

Title: CD8⁺ T-cell transcription and DNA methylation show age specific differences and lack correlation with clinical outcome in pediatric Inflammatory Bowel Disease

Short Title: CD8⁺ T-cell gene expression and DNA methylation in childhood onset IBD

Authors: **Gasparetto M**^{1,2}, **Payne F**^{1,2}, Nayak K¹, Kraiczy J¹, Glemas C^{1,2}, Philip-McKenzie Y¹, Ross A^{1,3}, Edgar RD⁴, Zerbino D⁴, Salvestrini C², Torrente F², Sarkies P⁵, **Heuschkel R**², **Zilbauer M**^{1,2,3}.

Author names in bold designate shared co-first authorship

¹ Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

² Department of Paediatric Gastroenterology, Hepatology and Nutrition, Cambridge University Hospitals, Addenbrooke's, Cambridge, UK

³ Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK

⁴ European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

⁵ MRC London Institute of Medical Sciences, Imperial College London, UK

Grant support

This study was supported by grants from the Broad Medical Research Program at Crohn's and Colitis Foundation of America (BMRP, CCFA), Isaac Newton Trust / Wellcome Trust ISSF and University of Cambridge Joint Research Grant Scheme, Cambridge Biomedical Research Campus (BRC) pump priming grant and Action Medical Research UK. MG was supported by fellowships from the Crohn's in Childhood Research Association (CICRA) charity, the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and

the IG-IBD International Exchange Program. DZ and RDE were supported by core funding from the European Molecular Biology Laboratory (EMBL).

The authors would like to thank Natalia Savinykh Yarkoni at the at NIHR Cambridge BRC Cell Phenotyping Hub for her support with cell sorting. Furthermore, we express our gratitude towards all patients and their parents who have participated in this study. We thank Bella's Fund for their support and contribution towards raising awareness for childhood IBD and charity funding for the project. We would like to thank pediatric anaesthetists at Cambridge University Hospitals for their help with obtaining blood samples from patients. We 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, USA. 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.

Abbreviations

AAV	Anca Associated Vasculitis
CD	Crohn's Disease
CpG	Cytosine – phosphate - Guanine
DEG	differentially expressed gene
DNAm	DNA methylation
IBD	Inflammatory Bowel Disease
MACS	Magnetic Activated Cell Sorting
PBMC	Peripheral Blood Mononuclear Cells
PCDAI	Pediatric Crohn's Disease Activity Index
PDCD1	Programmed Cell Death 1
PUCAI	Pediatric Ulcerative Colitis Activity Index
PC	Principle Component

SLE	Systemic Lupus Erythematosus
UC	Ulcerative Colitis
WGCNA	weighted gene co-expression network analysis

Correspondence

Matthias Zilbauer MD PhD

University Department of Paediatrics

Level 8 Addenbrooke's Hospital, Box 116

Hills Road, Cambridge, CB2 0QQ

Phone: +44 (0) 1223 336889

e-mail: mz304@cam.ac.uk

Robert Heuschkel MD

Department of Paediatric Gastroenterology

Box 267, Addenbrookes Hospital

Cambridge University Hospitals NHS Trust

Hills Road, Cambridge, CB2 0QQ

Tel: 01223 274827

e-mail: robert.heuschkel@addenbrookes.nhs.uk

Disclosures

The authors declare no conflict of interest.

Transcript and DNA methylation Profiling

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

Writing Assistance

None

Author contributions

MZ and RH designed the study, secured funding, led data analyses and wrote the manuscript. MG obtained clinical information, contributed to the design, experimental work, analysis and writing of the manuscript. FP contributed to the analysis of the study and writing of the manuscript. KN, AR and JK contributed to the experimental work. CG and YP-M obtained and documented clinical information. CS and FT recruited patients, obtained blood samples and contributed to documentation of clinical information. PS, DZ and RE contributed to data analyses. All authors contributed to experimental design, critical discussion of the findings and to the final manuscript. MG and FP contributed equally to this work.

Abstract

Background & Aims: CD8⁺ T-cell gene expression has been implicated in the pathogenesis of Inflammatory Bowel Diseases (IBD) and has been shown to correlate with disease outcome in adult patients. Moreover, CD8⁺ T-cell exhaustion was identified as a contributing mechanism that impacts on disease behaviour. We aimed to explore CD8⁺ T-cell gene expression and DNA methylation in children newly diagnosed with IBD and test their correlation with disease outcome.

Methods: We prospectively recruited 112 children with IBD at the point of diagnosis and 19 non-IBD controls. Follow-up samples were obtained from a subset of patients at 3-month intervals (n=62). CD8⁺ T-cells were purified from peripheral blood samples using magnetic bead sorting and genome-wide transcriptional (n=192) and DNA methylation (n=66) profiles were generated using Affymetrix and Illumina arrays respectively. Publicly available adult CD8⁺ T-cell transcriptomes and DNA methylomes were included in data analyses to investigate age dependant differences.

Results: Variation amongst CD8⁺ T-cell transcriptomes obtained from children showed association with disease, systemic inflammation, age and gender but lacked correlation with disease outcome in pediatric IBD. In contrast to CD8⁺ T-cell transcriptomes in adult Crohn's Disease (CD), samples from pediatric patients did not show variation within genes forming part of the previously reported prognostic expression or T-cell exhaustion signatures. Pediatric CD patient derived DNA methylation profiles also lacked correlation with disease outcome but in comparison to adult CD8⁺ methylomes showed a higher predicted proportion of CD8⁺ naïve T-cells.

Conclusions: Our findings indicate age-related differences in IBD pathogenesis and highlight the importance of validating adult clinical biomarkers in pediatric cohorts.

Keywords

CD8⁺ T-cells; pediatric IBD; DNA methylation; gene transcription; prognostic biomarker

Introduction

Inflammatory bowel diseases (IBD) such as Crohn's Disease (CD) and Ulcerative Colitis (UC) are complex conditions which vary vastly in their phenotype. Amongst the most prominent factors known to impact on disease presentation and behaviour is the age of onset, as patients diagnosed in childhood tend to suffer from a much more aggressive and treatment resistant disease¹. Although this strongly suggests that key factors involved in the pathogenesis of IBD differ according to when the disease first manifests, our understanding of specific cell types or molecular mechanisms involved remains limited. CD8⁺ T-cells have been implicated in the pathogenesis of several immune mediated diseases including IBD^{2, 3}. 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 (SLE), Anca Associated Vasculitis (AAV) as well as IBD^{2, 4}. Specifically, expression of distinct signature genes was found to divide patients into two groups which differed in their disease behaviour. Moreover, T-cell exhaustion was proposed as an underlying mechanism as patients displaying an 'exhausted' expression signature were shown to suffer from a milder disease course⁵. 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^{6, 7}.

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, 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). Additionally, 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 analyse these datasets with a view to identify variation amongst patients and potential correlation with clinical outcome. Importantly, the use of publicly available datasets allowed us to test for the presence of age-of-onset specific differences in CD8⁺ T-cell molecular profiles.

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 following 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 (median 13) years were recruited prospectively between March 2013 and March 2016 at Cambridge University Hospitals NHS Foundation Trust in the Department of Paediatric Gastroenterology, Hepatology and Nutrition. Diagnosis of Inflammatory Bowel Disease (IBD) was made according to current international guidelines (revised Porto criteria)⁸ and included measurement of serum and stool inflammatory markers, upper and lower gastrointestinal endoscopies, radiological and histological examination. Any patient with gastrointestinal and/or extra-intestinal 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 clinical data was prospectively recorded using the hospital's patient electronic database (EPIC). This included phenotypic parameters at diagnosis (e.g. presence of diarrhoea, rectal bleeding, weight loss, extra-intestinal manifestations, perianal disease); disease activity scores: Pediatric Crohn's Disease Activity Index (PCDAI)⁹ and Pediatric Ulcerative Colitis Activity Index (PUCAI)¹⁰; 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 which considered the number of treatment escalations, escalation to treatment with biologics, peri-anal disease, surgery as well as unplanned/urgent inpatient admissions (Supplementary Table 2). The score was calculated at 18 months from diagnosis and patients were categorised into mild, moderate and severe groups following 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 PCDAI⁹ and PUCAI¹⁰ 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 (PBMCs) were isolated by density centrifugation over Ficoll (Histopaque 1077) and CD8⁺ T-cells were separated by magnetic cell sorting using anti CD8 microbeads (Miltenyi Biotech) 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 an average 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), according to the manufacturer's instructions. DNA and RNA quality were assessed using an Agilent Bioanalyser 2100 and quantified by spectrophotometry using a NanoDrop ND-1000 spectrophotometer. DNA was bisulfite-converted using Zymo DNA methylation Gold kit (Zymo Research).

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). 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 XXXX (methylation data). DNA was bisulfite- converted using Zymo DNA methylation Gold kit (Zymo Research).

Publicly available datasets

Gene expression microarray data (Affymetrix Human Gene ST v1.0 and 1.1) from a previously published adult patient cohort diagnosed with CD (n=35) and UC (n=32) was downloaded from ArrayExpress (E-MTAB-331)⁴. Clinical phenotype data available for this cohort included age, gender and disease severity classed as either severe (termed IBD1) or mild (termed IBD2). Secondly, we obtained publicly available CD8⁺ T-cell derived genome wide DNA methylation profiles (K450 Illumina DNA methylation arrays) from two previously published adult patient cohort. 'Adult cohort 1' (GSE87640)¹¹, contained a total of 56 CD8⁺ T-cell samples; 18 CD, 19 UC and 19 healthy control individuals (age range 18 to 63 years). 'Adult cohort 2' included samples obtained from a younger (age 22–34, n=50) and older (age 73–84, n=50) group of healthy adults (GSE59065)¹².

Bioinformatic analyses

A detailed description of bioinformatic methods used in this publication is provided in the supplemental methods section. Briefly, all analyses were performed in R version 3.5.2. Pre-processing of Affymetrix gene expression array data included normalization of raw signal intensity data using the variance stabilization and calibration with robust multi-array average (VSNRMA) method as part of the *affy* package (v1.56.0)¹³, and quality control assessment using *arrayQualityMetrics* (v3.34.0)¹⁴. Samples failing quality control were removed, and batch correction was performed using the "ComBat" function as part of the *sva* package (v3.26.0)¹⁵. Data was annotated using the *hugene20sttranscriptcluster.db* package. A total of 67 CD, 40 UC, 19 control and 62 follow-up pediatric patient samples were retained for downstream analysis. WGCNA analyses were performed on normalised and batch corrected datasets using the *WGCNA* package (v1.63)¹⁶ and resulting modules were correlated with clinical phenotypes as described previously¹⁷. DNA methylation data was processed using the *minfi* package (v.1.28.0)¹⁸ and included functional normalization¹⁹ and quality control assessment as described previously²⁰. 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 Horvath²¹.

Results

Variation of CD8⁺ T-cell gene transcription shows association with disease, age, systemic inflammation and gender

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^{22, 23}. In order to determine the degree of variation in CD8⁺ T-cell gene transcription within our sample cohort, we first performed principal component analyses (PCA) and tested the correlation between observed variance and phenotype at diagnosis. For these analyses we included samples passing quality control obtained from children at the point of diagnosis (treatment naïve, UC n=40, CD n=67, non-IBD controls n=19). Variation in CD8⁺ T-cell gene transcription was found to be significantly associated with diagnosis (i.e. difference between IBD and non-IBD controls; Figure 1A) and age (Supplementary Figure 1A), both contributing towards the largest proportion of variance observed in PC1. Indeed, although a major overlap between IBD and control patient derived samples was observed with 89% (i.e. 17 out of 19) of control samples clustering closely together, a proportion of IBD samples separated in PC1 (Figure 1B). Additional associations were identified for disease distribution, disease subtype (i.e. UC versus CD) as well as albumin and a number of serum inflammatory markers such as platelet count, Erythrocyte Sedimentation Rate (ESR) and C-Reactive Protein (CRP). Furthermore, gender was found to be strongly associated with variance observed in PCs 8 -10 (Figure 1A, Supplementary Figure 1B). 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), first and second relapse (Supplementary 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 (Figure 1C), further suggesting an impact of systemic inflammation on CD8⁺ T-cell transcription.

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

CD8⁺ T-cell transcription in childhood IBD lacks correlation with disease outcome

CD8⁺ T-cell gene transcription has been shown to correlate with disease outcome in adult patients with several immune mediated diseases including IBD^{2, 4}. Using publicly available adult CD patient derived CD8⁺ transcriptomes, we reproduced previously reported analyses and generated the prognostic expression signature (Supplementary methods)⁴. As shown in Figure 2A, unsupervised clustering of these samples according to expression of signature genes separated patients into two distinct groups that have been reported to vary in their disease outcome (termed IBD1=severe; IBD2=mild, based on number of treatment escalations). We next applied the same prognostic expression signature to our pediatric dataset, selecting genes comprising the adult signature and subjecting the resulting expression profiles to unsupervised clustering (Supplementary methods). As shown in Figure 2B, in contrast to the adult dataset, CD8⁺ T-cell expression of signature genes in pediatric CD (n=67) did not lead to any significant clustering. Similar results were obtained for UC patient derived samples (n=40, Figure 2C), and when combining all samples, including CD, UC, non-IBD controls as well as longitudinal samples (n=188, Figure 2D).

Given the absence of the previously reported adult expression signature, we next tested our entire dataset for the presence of a potential childhood onset specific prognostic signature by

correlating expression data with clinical outcome. Amongst other strategies we performed weighted gene co-expression network analyses (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 revealed several strong gene expression modules, some of which significantly correlated with reported patient outcome status (i.e. IBD1/IBD2, Figure 3Ai). A major overlap of the genes forming these modules with the previously reported prognostic expression signature was observed (Figure 3Aii). 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 3Aiii). 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 for details; Figure 3Bi). The validity of our approach in the pediatric sample cohort was confirmed by the identification of module-trait relationships for gender and age, although the latter only reached statistical significance in UC patient derived samples (Figure 3Bii and Supplementary Figure 2).

Taken together, our results demonstrate a lack of correlation between CD8⁺ T-cell transcription and disease outcome parameters in children newly diagnosed with IBD and highlight distinct differences to previously published datasets derived from adult IBD patients.

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 behaviour⁵. To investigate these findings in our patient cohort, we performed unsupervised clustering analyses based on genes associated with T-

cell exhaustion as reported previously⁵. Whilst expression of exhaustion related genes²⁴ clustered adult CD patient derived transcriptomes into distinct sub-groups (Figure 4A), we did not observe any significant clustering in CD8⁺ T-cell transcriptomes of children newly diagnosed with CD, UC or when combining all pediatric samples (n=188, Figure 4B and Supplementary Figure 3).

Together these findings suggest that T-cell exhaustion associated gene transcription does not vary significantly in CD8⁺ T-cells obtained from children diagnosed with IBD or non-IBD controls.

CD8⁺ T-cell DNA methylation shows association with age but not with disease outcome in children newly diagnosed with CD

DNA methylation has been shown to regulate key aspects of CD8⁺ T-cell function, including cellular differentiation and exhaustion⁷. Moreover, epigenetically mediated alterations in T-cell function have been implicated in disease pathogenesis of several immune mediated conditions^{6, 25, 26}. We performed genome-wide DNA methylation profiling of CD8⁺ T-cells derived from children newly diagnosed with CD (n=66). Variation observed amongst these samples showed significant associations with disease distribution, gender, age and inflammatory markers (Figure 5A). 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 5B 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^{11, 12}, one of which included samples obtained from adult IBD patients (Adult Cohort 1)¹¹. Firstly we applied an 'epigenetic clock' algorithm²¹ to these datasets and confirmed a significant correlation between the predicted age based on DNA methylation and chronological age ($p < 0.005$; Figure 5C). 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 5D, 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-dependant increase in the proportion of memory CD8⁺ T-cells in Adult cohort 2 ($p < 0.005$). Interestingly, amongst the CpGs which were found to display age dependant methylation changes, 265 were located near genes implicated in T-cell exhaustion and/or cell differentiation such as CD160 and Programmed Cell Death 1 (PDCD1, Supplementary Table 3, Supplementary Figure 4). 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 personalised treatment, by predicting clinical outcomes in individual patients. However, the search for parameters that reliably predict disease behaviour 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 patients²⁷. Interestingly, the reported prognostic expression signature was also found to be present in healthy adult control cohorts², 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, whilst being more susceptible to infectious diseases⁵.

Here, we investigated CD8⁺ T-cell transcription in a large, prospectively recruited cohort of children, newly diagnosed with IBD. Our findings stand in clear contrast to those described in adult patient populations, as we were unable to replicate the previously reported CD8⁺ T-cell transcriptional signature in any of our 193 genome wide profiles. Moreover, we were unable to find any correlation between CD8⁺ T-cell gene transcription or DNA methylation and clinical outcomes. This was despite including a wide range of clinically meaningful pediatric outcome measures, those used in previous adult studies (e.g. number of treatment escalations), as well as a new summary outcome score, in which we combined key aspects of disease severity. Defining disease outcome is complex. There continues to be a lack of consensus on outcomes in the search for clinical biomarkers in IBD, as described recently by Dulai and colleagues²⁸. Nevertheless, despite the presence of variation in disease outcome measures within our patient cohort, the clear lack of correlation with variation of CD8⁺ T-cell gene expression or DNA methylation does not support their use as a prognostic biomarker in childhood onset IBD. Given the near-identical methodology, and the application of additional novel analytical approaches to both adult and pediatric datasets, the differences we observe are most likely to be due to the age difference between cohorts. Indeed, human ageing is well known to have a major effect on immunocompetence, generally leading to a gradual decline²⁹. 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 time³⁰. 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^{30, 31}. 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, several were associated with T-cell exhaustion and/or cellular differentiation. As expected, age related DNA methylation changes were most apparent in datasets from patient cohorts that covered a wide age range^{11, 12}. Similarly, in contrast to adult CD patient derived CD8⁺ T-cells, we did not observe major variation in gene transcription of T-cell exhaustion related genes within the pediatric dataset. We therefore speculate that our findings indicate a relative absence of CD8⁺ T-cell exhaustion

in children, perhaps contributing to the commonly seen severe and extensive phenotype in childhood onset IBD. Further studies are required to address this concept in more detail.

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. Hence, observed age-related differences in CD8⁺ T-cell molecular profiles require confirmation in future studies, ideally containing IBD patients of all ages.

Conclusions

In summary, our study reveals distinct differences in CD8⁺ T-cell transcriptomes between adult and childhood onset IBD, with the latter lacking correlation with disease outcome. The observed correlation between age and CD8⁺ T-cell gene transcription as well as DNA methylation suggests that our findings may reflect distinct differences in disease pathophysiology between adult and childhood onset IBD. Importantly, our results highlight the need to validate adult clinical biomarkers in independent pediatric patient cohorts.

Table 1: Summary of patients, samples and molecular profiles

	CD	UC	Control
Number of patients	n = 71	n = 41	n = 19
Median age (years) ^a	13 (6 – 17)	13 (5 – 16)	13 (4 – 16)
Gender (% male) ^b	46 (70%)	45 (51%)	8 (42%)
Disease distribution: ^c			
CD – L1: 7 L1+ L4: 3 ^b	10 (14%)	-	-
CD – L2: 13 L2+L4: 10 ^b	23 (32%)	-	-
CD – L3: 23 L3+L4: 14 ^b	33 (52%)	-	-
CD – L4 ^b	1 (1%)		
UC – E1 ^b	-	5 (12%)	-
UC – E2 ^b	-	7 (17%)	-
UC – E3 ^b	-	6 (15%)	-
UC – E4 ^b	-	23 (56%)	-
CD8 purified samples:			
At diagnosis ^d	71 (67)	41 (40)	19 (19)
Longitudinal ^d	36 (36)	26 (26)	-
With methylation data ^d	66 (66)	-	-

CD, Crohn's disease (107 samples from 71 patients)

UC, ulcerative colitis (67 samples from 41 patients)

IBD, inflammatory bowel disease

^a Age range in parentheses

^b Percentage of samples in parentheses

^c Disease location for CD and UC at diagnosis as per Paris Classification for pediatric IBD⁸

^d Number of samples after quality control in parentheses

References

1. Rosen MJ, Dhawan A, Saeed SA. Inflammatory Bowel Disease in Children and Adolescents. *JAMA Pediatr* 2015; 169:1053-60.
2. McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat Med* 2010; 16:586-91, 1p following 591.
3. Gravano DM, Hoyer KK. Promotion and prevention of autoimmune disease by CD8+ T cells. *J Autoimmun* 2013; 45:68-79.
4. 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; 121:4170-9.
5. McKinney EF, Lee JC, Jayne DR, et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* 2015; 523:612-6.
6. 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.
7. 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.
8. 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.
9. Hyams JS, Ferry GD, Mandel FS, et al. Development and validation of a pediatric Crohn's disease activity index. *J Pediatr Gastroenterol Nutr* 1991; 12:439-47.
10. 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; 133:423-32.

11. 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.
12. 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.
13. Gautier L, Cope L, Bolstad BM, et al. *affy*--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; 20:307-15.
14. Kauffmann A, Gentleman R, Huber W. *arrayQualityMetrics*--a bioconductor package for quality assessment of microarray data. *Bioinformatics* 2009; 25:415-6.
15. 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; 28:882-3.
16. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; 4:Article17.
17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9:559.
18. 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; 30:1363-9.
19. 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.
20. 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.
21. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013; 14:R115.

22. 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.
23. 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; 22:1596-608.
24. Wherry EJ, Ha SJ, Kaech SM, et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* 2007; 27:670-84.
25. 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.
26. Li Y, Gorelik G, Strickland FM, et al. Oxidative stress, T cell DNA methylation, and lupus. *Arthritis Rheumatol* 2014; 66:1574-82.
27. Biasci D, Lee JC, Noor NM, et al. A blood-based prognostic biomarker in IBD. *Gut* 2019; 68:1386-1395.
28. 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.
29. 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.
30. Briceno O, Lissina A, Wanke K, et al. Reduced naive CD8(+) T-cell priming efficacy in elderly adults. *Aging Cell* 2016; 15:14-21.
31. 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.

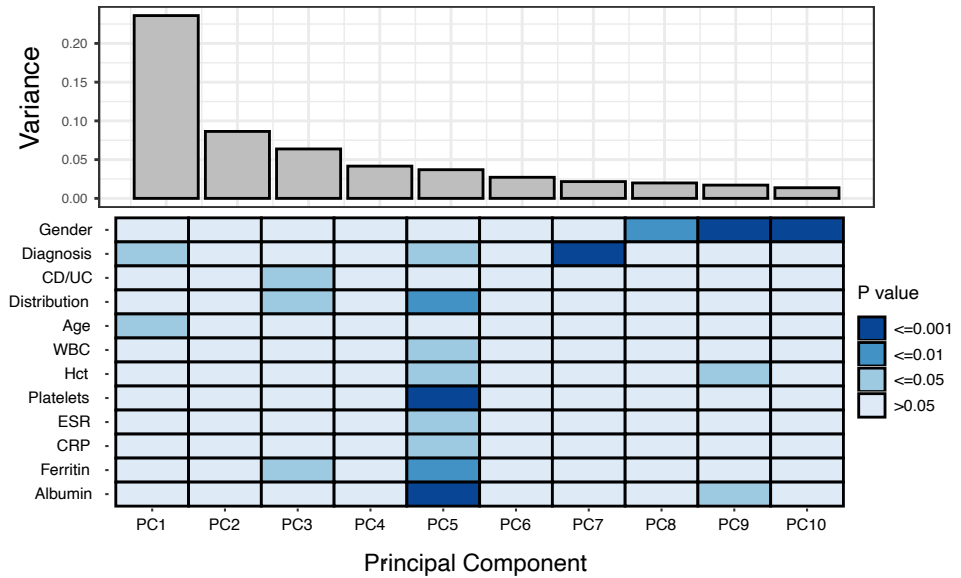
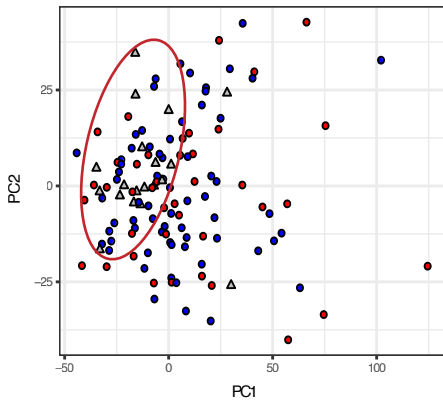
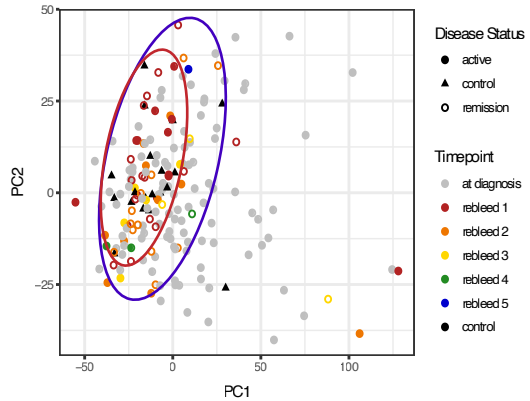
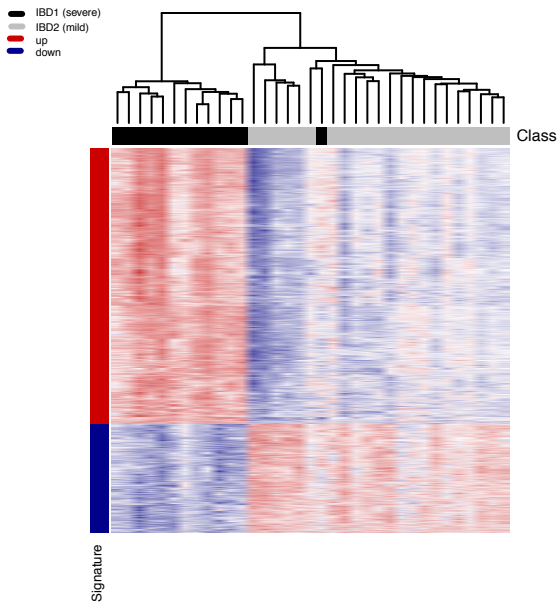
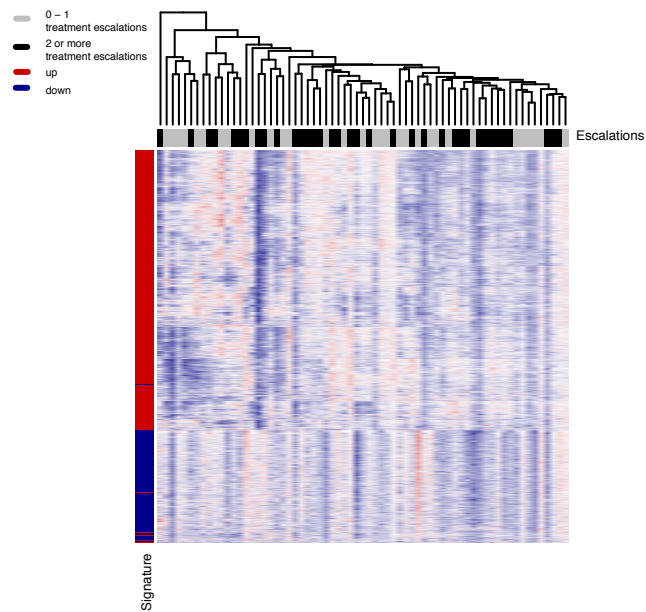
A.**B.****C.**

Figure 1: Genome wide transcriptional profiles of CD8⁺ T-cells obtained from children newly diagnosed with Crohn's Disease (n=67), Ulcerative Colitis (n=40) and healthy controls (n=19). Samples were obtained at diagnosis (A&B, n=126) as well as during follow up (n=62, included in C). **A)** Observed variance within CD8⁺ T-cell transcriptomes (top panel) in each principle component (PC). Heat map displaying correlation between observed transcriptional variance and phenotype as well as selected serum markers at diagnosis (bottom panel). P values were generated with a Kendall correlation for continuous variables, or an ANOVA for categorical. **B)** PCA plot of CD8⁺ T-cell transcriptomes illustrating close clustering of samples derived from non-IBD controls (red circle containing 17/19 control samples). **C)** PCA plot of samples obtained from patients at diagnosis and during follow up, illustrating overlap between the majority of samples derived from patients in clinical remission (blue circle, 30/32) and non-IBD controls (red circle 17/19).

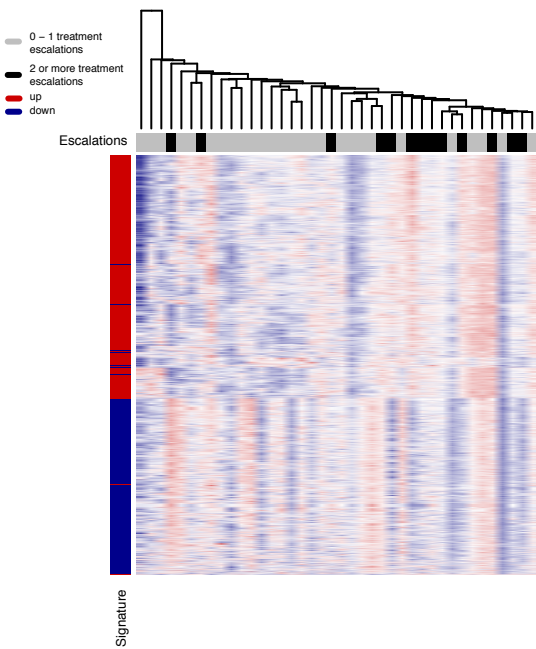
A. Adult Signature in Adult CD Samples (n=35)



B. Adult Signature in Pediatric CD Samples (n=67)



C. Adult Signature in Pediatric UC Samples (n=40)



D. Adult Signature in ALL Pediatric Samples (n=188)

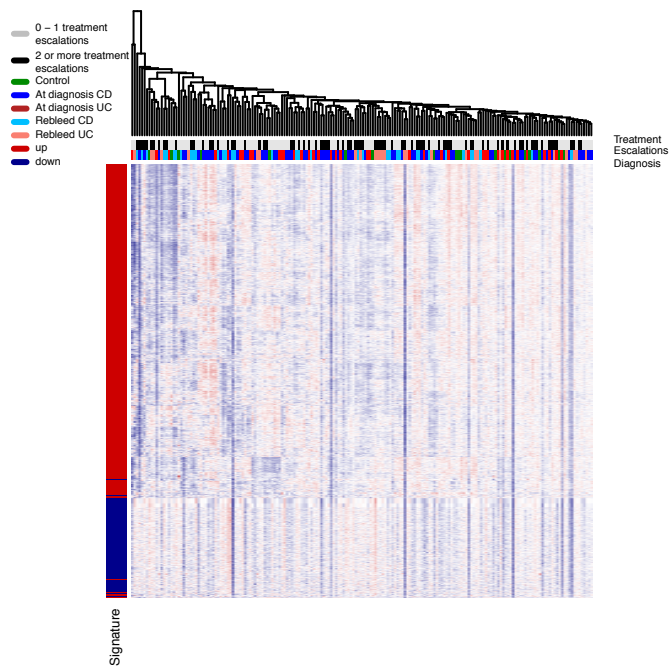
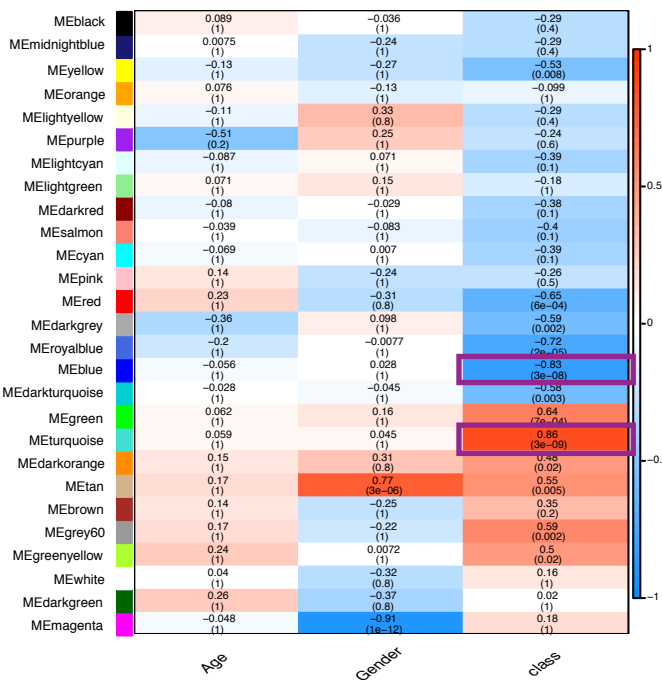


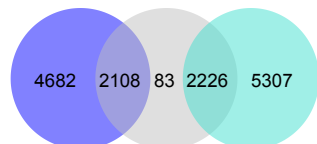
Figure 2: Disease prognostic transcriptional signature in adult and pediatric CD8⁺ T-cell samples. Genes forming the transcriptional signature were selected from genome-wide transcriptomes and subjected to hierarchical clustering. **A)** Heat map and hierarchical clustering of genes forming the IBD prognostic expression signature in CD8⁺ T-cells obtained from adult patients diagnosed with CD (n=35). **B-D)** Heat map and hierarchical clustering tree for genes forming part of the adult IBD prognostic expression signature applied to CD8⁺ T-cell transcriptomes of children newly diagnosed with CD (n=67, B), UC (n=40, C) and all generated pediatric transcriptional profiles (n=188, D). Clustering was tested for statistical significance using M3C.

Ai.

Module-trait relationships for adult CD samples (n=35)

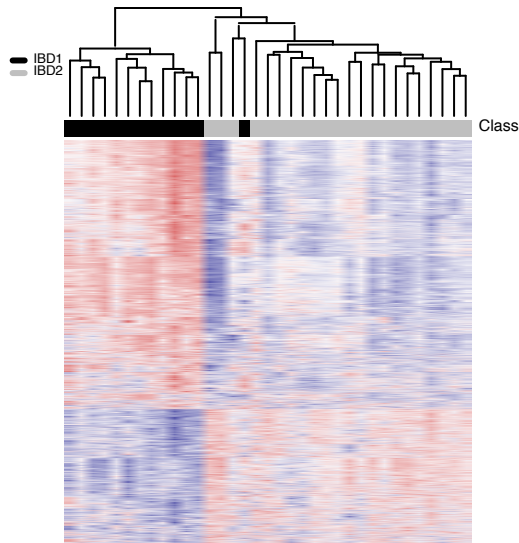


Aii.



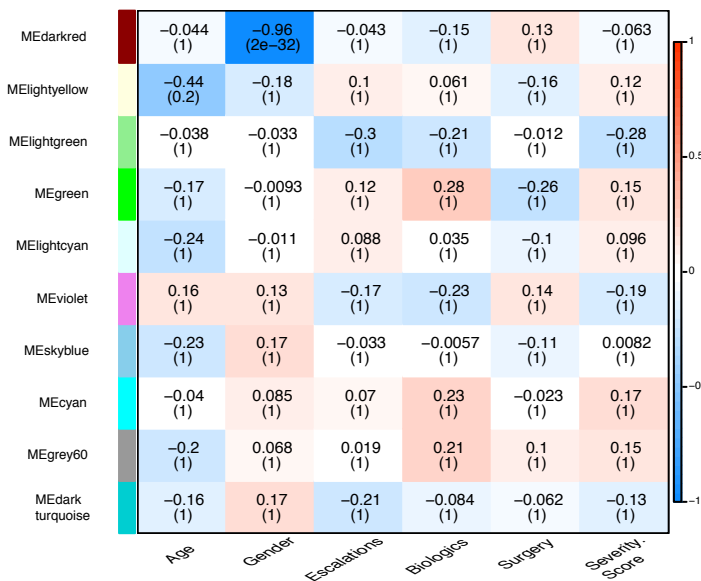
Aiii.

WGCNA blue module heatmap in all adult CD samples (n=35)



Bi.

Top 10 Module-trait relationships for Pediatric CD Samples (n=60)



Bii.

WGCNA gender module Heatmap in All Pediatric CD Samples (n=67)

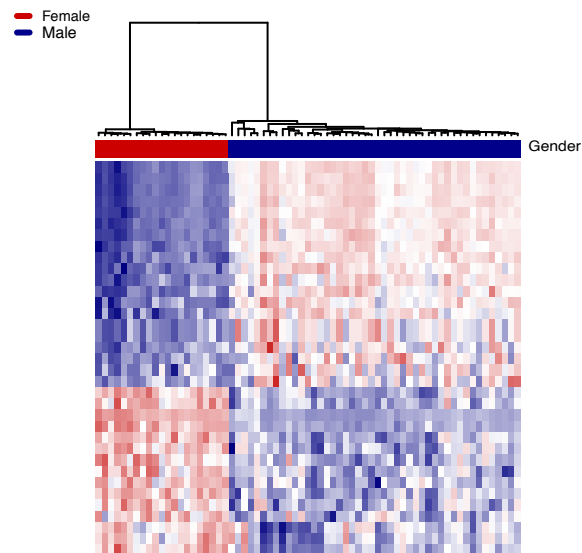


Figure 3: Weighted gene co-expression network analyses (WGCNA) of CD8⁺ T-cell transcriptomes derived from adult **(A)** and childhood **(B)** onset Crohn's Disease. **Ai)** Module-trait relationship heatmap displaying gene modules and their correlation with age, gender and outcome class (severe versus 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 1A (grey) and top correlated modules (Blue and Turquoise). **Aiii)** Hierarchical clustering and heatmap of genes forming the Blue module. **Bi)** 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 gender (Darkred). Samples of patients commenced on treatment with biologics at diagnosis were excluded from these analyses.

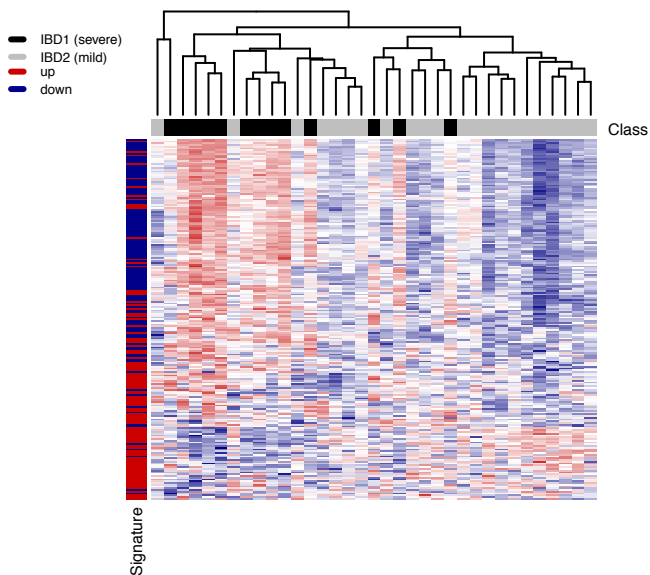
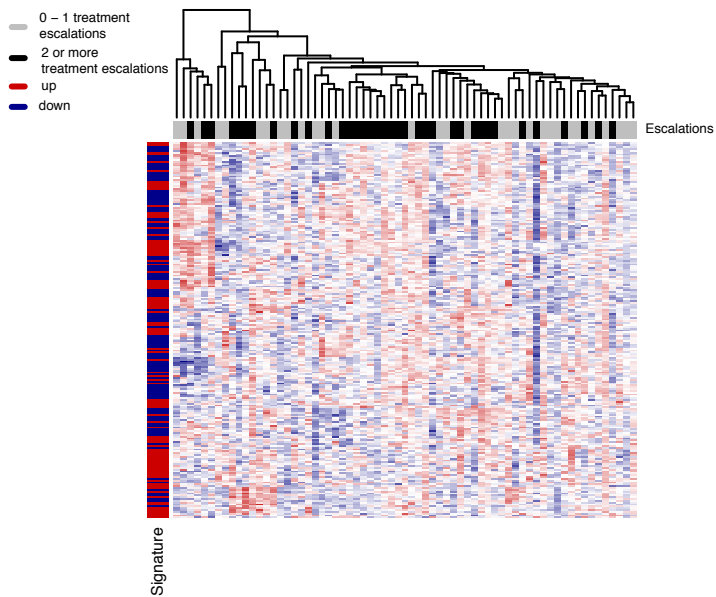
A.**Exhaustion signature in Adult CD Samples (n=35)****B.****Exhaustion signature in Pediatric CD Samples (n=67)**

Figure 4: Transcriptional variation of genes associated with T-cell exhaustion in CD8⁺ T-cells. Genes forming a transcriptional T-cell exhaustion signature²⁴ were selected from genome-wide transcriptomes and subjected to hierarchical clustering. Heatmap and hierarchical clustering trees are displayed for samples obtained from adults (**A**, n=35) and children (**B**, n=67) diagnosed with Crohn's Disease.

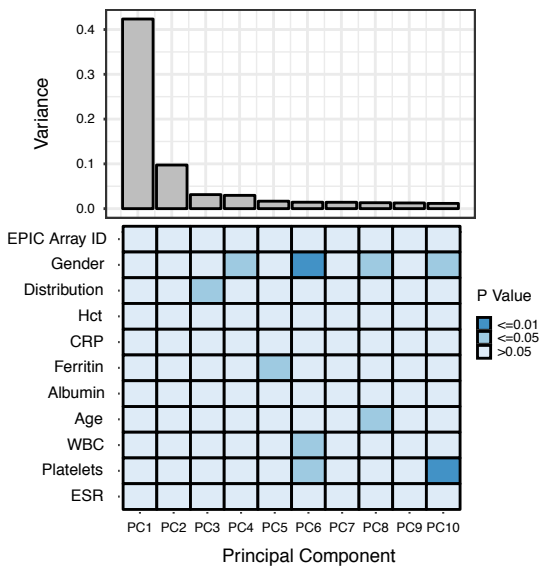
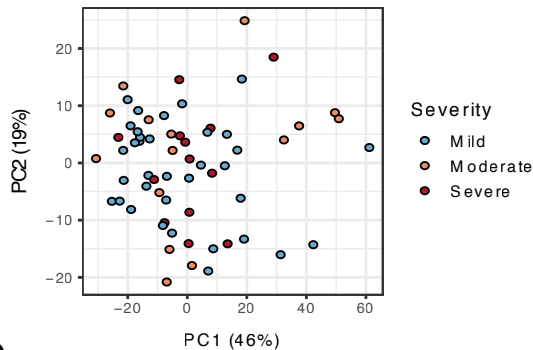
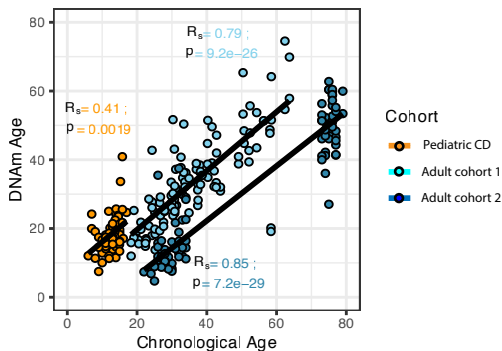
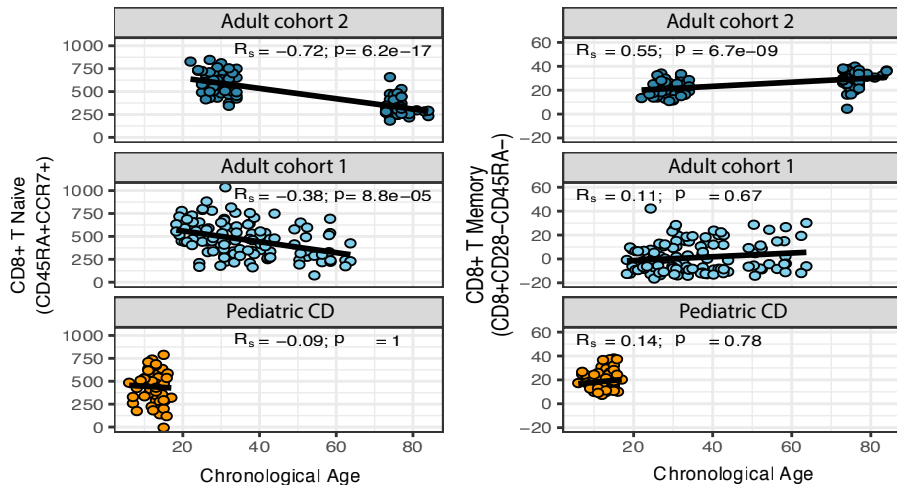
A.**B.****C.****D.**

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 Crohn's Disease (n=66). Degree of variance in each principal component (PC – 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 ANOVA for categorical. **B)** PCA plot of childhood CD patient derived CD8⁺ T-cell DNA methylomes. Samples labelled 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 two publicly available datasets (Adult cohorts 1 and 2). Adult cohort 1 includes samples derived from adult patients diagnosed with IBD and healthy controls¹¹. **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)¹¹, Adult cohort 2 (n=100)¹²