Variants in a Novel Epidermal Collagen Gene *(COL29A1)* Are Associated with Atopic Dermatitis

Cilla Söderhäll[1,2☯], Ingo Marenholz[1,2☯], Tamara Kerscher[1,2], Franz Rüschendorf[2], Jorge Esparza-Gordillo[1,2], Margitta Worm[3], Christoph Gruber[1], Gabriele Mayr[4], Mario Albrecht[4], Klaus Rohde[2], Herbert Schulz[2], Ulrich Wahn[1], Norbert Hubner[2], Young-Ae Lee[1,2*]

1 Department of Pediatric Pneumology and Immunology, Charité Universitätsmedizin Berlin, Germany, **2** Max-Delbrück-Centrum for Molecular Medicine, Berlin, Germany, **3** Department of Dermatology and Allergology, Charité Universitätsmedizin Berlin, Germany, **4** Max Planck Institute for Informatics, Saarbrücken, Germany

Atopic dermatitis (AD) is a common chronic inflammatory skin disorder and a major manifestation of allergic disease. AD typically presents in early childhood often preceding the onset of an allergic airway disease, such as asthma or hay fever. We previously mapped a susceptibility locus for AD on Chromosome 3q21. To identify the underlying disease gene, we used a dense map of microsatellite markers and single nucleotide polymorphisms, and we detected association with AD. In concordance with the linkage results, we found a maternal transmission pattern. Furthermore, we demonstrated that the same families contribute to linkage and association. We replicated the association and the maternal effect in a large independent family cohort. A common haplotype showed strong association with AD ($p = 0.000059$). The associated region contained a single gene, *COL29A1,* which encodes a novel epidermal collagen. *COL29A1* shows a specific gene expression pattern with the highest transcript levels in skin, lung, and the gastrointestinal tract, which are the major sites of allergic disease manifestation. Lack of *COL29A1* expression in the outer epidermis of AD patients points to a role of collagen XXIX in epidermal integrity and function, the breakdown of which is a clinical hallmark of AD.

Citation: Söderhäll C, Marenholz I, Kerscher T, Rüschendorf F, Esparza-Gordillo J, et al. (2007) Variants in a novel epidermal collagen gene *(COL29A1)* are associated with atopic dermatitis. PLoS Biol 5(9): e242. doi:10.1371/journal.pbio.0050242

Introduction

Atopic dermatitis (AD) is a chronic inflammatory skin disease that is characterized by intensely itchy skin lesions. AD is one of the most common chronic diseases in childhood affecting 10%–20% of children in industrialized societies [1], with a steady increase over the past decades [2,3]. Along with asthma and hay fever, AD is commonly associated with an abnormal immune response and the formation of allergy antibodies (IgE) against innocuous environmental allergens.

AD is often the first clinical manifestation of allergic disease. The onset of disease is typically observed during the first two years of life [4]. For the majority of affected children, AD heralds a lifetime of allergic disease. A susceptible child commonly passes a characteristic sequence of transient or persistent disease stages that is known as the "atopic march" which begins with AD and food allergy in the young infant and continues with the development of allergic airways disease later in childhood and adulthood [5]. The close familial and intra-individual association of the allergic disorder strongly suggests shared genetic etiology.

A strong genetic component in allergic disorders has been recognized almost a century ago. Cooke and van der Veer first reported that the relatives of patients are at significantly increased risk of developing allergic disease [6]. Even today, a positive family history for allergic disorders is the single strongest predictor for the development of AD [7]. Additional evidence for the importance of genetic factors in atopic disease comes from twin studies. The concordance rate for AD among monozygotic twins of about 80% far exceeds the concordance rate of 20% observed among dizygotic twins [8,9]. These data clearly indicate that the genetic contribution to the expression of AD is substantial. Furthermore, studies on the vertical transmission of AD and atopic disease show that children are more likely to inherit these disorders if the mother is affected (parent-of-origin effect) [10]. The predominance of maternal inheritance may be due to environmental factors such as uterine milieu or breast feeding, but they may also arise due to genetic mechanisms such as parent-specific gene expression (genomic imprinting) [11].

AD and atopic disorders are regarded as multifactorial conditions, the onset and severity of which are influenced by both genetic and environmental factors. The data are consistent with an immune etiology shared by all allergic diseases and a congenital target organ defect, the penetrance of which is modified by multiple environmental factors

Academic Editor: Gonçalo Abecasis, University of Michigan, United States of America

Received February 9, 2007; **Accepted** July 9, 2007; **Published** September 11, 2007

Copyright: © 2007 Söderhäll et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: AD, atopic dermatitis; cM, centiMorgan; *COL29A1*, collagen XXIX gene; ECM, extracellular matrix; LD, linkage disequilibrium; RT-PCR, reverse-transcriptase PCR; SNP, single nucleotide polymorphism; vWA, von Willebrand factor–type A domain

* To whom correspondence should be addressed. E-mail: yolee@mdc-berlin.de

☯ These authors contributed equally to this work.

Author Summary

Atopic dermatitis (AD, eczema) is a common chronic inflammatory skin disorder and a major manifestation of allergic disease. Typically, AD first occurs in early childhood, often preceding the onset of allergic airways disease, such as asthma and hay fever. A family history of allergic disorders is the single strongest predictor for AD, showing that genetic factors play a major role in the disease development. We have previously mapped a disease locus for AD on Chromosome 3q21, Now we have used a dense map of microsatellite markers and single nucleotide polymorphisms (SNPs) to find the underlying disease gene. We identified genetic markers in a subregion that showed association with AD, and replicated this finding in a large independent family cohort. The associated region contained a single gene, *COL29A1,* which encodes a novel collagen. We demonstrate that AD patients lack *COL29A1* expression in the outer epidermis, implicating collagen XXIX in epidermal integrity and function. The gene expression pattern of *COL29A1* in other organs, including the lung and the gut, suggests that this gene could have a role in a wider spectrum of allergic diseases and may provide a molecular link between AD and respiratory airways disease and food allergies.

during early childhood. The identification of genes underlying AD and allergic disorders has the capacity to define primary physiologic mechanisms, thereby clarifying disease pathogenesis, identifying pathways and targets for therapeutic intervention.

Although several genome-wide linkage screens for AD have been conducted [12–14], there was no substantial overlap between the identified regions of highest linkage, and the underlying genes remained elusive [15]. We have previously mapped a major susceptibility locus for AD on Chromosome 3q21 [16]. Here we report the identification and characterization of a novel epidermal collagen gene as the underlying disease gene.

Results

To narrow the candidate region spanning 12.75 centiMorgans (cM) (13.5 Mb) on Chromosome 3q21 (Figure 1A), 96 additional microsatellite markers at an average distance of 140 kb were genotyped in 199 affected sibling families with AD from the original linkage scan. Linkage analysis yielded a 1-lod support interval of 5.4 Mb between the markers M3CS075 and M3CS233 (Figure 1B). Subsequently, we conducted an association scan of the 5.4-Mb region using 212 single nucleotide polymorphisms (SNPs) at an average distance of 25.5 kb (Table S1). The SNPs were selected primarily to cover known and predicted genes. Because we had observed a strong maternal effect in the linkage study [16], we chose a family-based association analysis that allowed us to search for risk alleles preferentially transmitted from the mother [17]. Two adjacent SNPs, rs5852593 ($p = 0.0079$) and rs1497309 ($p = 0.016$), located 36 kb apart, were associated with AD (Table S1 and Figure 1B). To define the critical region, we typed 16 additional SNPs. Eight of these markers showed association with AD and a maternal transmission pattern to affected children, which was consistent with our previous linkage results (Table 1). The strongest association with a single marker was observed for rs4688761 ($p_{all} = 0.0016$, $p_{maternal} = 0.0006$). We selected eight markers spanning 96 kb that were associated with AD and carried nonredundant information based on the linkage disequilibrium (LD) in the region, and we performed haplotype analysis which confirmed the results (transmissions $(T)_{all}$:non-transmissions $(NT)_{all}$ = 222:168, $p_{all} = 0.0076$, T_{mat}:NT_{mat} = 105:68, $p_{maternal} = 0.0070$). In addition, we assessed the significance of the difference in maternal-versus-paternal haplotype over transmissions empirically using the permutation procedure for parent-of-origin transmission disequilibrium test (TDT) implemented in PLINK ($P_POO_{maternal\ emp} = 0.014$) [18].

Next, we investigated whether the observed association accounted for the linkage in the region. We used the marker that had shown the strongest association, rs4688761, to identify 73 of the 199 families from the original linkage scan in whom the disease-associated allele had been transmitted to affected offspring. Nonparametric linkage analysis in the 73 associated families yielded significant evidence for linkage ($Z_{all} = 4.18$ versus 4.31 in the complete cohort), demonstrating that the majority of the linkage signal was attributable to these families. The significance of this finding was assessed by performing nonparametric linkage analysis in 10,000 random selections of 73 families. An empirical significance level was calculated as the proportion of replicates for which the maximum Z_{all} score was equal or greater than that obtained in the actual analysis. The probability of obtaining a Z_{all} score of ≥ 4.18 by chance in a random selection of 73 families was estimated by 10,000 simulations to be 0.027.

To confirm the association result, we used a large independent replication set consisting of 292 complete nuclear families including 481 children with AD. We genotyped the selected eight SNPs covering an interval of 96 kb that were associated with AD in the discovery dataset. We confirmed the association with AD across all markers with the strongest association with AD observed for marker A36637742 ($p = 0.00038$), which also showed a significant overtransmission of the maternal allele ($p = 0.0013$) (Table 2). The association remained significant after correction for multiple testing. For each marker, it was the more common allele that was overtransmitted. Haplotypes were constructed over the region, which confirmed the association with AD and showed that this phenotype is associated with the overtransmission of the most common haplotype (haplotype frequency 65%, Table 3), and of the maternal allele ($p_{maternal} = 0.000025$ for AD, $P_POO_{maternal\ emp} = 0.033$). Next, we compared the AD status among the parents: in the discovery cohort, significantly more mothers ($n = 63$) than fathers ($n = 19$) suffered from AD (odds ratio [OR] 4.39, 95% confidence interval [CI] 2.51 – 7.68, $p = 5.46 \times 10^{-8}$). Similarly, in the replication cohort, mothers ($n = 83$) were significantly more frequently affected with AD than fathers ($n = 55$) (OR 1.71, 95% CI 1.16 – 2.52, $p = 8.3 \times 10^{-3}$). To assess whether the parent-of-origin effect that was observed originated from the discrepancy in AD prevalence among the parents, we compared haplotype transmissions from affected and unaffected mothers. In the discovery cohort, the excess in maternal transmissions was not predominantly attributable to affected mothers (T:UT = 24:21, $p = 0.14$) compared to unaffected ones (T:UT = 81:47, $p = 0.0059$). Similarly, affected mothers (T:UT = 25:17, $p = 0.14$) of the replication cohort did not contribute a larger transmission excess than the unaffected ones (T:UT = 70:29, $p = 0.00013$). We conclude that the observed maternal over-

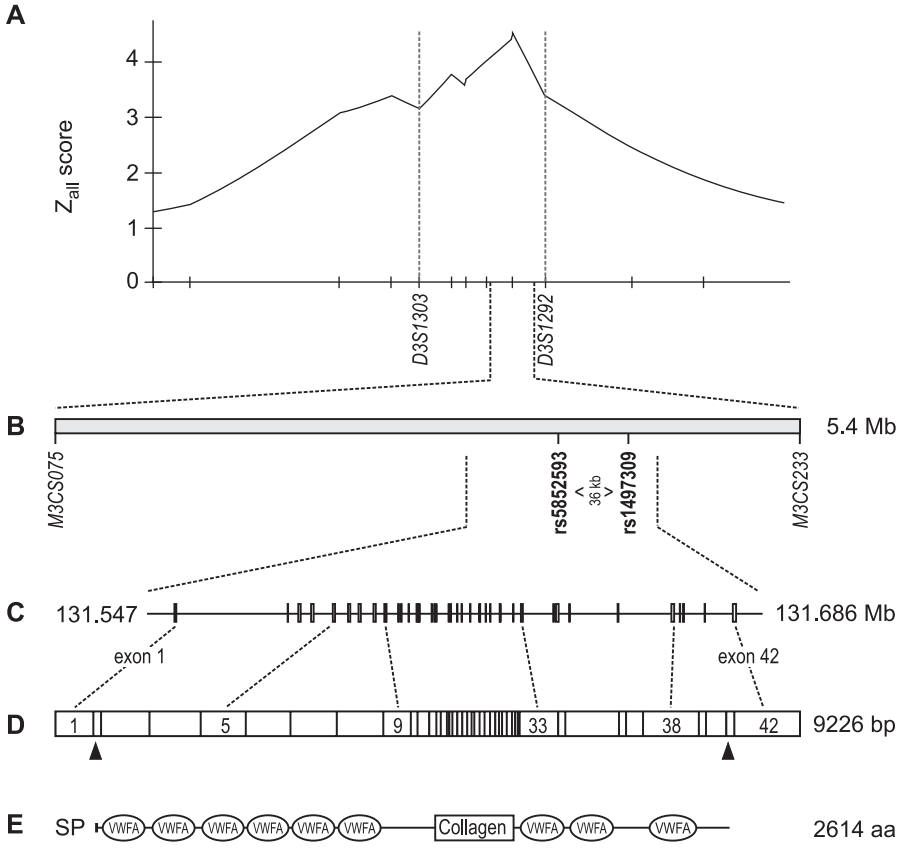


Figure 1. Positional Cloning Strategy for the AD Disease Gene on Chromosome 3q21

(A) The candidate region spanned 12.75 cM between markers *D3S1303* and *D3S1292*. The *y*-axis depicts the *GENEHUNTER* nonparametric Z_{all} as previously reported [16].

(B) Fine mapping with 96 microsatellite markers narrowed the interval to 5.4 Mb between markers *M3CS075* and *M3CS233*. An association scan using 212 SNPs of the region revealed association of AD with two adjacent SNPs, rs5852593 and rs1497309. Genotyping of 16 additional SNPs refined the associated region.

(C) Genomic positions of the 42 exons of *COL29A1* are shown. The gene entirely encompasses the associated region.

(D) The *COL29A1* mRNA consists of 9226 bp. Translation start site and stop codon are indicated.

(E) The predicted open reading frame encodes a protein of 2614 amino acids including a secretion peptide (SP), six N-terminal and three C-terminal vWAs, flanking a short collagen triple helix.

doi:10.1371/journal.pbio.0050242.g001

transmission pattern in both cohorts was not due to the higher AD prevalence among mothers.

A database search within the associated 96-kb interval revealed a single predicted gene, *FLJ35880*, extending 11.6 kb into the associated region. No other expressed sequence tag was detected. To identify any additional transcripts, we used putative exons predicted with The National Center for Biotechnology Information (NCBI) Modelmaker within and bordering the critical region to perform rapid amplification of cDNA ends (RACE) from human skin mRNA. We thus identified a single transcript of 9226 bp that consisted of 42 exons (Figure 1D) and included all eight exons of *FLJ35880*. The corresponding gene spanned 139 kb of genomic sequence and completely encompassed the associated region. Pairwise LD measures (D′) of the genotyped markers indicated that the gene was contained within one 170-kb region of increased LD (Figure 2), whereas the neighboring genes, *LOC440978* and *LOC131873*, were located in separate blocks. The LD structure was consistent with the data for the European population in the HapMap database (Figure S1) [19]. The size of the transcript was confirmed by Northern blot hybridization with an *FLJ35880*-specific probe detecting a single transcript of 9.6 kb in human skin mRNA (unpublished data), which was in good agreement with the RACE experiments.

The open reading frame yielded a protein of 2614 amino acids with an estimated molecular weight of 289.9 kDa. The predicted protein contained a collagenous domain in the central part and was therefore classified as a new member of the collagen superfamily, collagen XXIX. A BLAST search revealed the human collagen VI alpha 3 chain as its closest neighbor (32% identity). Homology with collagen VI alpha 3 was further strengthened by a similar domain architecture consisting of six N-terminal and three C-terminal von Willebrand factor–type A domains (vWAs) flanking a short collagen triple helix (Figure 1E), and an 18–amino acid secretion signal [20]. Sequence analysis of all 42 exons in 46 unrelated children with AD and 2 unaffected individuals revealed 13 common and six rare sequence variations (frequency <2%) predicted to cause nonsynonymous amino acid substitutions (Table S2). Four variants were located within the triple helix repeat of a collagen domain, however none of them changed the first glycine residue in the repeating sequence patterns (Gly-X-Y), which is critical for

Table 1. Results of the ASPEX sib_TDT for SNPs in the Region of *COL29A1* for AD in the Discovery Cohort

Locus	Position on Chromosome 3	Location in *COL29A1*	MAF	Major/ Minor Allele	All Transmissions				Paternal				Maternal			
					T	NT	χ2	*p*-Value	T	NT	χ2	*p*-Value	T	NT	χ2	*p*-Value
rs1349747	131,543,410	5′	18.0	A/G	129	99	3.95	n.s.	48	43	0.27	n.s.	72	47	5.25	n.s.
rs1869319	131,547,433	exon 1	9.4	C/T	80	63	2.02	n.s.	36	32	0.24	n.s.	80	63	2.38	n.s.
rs13095825	131,549,247	intron 1	36.6	T/C	222	167	7.78	0.0072	84	56	5.6	0.024	99	72	4.26	0.047
rs1453243	131,575,875	intron 2	30.5	G/A	187	149	4.3	n.s.	68	49	3.09	n.s.	91	72	2.21	n.s.
rs5852593	131,580,335	intron 3	24.3	del/T	157	111	7.9	0.0079	71	45	5.83	0.026	67	47	3.51	0.045
rs16845861	131,581,039	exon 4	24.4	G/A	166	123	6.4	0.018	68	49	3.09	n.s.	79	55	4.3	0.037
rs1453242	131,586,229	intron 4	24.3	C/T	171	123	7.84	0.0099	75	51	4.57	0.045	79	55	4.3	0.036
rs10212372	131,586,896	exon 5	24.4	A/G	165	123	6.12	0.021	68	50	2.75	n.s.	78	54	4.36	0.037
A36603217	131,590,753	intron 6	22.6	A/G	161	118	6.63	0.015	73	56	2.24	n.s.	72	46	5.73	0.02
rs10934938	131,596,325	intron 7	22.0	T/G	156	114	6.53	0.013	71	54	2.31	n.s.	72	47	5.25	0.019
rs4688761	131,607,679	exon 16	28.6	T/C	187	128	11.05	0.0016	76	56	3.03	n.s.	79	40	12.78	0.0006
A36624476	131,612,012	exon 20	13.5	G/C	109	83	3.52	n.s.	50	44	0.38	n.s.	50	30	5	0.028
rs1497309	131,616,310	intron 22	28.7	C/T	185	137	7.16	0.016	76	57	2.71	n.s.	76	47	6.84	0.013
rs7628563	131,617,253	intron 23	15.1	A/G	123	97	3.07	n.s.	53	38	2.47	n.s.	53	42	1.27	n.s.
A36637742	131.625.270	intron 28	27.8	del/T	184	146	4.4	0.05	75	62	1.44	n.s.	75	53	4.43	0.040
rs1566136	131,634,657	intron 33	28.4	A/G	191	154	3.97	n.s.	85	64	2.96	n.s.	76	60	1.88	n.s.
rs9883988	131,645,085	exon 36	6.3	G/A	56	35	4.85	0.049	31	18	3.45	n.s.	19	11	2.13	n.s.
rs322113	131,670,507	exon 38	7.8	C/T	46	52	0.37	n.s.	27	25	0.08	n.s.	17	25	1.52	n.s.
rs322115	131,672,716	intron 39	18.2	A/T	116	99	1.34	n.s.	48	45	0.1	n.s.	52	38	2.18	n.s.
rs819091	131,678,260	intron 40	23.7	G/A	153	118	4.52	n.s.	67	55	1.18	n.s.	62	39	5.24	0.025
rs7624712	131,686,432	3′	11.6	A/G	75	74	0.01	n.s.	33	38	0.35	n.s.	37	31	0.53	n.s.

MAF, minor allele frequency.
T, number of transmissions.
NT, number of nontransmissions.
n.s., not significant.
doi:10.1371/journal.pbio.0050242.t001

triple helix formation [21]. In addition, we performed in silico comparisons of the six variants located within a vWA with the crystal structure of the human vWF A3 domain. One variant, V669G, affects an amino acid within a highly conserved stretch of eight amino acids. The mutation of an adjacent amino acid within this conserved region of a vWA in the homologous gene *COLVIA1* has been reported to cause the monogenic muscle disorder Bethlem myopathy [22]. This variant, however was rare (allele frequency 1,3%). One additional variant (E455K) is located in a region predicted to be important for integrin–collagen interaction, and another one (M56T) is located in a helix near a rather conserved region that may also affect protein–protein interaction. All coding SNPs were genotyped in the discovery cohort. Four of them showed a positive association with AD that did, however, not account for the observed haplotype association (Table S2).

Gene expression analysis in human tissues revealed a tissue-specific expression pattern of *COL29A1*. The highest expression was observed in the skin, but also in the lung, small intestine, colon, and testis (Figure 3A). Overall, *COL29A1* expression is moderately low compared to more abundant epidermal transcripts such as keratin 10 (unpublished data).

To specify the expression sites in the layers of the skin and to assess the role of collagen XXIX in AD, we performed in situ hybridization using a *COL29A1*-specific cRNA-probe on

Table 2. Association Analysis of Single Markers for Atopic Dermatitis in the Replication Dataset

Locus	Major/ Minor Allele	All Transmissions				Paternal				Maternal			
		T	NT	χ2	*p*-Value[a]	T	NT	χ2	*p*-Value	T	NT	χ2	*p*-Value
rs13095825	T/C	218	174	4.94	0.028	81	66	1.53	0.18	95	66	5.22	0.022
rs16845861	G/A	177	119	11.36	0.00077*	68	46	4.25	0.038	82	46	10.12	0.0014*
rs10212372	A/G	174	117	11.16	0.00093*	68	46	4.25	0.035	81	46	9.65	0.0017*
A36603217	A/G	180	118	12.9	0.00040*	81	52	6.32	0.013	77	44	9	0.0023*
rs10934938	T/G	182	122	11.84	0.00063*	83	54	6.14	0.019	76	45	7.94	0.005*
rs4688761	T/C	208	145	11.24	0.00082*	94	67	4.53	0.03	88	52	9.26	0.0031*
A36637742	del/T	204	139	12.32	0.00038*	84	57	5.17	0.028	89	51	10.31	0.0013*
rs9883988	G/A	75	43	8.68	0.0019*	37	23	3.27	0.055	37	19	5.79	0.0091

[a] *p*-values of the ASPEX sib TDT for atopic dermatitis for all, paternal, and maternal transmissions for the major allele. None of the paternal transmissions were significant after simulation. * *p*-values significant after correction for multiple testing.
T, number of transmissions.
NT, number of nontransmissions
doi:10.1371/journal.pbio.0050242.t002

Table 3. Haplotype Analysis for AD in the Replication Dataset

rs13095825	rs16845861	*rs*10212372	A36603217	*rs*10934938	*rs*4688761	A3667742	rs9883988	Atopic Dermatitis						
								freq.	fam.	S	E(S)	Var(S)	Z	*p*-value
T	G	A	A	T	.		.	0.62	127	429.5	403.4	95.8	2.66	0.0078*
.	G	A	A	T	T		.	0.68	120	432.3	397.8	89.2	3.65	0.00025*
		A	A	T	T	del		0.66	115	428.3	390.7	87.2	4.02	0.000059*
.	.		A	T	T	del	G	0.70	114	415.9	381.4	89.20	3.65	0.00026*

fbat statistic (Z) for 5-marker haplotype transmissions in the replication set adjusted for the correlation among siblings, reporting haplotype frequency (freq.) the number of informative families (fam.), the distribution (S), the expected distribution (E(S)), and the variance (Var(S)). * *p*-values significant after correction for multiple testing.
doi:10.1371/journal.pbio.0050242.t003

skin biopsies of five patients with AD and five healthy patients (controls). In normal skin, *COL29A1* was exclusively expressed in the epidermis with the strongest staining in the suprabasal viable layers. In contrast, the skin of patients with AD revealed a striking absence of *COL29A1* mRNA staining in the most differentiated upper spinous and granular layers (Figure 3B and Figure S2). No significant difference in the expression of *COL29A1* was observed comparing patients with AD to controls (1.28- ± 0.53-fold down-regulation in AD patients, $p = 0.41$) using quantitative Taqman reverse-transcriptase

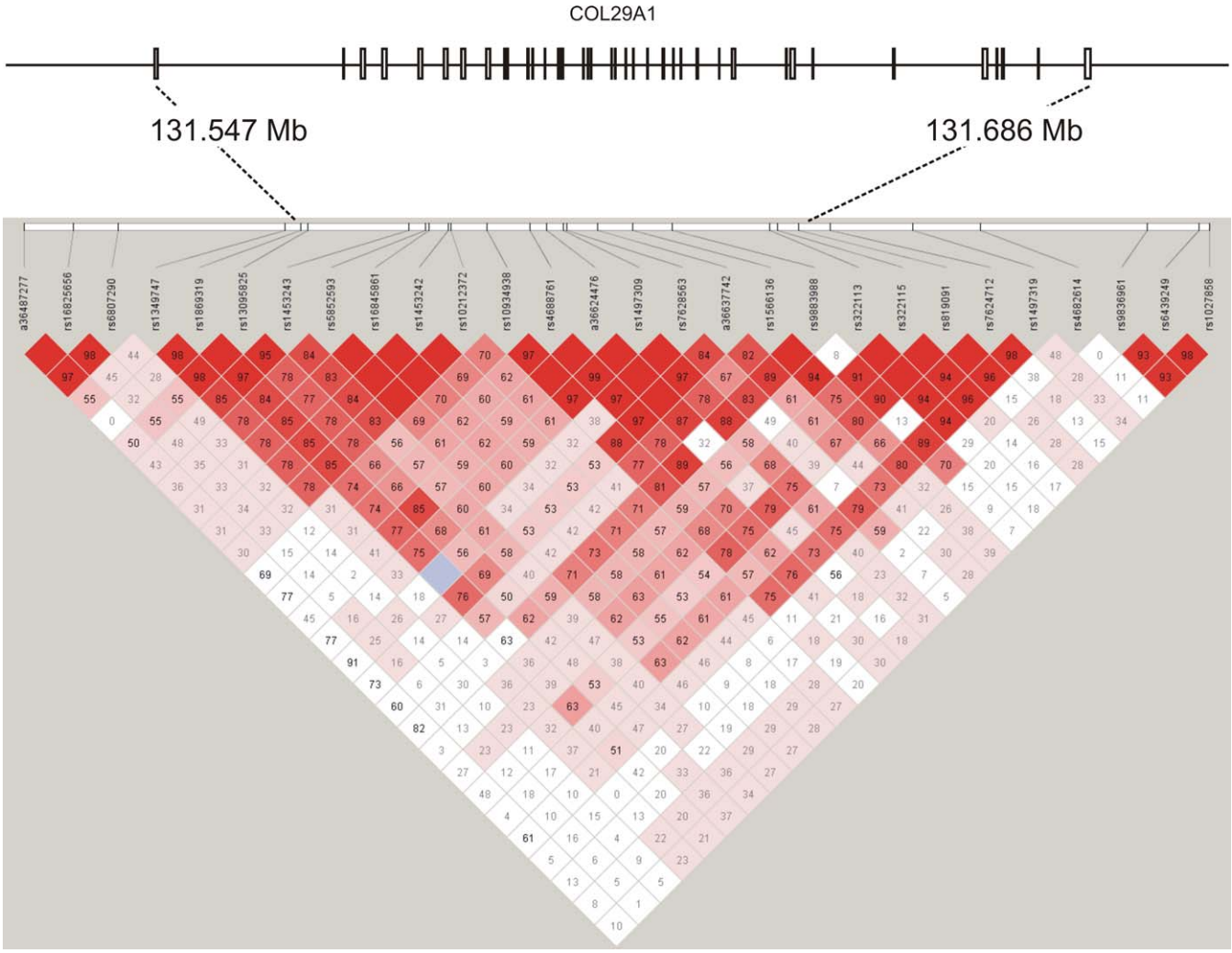


Figure 2. Pairwise LD Values (D′) Between 28 SNPs Based on Genotypes of the Founders in the Discovery Cohort

Boxes contain the LD values (D′) between the respective markers indicated on top. Higher LD values correspond to a darker shade of red. Positions on Chromosome 3 are given in Mb; 131.547 denotes the start and 131.686 the end of *COL29A1* on the genomic sequence. Boxes on the horizontal bar represent the 42 exons of *COL29A1*.
doi:10.1371/journal.pbio.0050242.g002

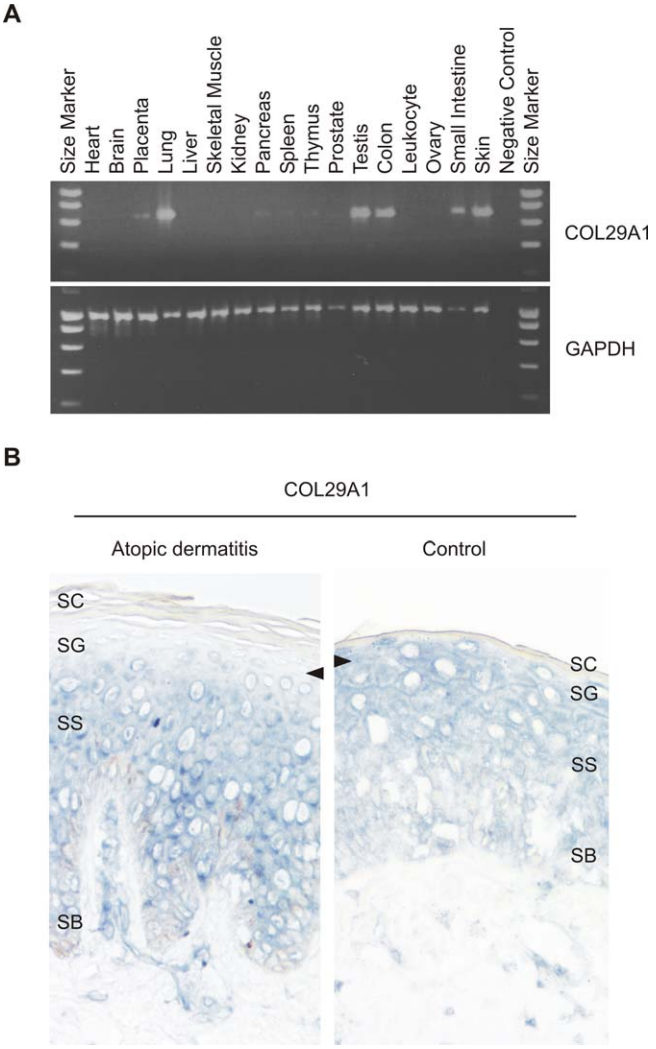


Figure 3. Gene Expression Analysis of *COL29A1*

(A) RT-PCR analysis of collagen XXIX in human tissues. (B) In situ gene expression analysis of *COL29A1* in AD skin. In situ hybridization results of a *COL29A1*-specific antisense probe on cryostat sections (5 μm thick) of an AD skin biopsy (left) and a normal human control (right) are shown. *COL29A1* mRNA detected by the digoxigenin-labelled probe is stained in blue (BCIP/NBT staining). The arrows point the different gene expression in the upper spinous and granular layer of the epidermis of an AD patient and a normal human control. Stratum basale (SB), stratum spinosum (SS), stratum granulosum (SG), and stratum corneum (SC) are indicated.

doi:10.1371/journal.pbio.0050242.g003

(RT)-PCR. These results indicate that while differences in mRNA quantity were not detected, AD patients show a distinct abnormal cellular distribution pattern of *COL29A1* expression in the differentiated outer epidermis.

We generated a polyclonal antibody to visualize the collagen XXIX protein in the skin of five patients with AD and five normal controls, including four and three of each group, respectively, in whom in situ analysis was performed. Consistent with the in situ findings, we observed collagen XXIX staining in the differentiated suprabasal layers of the epidermis in normal human skin and a remarkable absence of staining in the most differentiated upper spinous and granular layers (Figure 4 and Figure S3).

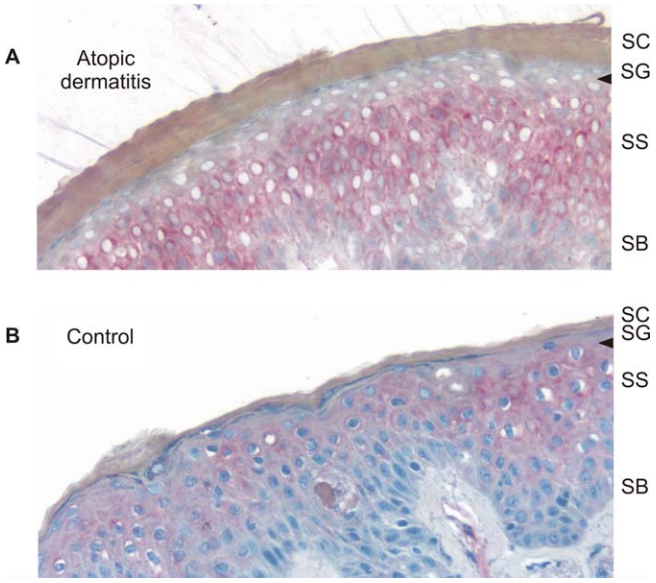


Figure 4. Immunohistochemical Analysis of Collagen XXIX Expression in AD Skin

Cryostat sections (5μm thick) of an AD skin biopsy (A) and a normal human control (B) are shown. In normal, human skin, collagen XXIX is expressed in the epidermis. In patients with atopic dermatitis, a striking lack of collagen XXIX staining was observed in the viable outermost spinous and granular layers of the epidermis (arrow). Stratum basale (SB), stratum spinosum (SS), stratum granulosum (SG), and stratum corneum (SC) are indicated. Collagen XXIX was stained with fuchsin (red). Sections were counterstained with hematoxylin (blue).

doi:10.1371/journal.pbio.0050242.g004

Discussion

In a whole-genome linkage scan for AD, we previously identified a susceptibility locus on human Chromosome 3q21. The candidacy of this chromosomal region was further supported by Bradley et al., who mapped a locus for AD severity in close proximity (3q14) in a Swedish population [13]. By positional cloning, we have now identified the disease-causing gene, *COL29A1,* which encodes a novel epidermal collagen.

The disease gene was located in a two-staged investigation consisting of systematic linkage and association scanning of the region and subsequent confirmation of the association in a large, independent, replication dataset. In the first stage, we genotyped additional microsatellite markers in the candidate region, which narrowed the initial 12.75-cM linkage interval to 5.4 Mb. The association analysis with an average marker distance of 25.5 kb, finally, revealed an association that was confined to a haplotype block of 170 kb, which included a single gene, *COL29A1*. Pairwise LD measures indicated LD across the entire gene and defined two subblocks of increased LD. The strongest association was observed within the 96-kb subblock encoding the collagenous domain and the C terminus of collagen XXIX. A rapid decay of LD at the borders of the *COL29A1* haplotype block and lack of association of the SNPs located within the neighboring genes clearly limited the association to *COL29A1*. In addition, we demonstrated in the discovery dataset that the families that contributed to the association of SNP rs4688761 with AD also accounted for most of the linkage signal. This finding corroborated that variants in *COL29A1* explained the

previously reported linkage of AD to 3q21 [16]. Finally, we confirmed association in a large independent family cohort, making *COL29A1* the first AD susceptibility gene, to our knowledge, that is identified by positional cloning.

Consistent with the linkage analysis, we found a maternal transmission pattern to affected offspring in both family cohorts. Although the sexes were equally represented among the affected children of both cohorts, we observed a marked maternal preponderance in AD status among the parents in both cohorts. This finding clearly supports the notion of a maternal effect. It may, however, also raise the question whether the maternal overtransmission pattern observed for the *COL29A1* haplotype was due to the different prevalence of AD in mothers and fathers. This is unlikely to be the case, because the analytical tools used in this study, nonparametric linkage and TDT, do not take into account the parental phenotype. Furthermore, we showed that the observed maternal effect was not predominantly attributable to transmissions from affected rather than unaffected mothers. Although parent-of-origin effects have not previously been reported for genes from Chromosome 3q21, they have been observed at other loci influencing allergic disease [23]. Tissue-specific imprinting of genes encoding extracellular matrix (ECM) proteins has been reported in the mouse [24], and their disruption has been shown to impair skin structure and function [25]. Interestingly, *COL29A1* is expressed in human placenta, an organ of embryonic origin. Apart from classic genomic imprinting mechanisms, maternal effects may be due to an interaction of the child's genotype with the maternal environment during prenatal life.

Sequencing of the exons and promoter region revealed 19 nonsynonymous coding SNPs, six of which were located within a vWA. In silico analysis of these variants revealed only one rare mutation altering a highly conserved amino acid, but all of them may affect protein–protein interaction. None of the nonsynonymous coding SNPs explained the observed association on its own. It has been demonstrated in other complex diseases that multiple independent variants may occur in a single disease gene [26]. It is therefore possible that several variants or combinations thereof which are associated with the most common haplotype of *COL29A1* are involved in the disease pathogenesis. The functional influence of the associated variants remains to be determined.

Involvement of *COL29A1* in AD is further supported by its tissue- and cell-specific expression pattern. Like *COL29A1*, a growing number of collagens recently identified show a restricted expression pattern. These are not mainly found in fibrous connective tissue, but rather in the ECM of more specialized tissue structures pointing to a distinctive function of these proteins [27]. Highest *COL29A1* expression was observed in the skin, but also in other epithelial tissues such as the lung, small intestine, and colon, which are the main manifestation sites of allergic disorders, including asthma and food allergies. This gene expression pattern might indicate a role of collagen XXIX in a wider spectrum of allergic diseases and suggests a molecular link between AD, respiratory airways disease, and food allergies, which are epidemiologically closely associated [28,29].

In human skin, collagen XXIX was detected throughout the viable layers of the epidermis with an increase toward the differentiated cells of the granular layer. Comparative expression analysis of *COL29A1* by in situ hybridization and immunohistochemistry in skin biopsies of patients with AD and healthy controls revealed a distinct lack of *COL29A1* mRNA and protein in the outer viable layers of the epidermis. The process of epidermal stratification is tightly regulated by an increasing gradient of extracellular Ca^{2+} concentration and a specific special and temporal expression pattern of transcriptional regulators [30]. Our findings indicate that the specific cellular milieu acquired during terminal epidermal differentiation affect the regulation or degradation of *COL29A1* mRNA in the outer epidermis. However, our findings do not allow us to distinguish between these two processes.

Lack of collagen XXIX in the outer epidermis of AD patients indicates that a defective ECM may give rise to the disease, proposing a new pathomechanism for AD. Collagens are the most abundant ECM proteins in vertebrates and play a crucial role in maintaining tissue integrity. Their importance for tissue function has been highlighted by the wide spectrum of human diseases caused by mutations in collagen genes [31]. Although a large number of collagens in the connective tissue–rich dermis have been characterized, little is known about collagens in the ECM of the epidermis [32,33]. Collagen XXIX belongs to the vWA containing collagens. They form filaments with globular domains containing the vWA motifs, which are involved in protein–ligand interactions for the organization of tissue architecture and cell adhesion [34]. It is therefore conceivable that collagen XXIX plays an important role in keratinocyte cohesion. Lack of collagen XXIX may facilitate antigen penetration through the skin, which may explain the association found between *COL29A* variants and allergic sensitization, a common feature in AD patients [35]. Recent findings indicate that structural and functional integrity of the epidermis is a key factor in the development of AD [36] and in the disease progression to allergic airways disease [37].

Furthermore, ECM collagens influence the migration of epidermal antigen-presenting Langerhans cells and T cells [38,39] and may thus play an important role in the initiation and maintenance of cutaneous immune responses. In addition, ECM collagens participate in immune regulation by binding to inhibitory immune receptors [40], rendering collagen XXIX an interesting novel susceptibility gene for AD. Impairment of the immune defense function of the skin is a clinical hallmark of AD. Patients with AD show a diminished resistance against microbes resulting in the colonization of nonlesional skin with *Staphylococcus aureus* in nearly 90% of patients and an increased susceptibility to bacterial and viral skin infections [41]. Our findings led to the identification of collagen XXIX as a novel component of the epidermal ECM and propose a new disease mechanism in the etiology of atopic dermatitis and allergies.

Materials and Methods

Human subjects. The diagnosis of AD was made according to standard criteria, as previously described [16]. Recruitment was restricted to patients with an age of onset below 2 y and moderate to severe disease expression. Total IgE levels and levels of specific IgE against 12 common environmental allergens were determined using ImmunoCAP (Phadia AB; http://www.phadia.com/). Allergic sensitization was defined as either the presence of specific IgE to at least one allergen (detection limit 0.35 kU/l) or a total serum IgE level elevated above the age-specific norm. The institutional review boards of the

participating centers approved the study protocol and informed consent was obtained from all probands or their legal guardians.

One discovery study sample and one replication sample were investigated. All families were of European origin. The discovery data set consisted of 199 complete affected sibling families composed of 427 children with AD that were studied in the original genome scan. [16] The replication set consisted of 292 families including 481 children with AD with an age of onset $\leq$2 y and moderate to severe disease expression. Among AD patients in the discovery and the replication cohorts, the proportion of boys was 52% and 50.9%, and the proportion of children with allergic sensitization was 74% and 72.1%, respectively.

Punch biopsies of human skin were obtained from six patients with AD and 7 healthy donors aged 24–55 y with written informed consent. Specimens were prepared for in situ hybridization and immunohistochemistry as described below.

Microsatellite genotyping. In the first stage, fine mapping with microsatellite markers was performed in the discovery dataset. 96 short tandem repeat markers were selected within the interval between *D3S1303* (126.07 cM) and *D3S1292* (138.82 cM) from the Genethon (http://www.genethon.fr/) and Marshfield (http://research.marshfieldclinic.org/genetics/) databases, or were identified by screening human genome sequence data for short tandem repeats. Fluorescence-based semi-automated genotyping was performed as previously described [16]. Primer sequences, amplification conditions, and allele size are available on request.

Association scan with SNPs. We performed an association scan of the 5.4-Mb region using 212 SNPs at an average distance of 25.5 kb. The SNPs were selected from the NCBI database (http://www.ncbi.nlm.nih.gov/projects/SNP/) to cover known and predicted genes. To determine the allele frequency of the polymorphisms, we amplified 600 to 800 bp surrounding each SNP by PCR for resequencing in 31 unrelated Caucasian individuals. Markers with a minor allele frequency (MAF) of >5% were selected for genotyping. To identify functional variants in collagen XXIX, we sequenced all 42 exons including the exon–intron boundaries and 5.1 kb of the promoter region in 46 unrelated patients with AD and two controls. Sequencing was performed on an ABI3730 DNA sequencer (Applied Biosystems; http://www.appliedbiosystems.com) using standard procedures.

We carried out SNP genotyping using TaqMan allelic discrimination, with probes and primers designed and synthesized by the supplier (Applied Biosystems), or by pyrosequencing using PSQ HS 96A (Pyrosequencing AB; http://www.biotagebio.com/). The averages genotyping success rate was 97.9%. All primer and probe sequences are available on request.

Statistical analysis. We performed linkage analysis of the microsatellite data using Genehunter V. 2.1 [42]. Each SNP was checked for compliance with Hardy-Weinberg equilibrium (HWE) in the parent population using a χ^2 test, and those markers that were not in HWE were excluded from the analysis. We calculated pairwise LD between each marker pair using the D′ statistic.

In view of the strong imprinting effect at our locus, we conducted family-based association tests, because this strategy allows us to determine the parental origin of an associated allele. In the affected sibling families, we used the sib__TDT of the ASPEX software that performs a permutation procedure to calculate empirical *p*-values that are independent of linkage within families [17]. Furthermore, to assess the significance of the maternal effect, we calculated empirical *p*-values for the difference in maternal versus paternal haplotype transmissions using the parent-of-origin TDT implemented in PLINK [18].

The sib__TDT was also used in the analysis of eight markers in the replication dataset. The significance level of the replication results was assessed empirically. Using all pedigrees and all genetic markers used in the actual analysis, we generated 10,000 replicates using Allegro V1.2c [43] and conducted an association analysis as in the original dataset. An empirical significance level was calculated as the proportion of replicates for which the maximum χ^2 score was greater than that obtained in the real dataset. All *p*-values are two-sided, significance was defined as statistical evidence expected to occur 0.05 times at random in the analysis. For multipoint analysis, we used the FBAT tools package [44] to generate haplotypes and performed family-based association tests for five marker haplotypes using the empirical variance option to adjust for correlation among sibling genotypes. To evaluate parent-of-origin effects in the multipoint analysis, we estimated haplotypes using MERLIN, recoded the haplotypes as alleles, and performed the sib__TDT using ASPEX.

In silico protein analysis. Protein sequences were retrieved from the UniProt (http://www.uniprot.org) and Ensembl (http://www.ensembl.org) databases. The domain architecture of the collagen XXIX protein was retrieved from the NCBI conserved domain search website (http://www.ncbi.nlm.nih.gov:80/Structure/cdd/wrpsb.cgi). The following domains were found in collagen XXIX and analyzed further: cd01472 (vWA__collagen), cd01450 (vWFA__subfamily__ECM), cd01470 (vWA__complement__factors), cd01465 (vWA__subgroup), and Pfam01391 (collagen triple helix repeat). To predict the 3D structure of the vWF protein domains in collagen XXIX, we explored the structure prediction results returned by the web servers GenTHREADER (http://bioinf.cs.ucl.ac.uk/psipred/) and FFAS03 (http://ffas.ljcrf.edu). Based on their very similar predictions, the human vWF A3 domain was chosen as the structural template to analyze the position of *COL29A1* SNPs in 3D and to predict their potential effect on protein function.

Exon identification and characterization of the collagen XXIX gene. A 2447-bp sequence from a human testis cDNA library which covered eight exons was the starting point for the characterization of *COL29A1*. Using rapid amplification of cDNA ends together with the Model Maker of NCBI, we identified a total of 42 exons and determined the transcription start site as well as the 3′ end of the *COL29A1* transcript in cDNA from human skin. The complete sequence of the transcript was confirmed by cloning and sequencing of the full-length cDNA. To explore the potential gene function of *COL29A1*, the protein sequence was predicted (http://us.expasy.org/tools/dna.html), a domain search was performed (http://www.sanger.ac.uk/Software/Pfam/search.shtml) [45], and the presence and location of signalling peptides was analyzed (http://www.cbs.dtu.dk/services/SignalP/) [46].

Gene expression analysis. We examined tissue-specific expression of *COL29A1* using oligo(dT)-primed cDNA of 17 different human tissues. cDNA samples of 16 tissues were from the human MTC Panels I and II, which are standardized for the expression of GAPDH (BD Biosciences; http://www.bdbiosciences.com). In addition, human skin poly(A)+ RNA (BD Biosciences) was transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics; http://www.roche.com). PCR was performed using *COL29A1*-specific primers 5′-GTTCTAACCAGAATGTATAATCATC (forward) and 5′-TAATTCCCAAGAACATCTCTGGT (reverse), yielding a product of 694 bp, and the GAPDH control primers supplied with the MTC panels.

For in situ hybridization, we generated a plasmid by cloning a *COL29A1*-specific PCR product amplified from human skin cDNA [5′-ACCTTAGGAGACAGGGTTGCT (forward); 5′-AGTTCCAATCTGGCTTGTGG (reverse)] into the pCRII vector (Invitrogen; http://www.invitrogen.com). We synthesized antisense and sense digoxigenin-labeled riboprobes using the Dig RNA Labeling Kit (Roche Diagnostics). Punch biopsies of human skin were obtained from five AD patients and five healthy donors, immediately fixed in 4% paraformaldehyde for 4 h, cryoprotected in 30% sucrose overnight, and embedded in Tissue-Tek (Sakura; http://www.sakura.com) for cryosectioning. 10 μm cryosections were mounted on slides and dried for 15 min at 50 °C. Sections were postfixed in 4% paraformaldehyde for 7 min at 4 °C and acetylated with 0.25% acetic acid for 10 min. Sections were prehybridized for 3 h and hybridized overnight at 50 °C with digoxigenin-labeled riboprobes. After hybridization, sections were washed twice with 2 × SSC at 53 °C and once with 0.1 × SSC at 58 °C. For detection of the hybridized probe, slides were incubated with BCIP/NBT staining solution (Roche Diagnostics) for 4 d according to the manufacturer's recommendations.

To quantify gene expression in skin specimens, total RNA was isolated from 160 μm cryosections of skin biopsies using the RNeasy Mini Kit (Qiagen; http://www.qiagen.com). RNA was reverse transcribed into cDNA with random hexamer primers using the Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics). Taqman real-time PCR was performed with iTaq SYBR Green (BioRad; http://www.biorad.com) and gene-specific PCR products were detected on the ABI PRISM 7900 sequence detection system (Applied Biosystems). All measurements were performed in duplicate. *COL29A1* expression was normalized for 18S rRNA expression. Differences in gene expression were calculated using the ΔΔct method and were expressed as fold change. Gene-specific primers were as follows; *COL29A1*-forward, 5′-CCACCCTCTGGATCATCACT, *COL29A1*-reverse, 5′-GTTTTCTGTGCCACCGTTCT, KRT10-forward, 5′-CTGAAACCGAGTGCCAGAAT, KRT10-reverse, 5′-GCCTCCGGAACTTCCCTCT, 18S rRNA-forward, 5′-GGATGCGTGCATTTATCAGA, 18S rRNA-reverse, 5′-GATCAGCCCGAGGTTATCTA. The sizes of the PCR products were confirmed by gel electrophoresis and the specificity of the reaction was confirmed by melting curve analysis of the PCR products. For statistical analysis, the unpaired *t*-test was used.

Antibodies and immunohistochemistry. A polyclonal antibody against human collagen XXIX protein was raised by immunizing

rabbits with a collagen XXIX specific peptide (SLGSTRKDDMEELAC, residues 2115–2128) (Eurogentec; http://www.eurogentec.com). The specificity of the antibodies purified by affinity chromatography was tested by comparing their reactivity against recombinant proteins by Western blotting and by blocking experiments.

For immunohistochemistry, freshly isolated skin specimens from five AD patients and five healthy individuals were embedded in Tissue-Tek. Cryosections of 5 μm thickness were prepared and fixed with acetone. Sections were incubated with anti-collagen XXIX antibodies followed by dextran-coupled anti-rabbit antibody, an alkaline phosphatase labelled amplification polymer (DAKO EnVision System; http://www.dako.com) and detection with fuchsin (DAKO). Nuclei were counterstained with Mayer's hematoxylin solution (Sigma-Aldrich; http://www.sigmaaldrich.com). Omission of primary antibody and preincubation with equimolar amount of peptide used for generation of the antibody in rabbits were used as negative controls for parallel sections. The results were consistent among the AD patients on one hand, and among the control biopsies on the other.

Supporting Information

Figure S1. Hapmap LD Data for the *COL29A1* Region on Human Chromosome 3

LD plot for the CEPH population (Utah residents with ancestry from northern and western Europe) is shown [19]. The intensity of the color of the red boxes represents the pairwise LD values (r^2) between two markers. Positions of the markers genotyped by the HapMap Project are indicated by vertical dashes on the horizontal bar. The scale indicates the physical position on Chromosome 3 in kb. Gene loci were taken from the human genome NCBI build 36.2 and correspond to the largest mRNA identified (FLJ35880 and LOC131873) or were predicted (LOC646300).

Found at doi:10.1371/journal.pbio.0050242.sg001 (1.3 MB TIF).

Figure S2. Gene Expression Analysis

Gene expression analysis of *COL29A1* in the epidermis of AD patients (A) and of non-affected individuals (B). Cryosections (5 μm thick) of skin biopsies of five AD patients (AD1 to AD5) and of five normal human control individuals (NC1 to NC5) were hybridized with a digoxigenin-labeled antisense probe specific for *COL29A1* mRNA (on the left) and with a digoxigenin-labeled sense probe as a negative control (on the right). Staining of the detected transcripts was performed with the BCIP/NBT staining solution (blue). No staining was observed in the upper spinous and granular layer (indicated by arrows) of AD patients. The mRNA expression pattern was in good agreement with the results of the immunohistochemical staining of the protein. Sections were photographed at 200×magnification (AD5, 100×).

Found at doi:10.1371/journal.pbio.0050242.sg002 (15 MB TIF).

Figure S3. Immunohistochemical Analysis

Immunohistochemical analysis of collagen XXIX expression in the epidermis of AD patients (A) and of non-affected individuals (B). Cryosections (5 μm thick) of skin biopsies of five AD patients (AD1–AD4, AD6) and of five normal human control individuals (NC1–NC3, NC6, NC7) were incubated with a rabbit polyclonal antibody against ColXXIX epitope and stained with fuchsin (red). Counterstaining was performed with hematoxylin (blue). Negative controls (on the right) were lacking the primary antibody. In normal, human skin, collagen XXIX is expressed in the epidermis with a slight increase towards the more differentiated layers. AD patients lack the protein in the upper spinous and granular layer as indicated by arrows. Sections were photographed at 200× magnification (AD5, 100×).

Found at doi:10.1371/journal.pbio.0050242.sg003 (18 MB TIF).

Table S1. SNPs Used in the Association Scan in the 5.4-Mb Region Between Markers *M3CS075* and *M3CS233*

The average typing rate was $\geq$ 97.9%

Found at doi:10.1371/journal.pbio.0050242.st001 (564 KB DOC).

Table S2. Nonsynonymous Coding SNPs in *COL29A1*

vWA, von Willebrand factor domain; Col, collagen domain

Found at doi:10.1371/journal.pbio.0050242.st002 (113 KB DOC).

Table S3. Results of the ASPEX sib__TDT for SNPs in the Region of *COL29A1* for Allergic Sensitization in the Discovery Cohort

Found at doi:10.1371/journal.pbio.0050242.st003 (112 KB DOC).

Table S4. Association Analysis of Single Markers for Allergic Sensitization in the Replication Dataset

Found at doi:10.1371/journal.pbio.0050242.st004 (58 KB DOC).

Table S5. Haplotype Analysis for Allergic Sensitization in the Replication Dataset

Found at doi:10.1371/journal.pbio.0050242.st005 (49 KB DOC).

Accession Numbers

The GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) accession numbers for genes and proteins discussed in this paper are: AK093199 *(FLJ35880)*, NM 153264 *(FLJ35880)*, and NP 476507 (human collagen VI alpha 3 chain),

The Protein Databank (http://www.pdb.com) accession number for the human vWF A3 domain is 1atz.

Acknowledgments

We thank the patients with atopic dermatitis and their families for their participation; Monika Schwarz, Jenny Pech, Christina Flachmeier, and Sabine Enigk for technical assistance; and Dr. Sylke Schneider and Dr. Claudia Rasche for performing the skin biopsies.

Author contributions. YAL designed the experiments and directed the study. CS, IM, TK, JEG, and NH performed the experiments. FR, KR, and HS performed the statistical analyses. GM and MA performed the in silico protein analysis. CG, UW, MW, and YAL were responsible for family recruitment, clinical phenotyping, and taking of the human skin biopsies. CS, IM, NH, and YAL wrote the paper.

Funding. This work was supported by grants from the German Ministry for Education and Research through the Clinical Research Group for Allergy and the German National Genome Research Network "Environmental Diseases" and the "DNA Platform" to YAL and NH. Jorge Esparza-Gordillo is supported by an EMBO long-term fellowship (LTF 522-2006).

Competing interests. The authors have declared that no competing interests exist.

References

1. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee (1998) Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. Lancet 351: 1225–1232.
2. Kay J, Gawkrodger DJ, Mortimer MJ, Jaron AG (1994) The prevalence of childhood atopic eczema in a general population. J Am Acad Dermatol 30: 35–39.
3. Taylor B, Wadsworth J, Wadsworth M, Peckham C (1984) Changes in the reported prevalence of childhood eczema since the 1939–45 war. Lancet 2: 1255–1257.
4. Rajka G (1960) Prurigo Besnier (atopic dermatitis) with special reference to the role of allergic factors. Acta Derm Venereol (Stockh) 40: 285–306.
5. Wahn U (2000) What drives the allergic march? Allergy 55: 591–599.
6. Cooke RA, Vander Veer A (1916) Human sensitization. J Immunol 1: 201–305.
7. Zutavern A, Hirsch T, Leupold W, Weiland S, Keil U, et al. (2005) Atopic dermatitis, extrinsic atopic dermatitis and the hygiene hypothesis: results from a cross-sectional study. Clin Exp Allergy 35: 1301–1308.
8. Larsen FS, Holm NV, Henningsen K (1986) Atopic dermatitis. A genetic-epidemiologic study in a population-based twin sample. J Am Acad Dermatol 15: 487–494.
9. Schultz Larsen F (1993) Atopic dermatitis: A genetic-epidemiologic study in a population-based twin sample. J Am Acad Dermatol 28: 719–723.
10. Chavanas S, Bodemer C, Rochat A, Hamel-Teillac D, Ali M, et al. (2000) Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. Nat Genet 25: 141–142.
11. Wilkins JF, Haig D (2003) What good is genomic imprinting: The function of parent-specific gene expression. Nat Rev Genet 4: 359–368.
12. Cookson WO, Ubhi B, Lawrence R, Abecasis GR, Walley AJ, et al. (2001) Genetic linkage of childhood atopic dermatitis to psoriasis susceptibility loci. Nat Genet 27: 372–373.
13. Bradley M, Soderhall C, Luthman H, Wahlgren CF, Kockum I, et al. (2002) Susceptibility loci for atopic dermatitis on chromosomes 3, 13, 15, 17 and 18 in a Swedish population. Hum Mol Genet 11: 1539–1548.
14. Haagerup A, Bjerke T, Schiotz PO, Dahl R, Binderup HG, et al. (2004)

Atopic dermatitis—a total genome-scan for susceptibility genes. Acta Derm Venereol 84: 346–352.
15. Morar N, Willis-Owen SA, Moffatt MF, Cookson WO (2006) The genetics of atopic dermatitis. J Allergy Clin Immunol 118: 24–34.
16. Lee YA, Wahn U, Kehrt R, Tarani L, Businco L, et al. (2000) A major susceptibility locus for atopic dermatitis maps to chromosome 3q21. Nat Genet 26: 470–473.
17. Hinds DA, Risch N (1996) The ASPEX package: Affected sib-pair exclusion mapping. Available: http://aspex.sourceforge.net/. Accessed 31 July 2007.
18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A tool set for whole genome association and population-based linkage analyses. Am J Hum Genet. E-pub ahead of print.
19. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299–1320.
20. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795.
21. Persikov AV, Ramshaw JA, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. J Biol Chem 280: 19343–19349.
22. Pan TC, Zhang RZ, Pericak-Vance MA, Tandan R, Fries T, et al. (1998) Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. Hum Mol Genet 7: 807–812.
23. Cox HE, Moffatt MF, Faux JA, Walley AJ, Coleman R, et al. (1998) Association of atopic dermatitis to the beta subunit of the high affinity immunoglobulin E receptor. Br J Dermatol 138: 182–187.
24. Mizuno Y, Sotomaru Y, Katsuzawa Y, Kono T, Meguro M, et al. (2002) Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. Biochem Biophys Res Commun 290: 1499–1505.
25. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, et al. (1997) Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. J Cell Biol 136: 729–743.
26. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, et al. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411: 599–603.
27. Cookson W (2004) The immunogenetics of asthma and eczema: A new focus on the epithelium. Nat Rev Immunol 4: 978–988.
28. Spergel JM, Paller AS (2003) Atopic dermatitis and the atopic march. J Allergy Clin Immunol 112: S118–S127.
29. Sicherer SH, Sampson HA (1999) Food hypersensitivity and atopic dermatitis: Pathophysiology, epidemiology, diagnosis, and management. J Allergy Clin Immunol 104: S114–S122.
30. Koster MI, Roop DR (2007) Mechanisms regulating epithelial stratification. Annu Rev Cell Dev Biol
31. Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20: 33–43.
32. Edfors-Lubs ML (1971) Allergy in 7000 twin pairs. Acta Allergol 26: 249–285.
33. Koch M, Veit G, Stricker S, Bhatt P, Kutsch S, et al. (2006) Expression of type XXIII collagen mRNA and protein. J Biol Chem 281: 21546–21557.
34. Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: Widely dispersed domains with roles in cell adhesion and elsewhere. Mol Biol Cell 13: 3369–3387.
35. Leung DY, Boguniewicz M, Howell MD, Nomura I, Hamid QA (2004) New insights into atopic dermatitis. J Clin Invest 113: 651–657.
36. Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, et al. (2006) Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet 38: 441–446.
37. Marenholz I, Nickel R, Ruschendorf F, Schulz F, Esparza-Gordillo J, et al. (2006) Filaggrin loss-of-function mutations predispose to phenotypes involved in the atopic march. J Allergy Clin Immunol 118: 866–871.
38. Gunzer M, Friedl P, Niggemann B, Brocker EB, Kampgen E, et al. (2000) Migration of dendritic cells within 3-D collagen lattices is dependent on tissue origin, state of maturation, and matrix structure and is maintained by proinflammatory cytokines. J Leukoc Biol 67: 622–629.
39. Dustin ML, de Fougerolles AR (2001) Reprogramming T cells: The role of extracellular matrix in coordination of T cell activation and migration. Curr Opin Immunol 13: 286–290.
40. Lebbink RJ, de Ruiter T, Adelmeijer J, Brenkman AB, van Helvoort JM, et al. (2006) Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. J Exp Med 203: 1419–1425.
41. Mempel M, Lina G, Hojka M, Schnopp C, Seidl HP, et al. (2003) High prevalence of superantigens associated with the egc locus in *Staphylococcus aureus* isolates from patients with atopic eczema. Eur J Clin Microbiol Infect Dis 22: 306–309.
42. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: A unified multipoint approach. Am J Hum Genet 58: 1347–1363.
43. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nature Genet 25: 12–13.
44. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, et al. (2004) Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 26: 61–69.
45. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al. (2004) The Pfam protein families database. Nucleic Acids Res 32: D138–D141.
46. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005–1016.

Figure S1

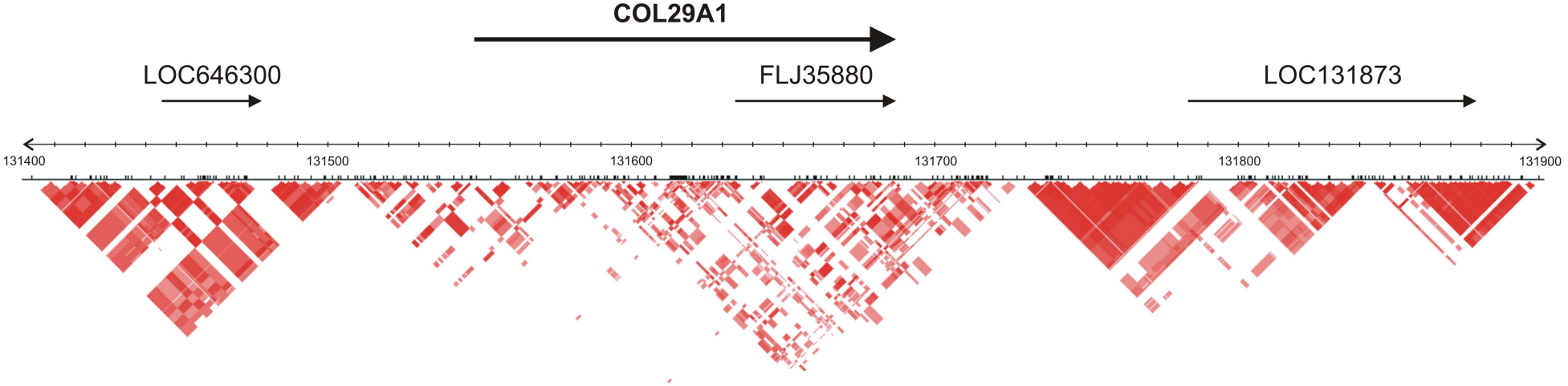

COL29A1
LOC646300
FLJ35880
LOC131873
131400
131500
131600
131700
131800
131900

Figure S2

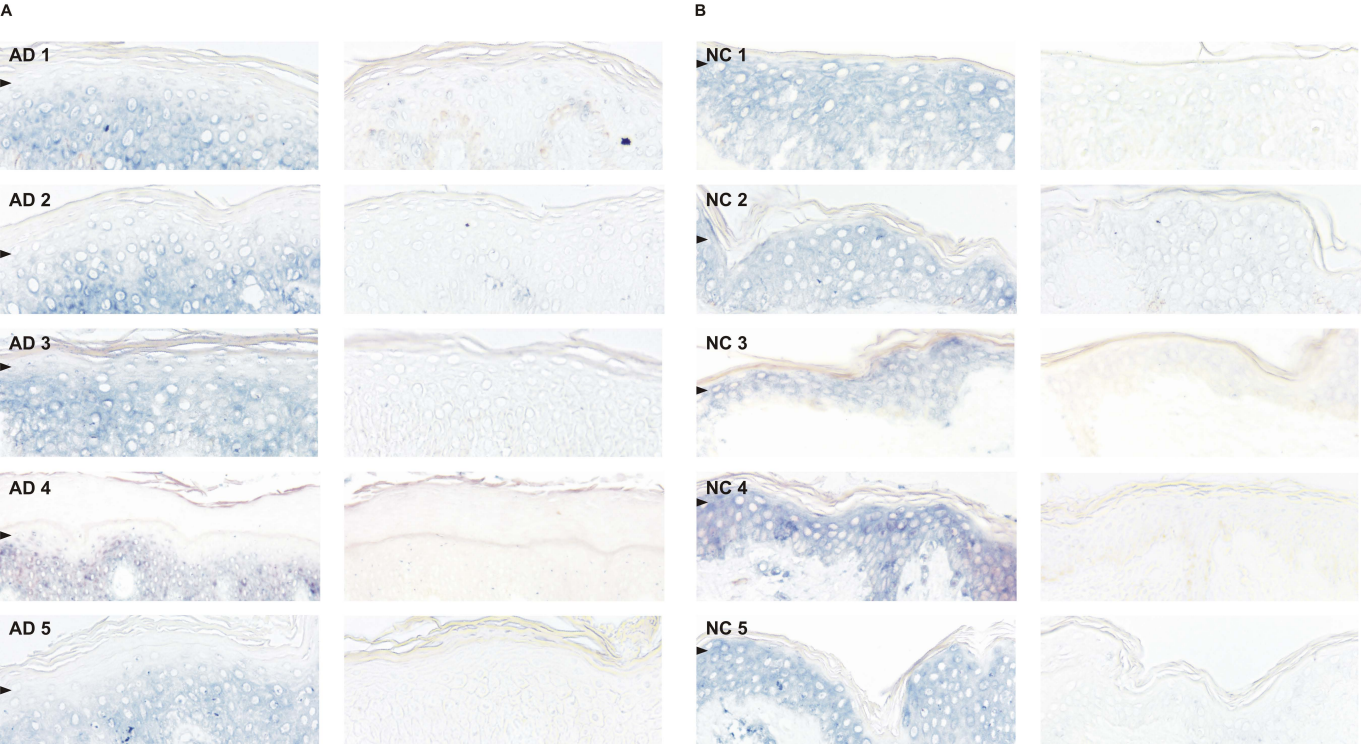
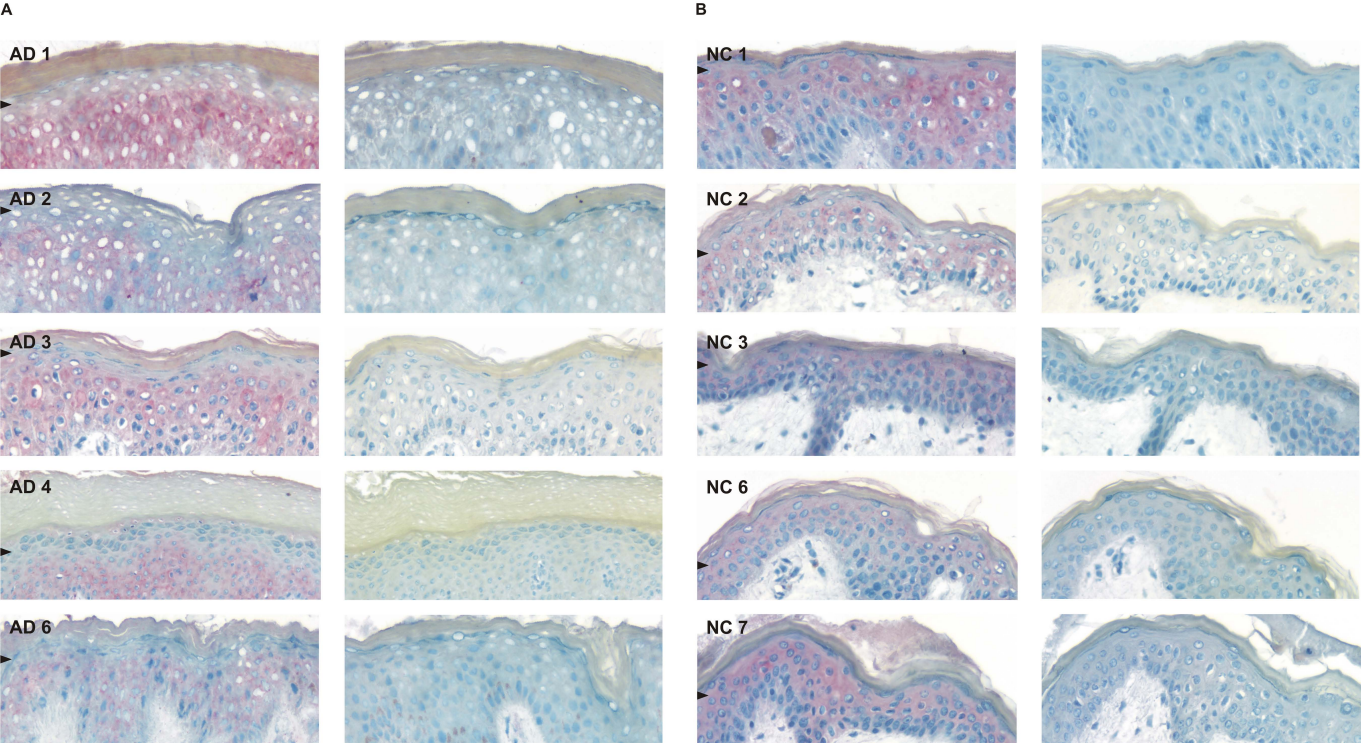


Figure S3



Supporting Material

Table S1

SNPs used in the association scan in the 5,4 Mb region between markers M3CS075 and M3CS233. The average typing rate was ≥ 97,9%.

	SNP	Chromosome 3 NCBI Build 36.1	MAF	Major/ Minor allele	Atopic dermatitis			Allergic sensitization		
					All P value	Paternal P value	Maternal P value	Paternal P value	Maternal P value	All P value
1	rs1532602	126,107,810	42.7	C/T	0.97	0.21	0.33	0.95	0.34	0.3
2	rs2270780	126,114,817	22.7	A/G	0.42	0.23	0.82	0.57	0.22	1
3	rs2332924	126,127,555	24	A/G	0.35	0.039	0.68	0.71	0.11	0.31
4	pos205149	126,277,468	9.4	C/G	0.22	0.65	0.29	0.27	0.64	0.36
5	rs2333034	126,277,567	25.3	G/T	0.068	0.12	0.15	0.099	0.068	0.32
6	rs2076729	126,278,274	35.5	T/C	0.13	0.71	0.025	0.31	0.92	0.089
7	rs2981482	126,285,578	47	A/G	0.02	0.043	0.065	0.051	0.073	0.17
8	rs651630	126,294,004	43.7	A/G	0.67	0.3	0.9	0.4	0.57	0.38
9	rs2076727	126,299,784	47.2	A/G	0.12	0.11	0.36	0.25	0.12	0.74
10	rs13325275	126,303,160	8.3	G/A	0.054	0.2	0.079	0.17	0.62	0.11
11	rs12487908	126,313,738	15.7	C/T	0.58	0.43	0.8	0.81	0.32	0.73
12	rs591803	126,318,469	29.4	G/A	0.92	0.55	0.79	1	0.24	0.33
13	rs1554241	126,323,002	37.2	T/A	0.073	0.18	0.091	0.25	0.41	0.29
14	rs12492332	126,324,170	16.8	C/T	0.42	0.38	0.72	0.64	0.31	0.85
15	rs1971408	126,328,091	34.5	T/A	0.43	0.2	0.83	0.67	0.26	0.94
16	rs516355	126,330,592	50	A/C	0.57	0.85	0.45	1	0.86	0.83
17	rs2228677	126,331,445	48.4	A/G	0.89	0.69	0.49	0.48	1	0.31
18	rs702045	126,332,460	10.3	A/G	0.67	0.32	0.73	0.65	0.18	0.53

19	rs2981503	126,337,326	35.8	T/A	0.79	0.41	0.86	1	0.6	0.69
20	rs504255	126,338,381	20.5	C/T	0.45	0.26	0.86	0.86	0.21	0.44
21	rs531740	126,339,080	19.8	A/G	0.87	0.7	1	1	0.44	0.55
22	rs6765325	126,345,375	18.2	C/T	0.55	0.44	0.79	0.68	0.35	0.84
23	rs631691	126,357,868	50	T/C	0.48	0.57	0.098	0.37	0.6	0.063
24	rs1398759	126,371,595	38.4	G/C	0.29	0.51	0.21	0.3	0.77	0.093
25	rs2993631	126,379,358	5.3	G/A	0.5	0.32	0.9	0.3	0.23	0.46
26	a304958	126,389,178	19.4	A/T	0.36	0.82	0.12	0.084	0.55	0.042
27	c171945	126,391,761	1.3	C/T	0.18	0.26	0.62	0.25	0.31	0.61
28	rs1980079	126,395,237	38.9	A/G	0.91	0.95	0.73	0.14	0.69	0.052
29	rs2137599	126,397,335	38.9	A/G	0.87	1	0.76	0.13	0.64	0.05
30	a1486438	126,407,361	11	A/C	0.58	0.49	0.81	0.23	0.46	0.3
31	rs6778311	126,410,481	40.4	T/C	0.7	0.56	0.83	0.38	0.53	0.32
32	c194270	126,413,098	44	G/A	0.97	0.48	0.56	0.36	0.22	0.62
33	rs655465	126,414,557	43.7	A/C	0.55	0.77	0.2	1	0.47	0.51
34	rs1053738	126,434,511	24.1	T/C	0.18	0.21	0.41	0.055	0.073	0.24
35	rs1850506	126,448,618	24.1	C/T	0.51	0.49	0.62	0.25	0.29	0.52
36	pos400840	126,473,159	10.5	T/C	0.2	0.62	0.076	0.51	0.6	0.4
37	rs1999921	126,473,528	24.4	A/G	0.34	0.45	0.52	0.13	0.13	0.39
38	rs7615139	126,476,877	24.2	T/C	0.41	0.46	0.61	0.16	0.14	0.46
39	rs1975760	126,513,071	24.6	C/T	0.39	0.45	0.48	0.24	0.22	0.47
40	a1486427	126,522,281	31.1	C/A	0.15	0.33	0.2	0.39	0.76	0.27
41	rs4679167	126,525,957	24.2	C/A	0.41	0.46	0.51	0.16	0.16	0.41
42	rs12485473	126,532,626	40.4	C/T	0.14	0.23	0.22	0.3	0.15	0.82
43	rs2076725	126,604,771	36.4	T/C	0.91	0.31	0.61	0.46	0.15	0.93
44	rs1839063	126,726,432	18.7	G/T	0.3	0.78	0.16	0.48	1	0.21
45	rs2294053	126,732,030	9.3	G/A	0.32	0.069	0.77	0.35	0.017	0.52
46	rs2922172	126,770,206	23.3	A/G	0.18	0.3	0.21	0.25	0.29	0.42

47	pos698044	126,770,363	9	G/A	0.75	0.15	0.41	0.79	0.049	0.31
48	pos708682	126,781,001	14.1	C/T	0.3	0.32	0.59	0.78	0.45	0.84
49	rs3829007	126,781,202	9.2	T/C	0.22	0.044	0.92	0.21	0.024	0.91
50	rs2979306	126,794,820	13.7	C/A	1	0.85	0.76	0.82	0.81	0.83
51	rs2976713	126,882,374	13.1	A/C	0.8	0.61	0.83	1	0.64	0.71
52	rs12633485	127,088,267	17.9	C/T	0.36	0.52	0.3	0.12	0.16	0.2
53	rs10934738	127,162,739	40.2	G/C	0.73	0.85	0.62	0.77	0.53	0.32
54	rs11925047	127,181,634	27	G/A	0.54	0.5	0.65	0.46	0.68	0.34
55	rs11543284	127,208,145	10	G/C	0.21	0.44	0.19	0.38	0.78	0.32
56	rs1044215	127,208,249	20.2	C/G	0.054	0.39	0.034	0.01	0.2	0.01
57	rs2279820	127,216,962	26.7	G/A	1	0.94	0.93	0.72	1	0.49
58	rs2279821	127,216,998	20	G/A	0.074	0.34	0.059	0.013	0.15	0.011
59	rs868149	127,233,547	25.8	C/A	0.79	0.55	0.89	0.49	0.54	0.66
60	rs1503073	127,241,922	27	G/C	0.69	0.72	0.69	0.65	0.94	0.49
61	rs4077563	127,252,447	32.2	C/T	0.77	0.8	0.46	0.73	0.5	0.93
62	rs4449265	127,262,342	31.8	G/A	0.45	0.63	0.45	0.69	0.85	0.45
63	rs4493383	127,266,380	33.2	T/C	0.84	0.71	1	0.92	1	0.82
64	c32184851	127,268,005	32.6	A/G	0.84	0.67	1	1	0.86	0.81
65	c32197449	127,280,603	28.6	G/A	0.77	0.59	0.94	0.46	0.77	0.3
66	rs4646760	127,305,561	39.7	A/G	0.56	0.77	0.22	0.49	0.87	0.19
67	rs2276729	127,308,904	19.2	G/A	0.21	0.5	0.17	0.047	0.22	0.066
68	rs2305230	127,339,385	15.5	C/A	0.094	0.12	0.3	0.58	0.41	0.82
69	rs1868121	127,388,343	36.2	C/T	0.11	0.2	0.15	0.32	0.33	0.47
70	rs747037	127,475,849	6.1	C/A	0.58	1	0.28	0.74	0.53	0.19
71	pos1473510	127,545,819	12.7	C/G	0.68	1	0.46	0.95	0.9	0.74
72	rs777485	127,689,359	26.9	G/C	0.68	0.29	0.73	0.87	0.73	0.4
73	rs777484	127,689,405	47.7	A/C	0.09	0.14	0.16	0.3	0.42	0.34
74	rs1056522	127,744,035	29.9	G/A	0.53	0.64	0.47	0.94	0.64	0.52

75	rs3732534	127,823,243	37.5	A/G	0.1	0.032	0.55	0.11	0.0033	0.95
76	rs4679128	127,863,494	22.5	G/A	0.096	0.024	0.71	0.59	0.071	0.42
77	rs2053820	127,977,996	34	T/C	0.27	0.54	0.21	0.42	0.45	0.53
78	rs1014081	127,982,798	33.9	G/A	0.16	0.45	0.12	0.53	0.64	0.55
79	rs17693923	127,999,896	23.8	T/G	0.14	0.089	0.55	0.78	0.49	0.79
80	rs1542234	128,034,762	9.7	G/A	1	0.63	0.61	0.31	0.56	0.25
81	rs11714878	128,056,685	13.1	T/C	0.31	0.13	0.93	0.079	0.018	0.51
82	rs7633939	128,091,288	27.6	A/G	0.26	0.41	0.33	0.47	0.55	0.63
83	rs1008175	128,102,704	18.1	G/A	0.82	0.42	0.68	0.31	0.93	0.079
84	rs9683346	128,116,700	32.2	A/G	0.21	0.33	0.29	0.47	0.48	0.7
85	rs734506	128,127,236	32	A/G	0.19	0.27	0.32	0.51	0.47	0.75
86	rs4679309	128,148,697	31.2	A/G	0.17	0.35	0.18	0.42	0.38	0.62
87	rs6777746	128,160,859	13.2	T/C	0.45	0.23	0.94	0.097	0.033	0.45
88	rs9864643	128,186,164	8.6	A/G	0.63	0.36	0.91	0.56	0.77	0.49
89	rs1347003	128,214,737	27.6	A/G	0.25	0.3	0.44	0.53	0.64	0.6
90	rs2241675	128,231,707	8.4	T/C	0.94	0.75	1	0.16	0.22	0.34
91	rs900422	128,350,249	31.1	T/C	0.28	0.12	0.82	0.46	0.31	0.75
92	rs977683	128,396,814	26.2	C/G	0.63	0.32	0.95	0.68	0.84	0.4
93	rs6776267	128,501,229	40.1	G/A	0.097	0.063	0.41	0.28	0.14	0.61
94	rs6776659	128,501,436	21.3	G/A	0.1	0.026	0.63	0.39	0.1	1
95	rs6789659	128,501,526	18	A/T	0.58	0.8	0.41	0.14	0.55	0.088
96	rs1001942	128,614,211	34.3	A/C	0.88	0.88	0.91	0.63	0.73	0.58
97	rs1072711	128,670,101	21.3	G/A	0.95	1	0.83	0.45	0.28	0.72
98	rs2720251	128,703,677	17	T/C	0.68	0.26	0.7	0.74	0.36	0.79
99	rs2720244	128,709,235	17	T/C	0.71	0.18	0.64	0.74	0.32	0.73
100	rs2720240	128,730,064	16.7	A/G	0.64	0.2	0.63	0.73	0.33	0.73
101	rs768853	128,747,382	47.1	A/G	0.14	0.11	0.42	0.081	0.15	0.12
102	rs768854	128,747,599	24.5	C/T	0.76	0.71	0.38	0.6	0.87	0.27

103	rs893293	128,806,520	25.5	T/C	0.27	0.94	0.1	0.75	0.5	0.2
104	rs782454	128,875,420	24.6	A/G	0.43	0.89	0.16	0.85	0.42	0.34
105	rs1658037	128,878,910	24.8	C/T	0.24	0.81	0.075	0.71	0.43	0.21
106	rs782456	128,880,866	25	G/A	0.13	0.57	0.048	0.38	0.85	0.16
107	rs782441	128,895,263	20.8	C/T	1	0.59	0.59	0.6	0.85	0.27
108	rs782440	128,898,832	45.9	C/T	0.58	0.21	0.74	1	0.49	0.55
109	rs782438	128,899,738	36	C/T	0.34	0.26	0.69	0.96	0.62	0.77
110	rs782451	128,910,158	46.3	C/G	0.62	0.26	0.73	1	0.47	0.53
111	rs1437553	128,910,282	41.8	C/A	0.88	0.48	0.71	1	0.62	0.64
112	rs1010174	128,912,257	32.3	C/G	0.33	0.57	0.32	0.72	0.79	0.34
113	rs3821581	128,912,615	42.3	G/A	0.84	0.78	0.56	0.54	0.73	0.48
114	rs3773131	128,912,668	9.3	C/G	0.61	0.27	0.83	1	0.66	0.68
115	rs3773135	128,914,091	41.6	C/T	1	0.73	0.72	0.77	1	0.57
116	rs3773136	128,914,119	10	T/A	0.79	0.22	0.35	1	0.33	0.36
117	rs3773137	128,914,333	32.4	G/A	0.47	0.73	0.42	0.92	0.51	0.76
118	rs597956	128,915,249	10.4	G/A	1	0.4	0.39	1	0.54	0.53
119	rs534018	128,916,232	47.7	C/T	0.81	0.64	0.29	0.64	1	0.44
120	rs477554	128,917,546	47.3	C/T	1	0.6	0.55	0.85	0.94	0.68
121	rs479504	128,917,790	19.7	G/T	1	1	0.92	0.71	0.73	0.62
122	rs3773146	128,917,881	43.2	C/T	0.76	0.32	0.6	1	0.55	0.48
123	rs3773154	128,921,215	17.9	C/T	0.63	0.89	0.53	0.42	0.42	0.63
124	rs3773155	128,921,218	11.2	A/G	1	0.46	0.6	0.92	0.38	0.64
125	rs9811688	128,921,577	32	T/C	1	0.79	0.72	0.57	0.47	0.88
126	rs526862	128,922,071	45.7	A/G	1	0.31	0.27	0.61	0.88	0.32
127	rs555183	128,922,832	45.8	A/G	0.96	0.41	0.31	0.83	0.72	0.49
128	rs1466571	128,936,220	32.1	G/A	0.66	0.74	0.61	0.96	0.71	0.86
129	rs1046821	128,945,878	41.4	C/T	0.81	0.89	0.62	0.57	0.31	0.94
130	rs11915484	128,956,517	17.1	G/A	0.35	0.76	0.24	0.36	0.53	0.45

131	rs664910	128,956,720	28.3	A/G	0.38	1	0.15	0.75	0.77	0.36
132	pos480745	128,964,601	12.7	G/T	0.38	1	0.12	0.29	0.69	0.16
133	rs549662	128,964,726	17.1	A/G	0.49	0.37	0.71	0.83	0.74	0.91
134	rs541042	128,967,121	13.3	T/C	0.83	0.45	0.22	0.64	0.74	0.2
135	rs567384	128,976,554	11.8	T/C	1	0.75	0.85	0.73	0.46	0.19
136	rs587779	128,978,605	14.2	T/C	1	0.85	0.91	0.86	0.49	0.71
137	rs541855	128,978,766	17.6	C/T	1	0.33	0.27	1	0.9	0.82
138	rs586919	128,978,778	17.5	G/C	1	0.33	0.27	0.92	0.9	0.71
139	rs9289321	129,011,459	15.7	G/A	0.73	0.66	0.82	0.62	0.81	0.35
140	rs1543272	129,070,913	10.7	G/A	0.95	0.43	0.2	0.66	0.92	0.28
141	rs1014470	129,113,596	6.9	G/C	0.41	1	0.087	0.55	0.7	0.11
142	rs2217628	129,134,319	14.4	C/T	0.53	0.73	0.42	0.8	1	0.64
143	rs1469656	129,161,494	45.8	A/G	1	0.9	0.83	0.67	0.73	0.73
144	rs4857926	129,192,982	34.4	C/T	0.91	0.89	0.95	0.26	0.28	0.47
145	rs1042907	129,271,862	26.6	C/G	0.69	0.58	0.22	0.66	0.46	0.22
146	rs7632756	129,300,198	26.3	T/C	0.19	0.62	0.11	0.26	0.78	0.084
147	rs2955114	129,360,446	16.6	A/C	0.29	0.87	0.084	0.16	0.85	0.032
148	rs6777054	129,428,039	25.8	A/G	0.17	0.47	0.12	0.3	0.73	0.17
149	rs2955103	129,497,926	37.3	T/C	0.13	0.9	0.018	0.81	0.46	0.25
150	rs1735538	129,574,792	23.7	A/G	0.21	0.57	0.15	0.41	0.78	0.23
151	rs2981015	129,672,076	22.3	A/G	0.92	0.28	0.16	0.68	0.83	0.39
152	rs3803	129,682,070	16.5	G/A	0.23	0.53	0.15	0.24	0.62	0.21
153	rs2713603	129,683,224	39.8	G/A	0.48	0.85	0.27	0.73	0.71	0.71
154	rs2860228	129,692,357	44.9	C/T	0.42	0.081	0.75	0.047	0.038	0.18
155	rs11712335	129,693,240	27.8	T/C	0.48	0.53	0.53	0.056	0.26	0.048
156	rs4857914	129,833,374	35.1	C/T	0.63	1	0.36	0.32	0.4	0.38
157	a488884	129,947,416	18.5	G/C	0.12	0.17	0.27	0.14	0.35	0.17
158	rs7642199	129,998,911	18.5	A/G	0.5	0.79	0.4	0.67	0.61	0.65

159	rs789230	130,085,248	23.5	G /A	0.18	0.081	0.68	0.5	0.096	0.7
160	rs883602	130,136,642	25.3	A/G	0.032	0.042	0.16	0.23	0.05	0.8
161	rs12695512	130,204,772	43.6	G/T	0.74	0.27	0.7	0.53	0.19	0.88
162	rs11709852	130,389,647	12.2	G/A	0.52	0.23	1	0.59	0.51	0.62
163	rs140695	130,634,834	28	C/T	0.82	1	0.69	0.73	0.8	0.73
164	a3140797	130,706,883	9.5	C/T	0.63	0.54	0.14	0.71	0.63	0.22
165	rs2625967	130,749,949	9.9	G/A	0.74	1	0.43	1	0.57	0.7
166	rs2255703	130,775,946	36.8	T/C	0.67	0.78	0.69	0.64	0.82	0.29
167	rs10804590	130,816,761	17	C/T	0.8	0.85	0.8	1	1	0.92
168	rs2717250	131,057,162	14.8	A/G	0.75	0.58	0.86	0.59	0.33	1
169	rs9289347	131,087,680	14.2	A/G	0.8	0.53	0.93	0.76	0.31	0.7
170	rs938194	131,156,190	22.3	C/G	0.59	0.73	0.62	0.64	0.62	0.22
171	rs7630899	131,167,973	42.5	C/T	0.76	0.94	0.52	0.38	0.87	0.11
172	rs6791923	131,189,082	39.8	A/G	0.63	0.69	0.65	0.64	0.51	0.83
173	rs9836814	131,417,587	28.8	A/G	0.15	0.57	0.1	0.091	0.13	0.22
174	rs6802606	131,422,477	13	G/A	0.69	0.84	0.65	0.48	1	0.31
175	rs2896565	131,432,956	7.8	G/A	0.36	1	0.15	0.57	0.28	0.13
176	rs9871734	131,442,289	21.1	A/G	0.51	0.62	0.48	0.46	0.097	0.76
177	a36458361	131,445,897	1.1	G/A	0.37	1	0.5	0.77	1	1
178	rs6777871	131,452,507	10.9	G/A	0.17	0.64	0.065	0.69	0.31	0.19
179	rs6804874	131,458,738	20.6	C/A	0.32	0.48	0.36	0.14	0.046	0.7
180	rs1453255	131,460,348	32	G/A	0.084	0.11	0.18	0.19	0.11	0.43
181	rs9832769	131,468,386	15.4	A/G	0.054	0.32	0.026	0.44	0.65	0.11
182	a25930206	131,468,667	4.6	G/C	0.46	0.27	0.67	0.57	0.42	0.62
183	a36487080	131,474,616	4.5	A/G	0.28	0.19	0.5	0.33	0.27	0.44
184	a36487277	131,474,813	9.1	C/T	0.12	0.42	0.11	0.1	0.12	0.55
185	rs7648240	131,479,302	10.9	G/T	1	0.89	0.92	0.93	0.89	0.89
186	rs16825656	131,487,758	27.3	A/G	0.16	0.27	0.2	0.13	0.082	0.39

187	rs6807290	131,499,485	26.7	G/A	0.16	0.28	0.25	0.11	0.081	0.35
188	rs1349747	131,543,410	18	A/G	0.081	0.65	0.053	0.27	0.9	0.12
189	rs5852593	131,580,335-131,580,336	24.3	del/T	0.0079	0.026	0.045	0.0082	0.018	0.095
190	rs1497309	131,616,310	28.7	C/T	0.016	0.14	0.013	0.0055	0.042	0.011
191	rs9883988	131,645,085	6.3	G/A	0.049	0.069	0.19	0.056	0.1	0.23
192	rs819091	131,678,260	23.7	G/A	0.05	0.3	0.025	0.065	0.31	0.06
193	rs1497319	131,708,045	29.4	T/A	0.065	0.3	0.037	0.07	0.35	0.045
194	rs4682614	131,725,939	20.7	G/A	0.91	0.7	0.87	0.94	0.93	0.93
195	rs9836961	131,769,576	16.4	C/T	0.19	0.15	0.5	0.28	0.29	0.42
196	rs1508520	131,780,349	18.8	T/C	0.95	0.84	0.69	0.87	1	0.65
197	rs6439249	131,783,340	11.8	C/T	0.67	0.34	0.78	0.88	0.48	0.71
198	rs1027858	131,785,876	11.8	A/C	0.78	0.34	0.63	0.94	0.4	0.57
199	rs9823913	131,829,269	20.6	A/G	0.52	0.88	0.23	0.65	0.85	0.42
200	rs11710068	131,877,537	22.8	C/T	0.6	0.75	0.54	0.72	0.85	0.39
201	rs12495186	131,881,158	22.6	T/C	0.46	0.79	0.32	0.8	0.63	0.32
202	rs4682627	131,893,100	22.9	T/C	0.41	0.49	0.38	0.54	1	0.3
203	rs900989	131,902,071	23.4	A/G	0.43	0.66	0.38	0.59	0.86	0.32
204	rs10934955	131,933,044	23.5	A/G	0.5	0.72	0.43	0.6	0.86	0.31
205	rs9824635	132,030,056	44.3	G/T	0.8	0.59	0.35	0.85	0.52	0.34
206	rs712984	132,131,488	40.2	G/A	0.84	0.4	0.64	0.46	0.29	0.8
207	rs218488	132,178,652	41.7	C/T	0.85	0.73	0.38	0.91	0.42	0.57
208	rs218503	132,220,322	32.1	G/A	0.48	0.14	0.93	0.73	0.19	0.48
209	rs17331416	132,279,027	20.4	G/T	1	0.27	0.24	0.95	0.13	0.16
210	rs1400013	132,356,413	43.7	G/T	0.45	0.71	0.29	0.7	0.94	0.51
211	rs6800866	132,424,091	21.5	T/C	0.43	0.87	0.23	0.61	0.85	0.3
212	rs6439279	132,491,823	22.2	C/T	0.74	0.57	0.27	0.64	0.79	0.23

Table S2. Association analysis of non-synonymous coding SNPs in *COL29A1* for atopic dermatitis in the discovery cohort.

							All				Paternal				Maternal			
SNP ID	Minor allele %	Alleles	Amino acid exchange	Amino acid position	*COL29A1* Location	Collagen XXIX Domain	T	NT	chi2	P value[a]	T	NT	chi2	P value	T	NT	chi2	P value
a36590333	1.9	**T**/C	Met/Thr	56	exon 3	VWFA1[d]	9	22	5.45	0.041	6	10	1	0.38	2	11	6.23	0.034
rs1453241	24.0	**G**/A	Glu/Lys	455	exon 5	VWFA3	166	124	6.08	0.023	68	50	2.75	0.11	79	55	4.3	0.038
a36602721	1.3	**T**/G	Val/Gly	669	exon 6	VWFA4	13	8	1.19	0.38	5	5	0	1	8	3	2.27	0.24
rs16827168	1.9	**A**/G	His/Arg	805	exon 6		16	14	0.13	0.84	7	6	0.08	1	9	8	0.06	1
rs11917356	16.6	**A**/G	Asp/Gly	982	exon 7	VWFA5	121	99	2.2	0.19	50	40	1.11	0.32	59	47	1.36	0.25
rs1353613	<1.0	**C**/G	Ile/Met	1114	exon 8	VWFA6	4	4	0	1								
a36611838	7.8	**G**/A	Val/Ile	1276	exon 9		57	54	0.08	0.87	31	23	1.19	0.36	23	28	0.49	0.53
rs12488457	29.9	**C**/A	Pro/Thr	1280	exon 9		201	155	5.94	0.025	82	62	2.78	0.11	85	59	4.69	0.042
rs1497312	23.7	**G**/C	Cys/Ser	1477	exon 17	Col [e]	162	137	2.09	0.2	56	52	0.15	0.7	74	53	3.47	0.077
a36620315	<1.0	**C**/G	Ser/Cys	1492	exon 17	Col	6	7	0.08	1								
a36624476	13.5	**C**/G	Thr/Ser	1539	exon 20	Col	109	83	3.52	0.086	50	44	0.38	0.59	50	30	5	0.027
rs16827497	23.1	**C**/T	Pro/Ser	1589	exon 23	Col	159	127	3.58	0.086	60	50	0.91	0.33	70	48	4.1	0.065
a36645464	5.0	**A**/C	Glu/Asp	1750	exon 33		55	38	3.11	0.078	30	20	2	0.16	20	13	1.48	0.29
a36654694	<1.0	**C**/T	Arg/Cys	2120	exon 35	VWFA 8	9	8	0.05	1								
rs9883988	6.3	**G**/A	Arg/Gln	2188	exon 36		56	35	4.85	0.05	31	18	3.45	0.068	19	11	2.13	0.19
rs819085	6.3	**A**/G	Asp/Gly	2205	exon 37		56	35	4.85	0.049	31	18	3.45	n.s.	19	11	2.13	n.s.
rs322117	18.2	**A**/G	Asn/Ser	2560	exon 40		116	99	1.34	0.3	48	45	0.1	0.73	52	38	2.18	0.17
a36685870	6.9	**G**/A	Ala/Thr	2589	exon 40		54	46	0.64	0.52	25	22	0.19	0.71	26	21	0.53	0.49
rs11355796	23.9	**_**/T	Stop/Leu	2591	exon 40		133	109	2.38	0.16	56	52	0.15	0.73	54	34	4.55	0.031

[a] *P* values of the ASPEX sib TDT for atopic dermatitis for all, paternal, and maternal transmissions.

[b] T number of transmissions

[c] NT number of non-transmissions

[d] VWA von Willebrand factor domain

[e] Col collagen domain

Supporting Table S3
Results of the ASPEX sib_TDT for SNPs in the region of *COL29A1* for allergic sensitization in the discovery cohort.

Locus	Position on chromosome 3	Location in COL29A1	MAF[a]	Major/ minor allele	All transmissions				Paternal				Maternal			
					Tb	NTc	$\chi 2$	P value	T	NT	$\chi 2$	P value	T	NT	$\chi 2$	P value
rs1349747	131.543.410	5’	18.0	**A**/G	91	75	1.54	0.27	34	35	0.01	0.90	52	35	3.32	0.12
rs1869319	131,547,433	exon 1	9.4	**C**/T	53	48	0.25	0.73	19	24	0.33	0.67	31	21	1.53	0.33
rs13095825	131,549,247	intron 1	36.6	**T**/C	164	122	6.17	0.02	58	42	2.56	0.081	78	52	5.2	0.036
rs1453243	131,575,875	intron 2	30.5	**G**/A	141	107	4.66	0.048	51	36	2.59	0.1	67	48	3.14	0.11
rs5852593	131,580,335	intron 3	24.3	**del**/T	117	78	7.8	0.0085	55	31	6.7	0.02	48	33	2.78	0.095
rs16845861	131,581,039	exon 4	24.4	**G**/A	123	86	6.55	0.017	52	33	4.25	0.052	57	39	3.38	0.081
rs1453242	131,586,229	intron 4	24.3	**C**/T	126	86	7.55	0.011	58	35	5.69	0.027	56	39	3.04	0.095
rs10212372	131,586,896	exon 5	24.4	**A**/G	122	86	6.23	0.02	52	34	3.77	0.065	56	38	3.45	0.078
A36603217	131,590,753	intron 6	22.6	**A**/G	120	87	5.26	0.031	56	44	1.44	0.27	53	32	5.19	0.026
rs10934938	131,596,325	intron 7	22.0	**T**/G	118	82	6.48	0.017	57	41	2.61	0.12	52	32	4.76	0.031
rs4688761	131,607,679	exon 16	28.6	**T**/C	146	91	12.76	0.0007	62	38	5.76	0.018	58	27	11.31	0.0013
A36624476	131,612,012	exon 20	13.5	**G**/C	79	65	1.36	0.29	36	33	0.13	0.65	35	24	2.05	0.16
rs1497309	131,616,310	intron 22	28.7	**C**/T	144	98	8.74	0.0052	60	40	4.0	0.045	58	32	7.51	0.011
rs7628563	131,617,253	intron 23	15.1	**A**/G	96	63	6.85	0.017	46	25	6.21	0.01	37	25	2.32	0.16
A36637742	131.625.270	intron 28	27,8	**del**/T	141	104	5.59	0.028	61	43	3.12	0.09	55	36	3.97	0.063
rs1566136	131,634,657	intron 33	28.4	**A**/G	140	112	3.11	0.096	66	49	2.51	0.11	53	42	1.27	0.24
rs9883988	131,645,085	exon 36	6.3	**G**/A	44	26	4.63	0.057	26	15	2.95	0.10	15	8	2.13	0.23
rs322113	131,670,507	exon 38	7.8	**C**/T	29	26	0.16	0.8	16	13	0.31	0.77	13	13	0	1.0
rs322115	131,672,716	intron 39	18.2	**A**/T	80	64	1.78	0.26	37	30	0.73	0.43	36	27	1.29	0.36
rs819091	131,678,260	intron 40	23.7	**G**/A	112	84	4.0	0.062	52	41	1.3	0.31	47	30	3.75	0.061
rs7624712	131,686,432	3’	11.6	**A**/G	48	47	0.01	1	24	22	0.09	0.66	22	23	0.02	0.80

[a] MAF minor allele frequency

[b] T number of transmissions

[c] NT number of non-transmissions

Table S4. Association analysis of single markers for allergic sensitization in the replication data set.

Locus	Major/ minor allele	All transmissions				Paternal				Maternal			
		T[b]	NT[c]	$\chi 2$	P value[a]	T	NT	$\chi 2$	P value	T	NT	$\chi 2$	P value
rs13095825	T/C	159	122	4.87	0.03	57	46	1.17	0.28	67	41	6.26	0.0073
rs16845861	G/A	128	89	7.01	0,0072*	54	36	3.6	0.05	55	34	4.96	0.02
rs10212372	A/G	127	89	6.69	0.0097*	54	36	3.6	0.05	54	34	4.55	0.024
A36603217	A/G	122	81	8.28	0,0044*	62	39	5.24	0.024	45	27	4.5	0.02
rs10934938	T/G	123	83	7.77	0,0056*	63	40	5.14	0.023	45	28	3.96	0.026
rs4688761	T/C	138	96	7.54	0,0065*	73	48	5.17	0.027	49	32	3.57	0.055
A36637742	del/T	170	115	10.61	0,0010*	74	53	3.47	0.07	71	37	10.7	0.00072*
rs9883988	G/A	56	31	7.18	0,0073*	32	17	4.59	0.035	23	13	2.78	0.072

[a] *P* values of the ASPEX sib TDT for allergic sensitization for all, paternal, and maternal transmissions. None of the paternal transmissions were significant after simulation. * *P* values significant after correction for multiple testing.

Table S5. Haplotype analysis for allergic sensitization in the replication data set.

										Allergic sensitization				
rs13095825	rs16845861	*rs*10212372	A36603217	*rs*10934938	*rs*4688761	A3667742	rs9883988	freq.	fam.	S	E(S)	Var(S)	Z	*P* value
2	1	1	1	2	.		.	0.62	112	360.0	332.7	76.8	3.12	0.0018*
.	1	1	1	2	1		.	0.68	103	352.7	324.3	69.2	3.42	0.00062*
		1	1	2	1	1		0.66	100	350.7	320.2	67.7	3.70	0.00021*
.	.		1	2	1	1	2	0.70	96	313.9	313.9	70.5	3.9	0.000085*

fbat statistic (Z) for 5-marker haplotype transmissions in the replication set adjusted for the correlation among siblings, reporting haplotype frequency (freq.) the number of informative families (fam.), the distribution (S), the expected distribution (E(S)), and the variance (Var(S)). * *P* values significant after correction for multiple testing.