Genome-wide association analysis of eosinophilic esophagitis provides insight into the tissue specificity of this allergic disease

Supplementary Information

Methods

Nature Genetics: doi:10.1038/ng.3033

Supplementary Figure 7. Two genes, including *CAPN14*, are replicated in EoE genetic, expression, and epigenetic analyses
Supplementary Figure 8. Proposed model of increased EoE risk at *CAPN14* locus
Supplementary Figure 9. Association of variants identified in loci previously reported to be associated with allergic sensitization with EoE risk
Supplementary Figure 10. Flow chart of the EoE GWAS analytical pipeline
Supplementary Figure 11. The Genome-Studio cluster plot for rs8041227 (also known as kgp344467) at 15q13 from the OMNI5 analysis (all EoE cases and local controls)

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 1. Subjects in the EoE GWAS

	Local EoE cohort		External CoFAR EoE		Local controls		External University of Michigan controls	
	Males	Females	Males	Females	Males	Females	Males	Females
CCHMC cohort	376	138			376	384	2,459	3,539
CoFAR cohort			150	72			1,020	1,468
Combined analysis	514		222		760		8,486	

EoE, Eosinophilic Esophagitis; CoFAR, NIH Consortium of Food Allergy Research [CoFAR].

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 2. Variants that demonstrated or failed to demonstrate association in CCHMC cohort, CoFAR cohort, or local analyses. We required that variants demonstrate association ($p < 0.05$) in the CCHMC and CoFAR cohorts (Supplementary Table 1) and a local cohort analysis including all of the EoE cases in the CCHMC cohort and the locally recruited controls.

Variants that demonstrated association in all three analyses:

					All cases and controls (weighted Z-score combined P)		Cases from CCHMC and controls from CCHMC and University of Michigan (dbGAP)		Cases and controls from CCHMC		Cases from CoFAR and separate controls from University of Michigan (dbGAP)	
Chr	**SNP**	**BP**	**MAF cases**	**MAF controls**	**p value**	**OR**	**p value CCHMC cohort**	**OR CCHMC cohort**	**p value local**	**OR local**	**p value CoFAR cohort**	**OR CoFAR cohort**
1	rs2000260	108673405	0.37	0.43	$6.56x10^{-7}$	0.757	0.0001099	0.773	0.00128	0.766	0.00171	0.724
2	rs77569859	31411287	0.09	0.05	$3.30x10^{-10}$	1.982	$1.41x10^{-7}$	1.930	0.00046	1.787	0.00124	2.055
5	rs3806933	110406742	0.37	0.44	$2.00x10^{-8}$	0.731	$1.32x10^{-6}$	0.724	0.00088	0.759	0.00310	0.740
5	rs2055376	116181428	0.04	0.02	$7.12x10^{-8}$	2.296	$1.89x10^{-6}$	2.289	$6.24x10^{-6}$	3.323	0.00772	2.400
8	rs2898261	10958539	0.35	0.42	$4.84x10^{-8}$	0.735	$1.73x10^{-6}$	0.724	$1.34x10^{-5}$	0.696	0.00616	0.755
10	rs11819199	20865157	0.09	0.06	$2.89x10^{-7}$	1.62	$1.64x10^{-5}$	1.618	0.00032	1.735	0.00204	1.700
11	rs2155219	76299194	0.413	0.491	$3.65x10^{-7}$	0.729	$2.64x10^{-5}$	0.743	0.00050	0.742	0.00583	0.670
11	rs118086209	86104495	0.03	0.02	$2.35x10^{-7}$	2.19	0.000193	1.993	0.00106	2.479	$6.61x10^{-5}$	2.831
15	rs8041227	31538542	0.2	0.28	$6.34x10^{-10}$	0.657	$3.97x10^{-6}$	0.6932	0.00277	0.749	$3.38x10^{-5}$	0.581
21	rs17004598	45078556	0.03	0.01	$1.37x10^{-7}$	2.566	$4.48x10^{-6}$	2.592	0.00204	2.640	0.00329	2.754

BP, build 37 map position of the SNP; CCHMC, Cincinnati Children's Hospital Medical Center; CoFAR, NIH Consortium of Food Allergy Research); Chr and band, chromosome and cytogenetic band; MAF. minor allele frequency across cases or controls; OR, odds ratio for the minor allele; p value, the weighted Z-score method implemented in METAL was used to combine the p values for the CCHMC and CoFAR cohorts; SNP, rs ID of variant.

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 2 (continued)

Variants that failed to demonstrate association in one of the three analyses:

					All cases and controls (weighted Z-score combined P)		Cases from CCHMC and controls from CCHMC and University of Michigan (dbGAP)		Cases and controls from CCHMC		Cases from CoFAR and separate controls from University of Michigan (dbGAP)	
Chr	**SNP**	**BP**	**MAF cases**	**MAF controls**	**p value**	**OR**	**p value CCHMC cohort**	**OR CCHMC cohort**	**p value local**	**OR local**	**p value CoFAR cohort**	**OR CoFAR cohort**
1	rs28530674	19234134	0.07	0.04	$3.43x10^{-7}$	1.826	$1.24x10^{-8}$	2.074	$6.58x10^{-5}$	2.015	**0.73**	**1.112**
1	rs2296225	21031042	0.13	0.08	$1.08x10^{-7}$	1.626	$1.03x10^{-7}$	1.713	0.0018	1.505	**0.32**	**1.24**
11	rs77301713	76832446	0.04	0.02	$1.46x10^{-7}$	2.22	$4.06x10^{-10}$	2.669	$1.72x10^{-5}$	2.766	**0.38**	**0.602**
17	rs3744790	76893135	0.14	0.2	$8.16x10^{-7}$	0.651	$2.50x10^{-6}$	0.623	$2.58x10^{-5}$	0.609	**0.097**	**0.748**
22	rs2075277	21382482	0.13	0.09	$9.36x10^{-7}$	1.544	$2.28x10^{-6}$	1.599	0.0011	1.526	**0.089**	**1.39**
1	rs11206830	56960123	0.04	0.02	$7.92x10^{-8}$	2.162	0.0017	1.775	**0.18**	**1.377**	$1.49x10^{-7}$	3.36
18	rs9956738	49940973	0.03	0.01	$3.53x10^{-7}$	2.472	0.0038	1.913	**0.11**	**1.613**	$2.59x10^{-7}$	4.321

BP, build 37 map position of the SNP; CCHMC, Cincinnati Children's Hospital Medical Center; CoFAR, NIH Consortium of Food Allergy Research); Chr and band, chromosome and cytogenetic band; MAF. minor allele frequency across cases or controls; OR, odds ratio for the minor allele; p value, the weighted Z-score method implemented in METAL was used to combine the p values for the CCHMC and CoFAR cohorts; SNP, rs ID of variant.

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 3. Association of the top replicated loci in a logistic regression adjusted for atopy using all cases (n = 736) and local controls (n = 760). Bolded loci reach genome-wide significance in the combined analysis. The difference between the odds ratio of the unadjusted association and the association with atopy adjustment is given in the column "Difference in OR".

					Association with atopy adjustment		
Chr	**SNP**	**BP**	**Band**	**Nearest Gene**	**p value**	**Odds Ratio**	**Difference in OR**
1	rs2000260	108673405	1p13	*SLC25A24*	7.27×10^{-4}	0.7555	0.002
2	**rs77569859**	**31411287**	2p23	***CAPN14***	$\mathbf{3.10 \times 10^{-4}}$	**1.979**	0.003
5	**rs3806933**	**110406742**	5q22	***TSLP/WDR36***	$\mathbf{4.98 \times 10^{-5}}$	**0.7116**	0.019
5	rs2055376	116181428	5q23	Near *SEMA6a*	5.56×10^{-3}	2.236	0.060
8	**rs2898261**	**5001364**	8p23	***XKR6***	$\mathbf{6.23 \times 10^{-2}}$	**1.401**	-0.666
10	rs11819199	139337546	10p12	*MIR4675*	3.39×10^{-5}	2.536	-0.916
11	rs2155219	20865157	11q13	Between *C11orf30* and *LRRC32*	2.95×10^{-4}	1.768	-1.039
11	rs118086209	86104495	11q14	*CCDC81*	1.36×10^{-3}	2.553	-0.363
15	**rs8041227**	**31538542**	15q13	**Between *LOC283710* and *KLF13***	$\mathbf{3.96 \times 10^{-4}}$	**0.7097**	-0.053
21	rs17004598	45078556	21q22	*HSF2BP*	7.25×10^{-4}	3.059	-0.493

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 4. Meta-analysis of the association of SNPs at 5q22. The results of the previous genome-wide association study of EoE[9] was combined with the association of the same variants in the current study using the Fisher's method to combined p values. Subjects included in both studies were removed from the present study before the p value was calculated.

SNP	Position	EoE Genome-wide study p value		Combined p value
		2009	2014	
rs3806932	110,433,574	$3.1x10^{-9}$	$7.2x10^{-8}$	$2.3x10^{-16}$
rs7723819	110,455,246	$7.6x10^{-9}$	$2.5x10^{-8}$	$1.9x10^{-16}$
rs10051830	110,480,744	$2.3x10^{-8}$	$2.9x10^{-7}$	$6.6x10^{-16}$

Position, build 37 map position of the SNP; Chr, chromosome and cytogenetic band; SNP, rs ID of variant

Nature Genetics: doi:10.1038/ng.3033

Supplementary Table 5. Expression quantitative trait loci (eQTL) analysis of top EoE risk variants in a public database with expression from peripheral blood mononuclear cells.

CHR	SNP	SNP Position	Probe	Z-score	Gene name	p value
1	rs2000260	108474928	7510681	6.07	*SLC25A24*	1.25×10^{-9}
8	rs2898261	10995949	4730672	5.32	*XKR6*	1.02×10^{-7}

Chr, chromosome and cytogenetic band; SNP, rs ID of variant; SNP Position, build 37 map position of the SNP.

Nature Genetics: doi:10.1038/ng.3033

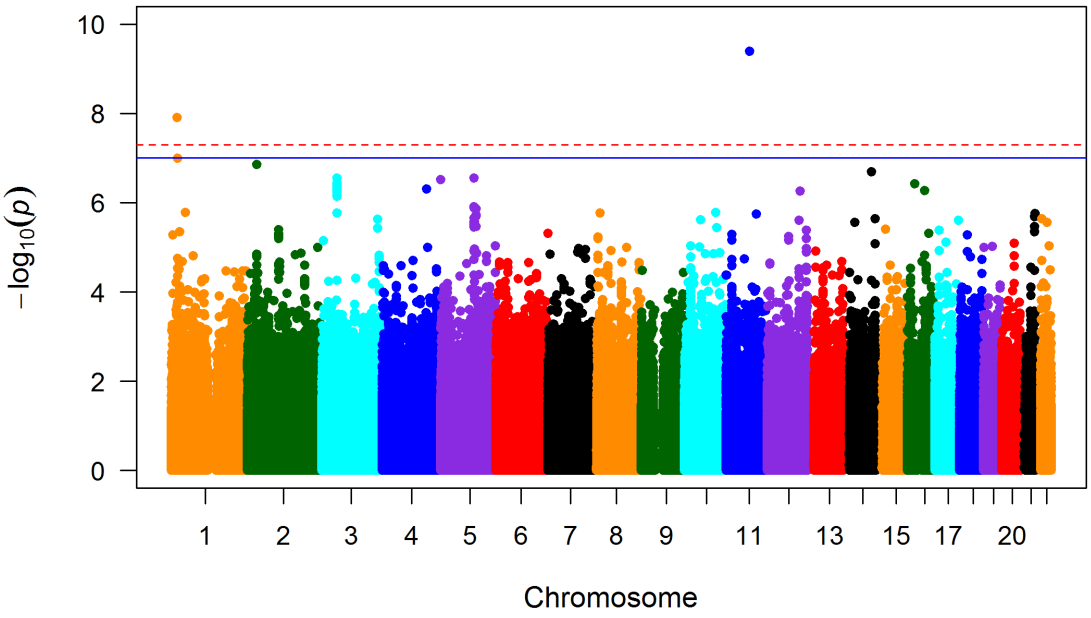
Supplementary Table 6. EoE genetic linkage with allergic sensitization markers[+]

Chr	Marker	Position	Gene	Odds Ratio AS	p value	Odds Ratio EoE
1	rs2056417	10581658	*PEX14*	1.07	0.26	0.927
2*	rs10174949	8442248	*ID2*	1.07	0.32	0.95
2	rs10189629	102879464	*IL1RL2/IL1RL1*	1.16	0.84	1.02
2*	**rs10497813**	**198914072**	***PLCL1***	**0.92**	**0.031**	**0.888**
3*	**rs9860547**	**188128979**	***LPP***	**1.08**	**5.0×10^{-7}**	**1.32**
4*	rs2101521	38811551	*TLR1/TLR6*	1.12	0.43	0.948
4	rs17388568	123329362	*ADAD1*	1.08	0.11	1.11
5	rs7720838	40486896	*PTGER4*	1.08	0.91	1.00
5*	**rs1438673**	**110467499**	***WDR36/CAMK4***	**0.89**	**5.1×10^{-11}**	**0.685**
6	rs9266772	31352113	*HLA-C/MICA*	1.11	0.28	1.086
6	rs6906021	32626311	*HLA-DQA1/HLA-DQB1*	1.10	0.26	1.08
8	**rs6473223**	**81268155**	***TPD52/ZBTB10***	**1.07**	**0.00019**	**1.26**
9	**rs7032572**	**6172380**	***RANBP6/IL33***	**1.12**	**0.048**	**1.17**
10*	**rs962993**	**9053132**	***GATA3***	**0.93**	**0.019**	**0.877**
11	**rs2155219**	**76299194**	***C11orf30/LRRC32***	**0.90**	**1.1×10^{-6}**	**0.759**
11	rs10893845	128186882	*ETS1*	1.06	0.73	0.979
14*	rs1998359	38077148	*FOXA1/TTC6*	1.08	0.66	1.03
15*	**rs17228058**	**67450305**	***SMAD3***	**1.08**	**0.0032**	**1.20**
16*	rs2107357	27410829	*IL4R/IL21R*	1.09	0.53	0.948
17*	rs9303280	38074031	*GSDMB*	1.07	0.51	0.964
16	**rs7203459**	**11230703**	***CLEC16A***	**0.93**	**4.6×10^{-5}**	**0.731**
20	rs6021270	50141264	*NFATC2*	1.16	0.57	1.07

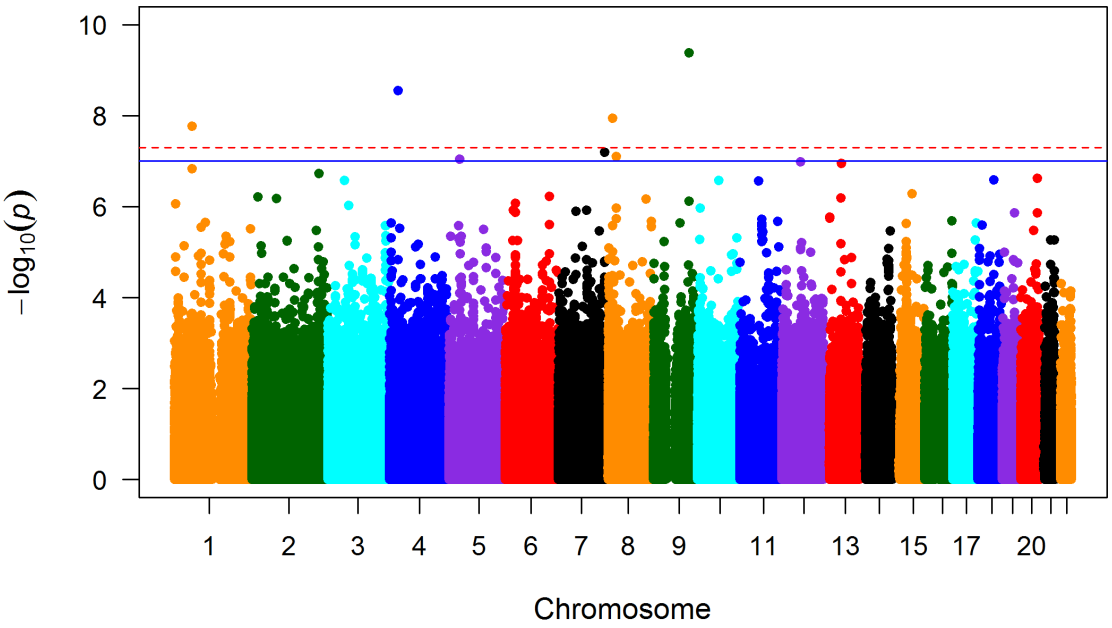
[+]Association of variant previously reported at genome-wide significant in two GWAS of allergic sensitization (AS)[39,40]. An asterisk by the chromosome indicates that the variant was imputed from the EoE GWAS. The odds ratios are given for the allele with the smallest frequency (the minor allele) in the EoE analysis.

Nature Genetics: doi:10.1038/ng.3033

a. **CCHMC cohort: 514 cases and 6,758 controls**

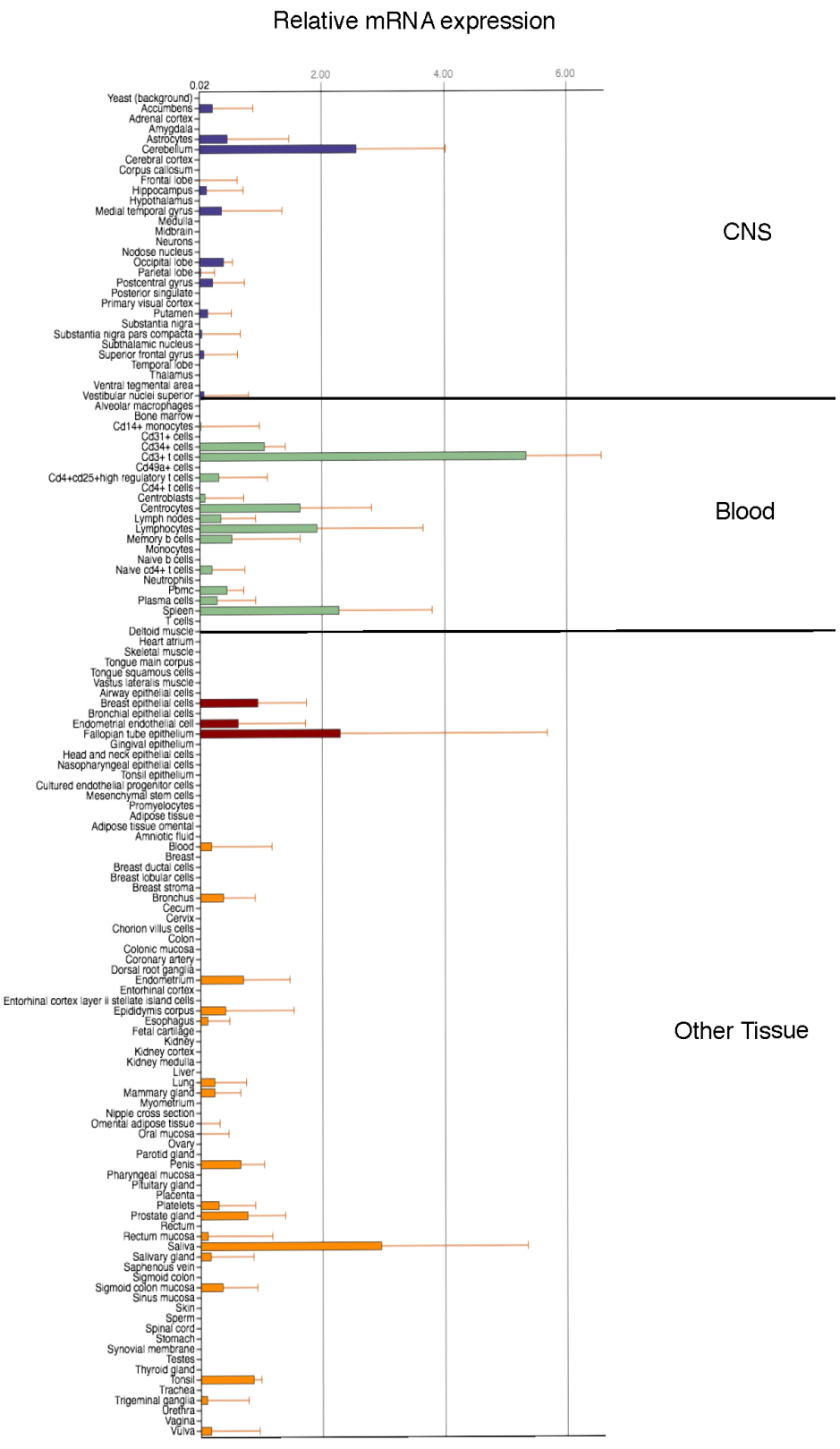


b. **CoFAR cohort: 222 cases and 2,488 controls**



Supplementary Figure 1. Manhattan plot of the p values obtained from the genome-wide association analysis. Data is from 1,468,075 genetic variants with minor allele frequencies greater than 1% in the subjects with EoE in 514 EoE cases and 6,758 controls **(a)** and 222 cases and 2,488 controls **(b)**. The –log of the probability is shown as a function of the genomic position on the autosomes. Genome-wide significance (red dotted line, $p \leq 5\times10^{-8}$) and suggestive significance (blue line, $p \leq 10^{-7}$) are indicated.

Nature Genetics: doi:10.1038/ng.3033

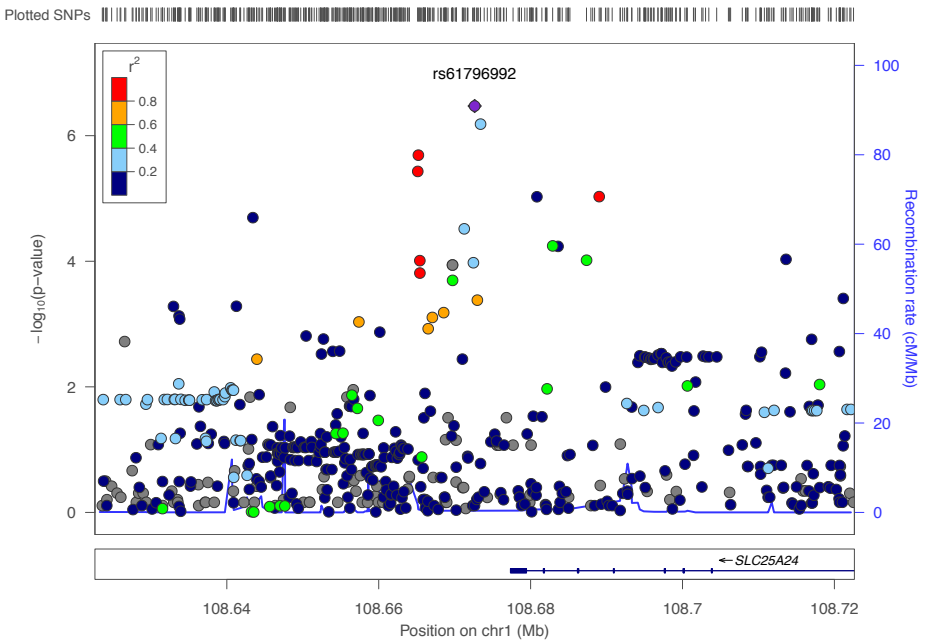


Supplementary Figure 2. The expression of *XKR6* (data collected form www.biogps.org)[52,53]. The colors of the bars correspond to the tissue source.

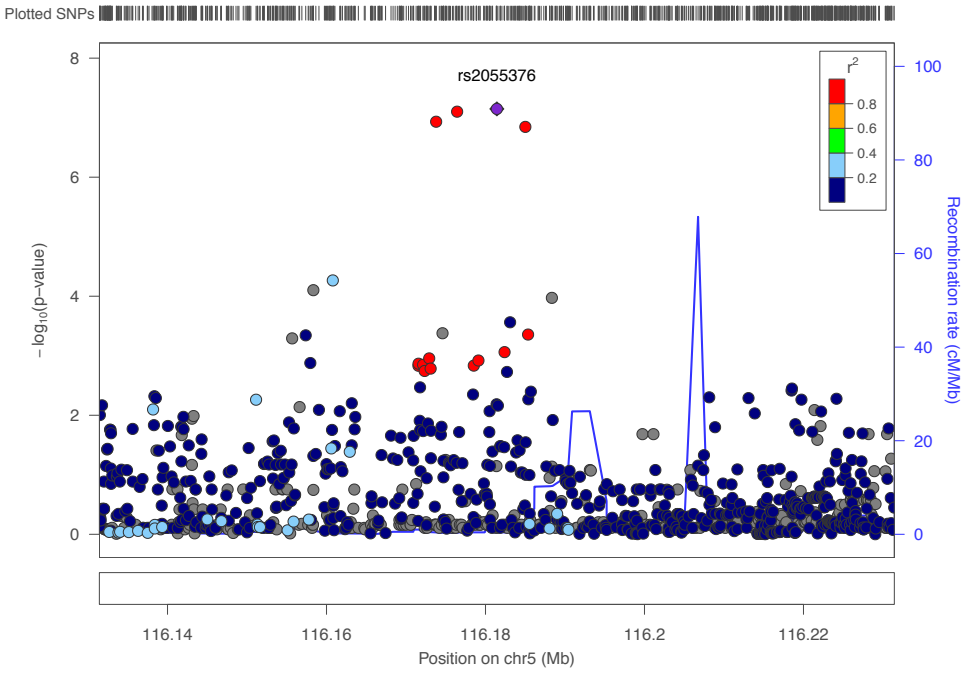
Nature Genetics: doi:10.1038/ng.3033

Supplementary Figure 3

1q13



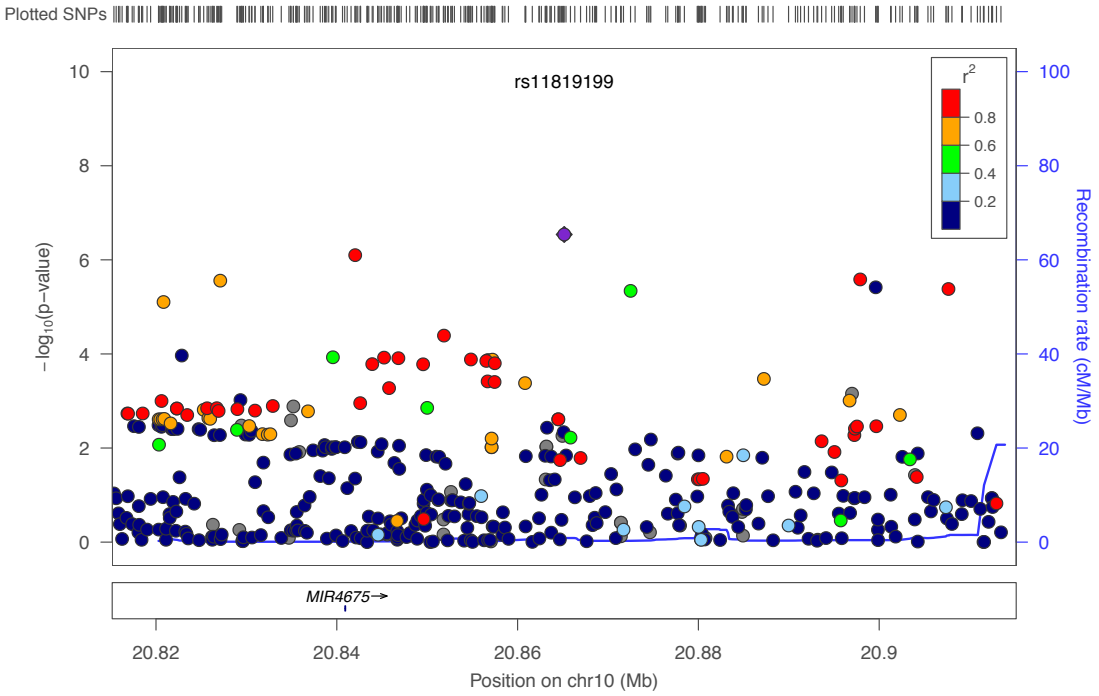
5q23



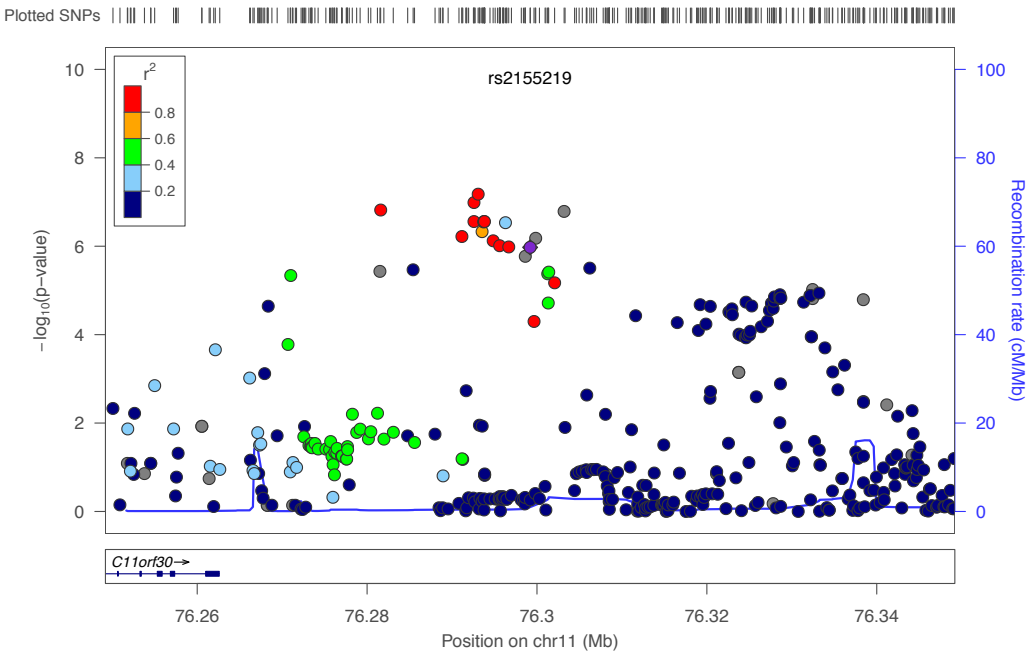
Nature Genetics: doi:10.1038/ng.3033

Supplementary Figure 3 (continued)

10p12



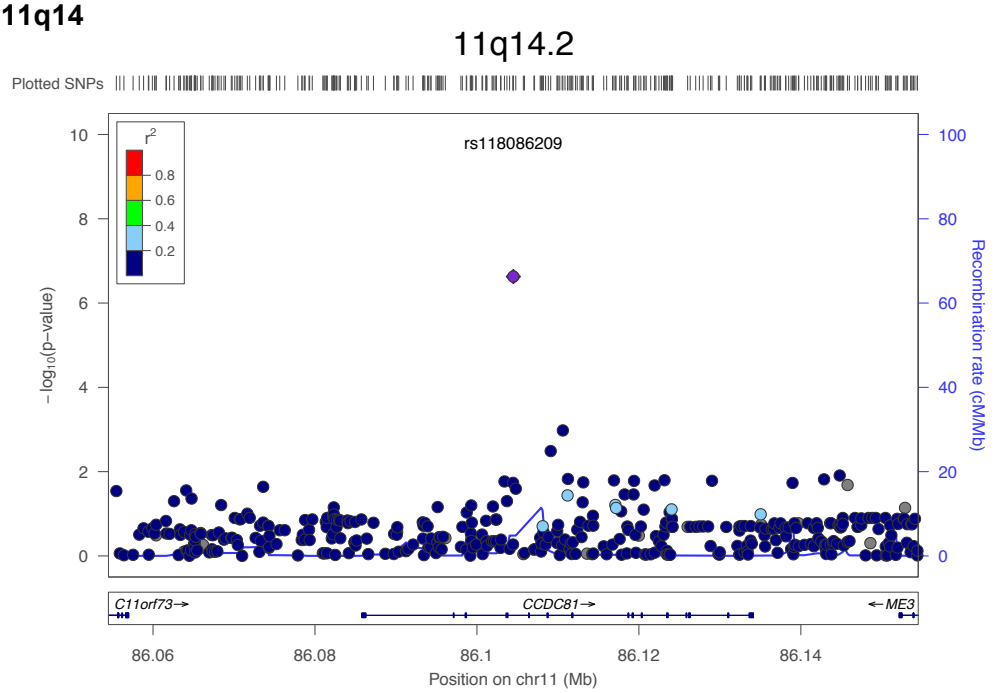
11q13



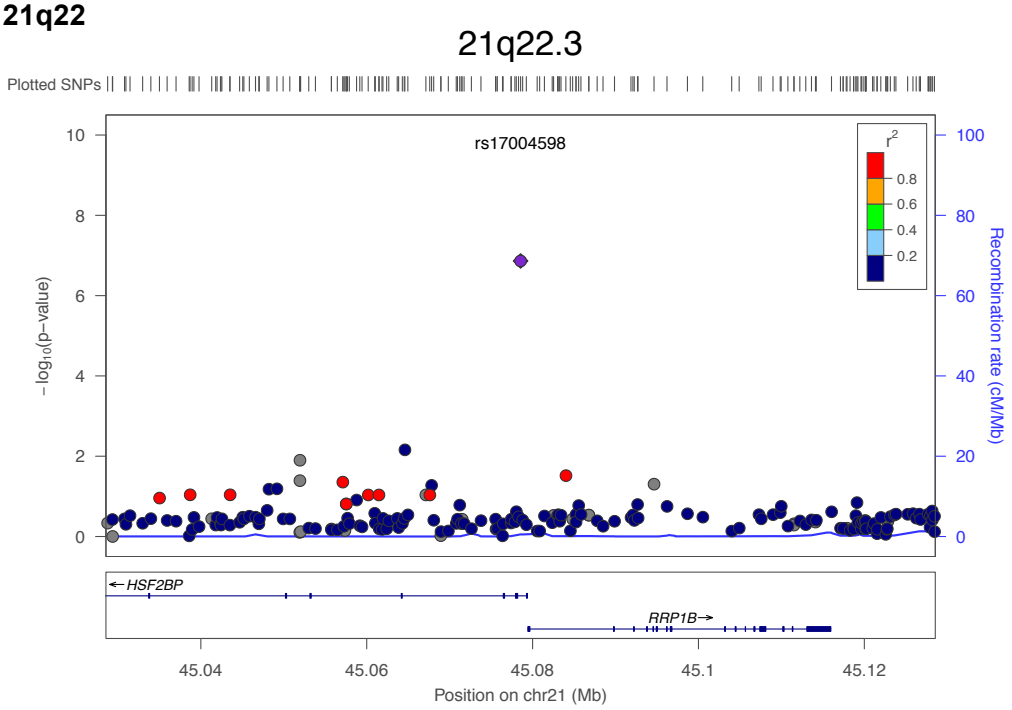
Nature Genetics: doi:10.1038/ng.3033

Supplementary Figure 3 (continued)

11q14

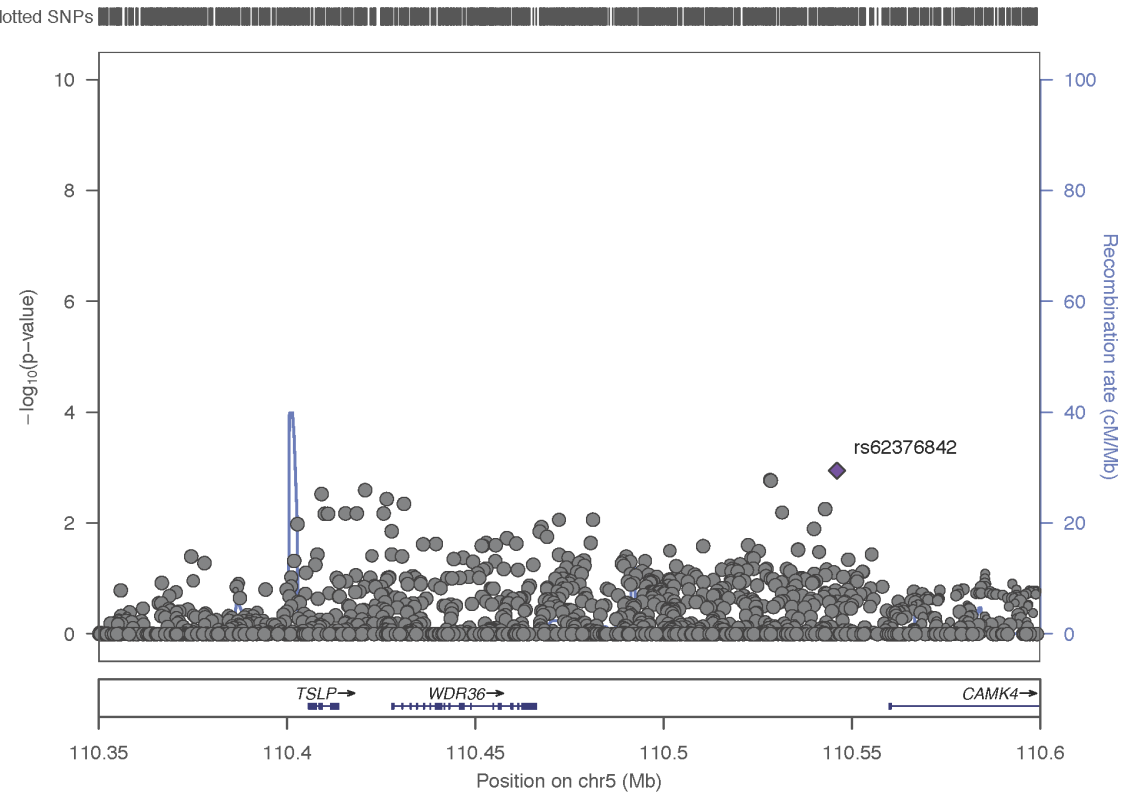


21q22



Supplementary Figure 3. Fine mapping of loci associated (replicated and $5x10^{-8} < p < 10^{-6}$) with EoE risk. P values ($-\log_{10}$) of the genetic association analysis of imputed variants on the *y* axis are plotted as a function of genomic positions of each genotyped and imputed SNPs (MAF > 0.05). Genes in the region are shown below. The LD values (r^2) between the lead SNP and the other SNPs as assessed in the March 2012 release of the 1,000 genomes project are indicated in different colors. The blue lines indicate the recombination rates in cM per Mb using HapMap controls.

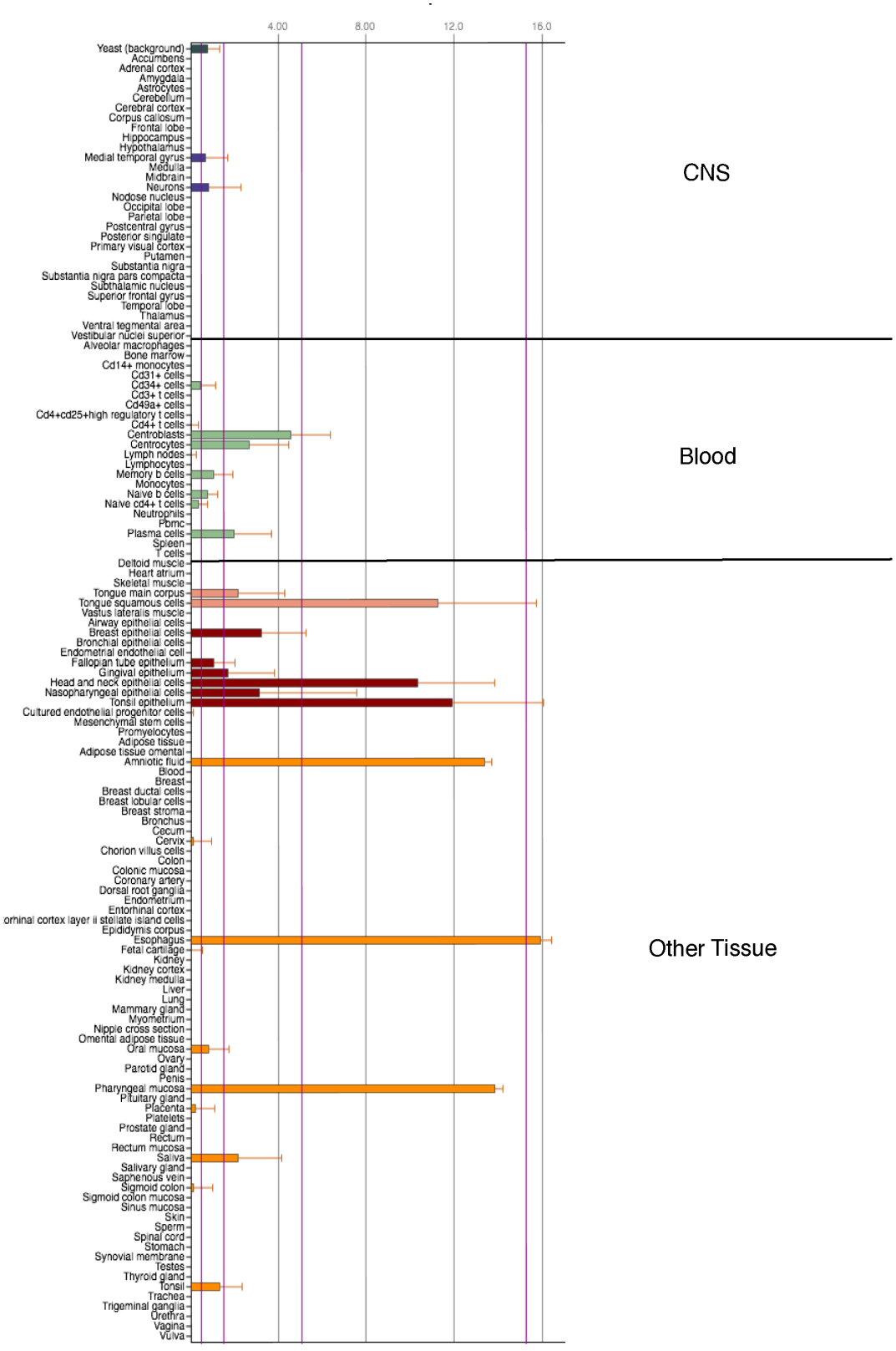
Nature Genetics: doi:10.1038/ng.3033



Supplementary Figure 4. 5q22. After adjusting for any of the most highly associated variants in the region (rs1438672), a group of 12 variants between *TSLP* and *WDR36* demonstrate residual association 0.001< p < 0.05. After accounting for multiple testing, we cannot reject the null hypothesis that there is one independent genetic effect in the region. *P* values ($-\log_{10}$) of genetic association analysis are plotted as a function of genomic positions of each imputed variant (minor allele frequency (MAF) > 0.01) on chromosome 5 (Chr5) using a logistic regression adjusting the genotype of rs1438672. Genes in the region are shown below. The blue lines indicate the recombination rates in cM per Mb using HapMap controls.

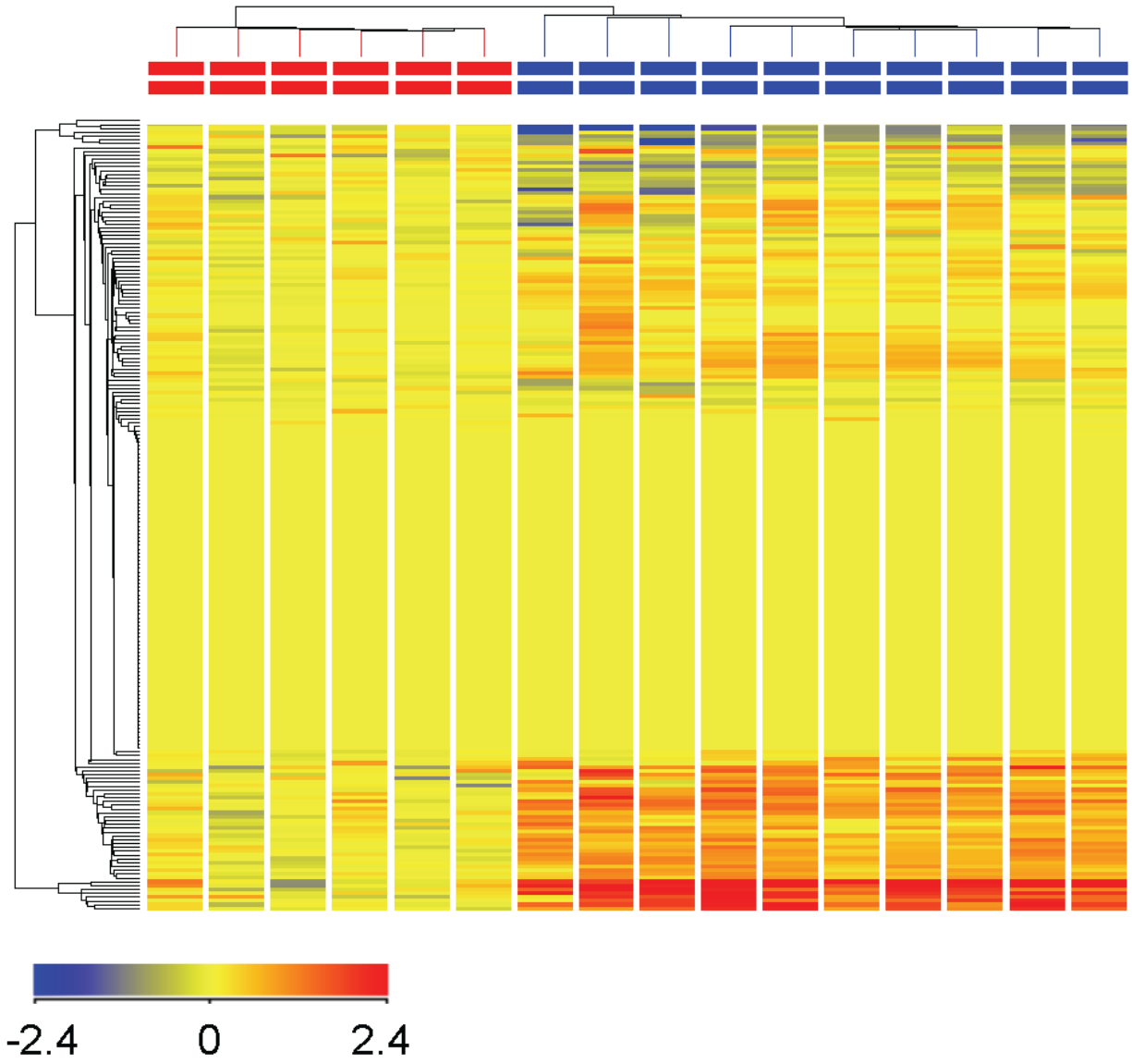
.

Nature Genetics: doi:10.1038/ng.3033



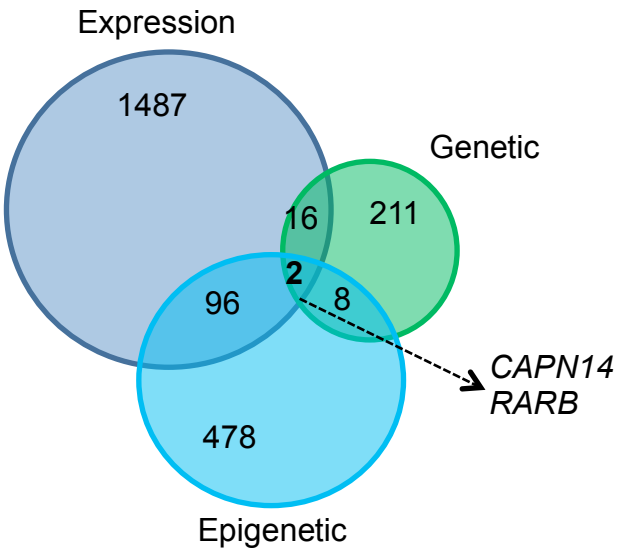
Supplementary Figure 5. Expression of *CAPN14* (data collected from www.biogps.org)[52,53]. The colors of the bars correspond to the tissue source.

Nature Genetics: doi:10.1038/ng.3033



Supplementary Figure 6. Expression of genes near EoE-associated genetic variants ($p < 10^{-4}$) in the esophageal biopsies of EoE patients vs. controls. In this dataset, 9 of the 215 transcripts (representing 208 genes) have an average difference of ≥2 FPKMs between esophagi from active EoE cases and controls. The differences in expression were sufficient to segregate EoE cases from controls. The 8 transcripts with differences in expression greater than 2-fold can be found in **Figure 4**. Subjects with EoE are noted with blue bars, and subjects without EoE are indicated with red bars. All expression is normalized to the average expression in esophageal biopsies without EoE.

Nature Genetics: doi:10.1038/ng.3033

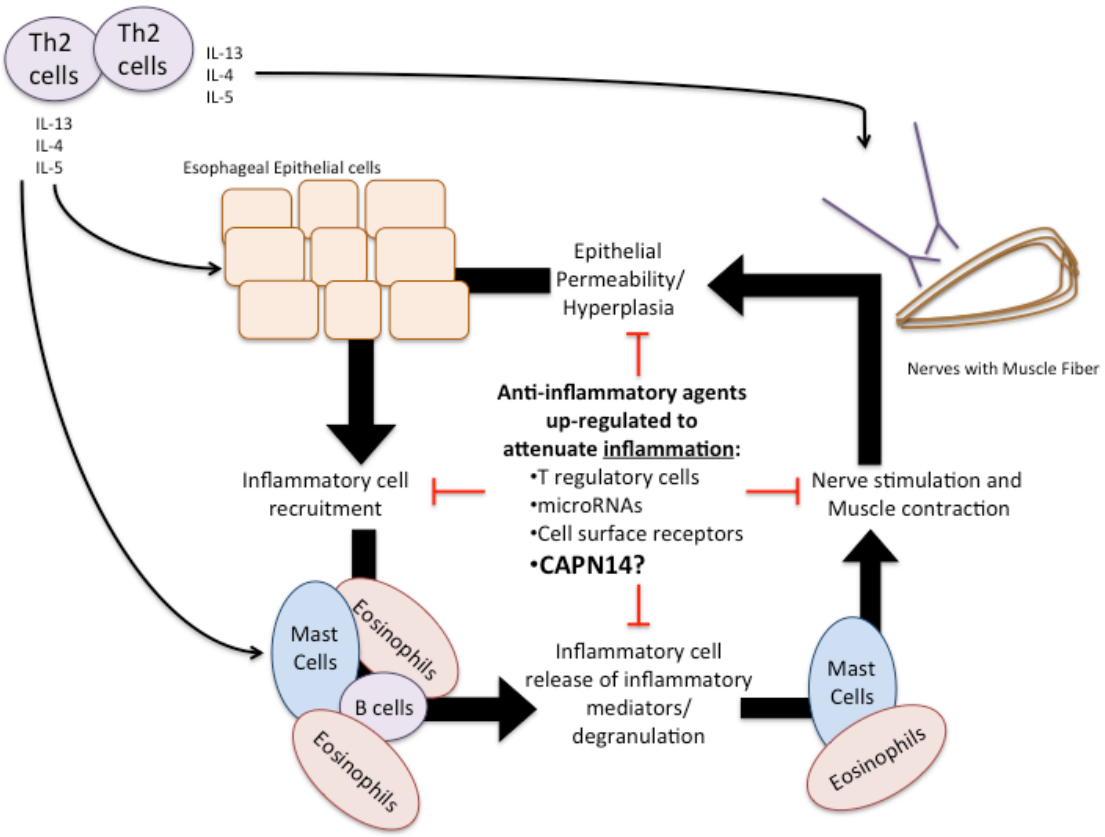


	H3K27Ac		
	Fold change	M value	P value
CAPN14	6.02	-3.08	5.8×10^{-5}
RARB	5.19	-1.5	0.0001

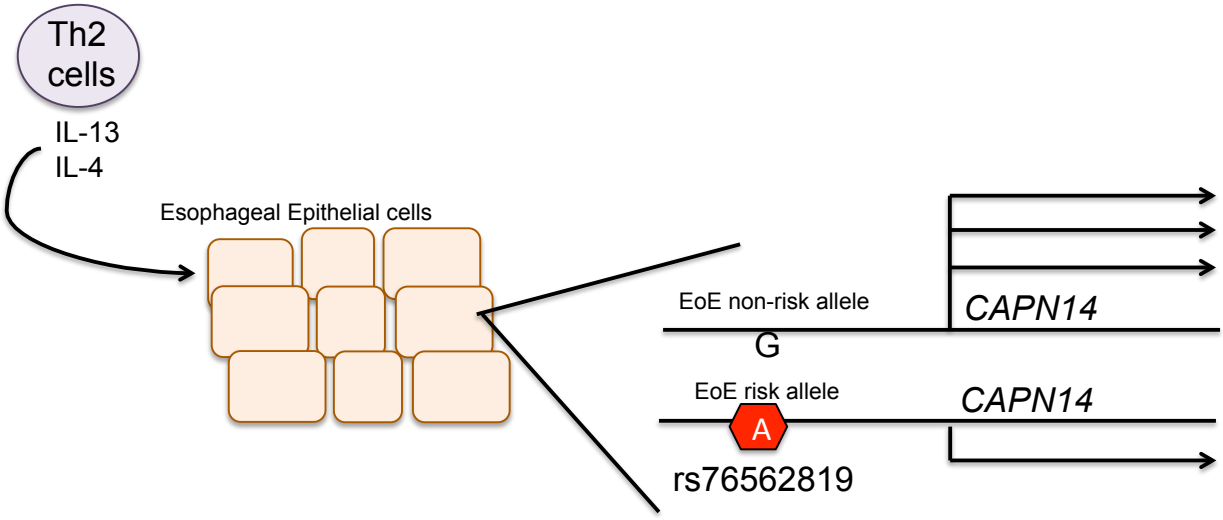
Supplementary Figure 7. Two genes including *CAPN14* are replicated in EoE genetic, expression, and epigenetic analyses. We assessed overlap from analyses of genes within 25 kb of genetic variants with $p < 10^{-4}$, genes with increased H3K27Ac after 6-hour exposure to IL-13 ($p < 0.01$, $M < -1$ at 5 kb from TSS), and genes with increased esophageal expression in EoE (2-fold increase in expression in esophageal biopsy in EoE compared to control, corrected $p < 0.05$).

Nature Genetics: doi:10.1038/ng.3033

a.



b.



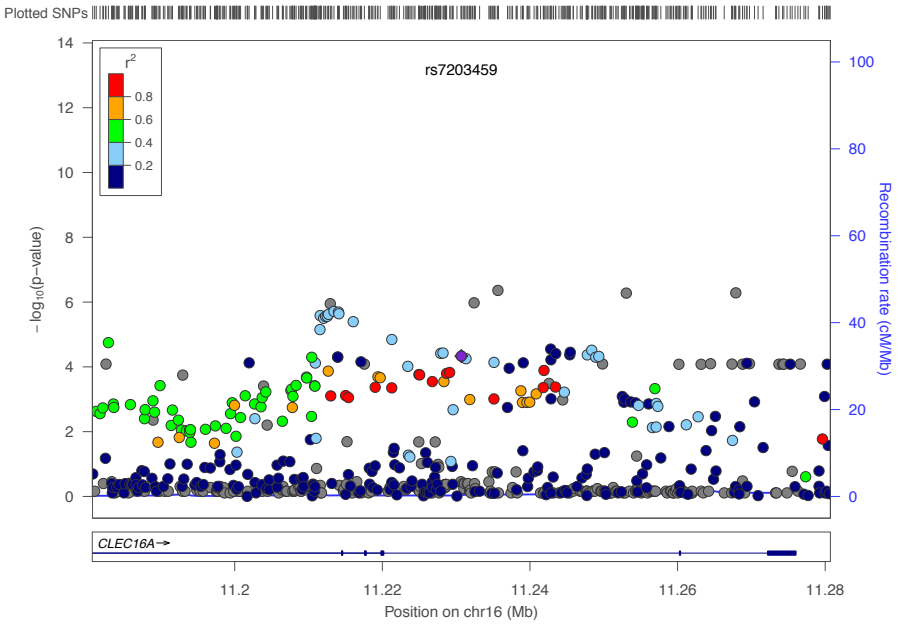
Supplementary Figure 8. Proposed model of increased EoE risk at *CAPN14* locus. *CAPN14* is expressed specifically in the esophagus (**Fig 2a**). Allergic inflammatory mediators including IL-13 and IL-4 induce *CAPN14* expression and activity (**Figs 2h** and **3a** and refs [36,37]), and the regulation of the increased expression is mediated in part through the acetylation of histones (Figure 2i). The abundance of IL-13 and IL-4 in the esophagus of patients with EoE[5-8] results in increased *CAPN14* expression (**Figs 2b-d**) and activity (**Fig 3b**), and the calpain activity of CAPN14 potentially attenuates further inflammation by digesting endogenous proteins. The genetic variants

Nature Genetics: doi:10.1038/ng.3033

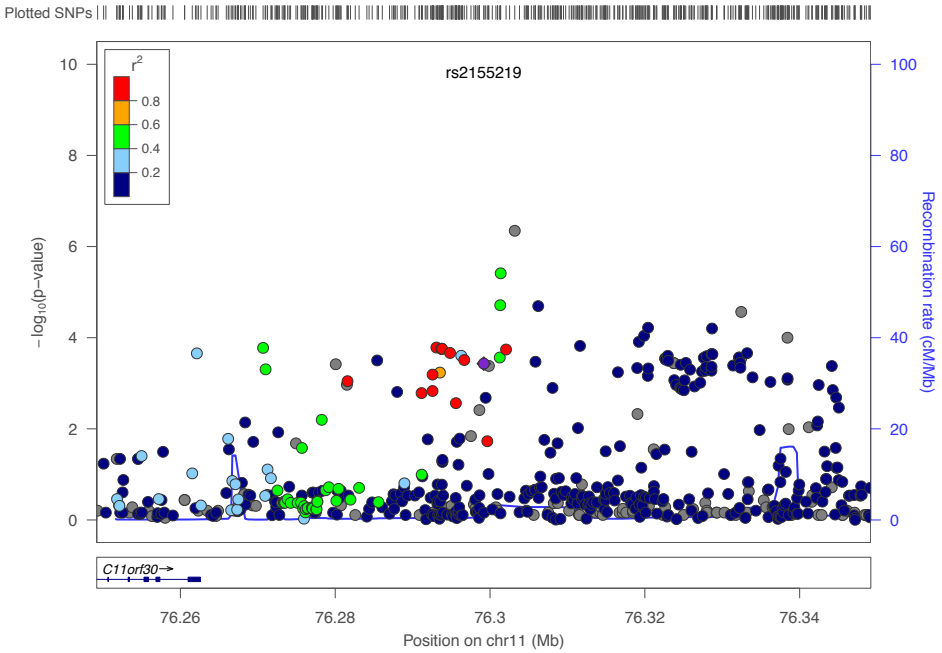
associated with EoE risk at the *CAPN14* locus lead to decreased *CAPN14* expression (**Fig 2e**) feasibly through the binding of a protein (**Fig 2j**), a transcription factor that potentially acts as a transcriptional repressor. Together, our data is consistent with a model **(a)** in which CAPN14 is induced along with other anti-inflammatory agents such as microRNAs, cell surface receptors, and T regulatory cells. **b.** When exposed to IL-13 and IL-4, chromosomes in epithelial cells with the *CAPN14* EoE risk allele are unable to induce *CAPN14* expression to the extent of chromosomes with the non-risk allele. We propose a model in which the allelic change in *CAPN14* expression dysregulates a critical negative feedback loop in the esophagus resulting in increased risk of pathology and EoE.

Nature Genetics: doi:10.1038/ng.3033

CLEC16A



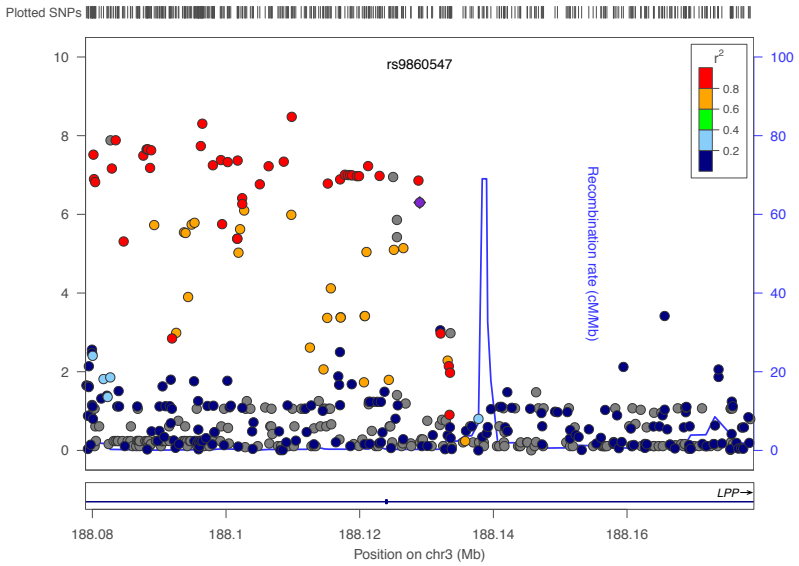
C11orf30/LRRC32



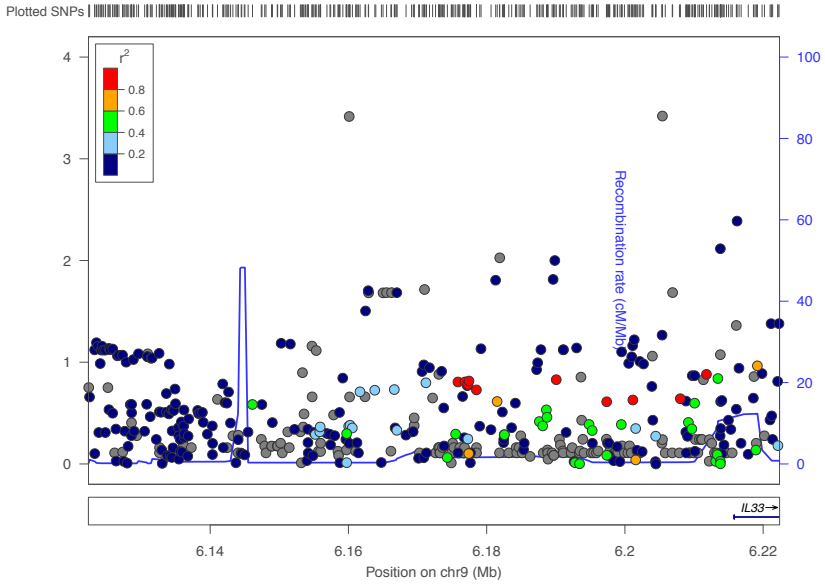
Nature Genetics: doi:10.1038/ng.3033

Supplementary Figure 9 (continued)

LPP

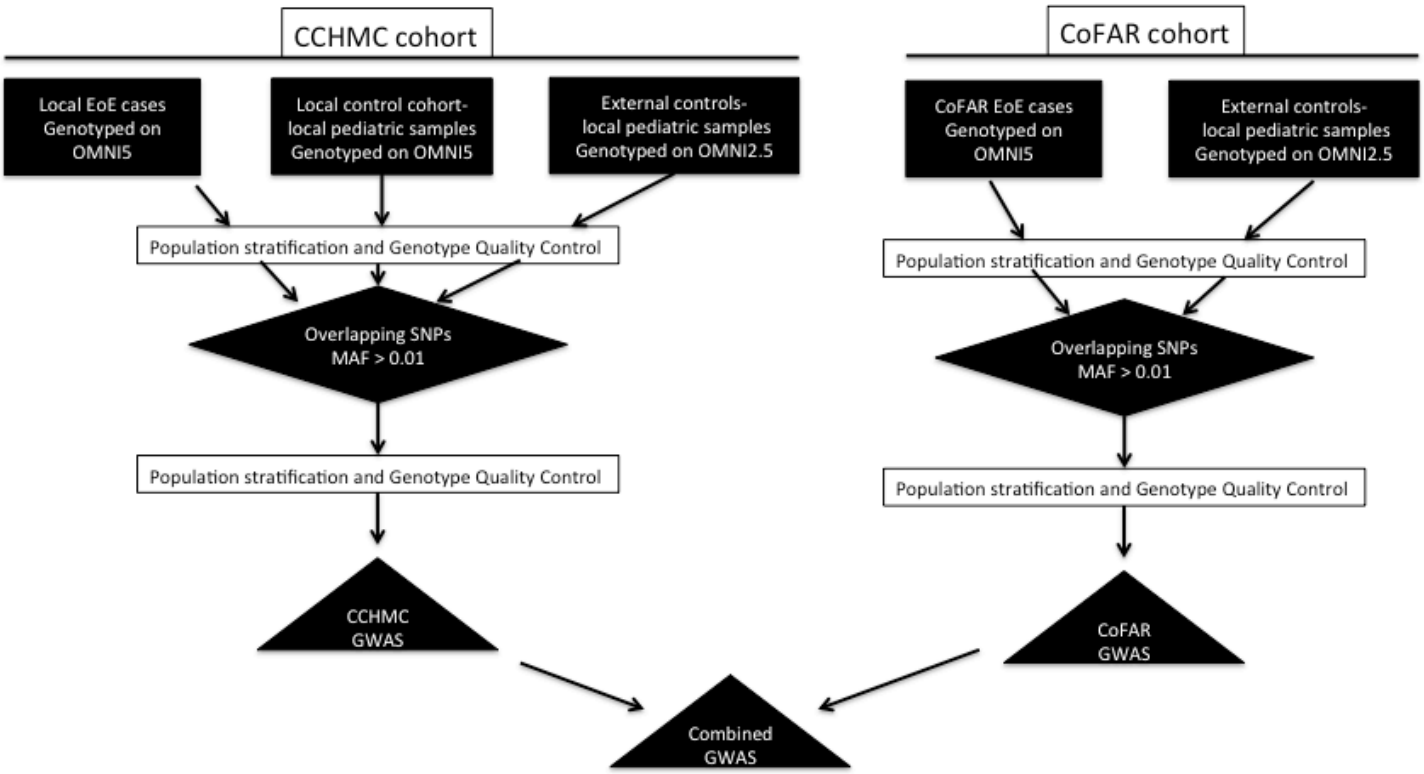


IL33



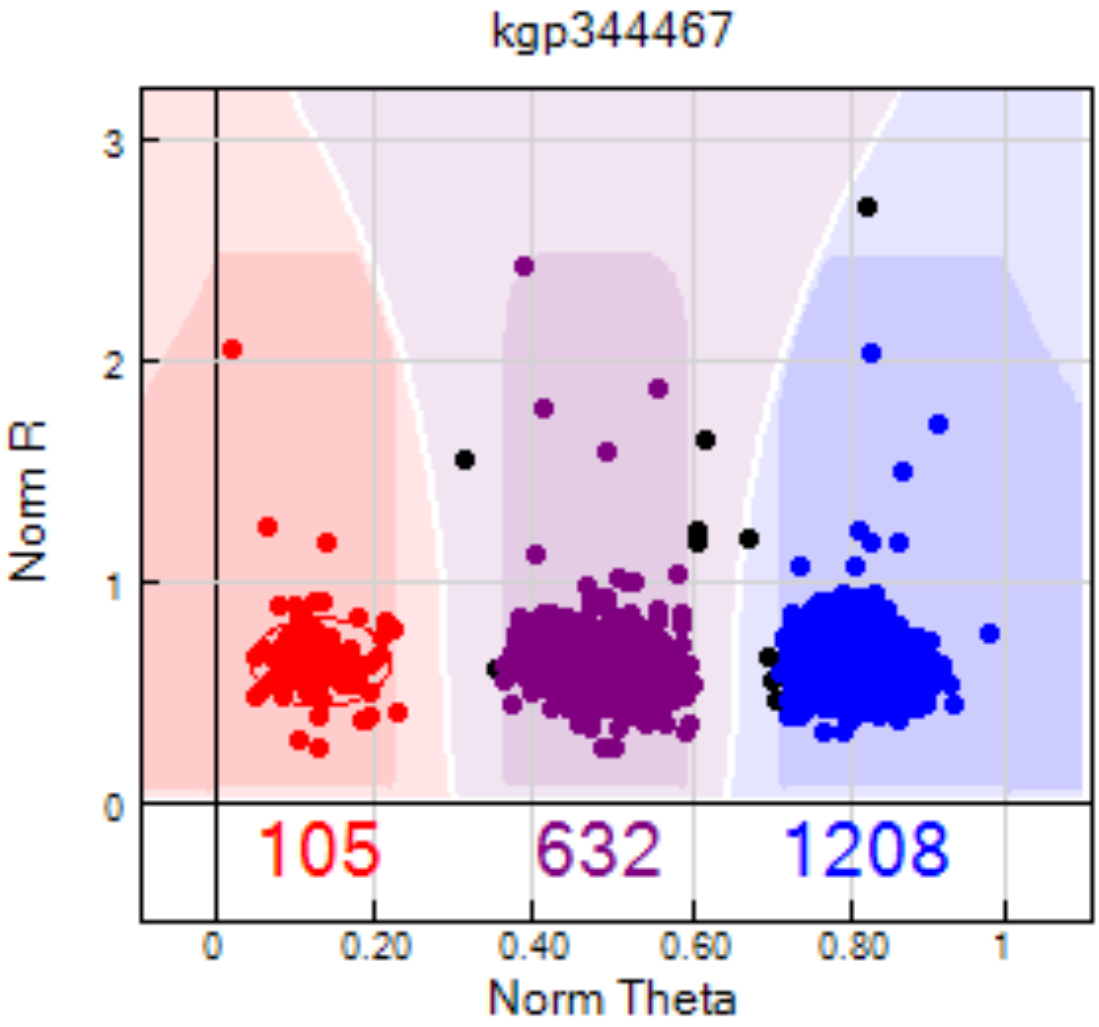
Supplementary Figure 9. Association of variants in loci previously reported to be associated to be associated with allergic sensitization with EoE risk. The variant that was most highly associated in the allergic sensitization study is indicated by a purple diamond and given in each panel. The LD values (r^2) between the lead SNP and the other SNPs as assessed in the March 2012 release of the 1,000 genomes project ar indicated in different colors. The blue lines behind the genetic variants indicate the recombination rates in cM per Mb using HapMap controls.

Nature Genetics: doi:10.1038/ng.3033



Supplementary Figure 10. Flow chart of the EoE GWAS analytical pipeline. External controls from University of Michigan (acquired through dbGAP) were randomly assigned to CCHMC and CoFAR cohort. CoFAR, NIH Consortium of Food Allergy Research [CoFAR]; EoE, Eosinophilic Esophagitis; GWAS, genome-wide association study; MAF, minor allele fre quency; SNP, single nucleotide polymorphism.

Nature Genetics: doi:10.1038/ng.3033



Supplementary Figure 11. The Genome-Studio cluster plot for rs8041227 (aka kgp344467) at 15q13 on chromosome 15 from the OMNI5 analysis (all EoE cases and local controls). The normalized theta (Norm Theta) is the ratio of signal intensities assayed for the A and B of the rs8041227 allele against normalized R (signal intensity; Norm R). Subjects in this graph are homozygous risk (red), heterozygotes (purple), or homozygous non-risk (blue).

Nature Genetics: doi:10.1038/ng.3033