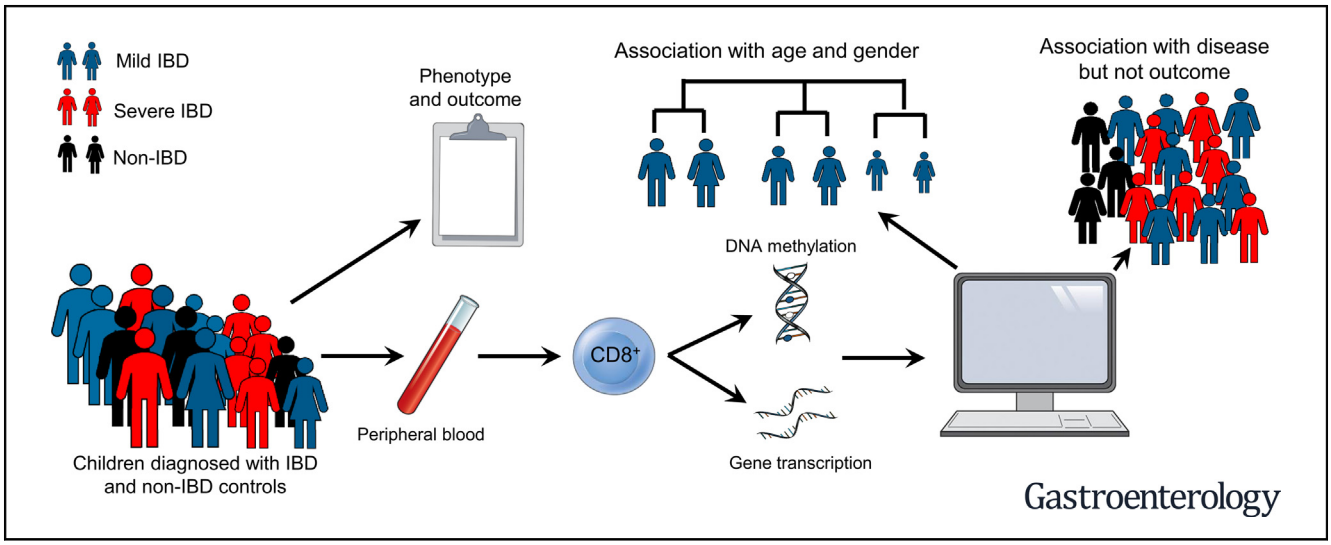
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Gastroenterology 2021;160:232–244

Transcription and DNA Methylation Patterns of Blood-Derived CD8$^+$ T Cells Are Associated With Age and Inflammatory Bowel Disease But Do Not Predict Prognosis

Marco Gasparetto,[1,2,*] **Felicity Payne**,[1,2,*] Komal Nayak,[1] Judith Kraiczy,[1] Claire Glemas,[1,2] Yosef Philip-McKenzie,[1] Alexander Ross,[1,3] Rachel D. Edgar,[4] Daniel R. Zerbino,[4] Camilla Salvestrini,[2] Franco Torrente,[2] Nicholas T. Ventham,[5] Rahul Kalla,[6] Jack Satsangi,[7] Peter Sarkies,[8,8] **Robert Heuschkel**,[2,*] and **Matthias Zilbauer**[1,2,3,*]

[1]Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom; [2]Department of Paediatric Gastroenterology, Hepatology and Nutrition, Cambridge University Hospitals, Addenbrooke's, Cambridge, United Kingdom; [3]Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom; [4]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom; [5]Academic Coloproctology, Western General Hospital, Edinburgh, Edinburgh, United Kingdom; [6]Medical Research Council Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom; [7]Translational Gastroenterology Unit, University of Oxford, Oxford, United Kingdom; [8]Medical Research Council London Institute of Medical Sciences, United Kingdom; and [8]Institute of Clinical Sciences, Imperial College London, United Kingdom



See editorial on page 21.

BACKGROUND & AIMS: Gene expression patterns of CD8$^+$ T cells have been reported to correlate with clinical outcomes of adults with inflammatory bowel diseases (IBD). We aimed to validate these findings in independent patient cohorts. **METHODS:** We obtained peripheral blood samples from 112 children with a new diagnosis of IBD (71 with Crohn's disease and 41 with ulcerative colitis) and 19 children without IBD (controls) and recorded medical information on disease activity and outcomes. CD8$^+$ T cells were isolated from blood samples by magnetic bead sorting at the point of diagnosis and during the course of disease. Genome-wide transcription (n = 192) and DNA methylation (n = 66) profiles were generated using Affymetrix and Illumina arrays, respectively. Publicly available transcriptomes and DNA methylomes of CD8$^+$ T cells from 3 adult patient cohorts with and without IBD were included in data analyses. **RESULTS:** Previously reported CD8$^+$ T-cell prognostic expression and exhaustion signatures were only found in the original adult IBD patient cohort. These signatures could not be detected in either a pediatric or a second adult IBD cohort. In contrast, an association between CD8$^+$ T-cell gene expression with age and sex was detected across all 3 cohorts. CD8$^+$ gene transcription was clearly associated with IBD in the 2 cohorts that included non-IBD controls. Lastly, DNA methylation profiles of CD8$^+$ T cells from children with Crohn's disease correlated with age but not with disease outcome. **CONCLUSIONS:** We were unable to validate previously reported findings of an association between CD8$^+$ T-cell gene transcription and disease outcome in IBD. Our findings reveal the challenges of developing prognostic biomarkers for patients with IBD and the importance of their validation in large, independent cohorts before clinical application.

Keywords: Epigenetic; Prognosis; Biomarker; Validation.

Inflammatory bowel diseases (IBD) such as Crohn's disease (CD) and ulcerative colitis (UC) are complex conditions that vary vastly in their phenotype and disease course. Among the most prominent factors to impact on disease presentation and behavior is the age of onset, with distinct phenotypic differences reported particularly in children diagnosed younger than 6 years.[1–3] In older children, who comprise the vast majority of childhood-onset IBD, phenotype and disease behavior are considered to be more similar to adult-onset IBD.[2] Nevertheless, our knowledge of factors contributing to phenotypic differences related to age of disease onset in IBD, such as affected cell types and specific molecular mechanisms involved, remains limited. $CD8^+$ T cells have been implicated in the pathogenesis of several immune-mediated diseases, including IBD.[4,5] An important finding linking $CD8^+$ T-cell biology to the pathogenesis of immune-mediated diseases was the discovery of a prognostic transcriptional signature in adult patients diagnosed with several conditions, including systemic lupus erythematosus, antineutrophil cytoplasmic antibody–associated vasculitis, as well as IBD.[4,6] Specifically, expression of distinct signature genes was found to divide patients into 2 groups that differed in their disease behavior. Moreover, T-cell exhaustion was proposed as an underlying mechanism as patients displaying an "exhausted" expression signature were shown to have a milder disease course.[7] In addition to gene transcription, increasing evidence points towards a major role for epigenetic mechanisms, such as DNA methylation, in regulating fundamental aspects of $CD8^+$ T-cell function, including proliferation, activation, and T-cell exhaustion.[8,9]

Given that existing evidence in this area has been exclusively derived from studies performed in adult IBD populations combined with a large body of evidence supporting age-related differences in T-cell function,[10,11] we set out to investigate $CD8^+$ T-cell biology in childhood-onset IBD. We prospectively recruited a cohort of 131 children newly diagnosed (treatment-naïve) with IBD (n = 112) and non-IBD controls (n = 19, Table 1) and isolated $CD8^+$ T-cells from peripheral blood samples. All patients were followed-up for a minimum of 18 months from diagnosis and detailed clinical phenotype as well as outcome and treatment data were documented (Supplementary Table 1). In addition, longitudinal samples were obtained from a subset of patients at 3-month intervals (n = 62 samples). In total, we generated 193 $CD8^+$ T-cell transcriptomes and 66 DNA methylomes and set out to analyze these datasets with a view to identify variation amongst patients and potential correlation with clinical outcome. Importantly, the use of publicly available and previously published datasets, including a further independent adult IBD cohort, allowed us to test for the presence of age-of-onset and cohort-specific differences in $CD8^+$ T-cell–derived molecular profiles.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Gene expression patterns of $CD8^+$ T cells was reported to correlate with outcomes of adults with inflammatory bowel diseases (IBD). T-cell exhaustion has been associated with IBD progression.

NEW FINDINGS

Transcription and DNA methylation patterns of $CD8^+$ T cells vary with patient age, sex, and IBD vs non-IBD but do not correlate with outcome.

LIMITATIONS

This was a study of 112 pediatric IBD patients and 19 controls; larger studies are needed.

IMPACT

The findings reveal the challenges of developing tests that determine prognoses of patients with IBD. All proposed prognostic markers must be validated in large, independent cohorts before they can be used in the clinic.

Methods

Ethical Approval

Ethical approval was obtained from the local research committee (REC 12/EE/0482 and REC 17/EE/0265) and patients were prospectively recruited after informed consent. All investigations were carried out according to the Declaration of Helsinki and Good Clinical Practice Guidelines.

Patient Recruitment and Clinical Data Recording

A total of 131 children aged between 4 and 17 years (median 13 years) were recruited prospectively between March 2013 and March 2016 at Cambridge University Hospitals National Health Service Foundation Trust in the Department of Paediatric Gastroenterology, Hepatology and Nutrition. Diagnosis of Inflammatory Bowel Disease was made according to current international guidelines (revised Porto criteria)[12] and included measurement of serum and stool inflammatory markers, upper and lower gastrointestinal endoscopies, and radiologic and histological examination. Any patient with gastrointestinal and/or extraintestinal diseases other than IBD was excluded from the study. In total, 71 patients were diagnosed with CD, 41 with UC, and 19 were classed as non-IBD, healthy controls. The latter were defined as patients who underwent endoscopic examination as part of their routine clinical care and were found to have normal macroscopic and histological appearance for their intestinal mucosa and complete resolution of any gastrointestinal symptoms.

All patients were followed up for a minimum of 18 months in the Cambridge pediatric gastroenterology unit, and detailed

* Authors share co-first authorship.

Abbreviations used in this paper: CD, Crohn's disease; IBD, inflammatory bowel disease; UC, ulcerative colitis; WGCNA, weighted gene co-expression network analysis.

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2020.08.017

Table 1. Summary of Pediatric Patients, Samples, and Molecular Profiles

Variable	CD[a]	UC[b]	Control
No. of patients	71	41	19
Age, *y*, median (range)	13 (6–17)	13 (5–16)	13 (4–16)
Sex, male, n (%)	46 (70)	45 (51)	8 (42)
Disease distribution, n (%)[c]			
CD–L1: 7 L1+ L4: 3	10 (14)	—	—
CD–L2: 13 L2+L4: 10	23 (32)	—	—
CD–L3: 23 L3+L4: 14	33 (52)	—	—
CD–L4	1 (1)		
UC–E1	—	5 (12)	—
UC–E2	—	7 (17)	—
UC–E3	—	6 (15)	—
UC–E4	—	23 (56)	—
CD8 purified samples, n			
At diagnosis[d]	71 (67)	41 (40)	19 (19)
Longitudinal[d]	36 (36)	26 (26)	—
With methylation data[d]	66 (66)	—	—

[a]107 samples from 71 patients.
[b]67 samples from 41 patients.
[c]Disease location for CD and UC at diagnosis as per Paris Classification for pediatric IBD.[12]
[d]Number of samples after quality control in parentheses.

clinical data was prospectively recorded using the hospital's patient electronic database (EPIC). This included phenotypic parameters at diagnosis (eg, presence of diarrhea, rectal bleeding, weight loss, extraintestinal manifestations, and perianal disease); disease activity scores: Pediatric Crohn's Disease Activity Index[13] and Pediatric Ulcerative Colitis Activity Index[14]; and information on disease course and outcomes. The latter covered number of treatment escalations, treatment history, response to treatment, requirement for surgical intervention, treatment with biologics and the number of unplanned inpatient admission days (Supplementary Table 1). In order to account for the fact that disease outcome is not restricted to a single measure, we designed a severity score that considered the number of treatment escalations, escalation to treatment with biologics, perianal disease, and surgery, as well as unplanned/urgent inpatient admissions (Supplementary Table 2). The score was calculated at 18 months from diagnosis and patients were categorized into mild, moderate, and severe groups after blinded scoring by at least 2 consultant pediatric gastroenterologists.

Blood Sampling and Processing

A peripheral blood sample (volume varied according to age from 10 to 25 mL) for the purification of $CD8^+$ T cells was taken on the same day as diagnostic endoscopy was performed. All children were treatment-naïve at this time point. Additional longitudinal samples (n = 62) were taken from a subset of children at 3-month intervals post diagnosis. Disease status was recorded as either "in remission" or "active disease" based on their clinical and biochemical disease activity scores Pediatric Crohn's Disease Activity Index[13] and Pediatric Ulcerative Colitis Activity Index[14] and the treatment history recorded (Supplementary Table 1). All samples were processed immediately and $CD8^+$ T cells isolated using magnetic bead sorting as detailed below.

Magnetic Bead Sorting

A peripheral blood sample of 10 mL (age 4–10 years) or 25 mL (age 10–17 years) was obtained and $CD8^+$ T cells extracted using magnetic bead sorting. Briefly, peripheral blood mononuclear cells were isolated by density centrifugation over Ficoll (Histopaque 1077; GE Healthcare, Chicago, IL) and $CD8^+$ T cells were separated by magnetic cell sorting using anti CD8 microbeads (Miltenyi Biotech, Bergisch Gladbach, Germany) as described by the manufacturer. Separation was performed on an AutoMACs Pro Separator (Miltenyi Biotech). Cell purity was regularly assessed on a subset of random samples using flow cytometry (see Supplementary Methods). With a mean of 84% (72.5%–93.8 %), purity of our samples was found to be similar or higher compared with other published datasets.

DNA and RNA Extraction

DNA and RNA were extracted simultaneously from isolated $CD8^+$ T-cell samples using AllPrep MiniKit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. DNA and RNA quality were assessed using an Agilent Bioanalyser 2100 (Agilent, Santa Clara, CA) and quantified by spectrophotometry using a NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA). DNA was bisulfite-converted using Zymo DNA methylation Gold kit (Zymo Research, Irvine, CA).

Genome Wide Transcriptional and DNA Methylation Profiling

Whole genome transcript analysis was performed on 200 ng of total RNA using the Affymetrix Human Gene ST version 2.0 Array (Affymetrix, Santa Clara, CA). Genome-wide DNA methylation was profiled using the Illumina EPIC platform (Illumina, Cambridge, UK). All microarray data have been deposited in ArrayExpress, accession numbers: E-MTAB-7923 (expression data) and E-MTAB-8925 (methylation data).

Publicly Available Datasets

Gene expression microarray data from 2 previously published adult patient cohorts were included. For adult cohort 1 (GSE87650),[15] a total of 52 $CD8^+$ T-cell–specific genome-wide transcriptomes (profiled on Illumina HumanHT-12, version 4) were available (n = 14 healthy controls, n = 19 CD, n = 19 UC). A summary of this patient cohort is provided in Supplementary Table 3. Adult cohort 2 included a total of 67 $CD8^+$ T-cell–specific transcriptomes (profiled on Affymetrix Human Gene ST, versions 1.0 and 1.1) with samples derived from patients diagnosed with CD (n = 35) and UC (n = 32) (E-MTAB-331).[6] Clinical phenotype data available for the latter cohort included age, sex, and disease severity classed as either severe (IBD1) or mild (IBD2). We obtained publicly available $CD8^+$ T-cell–derived genome-wide DNA methylation profiles (K450 Illumina DNA methylation arrays) from 2 previously published adult patient cohorts. Adult cohort 1 (GSE87640)[15] contained a total of 56 $CD8^+$ T-cell samples—18 CD, 19 UC, and 19 healthy control individuals (age range, 18–63 years). Adult cohort 3 included samples obtained from a younger (age range, 22–34

years, n = 50) and older (age range, 73–84 years, n = 50) group of healthy adults (GSE59065).[16]

Data analyses were performed separately in each dataset following the same analytical approach as outlined below, with adjustments made to account for differences in array type. Datasets derived from different studies were never combined or analyzed together.

Bioinformatic Analyses

A detailed description of bioinformatic methods used in this publication is provided in the Supplementary Methods section. Briefly, all analyses were performed in R, version 3.5.2. Preprocessing of Affymetrix gene expression array data included normalization of raw signal intensity data using the variance stabilization and calibration with robust multi-array average method as part of the *affy* package (version 1.56.0).[17] Preprocessing of Illumina gene expression array data was performed using the *lumi* package (version 2.34.0)[18] and normalized using robust spline normalization of log-transformed raw data. Quality-control assessment of all datasets was performed independently, using *arrayQualityMetrics* (version 3.34.0).[19] Samples failing quality control were removed, and batch correction was performed using the "ComBat" function as part of the *sva* package (version 3.26.0).[20] Our study design accounted for expected technical variation, including batch, by ensuring a balanced distribution of cases (ie, UC and CD) and controls between batches, as well as the inclusion of technical replicates. Successful batch correction was confirmed on technical replicates as well as principal variance component analyses. The latter was also used to demonstrate retention of biologic signals, such as sex, diagnosis, and age. Analyses were also performed on samples within individual batches and confirmed the results of combined batches. Data was annotated using the *hugene20sttranscriptcluster.db, hugene10sttranscriptcluster.db* or *lumiHumanIDMapping* package, dependent on array version. A total of 67 CD, 40 UC, 19 control, and 62 follow-up pediatric patient samples were retained for downstream analysis. Weighted gene co-expression network analysis (WGCNA) analyses were performed on normalized and batch-corrected datasets using the *WGCNA* package (version 1.63)[21] and resulting modules were correlated with clinical phenotypes as described previously.[22]

DNA methylation data was processed using the *minfi* package (version 1.28.0)[23] and included functional normalization[24] and quality-control assessment as described previously.[25] Published datasets included in this study were subjected to the same analyses. Epigenetic age and T-cell abundances were calculated using an established method developed by Horvath.[26]

Results

Variation of $CD8^+$ T-Cell Gene Transcription Shows Association With Disease, Age, Systemic Inflammation, and Sex

Transcriptional plasticity of $CD8^+$ T cells occurs during systemic inflammation and distinct differences have been reported in patients diagnosed with chronic inflammatory conditions, including IBD.[27,28] In order to determine the degree of variation in $CD8^+$ T-cell gene transcription within our sample cohort, we first performed principal component (PC) analyses and tested the correlation between observed variance and phenotype at diagnosis. For these analyses, we included samples obtained from children at the point of diagnosis (treatment naïve, UC n = 40, CD n = 67), as well as non-IBD controls (n = 19). Variation in $CD8^+$ T-cell gene transcription was found to be significantly associated with diagnosis (ie, difference between IBD and non-IBD controls; Figure 1*A*) and age (Supplementary Figure 1*A*), both contributing to the largest proportion of variance observed in PC1. Although a major overlap between IBD and control patient–derived samples was observed with 89% (ie, 17 of 19) of control samples clustering closely together, a proportion of IBD samples separated in PC1 (Figure 1*B*). Additional associations were identified for disease distribution, disease subtype (ie, UC vs CD), as well as albumin and a number of serum inflammatory markers, such as platelet count, erythrocyte sedimentation rate, and C-reactive protein. Furthermore, sex was found to be strongly associated with variance observed in PCs 8–10 (Figure 1*A*, Supplementary Figure 1*B*). Next, we performed similar analyses on previously published $CD8^+$ T-cell transcriptomes obtained from an adult patient cohort, including both IBD patients (n = 38) and healthy controls (n = 14, adult cohort 1, Supplementary Table 2).[15] As shown in Figure 1*C* and *D*, variation in $CD8^+$ T-cell–specific gene transcription in this adult patient cohort was also associated with diagnosis, age, and sex. Similar to the pediatric samples, the majority of adult control samples clustered closely together (78% of control samples) with a subset of IBD patient–derived $CD8^+$ T-cell transcriptomes separated out (Figure 1*D*). An association between $CD8^+$ T-cell transcription with age and sex was also confirmed in a second adult IBD patient cohort (adult cohort 2, Supplementary Figure 1*C*).

In order to investigate potential transcriptional changes over time and in response to treatment, we obtained longitudinal blood samples (n = 62) and isolated $CD8^+$ T cells from a subset of patients at various stages post diagnosis, including during early remission (3 months post induction), sustained remission (6 months post induction), and first and second relapse (Table 1). Although we did not observe major differences in $CD8^+$ gene expression based on specific treatment received (data not shown), samples obtained from patients in clinical remission appeared to cluster more closely to non-IBD controls (Supplementary Figure 1*D*), suggesting an impact of systemic inflammation on $CD8^+$ T-cell transcription.

In summary, variation observed in $CD8^+$ T-cell transcriptional profiles obtained from children and adults with IBD and controls showed significant association with diagnosis, age, systemic inflammation, and sex.

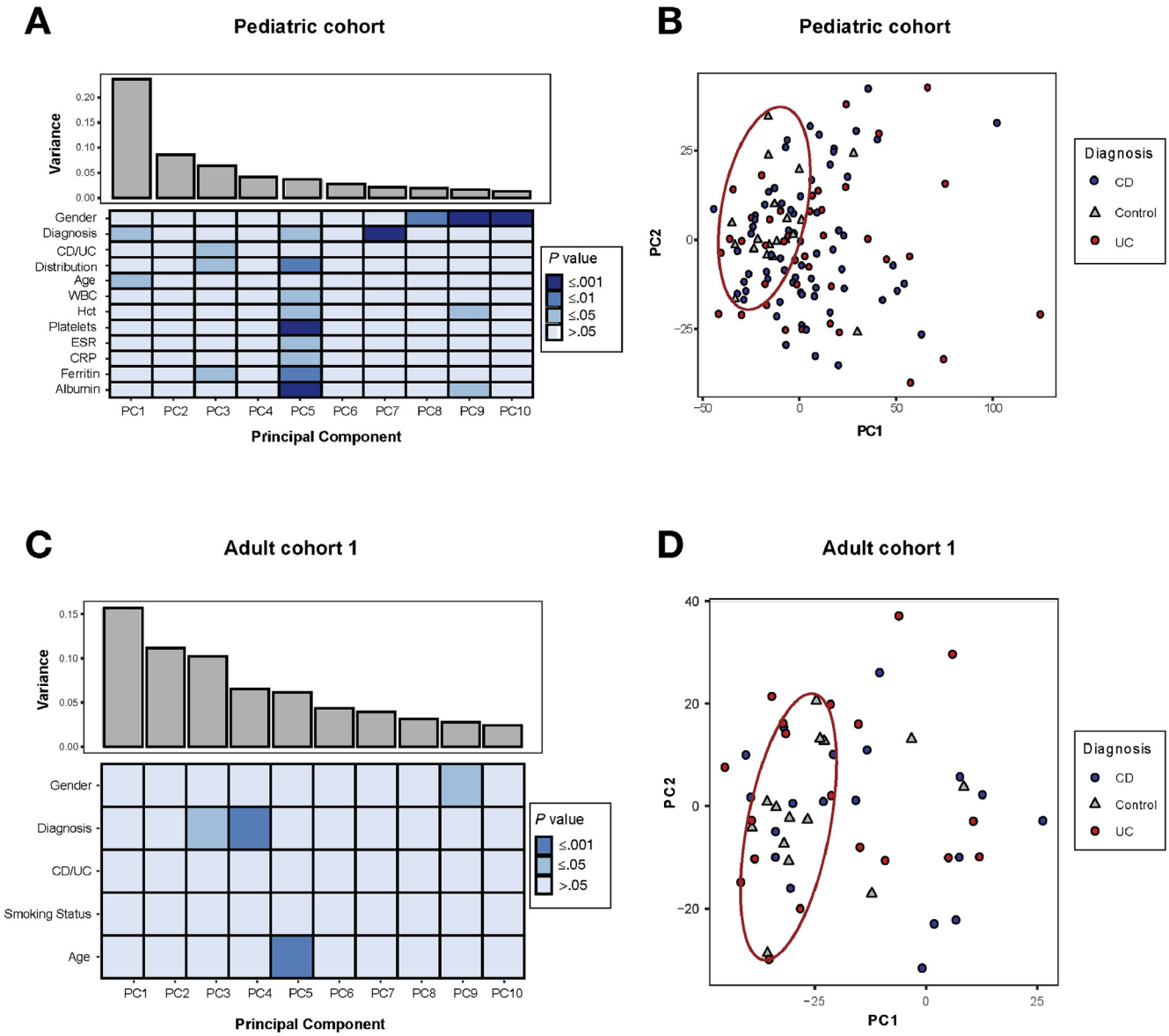


Figure 1. Genome-wide transcriptional profiles of CD8$^+$ T cells obtained from children newly diagnosed with CD (n = 67), UC (n = 40), and healthy controls (n = 19), as well as from a previously published adult patient cohort (CD n = 19, UC n = 19, non-IBD controls n = 14, adult cohort 1).[15] (*A*) Observed variance within pediatric CD8$^+$ T-cell transcriptomes (*top panel*) in the first 10 PCs. *Heatmap* displaying correlation between observed transcriptional variance and phenotype as well as selected serum markers at diagnosis (*bottom panel*). (*B*) *PC analysis plot* of pediatric CD8$^+$ T-cell transcriptomes illustrating close clustering of samples derived from non-IBD controls (*red circle* containing 17 of 19 control samples). (*C*) Variance in adult CD8$^+$ T-cell transcriptomes also shows association with diagnosis, sex, and age. (*D*) *PC analysis plot* of adult CD8$^+$ T-cell transcriptomes showing a similar distribution with close clustering of most non-IBD samples (*red circle* containing 11 of 14 control samples). *P* values were generated with a Kendall correlation for continuous variables, or an analysis of variance for categorical.

Testing for the Presence of a Previously Reported Prognostic CD8$^+$ T-Cell–Specific Expression Signature in Adult and Pediatric Inflammatory Bowel Disease Patient Cohorts

A CD8$^+$ T-cell–specific gene transcription signature has been shown to correlate with disease outcome in adult patients with several immune-mediated diseases, including antineutrophil cytoplasmic antibody–associated vasculitis, systemic lupus erythematosus, as well as adult patients diagnosed with CD and UC.[4,6] We analyzed 2 previously published adult (adult cohort 1 and 2) and our pediatric IBD patient–derived sample cohorts for the presence of the previously reported CD8$^+$ T-cell–specific, prognostic transcriptional signature.

Using the same analytical approach (Supplementary Methods),[6] we confirmed the presence of a CD8$^+$ T-cell expression signature in the adult CD and UC patient cohort in which the signature was first discovered (adult cohort 2).[6] As shown in Figure 2*A* and Supplementary Figure 2*A*, unsupervised clustering according to expression of signature genes separated CD and UC patient–derived CD8$^+$ T-

cell samples into 2 distinct groups that were reported to vary in disease outcome (IBD1 = severe; IBD2 = mild) based on number of treatment escalations. A major overlap between the CD and UC cohort–derived expression signatures was also confirmed (Supplementary Figure 2*B*). We next applied the same expression signature to our pediatric dataset, selecting genes of the core adult IBD signature and subjecting the resulting expression profiles to unsupervised clustering (Supplementary Methods). As shown in Figure 2*B*, $CD8^+$ T-cell expression of signature genes in pediatric IBD samples obtained at the point of diagnosis (treatment-naïve, active disease, n = 107) did not lead to any significant clustering. Given the major similarities in disease phenotype, behavior and pathogenesis between adolescent (12–16 years) and adult onset IBD, we performed additional analyses to test for the presence of the previously reported prognostic transcriptional signature in older children (ie, children diagnosed aged 12–16 years). As shown in Figure 2*C*, omitting younger IBD patients from the analyses yielded similar results, with no clear clustering observed. Next, we included a second adult IBD patient cohort (adult cohort 1) in our analyses and applied the previously reported expression signature. Similar to pediatric patient–derived $CD8^+$ T-cell transcriptomes, no clear clustering based on the expression of signature genes was observed in a total of 38 adult IBD patients (Figure 2*D*). No clustering could be identified when performing analyses after separating pediatric samples into CD and UC applying disease subtype–specific signatures (Supplementary Figure 2*C* and *D*), and when combining all samples, including CD, UC, non-IBD controls, as well as longitudinal samples (n = 188, Supplementary Figure 2*E*).

Together, our extensive analyses were unable to confirm the presence of the previously reported prognostic $CD8^+$ T-cell–derived transcriptional signature in our large pediatric or second adult IBD cohort.

$CD8^+$ T-Cell Transcription in Childhood Inflammatory Bowel Disease Lacks Correlation With Disease Outcome

In the absence of the previously reported adult expression signature, we next tested our entire pediatric dataset for the presence of a potential childhood-specific prognostic signature by correlating expression data with clinical outcome. Among other strategies we performed WGCNA, allowing the identification of gene groups (called eigengenes or modules) that correlate with clinical outcome measures as described previously (Supplementary Methods).[10] Applying WGCNA to adult CD patient–derived expression data (adult cohort 2) revealed several strong gene expression modules, some of which significantly correlated with reported patient outcome status (ie, IBD1/IBD2, Figure 3*Ai* and Supplementary Table 4). A major overlap of the genes forming these modules with the previously reported prognostic expression signature was observed (Figure 3*Aii*). Moreover, expression of genes defining the top module was found to cluster patients according to outcome status, confirming the validity of this analytical approach (Figure 3*Aiii*). In contrast to results obtained from the adult CD dataset, none of the gene modules resulting from the analysis of the pediatric CD dataset correlated significantly with any of the clinical outcome parameters tested, including number of treatment escalations, requirement for treatment with biologics, surgery, or with a summary disease outcome score designed to account for the fact that disease outcome is not defined by a single measurement (see Supplementary Methods and Supplementary Table 2 for details; Figure 3*Bi*). The validity of our approach in the pediatric sample cohort was confirmed by the identification of module–trait relationships for sex and age, although the latter only reached statistical significance in UC patient–derived samples (Figure 3*Bii*, Supplementary Table 4 and Supplementary Figure 3*A*). Lastly, we preformed WGCNA analyses on $CD8^+$ T-cell transcriptomes obtained from the older cohort of children diagnosed with CD or UC (aged 12–16 years). When omitting samples derived from children younger than 12 years, none of the resulting gene modules correlated significantly with any of the disease outcome measures we tested (Supplementary Figure 3*B* and *C*).

Taken together, our results demonstrate a lack of correlation between $CD8^+$ T-cell transcription and disease outcome parameters in children newly diagnosed with IBD.

Expression of Genes Associated With $CD8^+$ T-Cell Exhaustion Show Limited Variation in Pediatric Patient Samples

T-cell exhaustion has been reported as an underlying mechanism contributing to the observed differences in adult onset IBD disease behavior.[7] To investigate these findings in our pediatric patient cohort, we performed unsupervised clustering analyses based on genes associated with T-cell exhaustion as reported previously.[7] Although expression of exhaustion-related genes[29] clustered the original adult CD patient cohort[6]–derived transcriptomes into distinct subgroups (adult cohort 2, Figure 4*A*), we did not observe any significant clustering in $CD8^+$ T-cell transcriptomes of children newly diagnosed with CD, UC, all IBD, or when combining all pediatric samples (n = 188, Figure 4*B* and Supplementary Figure 4*A*–*C*). Furthermore, no clear clustering based on the expression of T-cell exhaustion–related genes was observed in $CD8^+$ T-cell transcriptomes of a second adult IBD cohort (adult cohort 1, Supplementary Figure 4*D*).

Together these findings suggest that T-cell exhaustion associated gene transcription in IBD patient–derived $CD8^+$ T–cells does not vary sufficiently to detect a reproducible clustering of patient samples in the cohorts we tested.

$CD8^+$ T-Cell DNA Methylation Shows Association With Age But Not With Disease Outcome in Children Newly Diagnosed With Crohn's Disease

DNA methylation has been shown to regulate key aspects of $CD8^+$ T-cell function, including cellular

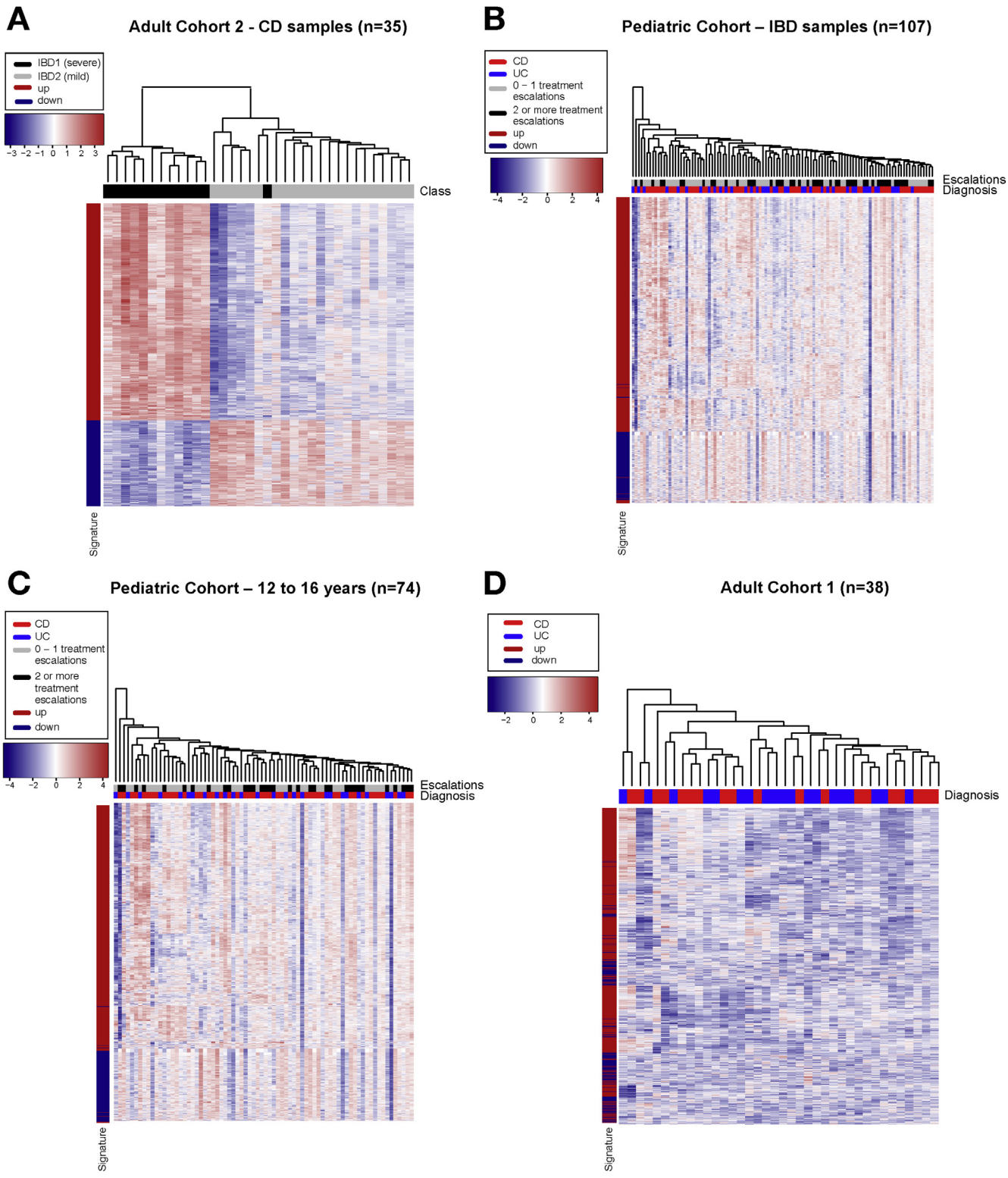


Figure 2. Disease prognostic transcriptional signature in adult and pediatric CD8$^+$ T-cell samples. Genes forming the previously reported transcriptional signature were selected from genome-wide transcriptomes and subjected to hierarchical clustering. (*A*) *Heatmap* and hierarchical clustering of genes forming the IBD prognostic expression signature in CD8$^+$ T cells obtained from adult patients diagnosed with CD (n = 35, adult cohort 2). (*B*, *C*) *Heatmap* and *hierarchical clustering tree* for genes forming the core adult IBD prognostic expression signature applied to CD8$^+$ T-cell transcriptomes of (*B*) all children newly diagnosed with IBD (n = 107), and (*C*) children diagnosed aged 12 to 16 years (n = 74). (*D*) Prognostic signature applied to a second adult IBD patient cohort (adult cohort 1, n = 38). Clustering was tested for statistical significance using M3C.

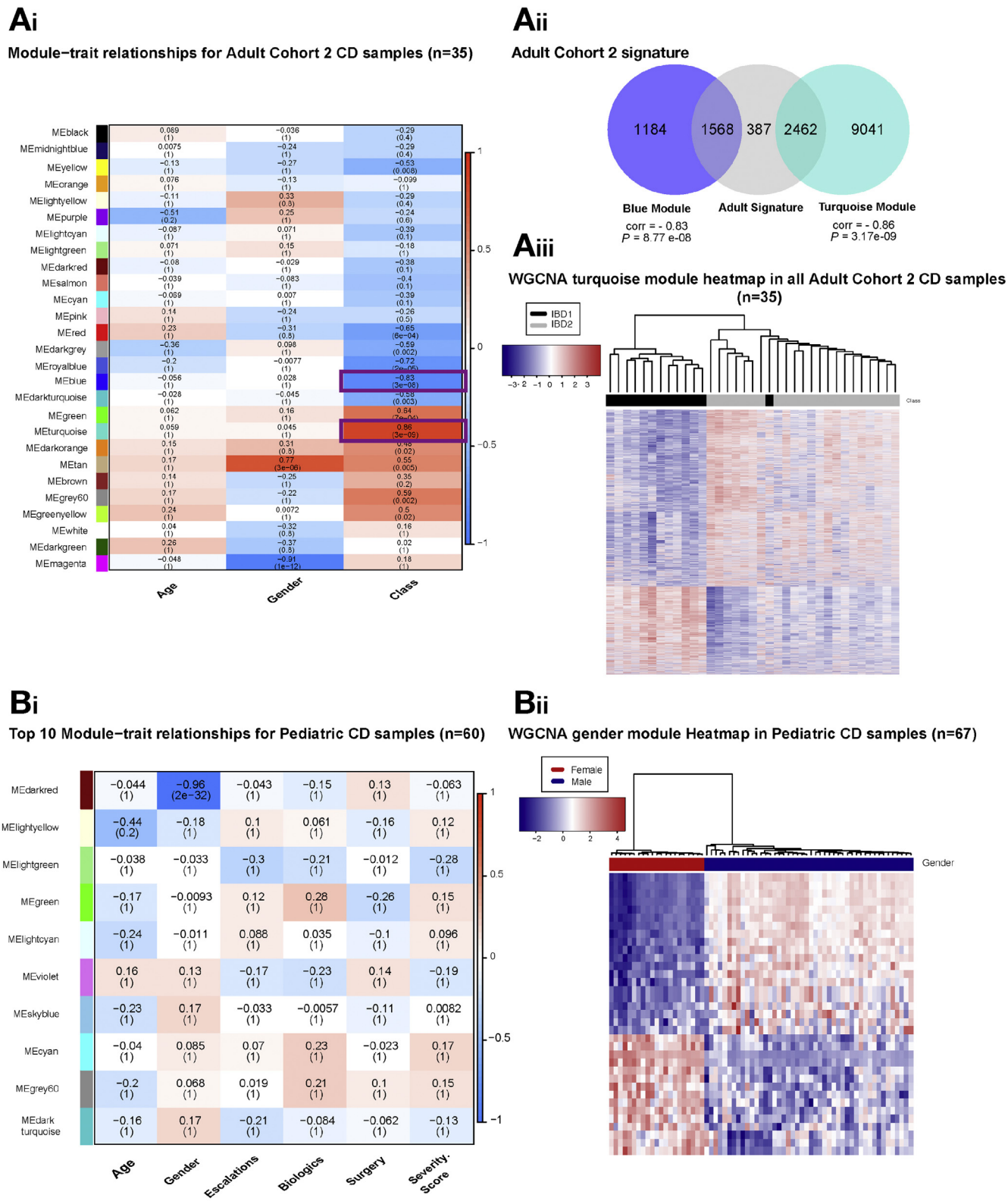


Figure 3. WGCNA of CD8$^+$ T-cell transcriptomes derived from adult-onset (adult cohort 2) (*A*) and childhood-onset (*B*) CD. (*Ai*) Module–trait relationship *heatmap* displaying gene modules and their correlation with age, sex, and outcome class (severe vs mild). *Numbers* indicate degree of correlation between module and trait (*top*) and corrected *P* value evaluating statistical significance of association (*in parentheses*). (*Aii*) Overlap between genes forming the adult IBD prognostic expression signature derived from Figure 1*A* (*gray*) and top correlated modules (*blue and turquoise*). (*Aiii*) Hierarchical clustering and *heatmap* of genes forming the *blue* module. (*Bi*) Top 10 module–trait relationship heatmap of CD8$^+$ T cells derived from children newly diagnosed with CD (n = 60). (*Bii*) Hierarchical clustering and *heatmap* based on genes forming the gene module associated with sex (*dark red* module). Samples of patients commenced on treatment with biologics at diagnosis were excluded from these analyses.

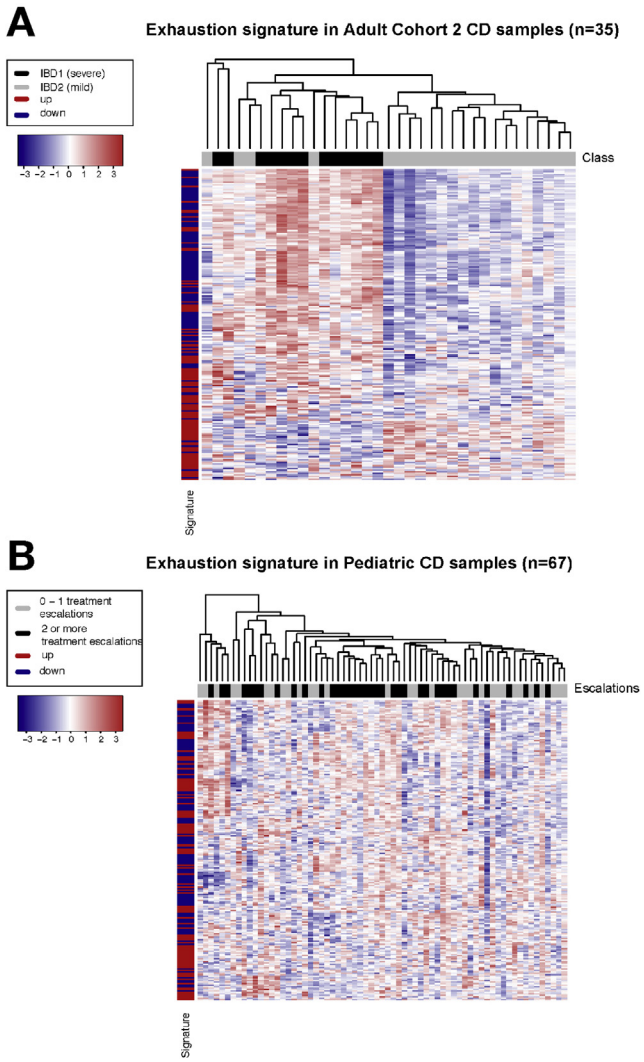


Figure 4. Transcriptional variation of genes associated with T-cell exhaustion in $CD8^+$ T cells. Genes forming a transcriptional T-cell exhaustion signature[29] were selected from genome-wide transcriptomes and subjected to hierarchical clustering. *Heatmap* and *hierarchical clustering trees* are displayed for samples obtained from adults (adult cohort 2, *A*, n = 35) and children (*B*, n = 67) diagnosed with CD.

differentiation and exhaustion.[9] Moreover, epigenetically mediated alterations in T-cell function have been implicated in disease pathogenesis of several immune-mediated conditions.[8,30,31] We performed genome-wide DNA methylation profiling of $CD8^+$ T cells derived from children newly diagnosed with CD (n = 66). Variation observed among these samples showed significant associations with disease distribution, sex, age, and inflammatory markers (Figure 5*A*). Similar to $CD8^+$ T-cell transcriptomes, we did not detect any significant correlation between DNA methylation and clinical outcome using various analytical approaches, including differential methylation and variance decomposition analyses (Figure 5*B* and data not shown). In order to investigate potential age-related changes in $CD8^+$ T-cell DNA methylation, we included publicly available genome wide methylomes obtained from 2 adult cohorts,[15,16] one of which included samples obtained from adult IBD patients (adult cohort 1).[15] Firstly, we applied an "epigenetic clock" algorithm[26] to these datasets and confirmed a significant correlation between the predicted age based on DNA methylation and chronological age ($P < .005$; Figure 5*C*). As DNA methylation has been used as a highly accurate measure to predict cellular composition, we next estimated the proportion of naïve and memory $CD8^+$ T cells in our dataset as well as the published adult datasets. As shown in Figure 5*D*, in both adult datasets we observed a significant inverse correlation for the estimated proportion of naïve $CD8^+$ T cells with age, as well as an age-dependent increase in the proportion of memory $CD8^+$ T cells in adult cohort 3 ($P < .005$).

Together, these results further support the lack of $CD8^+$ T-cell–derived molecular profiles with disease outcome in children newly diagnosed with CD and highlight age as a likely factor impacting on $CD8^+$ T-cell function and accounting for distinct differences between patient cohorts.

Discussion

Accurate disease prognostic biomarkers in IBD remain an essential requirement for effective personalized treatment by predicting clinical outcomes in individual patients. However, the search for parameters that reliably predict disease behavior at diagnosis remains challenging, with most studies identifying markers of short-term disease activity rather than longer-term clinical outcomes. An innovative and elegant approach of combining purification of individual leukocyte subsets with genome-wide transcriptional profiling led to the discovery of a prognostic expression signature in $CD8^+$ T cells and, more recently, the development of a whole blood–based clinical biomarker in adult IBD patients.[32] Interestingly, the reported prognostic expression signature was also found to be present in healthy adult control cohorts,[4] strongly suggesting differences in $CD8^+$ T-cell function are unrelated to chronic inflammation. Indeed, T-cell exhaustion was later identified to contribute to the observed differences. Specifically, these findings suggested that individuals displaying exhausted $CD8^+$ T-cell expression phenotypes experience a milder immune-mediated disease course, while being more susceptible to infectious diseases.[7]

Here, we investigated $CD8^+$ T-cell transcription in a large, prospectively recruited cohort of children, newly diagnosed with IBD and compared our dataset with 2 previously published adult IBD cohorts. Our results identified major differences between cohorts. Specifically, although we were able reproduce findings of the previously reported $CD8^+$ T-cell transcriptional signature in the original patient cohort (adult cohort 2),[6] we could not replicate these results in either our pediatric or a second adult IBD patient cohort (adult cohort 1)[15]–derived genome-wide profiles. Moreover, we were unable to find any correlation between $CD8^+$ T-cell gene transcription or DNA methylation and clinical outcomes in our pediatric patient cohort. This was despite including a wide range of clinically meaningful pediatric outcome measures, those used in previous adult studies (eg,

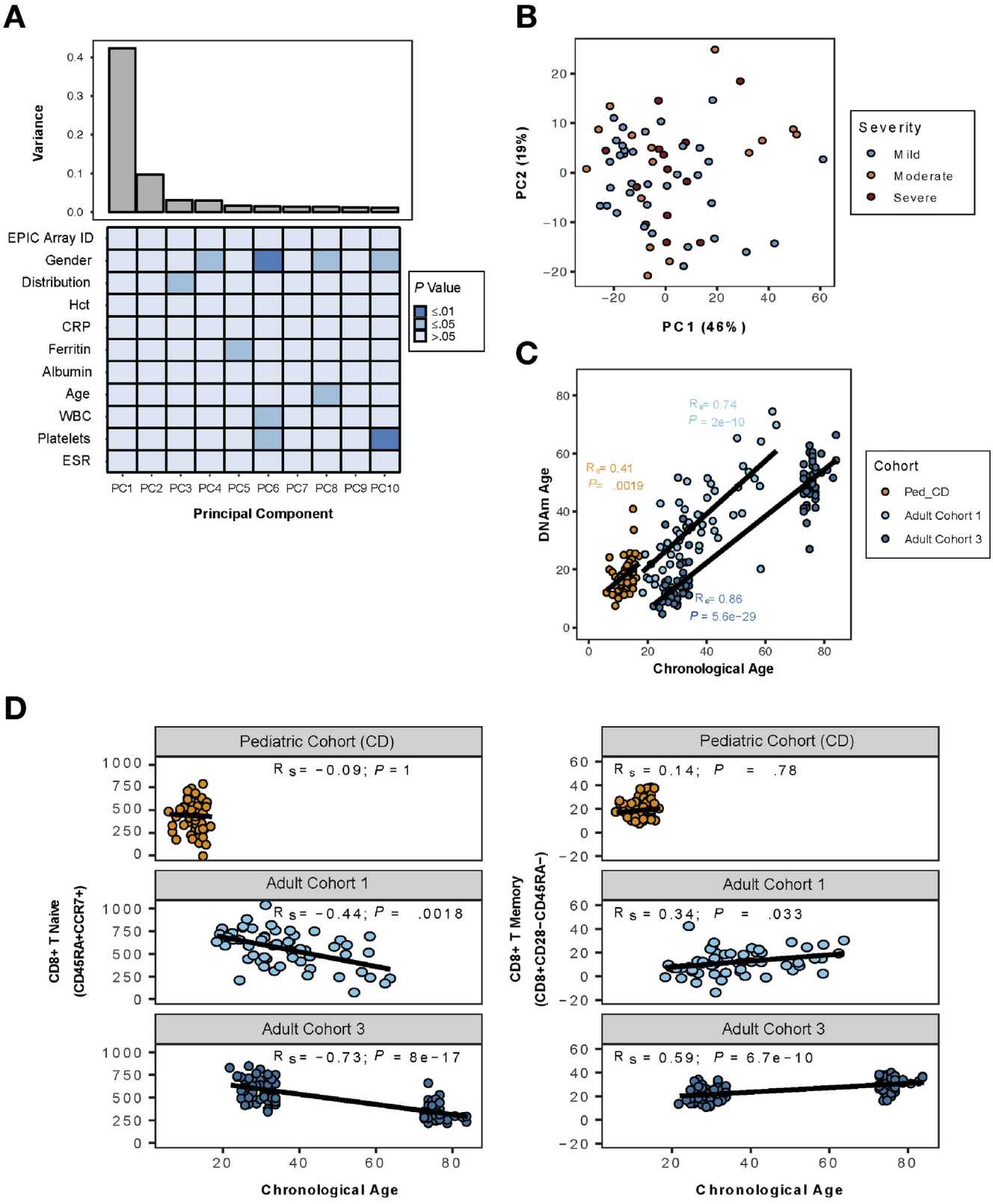


Figure 5. Genome-wide DNA methylation profiles derived from CD8$^+$ T cells. (*A*) Variation observed in CD8$^+$ T-cell DNA methylomes obtained from children newly diagnosed with CD (n = 66). Degree of variance in the first 10 PCs (*top panel*) and association with phenotype/serum parameters are indicated (*heatmap, bottom panel*). *P* values were generated with a Spearman correlation for continuous variables or an analysis of variance for categorical. (*B*) *PC analysis plot* of childhood CD patient–derived CD8$^+$ T-cell DNA methylomes. Samples labeled according to disease outcome (mild, moderate, severe). (*C*) Epigenetic clock applied to CD8$^+$ T-cell DNA methylomes derived from children newly diagnosed with CD, as well as 2 publicly available datasets (adult cohorts 1 and 3). Adult cohort 1 includes samples derived from adult patients diagnosed with IBD and healthy controls.[15] (*D*) Estimated proportion of naïve (*left*) and memory (*right*) CD8$^+$ T cells based on DNA methylomes demonstrating correlation with age in adult cohorts. Adult cohort 1 (n = 56),[15] adult cohort 3 (n = 98).[16]

number of treatment escalations), as well as a new summary outcome score, in which we combined key aspects of disease severity.

The differences observed in $CD8^+$ T-cell derived transcriptional profiles of 2 adult IBD patient cohorts in regard to the expression of genes forming the previously reported prognostic- or T-cell exhaustion signature are unexplained, as samples only formed distinct clusters in the original adult IBD cohort (ie, adult cohort 2). Furthermore, the clear clustering pattern of these signatures was also absent in $CD8^+$ T-cell–derived transcriptomes of older children (ie, diagnosed at age 12–16 years) in whom disease pathogenesis and behavior is considered very similar to adult-onset IBD.[2] Although the lack of clustering in 2 of 3 cohorts cannot completely exclude the presence of a subtler transcriptional signature, our results highlight the importance of validating molecular predictors of such clinical importance in independent cohorts. In this regard, the lack of consensus on defining disease outcome in IBD, as described recently by Dulai and colleagues,[33] makes the comparability of such studies difficult, and further adds to the challenges of developing a reliable, clinically relevant biological predictor. Hence, development of a validated clinical outcome score in IBD patients of all ages should be considered essential for the validation of existing and future biomarkers.

In contrast to disease outcome, our analyses identified significant and reproducible associations between $CD8^+$ T-cell gene transcription with sex and age across all 3 sample cohorts. Although the former is unlikely to be of major relevance to IBD pathogenesis, age associated changes in T-cell function might contribute to phenotypic differences between childhood- and adult-onset IBD. These findings are in keeping with a large body of existing evidence suggesting a major effect of human aging on immunocompetence.[34] Specifically, $CD8^+$ T-cell populations have been reported to be particularly affected by aging, as their ability to generate primary $CD8^+$ T-cell responses to newly encountered antigens decreases over time.[10] This effect can be partly attributed to a decrease in the number of naïve $CD8^+$ T cells with advancing age, as well as an increase in T-cell exhaustion.[10,11] Our data are in keeping with these reports, as we demonstrate a reduction in the estimated proportion of naïve T cells, as well as an increase in memory T cells with age. As expected, age-related DNA methylation changes were most apparent in those datasets from patient cohorts that covered a wide age range.[15,16] Beyond IBD, our findings are also relevant to age- and sex-related susceptibility to viral infections (eg, severe acute respiratory syndrome coronavirus 2).[35,36]

A limitation of our study was that we were unable to combine adult and pediatric datasets due to differences in array types, as well as the lack of matching, publicly available, clinical metadata. However, our study illustrates the power of comparing results obtained from the analyses of independent sample cohorts in order to validate biological and clinically relevant molecular signatures.

In summary, our study reveals an association between $CD8^+$ T-cell gene transcription and DNA methylation with age in 3 independent patient cohorts, suggesting a potential contribution of this mechanism towards age-associated differences in IBD phenotypes. In contrast, the presence of a previously reported prognostic expression signature was confirmed in only in 1 of 3 sample cohorts tested. In the absence of any significant association between $CD8^+$ T-cell gene transcription and DNA methylation and clinical outcomes in childhood-onset IBD, our results do not currently support the use of these molecular profiles as prognostic biomarkers in children. Importantly, our results highlight the need to validate clinical biomarkers in independent patient cohorts using reliable, clinically relevant outcome measures across all age groups.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dxdoi.org/10.1053/j.gastro.2020.08.017

References

1. Rosen MJ, Dhawan A, Saeed SA. Inflammatory bowel disease in children and adolescents. JAMA Pediatr 2015;169:1053–1060.
2. Herzog D, Fournier N, Buehr P, et al. Age at disease onset of inflammatory bowel disease is associated with later extraintestinal manifestations and complications. Eur J Gastroenterol Hepatol 2018;30:598–607.
3. Kerur B, Benchimol EI, Fiedler K, et al. Natural history of very early onset inflammatory bowel disease in north america: a retrospective cohort study. Inflamm Bowel Dis 2020 May 9 [Epub ahead of print].
4. McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. Nat Med 2010;16:586–591; 1p following 591.
5. Gravano DM, Hoyer KK. Promotion and prevention of autoimmune disease by CD8+ T cells. J Autoimmun 2013;45:68–79.
6. Lee JC, Lyons PA, McKinney EF, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. J Clin Invest 2011;121:4170–4179.
7. McKinney EF, Lee JC, Jayne DR, et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. Nature 2015;523:612–616.
8. Deng Q, Luo Y, Chang C, et al. The emerging epigenetic role of CD8+T cells in autoimmune diseases: a systematic review. Front Immunol 2019;10:856.
9. Ghoneim HE, Fan Y, Moustaki A, et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. Cell 2017;170:142–157.e19.
10. **Briceno O, Lissina A**, Wanke K, et al. Reduced naive CD8(+) T-cell priming efficacy in elderly adults. Aging Cell 2016;15:14–21.
11. **Lee KA, Shin KS**, Kim GY, et al. Characterization of age-associated exhausted CD8(+) T cells defined by increased expression of Tim-3 and PD-1. Aging Cell 2016;15:291–300.

12. **Levine A, Koletzko S**, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. J Pediatr Gastroenterol Nutr 2014;58:795–806.
13. Hyams JS, Ferry GD, Mandel FS, et al. Development and validation of a pediatric Crohn's disease activity index. J Pediatr Gastroenterol Nutr 1991;12:439–447.
14. Turner D, Otley AR, Mack D, et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. Gastroenterology 2007;133:423–432.
15. Ventham NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat Commun 2016;7:13507.
16. **Tserel L, Kolde R**, Limbach M, et al. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Sci Rep 2015;5:13107.
17. Gautier L, Cope L, Bolstad BM, et al. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307–315.
18. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–1548.
19. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics 2009; 25:415–416.
20. Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28:882–883.
21. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4(Article17).
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
23. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363–1369.
24. Fortin JP, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 2014;15:503.
25. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17:208.
26. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol 2013;14:R115.
27. **Funderburg NT, Stubblefield Park SR**, Sung HC, et al. Circulating CD4(+) and CD8(+) T cells are activated in inflammatory bowel disease and are associated with plasma markers of inflammation. Immunology 2013; 140:87–97.
28. Tom MR, Li J, Ueno A, et al. Novel CD8+ T-cell subsets demonstrating plasticity in patients with inflammatory bowel disease. Inflamm Bowel Dis 2016;22:1596–1608.
29. Wherry EJ, Ha SJ, Kaech SM, et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. Immunity 2007;27:670–684.
30. **Maltby VE, Graves MC**, Lea RA, et al. Genome-wide DNA methylation profiling of CD8+ T cells shows a distinct epigenetic signature to CD4+ T cells in multiple sclerosis patients. Clin Epigenetics 2015;7:118.
31. Li Y, Gorelik G, Strickland FM, et al. Oxidative stress, T cell DNA methylation, and lupus. Arthritis Rheumatol 2014;66:1574–1582.
32. **Biasci D, Lee JC**, Noor NM, et al. A blood-based prognostic biomarker in IBD. Gut 2019;68:1386–1395.
33. Dulai PS, Peyrin-Biroulet L, Danese S, et al. Approaches to integrating biomarkers into clinical trials and care pathways as targets for the treatment of inflammatory bowel diseases. Gastroenterology 2019;157:1032–1043.e1.
34. Quinn KM, Fox A, Harland KL, et al. Age-related decline in primary CD8(+) T cell responses is associated with the development of senescence in virtual memory CD8(+) T Cells. Cell Rep 2018;23:3512–3524.
35. **Chen J, Zhang ZZ, Chen YK, Long Q-X, Tian W-G**, et al. The clinical and immunological features of pediatric COVID-19 patients in China. Genes Dis 2020 Oct 20 [Epub ahead of print].
36. **Wu Y, Guo W**, Liu H, et al. Clinical outcomes of 402 patients with COVID-2019 from a single center in Wuhan, China. J Med Virol 2020 Jun 12 [Epub ahead of print].

Author names in bold designate shared co-first authorship.

Received March 31, 2020. Accepted August 11, 2020.

Correspondence
Address correspondence to: Matthias Zilbauer, MD, PhD, University Department of Paediatrics, Level 8 Addenbrooke's Hospital, Box 116, Hills Road, Cambridge, CB2 0QQ United Kingdom. e-mail: mz304@cam.ac.uk; fax: 00 44 (0) 1223 336996 or Robert Heuschkel, MBBS, MRCPCH, Department of Paediatric Gastroenterology, Box 267, Addenbrookes Hospital, Cambridge University Hospitals National Health Service Trust, Hills Road, Cambridge, CB2 0QQ United Kingdom. e-mail: robert.heuschkel@addenbrookes.nhs.uk; fax: 00 44 (0) 1223 336996.

Acknowledgments
The authors would like to thank Natalia Savinykh Yarkoni at the National Institute for Health Research Cambridge BRC Cell Phenotyping Hub for her support with cell sorting. The authors express their gratitude towards all patients and their parents who have participated in this study. The authors thank Bella's Fund for their support and contribution towards raising awareness for childhood IBD and charity funding for the project and the pediatric anesthetists at Cambridge University Hospitals for their help with obtaining blood samples from patients. The authors are also very grateful for the helpful discussions and guidance on data analysis provided by Theo Bammler and James MacDonald The University of Washington Department of Environmental and Occupational Health Sciences, Seattle, WA. Lastly, the authors express their gratitude to Paul Lyons (University of Cambridge, Department of Medicine) for his support in obtaining funding for the project and providing advice on data analyses.

Author contributions
Matthias Zilbauer and Robert Heuschkel designed the study, secured funding, led data analyses and wrote the manuscript. Marco Gasparetto obtained clinical information, contributed to the design, experimental work, analysis and writing of the manuscript. Felicity Payne contributed to the analysis of the study and writing of the manuscript. Komal Nayak, Alexander Ross, and Judith Kraiczy contributed to the experimental work. Claire Glemas and Yosef Philip-McKenzie obtained and documented clinical information. Camilla Salvestrini and Franco Torrente recruited patients, obtained blood samples and contributed to documentation of clinical information. Nicholas T. Ventham, Rahul Kalla, and Jack Satsangi provided data and clinical information for an independent adult IBD validation cohort. Peter Sarkies,

Daniel R. Zerbino, and Rachel D. Edgar contributed to data analyses. All authors contributed to experimental design, critical discussion of the findings and to the final manuscript. Marco Gasparetto and Felicity Payne contributed equally to this work.

Conflicts of interest
The authors disclose no conflicts.

Funding
This study was supported by grants from the Broad Medical Research Program at Crohn's and Colitis Foundation of America, Isaac Newton Trust/Wellcome Trust ISSF and University of Cambridge Joint Research Grant Scheme, Cambridge Biomedical Research Campus pump priming grant and Action Medical Research UK. Marco Gasparetto was supported by fellowships from the Crohn's in Childhood Research Association charity, the European Society for Paediatric Gastroenterology, Hepatology and Nutrition, and the IG-IBD International Exchange Program. Daniel R. Zerbino and Rachel D. Edgar were supported by core funding from the European Molecular Biology Laboratory. Peter Sarkies was supported by a grant from the UK Medical Research Council (Epigenetics and Evolution).

Transcript and DNA methylation profiling
All generated array data have been deposited in ArrayExpress, accession E-MTAB-7923 (gene transcription arrays), and E-MTAB-8925 (DNA methylation data).

Supplementary Methods

Flow Cytometry and Cell Purity Assessment

Flow cytometry was performed on a subset of samples pre- and post-magnetic cell sorting in order to assess purity using FACSCalibur (BD Biosciences, San Jose, CA). Briefly, after $CD8^+$ T-cell isolation, CD3-PE and CD8-APC antibodies were used (BD Pharmingen, San Diego, CA), along with the Zombie Aqua Fixable Viability Kit (Biolegend, San Diego, CA). A total of 8 randomly selected samples were subjected to flow cytometry analysis, and the mean cell purity for $CD8^+$ T cells was 84% (range, 72.5%–93.8%). Furthermore, we assessed gene expression of lymphocyte subset markers (eg, CD8, CD14, CD19, and CD16) using expression array data, and estimated cellular composition using DNA methylation data. Comparison between our data and published adult datasets confirmed at least similar if not higher purity of our samples (data not shown).

Bioinformatic Analyses

Affymetrix gene expression arrays. *Data preprocessing and quality control.* The raw signal intensity data were normalized using the variance stabilization and calibration with robust multi-array average method using *affy*, version 1.56.0[17] and quality control was performed using *arrayQualityMetrics*, version 3.34.0.[19] Samples failing quality control were removed and batch correction was performed using "ComBat" (part of *sva*, version 3.26.0).[20] Pediatric data were annotated using the *hugene20sttranscriptcluster.db* annotation package and the publicly available adult data (adult cohort 2) were annotated using the *hugene10sttranscriptcluster.db* annotation package. A total of 67 CD, 40 UC, 19 control, and 62 follow-up pediatric patient samples were retained for downstream analysis. Our study design accounted for expected technical variation, including batch, by ensuring a balanced distribution of cases (ie, UC and CD) and controls between batches, as well as the inclusion of technical replicates. Successful batch correction was confirmed using technical replicates and principal variance component analyses. The latter was also used to demonstrate retention of biologic signals, such as sex, diagnosis, and age. Analyses were also performed on samples within individual batches and confirmed the results of combined batches.

Illumina gene expression arrays. Publicly available raw expression bead array data, GSE87650[15] were downloaded from GEO DataSets using *GEOquery*, version 2.50.5.[e1] Robust spline normalization of log-transformed raw data was performed using the *lumi* package, version 2.34.0[18] and quality control was performed using *arrayQualityMetrics*, version 3.34.0.[19] Samples failing quality control were removed and batch correction was performed using "ComBat" (part of *sva*, version 3.26.0).[20] Data were annotated using the *lumiHumanIDMapping* package. A total of 19 CD, 19 UC, and 14 control patient samples were retained for downstream analysis.

Variance Decomposition Analysis

After quality control, in order to explore the main factors associated with variation within the dataset, PC analysis was performed on the normalized, batch-corrected pediatric dataset, and the first 10 PCs were examined for correlation with the sample clinical data using Kendall's test statistic for continuous variables and analysis of variance for categorical variables (Figure 1*A*). In addition, we investigated the distribution of sample variation with clinical phenotype more closely by plotting PCs (Figures 1*B* and *D*, Supplementary Figures 1*A*, *B*, and *D*).

Reproduction of the Adult $CD8^+$ T-Cell Prognostic Transcriptional Signature

In order to reproduce the previously reported adult $CD8^+$ T-cell prognostic transcriptional signature described by Lee et al,[6] we followed the protocol as described in their Supplementary Methods. Briefly, we used consensus clustering (merging results from k-means and hierarchical clustering from 5000 iterations with an 80% subsampling ratio) to identify optimal clustering within the published dataset described above (*clusterCons*, version 1.0). Differential expression analysis was then performed on the 2 groups of samples identified (*limma*, version 3.38.3) and a prognostic signature determined from probes significantly differentially expressed after stringent correction for multiple testing ($P < .05$, Holm).

Application of Adult $CD8^+$ T-Cell Prognostic and T-Cell Exhaustion Signature

Lists of probes corresponding to the reproduced adult $CD8^+$ T-cell prognostic signature[6] and the exhaustion signature described by Wherry et al.[29] (Supplementary Table 5) were used to extract data subsets. Hierarchical clustering was then performed using a Pearson correlation distance metric and average linkage analysis in R using *hclust* (part of the core statistics package) and *heatmap3* version 1.1.1 (Figures 2*A*–*D*).

Testing for Significance of Unsupervised Clustering Results

In order to assess clustering patterns for significance, we used the Monte Carlo Reference-based Consensus Clustering package (*M3C*),[e2] which uses the Relative Cluster Stability Index and Monte Carlo adjusted *P* values to select the optimal number of clusters (k) rejecting the null hypothesis ($k = 1$).

Weighted Gene Co-Expression Network Analysis

WGCNA was used to investigate correlations between gene expression profiles and clinical information, including disease outcomes (Supplementary Table 1) using *WGCNA*, version 1.63[21] as described previously.[22] Samples from patients given biologics at diagnosis were removed before analysis (60 CD and 39 UC samples remaining).

In brief, WGCNA uses unsupervised hierarchical clustering of normalized data to assess pair-wise correlation between gene expression profiles, implementing soft thresholding to assign a connection weight to each gene pair,[21] in order to reduce the dimensions of the dataset by grouping highly correlated genes into modules (eigengenes). Eigengene significance (correlation between sample trait and eigengene) and *P* values are then calculated for the whole module instead of individual genes, greatly alleviating multiple testing.

For our analysis, soft thresholding power was chosen based on the scale-free topology model parameters (scale independence and mean connectivity). Gene networks were constructed and modules identified from the resulting topologic overlap matrix (dissimilarity correlation threshold = 0.01, minimum module size: n = 30; deepSplit = 2). The resulting modules (eigengenes) were aligned to the clinical information, tested for correlation (using Pearson correlation) and Student asymptoptic *P* values calculated. False discovery rate correction for multiple testing of the resulting *P* values was performed using the Benjamini-Hochberg procedure.

DNA Methylation Analysis

Data pre-processing and quality control. Illumina EPIC arrays provided quantitative measures of DNA methylation (DNAm) for $CD8^+$ T cells from 66 patients, at single CpG resolution (>850,000 sites) covering the whole genome. DNAm data was processed using the *minfi* package, version 1.28,[23] specifically the "read.metharray" function, to extract beta values from raw IDAT files. Data was then normalized based on control probes on each array using functional normalization.[24] Starting with a total of 866,238 probes present on the EPIC array, the following probes were filtered out: polymorphic CpG,[25] located on sex chromosome, potential to cross hybridize to several regions of the genome,[25] poor quality as measured by a detection $P > .05$ in at least 1 % of samples. This filtering left 792,401 CpGs for analysis. Batch correction was performed using "ComBat" on array ID (part of *sva*, version 3.30.0).[20]

Correlating DNA Methylation With Clinical Outcome Data

To explore the main factors associated with variation in DNAm data, PC analysis was performed on the normalized, batch-corrected dataset. The first 10 PCs were examined for correlation with clinical data using Spearman's correlation for continuous variables and analysis of variance for categorical variables (Figure 5*A*). In addition, the distribution of PC1 and PC2 was compared with clinical data to explore the major contributors to variation in DNAm (Figure 5*B* and data not shown).

Differential Methylation Analysis

Differential methylation with severity score was tested at each CpG using *limma*, version 3.38.3.[e3] Models included covariates for age and sex. To be considered significantly differentially methylated, CpGs needed to have an association *P* value $<9 \times 10^{-8}$ (selected according to current recommendations)[e4] and an absolute methylation difference between the most severe and least severe of 0.05 ($\Delta\beta > 0.05$).

Differential methylation with age was tested in adult cohort 2 (GSE59065), as it had the greatest age range. Published DNAm data were processed as above, using annotation for the Illumina 450K array.[e5] Differential methylation at individual CpGs was tested using *limma*, version 3.38.3,[e3] while covarying for sex, using the same threshold criteria as described above.[e4]

To compare differential DNAm to the known lists of $CD8^+$ T-cell gene expression signatures, CpGs were associated with adjacent genes. A CpG was assigned to a gene based on proximity to a transcript from Ensembl Genes 99, GRCh37.p13, collected from BiomaRt.[e6] CpGs were classed as being associated with a gene if they were located between 1500 bp upstream of the transcription start and 300 bp downstream of the transcript end. CpG to transcript associations were then aggregated by Ensembl Gene stable ID.

Epigenetic Clock and Cell Subset Estimation Analysis

The normalized β values as provided under GSE87640 (adult cohort 1) and GSE59065 (adult cohort 3) were used to estimate epigenetic age. Using the Horvath epigenetic clock,[26] with the clock's normalization applied, the DNAm based age of each sample was estimated. As an extension, the estimated abundance levels of $CD8^+$ T-naïve and $CD8^+$ T-memory cells ($CD8^+CD28^-CD45RA^-$ T cells) were obtained from the "advanced blood analysis" of the online DNAm age predictor.[26] These measures indicate ordinal abundances from a regression with flow-sorted counts from other datasets. Hence, resulting data allows comparison of cell-subset proportions across samples over chronologic age within a cohort.

Supplementary References

e1. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–1847.

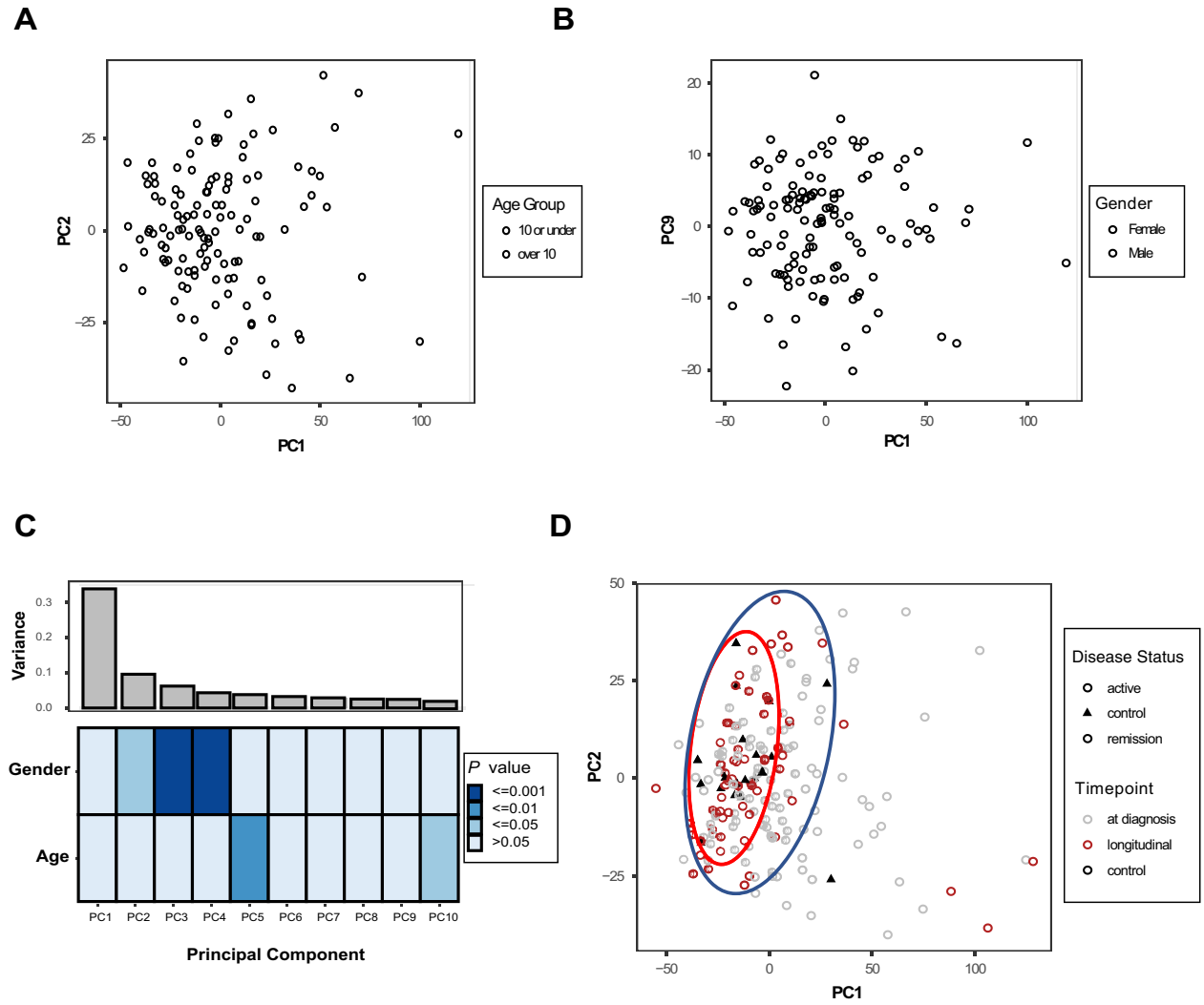
e2. John CR, Watson D, Russ D, et al. M3C: Monte Carlo reference-based consensus clustering. Sci Rep 2020; 10:1816.

e3. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

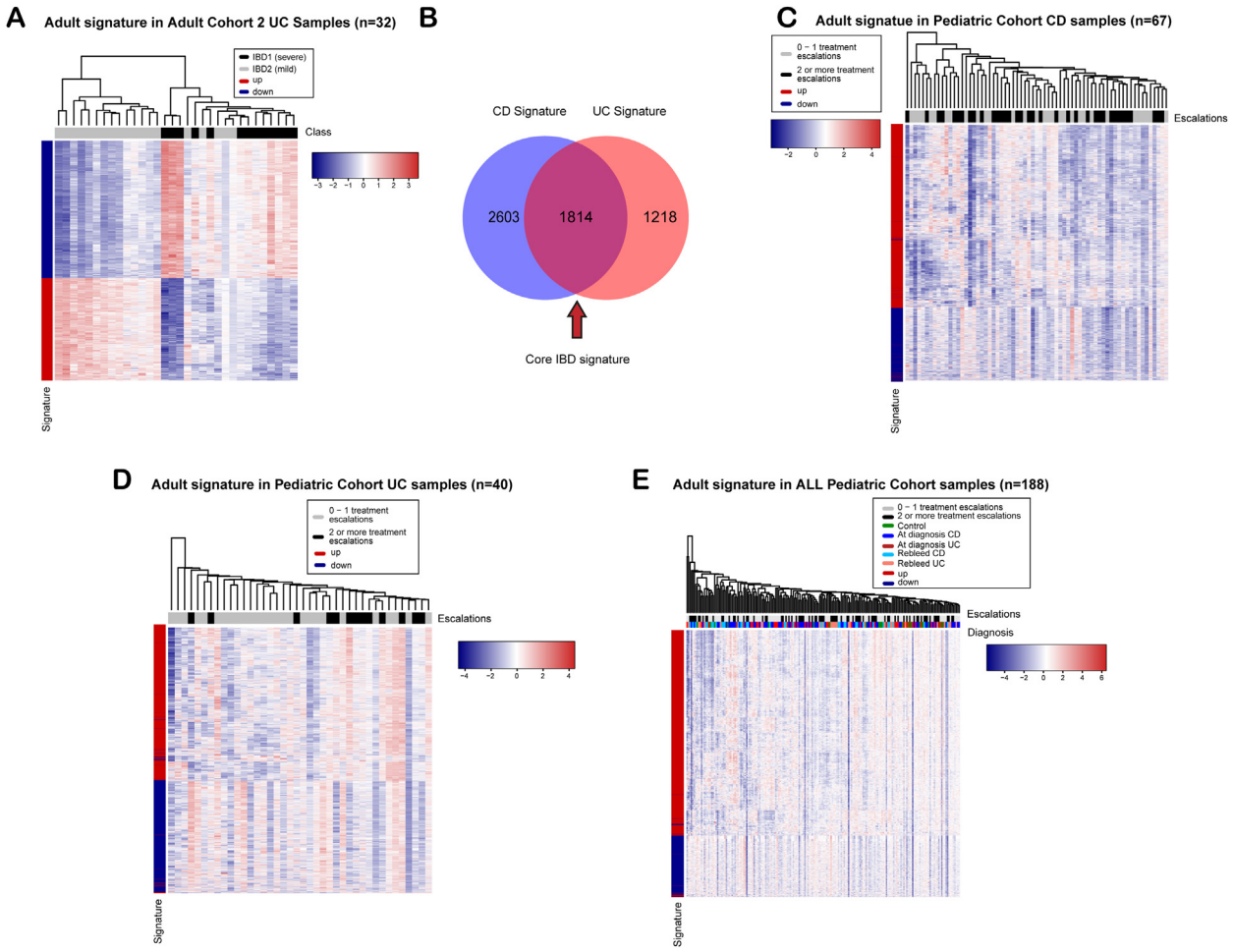
e4. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC Genomics 2019;20:366.

e5. Price ME, Cotton AM, Lam LL, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenet Chromatin 2013;6:4.

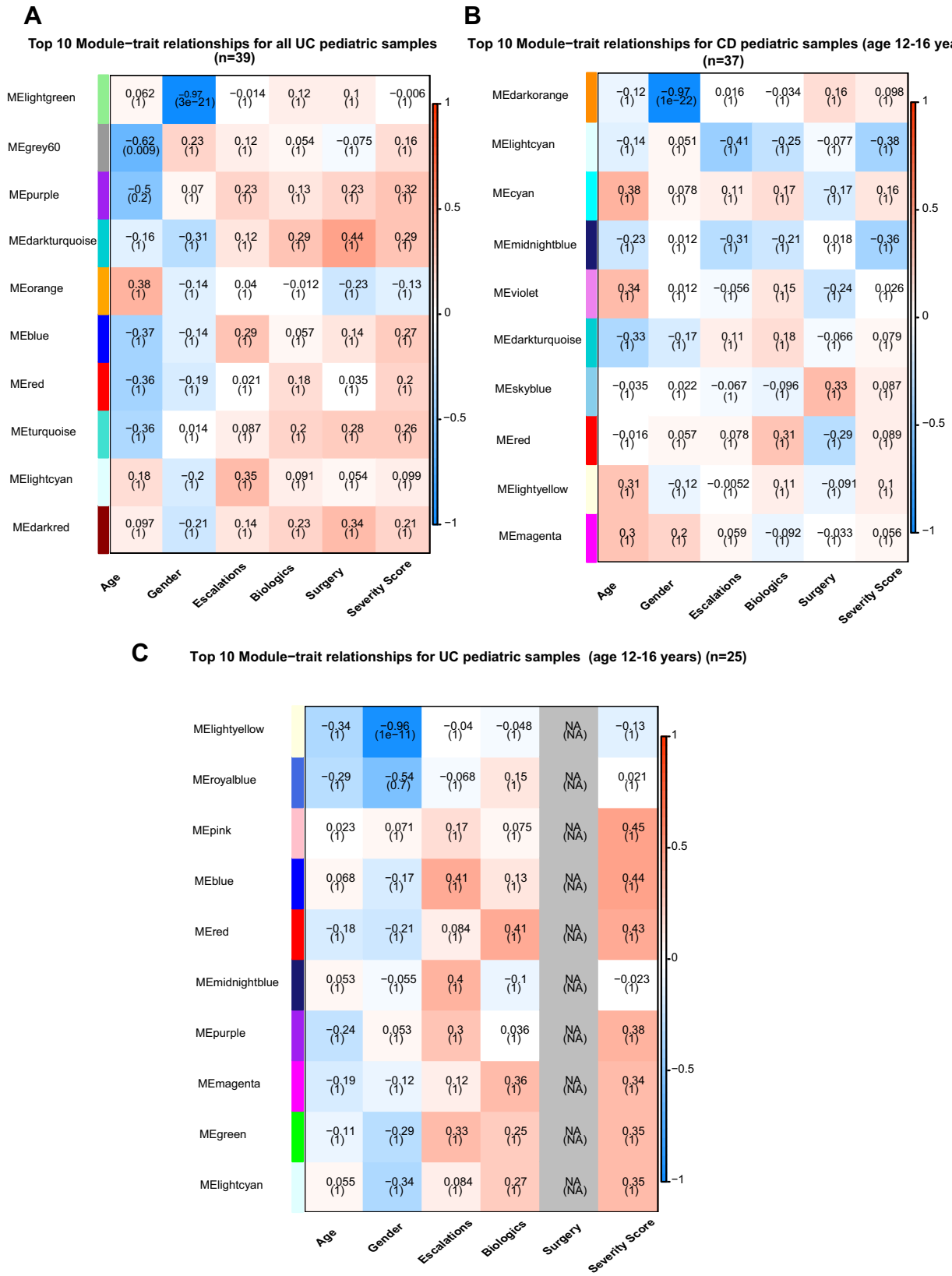
e6. Kinsella RJ, Kahari A, Haider S, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011;2011:bar030.



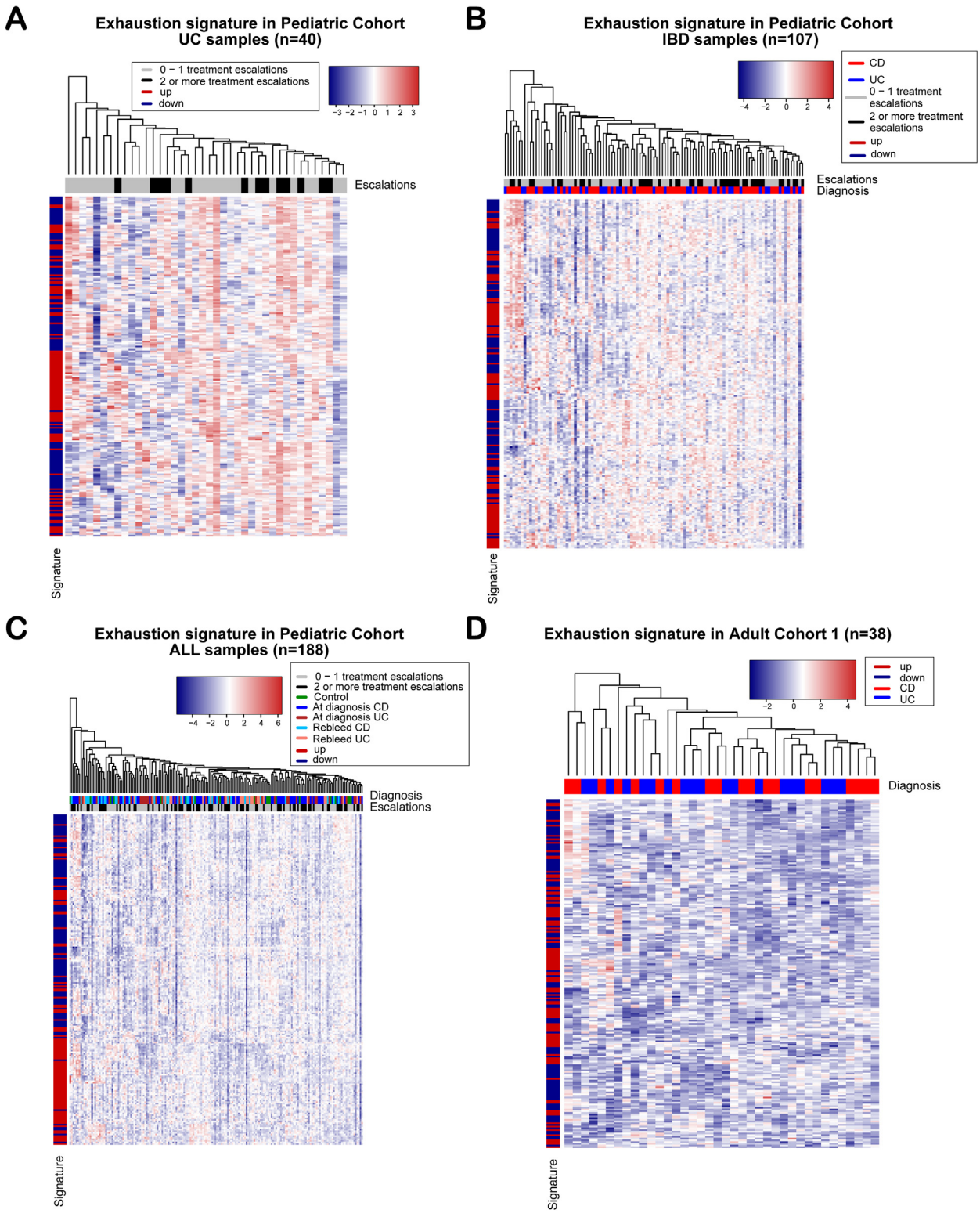
Supplementary Figure 1. (*A, B*) Genome-wide transcriptional profiles of $CD8^+$ T cells obtained from children newly diagnosed with CD (n = 67), UC (n = 40), and healthy controls (n = 19). PC analysis plots were generated after pre-processing and batch correction. (*A*) PC1 vs PC2 with samples labeled by age: 10 years or younger (n = 26, *red*) or older than 10 years (n = 100, *blue*). (*B*) Showing PC1 vs PC9 with samples labeled by sex: male (n = 74, *blue*), female (n = 52, *red*). (*C*) Assessment of variance within $CD8^+$ T-cell transcriptomes generated from adult patients diagnosed with CD (adult cohort 2, n = 35) in each PC (*top panel*). *Heatmap* displaying correlation between observed transcriptional variance with sex and age (*bottom panel*). (*D*) PC analysis plot of samples obtained from pediatric patients at diagnosis and during follow-up, illustrating overlap between the majority of samples derived from patients in clinical remission (*blue circle*, 30 of 32) and non-IBD controls (*red circle*, 17 of 19).



Supplementary Figure 2. Application of previously reported disease prognostic transcriptional signature to adult and pediatric CD8$^+$ T-cell samples. Genes forming the previously reported transcriptional signature were selected from genome-wide transcriptomes and subjected to hierarchical clustering. (*A*) *Heatmap* and hierarchical clustering of genes forming the UC-specific prognostic expression signature in CD8$^+$ T cells obtained from adult patients diagnosed with UC (n = 32, original adult cohort 2). (*B*) *Venn diagram* illustrating overlap of transcriptional signatures identified in adult patients diagnosed with CD (n = 35) and UC (n = 32) forming the core IBD signature. (*C–E*) *Heatmap* and hierarchical clustering tree for genes forming adult IBD prognostic expression signature applied to CD8$^+$ T-cell transcriptomes of (*C*) all children newly diagnosed with CD (n = 67, adult CD-specific signature applied), (*D*) all children newly diagnosed with UC (n = 40, adult UC-specific signature applied), and (*E*) all pediatric patient derived CD8$^+$ T-cell transcriptomes, including non-IBD controls and longitudinal samples (n = 188, adult core IBD signature applied).



Supplementary Figure 3. Application of WGCNA to CD8$^+$ T-cell transcriptional profiles derived from pediatric patients diagnosed with (*A*) UC (all ages, n = 39), (*B*) CD (age of disease onset 12–16 years, n = 37) and (*C*) UC (age of disease onset 12–16 years, n = 25). Displayed are the top 10 gene modules and their association with clinical phenotype (age at diagnosis, sex, number of treatment escalations, treatment escalation to biologics, surgery, and summary severity score). Statistically significant correlations were observed for sex ($P < .05$) in all cohorts and with age when including pediatric UC samples of all age groups (*A*). *Numbers* indicate degree of correlation between module and trait (*top*) and corrected *P* value evaluating statistical significance of association (in parentheses).



Supplementary Figure 4. Transcriptional variation of genes associated with T-cell exhaustion in $CD8^+$ T cells. Genes forming a transcriptional T-cell exhaustion signature (Wherry et al[29]) were selected from genome-wide transcriptomes and subjected to hierarchical clustering. *Heatmap* and *hierarchical clustering trees* are displayed for samples obtained from (*A*) children newly diagnosed with UC (n = 40), (*B*) all pediatric patient derived $CD8^+$ T-cell transcriptomes (n = 188, including controls and longitudinal follow-up samples) and (*C*) adult IBD patients (adult cohort 1, n = 38). No significant clusters were identified and patients with mild (ie, 0–1 treatment escalations in *gray*), or moderate to severe (ie, 2 or more treatment escalations in *black*) outcome are distributed equally in pediatric patients. The y-axis indicates expression of genes according to reference signature (*red* = up-regulated, *blue* = down-regulated).

Supplementary Table 2. Disease Outcome Severity Score

CD	UC
No. of treatment escalations 0–1: 0 2: 1 ≥3: 2	No. of treatment escalations 0–1: 0 2: 1 3: 2
Biologics No: 0 Yes: 2	Biologics No: 0 Yes: 2
CD-related surgery No: 0 Yes: 2	UC-related surgery (colectomy) No: 0 Yes: 2
Perianal disease Absent: 0 Medical management: 1 Surgical management: 2	Steroid-free remission at 3 mo from diagnosis (PUCAI score <10) Yes: 0 No: 2
Unplanned/urgent inpatient days 0–2: 0 3–4: 1 ≥5: 2	Unplanned/urgent inpatient days 0–2: 0 3–4: 1 ≥5: 2
Outcome based on total score 0–1: Mild 2–4: Moderate 5–10: Severe	Outcome based on total score 0–1: Mild 2–4: Moderate 5–10: Severe

PUCAI, Pediatric Ulcerative Colitis Activity Index.

Supplementary Table 3. Adult Cohort 1: Summary of Patients Passing Quality Control, Samples, and Molecular Profiles

Variable	CD	UC	Control
No. of patients	19	19	14
Age, *y*, median (range)	30 (18–63)	33 (22–62)	35 (20–58)
Sex, male, n (%)	11 (58)	13 (68)	10 (71)
Smoking status, n (%)			
Current smoker	8 (42)	1 (5)	2 (14)
Ex smoker	4 (21)	9 (47)	8 (57)
Never smoked	7 (37)	8 (42)	2 (14)
Unknown	0	1 (5)	2 (14)
Disease distribution,[a] n (%)			
CD–L1	3 (16)	—	—
CD–L2	7 (37)	—	—
CD–L3: 5 L3+L4: 1	6 (32)	—	—
UC–E2	—	6 (32)	
UC–E3	—	4 (21)	—
UC–E4	—	4 (21)	—
Unknown	3 (16)	5 (26)	—

[a]Disease location for CD and UC at diagnosis.