Supplementary Information

Methods

Eosinophilic Esophagitis (EoE) Cohort Identification

Subjects with EoE and their parents were systematically recruited for this study from August 2010 through May 2011 at Cincinnati Children's Hospital Medical Center (CCHMC). These pediatric patients were restricted to those with a confirmed diagnosis of EoE, which was defined as the presence of upper gastrointestinal tract symptoms and an endoscopy with ≥15 eosinophils/hpf in the proximal or distal esophageal tissue biopsies; 87% had proton pump–inhibitor therapy prior to the diagnostic endoscopic. Patients were systematically recruited with all degrees of treatment (*de novo*, untreated patients and prior patients who were untreated or treated); fifty-one patients were screened, and three did not have a biopsy sample. Given the typical EoE demographic and our relatively small sample cohort, patients in this analysis were restricted to males (two females excluded) to minimize heterogeneity. Upon initial review, all participants self-reported race as white. However, a single patient reported his race as mixed, which was noted during our data cleaning process, and was retained to maximize sample size.

Clinical Symptom Questionnaire

In addition to the use of validated metrics to capture symptoms and health-related quality of life (HRQOL) that are described below, we also captured general clinical information relevant to EoE during research interviews at the time of endoscopy. Dichotomous yes/no answers regarding gastrointestinal symptoms (food impaction, choking/gagging, clinical dysphagia, difficulty swallowing, heartburn, reflux, poor appetite, food aversions, weight loss, abdominal pain, early satiety, poor weight gain, chest pain, nausea, emesis, diarrhea, bloody stools) were gathered from parents about their children. Information on diet therapy, medication therapy, atopic history (allergies, asthma, eczema, urticaria) and general clinical parameters (cough, irritability, headache, migraines) at the time of the study was also collected.

Endoscopic Sample Collection

Patients undergoing diagnostic endoscopy for ongoing clinical care consented to provide additional esophageal biopsy specimens for research in addition to our standard clinical practice of obtaining 3 proximal

and 3 distal esophageal endoscopic biopsies. Research biopsy specimens were obtained after specimens obtained for clinical purposes. The tissues obtained for research were then placed in RNA-Later™ buffer (76104; Qiagen, Valencia, CA) and later processed for RNA extraction.

Histopathology Staining and Analysis

Biopsies obtained for clinical purposes were fixed in 10% formalin and routinely processed. Sections were cut at 5 microns and stained with hematoxylin and eosin. All clinical biopsy specimens were reviewed in a blinded fashion by a pathologist at CCHMC (M.H.C.) who has extensive experience evaluating EoE esophageal biopsies. Eosinophil counts were expressed as eosinophils/hpf (400x, 0.3 mm^2) and recorded as peak eosinophil counts in the proximal and distal biopsies.

Additional sections were stained with antibody to tryptase (Cell Marque, Rocklin, CA; predilute, EDTA with 8 minutes antigen retrieval) or chymase (AbD Serotec, Raleigh, NC, 1:50, EDTA with 36 minutes antigen retrieval). The most densely inflamed areas were identified, and peak counts of tryptase-positive and chymase-positive cells in the most inflamed hpf were recorded.

Eosinophil Peroxidase (EPX)–based Immunohistochemistry

EPX-based immunohistochemistry of esophageal tissues sections from formalin-fixed, paraffin-embedded biopsies was performed as previously described.[5] Briefly, infiltrating, intact eosinophils and evidence of eosinophil degranulation (i.e., the presence of free cytoplasmic granules and/or extracellular matrix deposition of EPX) were assessed by immunohistochemistry by using an eosinophil-specific mouse monoclonal antibody (Clone: MM25-82.2.1). Immunocytochemical staining was performed with Dako detection/visualization reagents purchased from Dako Cytomation (Carpinteria, CA). Eosinophil-containing sections from patients identified by traditional pathologic assessments were used as positive control slides. Negative control slides (i.e., antibody isotype controls and negative tissue control sections) were also included as part of the processing of each group of slides examined. Permanent Red Substrate-Chromogen (Dako Cytomation, Cat. No. K0695) was used to visualize EPX-specific staining. Slides were counterstained with methyl green prior to

being mounted on coverslips and photomicroscopy with a Zeiss Axiophot microscope equipped with an AxioCam MRc5 digital camera.

RNA Extraction and Reverse Transcription

Total RNA was extracted from single, distal esophageal biopsies collected in RNAlater (Qiagen, 76104) using the miRNeasy RNA Extraction Kit (Qiagen, 217004) following the manufacturer's manual protocol. After RNA quantity and quality analysis by NanoDrop spectrometer, an aliquot of 500 ng of RNA was acquired for reverse transcription by the iScript cDNA Synthesis Kit (BioRad 170-8891) following the manufacturer's manual protocol. Briefly, 500 ng of RNA was mixed with reaction mix and the reverse transcriptase enzyme in a total volume of 20 µl, incubated at 25ºC for 5 minutes, 42ºC for 30 minutes and 85ºC for 5 minutes and then kept at 4ºC and later -20ºC for storage.

PCR Amplification of Representative EoE genes

A selection of representative EoE genes was amplified from cDNA stock generated as mentioned above. The Taqman reagents for amplification of major EoE signature genes[6, 11] were obtained from Applied Biosystems, and Taqman real-time PCR amplification was performed on an ABI 7900HT System (Applied Biosystems). The amplification protocol consisted of a hot start of 95^{o}C for 10 minutes followed by 40 cycles of 15 seconds at 95^{o}C and 1 minute at 60^{o}C. After the quantitative PCR was complete, raw Ct values for each sample / each gene were exported into GeneSpring GX 11.5 (Agilent Technologies) for statistical analysis and heat map generation. *GAPDH* was used as an expression control for all of the genes analyzed.

The Pediatric EoE Symptom Score, Version 2.0 (PEESS® v2.0) and Domains

The PEESS® v2.0 has 20 questions: 11 ask the participants to rank the frequency of symptoms, and 9 ask the participants to rank the severity of symptoms. Both frequency and severity are scored from 0 to 4, with 4 being the worst (Table 1). In the validation phase of the PEESS® v2.0, parents of patients with EoE were asked what the questions were about and how the items were related. From these parent participant interviews, four domains for these 20 questions were established: dysphagia, gastroesophageal reflux disease (GERD),

nausea/vomiting and pain. We created total and domain scores by summing the individual elements and dividing by the maximum total score. Because only a subset of children were eligible to fill out the PEESS® v2.0 (because of age restrictions), all validation of the PEESS® v2.0 scores were conducted using the parent proxy–report scores.

Data Management and Statistical Analyses

Data Management

All clinical data were collected on a paper form and entered into one of two electronic datasets. Basic research data was collected through the use of a Structured Query Language (SQL) Server database developed and maintained in the Cincinnati Center for Eosinophilic Disorders. These data, as was all clinical laboratory data, were captured using DocFlowSheets in an EPIC electronic medical record and were later extracted for analysis. Once the forms were all entered, the electronic record was compared to the paper forms to ensure that there were no discrepancies from data entry. Data extracted from the separate databases were then joined using statistical software.

Statistical Analysis

Prior to analyses, all variables were examined for distributional issues and plausibility of the data. Descriptive statistics were reported as means and SDs or medians, interquartile ranges (IQRs), and frequencies depending on the variable distributions.

Consistency Between Parent-Proxy Reports and Child Self-Report Scores

Self-reported outcomes are considered the goal standard for patient-reported outcomes (PROs), yet children below a certain age may not have the developmental capacity to answer such questionnaires[52]. Thus, we examined the subset of participants with both parent proxy–reported and child self-reported PEESS® v2.0 scores and used Spearman correlations (due to the non-normal distributions).

Interrelationship of PEESS® v2.0 Domains

To gain an understanding of how different instruments relate to each other in a pediatric EoE cohort, we first performed pairwise analysis between the PEESS® v2.0 domains using Wilcoxon signed-rank paired test to determine whether the EoE pediatric cohort had specific domains that they perceived as worse than others. This resulted in 10 statistical tests. To account for the possibility of false positives in these tests, we used a standard Bonferroni correction (0.05/10 = 0.005). Additionally, we used Spearman correlations to understand how these different measures co-varied with each other. In total, 10 correlations were performed; the p value for significance of the correlation in this analysis is $\alpha = 0.005$.

Validation of PEESS® v2.0 Domains with Respect to Clinical Features

To determine whether PEESS® v2.0 domains are associated with clinical symptoms in an anticipated manner, we tested whether individuals who reported a symptom had worse domain scores than those who did not using Wilcoxon Rank Sum. As many tests were performed, a multiple testing adjustment was included. After accounting for the 17 gastrointestinal symptoms with frequencies greater than 5%, a p value of 0.003 was required (Bonferroni correction 0.05/17 = 0.003) to reach statistical significance. Given the high level of correlation between the PEESS® v2.0 domains, no additional multiple testing corrections were applied. We also examined the relationship of the domains to allergic symptoms as a negative control, as we would not expect the PEESS® v2.0 domains to differ in individuals with and without allergic symptoms. As this was a negative control, the significance threshold was 0.05.

Validation of PEESS® v2.0 Domains with EoE Biological Features

To determine whether the *PEESS® v2.0* domains aligned with the biological features associated with EoE, we examined several aspects of EoE biology, including a diagnostic subset of the EoE transcriptome (EoE Diagnostic Panel [EDP]) and histological features (eosinophil and mast cell measures). As previous work identified a gene expression panel highly specific to active EoE (EDP),[11] we tested the hypothesis that the EDP (either as a whole or individual genes) would be associated with the PEESS® v2.0 domains. For this analysis, we performed Spearman correlation analysis between the gene levels on the EDP and the PEESS® v2.0 domains. In addition, we tested whether the median of the correlation (using the absolute value to account for

differences in the direction of the effect across genes) was enriched for any of the domains using Wilcoxon Rank Sum.

To test the hypothesis that the PEESS® v2.0 domains are related to eosinophil measures, we considered both the quantitative measure of peak eosinophil count and a disease activity classification schema defined as: none (0 eosinophils/hpf), low (1-5 eosinophils/hpf), intermediate (6-14 eosinophils/hpf), and active (≥15 eosinophils/hpf) using the peak eosinophil count (maximum of the proximal and distal counts). Additionally, we examined eosinophil activation as measured by EPX staining from both proximal and distal samples. Analyses using continuous eosinophil counts (proximal, distal and peak counts) were evaluated using Spearman correlations. For analyses using categorical eosinophil counts, Wilcoxon Rank Sum was used. For the analysis of EPX staining, Spearman rank correlations were used due to the non-normal distribution of the EPX staining values. To account for multiple testing within eosinophil measures, we used a Bonferroni correction adjusting for the overall correlation between these 15 measures ($\rho = 0.76$), yielding $\alpha = 0.03$.

To test the hypothesis that PEESS® v2.0 domains are associated with mast cells, we tested for correlation between PEESS® v2.0 domains and levels of tryptase, chymase and *CPA3* mRNA. To quantify mast cell markers, we first examined the gene expression from the distal esophagus for tryptase, chymase and *CPA3*. We also examined the tryptase and chymase quantitative staining from both distal and proximal esophagus biopsies with the amounts varying due to the number of samples (n = 30-40) collected from each esophageal location; mast cell levels were expressed as cell number per hpf. To account for the multiple testing within mast cell parameters, we used a Bonferroni correction adjusting for the overall correlation between these 7 measures ($\rho = 0.56$), yielding $\alpha = 0.02$.

Supplementary Tables

Table S1: Question Composition of Parent Proxy-Reported PEESS® v2.0 Domains

Domain	Question
Dysphagia	How often does your child have trouble swallowing?
	How bad is your child's trouble swallowing?
	How often does your child feel like food gets stuck in his/her throat or chest?
	How bad is it when your child gets food stuck in his/her throat or chest?
	How often does your child need to drink a lot to help swallow food?
	How bad is it when your child needs to drink a lot to help swallow food?
	How often does your child eat less than others?
	How often does your child need more time to eat than others?
GERD	How often does your child have heartburn (burning in the chest, mouth, or throat)?
	How bad is your child's heartburn (burning in the chest, mouth, or throat)?
	How often does your child have food come back up in his/her throat when eating?
	How bad is it when food comes back up in your child's throat?
Nausea/Vomiting	How often does your child vomit (throw up)?
	How bad is your child's vomiting (throwing up)?
	How often does your child feel nauseous (feel like throwing up, but doesn't)?
	How bad is your child's nausea (feeling like throwing up, but doesn't)?
Pain	How often does your child have chest pain, ache, or hurt?
	How bad is your child's chest pain, ache, or hurt?
	How often does your child have stomach aches or belly aches?
	How bad are your child's stomach aches or belly aches?

Table S2. Frequency of Allergic Conditions

	N with data available	Frequency
Drug Allergy	45	31.1
Environmental Allergy	43	67.4
Food Allergy	46	84.8
Urticaria	46	65.2
Eczema	46	58.7
Asthma	45	51.1

Table S3: EPX Staining and Mast Cell Measures+.

Measurement	Distal	Proximal
EPX Staining Values		
Total	12.5 (0-37.8, 0-50)	0 (0-32, 0-48)
Reproducibility	1.5 (0-4, 0-4)	0 (0-3, 0-4)
Patchiness	0 (0-6, 0-8)	0 (0-2, 0-8)
Degranulation	0 (0-3, 0-10)	0 (0-4, 0-8)
Intact	4.5 (0-12, 0-12)	0 (0-12, 0-12)
Randomness	2 (0-9, 0-16)	0 (0-8, 0-16)
Tryptase		
Epithelium	10.5 (6-24.8, 0-49)	9 (6-22, 1-65)
Chymase		
Epithelium	4.5 (2-10.8, 0-28)	5 (2-8.5, 0-43)
Gene Expression*		
Tryptase	-0.07 (-1.00-1.78, -3.8 -3.23)	--
Chymase	-0.04 (-7.8-1.03, -11.10-4.26)	--
CPA3	-0.05 (-1.5 – 2.07, -4.19 – 5.01)	--

+All values are reported as medians (IRQ; range)
*Gene expression levels normalized to housekeeping gene GAPDH.

Table S4: The relationship of PEESS scores and in individuals with active EoE (≥15 eos) and who are in remission (no eos). Data are presented as median and interquartile ranges. P-values are from Wilcoxon Rank Sum.

PEESS scores	Active (n = 33)	Remission (n = 3)	Non active (n = 13)	P value active vs remission	P value active vs non active
Total score	25.0 (10.0 – 37.5)	22.5 (8.8 – 23.4)	18.75 (8.8 – 41.8)	0.37	0.16
Dysphagia	25.0 (12.5 – 39.7)	25.0 (9.4 – 33.3)	12.5 (9.4 – 26.6)	0.89	0.14
GERD	18.8 (0 – 28.1)	6.3 (6.3 – 8.3)	18.75 (3.1 – 28.1)	0.23	0.62
Nausea/Vomiting	12.5 (0 - 40.6)	31.3 (12.5 – 37.5)	12.5 (0 – 34.4)	0.45	0.77
Pain	31.25 (12.5 – 43.8)	8.3 (6.3 – 18.8)	18.75 (10.4 – 31.3)	0.13	0.21

Table S5: **Spearman correlations between PEESS® v2.0 scores and a diagnostic subset of the eosinophilic esophagitis transcriptome.** Correlations by functional groupings of genes. The text within the cell indicates the Spearman's r. Values in boldface indicate nominal statistical significance ($p < 0.05$).

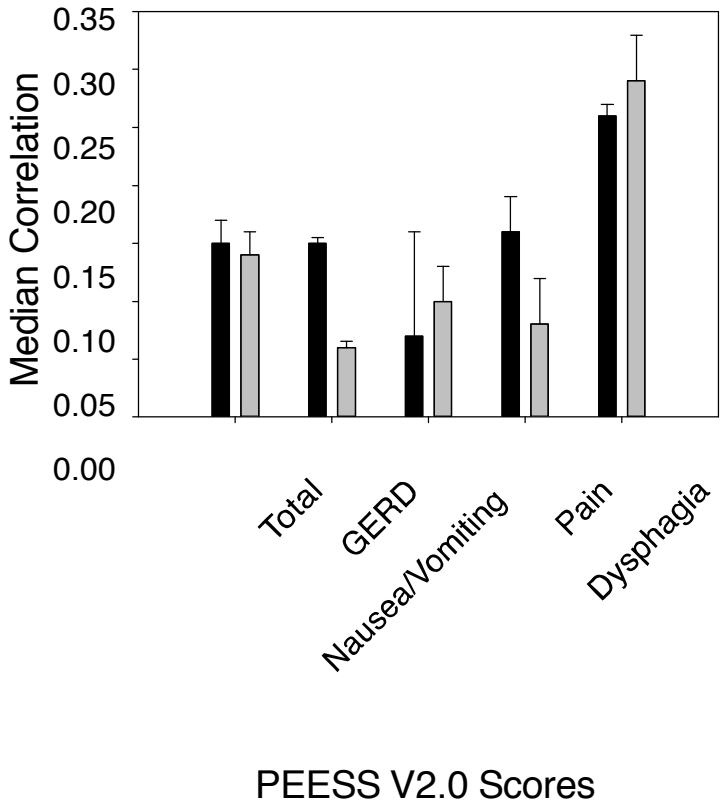
Function	Gene	Total	Dysphagia	GERD	NV	Pain
Cell Adhesion	*CDH20*	0.17	-0.15	-0.02	-0.02	0.01
	CDH26	0.00	0.27	-0.12	0.01	0.10
	CHL1	-0.05	0.17	-0.04	-0.09	0.05
	CLDN10	0.00	-0.23	-0.08	-0.10	-0.25
	CTNNAL1	0.07	-0.09	-0.06	-0.10	-0.12
	DSG1	-0.09	-0.12	0.07	0.05	-0.17
Chemokines	*CXCL1*	-0.04	0.23	0.02	0.00	0.07
	CXCL6	-0.06	0.28	-0.11	-0.03	0.09
Cytokines	*IL13*	0.00	0.26	-0.07	0.01	-0.01
	IL32	0.16	0.14	0.10	0.14	0.11
	IL33	-0.32	0.03	-0.15	-0.18	-0.12
	IL4	-0.13	**0.42**	0.03	0.21	0.18
	IL5	-0.02	0.21	0.00	-0.06	0.05
	IL8	-0.11	0.31	0.07	0.10	0.19
Eosinophilia	*CCR3*	0.01	0.28	0.01	-0.07	0.03
	CLC	0.02	**0.33**	0.02	0.00	0.11
	IL5RA	-0.11	**0.41**	0.16	0.09	0.09
Epithelial	*ACPP*	-0.04	-0.05	0.08	-0.11	-0.04
	CA2	-0.07	0.18	0.02	0.09	0.01
	CRISP2	0.02	-0.10	-0.02	-0.18	-0.08
	CRISP3	0.09	-0.10	0.08	-0.15	-0.09
	EPPK1	-0.11	0.20	0.04	0.01	0.01
	FLG	0.03	-0.14	0.08	-0.02	-0.13
	GCNT3	0.05	0.20	-0.10	-0.02	0.06
	MUC4	-0.04	0.26	-0.08	0.13	0.10
	PHLDB2	-0.07	0.21	0.01	0.06	0.03
	SPINK7	0.03	-0.14	-0.02	-0.08	-0.21
	UPK1A	0.00	-0.02	-0.02	0.23	0.02
	UPK1B	0.06	0.06	-0.09	-0.13	0.01
Inflammation	*ALOX12*	0.23	-0.31	-0.03	-0.08	-0.27
	ALOX15	0.05	0.15	-0.05	-0.01	0.03
	APOBEC3A	0.16	0.13	-0.07	-0.04	0.02
	ARG1	0.11	-0.13	0.12	-0.15	-0.12
	CCL26	-0.05	0.22	-0.03	0.00	0.09
	CD200R1	-0.10	0.21	0.05	0.00	0.04
	CD244	0.10	0.09	-0.20	-0.05	-0.05
	CFB	0.00	0.08	0.00	-0.03	0.04
	CFI	-0.01	0.22	0.04	-0.10	0.07
	CITED2	**0.42**	-0.10	-0.10	-0.07	-0.27
	FCGR3B	-0.16	0.31	0.01	0.00	0.05
	GPR44	-0.13	0.25	0.08	0.07	0.14
	GRK5	-0.01	0.11	-0.11	0.00	-0.08
	HPGDS	-0.09	**0.35**	0.12	0.04	0.12
	HRH1	-0.05	0.25	0.01	-0.06	0.05
	IGJ	0.07	0.15	-0.06	-0.02	-0.10
	MMP12	0.02	0.22	0.03	-0.02	0.06
	PMCH	-0.04	0.19	-0.05	-0.01	0.01
	PTGFRN	-0.02	-0.08	-0.07	-0.12	-0.11
	RUNX2	-0.12	0.24	-0.07	-0.06	-0.07
	SAMSN1	-0.16	**0.40**	0.11	0.01	0.16
	TNFAIP6	0.04	0.22	-0.12	-0.01	0.00
	TSLP	-0.18	0.14	0.14	0.03	0.11
	ZNF365	0.13	-0.07	-0.01	-0.05	-0.16
Ion Channels	*ANO1*	0.05	0.07	-0.09	-0.08	-0.04
	KCNJ2	-0.04	0.16	0.02	-0.06	0.01
	SLC16A6	-0.07	-0.11	0.06	-0.02	-0.07
	SLC26A4	-0.15	**0.35**	0.08	0.09	0.13
Mast Cells	*CMA1*	0.05	0.19	0.14	0.05	0.01
	CPA3	-0.08	**0.36**	0.04	0.00	0.12

	TPSB2	0.03	0.28	-0.04	-0.07	0.00
Neurosensory	*NEFL*	-0.07	0.20	0.06	-0.03	0.07
	NEFM	-0.09	0.18	0.05	-0.01	0.08
Other/Unknown	*C7orf68*	0.09	-0.11	0.04	0.07	-0.10
	CDA	-0.10	-0.19	0.00	0.15	-0.10
	EML1	0.26	-0.29	-0.14	-0.25	-0.27
	ENDOU	0.06	-0.21	0.02	-0.11	-0.13
	GAPDH	-0.07	0.16	-0.04	-0.03	0.00
	GLDC	-0.08	0.16	0.02	0.00	0.01
	GPR160	-0.16	**0.36**	0.11	0.02	0.08
	LRRC31	-0.18	**0.33**	0.02	0.09	0.17
	MSRB3	0.08	-0.23	-0.14	-0.17	-0.31
	PNLIPRP3	-0.03	0.02	0.02	0.00	-0.10
	SUSD2	0.03	0.14	0.02	-0.04	0.06
	SYNPO2	0.13	0.09	0.04	0.02	-0.09
	SYNPO2L	0.08	-0.21	0.12	-0.12	-0.12
Proliferation, growth, cell cycle	*CRYM*	0.14	-0.08	-0.04	-0.15	-0.05
	GYS2	0.08	-0.04	0.09	-0.07	-0.04
	IGFL1	-0.13	-0.06	0.05	0.02	-0.08
	MT1M	0.01	0.02	0.26	0.10	0.04
	UBD	0.03	0.17	-0.03	-0.06	-0.01
	GRPEL2	0.10	-0.06	-0.01	0.06	-0.17
	RTP4	-0.06	0.24	0.09	-0.02	0.07
Remodeling	*ACTG2*	0.29	-0.20	0.06	-0.01	-0.17
	COL8A2	-0.14	0.11	-0.06	0.12	0.01
	CTSC	-0.10	0.22	-0.02	0.03	0.05
	KRT23	0.13	0.17	0.01	0.00	0.04
	POSTN	0.02	0.06	-0.07	-0.04	0.00
	TGFB1	-0.06	-0.04	-0.01	0.07	-0.22
Steroid responding genes	*EPB41L3*	0.07	0.08	0.12	-0.11	0.10
	F3	0.02	0.04	-0.21	-0.05	0.00
	FKBP5	-0.05	-0.23	-0.18	-0.07	-0.16
	H19	0.00	0.04	0.27	0.08	0.02
	TRIM2	-0.04	0.27	-0.09	0.18	0.05
Vascular development	*TSPAN12*	-0.04	-0.29	-0.10	-0.14	-0.22
	VEGFA	-0.16	0.31	-0.06	0.03	0.14

Supplemental Figures

Figure S1. Median Correlation (and upper 75% quartile) between eosinophil measures (eosinophil count, EPX staining) and PEESS® V2.0 demonstrates specific association with dysphagia domain. Black bars are from distal esophagus and grey bars are from proximal esophagus. The median dysphagia domain is statistically higher than the total and the domain scores for both proximal and distal ($p \leq 0.0049$ for all except distal pain, $p = 0.022$).

Figure S1.


Median Correlation
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Total
GERD
Nausea/Vomiting
Pain
Dysphagia
PEESS V2.0 Scores