

Common variants at 5q22 associate with pediatric eosinophilic esophagitis

Marc E Rothenberg^{1,11}, Jonathan M Spergel^{2,3,11}, Joseph D Sherrill^{1,11}, Kiran Annaiah^{4,11}, Lisa J Martin^{5,11}, Antonella Cianferoni^{2,3}, Laura Gober², Cecilia Kim⁴, Joseph Glessner⁴, Edward Frackelton⁴, Kelly Thomas⁴, Carine Blanchard¹, Chris Liacouras^{3,6}, Ritu Verma^{3,6}, Seema Aceves⁷, Margaret H Collins⁸, Terri Brown-Whitehorn^{2,3}, Phil E Putnam⁹, James P Franciosi⁹, Rosetta M Chiavacci⁴, Struan F A Grant^{3,4,10}, J Pablo Abonia¹, Patrick M A Sleiman⁴ & Hakon Hakonarson^{3,4,10}

Eosinophilic esophagitis (EoE) is an allergic disorder characterized by the accumulation of eosinophils in the esophagus. We report association of EoE with variants at chromosome 5q22 encompassing *TSLP* and *WDR36* (rs3806932, combined $P = 3.19 \times 10^{-9}$). *TSLP* is overexpressed in esophageal biopsies from individuals with EoE compared with unaffected individuals, whereas *WDR36* expression is unaltered between the two groups. These data implicate the 5q22 locus in the pathogenesis of EoE and identify *TSLP* as the most likely candidate gene in the region.

EoE is a global health condition now reported on all continents except Africa, with an incidence of ~1 per 10,000 people¹. Symptoms of EoE include difficulty feeding, failure to thrive, vomiting, epigastric or chest pain, dysphagia and food impaction. Individuals with EoE (cases) are predominantly young males with a high rate of atopic disease, and the diagnosis of EoE is made by endoscopy and biopsy findings of isolated eosinophils in the esophagus¹.

Based on several findings, EoE is considered a food allergy-related disorder. The majority of EoE cases are atopic individuals with a high rate of food allergen sensitization as determined by skin prick and patch testing; in addition, EoE cases have a higher rate of food anaphylaxis compared with the general population¹. Moreover, nearly all EoE cases have complete disease remission following introduction of an elemental formula diet that lacks all allergens; conversely, the disease flares upon reintroduction of specific foods. Unlike the situation in classic anaphylaxis, which typically involves a limited set of

foods, in EoE affected individuals are often sensitized to a myriad of foods, including food groups not typically associated with eliciting anaphylaxis². Experimental modeling in mice has demonstrated a key role for adaptive immunity and Th2-cell cytokines (especially IL-5 and IL-13) in the disease process and a strong connection between allergic sensitization and inflammation in the respiratory tract and skin³.

Accumulating evidence suggests that EoE has a strong familial association⁴. Nearly 10% of parents of EoE cases have a history of esophageal strictures, and ~8% have biopsy-proven EoE⁴; in addition, at least 27 multiplex EoE families have been described to date⁵, wherein multiple family members are affected. EoE also exhibits a high sibling risk ratio (λ_s) of ~80 compared with related atopic diseases such as asthma ($\lambda_s \sim 2$)⁶. Although genetics is likely to have a large role in EoE susceptibility, there has been only one candidate gene identified to date, *CCL26* (encoding eotaxin-3), which is the most highly overexpressed gene in the esophagus of EoE cases based on genome-wide expression profiling⁷. However, the disease-associated allele is only present in 14% of EoE cases⁷, suggesting that additional risk variants are yet to be found.

We report the results of an ongoing genome-wide association study (GWAS) aimed at identifying susceptibility loci for EoE. We genotyped approximately 550,000 SNPs on the Illumina HumanHap550 BeadChip in a discovery cohort of 181 EoE cases from Cincinnati Children's Hospital Medical Center (CCHMC) and 1,974 unaffected individuals (controls) from the Children's Hospital of Philadelphia (CHOP), and in a replication cohort of 170 EoE cases and 1,130 controls from CHOP (see **Supplementary Methods** for quality-control measures and detailed descriptions of cases and controls). The study was approved by the institutional review boards of CHOP and CCHMC. Parental informed consent was obtained from all participants in this study for the purpose of DNA collection and genotyping.

Both cases and controls were chosen based on their self-reported ancestry; however, to avoid any population stratification, we screened all samples at ancestry informative markers using the STRUCTURE package (**Supplementary Methods** and URLs). Cases were subsequently genetically matched to controls by principal component analysis (**Supplementary Methods**). Single marker analyses for genome-wide data were carried out using the Cochran-Armitage trend test as implemented in PLINK.

In the discovery cohort, two SNPs remained significantly associated with EoE following correction for multiple testing (**Supplementary Fig. 1a**). One of the two SNPs, rs3806932 (uncorrected $P = 8.81 \times 10^{-8}$,

¹Division of Allergy and Immunology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio, USA.

²Division of Allergy and Immunology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ³Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. ⁴The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ⁵Division of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio, USA. ⁶Division of Gastroenterology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ⁷Division of Allergy and Immunology, Riley Hospital for Children, University of California at San Diego, San Diego, California, USA. ⁸Division of Pathology and ⁹Division of Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio, USA. ¹⁰Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ¹¹These authors contributed equally to this work. Correspondence should be addressed to H.H. (hakonarson@chop.edu).

Received 22 September 2009; accepted 11 February 2010; published online 7 March 2010; doi:10.1038/ng.547

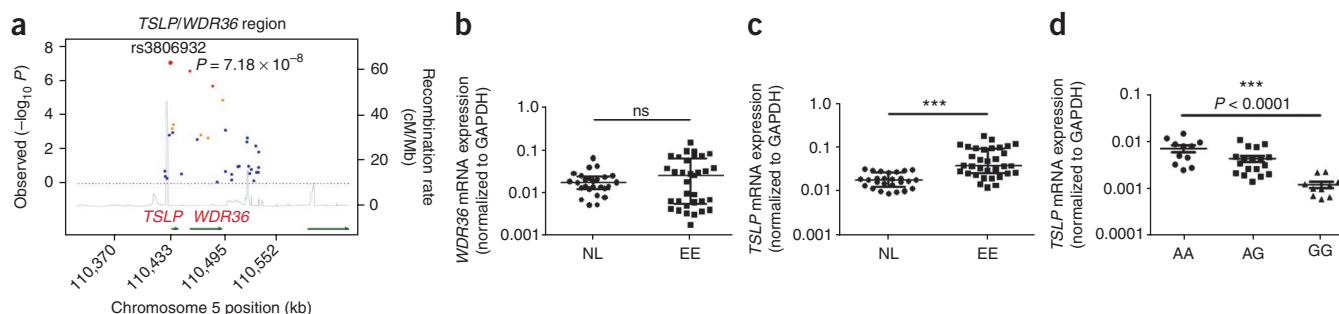


Figure 1 Regional plot of the 5q22 locus and follow-up expression studies. **(a)** Regional plot showing the associated SNPs in the *TSLP*-*WDR36* region. The relative position of each SNP is plotted against $-\log_{10} P$ values and the recombination rate in the region. **(b,c)** *TSLP* mRNA is increased in the esophagus in EoE cases. Quantitative PCR on esophageal biopsies from 26 unaffected individuals (NL) and 33 EoE cases show that mRNA levels of *TSLP* are significantly increased in EoE (**c**), but that *WDR36* (**b**) is not. Data are normalized to *GAPDH* mRNA levels. ***, $P < 0.0001$; ns, not significant. **(d)** *TSLP* expression level stratified by rs3806932 genotype. *TSLP* expression is decreased in individuals homozygous for the protective minor allele (G). Error bars in **b–d**, interquartile range.

minor allele frequency of 31.2% in cases and 45.8% in controls, OR = 0.53, 95% CI 0.42–0.67), maps to a linkage disequilibrium block on 5q22.1 that contains 11 other SNPs that were also associated with EoE (P value range of 4.2×10^{-3} to 4.7×10^{-7} ; OR range = 0.55–1.55) (Fig. 1a and Supplementary Table 1). The associated block spans the *TSLP*-*WDR36* gene region (pairwise r^2 values shown in Supplementary Fig. 2; averaged recombination rates from the CEU HapMap trios shown in Fig. 1).

We next sought to replicate our findings in the CHOP EoE cohort. Although none of the SNPs surpassed genome-wide significance (Supplementary Fig. 1b), of the 12 EoE-associated SNPs at the 5q22.1 locus in the discovery cohort, 4 were also associated with EoE in the replication cohort (P value range of 0.05 to 4.8×10^{-3} ; OR range = 0.70–1.3) (Table 1). Combining the P values across the two cohorts using Fisher's method, three SNPs surpassed genome-wide significance (P value range of 2.37×10^{-8} to 3.19×10^{-9}). The most significant SNP in the discovery cohort, rs3806932, was also significantly associated in the replication cohort ($P = 8 \times 10^{-3}$; minor allele frequency of 37.7% in cases and 45.2% in controls; OR = 0.73; combined P value for rs3806932 across CCHMC and CHOP cohorts = 3.19×10^{-9}).

In addition to the 5q22.1 locus, one SNP on 13q31.1, rs1873288, reached genome-wide significance in the discovery cohort ($P = 6.72 \times 10^{-8}$) but did not replicate in the CHOP cohort. We subsequently performed a genome-wide meta-analysis of the discovery and replication cohorts (Supplementary Fig. 1c). Besides the 5q22.1 locus, one SNP on chromosome 16 (rs371915, $P = 2.19 \times 10^{-8}$) also reached genome-wide significance; however, this association remains to be replicated in future studies. A summary of all the meta-analysis SNPs with $P < 1 \times 10^{-5}$ is shown in Supplementary Table 2; it is notable that the list includes SNPs that map to genes that have previously been implicated in immune function such as *STAT6* (rs167769) and *DSG1* (rs7236477).

Finally, as the proportion of EoE cases with asthma is high (~40% in the CCHMC cohort and ~20% in the CHOP cohort), we carried out a logistic regression analysis including asthma status as a covariate in the analysis of both cohorts to ensure that the result was not being driven by asthma rather than EoE. We observed consistent association at the 5q22 locus (Supplementary Table 1) with no significant heterogeneity in the odds ratios after adjusting for asthma status, which indicates that the association is independent of asthma.

We subsequently examined the expression levels of both *TSLP* and *WDR36* mRNA in esophageal biopsies from EoE cases and controls. *TSLP* was significantly overexpressed in EoE cases compared to controls following normalization with *GAPDH*, whereas *WDR36* was unaltered (Fig. 1b,c). We then investigated the expression of the two *TSLP* isoforms (Supplementary Fig. 3) in the esophageal biopsies. *TSLP* isoform 2 was significantly overexpressed in EoE cases compared to controls ($P < 0.0001$). Finally, *TSLP* expression levels were significantly correlated with rs3806932 genotype in the EoE cases; homozygote carriers of the protective minor allele showed less *TSLP* expression than heterozygotes and homozygotes for the ancestral allele (Fig. 1d and Supplementary Fig. 4). These results indicate that *TSLP* is the stronger candidate at the 5q22 locus in influencing EoE susceptibility.

Taken together, we have identified and replicated a genome-wide significant locus at 5q22.1 in EoE cases of European ancestry. Two genes map to this locus, *TSLP* and *WDR36*. *WDR36* is highly coregulated with the T-cell growth factor IL-2 and has also been linked to open-angle glaucoma⁸. *TSLP* encodes an IL-7–like cytokine that is expressed in the thymus and peripheral tissues and that regulates dendritic cell–mediated inflammatory Th2 responses⁹. As a key initiator of allergic inflammatory diseases, *TSLP* is overproduced in lesional skin in atopic dermatitis and the asthma-affected lung. In mouse models of allergic airway inflammation, enhanced Tslp production from skin keratinocytes correlates with increased Th2-attracting chemokine

Table 1 Results for the most strongly associated SNPs at 5q22

SNP	Position	Minor allele	Discovery cases MAF	Discovery controls MAF	Discovery P trend	Discovery OR	Replication cases MAF	Replication controls MAF	Replication P trend	Replication OR	Fisher's combined P
rs3806932	110,433,574	G	0.312	0.458	7.18×10^{-8}	0.54	0.377	0.452	0.008	0.73	3.19×10^{-9}
rs7723819	110,455,246	A	0.326	0.467	3.09×10^{-7}	0.55	0.382	0.464	0.004	0.71	7.67×10^{-9}
rs10051830	110,480,744	A	0.323	0.453	1.41×10^{-6}	0.58	0.371	0.454	0.003	0.71	2.37×10^{-8}
rs1043828	110,491,907	C	0.464	0.350	1.26×10^{-5}	1.61	0.418	0.346	0.010	1.36	5.99×10^{-7}

MAF, minor allele frequencies; P trend, Cochran-Armitage trend test P values; OR, odds ratio; discovery, CCHMC cohort; replication, CHOP cohort; Fisher's combined P values, both cohorts combined.

production (including the chemokines eotaxin-1 and eotaxin-2) and with enhanced disease severity and epicutaneous sensitization¹⁰. The higher expression of *TSLP* observed in EoE cases suggests a model in which *TSLP* acts as a driving force for the allergic phenotype. Indeed, individuals with EoE are highly atopic and highly sensitized to multiple food allergens, and *TSLP* plays a critical role in initiating allergen sensitization⁹, a process that is profoundly abnormal in EoE cases. In addition, IL-13 has been shown to co-induce *TSLP* in epithelial cells, and we have previously associated IL-13-driven epithelial responses with EoE pathogenesis¹¹. We hypothesize that increased *TSLP* expression as observed in the EoE cases leads to increased allergen sensitization and to development of EoE. It is noteworthy that the *TSLP* receptor (cytokine receptor-like factor 2, *CRLF2*) resides within the pseudoautosomal region 1 on the X and Y chromosomes (Xp22.3 and Yp11.3)¹², providing a plausible hypothesis to explain the higher incidence of EoE in males.

Consistent with our analysis, a recent GWAS has identified several SNPs associated with peripheral blood eosinophilia, including SNPs in the 5q22 locus¹³. EoE can associate with peripheral blood eosinophilia but only in a subset of individuals¹⁴. This suggests that EoE and peripheral eosinophilia are associated with the same locus. In conclusion, we report the first EoE susceptibility locus identified using a GWAS approach and implicate *TSLP* as a promising candidate gene underlying this association.

URLs. STRUCTURE, <http://pritch.bsd.uchicago.edu/structure.html>; PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all of the participating families and subjects. We thank B. Buckmeier, A. Ahrens, S. Jameson, M. Palazzolo, H. Foote, M. Mingler, M. Shuker and

E. Stucke for assistance with patient enrollment, DNA preparation and/or database management. This work was supported in part by US National Institutes of Health (NIH) U19 AI070235, NIH R01 DK076893, Food Allergy Project, Campaign Urging Research for Eosinophilic Disorders (CURED) Foundation, American Partnership for Eosinophilic Disorders (APFED), The Children's Hospital of Philadelphia Chair's Initiative, the Buckeye Foundation and the Coltswood Foundation. All genome-wide genotyping was funded by an Institute Development Award to the Center for Applied Genomics from the Children's Hospital of Philadelphia.

AUTHOR CONTRIBUTIONS

M.E.R., J.M.S., S.A., M.H.C., T.B.-W., P.E.P., J.P.F., S.F.A.G., J.P.A., P.M.A.S. and H.H. designed the study. J.M.S., J.D.S., L.J.M., A.C., L.G., C.K., J.G., E.F., K.T., C.B., C.L., R.V., R.M.C. and H.H. generated the data. K.A., J.D.S., L.J.M. and P.M.A.S. analyzed the data. M.E.R., J.M.S., J.D.S., P.M.A.S. and H.H. prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

1. Furuta, G.T. *et al. Gastroenterology* **133**, 1342–1363 (2007).
2. Spergel, J.M., Brown-Whitehorn, T., Beausoleil, J.L., Shuker, M. & Liacouras, C.A. *J. Allergy Clin. Immunol.* **119**, 509–511 (2007).
3. Akei, H.S., Mishra, A., Blanchard, C. & Rothenberg, M.E. *Gastroenterology* **129**, 985–994 (2005).
4. Noel, R.J., Putnam, P.E. & Rothenberg, M.E. *N. Engl. J. Med.* **351**, 940–941 (2004).
5. Collins, M.H. *et al. Clin. Gastroenterol. Hepatol.* **6**, 621–629 (2008).
6. Malerba, G. *et al. Am. J. Respir. Crit. Care Med.* **162**, 1587–1590 (2000).
7. Blanchard, C. *et al. J. Clin. Invest.* **116**, 536–547 (2006).
8. Challa, P. *Int. Ophthalmol. Clin.* **48**, 73–94 (2008).
9. Liu, Y.J. *Adv. Immunol.* **101**, 1–25 (2009).
10. Oyoshi, M.K., He, R., Kumar, L., Yoon, J. & Geha, R.S. *Adv. Immunol.* **102**, 135–226 (2009).
11. Blanchard, C. *et al. J. Allergy Clin. Immunol.* **120**, 1292–1300 (2007).
12. Tonoizuka, Y. *et al. Cytogenet. Cell Genet.* **93**, 23–25 (2001).
13. Gudbjartsson, D.F. *et al. Nat. Genet.* **41**, 342–347 (2009).
14. Konikoff, M.R. *et al. Clin. Gastroenterol. Hepatol.* **4**, 1328–1336 (2006).

Supplementary Methods

Detailed description of cases and controls

The EoE discovery cohort consisted of 181 clinically confirmed EE patients of European ancestry recruited at CCHMC. The subjects in the discovery cohort had an eosinophils/hpf (400X) count of ≥ 24 . The mean age of these cases was 11.3 ± 10.4 (SD) years. The replication cohort consisted of 170 cases recruited at CHOP. Cases were defined as those having an eosinophils/hpf (400X) count of ≥ 24 and on PPI therapy for at least 8 weeks. The mean age of the replication cohort cases was 7.8 ± 4.9 SD years. Most of the EE subjects in both the CCHMC and the CHOP cohorts were male making up 70% in the discovery cohort (CCHMC) and 75% in the replication cohort (CHOP). Moreover, 73% of the discovery cohort and 72% of the replication cohort had asthma, allergic rhinitis or atopic dermatitis.

The control samples used in both the discovery and replication phase were recruited within the CHOP Health Care Network by CHOP clinicians and nursing staff. They were all of self-reported Caucasian ancestry and screened negative for a diagnosis of asthma, dermatitis or other atopic conditions, as well as for other chronic medical conditions based on questionnaire responses. The mean age of the control group ($n=2,096$) used for the discovery phase was 8.54 ± 5.65 SD years and consisted of an equal percentage of males and females (50.4% and 50.6%, respectively). There was no overlap between patients or controls in the discovery and replication sets.

Genotyping QC measures

Quality control measures prior to association analysis resulted in the rejection of 16,942 SNPs with call rates $< 95\%$, 21,901 SNPs with a minor allele frequency (MAF) $< 1\%$ and 14,944 SNPs with Hardy Weinberg

equilibrium $P < 10^{-5}$ for the CCHMC analysis; the genomic inflation factor (GIF) for that study was 1.08. For the CHOP analysis we rejected 7,481 SNPs with call rates $< 95\%$, 18,755 SNPs with a MAF $< 1\%$ and 1,860 SNPs with Hardy Weinberg equilibrium $P < 10^{-5}$; the GIF was 1.05.

Structure analysis

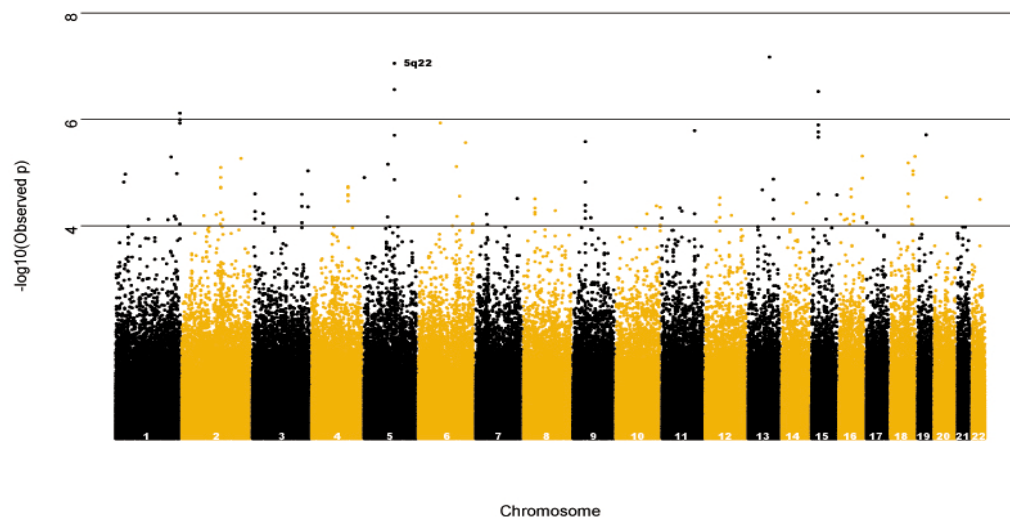
To reduce the risk of population stratification due to inaccurate self-reported ancestry all cases and controls were initially screened using the STRUCTURE package¹. We used 220 ancestry informative markers (AIMs) and spiked the test sample set with 90 CEPH, Yoruban and Chinese/Japanese individuals genotyped as part of the HapMap project to improve clustering. Samples were excluded if their inferred proportion of ancestry was less than 90% that of the CEPH cluster.

Genetic matching parameters

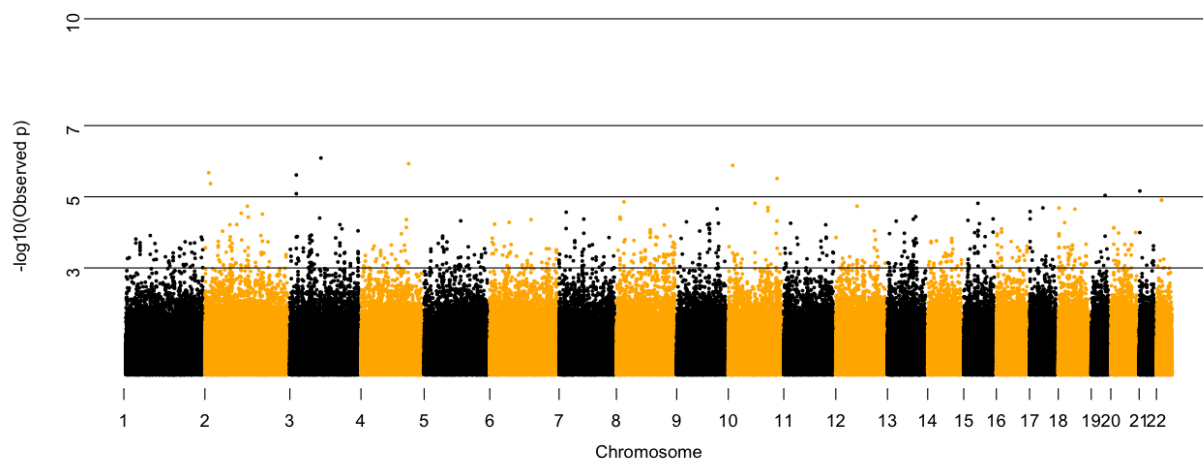
‘Genetic matching’ is a principal component analysis (PCA) based method of matching cases with controls that minimizes the effects of population stratification in GWA studies by selecting a set of optimally matched cases and controls and excluding outliers that would otherwise increase genomic inflation². We computed principal components for our dataset by running smartpca, a part of the EIGENSTRAT package, on 100,000 random autosomal SNPs and applied a matching algorithm implemented in MATLAB to the output. The matching algorithm assigns each sample a coordinate based on k eigenvalue-scaled principal components. It then matches each case to m unique controls within a distance d , keeping only cases that match exactly m controls. The distance thresholds were manually optimized for each data set to minimize GIF and maximize power (i.e. number of cases). In this study, we matched the CCHMC discovery cases to approximately 9 controls, using 3 principal components and a distance threshold of 0.05. The CHOP replication cases were matched to approximately 7 controls at a distance threshold of 0.05.

Expression Analysis: Quantitative real-time PCR analysis was performed on cDNA from distal esophageal biopsy RNA. All EoE patient biopsies showed active disease pathology at the time in which they were taken and all patients reported no steroid treatment at time of biopsy. The primers used for detection of TSLP amplified both TSLP isoforms (NM_033035 and NM_138551) were as described by Kato, *et al*³. The TSLP isoform-specific primers used were the following: isoform 1, forward (5'-GGCTGGTGTAACTTACGAC-3'), reverse (5'-ACGCCACAATCCTTGTAATTGTG-3'); isoform 2, forward (5'-CGTAAACTTTGCCGCCTATGA-3'), reverse (5'-ACGCCACAATCCTTGTAATTGTG-3'). Statistical testing for mRNA expression normalized to GAPDH was determined by Mann-Whitney U test using GraphPad Prism software.

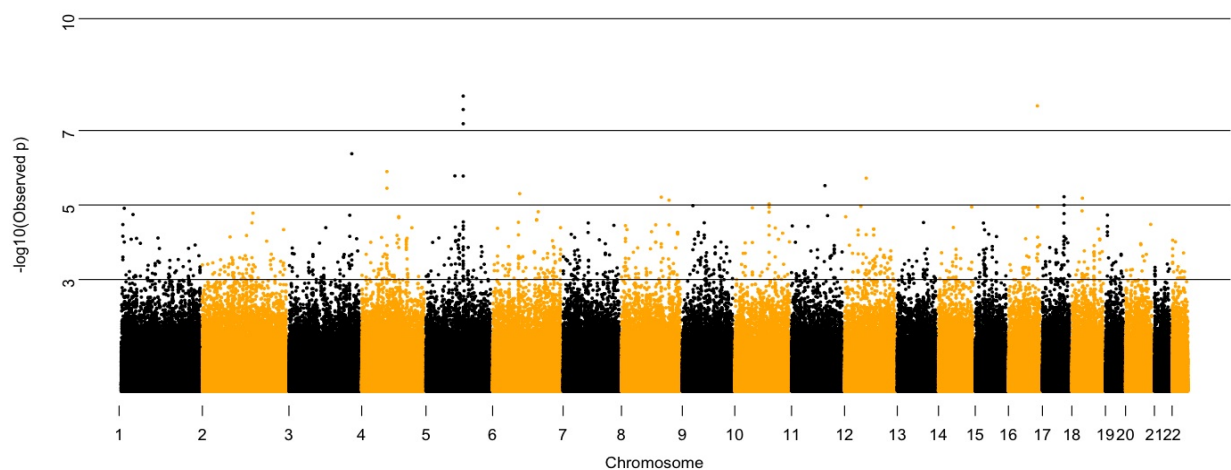
Supplementary Figure 1: Manhattan plots of A) the European ancestry EoE cases in the discovery (CCHMC) cohort. $-\log_{10} P$ -values are plotted against the physical distance. The locus at 5q22 was significantly associated with EoE following Bonferroni correction.



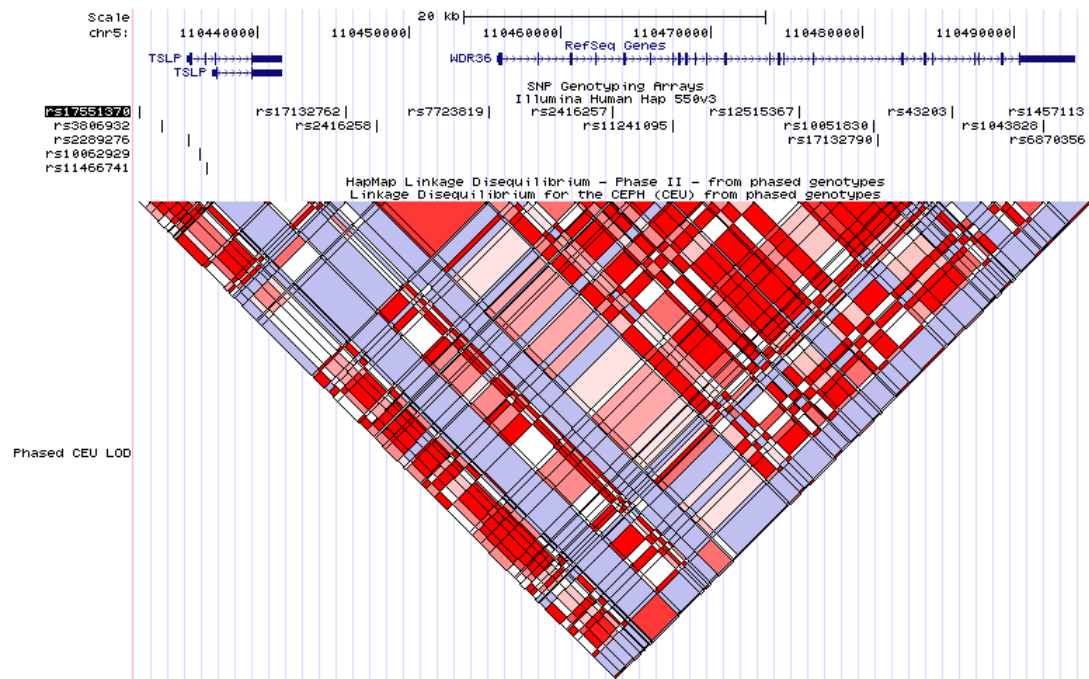
Supplementary Figure 1: Manhattan plots of B) The replication (CHOP) cohort genome wide association – $\log_{10} P$ -values plotted against physical distance.



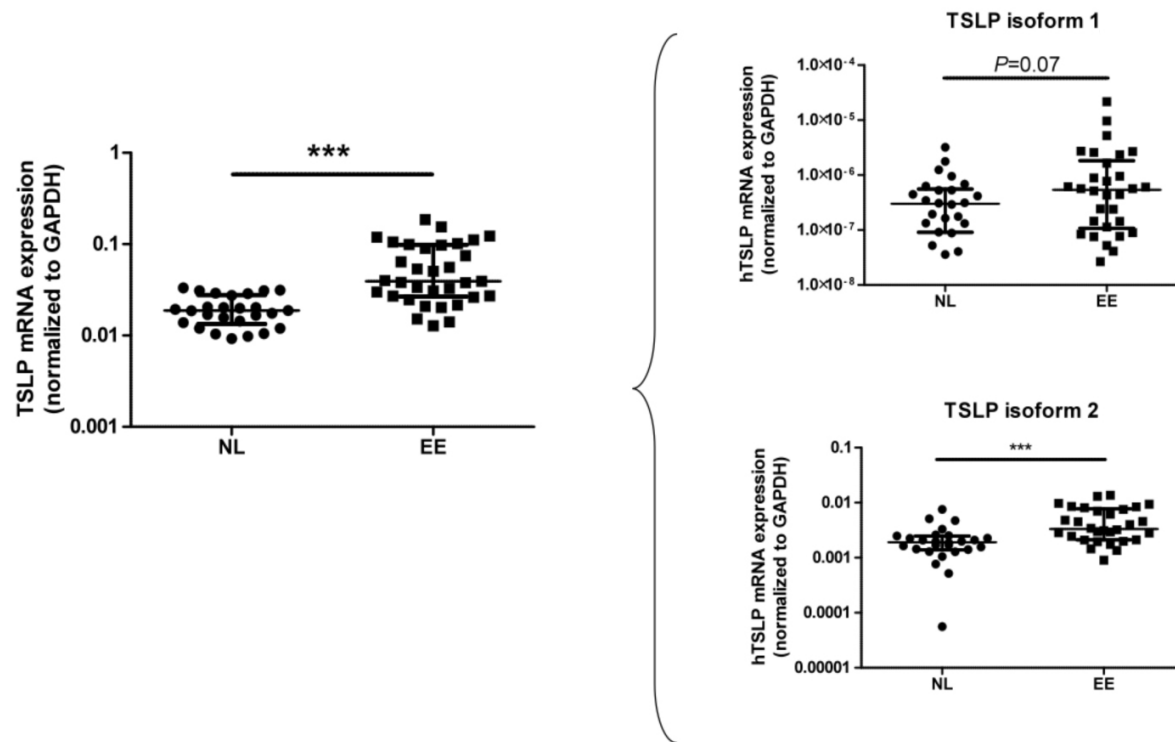
Supplementary Figure 1: Manhattan plots of **C**) The discovery (CCHMC) and replication (CHOP) cohorts meta-analysis. The $-\log_{10} P$ -values are plotted against physical distance.



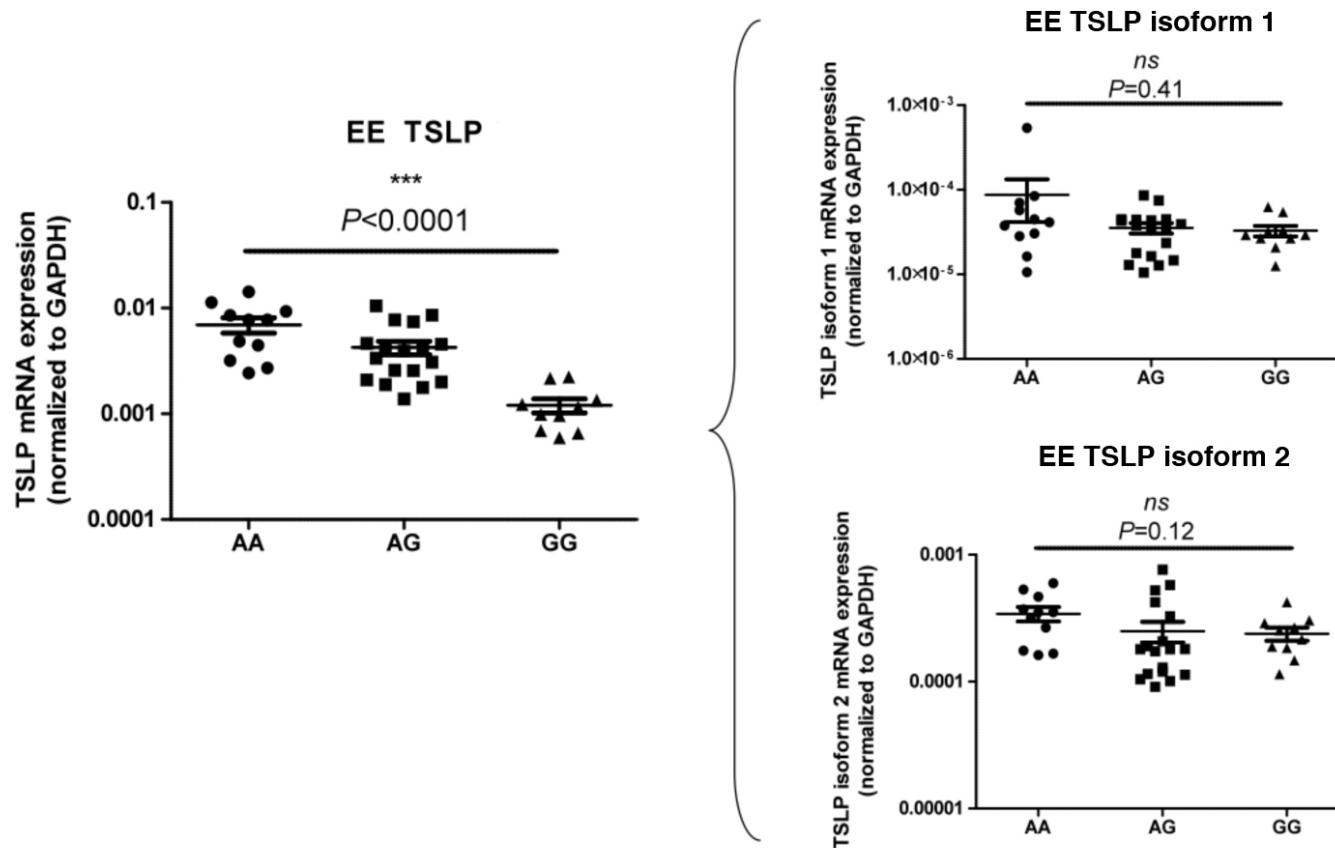
Supplementary Figure 2: Representation of the chromosome 5q22.1 associated interval. The figure shows pairwise r^2 LD values from the discovery cohort (CCHMC). The *TSLP* gene and its two isoforms and *WDR36* are drawn to scale in relation to the associated SNPs.



Supplementary Figure 3: TSLP isoform 1 and 2 overexpression in EoE esophagus. Quantitative PCR on esophageal biopsies from 26 normal (NL) and 33 EoE patients show mRNA levels of TSLP isoform 2 are significantly increased in EoE. TSLP isoform 1 shows a non-significant trend towards overexpression. Data shown are normalized to GAPDH mRNA levels (***, P -value <0.0001) EE: cases NL: controls. Error bars represent the interquartile ranges.



Supplementary Figure 4: TSLP expression level stratified by rs3806932 genotype. TSLP expression is decreased in homozygotes for the protective minor allele (G). Both TSLP isoforms showed a similar but non-significant trend in expression. Data shown are normalized to GAPDH mRNA levels (***, P -value <0.0001). Error bars represent the interquartile ranges.



Supplementary Table 1: Association results at the 5q22 locus. No covariate columns: Minor allele frequencies (MAF), Cochran-Armitage trend test *P*-values (*P*-trend) and odds ratios (OR) for the CCHMC (Disc) and CHOP (Rep) cohorts and the Fisher's combined *P*-values for the two cohorts combined. BP: base pair, A1: minor allele. **Asthma covariate columns:** *P*-values and odds ratios (OR) in the discovery (DISC) and replication (REP) cohorts following logistic regression with asthma status as a covariate. Abbreviations, BP: base pair, A1: minor allele.

SNP	BP	A1	No covariates									Asthma covariate			
			Disc Case MAF	Disc Control MAF	Disc <i>P</i> -trend	Disc OR	Rep Case MAF	Rep Control MAF	Rep <i>P</i> -trend	Rep OR	Fisher's Combined <i>P</i>	Disc <i>P</i>	Disc OR	REP <i>P</i>	REP OR
rs17551370	110432084	A	0.086	0.146	1.62x10 ⁻³	0.55	0.112	0.135	0.232	0.80	9.67x10 ⁻⁴	9.47x10 ⁻³	0.55	0.19	0.77
rs3806932	110433574	G	0.312	0.458	7.18x10 ⁻⁸	0.54	0.377	0.452	0.008	0.73	3.19x10 ⁻⁹	1.76x10 ⁻⁵	0.54	8.47x10 ⁻³	0.71
rs2289276	110435406	T	0.218	0.304	5.59x10 ⁻⁴	0.64	0.262	0.302	0.118	0.82	1.98x10 ⁻⁴	3.23x10 ⁻³	0.62	0.13	0.80
rs10062929	110436078	A	0.086	0.149	1.12x10 ⁻³	0.54	0.112	0.137	0.208	0.79	6.25x10 ⁻⁴	0.01	0.56	0.17	0.76
rs11466741	110436604	T	0.216	0.305	3.29x10 ⁻⁴	0.63	0.262	0.308	0.075	0.80	8.05x10 ⁻⁵	1.93x10 ⁻³	0.61	0.09	0.78
rs7723819	110455246	A	0.326	0.467	3.09x10 ⁻⁷	0.55	0.382	0.464	0.004	0.71	7.67x10 ⁻⁹	1.29x10 ⁻⁵	0.54	8.02x10 ⁻³	0.71
rs2416257	110463389	A	0.094	0.152	2.88x10 ⁻³	0.58	0.118	0.139	0.294	0.83	2.00x10 ⁻³	0.02	0.60	0.26	0.80
rs11241095	110467408	G	0.235	0.315	1.24x10 ⁻³	0.67	0.265	0.314	0.061	0.79	2.25x10 ⁻⁴	2.37x10 ⁻³	0.62	0.10	0.79
rs12515367	110475752	G	0.240	0.318	1.84x10 ⁻³	0.68	0.266	0.317	0.052	0.78	2.78x10 ⁻⁴	2.77x10 ⁻³	0.62	0.09	0.78
rs10051830	110480744	A	0.323	0.453	1.41x10 ⁻⁶	0.58	0.371	0.454	0.003	0.71	2.37x10 ⁻⁸	1.79x10 ⁻⁵	0.54	4.87x10 ⁻³	0.69
rs1043828	110491907	C	0.464	0.350	1.26x10 ⁻⁵	1.61	0.418	0.346	0.010	1.36	5.99x10 ⁻⁷	4.99x10 ⁻⁴	1.59	7.39x10 ⁻³	1.41
rs6870356	110495136	A	0.088	0.154	7.27x10 ⁻⁴	0.53	0.119	0.151	0.128	0.76	2.72x10 ⁻⁴	0.01	0.56	0.12	0.74

Supplementary Table 2: Loci with P -values $< 1 \times 10^{-5}$ in the meta-analysis of the discovery (CCHMC) and replication (CHOP) cohorts. Minor allele frequencies of the cases (F_A) and controls (F_U) as well as the P -values, odds ratios (OR) and the lower and upper 95% confidence intervals of the odds ratios (L95, U95) are presented for the two cohorts separately. The results of the fixed effects meta-analysis are presented in the final column (Meta-P). Abbreviations, BP: base pair, A1: minor allele.

Discovery (CCHMC)										Replication (CHOP)						Meta P
CHR	SNP	BP	A1	F_A	F_U	P	OR	L95	U95	F_A	F_U	P	OR	L95	U95	
16	rs371915	83135742	A	0.23	0.13	4.89x10⁻⁶	1.9	1.44	2.5	0.20	0.13	6.12x10 ⁻⁴	1.67	1.24	2.25	2.19x10⁻⁸
3	rs6799767	177512798	G	0.33	0.42	6.22x10⁻⁴	0.67	0.54	0.85	0.34	0.42	4.75x10 ⁻³	0.71	0.56	0.90	4.21x10⁻⁷
4	rs1986734	77777963	C	0.4	0.51	1.41x10⁻⁴	0.65	0.53	0.82	0.41	0.49	9.93x10 ⁻³	0.74	0.58	0.93	1.27x10⁻⁶
5	rs1032757	81975074	T	0.12	0.07	6.80x10⁻⁵	1.96	1.4	2.74	0.12	0.08	0.036	1.47	1.02	2.11	1.66x10⁻⁶
12	rs167769	55790042	T	0.44	0.37	5.34x10⁻³	1.36	1.1	1.69	0.47	0.35	1.84x10 ⁻⁵	1.65	1.31	2.07	1.90x10⁻⁶
11	rs1939875	95062515	T	0.35	0.26	1.93x10⁻⁴	1.54	1.22	1.93	0.33	0.26	0.016	1.35	1.06	1.72	3.02x10⁻⁶
4	rs13106227	77775860	G	0.28	0.38	4.50x10⁻⁴	0.66	0.52	0.83	0.29	0.36	0.017	0.74	0.58	0.95	3.55x10⁻⁶
6	rs9500256	58416914	A	0.26	0.42	1.16x10⁻⁶	0.49	0.37	0.66	0.37	0.44	0.023	0.75	0.58	0.96	4.97x10⁻⁶
17	rs6501384	65802728	T	0.41	0.33	1.97x10⁻³	1.41	1.13	1.76	0.42	0.36	0.044	1.27	1.01	1.60	6.01x10⁻⁶
8	rs13278732	99157757	T	0.33	0.27	0.021	1.31	1.04	1.65	0.37	0.27	1.35x10 ⁻⁴	1.59	1.25	2.02	6.16x10⁻⁶
18	rs7236477	27156320	G	0.06	0.03	6.74x10⁻⁴	2.22	1.39	3.55	0.06	0.04	0.035	1.67	1.03	2.70	6.57x10⁻⁶
8	rs11989782	121763831	A	0.32	0.23	3.60x10⁻⁴	1.53	1.21	1.93	0.30	0.21	4.87x10 ⁻⁴	1.57	1.22	2.02	7.44x10⁻⁶
10	rs2224865	85412834	G	0.4	0.31	1.27x10⁻³	1.44	1.15	1.79	0.40	0.32	7.66x10 ⁻³	1.38	1.09	1.74	9.35x10⁻⁶

Supplementary references

- 1 Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-87 (2003).
- 2 Luca, D. et al. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am J Hum Genet* **82**, 453-63 (2008).
- 3 Kato, A. et al. TLR3- and Th2 cytokine-dependent production of thymic stromal lymphopoietin in human airway epithelial cells. *J Immunol* **179**, 1080-7 (2007).