

A common variant on chromosome 11q13 is associated with atopic dermatitis

Jorge Esparza-Gordillo^{1,2,16}, Stephan Weidinger^{3,4,16}, Regina Fölster-Holst⁵, Anja Bauerfeind², Franz Ruschendorf², Giannino Patone², Klaus Rohde², Ingo Marenholz^{1,2}, Florian Schulz^{1,2}, Tamara Kerscher^{1,2}, Norbert Hubner², Ulrich Wahn¹, Stefan Schreiber^{6,7}, Andre Franke⁶, Rainer Vogler⁷, Simon Heath⁸, Hansjörg Baurecht^{4,9}, Natalija Novak¹⁰, Elke Rodriguez^{3,4}, Thomas Illig¹¹, Min-Ae Lee-Kirsch¹², Andrzej Ciechanowicz¹³, Michael Kurek¹⁴, Tereza Piskackova¹⁵, Milan Macek¹⁵, Young-Ae Lee^{1,2} & Andreas Ruether⁶

We conducted a genome-wide association study in 939 individuals with atopic dermatitis and 975 controls as well as 270 complete nuclear families with two affected siblings. SNPs consistently associated with atopic dermatitis in both discovery sets were then investigated in two additional independent replication sets totalling 2,637 cases and 3,957 controls. Highly significant association was found with allele A of rs7927894 on chromosome 11q13.5, located 38 kb downstream of *C11orf30* ($P_{\text{combined}} = 7.6 \times 10^{-10}$). Approximately 13% of individuals of European origin are homozygous for rs7927894[A], and their risk of developing atopic dermatitis is 1.47 times that of noncarriers.

Atopic dermatitis (MIM603165), or eczema, is a chronic inflammatory skin disorder and a major manifestation of allergic disease. In industrialized countries, the prevalence of atopic dermatitis is approximately 15%, and there has been a steady increase over the past decades¹. Genetic and environmental factors interact to determine disease susceptibility², and family and twin studies indicate that the genetic contribution is substantial³. The molecular mechanisms underlying eczema are not fully understood, although skin barrier defects as well as systemic and cutaneous immune dysfunction in response to allergens or bacterial products are thought to have an important role⁴.

To identify genetic variants contributing to atopic dermatitis, we conducted a genome-wide association (GWA) study in two German

study populations. We genotyped 939 atopic dermatitis cases and 975 controls (set 1) as well as an independent set of 270 complete nuclear families comprising 1,097 individuals (set 2) on Affymetrix Human mapping 500K and 5.0 arrays (Fig. 1 and Supplementary Methods online). A total of 342,303 markers fulfilled quality control criteria and were analyzed for association with atopic dermatitis. Individuals showing evidence of non-European admixture were excluded, and no additional evidence of genetic stratification was found by principal component analysis with Eigenstrat⁵. Comparison of the observed and expected distribution showed a moderate inflation of the test statistic in set 1 (Cochran-Armitage trend test; inflation factor $\lambda = 1.14$; Supplementary Fig. 1a online), which was corrected for by the genomic control method.

No single SNP achieved genome-wide significance after correction for the number of test performed ($P < 1.46 \times 10^{-7}$) in either set 1 or 2 (Supplementary Fig. 1b). We then focused on those genomic regions consistently associated with atopic dermatitis in both discovery sets. We required the primary associations to be significant at $P < 0.005$ in either set 1 or 2, and replicated with the same risk allele at $P < 0.05$ in the alternative discovery set (Fig. 1). When multiple markers in the same region fulfilled these criteria, we selected the SNP yielding the strongest association. We selected and genotyped 54 SNPs in an independent sample of 1,363 German atopic dermatitis cases and 2,739 German controls (set 3; Supplementary Table 1 online). We expected to observe less than three association results at a nominal P value of <0.05 under the null hypothesis of no association

¹Pediatric Pneumology and Immunology, Charité Universitätsmedizin Berlin, Berlin, Germany. ²Max-Delbrück-Centrum for Molecular Medicine (MDC), Berlin-Buch, Germany. ³Department of Dermatology and Allergy, Technische Universität München, Munich, Germany. ⁴Division of Environmental Dermatology and Allergy, Helmholtz Zentrum Munich and ZAUM-Center for Allergy and Environment, Technische Universität München, Munich, Germany. ⁵Clinic for Dermatology, Venerology and Allergology, University Hospital Schleswig-Holstein, Kiel, Germany. ⁶Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. ⁷POPGEN Biobank Project, Christian-Albrechts-University, Kiel, Germany. ⁸Centre National de Génotypage, Evry, France. ⁹Institute for Medical Statistics and Epidemiology IMSE, Technische Universität München, Munich, Germany. ¹⁰Department of Dermatology and Allergy, University of Bonn, Bonn, Germany. ¹¹Department of Epidemiology, Helmholtz Zentrum Munich-German Research Center for Environmental Health, Neuherberg, Germany. ¹²Klinik für Kinder- und Jugendmedizin, Technical University Dresden, Dresden, Germany. ¹³Department of Laboratory Diagnostics and Molecular Medicine, Pomeranian Medical University, Szczecin, Poland. ¹⁴Department of Clinical Allergology, Pomeranian, Pomeranian Medical University, Szczecin, Poland. ¹⁵Department of Biology and Medical Genetics, Charles University Prague, 2nd Medical School and Faculty, Hospital Motol, Prague, Czech Republic. ¹⁶These authors contributed equally to this work. Correspondence should be addressed to Y.A.L. (yolee@mdc-berlin.de).

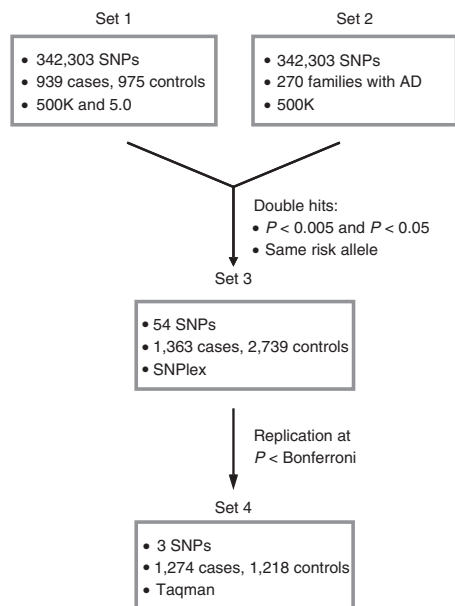


Figure 1 Study design. AD, atopic dermatitis.

($54 \times 0.05 = 2.7$); instead, we found nine SNPs associated with atopic dermatitis at this threshold, which is more than three times the number of positive results expected to occur by chance. Moreover, three SNPs located on chromosomes 1q21, 9p21 and 11q13 were significantly associated with atopic dermatitis in set 3 after Bonferroni correction for the 54 tests performed (corrected threshold = $0.05/54 = 0.00093$).

Results on 1q21 will be shown below. Markers on 9p21 and 11q13 were tested for association in an additional replication set of 1,274 cases and 1,218 controls of European ancestry (set 4). The association finding on chromosome 9p21 was not replicated in set 4 (rs7024096; **Supplementary Table 2** online). However, a strong association with atopic dermatitis, reaching genome-wide significance in the overall analysis of the discovery and follow-up groups, was observed for the A allele of rs7927894 on chromosome 11q13.5 (combined OR = 1.22, 95% CI = 1.15–1.30, $P = 7.6 \times 10^{-10}$; **Table 1**). We computed the genotype-specific OR for rs7927894 in order to investigate the mode of inheritance more carefully. Results from all groups combined demonstrated that association of rs7927894 with atopic dermatitis did not deviate from the additive model ($P = 0.37$). Compared to noncarriers, heterozygous and homozygous carriers of the risk allele had ORs of 1.16 (95% CI = 1.02–1.32) and 1.47 (95% CI = 1.29–1.68), respectively. Considering the average frequency of the risk allele in all control populations (0.363; **Table 1**), individuals homozygous for rs7927894 (AA) represent approximately 13% of the population.

Notably, the same atopic dermatitis risk allele reported here (rs7927894 allele A) has recently been identified as a susceptibility factor for Crohn's disease⁶. Crohn's disease is a complex chronic inflammatory bowel

disorder sharing many pathophysiological characteristics with atopic dermatitis, such as recurrent inflammation of the epithelial barrier, defective cutaneous or mucosal barrier function, and deficient innate immune responses against bacterial infections⁷. Present data suggest that rs7927894[A] confers susceptibility to atopic dermatitis and Crohn's disease jointly, which may contribute to the higher incidence of atopic dermatitis observed among individuals with Crohn's disease^{8,9}.

rs7927894 is located in an intergenic region 38 kb downstream of *C11orf30* (chromosome 11 open reading frame 30) and 68 kb upstream of *LRRC32* (leucine rich repeat containing 32). To identify a potential link to atopic dermatitis, we performed RT-PCR analysis, which showed that both *C11orf30* and *LRRC32* are ubiquitously expressed, including in tissues relevant to atopic dermatitis such as skin and peripheral blood mononuclear cells (**Supplementary Fig. 2** online). Additionally, we assessed a possible regulatory role of rs7927894 on *C11orf30* or *LRRC32* gene expression in the publicly available dataset generated by Dixon *et al.*¹⁰, who conducted a genome-wide association study for global mRNA expression in lymphoblastoid cell lines from asthmatic children. There was no evidence for a *cis*-regulatory effect of rs7927894. However, the possibility of a regulatory effect in another tissue or physiological state cannot be ruled out.

Linkage disequilibrium (LD) analysis in HapMap showed that rs7927894 is located in a 200-kb LD block containing *C11orf30* (**Fig. 2**), suggesting that the functional variant responsible for the observed association may be located within the *C11orf30* region. However, the possibility remains that rs7927894 may be in LD with a functional variant not genotyped in the HapMap project and located outside this LD block. *C11orf30* encodes the nuclear protein EMSY, which binds and inactivates the cancer susceptibility gene *BRCA2*. EMSY associates with chromatin regulators HP1 β and BS69, and localizes to sites of repair following DNA damage. It has been implicated in chromatin modification, DNA repair and transcriptional regulation. An increase in *C11orf30* copy number has been reported in epithelium-derived cancer of the breast and ovary¹¹. The potential involvement of *C11orf30* in multiple inflammatory and

Table 1 Association of rs7927894[A] on 11q13.5 with atopic dermatitis

Study population	Number		RAF or T:U		OR (95% CI)	P value
	Cases	Controls	Cases	Controls		
Discovery groups (GWA)						
Set 1 ^a	939	975	0.409	0.372	1.18 (1.03–1.33)	0.027
Set 2	529 (270 families)		294:226		1.30 (1.09–1.55)	0.001
Follow-up groups						
Set 3 ^a	1,363	2,739	0.407	0.363	1.21 (1.10–1.33)	2.46 × 10 ^{−4}
Set 4	1,274	1,218	0.415	0.357	1.29 (1.15–1.45)	1.94 × 10 ^{−5}
GWA ^b						3.83 × 10 ^{−4}
Follow-up groups ^b	2,637	3,957	0.411	0.361	1.24 (1.15–1.33)	9.62 × 10 ^{−8}
All combined ^{bc}	4,105	4,932	0.411	0.363	1.22 (1.15–1.30)	7.64 × 10 ^{−10}

All *P* values are two-sided. Shown are the number of cases and controls in each data set with the corresponding risk allele frequency (RAF). The number of individuals in set 2 refers to the number of atopic dermatitis cases (529) and the number of families (270). Odds ratios (OR) were calculated by logistic regression analysis and *P* values with the Cochran-Armitage trend test. For set 2, the number of transmissions versus nontransmissions are shown (T:U). OR in set 2 represents the ratio of transmitted/untransmitted alleles.

^aResults for set 1 and set 3 were adjusted by the genomic control method ($\lambda = 1.14$, **Supplementary Fig. 1**). ^bFor the combined analysis, the OR and *P* values were calculated using the inverse variance²⁸ and Fisher's method²⁷, respectively. ^cResults from the family-based set 2 were not included in the calculation of allele frequencies and OR for the combined population.

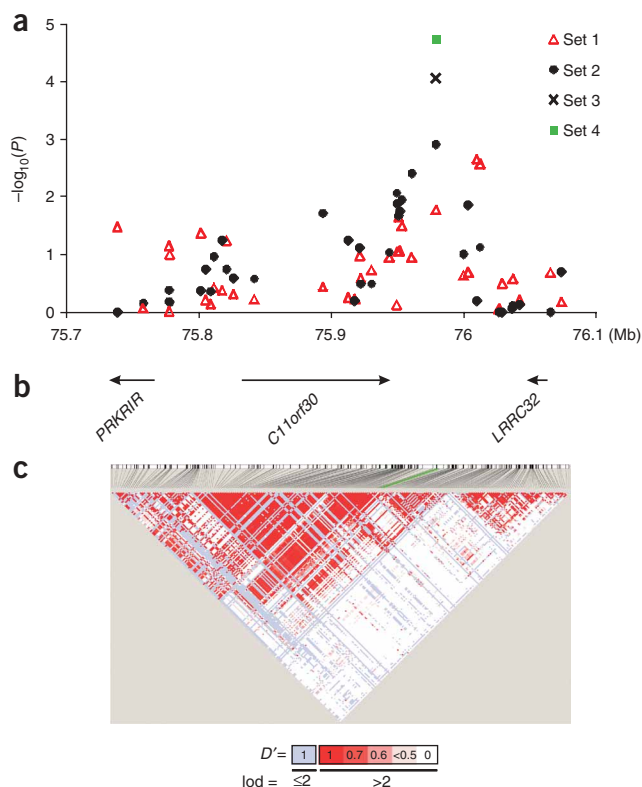


Figure 2 Association results and LD structure in the atopic dermatitis-associated region on 11q13.5. (a) Association results for all GWA markers tested in the region. Physical positions are based on NCBI build 36. (b) Genes in the region from the UniGene database. (c) LD in the CEU HapMap population. Disequilibrium coefficient values for Hapmap Phase II data (v. 22) were generated with Haploview. The lead SNP is indicated by a green line.

sets (Fig. 3 and Table 2). Marker rs877776 yielded the highest association ($P_{\text{set1}} = 0.0003$, $P_{\text{set2}} = 0.005$; Table 2) and was then genotyped in set 3 and set 4. Again, risk allele C of rs877776 showed an increased frequency in cases compared to controls after exclusion of *FLG* mutation carriers, although the trend was not significant ($P_{\text{replication sets 3, 4}} = 0.104$). Altogether, these data are consistent with that of a previous report¹⁶ and suggest that additional atopic dermatitis risk factors exist in the EDC apart from prevalent *FLG* mutations (rs877776, $P_{\text{combined}} = 3.5 \times 10^{-5}$). rs877776 is in LD with a number of candidates involved in skin barrier function (Fig. 3), including the genes encoding hornerin, flaggrin, flaggrin2, cornulin and late cornified envelope protein 5A, suggesting that the remaining atopic dermatitis risk factor may be located within or between any of these genes. Finally, no evidence for genetic interaction was observed after testing all possible combinations between the *FLG* mutations, rs877776, and the lead SNP on chromosome 11 (rs7927894).

Six additional loci were nominally replicated at $P < 0.05$ in set 3 but were not significant after multiple-testing correction and thus

malignant epithelial diseases (atopic dermatitis, Crohn's disease and adenocarcinoma) strongly suggests a role for *C11orf30* in epithelial immunity, growth and/or differentiation.

An additional association signal was observed in the epidermal differentiation complex (EDC) on chromosome 1q21, which contains a number of genes that are required for epidermal structure and function¹². Genotyping of rs6661961 in set 3 confirmed the association reaching genome-wide significance ($P_{\text{combined set 1,2,3}} = 1.2 \times 10^{-9}$; Table 2). Data from both discovery sets revealed the presence of multiple disease-associated markers spanning 600 kb within the EDC (Fig. 3). Because of the proximity of the well-established atopic dermatitis susceptibility gene *FLG* (filaggrin)¹³, we evaluated whether the observed association was due to LD with *FLG* mutations or to additional genetic factors. We genotyped all four *FLG* mutations previously reported in the German population (R501X, 2282del4, R2447X and S3247X)¹⁴ in all study samples. We observed markedly high LD between the *FLG* mutations and rs6661961, which is located 156 kb away (Supplementary Fig. 3 online). Exclusion of *FLG* mutation carriers abolished the observed association ($P > 0.1$; Table 2). This result indicates that the association signals in the EDC were due to *FLG* mutations, reinforcing their role in atopic dermatitis and supporting the robustness of the study. It has been shown that the spectrum of *FLG* mutations varies among human populations, with the more prevalent European mutations (R501X and 2282del4) absent in African and Asian populations^{13,15}. This suggests that the mutational events occurred after the separation of the respective human populations. The high LD observed between *FLG* mutations and SNPs located as far as 500 kb away (Supplementary Fig. 3) supports this hypothesis and allowed us to detect *FLG* in our scan.

Notably, a set of markers surrounding the *HRNR* (hornerin) gene, located 78 kb away from *FLG*, showed a strong association with atopic dermatitis after exclusion of *FLG* mutation carriers in both discovery

Table 2 Association results for the lead SNPs in the EDC before and after stratification for *FLG* mutations

		Full sample		Non- <i>FLG</i> mutation carriers	
Marker		rs6661961	rs877776	rs6661961	rs877776
Minor allele		A	C	A	C
GWA set 1	Cases	0.498	0.169	0.408	0.191
	Controls	0.430	0.137	0.409	0.139
	OR	1.31	1.28	0.99	1.46
	<i>P</i>	1.2×10^{-4}	0.01	0.939	2.8×10^{-4}
GWA set 2	T:U	281:217	164:122	150:134	120:75
	OR	1.30	1.34	1.12	1.60
	<i>P</i>	0.003	0.02	0.86	0.005
Follow-up set 3	Cases	0.484	0.156	0.406	0.175
	Controls	0.432	0.158	0.414	0.163
	OR	1.23	0.98	0.97	1.10
	<i>P</i>	2.9×10^{-5}	0.82	0.565	0.19
Follow-up set 4	Cases		0.163		0.181
	Controls		0.159		0.164
	OR		1.02		1.13
	<i>P</i>		0.77		0.13
All combined ^a	OR		1.06		1.20
	<i>P</i>		0.027		3.52×10^{-5}

Shown are the corresponding allelic frequencies of the minor allele and the number of transmitted versus untransmitted alleles for set 2. Non-*FLG* mutation carriers refers to the individuals remaining after exclusion of all individuals or families carrying any of the four genotyped mutations (665 cases, 873 controls in set 1; 161 families in set 2; 984 cases, 2,436 controls in set 3; 980 cases, 1,102 controls in set 4). Note that individuals or families with missing *FLG* mutation genotypes were also excluded.

^aResults from the family-based set 2 were not included in the calculation of allele frequencies and OR for the combined population.

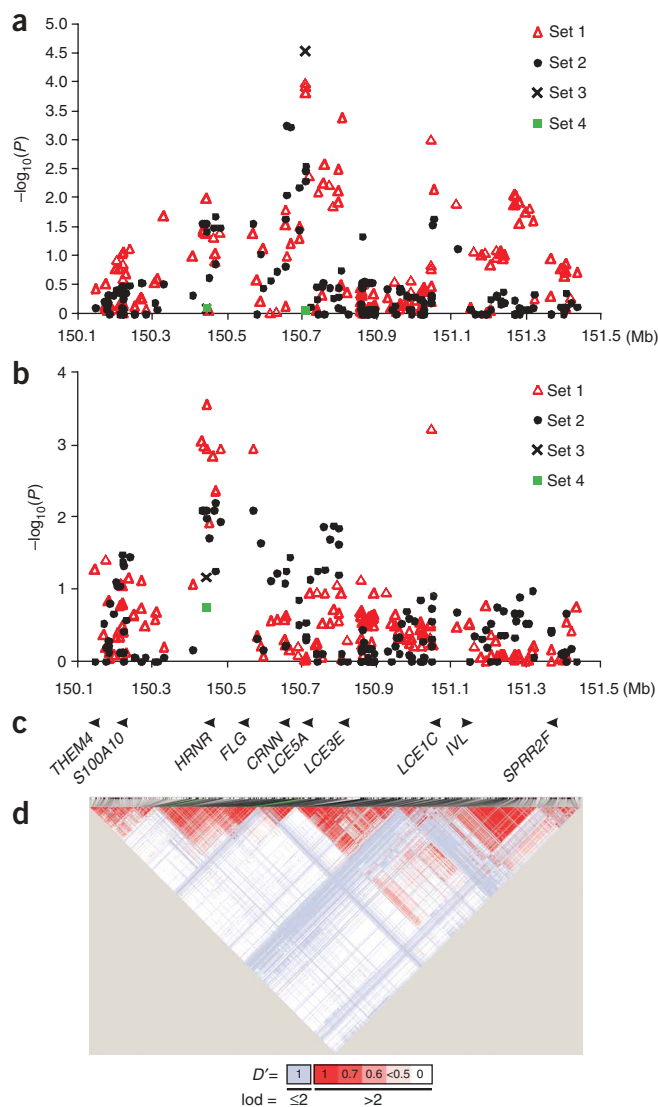


Figure 3 Association results and LD structure in the epidermal differentiation complex (EDC) on 1q21. **(a)** Association results for all GWA markers tested in the region. **(b)** Association results after exclusion of all individuals or families carrying any of the four tested *FLG* mutations. **(c)** Genes in the region from the UniGene database. Owing to the size of the genomic region shown (1.4 Mb) and the high number of genes contained, only selected genes are shown as a reference. **(d)** LD in the CEU HapMap population. Disequilibrium coefficient values for HapMap Phase II data (v. 22) were generated with Haploview. Markers rs877776 and rs6661961 are indicated by a green line.

METHODS

Study participants. We evaluated four independent study groups (**Supplementary Methods**). All cases and controls were of European origin. The institutional ethics review boards of the participating centers approved the study protocol, and informed consent was obtained from all probands or their legal guardians. In all four study groups, the physician's diagnosis of eczema was made according to standard criteria in the presence of a chronic or chronically relapsing pruritic dermatitis with the typical morphology and distribution^{20,21}.

All samples in sets 1, 2 and 3 were German. Set 1 included 939 atopic dermatitis cases and 975 controls; set 3 consisted of 1,363 atopic dermatitis cases and 2,739 controls. All atopic dermatitis cases from sets 1 and 3 were obtained from German university hospitals (Charité Universitätsmedizin Berlin, University of Kiel, Technical University Munich, and University of Bonn), and were of self-reported German ancestry. All German controls in sets 1 and 3 were derived from the previously described population-based Popgen Biobank²².

Samples in set 2 were recruited at Charité Universitätsmedizin Berlin and consisted of 270 complete nuclear families comprising 1,097 individuals and including 529 children with atopic dermatitis. Families were recruited through two affected siblings with an age of onset below two years of age and moderate-to-severe disease expression.

Set 4 consisted of 1,274 atopic dermatitis cases and 1,218 geographically matched controls from Germany, the Czech Republic and Poland. German atopic dermatitis cases ($n = 456$) were recruited at Charité Universitätsmedizin Berlin. All 360 Czech and 458 Polish atopic dermatitis cases had a positive family history of atopy and were recruited for the EPAAC (Early Prevention of Asthma in Atopic Children, $n = 795$) and ETAC (Early Treatment of the Atopic Child, $n = 23$) trials, two similar randomized, double-blind, placebo-controlled studies on the efficacy of levocetirizine or cetirizine, respectively, in the prevention of asthma^{23,24}. The Czech, Polish and German controls were obtained from blood donation programs in Prague, Szczecin and Dresden, respectively. A summary of the study groups is included in **Supplementary Methods**.

Genotyping and quality control of the data. Genotyping for the GWA study was done with Affymetrix 500K and 5.0 Arrays. According to the manufacturer's recommendations, we discarded samples with call rates $<93\%$ or 86% after preliminary quality control calling with the Dynamic Model algorithm for 500K and 5.0 arrays, respectively. Samples showing evidence of cryptic relatedness to other study participants or non-European ancestry were excluded. A summary of excluded samples is presented in **Supplementary Methods**. Principal components analysis with Eigenstrat⁵ revealed no evidence of population stratification.

We excluded SNPs that had low genotyping rate ($<95\%$), deviated from Hardy-Weinberg equilibrium ($P_{HWE} < 0.0005$ in controls), showed excess heterozygosity (>0.55) or were rare (minor allele frequency (MAF) $<1\%$). The genotype concordance rate between the 500K and 5.0 arrays was evaluated by genotyping 82 samples with both platforms. A group of 8,753 markers with more than two discordant genotypes in this set of 82 genotype pairs were excluded from further analysis, yielding a final genotype concordance rate of 99.9%. Additional quality control criteria are described in **Supplementary Methods**.

All replication genotyping in sets 3 and 4 was done with ligation-based SNPlex or Taqman following the assay manual (see **Supplementary Methods** for additional details).

were not genotyped in set 4 (**Supplementary Table 3** online). Some of these loci affect immune processes related to atopic dermatitis pathogenesis, such as the differentiation of type 2 T helper cells, mast cells, eosinophil and dendritic cell chemotaxis, and pruritus (*HRH4*)¹⁷, Toll-like receptor-mediated responses to bacteria (*LY86*, also called *MD-1*)¹⁸, and CD8⁺ T cell function (*EOMES*)¹⁹. Power calculations indicate that our discovery sets had $>84\%$ power to detect variants with a relative risk of 1.5 and a minor allele frequency of 20% at $P < 0.005$. The power drops to 12% for a relative risk of 1.2 and the same allele frequency (**Supplementary Methods**). These data suggest that additional low-risk factors remain to be discovered for atopic dermatitis, and bigger samples sizes will be required to identify and replicate them.

Here we present the first genome-wide association study for atopic dermatitis. We identified a common sequence variant on chromosome 11q13.5 associated with atopic dermatitis in all four study groups including more than 4,100 atopic dermatitis cases. Moreover, we provide a list of additional candidate genes, including a new susceptibility locus within the EDC. Further replication in additional cohorts, fine mapping and functional studies will be required to gain a better understanding of the physiological mechanisms underlying this common allergic disorder.

Statistical analysis. SNP association with atopic dermatitis in sets 1, 3 and 4 was analyzed using the Cochran-Armitage test (ATT) with additive trend as implemented in PLINK²⁵. In discovery set 1, we observed a moderate variance inflation ($\lambda = 1.14$) in the distribution of the test statistic (median χ^2_{1df} of the Cochran-Armitage trend test) when compared to the median of the expected distribution under the null hypothesis of no association (Supplementary Fig. 1a). After genomic control correction, no evidence of inflation remained in set 1, suggesting that no systematic bias exists in the dataset. To address potential genetic stratification in set 3, we adjusted the *P* values in set 3 with the variance inflation ($\lambda = 1.14$) observed in set 1 because both populations originated from the same study centers in Germany. Additionally, we used available genome-wide Illumina HapMap300 SNP genotypes²⁶ from control individuals originating from Germany, the Czech Republic and Poland to estimate the expected variance inflation (λ) in set 4. No evidence of variance inflation was observed in this analysis, and thus, no lambda correction was applied to the set 4 results (see Supplementary Methods for detailed description). Family-based association analysis in set 2 was done using the transmission disequilibrium test (TDT) implemented in PLINK²⁵. An empirical *P* value obtained with a permutation procedure was used to account for the non-independence of the two affected sibs included for each family. Odds ratios (OR) for sets 1, 3 and 4 are based on logistic regression without adjustment for covariates and represent the risk for atopic dermatitis associated with each copy of the risk allele. ORs for set 2 were calculated with PLINK and indicate the ratio of transmissions versus nontransmissions for the risk allele. Combined *P* values were obtained using Fisher's method²⁷. ORs were combined by the weighted average of study-specific estimates of the ORs using inverse variance weights²⁸. After combining the results from sets 1, 3 and 4, the multiplicative model provided an adequate fit for rs7927894 on chromosome 11. The risk predicted by the genotype-specific model was not significantly different from that predicted by the log-additive (or multiplicative) genetic model ($P = 0.37$), obtained with a likelihood-ratio test. We conducted logistic regression analyses to examine multiplicative gene–gene interaction between the *FLG* mutations and the markers rs7927894 (chromosome 11) and rs877776 (chromosome 1). Regarding the *FLG* genotypes, individuals were coded as noncarriers of any mutations (code 0), as heterozygote carriers (code 1) or as homozygote/composite heterozygote mutation carriers (code 2). For rs877776 and rs7927894, individuals were coded according to the number of minor alleles carried (0, 1 or 2).

Power calculations in case-control analysis were performed with the Genetic Power Calculation software²⁹, and a previously described method³⁰ was used to compute power for family-based set 2 (see Supplementary Methods).

Tissue expression pattern. We examined the tissue-specific expression of *C11orf30*, *LRRC32* and *GAPDH* 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). In addition, human skin RNA from a healthy donor was extracted using Trizol (Invitrogen), treated with DNase and transcribed into cDNA using the Transcriptor First Strand cDNA synthesis kit (Roche Diagnostics). The *GAPDH* PCR was carried out using the primers provided with the MTC panels, and the *C11orf30* and *LRRC32* PCRs were done using specific primers listed in Supplementary Table 4 online.

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

ACKNOWLEDGMENTS

We thank all subjects with eczema, families and physicians for their participation. We thank I.v. Gwinner-Bühring, F. Kussebi, C. Ockeloen, T. Schilling and I. Schulz for subject evaluation. We are grateful to C. Flachmeier, S. Kolberg, M. Schwarz, I. Szangolies, S. Blachut, G. Born, S. Schmidt, C. v.d. Lancken, T. Henke, B. Fedders and T. Wesse for technical assistance, and to M. Wittig for database management. We thank L. Ghys and M. De Longueville (UCB Pharma) and the following EPAAC and ETAC investigators for providing case samples: O. Rybníček (Brno, Czech Republic), J. Chládková, T. Chyba (Hradec Kralove, Czech Republic), F. Kopriva (Olomouc, Czech Republic), O. Škopková, V. Balčárek (Ostrava, Czech Republic), H. Honomichlova, P. Honomichlová-Houdkova, P. Honomichl (Plzeň Lohtin, Czech Republic), M. Špicáková, S. Kynclová, V. Petru, A. Carbolova, K. Kopecká, P. Pohunek, T. Svobodova, V. Špicák, K. Kopecka, M. Maltulka (Praha, Czech Republic), M. Kaczmarzki,

B. Cudowska, J. Wasilewska, E. Matuszewska (Białystok, Poland), T. Malaczynska, B. Klajna-Kraluk (Gdansk, Poland), J. Sokalska, E. Masnica-Wasylikowska, I. Biegun-Awramineko, M. Chlon (Gliwice Poland), T. Latos, E. Pomaranska, G. Gaszczyk, B. Makuch, J. Stanis, J. Bokiej (Karpacz, Poland), G. Lis, E. Cichocka-Jaros, I. Glodzik, T. Szczerbinski (Kraków, Poland), D. Chlebna-Sokol, A. Stanczyk, J. Wlazlowski, I. Ligenza, B. Kamer, K. Pyziak, R. Pasowska, J. Zwaigzne-Raczynska (Lodz, Poland), A. Emeryk, H. Milanowska, G. Zywicka, M. Bartkowiak-Emeryk, E. Chojna (Lublin, Poland), A. Swiatly, A. Szczawinska-Poplonyk, M. Kulesza-Kazecka, J. Alkiewicz, A. Breborowicz (Poznan, Poland), R. Kurzawa, A. Wojcik, U. Jedrys-Klucjasz, E. Urbanek-Jozwik (Rabka Zdrój, Poland), L. Dymek, A. Dymek, A. Bozek (Strzelce opolskie, Poland), D. Chmielewska-Szewczyk, J. Lange, J. Peradzynska, M. Kulus, E. Najberg, E. Nowicka, A. Chrupek (Warszawa, Poland), A. Boznanski, E. Willak-Janc, A. Latkowska, E. Sikorska, M.T. Jarlinska (Wroclaw, Poland). The study was funded by the German Ministry of Education and Research (BMBF) through the Clinical Research Group for Allergy at Charité Berlin, the National Genome Research Network (NGFN), by the PopGen biobank and infrastructure support through the DFG cluster of excellence "Inflammation at Interfaces". This work was supported by an EMBO long-term fellowship (522-2006) and by a Marie Curie EIF fellowship (039868) to J.E.G. N.N. was supported by grants from the German Research Council (DFG NO454/1-4, NO454/2-4, NO454/5-2, SFB704 TPA4) and BONFOR grants of the University of Bonn. S.W. was supported by research grants KKF-07/04 and KKF-27/05 of the University Hospital Rechts der Isar, Technische Universität München, a grant from the Wilhelm-Vaillant-Stiftung and a Heisenberg fellowship (DFG WE 2678/4-1) from the German Research Council. M.M. was supported by research grant VZFM00064203.

AUTHOR CONTRIBUTIONS

J.E.G., Y.A.L., S.W., A.R. and N.H. designed the study. Y.A.L. directed the study. Y.A.L., J.E.G., S.W., R.F.-H., U.W., S.S., N.N., M.L.-K., A.C., M.K., T.P. and M.M. provided the case and control samples. J.E.G., S.W., A.R., G.P., A.F., R.V., H.B., T.I. and Y.A.L. contributed to biobanking and management of clinical data. J.E.G., G.P., I.M., F.S. and T.K. performed 500K array and Taqman genotyping. A.R. did the SNPlex genotyping; E.R. contributed to DNA extraction, biobanking and *FLG* genotyping of Munich samples. A.B., F.R., K.R., J.E.G., N.H. and Y.A.L. conducted all statistical analyses and interpreted the data. S.H. estimated population stratification in set 4. J.E.G. and Y.A.L. wrote the manuscript.

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

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

- Taylor, B., Wadsworth, J., Wadsworth, M. & Peckham, C. Changes in the reported prevalence of childhood eczema since the 1939–45 war. *Lancet* **2**, 1255–1257 (1984).
- Cookson, W. The alliance of genes and environment in asthma and allergy. *Nature* **402**, B5–B11 (1999).
- Schultz Larsen, F. Atopic dermatitis: a genetic-epidemiologic study in a population-based twin sample. *J. Am. Acad. Dermatol.* **28**, 719–723 (1993).
- Bieber, T. Atopic dermatitis. *N. Engl. J. Med.* **358**, 1483–1494 (2008).
- Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
- Schreiber, S., Rosenstiel, P., Albrecht, M., Hampe, J. & Krawczak, M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat. Rev. Genet.* **6**, 376–388 (2005).
- Pugh, S.M. *et al.* Atopic disease in ulcerative colitis and Crohn's disease. *Clin. Allergy* **9**, 221–223 (1979).
- Hammer, B., Ashurst, P. & Naish, J. Diseases associated with ulcerative colitis and Crohn's disease. *Gut* **9**, 17–21 (1968).
- Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
- Hughes-Davies, L. *et al.* EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* **115**, 523–535 (2003).
- Mischke, D., Korge, B.P., Marenholz, I., Volz, A. & Ziegler, A. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J. Invest. Dermatol.* **106**, 989–992 (1996).
- Palmer, C.N. *et al.* Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat. Genet.* **38**, 441–446 (2006).
- Weidinger, S. *et al.* Filaggrin mutations, atopic eczema, hay fever, and asthma in children. *J. Allergy Clin. Immunol.* **121**, 1203–1209 (2008).
- Nomura, T. *et al.* Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J. Allergy Clin. Immunol.* **119**, 434–440 (2007).

16. Morar, N., Cookson, W.O., Harper, J.I. & Moffatt, M.F. Filaggrin mutations in children with severe atopic dermatitis. *J. Invest. Dermatol.* **127**, 1667–1672 (2007).
17. Thurmond, R.L., Gelfand, E.W. & Dunford, P.J. The role of histamine H1 and H4 receptors in allergic inflammation: the search for new antihistamines. *Nat. Rev. Drug Discov.* **7**, 41–53 (2008).
18. Nagai, Y. *et al.* Requirement for MD-1 in cell surface expression of RP105/CD180 and B-cell responsiveness to lipopolysaccharide. *Blood* **99**, 1699–1705 (2002).
19. Pearce, E.L. *et al.* Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science* **302**, 1041–1043 (2003).
20. Hanifin, J.M. & Rajka, G. Diagnostic features of atopic dermatitis. *Acta Derm. (Stockholm)* **92**(Suppl.), 44–47 (1980).
21. Williams, H.C. *et al.* The U.K. Working Party's diagnostic criteria for atopic dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. *Br. J. Dermatol.* **131**, 383–396 (1994).
22. Krawczak, M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **9**, 55–61 (2006).
23. ETAC Study Group. Allergic factors associated with the development of asthma and the influence of cetirizine in a double-blind, randomised, placebo-controlled trial: first results of ETAC. Early Treatment of the Atopic Child. *Pediatr. Allergy Immunol.* **9**, 116–124 (1998).
24. Hill, D.J. *et al.* Confirmation of the association between high levels of immunoglobulin E food sensitization and eczema in infancy: an international study. *Clin. Exp. Allergy* **38**, 161–168 (2008).
25. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
26. Heath, S.C. *et al.* Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet.* **16**, 1413–1429 (2008).
27. Fisher, R.A. *Statistical Methods for Research Workers* (Oliver & Boyd, Edinburgh, 1932).
28. Petitti, D.B. *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine* (Oxford, New York, 1994).
29. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
30. Chen, W.M. & Deng, H.W. A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genet. Epidemiol.* **21**, 53–67 (2001).

A common variant on chromosome 11q13 is associated with atopic dermatitis

Jorge Esparza-Gordillo^{1,2,16}, Stephan Weidinger^{3,4,16}, Regina Fölster-Holst⁵, Anja Bauerfeind², Franz Ruschendorf², Giannino Patone², Klaus Rohde², Ingo Marenholz^{1,2}, Florian Schulz^{1,2}, Tamara Kerscher^{1,2}, Norbert Hubner², Ulrich Wahn¹, Stefan Schreiber^{6,7}, Andre Franke⁶, Rainer Vogler⁷, Simon Heath⁸, Hansjörg Baurecht^{4,9}, Natalija Novak¹⁰, Elke Rodriguez^{3,4}, Thomas Illig¹¹, Min-Ae Lee-Kirsch¹², Andrzej Ciechanowicz¹³, Michael Kurek¹⁴, Tereza Piskackova¹⁵, Milan Macek¹⁵, Young-Ae Lee^{1,2}, Andreas Ruether⁶

¹ Pediatric Pneumology and Immunology, Charité Universitätsmedizin Berlin, Berlin, Germany

² Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin-Buch, Germany

³ Department of Dermatology and Allergy, Technische Universität München, Munich, Germany

⁴ Division of Environmental Dermatology and Allergy, Helmholtz Zentrum Munich and ZAUM-Center for Allergy and Environment, Technische Universität München, Munich, Germany

⁵ Clinic for Dermatology, Venerology and Allergology, University Hospital Schleswig-Holstein, Kiel, Germany

⁶ Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany

⁷ POPGEN Biobank Project, Christian-Albrechts-University, Kiel, Germany

⁸ Centre National de Génotypage, Evry, France

⁹ Institute for Medical Statistics and Epidemiology IMSE, Technische Universität München, Munich, Germany

¹⁰ Department of Dermatology and Allergy, University of Bonn, Bonn, Germany

¹¹ Department of Epidemiology, Helmholtz Zentrum Munich-German Research Center for Environmental Health, Neuherberg, Germany

¹² Klinik für Kinder- und Jugendmedizin, Technical University Dresden, Dresden, Germany

¹³ Dept. Laboratory Diagnostics and Molecular Medicine, Pomeranian Medical University, Szczecin, Poland

¹⁴ Department of Clinical Allergology, Pomeranian, Pomeranian Medical University, Szczecin, Poland

¹⁵ Department of Biology and Medical Genetics, Charles University Prague - 2. Medical School and Faculty, Hospital Motol, Prague, Czech Republic

¹⁶ These authors contributed equally to this work

Corresponding author:

Prof. Dr. Young-Ae Lee

Pediatric Pneumology and Immunology

Charité

Campus Virchow Klinikum

Augustenburger Platz 1

D-13353 Berlin

Germany

Phone: +49-30-9406 35 12

Fax: +49-30-9406 31 47

Email: yolee@mdc-berlin.de

1. Human research subject

2. Genotyping

3. Quality control of samples
4. Quality control of SNP data
5. Identification and correction of population stratification
6. Power calculations
7. References
8. Web resources
9. Supplementary tables and figures

1. Human research subjects

We evaluated 4 independent study groups. All cases and controls were of European origin. In all 4 study groups, the physician's diagnosis of eczema was made according to standard criteria in the presence of a chronic or chronically relapsing pruritic dermatitis with the typical morphology and distribution^{1,2}.

Study population	Number		Cases		Country of origin
	Cases	Controls	Mean age	Sex ratio (M/F)	
Discovery groups (GWA)					
Set 1	939	975	18	460/479	Germany
Set 2 ^a	529 (270 families)		8	259/270	Germany
Follow up groups					
Set 3	1,363	2,739	16	651/712	Germany
Set 4	1,274	1,218	6	664/610	Germany, Czech Republic, Poland
Total	4,105	4,932			

^a The number of individuals in set 2 refers to the number of AD cases (529) and the number of families (270).

2. Genotyping

All genotyping with Genome-Wide Human SNP array 5.0 was performed at Affymetrix service facility (South San Francisco, CA, USA) as a part of the German NGFN GWAS initiative (see press release 04-26-07 on http://www.ngfn.de/englisch/index_368.htm). Processing of the samples was done blind to case-control status and several HapMap samples with known genotype were included in each batch for quality control. Genotyping with Affymetrix Human mapping 500K arrays was done at Max-Delbrück Center for Molecular Medicine (Berlin, Germany). All data from 500K arrays correspond to the commercial release of the array.

Both arrays are based on an assay termed whole-genome sampling analysis (WGSA) developed for highly multiplexed SNP genotyping of complex DNA. This method reproducibly amplifies a subset of the human genome through a single primer amplification reaction using restriction enzyme digested, adapter-ligated human genomic DNA. The same set of 440,794 SNPs was genotyped with both 500K and 5.0 arrays. Although both genotyping platforms contain 59,744 additional SNPs (total 500,568 SNPs), these were masked in the array 5.0 default library file from Affymetrix due to poor performance. In this paper we will make reference exclusively to the high-quality set of 440,794 SNPs present in both 500K and 5.0 arrays.

Genotyping with the 500K array was performed following the protocol provided by the manufacturer. In brief, 2 aliquots of 250ng genomic DNA each were digested with NspI and StyI, respectively, an adaptor was ligated and molecules were PCR-amplified, purified, fragmented and biotin-labeled. Each enzyme preparation was then hybridized to the corresponding NspI or StyI array, containing 262,000 and 238,000 SNP-assays respectively. First analysis of the 500K data was performed using the Dynamic Model (DM) algorithm at $P = 0.33^3$. As recommended by Affymetrix, samples with a DM-call rate < 93% (per array) were repeated and excluded if they did not perform well after the second genotyping. After the DM quality check all samples were re-called with BRLMM (Bayesian Robust Linear Model with

Mahalanobis classifier), an improved calling algorithm subsequently developed by Affymetrix (Affymetrix white paper BRLMM).

Genotyping with 5.0 arrays was performed in Affymetrix service lab following the assay manual. In brief, the protocol was the same as described above with the following modifications: both NspI PCR product and StyI PCR product are combined prior to purification. The combined hybridization mix is injected into a single array. A preliminary DM-call rate of 86% was used to pass arrays for further analysis. Genotypes for passed arrays were then called with the BRLMM-P algorithm (Affymetrix white paper BRLMM-P). There were 4 unique positive controls corresponding to HapMap CEPH individuals that were genotyped a total of 41 times with the 5.0 array allowing us to calculate experimental reproducibility, trio accuracy, and HapMap concordance. The performance metrics were as follows (supplied by Affymetrix service facility):

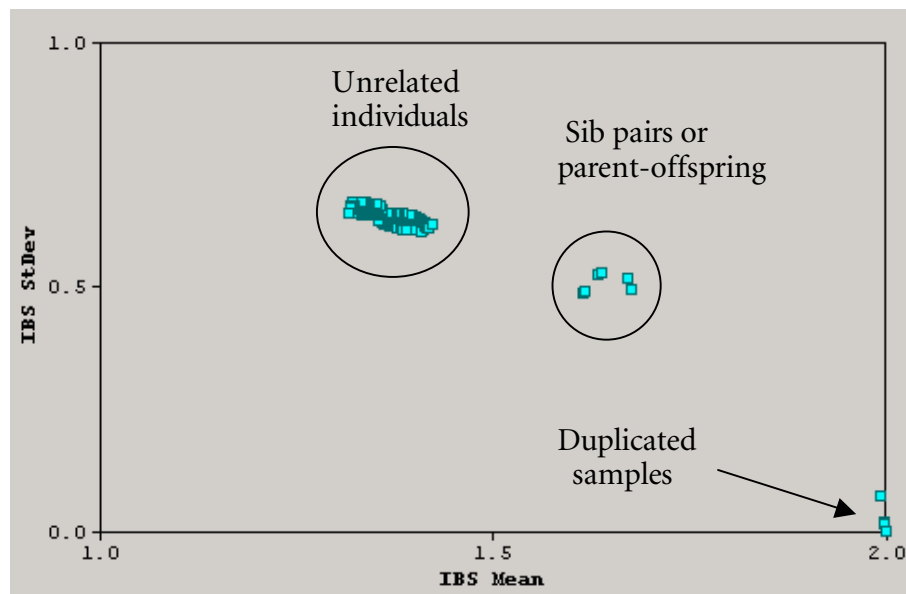
Average QC call rate	> 96.5%
Average BRLMM-p call rate	> 99.5%
Average reproducibility (controls)	> 99.2%
Average HapMap concordance (controls)	> 99.0%
Average Trio Accuracy (controls)	> 99.2%

Follow-up genotyping in set 3 was performed with ligation-based SNPlex as previously described⁴. Genotype assignments were verified visually using the Genemapper 4.0 (Applied Biosystems, Foster City, CA, USA) software. Samples in set 4 were genotyped using TaqMan allelic discrimination (Applied Biosystems, Foster City, CA, USA). Taqman assays for rs877776, rs7927894, and rs7024096 were obtained from Applied Biosystems by assay ids hCV1468787, hCV3243267, and hCV31290715, respectively. *FLG* mutations were genotyped using TaqMan allelic discrimination (Applied Biosystems, Foster City, CA, USA) as previously described^{5,6}.

3. Quality control of samples.

SNP data management, formatting and quality control was performed by a modified version of ALOHOMORA⁷. As recommended by Affymetrix, individual samples were excluded if the SNP call rate (% of completeness) was < 93% or < 86% after the initial analysis with the DM-algorithm for the samples genotyped with 500K and 5.0 arrays, respectively. Gender was assigned to every sample by counting the proportion of heterozygote calls on the X-chromosome outside the pseudoautosomal region. Samples were excluded when the sex from genotype data and clinical records was discordant.

The genetic relationship among family members and between unrelated individuals was estimated with the Graphical Representation of Relationships software (GRR)⁸. We selected 9,560 autosomal SNPs fulfilling quality control criteria (see below), separated by a minimal distance of 300 kb and with a minimal minor allele frequency of 0.1. For each pair of individuals GRR calculates the average and the standard deviation of the Identical-by-State (IBS) allele sharing over the 9,560 markers. The graphical plot of IBS mean versus IBS standard deviation allowed us to identify and exclude duplicate samples, individuals with strong evidence of non-European ancestry, and samples where the observed and expected genetic relationship were discordant (see figure below).



In this example we identified genetically related individuals ($IBS \cong 1.65$) and duplicated samples ($IBS = 2$) within a group of theoretically unrelated control individuals from set 1.

A more sensitive identification of individuals with non-European ancestry was done using the Structure software (see chapter 5). Complete information on the excluded samples is shown in the table below.

Set	Study population	Genotyping platform	Low call rate	Sex discordance	Population admixture	Duplicate samples	Related individuals	Total Excluded	Total Included
1	AD cases	5.0	7	4	18	11	5	45	939
	Popgen	5.0	na	0	3	0	7	10	482
	controls	500K	na	0	0	2	5	7	493
2	ASP families	500K	13	1	0	0	0	14	1097

Number of individuals excluded from each of the discovery sets and reason for exclusion. Low call rate: samples with $> 7\%$ or $> 14\%$ missing genotypes, for 500K and 5.0 arrays respectively. Sex discordance: discrepancy between clinical record and the gender derived from genotype data. Population admixture: samples with evidence of recent non-European admixture as identified with GRR and/or STRUCTURE software. ASP, Affected sib pairs. na, not available.

4. Quality control of SNP data

In order to minimize the number of false positives, testing for association with AD was done only after extensive quality control check of the genotype data. SNPs were excluded from further analysis according to 1) a minor allele frequency (MAF) < 1% in set 1 cases and controls (54,279 SNPs excluded), 2) SNP call rate < 95% in set 1 or 2 (22,747 SNPs excluded), 3) SNP call rate < 99% for markers with MAF between 1 and 5% (16,771 SNPs excluded), 4) deviation from Hardy-Weinberg equilibrium calculated with Pedstats⁹ ($P_{HWE} < 0.0005$ in set 1 controls; 5,592 SNPs excluded), 5) excess heterozygosity in controls ($H_z > 0.55$; 429 SNPs excluded), 6) markers located on chromosomes X or Y were excluded from the present analysis (9,786 SNPs excluded).

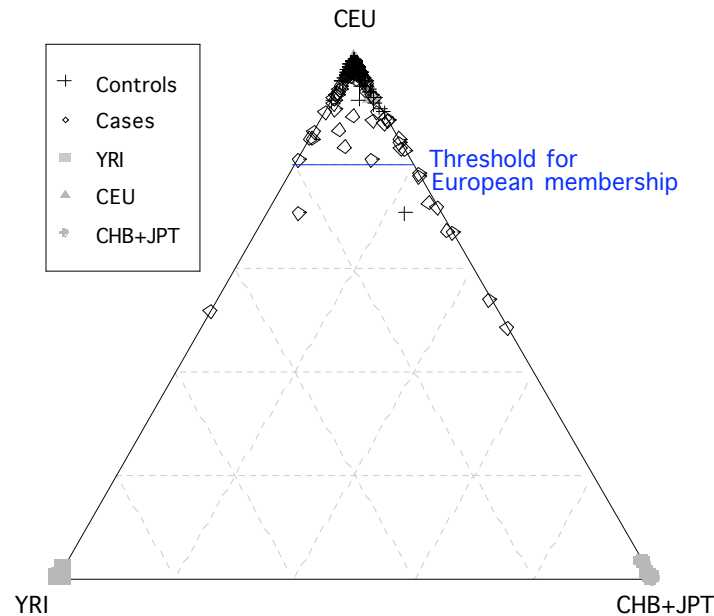
Additional quality control criteria were obtained from the family data set 2: 7) Mendelian errors were detected with PedCheck¹⁰ and erroneous genotypes were deleted. All SNPs with Mendelian errors in more than 9 out of 270 families were excluded from the study (3,377 SNPs excluded). 8) Unlikely genotypes (or double recombinants) were detected with Merlin¹¹ and deleted. All markers with unlikely genotypes in more than 9 families were excluded (4,965 exclusions). Note that any given marker may have been excluded by more than one of these filters.

The genotype concordance rate between the 500K and 5.0 arrays was evaluated by genotyping 82 samples with both platforms. These 82 pairs of DNAs were separate aliquots from the same DNA preparation and all yielded genotype calling rate above the correspondent quality threshold. We found an average concordance rate of 99.7 % (36,044,455 concordant genotype calls out of 36,145,108 total genotypes). This figure was in agreement with the 99.8% concordance rate reported from a similar analysis performed by Affymetrix with 270 HapMap samples (Affymetrix technical note). However, 8,753 markers with more than 2 discordant genotypes in this set of 82 genotype-pairs were excluded from further analysis, yielding a final genotype concordance rate of 99.9%. As an additional quality check we compared the genotype distribution between the set 1 Popgen individuals genotyped with 500K arrays versus those

genotyped with 5.0 arrays. Sixty markers yielding a Cochran-Armitage trend test $P < 0.00001$ were removed from the analysis. A total of 342,303 markers fulfilled quality control criteria and showed a 99.6% and 99.0% genotype completion rate in sets 1 and 2, respectively. SNP genotype cluster graphs were manually reviewed with Genotyping Console 1.0 (Affymetrix) for all SNPs included in set 3, in order to confirm quality of the genotype data.

5. Identification and correction of population stratification

In set 1 case-control analysis between autosomal SNPs and AD was performed using the Cochran-Armitage trend test. This test may yield inflated P values if hidden genetic structure exists within the analyzed sample (genetic differences between cases and controls not related to disease status). To correct for possible population stratification due to ancestry differences, cases and controls were analyzed jointly using the unsupervised clustering algorithm implemented in the program STRUCTURE (version 2.2)¹². All individuals were grouped into genetic clusters and each individual was given an estimated membership coefficient for each cluster corresponding to the fraction of his or her genome inferred to have ancestry in the cluster. To better discriminate between the different ethnicities potentially hidden within our data we enriched our dataset with genotypes obtained from individuals of the populations studied by the HapMap project¹³. We ran 30,000 burn-in periods followed by 50,000 Markov Chain Monte Carlo replications on two independent sets of 2,200 randomly selected SNPs each, distributed among all autosomes (100 SNPs/autosome). All SNPs fulfilled quality control criteria and had a call rate of 100%, $MAF > 5\%$, and were separated by at least 150 kb, and not associated with AD ($P_{ATT} > 0.05$). Both sets identified the same group of 13 individuals with a membership coefficient representing European ancestry of lower than 0.8 (80%) that were excluded from the analysis. Classification of individuals with STRUCTURE using one SNP set is shown below.



The plot shows the STRUCTURE result after analysis of 2,198 individuals (952 set 1 cases, 976 set 1 controls and 270 individuals from the HapMap project). HapMap data includes individuals from 4 different populations and/or geographic origins (YRI = Yoruba in Ibadan, Nigeria; JPT = Japanese in Tokyo, Japan; CHB = Han Chinese in Beijing, China; CEU or CEPH = Utah residents with ancestry from northern and western Europe). Ancestry of most case and control subjects was similar to that of the HapMap CEU individuals, i.e. the membership coefficient representing European ancestry (top corner) was higher than 0.80 (threshold; blue line). 13 individuals outside the CEU cluster were identified and excluded from further analysis.

Subsequent principal component analysis with EIGENSTRAT^{14,15} did not provide evidence of further (more subtle) genetic stratification of the data, i.e. no systematic differences between the assignment of cases and controls could be detected after removal of the previously identified outliers. However, we observed a moderate variance inflation ($\lambda = 1.14$) in the observed distribution of the test statistic (median χ^2_{1df} of the Cochran Armitage trend test) in set 1, when compared to the median of the expected distribution under the null hypothesis of no association (see probability-probability plot in Supplementary Fig. 1). Our findings are consistent with the results of Steffens et al.¹⁶ who showed in a detailed study of genetic substructure in the German population that individuals from different geographic regions of Germany could not be

separated via algorithms modelling ancestry differences (i.e. STRUCTURE), but nevertheless showed marginal heterogeneity measured by values of variance inflation. After genomic control correction no evidence of inflation remained in the set 1 results, suggesting that no systematic bias exist in the dataset.

All samples included in set 1 and 3 were of German origin and were recruited in the same study centers. Since genome wide genotypes were unavailable in set 3, we used the value of variance inflation derived in set 1 ($\lambda = 1.14$) as an approximation of population structure to correct P values in set 3.

In addition, to address population stratification in set 4, we estimated the variance inflation from control samples genotyped at the Centre National de Genotypage (Evry, France) on the Illumina Infinium HapMap 300 arrays for unrelated association studies. We estimated lambda based on 103,772 SNPs selected for high call rate, high informativity, and low LD between markers as previously reported¹⁷.

To simulate the sample composition in our set 4, available control individuals were randomly labelled as cases or controls to yield a German sample of 358 cases/297 controls, a Polish population of 311 cases/311 controls, and a Czech group of 179 cases/197 controls. We calculated the variance inflation factor lambda for each country pair and for the combined sample:

	lambda
Germany-Poland:	0.99
Germany-Czech Republic:	1.01
Czech Republic-Poland:	0.99
Czech Republic-Germany-Poland:	1.00

Although the genotype data used in this calculation did not correspond to the individuals in set 4, the observed lambda values (0.99 to 1.01) suggest that no major inflation in the test statistic is expected to occur in a combined case-control study with approximately equal number of cases and controls originating from Germany, Poland and the Czech Republic. Thus, we did not correct the results in set 4 for genetic stratification.

6. Power calculations

Power calculations were performed considering a 10% disease prevalence, 0.2 allele frequency and linkage disequilibrium D-prime of 0.95 between the genotyped marker and the risk variant. Power estimation was performed under an additive model with three alternative scenarios with different relative risks for the heterozygote (Aa) or homozygote carriers (AA) of the risk variant, respectively. Power calculation for sets 1 and 3 (case control study) was performed with the Genetic Power Calculation Software¹⁸. The method described by Chen et al.¹⁹ was used to compute power for set 2 (family-based sample). For all tree sets we calculated the power to detect a risk variant with 3 different type I error rates (alpha; see table below).

	Set 1			Set 2			Set 3			Set 4		
Relative risk Aa	1.2	1.4	1.5	1.2	1.4	1.5	1.2	1.4	1.5	1.2	1.4	1.5
Relative risk AA	1.44	1.96	2.25	1.44	1.96	2.25	1.44	1.96	2.25	1.44	1.96	2.25
alpha 0.05	0.60	0.99	1.00	0.36	0.88	0.97	0.86	1.00	1.00	0.71	1.00	1.00
alpha 0.005	0.27	0.91	0.99	0.12	0.62	0.84	0.60	1.00	1.00	0.38	0.97	1.00
alpha 0.00001	0.01	0.40	0.74	0.00	0.09	0.30	0.09	0.92	1.00	0.03	0.62	0.91

7. References

1. Hanifin, J. M. & Rajka, G. Diagnostic Features of Atopic Dermatitis. *Acta Derm.(Stockholm)* **92 (Suppl.)**, 44-47 (1980).
2. Williams, H. C. *et al.* The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. *Br.J.Dermatol.* **131**, 383-396 (1994).
3. Di, X. *et al.* Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* **21**, 1958-1963 (2005).
4. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat.Genet.* **39**, 207-211 (2007).
5. Palmer, C. N. *et al.* Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat.Genet.* **38**, 441-446 (2006).
6. Sandilands, A. *et al.* Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat.Genet.* **39**, 650-654 (2007).
7. Ruschendorf, F. & Nurnberg, P. ALOHOMORA: a tool for linkage analysis using 10K SNP array data. *Bioinformatics* **21**, 2123-2125 (2005).
8. Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. GRR: graphical representation of relationship errors. *Bioinformatics* **17**, 742-743 (2001).

9. Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445-3447 (2005).
10. O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am.J.Hum.Genet.* **63**, 259-266 (1998).
11. Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat.Genet.* **30**, 97-101 (2002).
12. Pritchard, J. K., Stephens, M., & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).
13. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
14. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat.Genet.* **38**, 904-909 (2006).
15. Patterson, N., Price, A. L., & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
16. Steffens, M. *et al.* SNP-based analysis of genetic substructure in the German population. *Hum.Hered.* **62**, 20-29 (2006).
17. Heath, S. C. *et al.* Investigation of the fine structure of European populations with applications to disease association studies. *Eur.J.Hum.Genet.* **16**, 1413-1429 (2008).
18. Purcell, S., Cherny, S. S., & Sham, P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-150 (2003).

19. Chen, W. M. & Deng, H. W. A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genet.Epidemiol.* **21**, 53-67 (2001).
20. Petitti, D. B. *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*Oxford University Press, New York (1994).
21. Fisher, R. A. *Statistical Methods for Research Workers*Oliver & Boyd, Edinburgh (1932).
22. Spielman, R. S., McGinnis, R. E., & Ewens, W. J. The transmission/disequilibrium test detects cosegregation and linkage. *Am.J.Hum.Genet.* **54**, 559-560 (1994).

8. Web Resources

Affymetrix support information:

Annotation files:

http://www.affymetrix.com/support/technical/byproduct.affx?product=genomewidesnp_5

<http://www.affymetrix.com/support/technical/byproduct.affx?product=500K>

Concordance across Affymetrix genotyping arrays:

http://www.affymetrix.com/support/technical/technotes/genotyping_concordance_technote.pdf

BRLMM: an Improved Genotyping Calling Method for the GenChip ♦ Human Mapping 500K Array set (Affymetrix white paper Version 1.0 2006)

http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf

BRLMM-P: a Genotype Calling Method for the SNP 5.0 array (Affymetrix white paper version 1.0 2007). http://www.affymetrix.com/support/technical/whitepapers/brlmp_whitepaper.pdf

Additional online resources:

ALOHOMORA software: <http://gmc.mdc-berlin.de/alohomora/>

GRR software: <http://www.sph.umich.edu/csg/abecasis/GRR/index.html>

Haploview software: <http://www.broad.mit.edu/mpg/haploview/index.php>

Merlin software: <http://www.sph.umich.edu/csg/abecasis/Merlin/>

PedCheck software: http://watson.hgen.pitt.edu/register/soft_doc.html

PLINK software: <http://pngu.mgh.harvard.edu/purcell/plink/>

UniGene database: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>

RefSeq database: <http://www.ncbi.nlm.nih.gov/RefSeq/>

NCBI dbSNP: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp&TabCmd=Limits>

Genetic power calculator: <http://pngu.mgh.harvard.edu/~purcell/gpc/>

Online Mendelian Inheritance in Man (OMIM): <http://www.ncbi.nlm.nih.gov/omim/>

9. Supplementary tables and figures

Supplementary Table 1. Association results for all SNPs tested in set 3

dbSNP	Chr	Position	P_{set1}	P_{set2}	P_{set3}	P_{set3} corrected	Upstream/ Downstream Gene
rs1292657	1	10815668	0.004	0.025	0.179	1	<i>CASZ1 / C1orf127</i>
rs2809964	1	24748077	0.004	0.009	0.096	1	<i>RCAN3 / C1orf130</i>
rs17424028	1	47990556	0.006	0.015	0.326	1	<i>FOXD2 / SLC5A9</i>
rs6675297	1	102320706	0.034	4.7×10^{-4}	0.087	1	<i>OLFM3</i>
rs6661961	1	150708913	1.2×10^{-4}	0.003	2.9×10^{-5}	0.002	<i>CRNN / LCE5A</i>
rs12468708	2	58740543	0.005	0.040	0.887	1	<i>FANCL</i>
rs1186830	2	61823751	0.049	0.004	0.718	1	<i>XPO1 / FAM161A</i>
rs10168064	2	66861815	0.004	0.012	0.203	1	<i>MEIS1</i>
rs17031438	2	105625456	2.1×10^{-4}	0.045	0.268	1	<i>FHL2 / NCK2</i>
rs17767728	2	134171845	0.040	0.004	0.385	1	<i>NAP5</i>
rs6744478	2	138727290	0.004	0.012	0.601	1	<i>HNMT / SPOPL</i>
rs6755553	2	202673605	0.015	0.003	0.777	1	<i>FZD7 / SUMO1</i>
rs2140502	3	781930	0.035	0.002	0.086	1	<i>CHL1 / CNTN6</i>
rs12498098	3	27573653	0.005	0.025	0.023	1	<i>SLC4A7 / EOMES</i>
rs17061003	3	59615236	0.016	0.003	0.298	1	<i>FHIT</i>
rs6578	3	195557301	0.045	0.005	0.241	1	<i>LRRC15</i>
rs2116294	3	198774555	0.019	2.6×10^{-4}	0.646	1	<i>BDH1</i>
rs4140904	4	17925090	0.005	0.023	0.178	1	<i>LCORL</i>
rs2348616	4	160859765	0.023	0.004	0.946	1	<i>RAPGEF2</i>
rs1960980	4	165615386	0.004	0.034	0.968	1	<i>ANP32C / TRIM61</i>
rs1518186	4	168831285	0.003	0.039	0.037	1	<i>SPOCK3 / ANXA10</i>
rs13123569	4	174626194	0.016	0.003	0.307	1	<i>SCRG1 / HAND2</i>

rs12200661	6	6363963	0.002	0.039	0.012	0.624	<i>F13A1 / LY86</i>
rs982569	6	77881910	0.004	0.013	0.611	1	<i>HTR1B</i>
rs10277702	7	141901402	0.011	0.005	0.674	1	<i>MOXD2 / PRSS1</i>
rs219247	7	150620300	0.003	0.034	0.331	1	<i>SMARCD3 / NUB1</i>
rs1157071	8	137017080	0.005	0.038	0.205	1	<i>KHDRBS3</i>
rs7024096	9	22389693	8.3×10^{-4}	0.044	4.4×10^{-4}	0.024	<i>CDKN2B / DMRTA1</i>
rs872875	10	8425250	0.008	0.043	0.561	1	<i>GATA3</i>
rs17665422	10	26193319	7.1×10^{-4}	0.016	0.414	1	<i>GPR158 / MYO3A</i>
rs2589446	10	56634203	6.9×10^{-4}	0.039	0.118	1	<i>PCDH15</i>
rs2116581	10	63684101	0.007	0.007	0.547	1	<i>RTKN2</i>
rs17749003	11	32463509	0.038	0.002	0.889	1	<i>WIT1 / EIF3M</i>
rs7927894	11	75978964	0.027	0.001	2.5×10^{-4}	0.013	<i>C11orf30 / LRRC32</i>
rs10790431	11	120660280	0.001	1.7×10^{-4}	0.238	1	<i>TECTA / SC5DL</i>
rs485868	11	127617875	0.004	0.047	0.485	1	<i>ETS1</i>
rs11169832	12	50094494	0.003	0.043	0.167	1	<i>GALNT6 / SLC4A8</i>
rs1106724	12	109559086	0.004	0.045	0.044	1	<i>TCTN1</i>
rs3825208	12	112771831	3.4×10^{-4}	0.029	0.437	1	<i>RMB19</i>
rs4770949	13	25663096	0.010	0.004	0.611	1	<i>SHISA2 / RNF6</i>
rs7321748	13	75260740	0.004	0.005	0.601	1	<i>LMO7</i>
rs3963390	13	87990321	0.001	0.005	0.662	1	<i>no gene</i>
rs2286409	14	70864498	1.0×10^{-4}	0.009	0.036	1	<i>PCNX / SIPA1L1</i>
rs4904889	14	91938114	0.0045	0.041	0.653	1	<i>SLC24A4</i>
rs1435831	15	24031805	8.3×10^{-4}	0.038	0.335	1	<i>ATP10A / GABRB3</i>
rs10518763	15	40427287	0.003	0.041	0.541	1	<i>GANC</i>
rs8023980	15	66720131	0.003	0.032	0.067	1	<i>CORO2B</i>

rs12440979	15	78407894	0.045	0.004	0.083	1	<i>FAH / ARNT2</i>
rs10492808	16	20639822	0.003	0.032	0.486	1	<i>ACSM1 / THUMPD1</i>
rs10853647	18	20697540	0.006	0.028	0.009	0.503	<i>HRH4 / ZNF521</i>
rs1542038	19	61676090	0.004	0.024	0.896	1	<i>ZNF667</i>
rs8115172	20	4338880	0.005	0.033	0.558	1	<i>ADRA1D / PRNP</i>
rs3787508	20	16656164	0.008	0.033	0.479	1	<i>KIF16B / SNRPB2</i>
rs2823642	21	16502160	0.039	0.003	0.968	1	<i>C21orf34</i>

Chromosomal location refers to the NCBI Build 36.1, March 2006 Human Genome assembly.

P_{set1} and P_{set3} indicates the P_{ATT} value after genomic control correction ($\lambda = 1.14$). $P_{\text{set3 corrected}}$

indicates the P_{ATT} value in set 3 after Bonferroni correction for the 54 tests performed.

Upstream/downstream gene refers to the nearest annotated gene in the RefSeq database. No gene indicates the absence of annotated genes in 1 Mb of genomic sequence surrounding the SNP. Six markers could not be typed with SNPlex and were replaced by a proxy SNP selected from Hapmap on the basis of high LD (rs1292657 was tagged by rs946684; rs1186830 by rs1186841; rs13123569 by rs1442158; rs1157071 by rs17546210; rs11169832 by rs11169831; rs4770949 by rs4770948).

Supplementary Table 2. Detailed association results for rs7024096 (9p21) and rs7927894 (11q13)

dbSNP	Chr	Position	dbSNP strand	MA	set	MAF or T:U		OR for RA (95% CI)	P value	Gene
						cases	controls			
rs7024096	9	22,389,693	Top	G	1 ^a	0.293	0.240	1.28 (1.12-1.49)	8.3 x 10 ⁻⁴	<i>DMRTA1</i>
					2	167 : 205		1.23 (1.00-1.52)	0.044	
					3 ^a	0.319	0.278	1.21 (1.10-1.34)	4.4 x 10 ⁻⁴	
					4	0.301	0.291	1.05 (0.93-1.19)	0.43	
					Combined ^{b,c}	0.310	0.273	1.18 (1.10-1.26)	9.06 x 10 ⁻⁶	
rs7927894	11	75,978,964	Top	A	1 ^a	0.372	0.409	1.18 (1.03-1.33)	0.027	<i>C11orf30</i>
					2	294 : 226		1.30 (1.09-1.55)	0.0013	
					3 ^a	0.363	0.407	1.21 (1.10-1.33)	2.5 x 10 ⁻⁴	
					4	0.357	0.415	1.29 (1.15-1.45)	1.94 x 10 ⁻⁵	
					Combined ^{b,c}	0.363	0.411	1.22 (1.15-1.30)	7.6 x 10 ⁻¹⁰	

Chromosomal location refers to the NCBI Build 36.1, March 2006 Human Genome assembly. dbSNP strand, DNA strand from the NCBI SNP database; MA, minor allele; MAF, minor allele frequency; T:U, number of transmissions versus non-transmissions of the minor allele. Odds ratios (OR) were calculated by logistic regression analysis and *P* values with the Cochran-Armitage trend test. For set 2, the number of transmissions versus non-transmissions are shown (T:U). OR in set 2 represents the ratio of transmitted/untransmitted alleles.

^aResults for set 1 and set 3 were adjusted by the genomic control method ($\lambda=1.14$, Supplementary Fig. 1). ^bFor the combined analysis the OR and *P* values were calculated using the inverse variance²⁰ and Fisher's method, respectively²¹. ^cResults from the family-based set 2 were not included in the calculation of allele frequencies and OR for the combined population. Gene refers to the nearest annotated gene in the RefSeq database.

Supplementary Table 3. Detailed association results for all loci nominally replicated in set 3

dbSNP	Chr	Position	dbSNP strand	MA	Set	MAF or T:U		OR for RA (95% CI)	P value	P _{combined} ^{b,c}	Gene
						Cases	Controls				
rs6661961	1	150708913	Top	A	1 ^a	0.498	0.430	1.30 (1.15-1.49)	1.17 x 10 ⁻⁴	3.95 x 10 ⁻⁹	<i>CRNN / LCE5A</i>
					2	281:217		1.30 (1.09-1.55)	0.0033		
					3 ^a	0.484	0.432	1.23 (1.12-1.35)	2.9 x 10 ⁻⁵		
rs12498098	3	27573653	Bottom	C	1 ^a	0.155	0.122	1.32 (1.10-1.59)	0.0055	2.96 x 10 ⁻⁴	<i>SLC4A7 / EOMES</i>
					2	161:125		1.29 (1.02-1.63)	0.0250		
					3 ^a	0.158	0.138	1.18 (1.03-1.33)	0.0232		
rs1518186	4	168831285	Top	G	1 ^a	0.292	0.245	1.26 (1.09-1.45)	0.0028	4.77 x 10 ⁻⁴	<i>SPOCK3 / ANXA10</i>
					2	224:181		1.24 (1.02-1.51)	0.0394		
					3 ^a	0.282	0.260	1.11 (1.01-1.23)	0.0368		
rs12200661	6	6363963	Top	A	1 ^a	0.388	0.442	1.26 (1.10-1.43)	0.0016	8.20 x 10 ⁻⁵	<i>F13A1 / LY86</i>
					2	199:245		1.23 (1.02-1.49)	0.0388		
					3 ^a	0.399	0.431	1.14 (1.04-1.25)	0.0115		
rs7024096	9	22389693	Top	G	1 ^a	0.240	0.293	1.28 (1.12-1.49)	8.30 x 10 ⁻⁴	2.91 x 10 ⁻⁶	<i>CDKN2B / DMRTA1</i>
					2	167:205		1.23 (1.00-1.52)	0.0439		

					3 ^a	0.278	0.319	1.21 (1.10-1.34)	4.4 x 10 ⁻⁴	
					1 ^a	0.409	0.372	1.18 (1.03-1.33)	0.0266	
rs7927894	11	75978964	Top	A	2	294:226		1.30 (1.09-1.55)	0.0013	1.61 x 10 ⁻⁶ <i>C11orf30 / LRRC32</i>
					3 ^a	0.407	0.363	1.21 (1.10-1.33)	2.4 x 10 ⁻⁴	
					1 ^a	0.207	0.249	1.28 (1.10-1.49)	0.0035	
rs1106724	12	109559086	Bottom	G	2	163:201		1.23 (1.00-1.52)	0.0450	7.41 x 10 ⁻⁴ <i>TCTN1</i>
					3 ^a	0.213	0.233	1.12 (1.00-1.25)	0.0436	
					1 ^a	0.022	0.006	3.70 (1.92-7.14)	1.01 x 10 ⁻⁴	
rs2286409	14	70864498	Bottom	C	2	23:09		2.56 (1.18-5.52)	0.0089	5.43 x 10 ⁻⁶ <i>PCNX / SIPA1L1</i>
					3 ^a	0.020	0.014	1.49 (1.05-2.13)	0.0363	
					1 ^a	0.450	0.497	1.20 (1.06-1.37)	0.0064	
rs10853647	18	20697540	Bottom	T	2	241:297		1.23 (1.04-1.47)	0.0279	1.71 x 10 ⁻⁴ <i>HRH4 / ZNF521</i>
					3 ^a	0.461	0.494	1.14 (1.04-1.26)	0.0093	

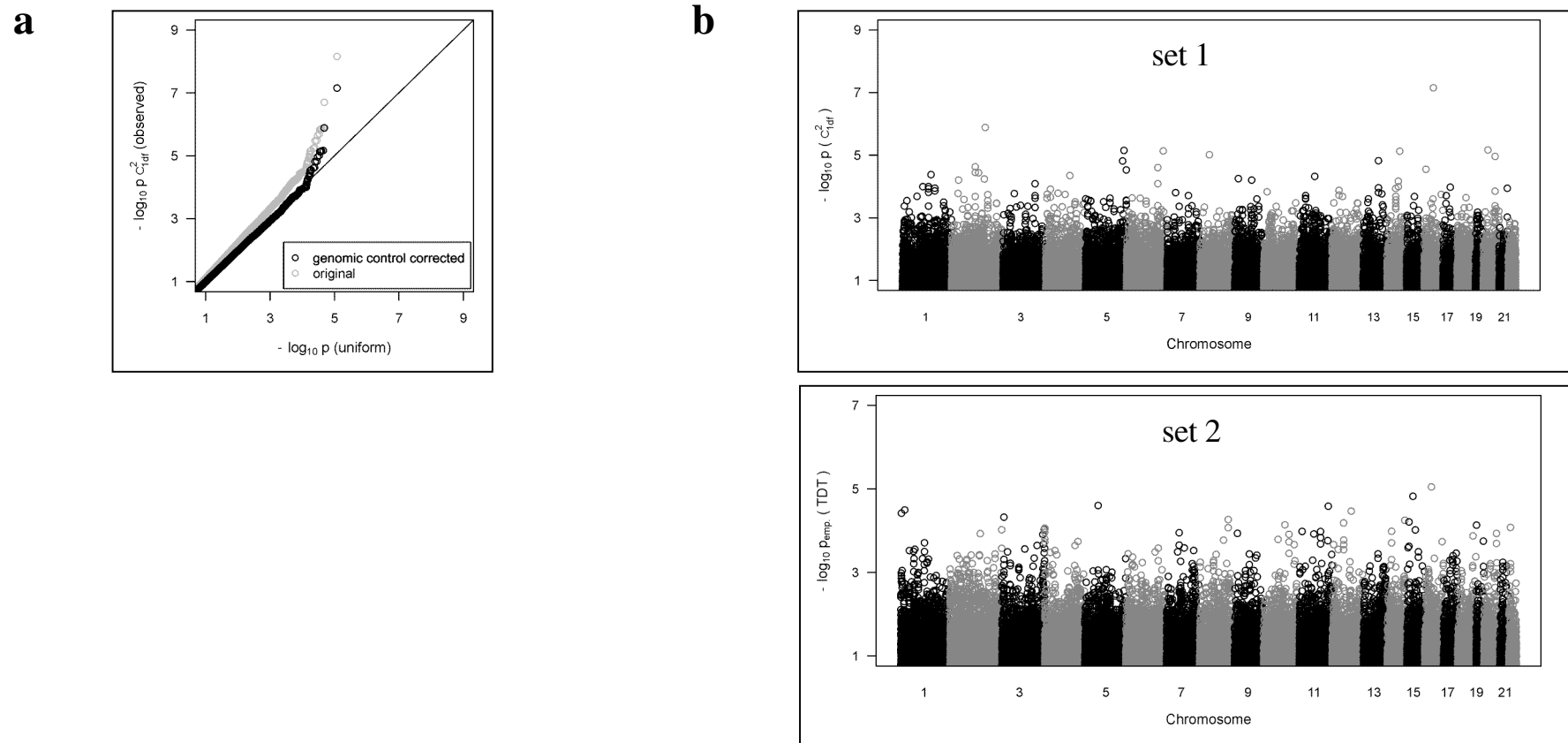
Chromosomal location refers to the NCBI Build 36.1, March 2006 Human Genome assembly. dbSNP strand, DNA strand from the NCBI SNP database; MAF, minor allele frequency; T:U, number of transmissions versus non-transmissions of the minor allele. Odds ratios (OR) were calculated by logistic regression analysis and *P* values with the Cochran-Armitage trend test. For set 2, the number of transmissions versus non-transmissions are shown (T:U). OR in set 2 represents the ratio of transmitted/untransmitted alleles.

^aResults for set 1 and set 3 were adjusted by the genomic control method ($\lambda=1.14$, Supplementary Fig. 1). ^bFor the combined analysis the OR and *P* values were calculated using the inverse variance²⁰ and Fisher's method, respectively²¹. ^cResults from the family-based set 2 were not included in the calculation of allele frequencies and OR for the combined population. Gene refers to the nearest annotated gene in the RefSeq database.

Supplementary Table 4. RT-PCR primers for *LRRC32* and *c11orf30*.

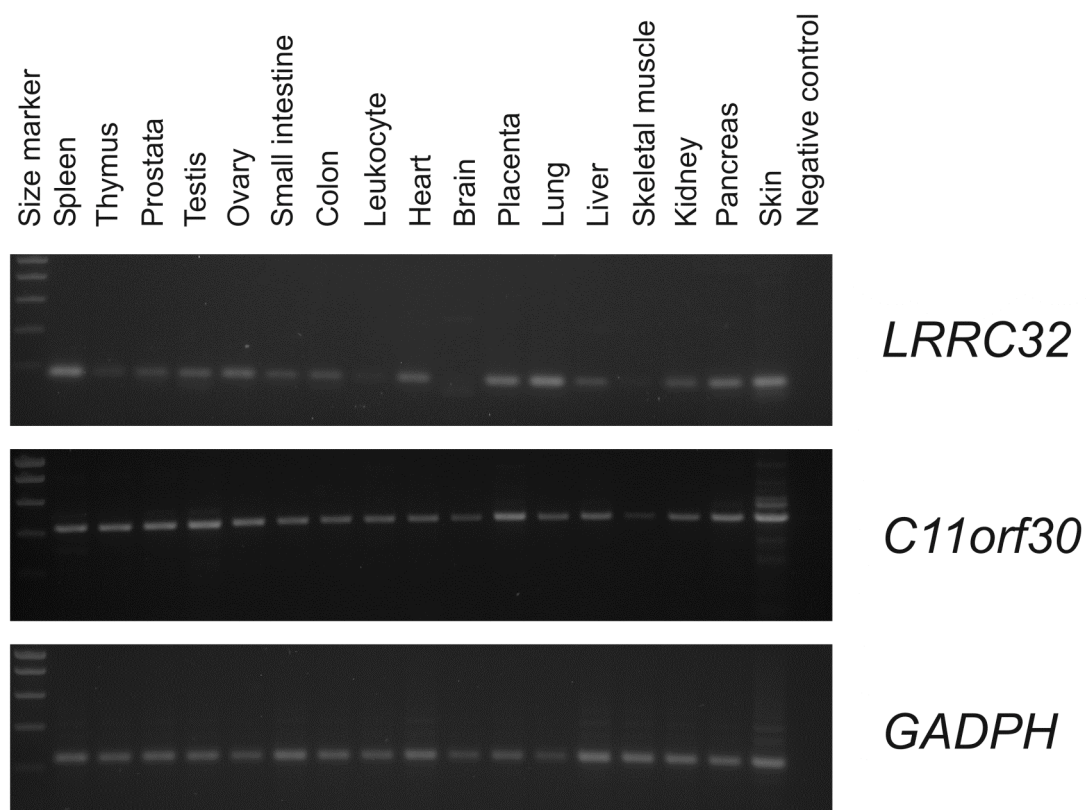
Gene	Orientation	Sequence 5' to 3'	Product size (bp)
<i>LRRC32</i>	FORWARD	GCTGCACAACACCAAGACAA	191
	REVERSE	AGGTCCAGGTGACGAAGTG	
<i>c11orf30</i>	FORWARD	CCCTCTCCAGCAAGAACAAG	415
	REVERSE	CTTCTCAGAGGCCTGAATGG	

Supplementary Fig. 1. Probability-probability plots (a) and genome-wide distribution of P values for association with AD (b)



Data represent all SNPs fulfilling quality control criteria. P values were calculated with the Cochran-Armitage trend test in set 1 (case-control). Family-based association in set 2 was tested using the TDT and empirical P values were obtained with a permutation procedure. a) Probability-probability plot comparing the observed P value distribution with the expected uniform distribution (null hypothesis) in set 1. The original P values from the Cochran-Armitage trend test are shown in grey and the genomic control-corrected P values (inflation factor = 1.14) are shown in black. Data from family-based set 2 was analyzed with the Transmission-Disequilibrium Test which is robust to population stratification²². b) Genome-wide plots of $-\log_{10} P$ of association with AD in set 1 (after genomic control correction) and set 2. Chromosomes are shown in alternating color for clarity.

Supplementary Fig. 2. Tissue expression pattern of *C11orf30* and *LRRC32*, and *GAPDH*



Gene expression analysis of *C11orf30*, *LRRC32*, and *GAPDH* by reverse transcription PCR analysis in 17 human tissues.

Supplementary Fig. 3. LD between *FLG* mutations and AD-associated markers in the EDC

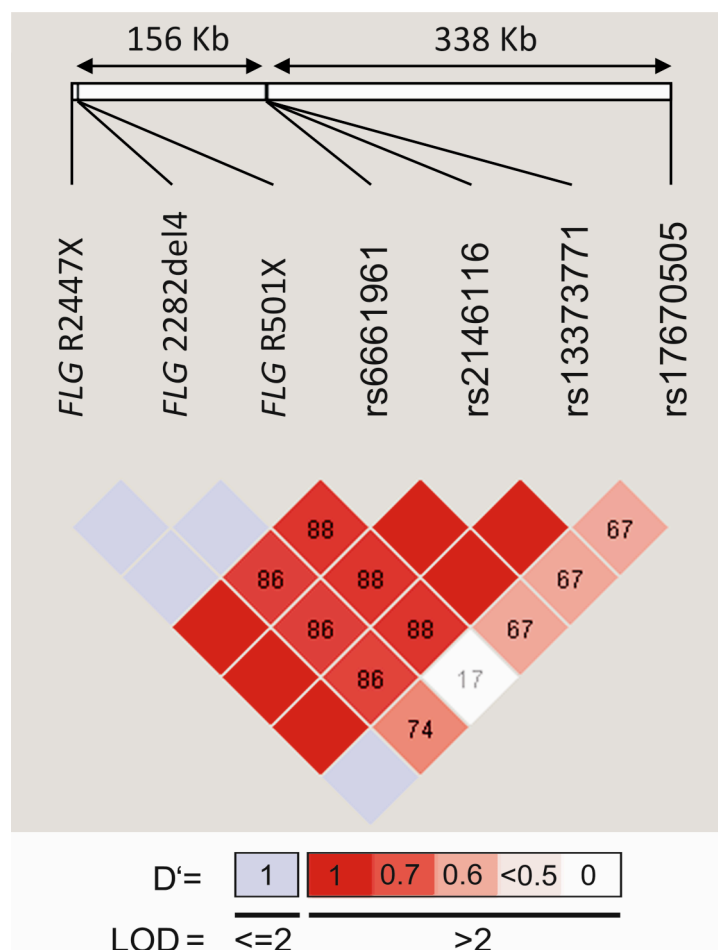


Figure shows the disequilibrium coefficient between *FLG* mutations and the 4 SNPs within the EDC yielding the strongest association results in set 1 and 2. Note that although the 4 SNPs fulfilled the criteria for inclusion in set 3 ($P < 0.005$ and $P < 0.05$), only one of them (rs6661961) was selected for genotyping. The plot was generated with haploview using set 2 data. The *FLG* mutation S3247X was not included since it is very rare and was not detected in set 2. Physical distance in kb is indicated at the top.