

Journal Pre-proof

Sex differences in blood transcriptional profiles and clinical phenotypes in pediatric patients with eosinophilic esophagitis

Elizabeth A. Erwin, MD, Lisa M. Jaramillo, BS, Bennett Smith, BS, Patrice G. Kruszewski, DO, Basil Kahwash, MD, Mitchell H. Grayson, MD, Asuncion Mejias, MD, PhD, Octavio Ramilo, MD

PII: S2213-2198(21)00774-1

DOI: <https://doi.org/10.1016/j.jaip.2021.06.043>

Reference: JAIP 3713

To appear in: *The Journal of Allergy and Clinical Immunology: In Practice*

Received Date: 20 November 2020

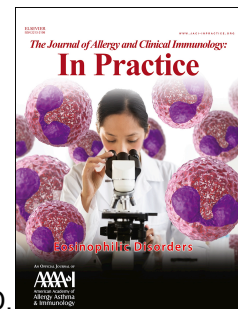
Revised Date: 22 June 2021

Accepted Date: 24 June 2021

Please cite this article as: Erwin EA, Jaramillo LM, Smith B, Kruszewski PG, Kahwash B, Grayson MH, Mejias A, Ramilo O, Sex differences in blood transcriptional profiles and clinical phenotypes in pediatric patients with eosinophilic esophagitis, *The Journal of Allergy and Clinical Immunology: In Practice* (2021), doi: <https://doi.org/10.1016/j.jaip.2021.06.043>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology



Sex differences in blood transcriptional profiles and clinical phenotypes in pediatric patients with eosinophilic esophagitis

Elizabeth A. Erwin MD^{1*}, Lisa M. Jaramillo BS², Bennett Smith BS², Patrice G. Kruszewski DO³, Basil Kahwash MD¹, Mitchell H. Grayson MD^{1,4}, Asuncion Mejias MD, PhD^{2,5}, Octavio Ramilo MD^{2,5}

*Contact author

¹ Division of Allergy and Immunology, Department of Pediatrics, Nationwide Children's Hospital and the Ohio State University College of Medicine, 700 Children's Dr, Columbus, Ohio 43205, Tel: 614-722-6884, Fax: 614-722-4423, Elizabeth.Erwin@nationwidechildrens.org

² Center for Vaccines and Immunity, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, Ohio

³ Division of Gastroenterology, Hepatology and Nutrition, Children's Healthcare of Atlanta, Department of Pediatrics, Emory University, Atlanta, Georgia

⁴ Center for Clinical and Translational Research, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, Ohio

⁵ Division of Infectious Diseases, Department of Pediatrics, Nationwide Children's Hospital and the Ohio State University College of Medicine, Columbus, Ohio

The authors have no conflicts to disclose.

Funding: National Institutes of Health grants K23-AI-059317 and UL1 TR001070 and Nationwide Children's Hospital intramural funds to AM and OR

Word count: 3774

1 Table/4 Figures

The authors have no conflicts of interest in the subject matter or materials discussed in this manuscript.

ABSTRACT

Background: Eosinophilic esophagitis (EoE) is an increasingly recognized, chronic inflammatory disease. Recent reports suggest clinical differences between males and females.

Objective: To define the relevant molecular pathways that could be related to clinical phenotypes in children with EoE.

Methods: We performed blood RNA expression analysis in newly diagnosed pediatric EoE males and females and matched, healthy controls, and applied bioinformatics tools to define EoE host immune biosignatures. Questionnaires and medical records were used to characterize symptoms, esophagogastroduodenoscopy results, and treatment response.

Results: Forty-one subjects (aged 2-17 years) were enrolled, and the cohort consisted of 27 males and 14 females. Patients were randomly divided into a discovery cohort (21 EoE patients and 12 controls) that identified 544 significant differentially expressed transcripts ($p \leq 0.01$, 1.25-fold change). Those 544 transcripts correctly classified most EoE patients in the validation cohort ($n=20$) from healthy controls. Global transcriptional perturbation relative to healthy controls, molecular distance to health (MDTH) scores were greater in EoE patients than controls ($p=0.003$). When we analyzed subjects based on age and sex, males ≥ 13 years were more likely to have food impactions ($p=0.033$) and had higher endoscopic severity scores ($p=0.036$).

Separate group comparisons according to gender identified 294 differentially expressed transcripts in males and 643 transcripts in female EoE patients. Of those, 37 genes were shared and similarly expressed irrespective of sex.

Conclusion: Whole blood transcriptional analysis represents a promising noninvasive tool to assess activity of the immune/inflammatory response in children with EoE. Male and female EoE patients showed robust differences in gene expression suggesting distinct pathogenic endotypes.

What is already known about this topic? The clinical presentation of eosinophilic esophagitis (EoE) is different in males and females.

What does this article add to our knowledge? EoE patients have a distinct gene expression profile from controls, and the transcriptome is not the same in males and females.

How does this study impact current management guidelines? This study provides a framework that can be used to classify EoE patients at diagnosis in order to measure severity and plan targeted therapy.

Key words: eosinophilic esophagitis, pediatrics, transcriptional analysis, phenotypes, sex differences

Abbreviations used

43 EoE: Eosinophilic esophagitis
44 PPI: Proton pump inhibitor
45 eos/hpf: eosinophils/high power field
46 EGD: esophagogastroduodenoscopy
47 TSS: Total Symptoms Score
48 EREFS: EoE Endoscopic Reference Score
49 QC: quality control
50 RIN: RNA integrity number
51 LIMMA: linear models for microarray data
52 FDR: false discovery rate
53 IPA: Ingenuity Pathway Analysis
54 MDTH: Molecular Distance to Health
55 IQR: interquartile range
56 ROC: receiver operating characteristic
57 Br: Branch (for hierarchical clustering)
58 GC-C: guanylate cyclase
59 RSV: respiratory syncytial virus
60 IVIG: intravenous immunoglobulin
61 CBC: complete blood count
62

INTRODUCTION

On a worldwide basis, incidence of eosinophilic esophagitis (EoE) has risen rapidly over the last 15-20 years, and the increase appears to be continuing.^{1,2} The definition of EoE as an antigen driven disorder of eosinophil predominant inflammation has been upheld in consensus documents.^{3,4} However, classifying patients into a single disease entity based on this definition has met with challenges of clinical disease variation such as proton pump inhibitor (PPI) responsiveness, “sparse” eosinophil counts, and even recently the proposal of an “EoE-like disease.”⁵⁻⁸

Some experts have highlighted the possibility of different EoE phenotypes and even proposed patient groupings, but those definitions have not been extended beyond observed clinical differences and have mostly related to the presence or absence of stenosis and food impactions or food allergies.⁹⁻¹³ Several recent studies suggest sex-specific differences in the clinical presentation of EoE with males having a higher risk of food impactions and strictures and females having more abdominal pain, regurgitation, and gastroesophageal reflux.¹⁴⁻¹⁷ In addition, another study showed differences in gene expression from esophageal samples in male versus female adults with EoE.¹⁸

Gene expression analysis has been used to uncover disease mechanisms and for biomarker identification. Transcriptional analysis of esophageal tissue has been helpful in EoE through i) identification of gene expression signatures that can distinguish EoE from other chronic inflammatory diseases of the esophagus and ii) by providing insights into disease mechanisms.¹⁹⁻²¹ Nevertheless, most gene expression studies in EoE have been focused on examination of esophageal tissue samples which might not represent the whole disease given that

recent genetic and epidemiologic evidence suggests that EoE has systemic features.²² Whole blood transcriptional profiling represents a useful marker of disease processes and has been used to understand complex host pathogen interactions in children with different infections, and also to study mechanisms associated with autoimmune diseases.²³⁻²⁵ To date, information about blood gene expression analysis in EoE has been limited.²⁶

We previously reported clinical and endoscopic outcomes for 44 newly diagnosed pediatric EoE patients who were treated with swallowed fluticasone or single food elimination of cow's milk.²⁷ From the majority of those patients (93%; n=41), blood samples were obtained at diagnosis for analysis of gene expression profiles. The objective of the study we now report, was to evaluate the clinical characteristics of the patients and to determine if gene expression signatures at baseline are associated with different clinical phenotypes of EoE.

METHODS

Study Design

Pediatric patients who presented with a new diagnosis of esophageal eosinophilia (>15 eosinophils/hpf [eos/hpf]) were recruited for a prospective study to monitor treatment response. Proton pump inhibitor (PPI) responsiveness was not determined prior to EoE diagnosis. Patients with increased eosinophils in other parts of the gastrointestinal tract and patients with other chronic gastrointestinal diagnoses were excluded. Patients were treated with either PPI and swallowed fluticasone or PPI and single food elimination of cow's milk. After 6-8 weeks of treatment, a subgroup of patients had a repeat esophagogastroduodenoscopy (EGD) and esophageal biopsies to evaluate histologic response. This study was approved by the Institutional Review Board at Nationwide Children's Hospital.

At baseline, demographic and clinical data were collected by using a combination of clinical questionnaires designed for the study and medical record review. Symptoms were evaluated using the Total Symptoms Score (TSS) obtained from the two symptom scales included in the standardized PedsQL EoE ModuleTM.²⁸ Visual changes in the esophagus were evaluated by a single pediatric gastroenterologist blinded to age, sex, and treatment. The results were recorded both categorically as yes or no to describe the presence of longitudinal furrowing, white plaques or rings and by using an EoE Endoscopic Reference (EREFS) Score. The EREFS score was calculated based on the standardized grading system used to quantify the presence and severity of edema, rings, exudates, furrows, and strictures in the esophagus.²⁹ Esophageal biopsy results obtained at follow up EGD during treatment were used to characterize patients as histologic responders (<15 eos/hpf) or nonresponders (≥ 15 eos/hpf).

Analysis of clinical data

Because we observed clinical differences in boys and girls in a separate EoE cohort¹⁷, patients were first compared based on sex. We anticipated age-based differences in symptoms and the median age was 13 years, so we stratified patients into four groups: males <13 and $\geq$ 13 years and females <13 and $\geq$ 13 years, a priori for the second part of the analysis. Fisher's exact test, or the chi square test of independence was used to compare categorical data. Continuous variables were assessed for normality using the Shapiro-Wilk test and Q-Q plot. Based on the results, the Mann-Whitney test or Kruskal Wallis test was used for comparisons. These analyses were performed using SPSS software version 24 (IBM, Armonk, NY).

Sample collection

Patients' blood samples were obtained at the time of study recruitment (baseline). As patients were newly diagnosed, they all had active EoE. We also enrolled 12 healthy controls from the outpatient clinical research unit who had no ongoing health conditions, no medications, and IgE <0.35 IU/ml to a panel of common food and environmental allergens. These children served as controls for all analyses. Whole blood (1 mL) samples for microarray analysis were collected in Tempus Tubes (Applied Biosystems), frozen, processed in batches and hybridized on Illumina Human HT-12v4 beadchips using microarray technology. Before hybridization, RNA underwent quality control (QC) to assess both quality using the Caliper instrument and quantity with the bioanalyzer. After globin reduction, the overall mean RNA integrity number (RIN) was 8.2 and the grRNA yield 3563 ng.³⁰⁻³²

Data import and normalization

Data were batch-corrected, normalized and extracted from BeadStudio with signal intensity values set to 10. First using linear models for microarray data (LIMMA) in R (R Foundation for Statistical Computing, Vienna, Austria), average sample signal intensity was normalized by dividing the result for each probe by the median intensity across each control group for that probe. Next, we selected transcripts present in 10% of samples (PAL10%) with signal precision <0.01 and minimum 2-fold expression up or down (2xUDALP) in EoE patients vs controls. The PAL10_2xUDALO or quality control gene test composed of 15,330 transcripts was used as the starting point for analysis.

Transcriptome data analysis

For supervised analysis, we used R software to perform LIMMA t-statistics (false discovery rate [FDR] <0.01) for comparisons across predefined study groups and a >1.25 -fold change in expression compared to the control group. We chose to use a higher p value for statistical comparisons combined with a 1.25-fold change in order to generate a complete list of transcripts for the blood transcriptional signature of the discovery group of EoE subjects. Unsupervised clustering was applied to the independent validation set. Hierarchical clustering was applied to the list of transcripts generated, by which each sample was assigned to its own cluster iteratively at each stage joining the most similar cluster, distance between the clusters was computed by the Lance-Williams dissimilarity formula according to complete linkage method to find similar clusters. For hierarchical clustering of genes we used the Pearson's correlation similarity measure that organizes transcripts according to their trend of expression across all samples. Participant samples were then clustered based on the similarity of their expression profiles using Euclidean correlation/distance similarity measure. Ingenuity Pathways Analysis (IPA) (Ingenuity Systems) was used to identify the top canonical pathways and lists of classifier

162 genes. For a numerical representation of degree of global RNA perturbation of a given sample
163 compared with healthy controls, Molecular Distance to Health (MDTH) scores were calculated
164 for each sample and the medians were compared between patient groups.^{23,30-33}

RESULTS

Clinical Characteristics

The demographic characteristics of the patients in the entire study have been published previously.²⁷ The median age of the 41 patients who were included in the current analyses was 13 years (range 2-17). Similar to other studies in EoE patients, the majority of our patients were white (95 %); however, the proportion of females was slightly higher than other well characterized pediatric cohorts at 14/41 (34 %).^{12,34-37}

First, we compared patient demographic and clinical characteristics based on sex. The median and distribution for age was remarkably similar in males (13 years [range 2-17]) and females (13 years [range 2-17]) ($p=0.92$) (Table I). The proportion of males with food impactions ($n=11$, 41%) was higher than females ($n=2$, 14%), but this difference was not significant ($p=0.16$). Symptoms of gastroesophageal reflux were more common in females at 43% versus 11% in males ($p=0.042$). Within each sex group, patients were further divided based on the median age: males <13 and ≥ 13 years and females <13 and ≥ 13 years. Compared with all other groups, food impactions were more common in males ≥ 13 years; 60% had impactions ($p=0.033$). The baseline median EoE EREFS Score was also higher in males (median of 3 [0-6]) compared to female patients (median of 2 [0-3]; $p=0.036$). The prevalence of associated allergic diagnoses of food allergy, asthma, and rhinitis was similar in male and female patients.

Eighty percent of patients (22 males; 11 females) had a repeat esophageal biopsy after 6-8 weeks of treatment with swallowed fluticasone or single food group elimination of cow's milk. The distribution of males and females in each treatment group was similar as 10 males (45%) and 5 females (45%) were treated with elimination of cow's milk and 12 males (55%) and 6

females (55%) were treated with fluticasone. Overall, 14/22 males (64%) and 10/11 females (90%) had <15 eos/hpf in the esophageal biopsy after treatment. The single female who was considered a nonresponder had a peak count of 15 eos/hpf after treatment. When treatment response was analyzed by age and sex, males ≥ 13 years had a trend towards lower rates of histologic remission (45%, $p=0.073$) (Fig 1). The difference in treatment response seemed mostly related to the increase in food impactions in males ≥ 13 years since 25% of males with food impactions (2/8) achieved histologic remission compared to 100% without food impactions (3/3) ($p=0.061$). The response rate to treatment with swallowed fluticasone and cow's milk elimination diet was similar within the group of males ≥ 13 years (50% and 43% respectively).

Biosignatures for EoE

To define the systemic immune response to EoE, we first divided the cohort of EoE patients randomly into separate discovery (training set $n=21$) and validation (test set $n=20$) cohorts and compared them with 12 age-matched, healthy controls that included 5 females and 7 males aged 2.9-14.3 years (median 6) (Table 2).³⁸ Group comparisons using LIMMA (FDR <0.01 , 1.25-fold change) between EoE patients and healthy controls in the training set identified 544 significant differentially expressed transcripts (Fig 2, A). Of those 544 differentially expressed RNA transcripts, 39.7% were underexpressed and 60.29% overexpressed. We applied the same list of differentially expressed genes ($n=544$) to the independent test set of EoE patients and healthy controls. Unsupervised hierarchical clustering correctly grouped most patients with EoE (Fig 2, B). To quantify group differences based on global gene expression, we calculated the MDTH genomic score, that ranged from 1.72-574 (Fig 2, C and D). Molecular distance to health (MDTH) scores (median [interquartile range [IQR]]) were significantly higher in EoE patients (73.3 [19.1-209]) compared with controls (9.6 [2.58-

25.3]) ($p=0.004$) (Fig 2, E). Using the MDTH score, we generated a receiver operating characteristic (ROC) curve for the validation cohort, and the area under the curve was 0.808.

Among the top ten overexpressed genes in EoE, we identified PRSS33, which encodes a serine protease expressed on eosinophils that colocalizes with granule proteins with staining (Table E1, A). Stimulation of fibroblasts by PRSS33 has been shown to increase expression of collagen mRNA.³⁹ The list of overexpressed genes also includes AHNAK, which encodes a structural scaffold protein that may be involved in neuronal cell differentiation and has been associated with Hirschsprung Disease. The top underexpressed genes included KIAA1211 for which the encoded protein is involved in epithelial barrier integrity through its role in maintaining the actin cytoskeleton, and PNKD, which is thought to play a role in myofibrillogenesis (Table E1, B). Ingenuity Pathway Analysis (IPA) showed that the most overexpressed genes were part of the PDGF signaling pathway (Fig E1), which plays a role in mesenchymal stem cell commitment promoting collagen deposition.

Gene expression profiles in EoE patients according to sex

We followed the same strategy to define the transcriptional profiles of EoE in males and females separately. Pairwise comparisons between female EoE patients and age/gender matched healthy controls identified 643 differentially expressed transcripts (Fig 3, A). Hierarchical clustering grouped 2 samples from EoE females with the controls (Branch [Br] 2). In contrast, among male EoE vs matched controls, statistical comparison identified 294 differentially expressed transcripts which grouped all 27 male EoE patients (Br 2) distinctly from healthy controls (Br 1) by hierarchical clustering (Fig 3, B).

Next, we analyzed the overlap between the male and female EoE signatures and found that only 37 genes were over- or underexpressed in a similar manner between the groups at baseline (Fig 3, C). To assess the biological significance of these findings, we used IPA. The genes that showed a similar pattern of dysregulation for males and females are shown in Tables E2, A and B. Among the top genes similarly overexpressed, we identified MS4A7 which encodes a receptor associated with monocytes and ANXA2P1 (Annexin A2 pseudogene) which encodes a calcium dependent phospholipid binding protein that has been identified as a specific antigen in IgG4 related disease. There were important differences between the two groups with females showing differential expression of genes associated with the inflammatory response such as MS4A1 (which plays a role in development of B cells into plasma cells) and infectious diseases such as GUCY2C (a receptor for E. coli enterotoxin). Males showed differential expression of genes associated with immunological pathways (HSP90AB1 which antagonizes inhibition of TGF- β signaling) and cardiovascular disease (AHNAK which is involved in cardiac calcium channel function). The lists of top differentially expressed genes in females and males with EoE are presented in Tables E3, A and B and Tables E4, A and B respectively.

Transcriptional profiles indicative of food impaction in EoE males

Next we analyzed the transcriptional profiles of males with EoE according to food impaction. This analysis could not be performed in females as only two of the female study patients experienced this complication. Group comparisons followed by unsupervised hierarchical clustering identified 359 differentially expressed transcripts in males with impactions (n=11) vs matched controls (63.78% overexpressed and 36.2% underexpressed) and grouped males with impaction distinctly from controls (Fig 4, A). In contrast, there were 217 differentially expressed transcripts among males without impactions (n=16) vs controls (40%

254 overexpressed and 60% underexpressed. By hierarchical clustering, 3 healthy controls grouped
255 with the EoE patients without impactions (Fig 4, B). There were 58 genes differentially
256 expressed in a similar manner in males with and without impactions (Fig 4, C). IPA was used to
257 generate separate lists of genes that are similarly differentially expressed (Tables E5, A and B)
258 and those that can be used to discriminate patients with and without impactions from healthy
259 controls (Tables E6, A and B and E7, A and B respectively). Among patients with impactions we
260 identified overexpression of AEN and RPAP3 genes which are associated with apoptosis. For
261 patients without impactions, we identified overexpression of several genes associated with the
262 innate immune system (VRK3 and UBE2G2, which is expressed in muscle).

DISCUSSION

EoE is an increasingly recognized chronic, inflammatory disease with a “heterogeneous” clinical presentation that includes food impactions, dysphagia, abdominal pain, vomiting, and food aversion or failure to thrive. These differences have seemed partially due to age⁴⁰ and disease duration¹¹ but also may be related to the sex of the patient.¹⁷ Our analysis of whole blood transcriptional profiles in a pediatric EoE cohort identified a reproducible biosignature in EoE patients compared with age-matched healthy controls. Subsequent analyses, however, showed striking differences between female and male gene expression profiles that were found in association with sex-related differences in clinical findings, suggesting at least two different phenotypes of EoE. The distinct transcriptomes identified in females versus males may further imply different immune pathways, and possibly different mechanisms for disease development.

The clinical differences we report here provide evidence for distinct phenotypes and suggest a differential response to treatment. In contrast to female patients who were more likely to have symptoms of gastroesophageal reflux, male patients age ≥ 13 years presented more often with food impactions and had higher EREFS scores. Furthermore, they were less likely to achieve histologic remission with treatment; although, there was one female who had a peak count of 15 eos/hpf during treatment (decrease of 43 eos/hpf from previous peak of 58 eos/hpf), and thus she was considered to be a nonresponder. These sex-based differences are similar to those previously identified in adult cohorts¹⁴⁻¹⁶ as well as in a separate pediatric EoE cohort at our own institution.¹⁷ The blood transcriptome results extend the work of Gonsalves *et al* who showed differential RNA expression in esophageal tissue in female and male adults.¹⁸ We observed differences in systemic gene expression resulting in distinct transcriptional profiles for pediatric female and male EoE patients.

The EoE blood transcriptomes for males and females were remarkably different with only 37 overlapping differentially expressed transcripts. Analyses of the top ten overexpressed genes in females showed associations with other gastrointestinal diseases and B cell responses. We observed increased expression of GUCY2C, which encodes guanylate cyclase (GC-C), an enterocyte receptor for bacterial endotoxins that is associated with dysmotility. Dysregulation of GUCY2C has been observed in Crohn's disease.⁴² In 2012, an activating mutation in GC-C was reported to be the genetic cause of familial diarrhea syndrome.⁴³ The gene MS4A1, which is involved in development and differentiation of B cells, was also overexpressed. Underexpressed genes in female EoE patients included, among others, IGKV1D-33 an immunoglobulin gene transcript that is reduced in certain subgroups of lupus patients. In contrast, males showed increased expression of HSP90AB1 which encodes a protein that antagonizes inhibition of TGF- β signaling. Additionally in males, a top overexpressed gene was PSMC1, which encodes a subunit of receptors that interact with viruses such as Hepatitis B and adenovirus. The relationship between viral infections and EoE is poorly understood, but there are case reports of herpes simplex virus and other infections preceding the diagnosis of EoE in immunocompetent children.⁴⁴⁻⁴⁶ It has been hypothesized that preceding viral infection may affect mucosal barrier function resulting in impaired immune tolerance and predisposing patients to EoE.

Molecular distance to health (MDTH) is a transcriptional score that converts global transcriptional perturbation of each individual patient sample into a numeric score relative to the median baseline in healthy controls.^{23,24} Studies applying this molecular tool to other diseases have shown that there is an association between MDTH scores and measures of clinical disease severity. In respiratory syncytial virus (RSV) infection, higher MDTH scores correlated with increased clinical disease severity scores, longer hospitalization and longer duration of

supplemental oxygen.²³ In Kawasaki's disease, MDTH was predictive of response to intravenous immunoglobulin (IVIG) such that higher values before treatment identified patients who did not respond to IVIG.⁴⁷ At the very least in our study, differences in MDTH scores between EoE patients and controls confirm distinct transcriptional profiles. Currently there is no generally accepted measure of severity in EoE. The use of symptom scores has been challenging due to variability of symptoms among pediatric patients and behavioral adaptations in older children and adults.^{48,49} Additionally, use of symptom scores has been further complicated by poor correlation between symptoms and inflammation with "low grade" eosinophilia, sometimes associated with food impactions and other times resulting in few symptoms and diagnostic delay.^{50,51} Esophageal compliance as measured by functional luminal imaging probe (Endoflip) may relate to fibrosis or food impactions but this does not include patients with other symptoms.⁵² The MDTH metric provides an objective measure that could be evaluated in larger cohorts as a potential marker of severity in EoE.

Previous studies of gene expression in EoE have been performed using esophageal biopsy samples.¹⁹⁻²¹ Those studies showed increased expression of mast cell and eosinophil-related genes as well as differential expression of genes involved in epithelial barrier function. Peripheral blood analysis extends these results to show that systemic gene expression could elucidate pathways of upstream inflammation and explain associations with other diseases resulting from immune dysregulation. Among EoE patients as a group, we observed differences in genes related to eosinophils (PRSS33), as well as in epithelial cell integrity (KIAA1211), and these findings are consistent with studies of gene expression in the esophagus. We also observed underexpression of PNKD a gene thought to play a role myofibrillogenesis and suspected to aggravate cardiac hypertrophy. This is noteworthy given the reported clinical association

between EoE and cardiomyopathy.⁵³ This initial study does not show the pathogenic mechanism of either EoE phenotype. However, it defines a unique systemic expression profile that could be the first step toward establishing a mechanistic pathway.

We report the first evidence of differential gene expression profiles in the peripheral blood in clinically different groups of EoE patients. The addition of transcriptional information associated with phenotypic differences suggests progress toward defining potential endotypes of EoE patients that may have a specific relationship with response to treatment. Wolf *et al* examined predictors of response to steroid therapy in adults with EoE and found that nonresponders were more likely to have baseline EGD that showed narrowing and required dilation.⁵⁴ Similarly in our study, older male patients with impactions did not respond to treatment as well as those without impactions.

This is a pilot study limited by its small sample size of patients and controls. With only two female patients with food impaction, we were unable to analyze gene expression in this group separately from females without food impactions. Additionally, whole blood gene expression does not separate cell specific components. It is possible the EoE blood transcriptional signature may represent altered cell composition or differences in gene expression in specific cell populations. However, the ease of sample collection makes it clinically applicable and further understanding of noninvasive biomarkers is a necessary part of determining relevant molecular pathways among EoE patients. Since we did not test our EoE patients for environmental allergies, it is possible that our signature could reflect atopy. However, it is striking that in spite of distinct differences from controls in peripheral blood gene expression, we did not observe an increase in genes associated with markers of allergy. In the present study,

despite certain limitations, we identified and validated blood biosignatures for EoE in independent patient cohorts.

In conclusion, transcriptome analysis of peripheral blood samples provides a noninvasive and attractive tool to assess active pathways of immune response and inflammation in EoE patients. Our initial study shows distinct transcriptomes in male and female EoE patients and suggests a potential role for transcriptome analysis for patient classification into distinct groups based on disease severity at the time of initial diagnosis.

References

1. van Rhijn BD, Verheij J, Smout AJ, Bredenoord AJ. Rapidly increasing incidence of eosinophilic esophagitis in a large cohort. *Neurogastroenterol Motil* 2013;25:47-52e5.
2. Giriens B, Yan P, Safroneeva E, Zwahlen M, Reinhard A, Nydegger A, et al. Escalating incidence of eosinophilic esophagitis in Canton of Vaud, Switzerland, 1993-2013: a population-based study. *Allergy* 2015;70:1633-9.
3. Liacouras CA, Furuta GT, Hirano I, Atkins D, Attwood SE, Bonis PA, et al. Eosinophilic esophagitis: updated consensus recommendations for children and adults. *J Allergy Clin Immunol* 2011;12:3-20.
4. Dellon ES, Gonsalves N, Hirano I, Furuta GT, Liacouras CA, Katzka DA. ACG Clinical Guideline: Evidence based approach to the diagnosis and management of esophageal eosinophilia and eosinophilic esophagitis (EoE). *Am J Gastroenterol* 2013;108:679-92.
5. Dellon ES, Liacouras CA, Molina-Infante J, Furuta GT, Spergel M, Zevit N, et al. Updated international consensus diagnostic criteria for eosinophilic esophagitis: Proceedings of the AGREE conference. *Gastroenterology* 2018;155:1022-33.
6. Sodikoff J, Hirano I. Proton pump inhibitor-responsive esophageal eosinophilia does not preclude food-responsive eosinophilic esophagitis. *J Allergy Clin Immunol* 2016;137:631-3.
7. Peterson KA, Cobell WJ, Clayton FC, Krishnamurthy C, Ying J, Pease LF, et al. Extracellular eosinophil granule protein deposition in ringed esophagus with sparse eosinophils. *Dig Dis Sci* 2015;60:2646-53.
8. Straumann A, Blanchard C, Radonjic-Hoesli S, Bussmann C, Hruz P, Safroneeva E, et al. A new eosinophilic esophagitis (EoE)-like disease without tissue eosinophilia found in EoE families. *Allergy* 2016;71:889-900.

9. Atkins D, Furuta GT, Liacouras CA, Spergel JM. Eosinophilic esophagitis phenotypes: Ready for prime time? *Pediatr Allergy Immunol* 2017;28:312-19.
10. Lin Z, Kahrilas PJ, Xiao Y, Nicodeme F, Gonsalves N, Hirano I, et al. Functional luminal imaging probe topography: an improved method for characterizing esophageal distensibility in eosinophilic esophagitis. *Therap Adv Gastroenterol* 2013;2:97-107.
11. Dellon ES, Kim HP, Sperry SL, Rybnicek DA, Woosley JT, Shaheen NJ. A phenotypic analysis shows that eosinophilic esophagitis is a progressive fibrostenotic disease. *Gastrointest Endosc* 2014;79:577-85.
12. Pelz BJ, Wechsler JB, Amsden K, Johnson K, Singh AM, Wershil BK, et al. IgE-associated food allergy alters the presentation of paediatric eosinophilic esophagitis. *Clin Exp Allergy* 2016;46:1431-40.
13. Chehade M, Jones SM, Pesek RD, Burks AW, Vickery BP, Wood RA, et al. Phenotypic characterization of eosinophilic esophagitis in a large multicenter patient population from the Consortium for Food Allergy Research. *J Allergy Clin Immunol Pract* 2018;6:1534-44.
14. Sperry SL, Sperry BA, Woosley JT, Shaheen NJ, Dellon ES. Influence of race and gender on the presentation of eosinophilic esophagitis. *Am J Gastroenterol* 2012;107:215-21.
15. Lynch KL, Dhalla S, Chedid V, Ravich WJ, Stein EM, Montgomery EA, et al. Gender is a determinative factor in the initial clinical presentation of eosinophilic esophagitis. *Dis Esophagus* 2016;29:174-8.
16. Moawad FJ, Dellon ES, Achem SR, Ljuldjuraj T, Green DJ, Maydonovitch CL, et al. Effects of race and sex on features of eosinophilic esophagitis. *Clin Gastroenterol Hepatol* 2016;14:23-30.

17. Erwin EA, Navalpakam A, Singla R, Bolender J, Workman LJ, Platts-Mills TA. Distinct clinical characteristics of boys and girls with eosinophilic esophagitis. *J Allergy Clin Immunol Pract* 2020;8:1452-5.
18. Gonsalves N, Berdnikovs S, Schroeder H, Zelewski A, Bryce PJ. Gender-specific differences in the molecular signatures of adult eosinophilic oesophagitis. *Clin Exp Allergy* 2017;47:969-71.
19. Wen T, Stucke EM, Grotjan TM, Kemme KA, Abonia JP, Putnam PE, et al. Molecular diagnosis of eosinophilic esophagitis by gene expression profiling. *Gastroenterology* 2013;145:1289-99.
20. Blanchard C, Mingler MK, Vicario M, Abonia JP, Wu YY, Lu TX, et al. IL-13 involvement in eosinophilic esophagitis: transcriptome analysis and reversibility with glucocorticoids. *J Allergy Clin Immunol* 2007;120:1292-300.
21. Azouz NP, Ynga-Durand MA, Caldwell JM, Jain A, Rochman M, Fischesser DM, et al. The antiprotease SPINK7 serves as an inhibitory checkpoint for esophageal epithelial inflammatory responses. *Sci Transl Med* 2018;10:eaap9736.
22. Abonia JP, Spergel JM, Cianferoni A. pro/con review-eosinophilic esophagitis: A primary disease of the esophageal mucosa. *J Allergy Clin Immunol Pract* 2017;5:951-5.
23. Mejias A, Dimo B, Suarez NM, Garcia C, Suarez-Arrabal C, Jartti T, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. *PLOS Med* 2013;10:e1001549.
24. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* 2007;109:2066-77.

25. Pascual V, Allantaz F, Arce E, Punaro M, Banchereau J. Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. *J Exp Med* 2005;201:1479-86.
26. Wen T, Aronow BJ, Rochman Y, Rochman M, Kc K, Dexheimer PJ, et al. Single-cell RNA sequencing identifies inflammatory tissue T cells in eosinophilic esophagitis. *J Clin Invest* 2019;130:2014-28.
27. Kruszewski PG, Russo JM, Franciosi JP, Varni JW, Platts-Mills TA, Erwin EA. A prospective, comparative effectiveness trial of cow's milk elimination and swallowed fluticasone for pediatric eosinophilic esophagitis. *Dis Esophagus* 2016;29:377-84.
28. Franciosi JP, Hommel KA, Bendo CB, King EC, Collins MH, Eby MD, et al. PedsQL Eosinophilic esophagitis module: Feasibility, reliability, and validity. *J Pediatr Gastroenterol Nutr* 2013;57:57-66.
29. Hirano I, Moy N, Heckman MG, Thomas CS, Gonsalves N, Achem SR. Endoscopic assessment of the oesophageal features of eosinophilic oesophagitis: Validation of a novel classification and grading system. *Gut* 2013;62:489-95.
30. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466:973-7.
31. Banchereau R, Jordan-Billegas A, Ardura M, Mejias A, Baldwin N, Xu H, et al. Host immune transcriptional profiles reflect the variability in clinical disease manifestations in patients with *Staphylococcus aureus* infections. *PLoS ONE* 2012;7:e34390.
32. Wittkowski KM, Song T. Nonparametric methods for molecular biology. *Methods Mol Biol* 2010;620:105-53.

33. Wittkowski KM, Lee E, Nussbaum R, Chamian FN, Krueger JG. Combining several ordinal measures in clinical studies. *Stat Med* 2001;23:1579-92.
34. Assa'ad AH, Putnam PE, Collins MH, Akers RM, Jameson SC, Kirby CL, et al. Pediatric patients with eosinophilic esophagitis: An 8-year follow-up. *J Allergy Clin Immunol* 2007;119:731-8.
35. Spergel JM, Beausoleil JL, Mascarenhas M, Liacouras CA. The use of skin prick tests and patch tests to identify causative foods in eosinophilic esophagitis. *J Allergy Clin Immunol* 2002;109:363-8.
36. Chadha SN, Wang L, Correa H, Moulton D, Hummell DS. Pediatric eosinophilic esophagitis: the Vanderbilt experience. *Ann Allergy Asthma Immunol* 2014;113:445-51.
37. Hiremath G, Byramji D, Pacheco A, Constantine G, Davis C, Shulman R, et al. Eosinophilic esophagitis in children and its relationship with parental allergies: Texas Children's Hospital experience. *Dig Dis Sci* 2016;61:501-6.
38. Mahajan P, Kuppermann N, Mejias A, Suarez N, Chaussabel D, Casper TC, et al. Association of RNA biosignatures with bacterial infections in febrile infants ages 60 days or younger. *JAMA* 2016;316:846-57.
39. Toyama S, Okada N, Matsuda A, Morita H, Saito H, Fujisawa T, et al. Human eosinophils constitutively express a unique serine protease, PRSS33. *Allergol Int* 2017;66:463-71.
40. Spergel JM, Brown-Whitehorn TF, Beausoleil JL, Franciosi J, Shuker M, Verma R, et al. 14 years of eosinophilic esophagitis: clinical features and prognosis. *J Pediatr Gastroenterol Nutr* 2009;48:30-6.

41. Slack MA, Erwin EA, Cho CB, Raveendran R, Phillips G, Ogbogu PU. Food and aeroallergen sensitization in adult eosinophilic esophagitis. *Ann Allergy Asthma Immunol* 2013;111:304-5.
42. Brenna O, Bruland T, Furnes MW, van Beelen Granlund A, Drozdov I, Emgard J, et al. The guanylate cyclase-C signaling pathway is down-regulated in inflammatory bowel disease. *Scand J Gastroenterol* 2015;50:1241-52.
43. Fiskerstrand T, Arshad N, Haukanes BI, Tronstad RR, Do-cong Pham K, Johansson S, et al. Familial diarrhea syndrome caused by an activating GUCY2C mutation. *N Engl J Med* 2012;366:1586-95.
44. Squires KA, Cameron DJ, Oliver M, da Fonseca Junqueira JC. Herpes simplex and eosinophilic oesophagitis: the chicken or the egg? *J Pediatr Gastroenterol Nutr* 2009;29:246-50.
45. Zaja Franulovic O, Lesar T, Basic N, Tesovic G. Herpes simplex primo-infection in an immunocompetent host with eosinophilic esophagitis. *Pediatr Int Official J Jpn Pediatr Soc* 2013;55:e38-41.
46. Srivastava MD. *M. pneumoniae* is a potential trigger for eosinophilic esophagitis. *J Allergy Clin Immunol* 131:AB177.
47. Jaggi P, Mejias A, Xu Z, Yin H, Moore-Clingenpeel M, Smith B, et al. Whole blood transcriptional profiles as a prognostic tool in complete and incomplete Kawasaki Disease. *PLoS One* 2018;13:e0197858.
48. Schoepfer AM, Straumann A, Panczak R, Coslovsky M, Kuehni CE, Maurer E, et al. Development and validation of a symptom-based activity index for adults with eosinophilic esophagitis. *Gastroenterology* 2014;147:1255-66.

49. Safroneeva E, Straumann A, Coslovsky M, Zwahlen M, Kuehni CE, Panczak R, et al. Symptoms have modest accuracy in detecting endoscopic and histologic remission in adults with eosinophilic esophagitis. *Gastroenterology* 2016;150:581-90.
50. Ravi K, Talley NJ, Smyrk TC, Katzka DA, Kryzer L, Romero Y, et al. Low grade esophageal eosinophilic in adults: an unrecognized part of the spectrum of eosinophilic esophagitis? *Dig Dis Sci* 2011;56:1981-6.
51. Schoepfer AM, Safroneeva E, Bussmann C, Kuchen T, Portmann S, Simon HU, et al. Delay in diagnosis of eosinophilic esophagitis increases risk for stricture formation in a time-dependent manner. *Gastroenterology*. 2013;145:1230-6.
52. Hassan M, Aceves S, Dohil R, Gharibans A, Newbury R, Proudfoot J. Esophageal compliance quantifies epithelial remodeling in pediatric patients with eosinophilic esophagitis. *J Pediatr Gastroenterol Nutr* 2019;68:559-65.
53. Davis BP, Epstein T, Kottyan L, Amin P, Martin LJ, Maddox A, et al. Association of eosinophilic esophagitis and hypertrophic cardiomyopathy. *J Allergy Clin Immunol* 2016;137:934-6.
54. Wolf WA, Cotton CC, Green DJ, Hughes JT, Woosley JT, Shaheen NJ, et al. Predictors of response to steroid therapy for eosinophilic esophagitis and treatment of steroid-refractory patients. *Clin Gastroenterol Hepatol* 2015;13:452-8.

Figure Legends

Fig 1. Patients are grouped by age and sex. For each patient, peak esophageal eosinophil counts are shown at baseline (labelled pre) and during (labelled post) treatment. Individual patient results are connected by a solid line. A dotted line indicates the threshold for remission (15 eosinophils/high power field). The proportion achieving remission is shown for each group.

Fig 2. (A) Group comparison was applied to the training set (n=21) to identify significant differentially expressed transcripts. Each column represents an individual patient and each row a single transcript. Expression levels are shown as overexpressed (red) or underexpressed (blue) compared to median expression of healthy controls. (B) Unsupervised hierarchical clustering of the 544 differentially expressed transcripts applied to the separate test set (n=20) grouped most EoE patients (identified by gray column labels) together. (C) Molecular distance to health (MDTH) scores for each subject are represented by bars below the expression profile for that sample in the training (white for controls and gray for EoE patients) and (D) test sets. (E) Bars show median MDTH score is significantly higher for EoE patients (gray) vs controls (white).

Fig 3. (A) Unsupervised hierarchical clustering of the 643 differentially expressed transcripts in female EoE patients and healthy female controls grouped 12 female samples (red column labels) into one group (Br 1). Two female EoE patients (red labels) clustered with healthy controls (white column labels) in Br 2. (B) Hierarchical clustering of the 294 differentially expressed transcripts in male EoE patients grouped all 27 male EoE patients (blue column labels) (Br 2)

distinctly from healthy controls (white column labels) (Br 1). (C) A Venn diagram shows mostly distinct male and female EoE transcriptomes.

Fig 4. (A) Unsupervised hierarchical clustering of the 359 differentially expressed transcripts grouped males with food impactions (red labels; Br 2) separately from male controls (white labels; Br 1). (B) Hierarchical clustering of 217 differentially expressed transcripts in males without impactions (blue labels) grouped EoE patients together in Br 1. Three male controls clustered near 2 of the EoE patients. (C) A Venn diagram shows the overlap between male impaction and no impaction transcriptomes.

Table I. Baseline demographic and clinical characteristics of the study patients.

Characteristic	All patients (n=41)	By Group				p value*
		Males <13 y (n=12)	Males ≥13 y (n=15)	Females <13 y (n=7)	Females ≥ 13 y (n=7)	
Sex, n (%)	27 M (66 %) 14 F (34 %)					
Median age, years (range)	13 (2.0-17)	7.5 (2-12)	15 (13-17)	5 (2-12)	15 (14-17)	p=0.92 [§]
Race, n (%)						
White	39 (95 %)	11 (92 %)	14 (93 %)	7 (100%)	7 (100%)	p=0.54 [‡]
Black or African American	2 (4.9 %)	1 (8.3 %)	1 (6.7 %)	0 (0)	0 (1)	
Symptoms						
Impaction	13 (32 %)	2 (17 %)	9 (60 %)	1 (14 %)	1 (14 %)	p=0.033
Dysphagia	20 (49 %)	7 (58 %)	8 (53 %)	3 (43 %)	2 (29 %)	p=0.60
Abdominal pain	19 (46 %)	5 (42 %)	5 (33 %)	4 (57 %)	5 (71 %)	p=0.35
Vomiting	9 (22 %)	3 (25 %)	1 (6.7 %)	3 (43 %)	2 (29 %)	p=0.22
Gastroesophageal reflux	9 (22 %)	1 (8.3 %)	2 (13 %)	3 (43 %)	3 (43 %)	p=0.15 [†]
Failure to thrive	8 (20 %)	4 (33 %)	2 (13 %)	2 (29 %)	0 (0)	p=0.28
Comorbid Allergic Conditions						
Asthma	18/40 [±] (45 %)	5 (42 %)	7 (43 %)	3 (43 %)	3 (43 %)	p=0.99
Rhinitis	21/31 [±] (68 %)	5 (42 %)	9 (60 %)	2 (29 %)	4 (57 %)	p=0.57
Food allergy	17/41 [±] (41 %)	5 (42 %)	5 (33 %)	3 (43 %)	3 (43 %)	p=0.85
PPI use before diagnosis	11 (26 %)	4 (33 %)	1 (6.7 %)	4 (57 %)	2 (29 %)	p=0.084
Median Total Symptoms Score	55 (23-100)	58 (43-100)	55 (23-95)	53 (36-89)	45 (40-68)	p=0.76 [‡]
Median eos/hpf	45 (15-150)	43 (15-150)	45 (25-100)	55 (20-90)	35 (20-58)	p=0.73 [‡]
Median EREFS	2.5 (0-6)	3 (0-6)	3 (1-5)	2 (0-3)	2 (2)	p=0.036[‡]
Median decrease in eos/hpf	35 (-10-148)	35 (1-148)	23 (-10-99)	57 (27-90)	35 (8-44)	p=0.38 [‡]
Histologic remission	24/33 [±] (73 %)	9/11 (82 %)	5/11 (45 %)	6/6 (100 %)	4/5 (80 %)	p=0.073

Abbreviations: M: male, F: female, PPI: proton pump inhibitor, eos/hpf: eosinophils/high power field

* p value by using the chi square test of independence except as noted.

§ p value for comparison of age in males vs females only was obtained using the Mann Whitney test.

¥ p value for comparison of sex in males vs females only was obtained using Fisher's Exact test.

† There was a significant difference in males and females ($p=0.042$). When age was incorporated into the analysis, there was a significant linear by linear association for gastroesophageal reflux and sex ($p=0.032$).

± Denominator shows the number of subjects for which the information was documented.

‡ p value obtained by using the Kruskal Wallis Test.

Table 2. Demographic characteristics of the healthy controls used as a reference population for microarray analysis.

Characteristic	EoE subjects (n=41)	Healthy controls (n=12)	p value
Sex, n (%)	27 M (66 %)/14 F (34 %)	7 M (58%)/5 F (42%)	p=0.73*
Median age, years (range)	13 (2.0-17)	6.0 (2.3-17.6)	p=0.09 [§]
Race, n (%)			
White	39 (95 %)	6 (50%)	p=0.001
Black or African American	2 (4.9 %)	6 (50%)	

Abbreviations: M: male, F: female

* p value by using Fisher's Exact test.

[§] p value obtained using the Mann Whitney test.

Table E2B. List of genes that were underexpressed in both females and males with EoE.

Gene	Expression	Function
ITGAX	-0.99	Encodes an integrin involved in endothelial adherence for neutrophils and monocytes.
GTF2IRD2P1	-0.56	Encodes a pseudogene. Diseases associated with GTF2IRD2P1 include oral squamous cell carcinoma.
TAGAP	-0.522	May function as a GTPase-activating protein and may play important roles during T-cell activation. Alterations in this gene may be associated with several diseases, including rheumatoid arthritis, celiac disease, and multiple sclerosis.
PPP2R5C	-0.517	Encoded protein implicated in the negative control of cell growth and division.
IFI27L2	-0.502	Plays a role in the apoptotic process and has a pro-apoptotic activity.
VWA1	-0.489	Encoded protein belongs to the von Willebrand factor A (VWFA) domain superfamily of extracellular matrix proteins and appears to play a role in cartilage structure and function.

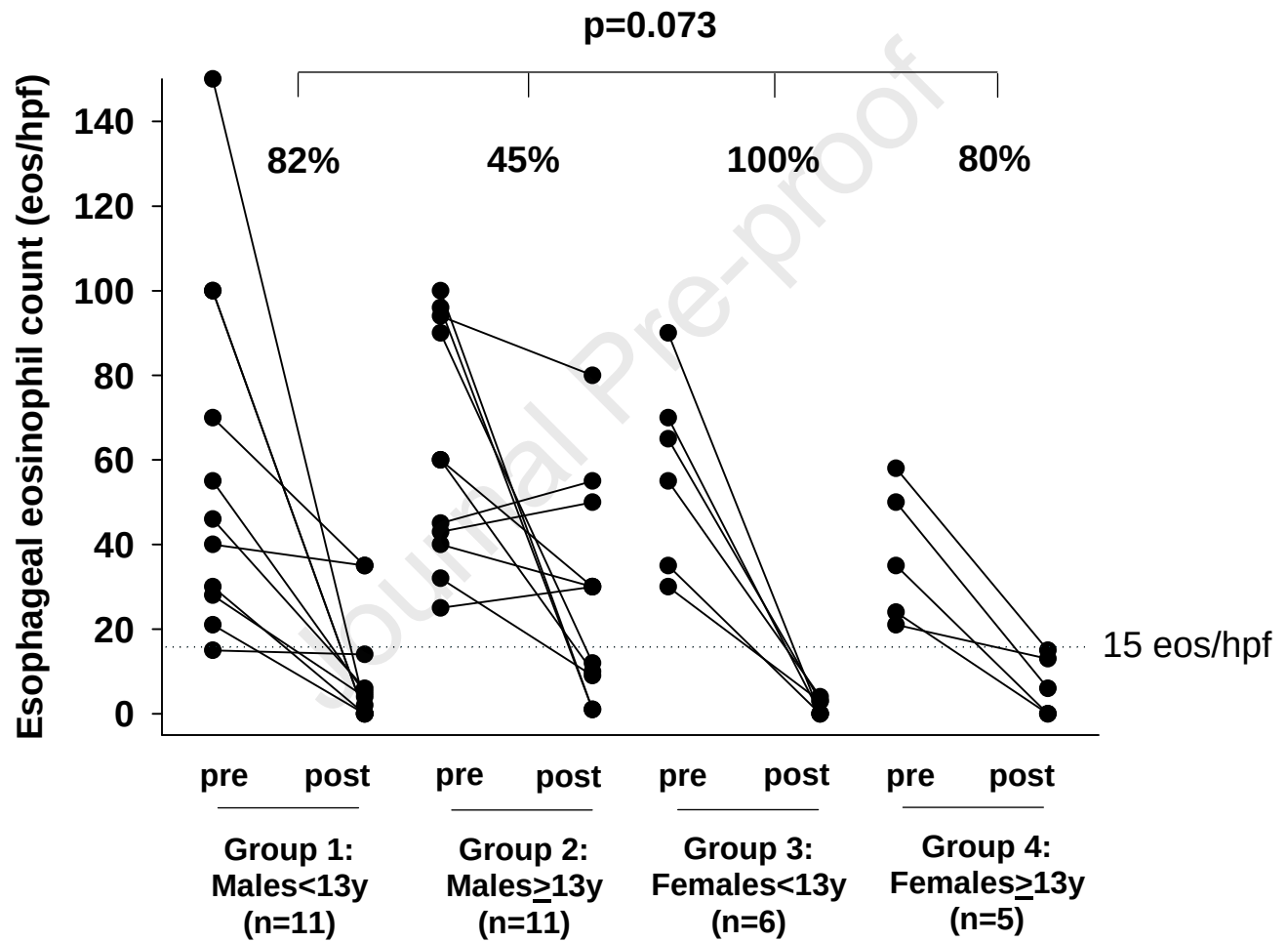


Fig. 2

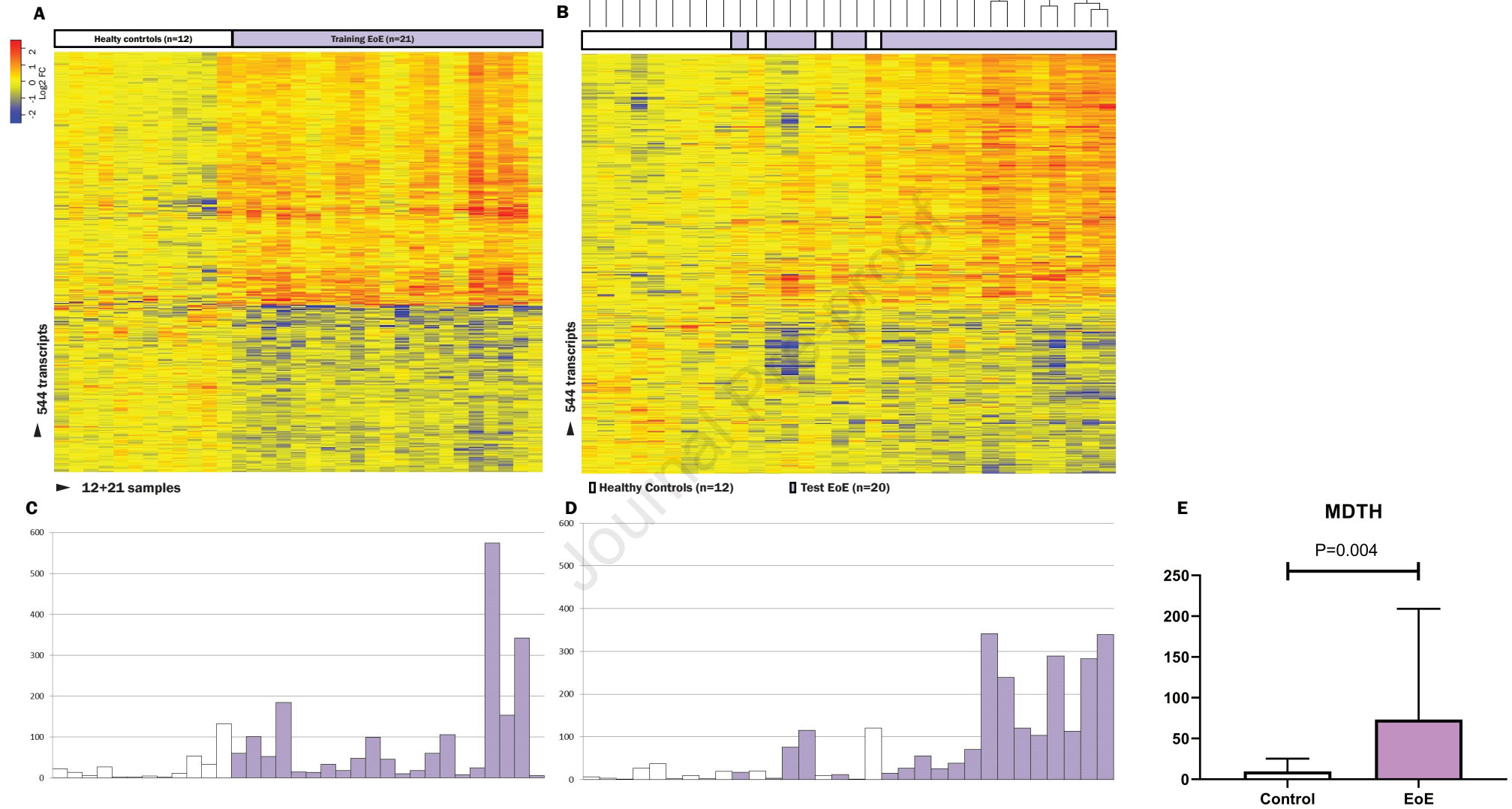


Fig. 3

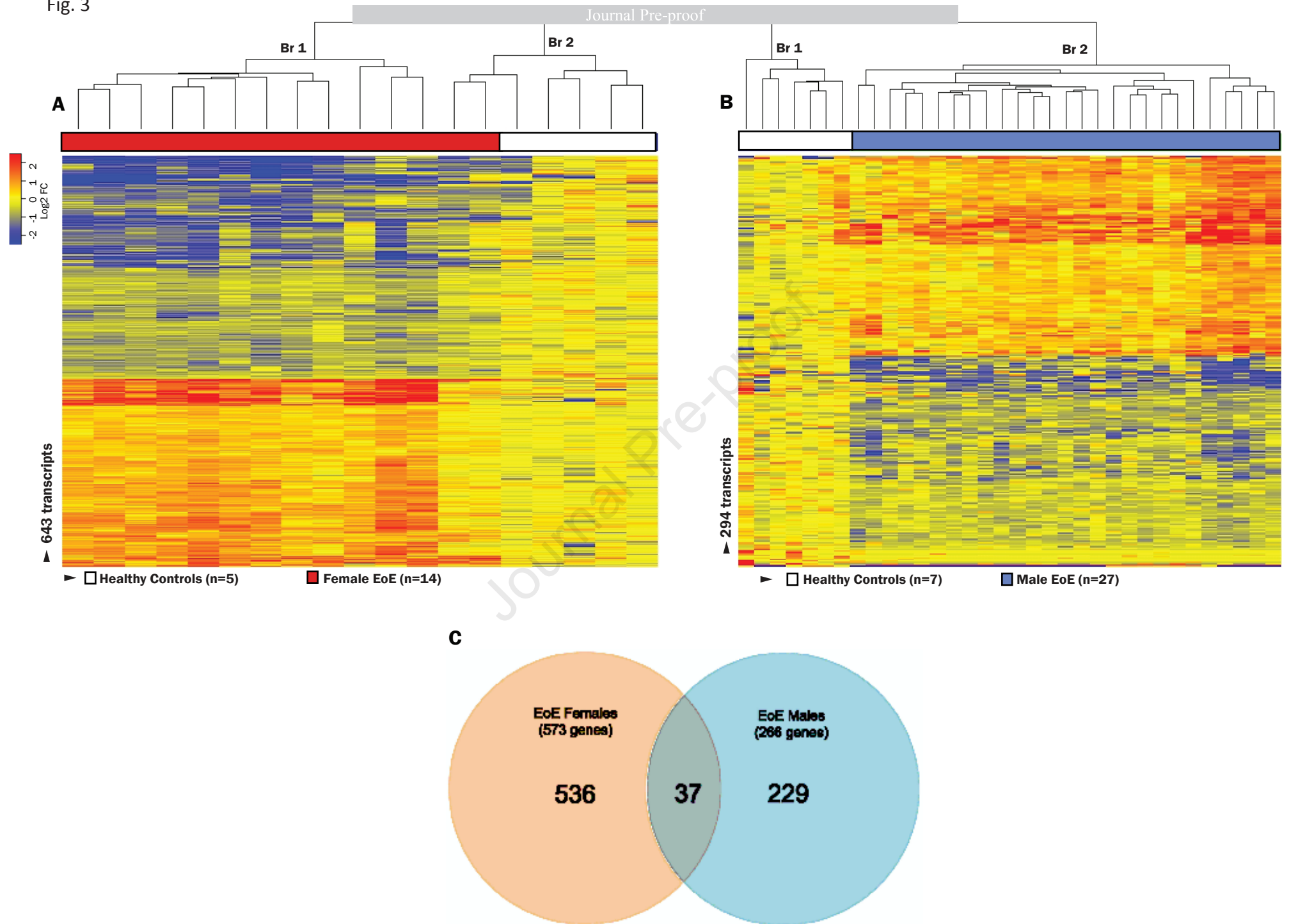
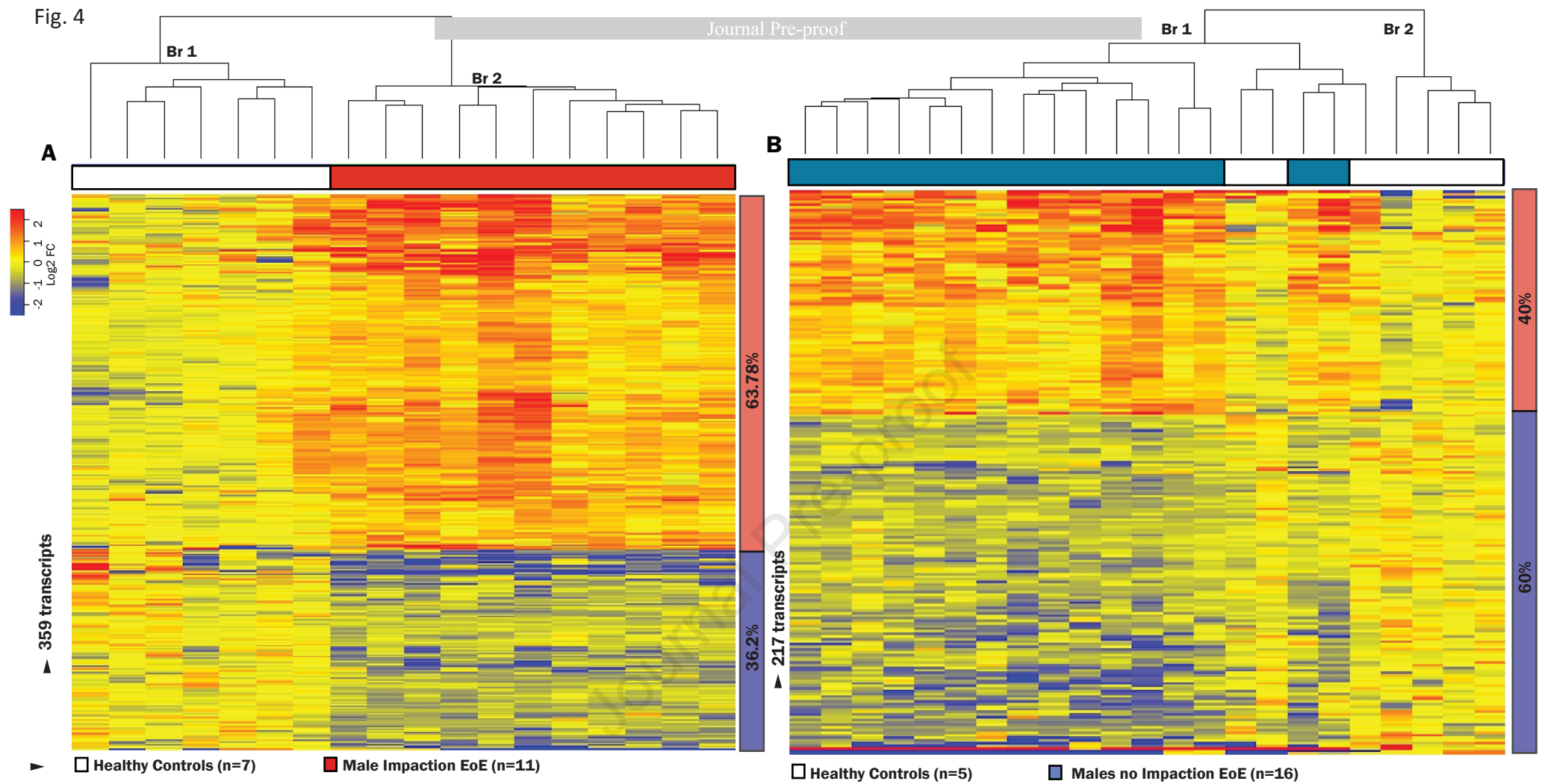


Fig. 4



C

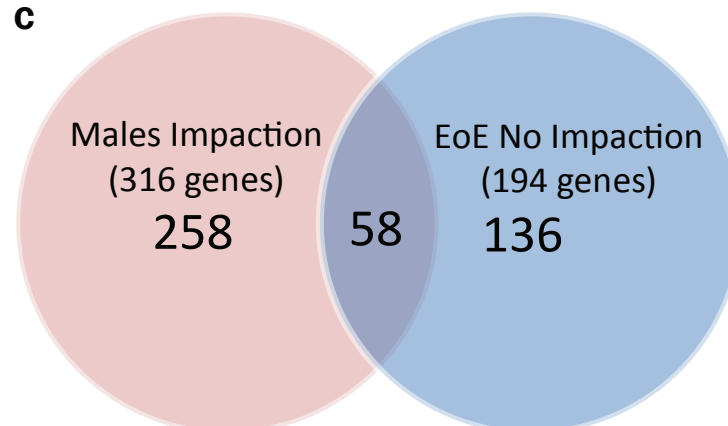


Figure Legends

Fig E1. IPA representation of top canonical pathways expressed in children with EoE.

Table E1A. List of top 10 overexpressed classifier genes that best discriminate EoE patients from healthy controls.

Gene	Expression	Function
TSPAN31	1.157	Cell surface protein that mediates signal transduction to play a role in the regulation of cell development, activation, growth and motility. Encoded protein is thought to be involved in growth-related cellular processes.
WIPI2	1.110	Encoded protein regulates assembly of multiprotein complexes. Related pathways include autophagy pathway and cellular senescence.
SGK1	1.070	Encodes a protein kinase that plays a role in cellular stress response. Associated diseases include gastric cancer and pseudohypoaldosteronism, Type I, autosomal dominant.
PRSS33	1.038	Serine protease with amidolytic activity.
NEPRO	1.031	As part of the NOTCH signaling pathway, may play a role in cortex development. Downstream may inhibit neuronal differentiation.
SARNP	0.989	Encodes a protein upregulated in response to various cytokines. May play a role in cell cycle progression.
LRCH4	0.984	Encoded protein resembles a receptor and is known to be involved in ligand binding.
STK38L	0.972	Encodes a protein involved in the regulation of differentiating and mature neuronal cells.
AHNAK	0.967	Encoded protein may play a role in diverse processes such as blood-brain barrier formation, cell structure and migration, cardiac calcium channel regulation, and tumor metastasis.
PTGES3P3	0.957	Prostaglandin E synthase 3 pseudogene.

Table E1B. List of top 10 underexpressed classifier genes that best discriminate EoE patients from healthy controls.

Gene	Expression	Function
KIAA1211	-1.575	Involved in epithelial cell integrity by acting on the maintenance of the actin cytoskeleton.
ABHD17B	-1.072	Encoded protein hydrolyzes fatty acids from proteins.
B4GALNT4	-0.956	Among its related pathways are metabolism. Related annotations include acetylgalactosaminyltransferase activity.
TRAPPC6A	-0.943	Encodes a component of the trafficking protein particle complex which tethers transport vesicles to the cis-Golgi membrane.
PNKD	-0.916	Gene thought to play a role in the regulation of myofibrillogenesis. Suspected aggravative role in cardiac hypertrophy.
ZNF765	-0.784	Among its related pathways are Herpes simplex virus 1 infection.
PDE7A	-0.782	Encoded protein belongs to the cyclic nucleotide phosphodiesterase (PDE) family and regulates transduction of cAMP. May have a role in muscle signal transduction.
SIMC1	-0.779	Associated diseases include anus cancer.
TRIM11	-0.777	This protein localizes to the nucleus and the cytoplasm. Its function is unknown. Among its related pathways are Class I MHC mediated antigen processing and presentation and innate immune system.
ZBTB2	-0.774	Encodes a transcriptional factor that represses autophagy.

Table E2A. List of top 10 genes that were overexpressed in both females and males with EoE.

Gene	Expression	Function
MS4A7	1.874	Encodes receptor associated with mature monocyte.
ANXA2P1	1.04	Annexin A2 pseudogene.
WIPI2	0.966	Encoded protein regulates assembly of multiprotein complexes. Related pathways include autophagy pathway and cellular senescence.
LOC112268105	0.952	Non-coding RNA.
BMP6	0.92	Encodes a secreted ligand of the TGF- β superfamily. These ligands bind TGF- β receptors leading to recruitment and activation of SMAD family transcription factors that regulate gene expression.
IMPDH1	0.81	Encodes a gene that catalyzes synthesis of guanine nucleotides and plays an important role in cell growth. Related pathways include innate immune system and metabolism of nucleotides.
LRCH4	0.7	Encoded protein resembles a receptor and is known to be involved in ligand binding.
FKBP3	0.573	Encoded protein is a member of the immunophilin family which plays a role in immunoregulation through inhibition of T-cell proliferation.
PTPRCAP	0.561	Encoded protein was identified as a transmembrane phosphoprotein associated with tyrosine phosphatase PTPRC/CD45, a key regulator of T and B lymphocyte activation.
PSMA3-AS1	0.549	Non protein coding RNA gene.

Table E3A. List of overexpressed classifier genes that best discriminate female EoE patients from healthy female controls.

Gene	Expression	Function
FAM206A	1.763	Encodes an actin-binding protein that regulates actin polymerization, filopodia dynamics, and increases branching of proximal dendrites of developing neurons.
TSPAN31	1.753	Cell surface protein that mediates signal transduction to play a role in the regulation of cell development, activation, growth and motility. Encoded protein is thought to be involved in growth-related cellular processes.
SARNP	1.660	Encodes a protein upregulated in response to various cytokines. May play a role in cell cycle progression.
GUCY2C	1.550	Encoded protein is a receptor for E. coli enterotoxin. Associated with diseases including meconium ileus and diarrhea.
NOTCH2	1.516	Encodes a member of the NOTCH family which play a role in intercellular signaling in a variety of developmental processes.
LRCH4	1.478	Encoded protein resembles a receptor and is known to be involved in ligand binding.
RPAP3	1.448	Encoded protein may function in transcriptional regulation and may regulate apoptosis.
MS4A1	1.432	Encodes a B lymphocyte surface molecule that plays a role in the development and differentiation of B cells into plasma cells.
CUTC	1.431	Members of the CUT family of copper transporters are associated with copper homeostasis. Related pathways are transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds and Ion channel transport.
REPS2	1.393	Gene product is part of a protein complex that regulates endocytosis of growth factor receptors. Related pathways include cytoskeleton remodeling.

Table E3B. List of underexpressed classifier genes that best discriminate female EoE patients from healthy female controls.

Gene	Expression	Function
S100BPB	-2.144	Encodes a protein that interacts with S100 calcium-binding protein P. Associated diseases include pancreatic ductal adenocarcinoma and Charcot-Marie-Tooth Disease.
RECQL	-1.876	Encoded protein is involved in DNA repair
MCM6	-1.864	Encodes a chromosome maintenance protein essential for genome replication. Single nucleotide polymorphisms of this gene have been associated with lactose intolerance.
DENND1A	-1.859	Encodes a protein that is related to internalization of proteins and lipids.
IGKV1D-33	-1.835	Immunoglobulin variable kappa 1D-33 is a protein coding gene. Gene Ontology (GO) annotations related to this gene include antigen binding.
TRAPPC6A	-1.816	Encodes a component of the trafficking protein particle complex, which tethers transport vesicles to the cis-Golgi membrane.
SIRPG	-1.739	Encoded protein belongs to immunoglobulin superfamily and is involved in negative regulation of signaling processes.
B4GALNT4	-1.730	Among its related pathways are metabolism. Related annotations include acetylgalactosaminyltransferase activity.
ACP5	-1.686	Encoded protein regulates activity of osteopontin which is involved in bone remodeling and found in macrophages and dendritic cells.
CSF3R	-1.672	Cytokine receptor that is essential for granulocyte maturation and may function in cell surface adhesion or recognition.

Table E4A. List of overexpressed classifier genes that best discriminate male EoE patients from healthy male controls.

Gene	Expression	Function
PSMC1	1.981	Encodes ATPase subunit (interacts with Hepatitis B virus and adenovirus).
PEX11B	1.341	Encodes a peroxisomal membrane protein.
ADARB1	1.081	Encodes enzyme responsible for pre-mRNA editing of viral and cellular RNAs.
AHNAK	1.057	Encoded protein may play a role in such diverse processes as blood-brain barrier formation, cell structure and migration, cardiac calcium channel regulation, and tumor metastasis.
HSP90AB1	1.032	Encoded protein is involved in signal transduction, protein folding and degradation and morphological evolution. Antagonizes STUB1-mediated inhibition of TGF-beta signaling. It is thought to play a role in gastric apoptosis and inflammation.
MARCH2	0.992	Encodes a member of the MARCH family of enzymes that regulates vesicular transport between membrane compartments.
PTGES3P3	0.979	Prostaglandin E synthase 3 pseudogene.
ZNF234	0.944	Among its related pathways are Herpes simplex virus 1 infection and gene expression.
PCTP	0.936	Encoded protein is involved in lipid transfer. Related pathways include glycerophospholipid biosynthesis and regulation of lipid metabolism.
PNPLA7	0.919	Encoded protein is a member of a family of phospholipases involved in regulation of adipocyte differentiation.

Table E4B. List of underexpressed classifier genes that best discriminate male EoE patients from healthy male controls.

Gene	Expression	Function
C11orf68	-1.087	Protein coding gene.
PCSK1N	-1.015	Encoded protein functions as an inhibitor which regulates proteolytic cleavage of neuroendocrine peptide precursors. A polymorphism may be associated with obesity.
CDC42EP3	-0.984	Encodes a protein involved in actin cytoskeleton re-organization (pseudopodia formation in fibroblasts).
TRIM11	-0.93	This protein localizes to the nucleus and the cytoplasm. Its function is unknown. Among its related pathways are Class I MHC mediated antigen processing and presentation and innate immune system.
C11orf71	-0.869	C11orf71 (Chromosome 11 Open Reading Frame 71) is an uncharacterized protein coding gene.
ZNF250	-0.836	Among its related pathways are Herpes simplex virus 1 infection and gene expression.
APOC1	-0.829	This gene encodes a member of the apolipoprotein C1 family. The encoded protein plays a central role in high density lipoprotein (HDL) and very low density lipoprotein (VLDL) metabolism.
PPAT	-0.825	Encoded protein is a member of the purine/pyrimidine phosphoribosyltransferase family that catalyzes the first step of de novo purine nucleotide biosynthesis.
ZNF765	-0.812	Among its related pathways are Herpes simplex virus 1 infection and gene expression.
ZKSCAN3	-0.81	Encodes a transcriptional factor that represses autophagy.

Table E5A. List of top 10 genes that are similarly overexpressed in EoE males with and without food impactions.

Gene	Expression	Function
PSMC1	1.944	Encodes ATPase subunit (interacts with Hepatitis B virus and adenovirus).
ADARB1	1.146	Encodes enzyme responsible for pre-mRNA editing of viral and cellular RNAs.
HSP90AB1	1.128	Encoded protein is involved in signal transduction, protein folding and degradation and morphological evolution. Antagonizes STUB1-mediated inhibition of TGF-beta signaling. It is thought to play a role in gastric apoptosis and inflammation.
WIPI2	0.985	Encoded protein regulates assembly of multiprotein complexes. Related pathways include autophagy pathway and cellular senescence.
AHNAK	0.974	Encoded protein may play a role in diverse processes such as blood-brain barrier formation, cell structure and migration, cardiac calcium channel regulation, and tumor metastasis.
PTGES3P3	0.928	Prostaglandin E synthase 3 pseudogene.
BMP6	0.848	Encodes a secreted ligand of the TGF- β superfamily. These ligands bind TGF- β receptors leading to recruitment and activation of SMAD family transcription factors that regulate gene expression.
ERC1	0.821	Encoded protein is a member of a family of proteins that regulate neurotransmitter release.
ACAD10	0.807	Gene encodes a member of the acyl-CoA dehydrogenase family of enzymes, which participate in the beta-oxidation of fatty acids in mitochondria.
SGSH	0.771	Encodes enzymes associated with lysosomal degradation.

Table E5B. List of top 10 genes that are similarly underexpressed in EoE males with and without food impactions.

Gene	Expression	Function
CDC42EP3	-0.968	Encodes a protein involved in actin cytoskeleton re-organization (pseudopodia formation in fibroblasts).
PPAT	-0.858	Encoded protein is a member of the purine/pyrimidine phosphoribosyltransferase family that catalyzes the first step of de novo purine nucleotide biosynthesis.
ZNF250	-0.835	Among its related pathways are Herpes simplex virus 1 infection and gene expression.
TRIM11	-0.822	This protein localizes to the nucleus and the cytoplasm. Its function is unknown. Among its related pathways are Class I MHC mediated antigen processing and presentation and innate immune system.
ELP5	-0.788	Encoded protein acts as subunit of the RNA polymerase II elongator complex involved in transcriptional elongation. May play a role in chromatin remodeling and is involved in acetylation of histones H3 and probably H4. Involved in cell migration (by similarity).
TRO	-0.774	Encoded protein participates in cell signaling during embryo implantation, and may also be involved in cancer formation. Gene is located near several other closely related genes on chromosome X.
NEK6	-0.759	Encodes a kinase required for progression through the metaphase portion of mitosis. Inhibition of encoded protein can lead to apoptosis. Protein also can enhance tumorigenesis by suppressing tumor cell senescence.
CCZ1P-OR7E38P	-0.75	Locus represents naturally-occurring readthrough transcription between two unprocessed unitary pseudogenes (likely does not encode functional protein).
ZNF765	-0.713	Among its related pathways are Herpes simplex virus 1 infection.
SOS1	-0.682	Membrane proteins that bind guanine nucleotides and participate in signal transduction pathways. Related pathway is gastric cancer.

Table E6A. List of top 10 genes that are differentially overexpressed in EoE males with food impactions.

Gene	Expression	Function
WWC3	1.476	Encodes a protein that is a subunit of the complex. Related Gene Ontology annotations include kinase binding and protein-containing complex scaffold activity.
PEX11B	1.468	Encodes a peroxisomal membrane protein.
PCTP	1.331	Catalyzes transfer of phosphatidycholine. Related pathways are metabolism and fatty-acyl CoA Biosynthesis.
LOC112268105	1.274	This gene is uncharacterized.
PARVB	1.249	Encodes an actin binding protein that plays a role in cytoskeleton organization and cell adhesion. These proteins contain calponin homolog domains that bind actin filaments.
PRSS33	1.244	Serine protease with amidolytic activity.
PNPLA7	1.164	Human patatin-like phospholipases, such as PNPLA7, have been implicated in regulation of adipocyte differentiation and have been induced by metabolic stimuli.
ZNF234	1.154	Among its related pathways are Herpes simplex virus 1 infection and gene expression.
AEN	1.091	Encoded protein is a nuclease that digests double-stranded DNA to form single-stranded DNA and amplifies DNA damage signals, leading to enhancement of apoptosis.
RPAP3	1.085	Encoded protein may function in transcriptional regulation and may also regulate apoptosis.

Table E6B. List of top 10 genes that are differentially underexpressed in EoE males with food impactions.

Gene	Expression	Function
COPB2	-1.530	Encoded protein essential for Golgi budding and vesicular trafficking.
LINC01662	-1.475	Non-protein coding RNA.
HEATR3	-1.374	The protein encoded plays a role in ribosomal protein transport. The encoded protein also may be involved in NOD2-mediated NF-kappaB signaling.
C1orf68	-1.295	Protein coding gene.
ITGAX	-1.258	Encodes a protein subunit of the leukocyte-specific integrin inactivated-C3b receptor 4 that is involved in adherence of neutrophils and monocytes to endothelium cells, and in the phagocytosis of complement coated particles. Related pathways include TGF-Beta pathway.
APOC1	-0.946	This gene encodes a member of the apolipoprotein C1 family. The encoded protein plays a central role in high density lipoprotein (HDL) and very low density lipoprotein (VLDL) metabolism.
ZBTB6	-0.932	May be involved in transcriptional regulation. Diseases associated with ZBTB6 include Myasthenic Syndrome, Congenital, 5
ZCCHC8	-0.929	Scaffolding subunit of the trimeric nuclear exosome targeting (NEXT) complex, a complex that directs a subset of non-coding short-lived RNAs for exosomal degradation.
ARPC2	-0.892	Encoded protein is a component of a multiprotein complex that mediates actin polymerization and mediates formation of branched actin networks in the cytoplasm, providing the force for cell motility. Also promotes homologous recombination repair in response to DNA damage.
NCK1	-0.870	Encoded protein is one of the signaling and transforming proteins containing Src homology 2 and 3 (SH2 and SH3) domains. Related pathway is IL-23 signaling pathway.

Table E7A. List of top 10 genes that are differentially overexpressed in EoE males without food impactions.

Gene	Expression	Function
MS4A7	1.918	Encodes receptor associated with mature monocyte.
VRK3	1.461	Diseases associated with VRK3 include Vaccinia. Among its related pathways are Activated TLR4 signaling and Innate Immune System.
ANXA2P1	1.121	This is a pseudogene.
MARCH2	1.055	Encodes a member of the MARCH family of enzymes that regulates vesicular transport between membrane compartments.
RIMS3	0.845	Encoded protein regulates synaptic membrane exocytosis. Related gene ontology annotations include ion channel binding.
METTL5	0.836	Encoded protein is a probable methyltransferase. Diseases associated with METTL5 include Familial Temporal Lobe, 4. Related Gene Ontology annotations include nucleic acid binding and protein methyltransferase activity.
WDR20	0.827	Encoded protein functions to preserve and regulate the activity of the USP12-UAF1 deubiquitinating enzyme complex.
UBE2G2	0.826	Protein modification with ubiquitin is an important cellular mechanism for targeting abnormal or short-lived proteins for degradation. Gene is ubiquitously expressed, with high expression seen in adult muscle. Related pathways also include innate immune system.
NOP16	0.753	Expression of this gene is induced by estrogens. It is a marker of poor patient survival in breast cancer.
ABCB1	0.697	Encoded protein is responsible for decreased drug accumulation in multidrug-resistant cells. Diseases associated with ABCB1 include Colchicine Resistance and Inflammatory Bowel Disease 13.

Table E7B. List of top 10 genes that are differentially underexpressed in EoE males without food impactions.

Gene	Expression	Function
GPALPP1	-3.282	Protein coding gene.
PCSK1N	-1.071	Encoded protein functions as an inhibitor which regulates proteolytic cleavage of neuroendocrine peptide precursors. A polymorphism may be associated with obesity.
PLGLA	-0.983	This is a pseudogene. It may bind plasminogen interfering with the binding of fibrin or alpha-2-antiplasmin to plasminogen resulting in the localization for extracellular matrix destruction.
ZKSCAN3	-0.968	Encodes a transcriptional factor that represses autophagy.
NCEH1	-0.938	Hydrolyzes 2-acetyl monoalkylglycerol ether, the penultimate precursor of the pathway for de novo synthesis of platelet-activating factor. May be responsible for cholesterol ester hydrolysis in macrophages, thereby contributing to the development of atherosclerosis.
RETREG3	-0.858	Mediates NRF1-enhanced neurite outgrowth.
PSMD9	-0.797	Encodes a subunit of the immunoproteasome which processes class I MHC peptides.
MARCH8	-0.761	Encoded protein induces internalization of several membrane glycoproteins. Related Gene Ontology annotations include ligase activity and MHC class II protein binding.
FAM13A	-0.756	Diseases associated with FAM13A include idiopathic pulmonary fibrosis. Among its related pathways are signaling by Rho GTPases and lung fibrosis.
MADD	-0.748	Protein encoded by this gene is a death domain-containing adaptor protein that interacts with the death domain of TNF-alpha receptor 1 to activate mitogen-activated protein kinase (MAPK) and propagate the apoptotic signal.