

The genetic etiology of eosinophilic esophagitis



Leah C. Kottyan, PhD,^{a,b,c} Sreeja Parameswaran, PhD,^c Matthew T. Weirauch, PhD,^{a,c,d}
Marc E. Rothenberg, MD, PhD,^{a,b} and Lisa J. Martin, PhD^{a,e}

Cincinnati, Ohio

Eosinophilic esophagitis (EoE) is a chronic allergic disease associated with marked mucosal eosinophil accumulation. Multiple studies have reported a strong familial component to EoE, with the presence of EoE increasing the risk for other family members with EoE. Epidemiologic studies support an important role for environmental risk factors as modulators of genetic risk. In a small percentage of cases, including patients who have Mendelian diseases with co-occurrent EoE, rare genetic variation with large effect sizes could mediate EoE and explain multigenerational incidence in families. Common genetic risk variants mediate genetic risk for the majority of patients with EoE. Across the 31 reported independent EoE risk loci ($P < 10^{-5}$), most of the EoE risk variants are located in between genes (36.7%) or within the introns of genes (42.4%). Although some variants do change the amino acid sequence of genes (2.2%), only 3 of the 31 EoE risk loci harbor an amino acid-changing variant. Thus most EoE risk loci are outside of the coding regions of genes, suggesting a key role for gene regulation in patients with EoE,

which is consistent with most other complex diseases. (J Allergy Clin Immunol 2020;145:9-15.)

Key words: Eosinophilic esophagitis, genetics, single nucleotide polymorphism, heritability, genome-wide association study

Eosinophilic esophagitis (EoE) is a chronic allergic disease associated with marked mucosal eosinophil accumulation.¹ EoE usually remits after removal of specific food types, and food reintroduction causes disease recurrence, including marked dysregulation of esophageal transcripts. One of the central questions in the EoE field and allergy in general is to understand why patients have certain manifestations of allergic disease, such as EoE. Defining a “typical” EoE case, beyond the presence of eosinophils in the esophagus, is challenging. Patients often exhibit marked heterogeneity in their presentation (eg, age of onset, comorbidities, and symptomology) and response to treatment. The diversity of presentations makes identification of etiology challenging.

From ^athe Department of Pediatrics, University of Cincinnati, College of Medicine and ^bthe Division of Allergy and Immunology, ^cthe Center for Autoimmune Genomics and Etiology, ^dthe Divisions of Biomedical Informatics and Developmental Biology, and ^ethe Division of Human Genetics, Cincinnati Children’s Hospital Medical Center.

L.C.K. is funded by grant R01 DK107502 and a Cincinnati Children’s Hospital Medical Center (CCHMC) Center for Pediatrics Genomics (CpG) award. M.T.W. is supported by a Cincinnati Children’s Research Foundation Endowed Scholar Award and a CCHMC CpG award. M.E.R.’s work is funded by National Institutes of Health grants R37 AI045898, U19 AI070235, R01 AI057803, and R01 DK107502; the Campaign Urging Research for Eosinophilic Disease (CURED); the Sunshine Charitable Foundation and its supporters, Denise and David Bunning. L.J.M. is supported by National Institutes of Health grants U19 AI070235, U54 AI117804, R01 AI124355, and R01 DK107502.

Disclosure of potential conflict of interest: L. C. Kottyan has no conflicts of interest to disclose and is funded by R01 DK107502 and a Cincinnati Children’s Hospital Medical Center (CCHMC) Center for Pediatrics Genomics (CpG) award. M. T. Weirauch has received the Cincinnati Children’s Research Foundation Endowed Scholar Award and a CCHMC CpG award. M. E. Rothenberg is a consultant for PulmOne, Spoon Guru, ClostraBio, Celgene, AstraZeneca, and Allakos; has an equity interest in the first three listed; received royalties from reslizumab (Teva Pharmaceuticals), PEESSv2 (Mapi Research Trust), and UpToDate; and is an inventor of patents owned by Cincinnati Children’s and his work is funded by National Institutes of Health (NIH) grants R37 AI045898, U19 AI070235, R01 AI057803, and R01 DK107502; the Campaign Urging Research for Eosinophilic Disease (CURED); and the Sunshine Charitable Foundation and its supporters, Denise and David Bunning. L.J.M. is supported by National Institutes of Health grants U19 AI070235, U54 AI117804, R01 AI124355, and R01 DK107502. S. Parameswaran declares no relevant conflicts of interest.

Received for publication September 27, 2019; revised November 15, 2019; accepted for publication November 15, 2019.

Corresponding author: Leah C. Kottyan, PhD, 3333 ML 15026 Burnet Ave, Cincinnati, OH 45229. E-mail: Leah.Kottyan@cchmc.org.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

0091-6749/\$36.00

© 2019 American Academy of Allergy, Asthma & Immunology

<https://doi.org/10.1016/j.jaci.2019.11.013>

Terms in boldface and italics are detailed in the glossary on page 10.

ETIOLOGY OF EoE: ENVIRONMENTAL AND GENETIC CONTRIBUTIONS

Multiple studies have reported a strong familial component to EoE, with the presence of EoE increasing the chance of having the disease for other family members.^{2,3} However, it is important to recognize that traits shared between family members can be the result of both genetic and environmental factors.

Any mention of genetics often evokes the pioneering work of Gregor Mendel using pea plants to elucidate inheritance. Mendelian inheritance focuses on single variants that completely explain a phenotype. For some families, single-gene variants can drive EoE inheritance. Indeed, multiple studies have shown EoE co-occurring with known Mendelian disorders, supporting a single genetic cause. However, like many diseases, the familial patterning in patients with EoE is often not consistent with Mendelian inheritance. Only approximately 2% of siblings of an EoE proband also have EoE.² Such a low recurrence risk of disease is not consistent with Mendelian modes of inheritance. Furthermore, because there is transmission from fathers to sons, X-linked inheritance is not supported.

Collectively, these results strongly support that for the most part, EoE is a disease with a complex genetic cause, and inheritance is due to the effects of multiple genetic *loci* that increase disease risk in the context of environmental disease risk-modifying factors. As with other complex traits, genetic loci that increase EoE risk are identified before the modified genes can be determined.

Comparing dizygotic and monozygotic twins is an excellent way to measure the contribution of genetic factors for a

Abbreviations used

EoE: Eosinophilic esophagitis
 GWAS: Genome-wide association study
 LD: Linkage disequilibrium
 PLAU: Plasminogen activator, urokinase

disease. The study by Alexander et al² found that monozygotic (ie, identical) twins had a 44% disease concordance, a 2-fold increase compared with dizygotic twins. Although the study had insufficient power to statistically distinguish the dizygotic and monozygotic concordance, these data are consistent with a complex genetic component that acts in the context of environmental exposures to increase the risk of EoE.

Because families live in the same household, they often share environmental exposures. Epidemiologic studies seek to identify those exposures that are shared by patients with EoE but not subjects without the disease.⁴ Leveraging the unique information provided by nuclear families and twins, Alexander et al² sought to address the total contribution of genetics to the familial inheritance observed in patients with EoE. Evaluating rates of EoE from the relatives of EoE probands, they found that dizygotic twins had a nearly 10-fold increase in EoE rate compared with siblings. Genetically, siblings and dizygotic twins would be expected to have similar shared genetic influences; thus the increased rate of EoE in dizygotic twins can be attributed to shared early-life environmental factors. The potential role of a shared environment was also noted by Allen-Brady et al,³ who observed that spouses of EoE probands were also at increased risk for EoE.

Jensen et al⁵ evaluated the role of early-life factors, finding that prenatal (maternal fever and preterm labor), intrapartum (cesarean section), and early-life (antibiotic use and acid suppressant use) factors were all associated with increased risk of EoE, whereas having a furry pet in infancy was associated with decreased risk. However, no one disease risk factor has been identified as sufficient for the development of EoE. Studies have identified environmental factors affecting both children and adults that have been associated with the risk of EoE (sensitization to galactose- α ,1,3-galactose) and protection (*Helicobacter pylori* infection).⁴ Taken together, these analyses support a complex genetic cause with contributions from both genetic and environmental factors.

GENETIC CAUSE OF EoE

To identify the regions of the genome that increase the risk of having EoE, the DNA of patients with and without EoE is genotyped on a microarray or sequenced using next-generation sequencing technology. Common and rare disease genetic variants that increase disease risk are identified through different analytic approaches. Variants are typically considered common if the less frequent minor allele can be found on at least 1% of chromosomes across a large population of patients; variants are considered rare when they occur at a frequency of less than 1%. For common variants, the typical approach is to evaluate a large number of patients with disease and compare them with subjects without disease to identify those variants whose minor alleles are enriched (ie, present at a higher frequency) in patients with the disease.⁶ These variants typically have modest *effect sizes*, with the genetic risk variant increasing disease risk by 30% to

GLOSSARY

EFFECT SIZE: Effect size aims to describe the magnitude of difference between groups or the magnitude of association between variables and not just whether a difference/association is statistically significant. Absolute effect size is the raw difference between groups. Standardized measures of effect or calculated indices of effect size are used when the measured differences have no intrinsic meaning or when different studies under evaluation might be using different scales, such as in a meta-analysis.

HAPLOTYPES: A set of closely linked genetic markers (alleles) present on 1 chromosome, which tend to be inherited together.

IL-4: A cytokine released by T_H2 cells, as well as basophils and mast cells. IL-4 is critical for T_H2 cell differentiation and promotes B-cell class-switching toward IgE production.

IL-5: A cytokine that promotes the growth and differentiation of eosinophils. IL-5 is released by T_H2 cells, basophils, and eosinophils.

IL-13: A cytokine with essentially the same features as IL-4 (see above).

INSULATORS: A class of DNA sequence elements that have the common property of shielding genes from counterproductive signals emanating from their surrounding environment. An insulator might, through its location between an enhancer and promoter, block the action of an enhancer on a promoter. Therefore the enhancer activates the promoter for genes on an unblocked side but is unable to activate gene expression on the blocked side, where the insulator is located. An insulator also might prevent advancement of condensed chromatin, thereby serving as a barrier for a protected gene.

LINKAGE DISEQUILIBRIUM: Where alleles (DNA markers) on different loci occur together more often than can be accounted for by chance

because of their physical proximity on a chromosome. This is a population-level characteristic that is a property of 2 loci and not their alleles.

LOCI: Plural for locus. The physical site or location of a specific gene on a chromosome. When more than 1 DNA sequence can occur at a particular locus, the sequences are called alleles. Typically, 1 allele is common and has a "normal" sequence (major allele), and other alleles (mutations) are rare (minor alleles).

MULTIPLE TESTING CORRECTION: A method to correct for random events that falsely appear significant. Bonferroni adjustment is the most common multiple testing correction and involves adjustment of the *P* value based on the number of separate tests performed (*n*). In other words, the Bonferroni adjustment deems a score significant only if the corresponding *P* value is α/n or less. The α value represents the probability of type I error or the significance level of a test.

SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 6 (STAT6): A DNA-binding transcription factor that can bind to phosphorylated cytokine receptors through their SH2 domain. Janus kinases phosphorylate STATs, which then dimerize and translocate to the nucleus. IL-4 and IL-13 activate STAT6.

THYMIC STROMAL LYMPHOPHOETIN (TSLP): A cytokine secreted by gastrointestinal, skin, and lung epithelial cells that promotes T_H2 differentiation through its effect on tissue dendritic cells and innate lymphoid cells. Its name arises from its initial identification in thymic stromal cells.

TYPE 1 ERROR: The error of rejecting a null hypothesis when it is actually true. In other words, a difference is observed when in fact there is none.

TABLE I. Reported EoE risk loci

EoE risk locus	Tag genetic variant	PMID	Genes at and near risk variants	Risk allele frequency	P value	Odds ratio
1p13.3	rs2000260	25017104	<i>SLC25A24</i>	0.57	7×10^{-7}	1.32
1p36.13	rs28530674	25017104	<i>KIF17</i>	0.04	3×10^{-7}	1.83
	rs2296225	25017104		0.08	1×10^{-7}	1.63
1p32.2	rs11206830	25017104	<i>AC119674.2</i>	0.02	8×10^{-8}	2.16
2p23.1	rs149864795	25407941	<i>CAPN14</i>	0.052	5×10^{-10}	2.22
	rs77569859	25017104		0.05	3×10^{-10}	1.98
3q26.32	rs6799767	20208534		0.58	4×10^{-7}	1.49
4q21.1	rs13106227	20208534	<i>SHROOM3</i>	0.62	4×10^{-6}	1.52
	rs1986734	20208534		0.49	1×10^{-6}	1.54
5q22.1	rs3806932	20208534	<i>WDR36, TSLP</i>	0.54	3×10^{-9}	1.85
	rs3806933	25017104		0.56	2×10^{-8}	1.37
	rs252716	25407941		0.447	4×10^{-14}	1.52
5q23.1	rs2055376	25017104	<i>SEMA6A</i>	0.02	7×10^{-8}	2.30
5q14.2	rs1032757	20208534		0.07	2×10^{-6}	1.96
6p11.2	rs9500256	20208534	<i>AL445250.1</i>	0.58	5×10^{-6}	2.04
8p23.1	rs2898261	25017104	<i>XKR6</i>	0.58	5×10^{-8}	1.35
8q24.12	rs11989782	20208534	<i>SNTB1</i>	0.23	7×10^{-6}	1.53
8q22.2	rs13278732	20208534	<i>ERICH5</i>	0.27	6×10^{-6}	1.31
10p12.31	rs11819199	25017104	<i>MIR4675</i>	0.06	3×10^{-7}	1.62
10q23.1	rs2224865	20208534	<i>MARK2P15-LINC02650</i>	0.31	9×10^{-6}	1.44
11q13.5	rs61894547	25407941	<i>LRRC32, EMSY, CAPN5</i>	0.043	4×10^{-11}	2.44
	rs2155219	25017104		0.51	4×10^{-7}	1.37
	rs77301713	25017104		0.02	1×10^{-7}	2.22
11q14.2	rs118086209	25017104	<i>CCDC81</i>	0.02	2×10^{-7}	2.19
11q21	rs1939875	20208534	<i>NR</i>	0.26	3×10^{-6}	1.54
12q13.3	rs167769	20208534	<i>STAT6</i>	0.37	2×10^{-6}	1.36
	rs167769	25407941		0.377	2×10^{-7}	1.35
14q12	rs8008716	25407941	<i>NOVA1</i>	0.087	7×10^{-8}	1.71
15q13.3	rs8041227	25017104	<i>LOC283710, KLF13</i>	0.72	6×10^{-10}	1.52
16p13	rs12924112	29904099	<i>CLEC16A</i>	0.301	2×10^{-9}	0.76
16q24.1	rs371915	20208534	<i>MEAK7</i>	0.87	2×10^{-8}	1.90
17q24.3	rs6501384	20208534	<i>CALM2P1-AC011990.1</i>	0.33	6×10^{-6}	1.41
17q25.3	rs3744790	25017104	<i>TIMP2, CEP295NL</i>	0.8	8×10^{-7}	1.54
18q12.1	rs7236477	20208534	<i>DSG1, DCC</i>	0.03	7×10^{-6}	2.22
	rs9956738	25017104		0.01	4×10^{-7}	2.47
19q13.11	rs3815700	25407941	<i>ANKRD27</i>	0.14	2×10^{-9}	1.62
21q22.3	rs17004598	25017104	<i>HSF2BP</i>	0.01	1×10^{-7}	2.57
22q11.21	rs2075277	25017104	<i>P2RX6</i>	0.09	9×10^{-7}	1.54

100%.⁷ In the context of EoE, if a male subject with European ancestry carries a single genetic risk variant, that increases the risk of EoE by 100% (ie, from 1:1000 to 1:500). Thus effect sizes of the currently identified EoE genetic risk variants support the model for genetic risk in which an accumulation of numerous genetic risk variants increases disease risk in the context of environmental factors.

For identification of rare variants, 2 approaches are generally used. The first involves testing whether a specific gene has an increased burden of rare mutations in patients with EoE compared with control subjects.⁸⁻¹⁰ The second approach is to evaluate whether a specific rare variant co-occurs with disease in families in which many members have the disease.¹¹⁻¹³ For both rare variant approaches, analytic filtering approaches are used to narrow down the considered variants.¹⁴ Beyond restricting variants to those that are rare, researchers often restrict to those considered to have the potential of a functional effect (either restricting to those that alter an amino acid or through bioinformatics pipelines are predicted to have a functional effect). The challenge with rare variants is that any given subject normally harbors too many rare variants with a predicted functional effect

for biologic follow-up. Focusing on variants present in multiple cases or located within genes that can be implicated in disease based on their known function can help reduce the number of variants considered.¹⁵ In addition, genes carrying rare variants associated with EoE might help elucidate the biologic processes underlying disease development.

A danger of focusing only on coding variants and genes “of biological interest” in the analysis of sequencing studies is that these approaches miss opportunities to identify new biology of noncoding variants. Similarly, focusing on genes with known function eliminates the opportunity to find new functions for genes in the pathoetiology of EoE. Although it is experimentally and logistically easier to focus on known genes that have been studied for decades in the context of allergic and epithelial pathways, there are tremendous opportunities for discovery by following the data and exploring any gene with rare variants identified in numerous families.

COMMON EoE GENETIC RISK LOCI

As a rare disease with a prevalence of approximately 1:2000 persons,¹⁶ EoE genetic studies aimed at identifying common

genetic risk variants have been limited to relatively small sample sizes. However, these studies have been informative for identifying several genome-wide significant susceptibility loci. Risk loci are identified based on single genetic variant statistical association that is replicated in 2 independent cohorts. Because genome-wide association studies (GWASs) assess variants across the genome, a **multiple testing correction** is used to reduce **type I error**. Based on a history of false-positive results that did not replicate in subsequent studies, a conservative Bonferroni correction is the standard multiple testing correction for GWASs. For genomes of subjects with European heritage, there are 1 million independently segregating **haplotypes**, and therefore a P value of 5×10^{-8} is required to reach genome-wide significance ($5 \times 10^{-8} = .05/10^6$). EoE genetic risk loci are named based on the chromosomal location and annotated with candidate genes located near each genetic risk locus (Table I).

In Table I we identify each of the 42 published EoE risk associations reported to date in which the “tag” single nucleotide polymorphism has a P value of less than 10^{-5} . Because EoE is a relatively rare disease, GWASs to date have been poorly to modestly powered to identify significant associations that are robust to a genome-wide multiple testing correction (ie, have a P value of less than 5×10^{-8}). Thus it is reasonable to consider suggestive genetic risk loci while awaiting replication in subsequent studies. From the published EoE risk loci, we can identify other genetic variants located near the reported “tag” variants that are in **linkage disequilibrium** (LD). Genetic variants in LD are inherited as a haplotype and therefore cannot be distinguished from one another through GWASs. Indeed, when considering LD, there are 543 genetic variants at these 31 independent risk loci across the genome.

As depicted in Fig 1,¹⁷ the majority of EoE genetic risk variants are located between genes (intergenic, 36.7%) or within the introns of genes (intronic, 42.4%). Although some variants do change the amino acid sequence of genes (coding, 2.2%), only 3 of the 31 risk loci harbor an amino acid changing variant. Thus most risk loci that increase the risk for EoE are outside of the coding regions of genes (Fig 1), suggesting a key role for genotype-dependent gene regulation in patients with EoE, which is consistent with most other complex diseases.

Among 3 reported GWASs, genetic variants at 4 loci have been consistently found at genome-wide significance (5q22 [*TSLP/WDR36*], 2p23 [*CAPN14*], 11q13 [*LRRC32/C11orf30*], and 12q13 [*STAT6*]).^{18–20} In single studies genome-wide significant associations have also been found at 19q13 (*ANKRD27*) and 16p13 (*CLEC16A*).²¹ Furthermore, an interplay of gene-gene (particularly between *IL4* and *TSLP*) and gene-environment interactions have been shown to be contributory.^{5,22,23} In a phenome-wide association study approach, a set of genetic variants are assessed in the context of many phenotypes. One such approach using a set of candidate variants identified *PTEN*, *TGFBR1/TGFBR2/PBN*, and *IL5/IL13* EoE genetic risk loci.²⁴

Only 5 EoE risk loci include a coding variant in the LD-expanded set of genetic risk variants. Because the majority of the common genetic risk loci for EoE that have been discovered to date have a genetic risk mechanism that results from genotype-dependent transcriptional regulation, further experimental assessment is needed to identify the biological mechanism driving EoE risk. Noncoding variants are found in gene promoters, introns, and regions of the genome with

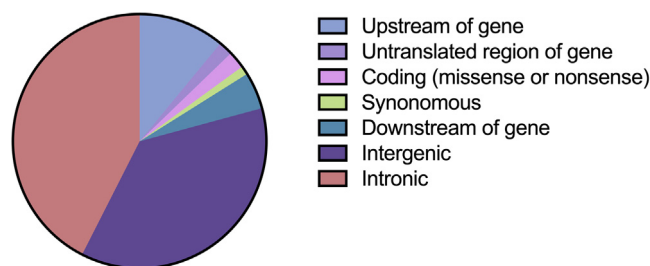


FIG 1. Annotation of genetic variants associated with EoE. All genetic variants in high LD with reported tag single nucleotide polymorphisms at EoE risk loci (see Table I) were identified by using European and European American haplotypes in the 1000 Genomes Project.¹⁷ Five hundred forty-four variants across the 31 independent risk loci were identified. The variants were annotated based on their position within the genome and in the context of genes.

regulatory activity (eg, enhancers, silencers, or **insulators**). These variants are hypothesized to affect gene expression through differential transcription factor binding and/or chromatin-based effects on gene regulation. Expression quantitative trait loci, cases in which the alleles of a given variant can be connected to genotype-dependent expression of a particular gene, have been identified at multiple EoE risk loci.

Sometimes, as in the case of the *CAPN14* EoE genetic risk locus, the expression quantitative trait locus is found specifically in patients with disease (Fig 2).¹⁸ EoE genetic risk variants in the promoter of *CAPN14* were associated with genotype-dependent expression of *CAPN14* in the esophageal biopsy specimens of patients with EoE. Miller et al²⁵ subsequently identified an EoE genetic risk variant in the *CAPN14* promoter that changed the promoter activity of *CAPN14* at a variant with genotype-dependent transcription factor binding.¹⁸ Similarly, variants at the *TSLP/WDR36* locus are associated with genotype-dependent expression of both *TSLP* and *WDR36*.²⁶ Taken together, these results point to an important but largely not understood role for genotype-dependent gene-regulatory mechanisms in the cause of EoE.

RARE EoE GENETIC RISK LOCI

The study of rare EoE genetic risk loci is an emerging field. Rochman et al²⁷ identified rare variants within 117 genes that exhibited altered expression in patients with EoE. Of the 33 patients examined, 25 harbored at least 1 rare variant that altered amino acid sequence. However, this study did not evaluate whether subjects without EoE would be expected to have similar variant burden in these genes.

In contrast, Sherrill et al²⁸ used a whole-exome approach to discovery. In this work Sherrill et al sequenced the coding regions of the genome (aka the exome) of 25 trios (probands along with parents without disease) and 12 multiplex families (including 38 patients with EoE and 8 family members without EoE). After initial filtering to identify high-quality variants that were rare and predicted to alter protein function, more than 10,000 variants were identified. Thus Sherrill et al focused on a single family with a mother and 3 offspring who have each received a diagnosis of EoE, identifying 90 rare protein coding variants that were present in the mother and her 3 offspring with the EoE diagnosis. Among these was a variant that caused a premature termination of the DHKTD1 protein (ie, a nonsense mutation). Importantly,

TABLE II. Mendelian diseases associated with EoE

Mendelian disease associated with EoE	Genetic mutation	Plausible etiologic mechanism
Loeys-Dietz syndrome (LDS)	Mutations in TGF- β receptors 1 and 2 (<i>TGFBR1</i> and <i>TGFBR2</i> , respectively)	Enhanced TGF- β signaling
Ehlers-Danlos syndrome, hypermobility type	Unknown; other subtypes of Ehlers-Danlos syndrome are caused by mutations in collagen genes.	Increased activity of TGF- β caused by altered binding by extracellular matrix
Severe atopy syndrome associated with metabolic wasting (SAM syndrome)	Homozygous mutations in desmoglein 1 (<i>DSG1</i>)	Disrupted epithelial barrier
Netherton syndrome	Loss-of-function mutations in skin protease inhibitor, Kazal-type 5 (<i>SPINK5</i>)	Unrestricted protease activity of kallikrein (KLK) 5 and KLK7
PTEN hamartoma tumor syndrome (PHTS)	Mutations in phosphatase and tensin homolog (<i>PTEN</i>)	Inhibited regulation of the phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) signaling pathway
Autosomal dominant hyper-IgE syndrome	Deleterious mutations in signal transducer and activator of transcription 3 (<i>STAT3</i>)	Dysregulated response to IL-6 and possibly IL-5
Autosomal recessive form of hyper-IgE syndrome	Loss-of-function mutations in dedicator of cytokinesis 8 (<i>DOCK8</i>)	Loss of T-cell homeostasis; lack of durable secondary antibody response against specific antigens
Erbin deficiency	Loss-of-function mutation in ERBB2-interacting protein (<i>ERBIN</i>)	Increased TGF- β pathway activation in T cells with increased T _H 2 responses

Adapted from Kottyan and Rothenberg²⁹ and O'shea et al.³⁰

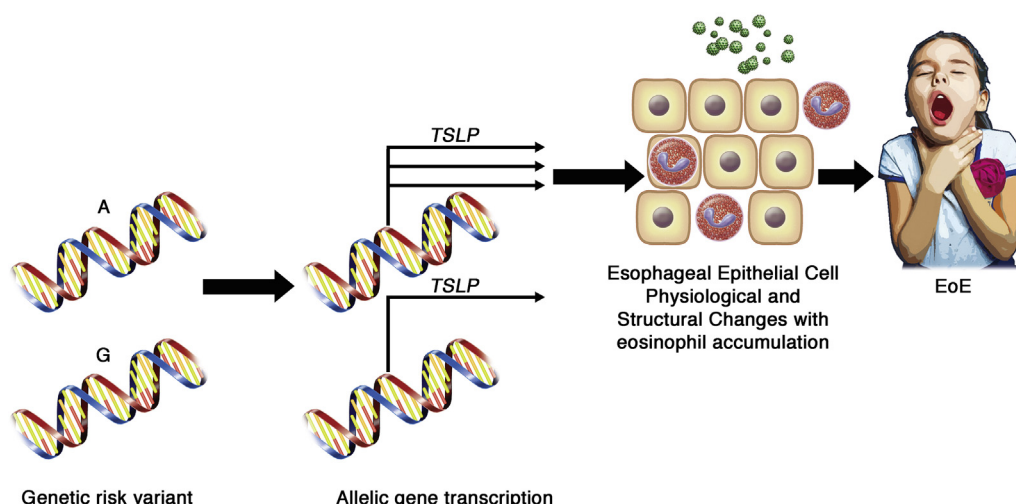


FIG 2. Genetic risk of EoE. Genetic risk variants are polymorphic places in the genome in which a risk allele is enriched in groups of patients with EoE compared with healthy control subjects. These variants are largely located in the noncoding region of the genome and do not affect amino acid use in proteins. Often, the way these noncoding variants increase the risk of EoE is through changing DNA regulatory activity, leading to genotype-dependent expression of a gene. In some cases these genes with allele-dependent expression (eg, *TSLP*) affect the physiology of epithelial cells in the esophagus. These physiologic changes can lead to EoE in patients.

additional variants inside of *DHKTDL* were also present in other families. Analysis at the gene level demonstrated that there was an enrichment in rare protein-coding variant results in *DHKTDL* in patients with EoE to a greater extent than would have been expected by chance. The results of this study highlight the large number of rare protein-coding variants identified by using exome analyses. However, this study also highlights the value of multiplex families to reduce the number of variants to a tractable number. Furthermore, although particular rare variants might not be shared between families, rare variants identified by using multiplex families can identify candidate genes for further evaluation by using genetic burden tests.

MENDELIAN DISEASES ASSOCIATED WITH EoE

For a small percentage of patients, a Mendelian disease that co-occurs with EoE can be accounted for by a single genetic “causal” variant (Table II).^{29,30} As previously reviewed,^{29,30} a larger percentage of patients with a connective tissue disease diagnosis have EoE than would be expected in a population of persons without the co-occurring disease. Patients with codiagnoses of Mendelian diseases and EoE also likely have additional common, non-Mendelian genetic alleles that modify the pathoetiology of EoE. These common disease modifiers are yet to be discovered and might overlap or be distinct to those genetic disease risk alleles that increase EoE risk in patients without these co-occurring diseases.

ONLY THE TIP OF THE ICEBERG: VARIATION EXPLAINED BY KNOWN EoE GENETIC LOCI IS ONLY A FRACTION OF WHAT IS EXPECTED

Family-based studies can be used to estimate the heritability of a disease (ie, the proportion of variation explained by genetics). The estimated genetic component to disease risk is larger than can be accounted for by genetic risk loci identified to date. There are several possible reasons for this.

First, GWASs are only powered to detect moderate-to-modest effect sizes, and thus they will not be able to detect subtle effects. The number of variants with subtle effects is likely to be substantial, perhaps numbering in the thousands. Although the effect of any one of these variants is miniscule, the cumulative effect could be substantial.

A second explanation is that GWASs of EoE have been relatively underpowered to identify and replicate all of the genetic risk loci with moderate effect sizes.

A third explanation is that genetic variants do not act in isolation. Rather, genomes, cells, organs, and bodies are intricate machines, with cross-talk and redundant systems. As a result, a single genetic risk variant in isolation might tell only a small part of the story. Thus it is critical to study genes in the proper context.

Martin et al²³ found that the strength of association between *TSLP* genetic variation and EoE was highly dependent on genetic variation at an additional locus, *IL4*. When subjects carried EoE genetic risk variants at both the *TSLP* and *IL4* loci, the odds ratio of EoE increased to 3.7 compared with 1.3 and 1.6 when either the EoE genetic risk variants at *TSLP* or *IL4* were carried in isolation, respectively. Because the *IL4* locus carries increased risk for multiple allergic diseases,³¹⁻³⁴ it was speculated that an underlying atopic phenotype might magnify the effects of *TSLP*. Likewise, Azouz et al³⁵ searched for interactions between EoE genetic risk variants at *TSLP* and those near other atopy-related genes, identifying a novel interaction between genetic variants near *TSLP* and *PLAU*. Having both risk variants resulted in a 2.7 odds for EoE compared with 1.6 for *TSLP* variants in the absence of *PLAU* variants and 1.3 odds for *PLAU* variants in the absence of *TSLP* variants. However, it is important to note that the context that might modify the association strength is likely not strictly genetic because environmental factors will likely also be critical for the disease association of many genetic risk loci.

The increased risk of EoE associated with some genetic loci might change as a function of specific environmental factors. Jensen et al²² explored whether the effects of EoE GWAS loci were conditional on early-life factors. Although the sample size on this study was small, an interaction between variants at *CAPN14* and breast-feeding was identified; the minor allele of the strongest *CAPN14* genetic risk variant increased the risk of EoE (odds ratio = 5.4) when the child had not been breast-fed compared with a decreased risk associated with that variant when the child had been breast-fed (odds ratio = 0.43).

GENETIC STUDIES OF OTHER EOSINOPHILIC GASTROINTESTINAL DISEASES AND NON-EUROPEAN COHORTS

To date, studies seeking to identify genetic risk variants for eosinophilic gastrointestinal diseases have been limited to patients with EoE who are of European ancestry.³⁶ Other

eosinophilic gastrointestinal diseases, including eosinophilic gastritis, eosinophilic duodenitis, and eosinophilic enterocolitis, are rarer than EoE. Thus identifying sufficient patients for genome-wide studies is challenging. Likewise, most of the patients that have been included in research-based tissue repositories have been of European ancestry. National Institutes of Health-funded studies are currently focused on increasing the number of patients in both of these categories, and forthcoming studies of the genetics in these important populations of patients will be critical for the understanding of disease cause.

CLINICAL APPLICATIONS OF EoE GENETIC RISK LOCI

With completion of the human genome, it was predicted that genomics would transform medicine. Although technological revolutions have enabled research to move forward much more rapidly, we are not yet positioned to use genetics for early detection or selection of optimal treatments for the vast majority of patients with an EoE diagnosis. As noted, most identified variants explain only a modest portion of disease risk. Genetic risk loci have already increased our understanding of the molecular pathophysiology of EoE through identification of critical genes and pathways. For example, identification of the *CAPN14* locus highlighted the role of this and other proteases in esophageal epithelial remodeling in patients with EoE.^{18,25,37-39} In addition to a clear disease role for esophageal epithelial cells, data support a role for immune cells, including T cells, eosinophils, and mast cells.^{30,37,40-44}

The cumulative genetic burden of identified EoE risk variants can be assessed by summing a person's risk alleles weighted by effect size to translate the genetic risk at individual loci into a patient-based assessment of genetic risk.^{45,46} These "polygenic risk scores" present an attractive opportunity to use genetic risk variants to identify those patients at the highest chance for the development of EoE. To date, these types of studies are of limited clinical usefulness for EoE, a disease with few genetic loci and a strong environmental component. Future studies will focus on the combination of genetic, transcriptomic, and environmental factors to predict those at the greatest probability for development of EoE.

CONCLUSIONS

Since the initial studies of EoE genetics nearly a decade ago, much progress has been made in the identification and understanding of genetic loci that increase the risk of EoE, as previously reviewed.^{29,30} To date, the data support a model in which genetic risk variants affect gene expression, leading to structural and physiologic changes in epithelial and immune cell function; these changes are hypothesized to lead to EoE in the context of other environment-mediated molecular and physiologic changes. In some cases rare genetic variation with large effect sizes could mediate EoE and explain multigenerational incidence in families. Future studies will focus on larger cohorts, patients of non-European ancestry, and the biological mechanisms that drive genotype-dependent changes in the development of EoE.

REFERENCES

1. 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;128:3-22.e6.

2. Alexander ES, Martin LJ, Collins MH, Kottyan LC, Sucharew H, He H, et al. Twin and family studies reveal strong environmental and weaker genetic cues explaining heritability of eosinophilic esophagitis. *J Allergy Clin Immunol* 2014;134:1084-92.e1.
3. Allen-Brady K, Firszt R, Fang JC, Wong J, Smith KR, Peterson KA. Population-based familial aggregation of eosinophilic esophagitis suggests a genetic contribution. *J Allergy Clin Immunol* 2017;140:1138-43.
4. Jensen ET, Dellon ES. Environmental factors and eosinophilic esophagitis. *J Allergy Clin Immunol* 2018;142:32-40.
5. Jensen ET, Kuhl JT, Martin LJ, Rothenberg ME, Dellon ES. Prenatal, intrapartum, and postnatal factors are associated with pediatric eosinophilic esophagitis. *J Allergy Clin Immunol* 2018;141:214-22.
6. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91-9.
7. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012;90:7-24.
8. Russo A, Di Gaetano C, Cugliari G, Matullo G. Advances in the genetics of hypertension: the effect of rare variants. *Int J Mol Sci* 2018;19.
9. Wagner MJ. Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits. *Pharmacogenomics* 2013;14:413-24.
10. Weissenkamp JD, Jiang Y, Eckert S, Jiang B, Li B, Liu DJ. Methods for the analysis and interpretation for rare variants associated with complex traits. *Curr Protoc Hum Genet* 2019;101:e83.
11. Toma C, Shaw AD, Allcock RJN, Heath A, Pierce KD, Mitchell PB, et al. An examination of multiple classes of rare variants in extended families with bipolar disorder. *Transl Psychiatry* 2018;8:65.
12. Peng G, Fan Y, Palculict TB, Shen P, Ruteshouser EC, Chi AK, et al. Rare variant detection using family-based sequencing analysis. *Proc Natl Acad Sci U S A* 2013;110:3985-90.
13. Kohli MA, Cukier HN, Hamilton-Nelson KL, Rolati S, Kunkle BW, Whitehead PL, et al. Segregation of a rare TTC3 variant in an extended family with late-onset Alzheimer disease. *Neurol Genet* 2016;2:e41.
14. Jalali Sefid Dashti M, Gamielidien J. A practical guide to filtering and prioritizing genetic variants. *Biotechniques* 2017;62:18-30.
15. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790-3.
16. Mansoor E, Cooper GS. The 2010-2015 Prevalence of eosinophilic esophagitis in the USA: a population-based study. *Dig Dis Sci* 2016;61:2928-34.
17. Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.
18. Kottyan LC, Davis BP, Sherrill JD, Liu K, Rochman M, Kaufman K, et al. Genome-wide association analysis of eosinophilic esophagitis provides insight into the tissue specificity of this allergic disease. *Nat Genet* 2014;46:895-900.
19. Sleiman PM, Wang ML, Cianferoni A, Aceves S, Gonsalves N, Nadeau K, et al. GWAS identifies four novel eosinophilic esophagitis loci. *Nat Commun* 2014;5:5593.
20. Rothenberg ME, Spergel JM, Sherrill JD, Annaiah K, Martin LJ, Cianferoni A, et al. Common variants at 5q22 associate with pediatric eosinophilic esophagitis. *Nat Genet* 2010;42:289-91.
21. Kottyan LC, Maddox A, Braxton JR, Stucke EM, Mukkada V, Putnam PE, et al. Genetic variants at the 16p13 locus confer risk for eosinophilic esophagitis. *Genes Immun* 2019;20:281-92.
22. Jensen ET, Kuhl JT, Martin LJ, Langefeld CD, Dellon ES, Rothenberg ME. Early-life environmental exposures interact with genetic susceptibility variants in pediatric patients with eosinophilic esophagitis. *J Allergy Clin Immunol* 2018;141:632-7.e5.
23. Martin LJ, He H, Collins MH, Abonia JP, Biagini Myers JM, Eby M, et al. Eosinophilic esophagitis (EoE) genetic susceptibility is mediated by synergistic interactions between EoE-specific and general atopic disease loci. *J Allergy Clin Immunol* 2018;141:1690-8.
24. Namjou B, Marsolo K, Carroll RJ, Denny JC, Ritchie MD, Verma SS, et al. Phenome-wide association study (PheWAS) in EMR-linked pediatric cohorts, genetically links PLCL1 to speech language development and IL5-IL13 to eosinophilic esophagitis. *Front Genet* 2014;5:401.
25. Miller DE, Forney C, Rochman M, Cranert S, Habel J, Rymer J, et al. Genetic, inflammatory, and epithelial cell differentiation factors control expression of human calpain-14. *G3 (Bethesda)* 2019;9:729-36.
26. Sherrill JD, Gao PS, Stucke EM, Blanchard C, Collins MH, Putnam PE, et al. Variants of thymic stromal lymphopoietin and its receptor associate with eosinophilic esophagitis. *J Allergy Clin Immunol* 2010;126:160-5.e3.
27. Rochman M, Travers J, Miracle CE, Bedard MC, Wen T, Azouz NP, et al. Profound loss of esophageal tissue differentiation in patients with eosinophilic esophagitis. *J Allergy Clin Immunol* 2017;140:738-49.e3.
28. Sherrill JD, Kc K, Wang X, Wen T, Chamberlin A, Stucke EM, et al. Whole-exome sequencing uncovers oxidoreductases DHTKD1 and OGDHL as linkers between mitochondrial dysfunction and eosinophilic esophagitis. *JCI Insight* 2018;3.
29. Kottyan LC, Rothenberg ME. Genetics of eosinophilic esophagitis. *Mucosal Immunol* 2017;10:580-8.
30. O'Shea KM, Aceves SS, Dellon ES, Gupta SK, Spergel JM, Furuta GT, et al. Pathophysiology of eosinophilic esophagitis. *Gastroenterology* 2018;154:333-45.
31. He JQ, Chan-Yeung M, Becker AB, Dimich-Ward H, Ferguson AC, Manfreda J, et al. Genetic variants of the IL13 and IL4 genes and atopic diseases in at-risk children. *Genes Immun* 2003;4:385-9.
32. de Guia RM, Ramos JD. The -590C/TIL4 single-nucleotide polymorphism as a genetic factor of atopic allergy. *Int J Mol Epidemiol Genet* 2010;1:67-73.
33. Baye TM, Butsch Kovacic M, Biagini Myers JM, Martin LJ, Lindsey M, Patterson TL, et al. Differences in candidate gene association between European ancestry and African American asthmatic children. *PLoS One* 2011;6:e16522.
34. Lu MP, Chen RX, Wang ML, Zhu XJ, Zhu LP, Yin M, et al. Association study on IL4, IL13 and IL4RA polymorphisms in mite-sensitized persistent allergic rhinitis in a Chinese population. *PLoS One* 2011;6:e27363.
35. 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.
36. Kottyan L, Spergel JM, Cianferoni A. Immunology of the ancestral differences in eosinophilic esophagitis. *Ann Allergy Asthma Immunol* 2019;122:443-4.
37. Lyles J, Rothenberg M. Role of genetics, environment, and their interactions in the pathogenesis of eosinophilic esophagitis. *Curr Opin Immunol* 2019;60:46-53.
38. Rochman M, Azouz NP, Rothenberg ME. Epithelial origin of eosinophilic esophagitis. *J Allergy Clin Immunol* 2018;142:10-23.
39. Litosh VA, Rochman M, Rymer JK, Porollo A, Kottyan LC, Rothenberg ME. Calpain-14 and its association with eosinophilic esophagitis. *J Allergy Clin Immunol* 2017;139:1762-71.e7.
40. 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;129:2014-28.
41. Schwartz JT, Morris DW, Collins MH, Rothenberg ME, Fulkerson PC. Eosinophil progenitor levels correlate with tissue pathology in pediatric eosinophilic esophagitis. *J Allergy Clin Immunol* 2019;143:1221-4.e3.
42. Martin LJ, Franciosi JP, Collins MH, Abonia JP, Lee JJ, Hommel KA, et al. Pediatric Eosinophilic Esophagitis Symptom Scores (PEESS v2.0) identify histologic and molecular correlates of the key clinical features of disease. *J Allergy Clin Immunol* 2015;135:1519-28.e8.
43. Rothenberg ME. Molecular, genetic, and cellular bases for treating eosinophilic esophagitis. *Gastroenterology* 2015;148:1143-57.
44. Azouz NP, Fukuda M, Rothenberg ME, Sagi-Eisenberg R. Investigating mast cell secretory granules; from biosynthesis to exocytosis. *J Vis Exp* 2015;95:52505.
45. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50:1219-24.
46. So HC, Sham PC. Exploring the predictive power of polygenic scores derived from genome-wide association studies: a study of 10 complex traits. *Bioinformatics* 2017;33:886-92.