

# Joint genotype and ancestry analysis identify novel loci associated with atopic dermatitis in African American population

Yadu Gautam,<sup>1</sup> Latha Satish,<sup>1</sup> Stephen Ramirez,<sup>1</sup> Brittany Grashel,<sup>1</sup> Jocelyn M. Biagini,<sup>1</sup> Lisa J. Martin,<sup>2</sup> Marc E. Rothenberg,<sup>3</sup> Gurjit K. Khurana Hershey,<sup>1</sup> and Tesfaye B. Mersha<sup>1,4,\*</sup>

## Summary

Atopic dermatitis (AD) is a chronic itchy inflammatory disease of the skin. Genetic studies have identified multiple risk factors linked to the disease; however, most of the studies have been derived from European and East Asian populations. The admixed African American (AA) genome may provide an opportunity to discover ancestry-specific loci involved in AD susceptibility. Herein, we present joint analysis of ancestry and genotype effects followed by validation using differential gene expression analysis on AD using 726 AD-affected individuals and 999 non-AD control individuals from the AA population, genotyped using Multi-Ethnic Global Array (MEGA) followed by imputation using the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA) reference panel. The joint analysis identified two novel AD-susceptibility loci, rs2195989 in gene *ANGPT1* (8q23.1) and rs62538818 in the intergenic region between genes *LURAP1L* and *MPDZ* (9p23). Admixture mapping (AM) results showed potential genomic inflation, and we implemented genomic control and identified five ancestry-of-origin loci with European ancestry effects. The multi-omics functional prioritization of variants in AM signals prioritized the loci *SLAIN2*, *RNF39*, and *FOXA2*. Genome-wide association study (GWAS) identified variants significantly associated with AD in the AA population, including *SGK1* (rs113357522, odds ratio [OR] = 2.81), *EFR3A* (rs16904552, OR = 1.725), and *MMP14* (rs911912, OR = 1.791). GWAS variants were common in the AA but rare in the European population, which suggests an African-ancestry-specific risk of AD. Four genes (*ANGPT1*, *LURAP1L*, *EFR3A*, and *SGK1*) were further validated using qPCR from AD and healthy skin. This study highlighted the importance of genetic studies on admixed populations, as well as local ancestry and genotype-ancestry joint effects to identify risk loci for AD.

## Introduction

Atopic dermatitis (AD) is a chronic itchy inflammatory skin disease characterized by clinical symptoms including rash, severe itchiness, oozing and crusting, and skin pain. Physiologically, it is a multifactorial disease involving defective skin barrier function, immune dysregulation, environmental exposures, and viral infections.<sup>1</sup> It is one of the most common skin diseases affecting children and adversely impacts the quality of life of affected individuals and their families.<sup>2</sup> AD often starts during infancy and progresses into other atopic diseases, including food allergy, asthma, and allergic rhinitis, a phenomenon known as the atopic march.<sup>3</sup> The prevalence and clinical presentation of AD varies worldwide, ranging from 0.2% to 24.6% depending on geographic location and ethnicity.<sup>4</sup> The single most predictive factor of AD development is a positive family history of AD or allergic diseases. Parental history of AD is associated with 3- to 5-fold increase in the risk of AD among offspring.<sup>5</sup> Twin studies showed up to 86% concordance rate in monozygotic twins and approxi-

mately 22% concordance in dizygotic twins.<sup>6,7</sup> The heritability of AD was estimated to be ~75%, underscoring the genetic basis of AD.<sup>8,9</sup>

Several well-powered genome-wide association studies (GWASs) have identified and replicated genetic-susceptibility variants associated with AD. The genetic variations in the epidermal differentiation complex on chromosome 1q21.3 that encodes the filaggrin (*FLG* [MIM: 134950]) gene were reported by GWAS and candidate gene approaches to be associated with AD. Several loss-of-function mutations in the *FLG* gene associated with the defects in the skin barrier functions are the most studied known genetic risk factors for AD.<sup>10</sup> The defective skin barriers lead to increased transepidermal water loss and skin dehydration, and increased penetration of allergens and infectious microbes into the skin, which trigger the AD symptoms.<sup>11</sup> Recently, a genetic variant in *SPRR2B* (MIM: 182268) was found to be associated with AD risk and the risk showed an increase with the presence of the loss-of-function *FLG* mutation.<sup>12</sup> Other well-established GWAS associations are *CD207* (MIM: 604862) on chromosome 2p13.3, the cytokine cluster

<sup>1</sup>Division of Asthma Research, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA;

<sup>2</sup>Division of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA;

<sup>3</sup>Division of Allergy and Immunology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA

<sup>4</sup>Lead contact

\*Correspondence: [tesfaye.mersha@cchmc.org](mailto:tesfaye.mersha@cchmc.org)

<https://doi.org/10.1016/j.xhgg.2024.100350>.

© 2024 The Author(s). Published by Elsevier Inc. on behalf of American Society of Human Genetics.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



on chr5q31.1 including *IL13* (MIM: 147683) and *IL4* (MIM: 147780), and the MHC region on 6p21.3. GWAS meta-analysis on African American (AA) participants identified a single locus near *RORC* (MIM: 602943) in chromosome 1.<sup>13</sup> Candidate gene analysis showed variants in *TSLP* (MIM: 607003) and *IL7R* (MIM: 146661) were associated with AD in European and AA participants.<sup>14</sup> Immunochip analyses have also identified intergenic regions on chromosome 4q27 and chromosome 16p13.13, and *ZNF652* (MIM: 613907) on 17q21.32 as risk loci associated with AD.<sup>15</sup> Grosche et al. identified multiple exonic variants in *DUSP1* (MIM: 600714), *NOTCH4* (MIM: 164951), and *SLC9A4* (MIM: 600531) associated with AD.<sup>16</sup> Several large-scale GWASs and meta-analyses have identified shared genetic risk among AD, asthma, and hay fever.<sup>17-19</sup> Mucha et al. identified several rare protein-coding variations in genes *IL4R* (MIM: 147781), *IL13* and *JAK1* (MIM: 147795), *JAK2* (MIM: 147796), and *TYK2* (MIM: 176941) that were associated with AD.<sup>20</sup> Genomic findings to date account for ~30% of AD heritability<sup>21</sup> and much more remains to be discovered.

Many epidemiological studies have found racial and ethnic variations in AD prevalence. In the US, data from the 2003 National Survey of Children's Health involving more than 102,000 families demonstrated an overall prevalence of AD of 10.7% of children and that AD is more prevalent (15.9% vs. 9.7%) in AA children than European American (EA) children.<sup>22</sup> After adjusting for potential confounders/covariates, African ancestry was still significantly associated with a higher prevalence of AD compared to European ancestry (odds ratio [OR], 1.70,  $p = 0.005$ ).<sup>22</sup> A similarly disparity in prevalence of AD was found between Black Caribbean children (16.3%) vs. White children (8.7%) was observed in a UK study based on dermatologist examination.<sup>23</sup> A clinic-based study in south-east London found AD to be the most frequently diagnosed dermatologic condition in Black children, affecting 36.5% of those seen.<sup>24</sup> A multi-country cross-sectional study also found high prevalence, up to 23%, of AD among several African countries.<sup>25</sup> Studies also demonstrated that children of African descendants are three times more likely than White children to attend a dermatology visit during which AD is diagnosed, even though they are significantly less likely to seek dermatological care.<sup>26</sup> A recent study on two large AA cohorts showed that self-identified race of AA was associated with 2-fold increase in AD prevalence; however, global African ancestry as determined by the overall proportion of genome from the African population was not associated with AD.<sup>27</sup> In addition to epidemiologic data, results from immuno-histochemistry and transcriptomic studies revealed that AD in the AA population manifests as a distinct clinical subtype demarcated by significantly higher allergen-specific serum immunoglobulin (Ig) E and attenuated T helper (Th) 1 and Th17 molecular signatures, which discriminates it from EA and

Asian molecular AD subtypes.<sup>28,29</sup> Lesional skin samples from AA patients can be characterized by decreased expression of innate immune, Th1-related, and Th17-related markers but increased dendritic cell infiltration compared to EA and Asian patient samples.<sup>28</sup> Taken together, multiple lines of evidence from epidemiologic, transcriptomic, and immunologic studies indicate that African ancestry is significantly associated with increase in prevalence of AD in the AA population who carry genetic contributions from parental genomes coming from African and European lineages. Genomic analysis could therefore unravel the molecular basis underlying the discordant prevalence and molecular subtype of AD in AA and other racial groups leading to improved personalized care for AD.

The major drawback in the genetics studies of AD so far is that the AD-associated genetic variations are based on studies with participants of European and East Asian populations and there is a lack of representation of other ancestral populations, such as African, Native Americans, and recently admixed populations such as AA and Latinos. A 2015 GWAS used samples from multi-ancestry populations, but only ~2% of cases and <1% of control individuals were from AA ancestry.<sup>30</sup> The genomes of admixed populations consist of mosaics of chromosomal segments from the founding populations. Even though GWAS can be used to unbiasedly detect disease-associated variants in admixed populations adjusted for genetic ancestry, it does not leverage the allele frequency differences across the founding populations. In admixed populations, the causal loci may be inherited in higher proportion from the founding population with higher disease prevalence. Admixture mapping (AM) leverages the admixture pattern in the admixed samples to detect abundance of the ancestry-of-origin loci associated with the disease. AM is more powerful for detecting the association when there are differences in disease prevalence and allele frequency across the founding populations.<sup>31</sup> Because of methodological issues, most researchers avoid mixed race or minorities in their analysis. The lack of racial or ethnic diversity in AD genetic research limits the discovery of ancestry-specific variants and potentially exaggerates the disparity in precision diagnosis and medical care for the under-represented groups. In this study, we analyzed 1,725 samples (726 AD-affected individuals and 999 non-AD control individuals) from the AA population genotyped using Illumina's Multi-Ethnic Global Array (MEGA). We use AM, which is a complementary statistical approach to GWAS, to leverage local ancestry estimates in the admixed population. We also use GWAS to identify ancestry-shared variants associated with AD in the AA population. We use a joint GWAS-AM analysis to identify risk variants that otherwise would be missed when studied separately.<sup>32</sup> To identify potential causal variants associated with AD, we perform fine mapping and functional prioritization.

## Materials and methods

### Study design and population

The study cohort consisted of 2,687 participants selected from two local cohorts, the Greater Cincinnati Pediatric Clinical Repository (GCPCR)<sup>33</sup> and the Cincinnati Genomic Control Cohort (GCC),<sup>34</sup> at the Cincinnati Children's Hospital Medical Center (CCHMC). The majority of the samples were from the greater Cincinnati metro area and recruited through the main campus as well as through the satellite locations. Clinical history including AD, skin conditions, asthma, allergy, and demographic information was ascertained through hospital visits, emergency visits, and inpatient visits. Cases were selected based on the reported medical history of AD. Control individuals consisted of subjects with a history of asthma, food allergy, and allergic rhinitis but with no history of AD or eczema. Parental informed consent was obtained for all participants under 18 years of age in the study for the purpose of DNA collection and genotyping, and from patients aged 18 and older. The study was approved by the institutional review boards at CCHMC.

### Genotyping

Genotyping was performed using MEGA, which contains SNP sets tailored toward admixed ancestry.<sup>35</sup> MEGA maximizes coverage and captures the genomic architecture of the AA population. Genotypes were called using the Genetrain2 algorithm in Illumina Genome Studio software following the Illumina's top-bottom strand orientation. Genotype data were available for over 1.43 million variants aligned with the genome assembly GRCh37.

### Quality control

Quality control (QC) analysis was performed in two stages: prior to imputation and post imputation. During the first stage, participants with sub-optimal genotype missing rate of  $\geq 10\%$  were removed. Also, duplicates or monozygotic twins were determined using identity by descent (IBD) score  $>0.90$  and only the sample with lowest missing rate was retained. Sex checks were performed and samples with discrepancy between predicted and reported sex were also removed. Out of the 2,687 samples with genotype data available, 2,382 samples passed the filtering criteria. In the next stage, SNPs were filtered for the sub-optimal missing of  $\geq 5\%$ , and significant deviation from Hardy-Weinberg Equilibrium (HWE) in controls at  $p < 1E-5$ . Genetically related samples were determined using IBD statistics from PLINK 2 (<https://www.cog-genomics.org/plink/2.0/>). IBD analysis was conducted using a set of LD-pruned SNPs and IBD score  $\hat{\pi} \geq 0.25$  was used as a cutoff for filtering potential duplicate and cryptic related samples. IBD score  $\hat{\pi} \geq 0.25$  potentially removes first- and second-degree relations.<sup>36</sup> Among the related samples, the sample with the lowest missing rate was selected, and others were removed from the analysis. With the IBD filtering criteria, 105 related samples were identified and further removed from the analysis. Principal-component analysis (PCA) was performed, and the top five PCs were extracted. For both IBD and PCA analyses, variants were pruned for linkage disequilibrium (LD) and only variants with LD  $< 0.2$  were used. All QC analyses were conducted using PLINK 2.<sup>37</sup>

### Genotype imputation

Imputation was carried out across the autosomal chromosomes using the Michigan Imputation Server, which implemented the min-

imac4 algorithm.<sup>38</sup> Strand-aligned genotype data were loaded into the server. Imputation was conducted using the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA) reference panel.<sup>39</sup> The output files with imputation dosage were converted to plink binary files using the "make-bed" option, resulting in the hard-call binary plink files that were subsequently used for the association analysis. All bi-allelic variants with imputation quality threshold of Rsq score  $\geq 0.3$  were reported.

### Local ancestry estimates

The chromosomes of admixed participants consist of a mosaic of chromosomal blocks from the ancestral populations, which are called local ancestry blocks. Since the true ancestries are unknown, ancestry at each locus or block could be inferred computationally based on appropriate reference ancestral populations. Local ancestry for the participants was inferred by modeling AA samples as a two-way admixture between African and European populations that occurred approximately eight generations prior.<sup>40</sup> Inference was carried out using RFMix v2, which uses a supervised conditional random forest method to optimally infer the ancestries of the alleles at a marker locus.<sup>41</sup> The CEU ( $n = 99$ ) and YRI ( $n = 108$ ) panels from the 1000 Genomes Project III were used as the reference populations for the European and African ancestry. The sample genotype data were checked for strand alignment using coform-gt (<https://faculty.washington.edu/browning/conform-gt.html>). After removing the SNPs that did not conform to the alignment of the reference panel, samples were phased using Beagle with the African and European populations as the reference panels.<sup>42</sup> The genetic map files were downloaded from the Beagle site ([http://bochet.gcc.biostat.washington.edu/beagle/genetic\\_maps/](http://bochet.gcc.biostat.washington.edu/beagle/genetic_maps/)). The number of generations since admixture was set to be eight and inference was carried out for each autosomal chromosome under the Expectation Maximization (EM) option with five iterations. Inference was carried out for each autosomal chromosome.

### Global ancestry estimates

Global ancestry, the proportions of genomic contribution from the ancestral populations in the entire genome of an admixed sample, was conducted using RFMix v2 with CEU and YRI as the reference populations for European and African ancestry; the genomic proportions of European and African ancestries were estimated for each autosomal chromosome.<sup>41</sup> Global ancestry proportion for each sample was estimated using the weighted sum of the chromosomal ancestry estimates where weights were proportional to the size of the chromosomes. Samples with global African-ancestry proportion  $<0.02$  were removed from further analysis.

### AM

To identify the association between AD and African ancestry, AM was performed using case-only and case-control analyses.<sup>43</sup> Estimation of local and global ancestry were carried out as described before. In the case-only analysis, each sample's local ancestry at a marker locus was compared to the respective global ancestry. In case-control analysis, deviations in local ancestries between the AD and control individuals were tested. The case-only and case-control AM approaches are described below. Superscript "c" and "d" are used to represent cases and controls, respectively, unless otherwise defined.

Let  $x_{ij}^c$  and  $x_{ij}^d$  be the proportion of African ancestry at a marker locus  $i$  and  $q_i^c$  and  $q_i^d$  be the global ancestry of  $i$ -th case and  $j$ -th control, respectively. Let  $n_1$  and  $n_2$  be the number of cases and controls.

Case only: define  $\bar{x}_i^c = \frac{1}{n_1} \sum_j x_{ij}^c$  as average local ancestry at marker locus  $i$  and  $\bar{q}^c = \frac{1}{n_1} \sum_j q_i^c$  as average global ancestry among all cases.

The test statistics for case only is defined as  $T_1 = \frac{\bar{x}_i^c - \bar{q}^c}{se_1}$ , where  $se_1 = \sqrt{\frac{\text{Var}(x_{ij}^c - q_i^c)}{n_1}}$  is the standard error.

Case-control: define  $\bar{x}_i^d = \frac{1}{n_2} \sum_j x_{ij}^d$  as average local ancestry at marker locus  $i$  and  $\bar{q}^d = \frac{1}{n_2} \sum_j q_i^d$  as average global ancestry among all control individuals. The test statistics for case-control analysis is defined as  $T_2 = \frac{(\bar{x}_i^c - \bar{q}^c) - (\bar{x}_i^d - \bar{q}^d)}{se_2}$ , where  $se_2 = \sqrt{\frac{\text{Var}(x_{ij}^c - q_i^c)}{n_1} + \frac{\text{Var}(x_{ij}^d - q_i^d)}{n_2}}$  is the standard error.

For large  $n_1, n_2$ , both  $T_1$  and  $T_2$  are approximated with the standard normal distribution.

### Post hoc power analysis

The statistical power of the study was calculated using the Power Analysis in Multi-ancestry Admixture Mapping (PAMAM) web tool.<sup>44</sup> First, we estimated the testing burden using the R package STEAM.<sup>45</sup> We followed the analytical approximation with # of generations since admixture,  $g = 8$ , and the map file thinned to have a variant selected for each 0.2 cM chunk. We selected these parameters to match the default setting used for the local ancestry estimation in the RFMix V2. The analysis resulted in a genome-wide significant  $p$  value threshold of  $1.47E-5$ . Next, we used the PAMAM for the power estimation, with sample size of 726 cases and 999 controls, a significance threshold of  $p < 1.47E-5$ , and the known African ancestry proportion of 0.78. Accordingly, this study achieved >80% power to detect an ancestral OR >1.7.

### Genomic inflation factor adjustment

Both AM and GWAS results were evaluated for the potential genomic inflation factors. Genomic control (GC) test statistics was used to correct the genomic inflation.<sup>46</sup> Briefly, the test statistics were adjusted using the GC factor  $\lambda$  estimated as the ratio of median of the  $\chi^2$  test statistic to the median of  $\chi^2$  with 1 degree of freedom. Finally, the adjusted test statistics were used to compute the inflation-adjusted  $p$  value.

### Fine-mapping analysis and SNP prioritization

A functional-mapping study of the significant AM region was conducted. All the SNP variants mapped to the significant AM regions with minor-allele frequency (MAF) >0.01 and imputation quality score (Rsq) >0.3 were accessed from the genome-wide imputed data. SNPs were tested for the association with AD using logistic regression test adjusted for age, sex, PCs, and the global ancestry. For each AM locus, the SNP with strongest association was reported. For each AM locus, the significance threshold used was  $p < 0.05/n$ , where  $n$  was determined after LD-pruning with LD cut-off of  $r^2 = 0.3$ .

To prioritize potential functional variants in the AM loci, SNPs were annotated using multi-omics functional annotation tools. First, the allele frequency difference  $\Delta = |f_{AFR} - f_{EUR}|$ , with allele frequency of the African ( $f_{AFR}$ ) and European ( $f_{EUR}$ ) were accessed from the 1000 Genomes project. Functional annotation of the variants with  $\Delta \geq 0.25$  was performed with CADD<sup>47</sup> and Regulo-

meDB.<sup>48</sup> The Variant Effect Predictor (VEP) from Ensembl (<http://grch37.ensembl.org/info/docs/tools/vep/index.html>) was used for the variant annotation including the CADD score and allele frequencies, whereas RegulomeDB 2.0.3 (<https://regulomedb.org/regulome-search>) was used for the RegulomeDB score. Overlap of variants with CADD score  $\geq 14$  and RegulomeDB score  $\leq 3$  are considered as the top prioritized variants associated to AD.

### Genome-wide association analysis

Genome-wide association analysis was performed under the logistic regression framework, between the binary AD phenotype (Y) and the genotype (X) adjusted for the covariates (W), which includes, sex, age, and top five PCs as  $\text{logit}(Y = 1) = \beta_0 + \beta_1 X + \beta_2 W + \epsilon$ , where  $\beta$ 's are the regression coefficients, and  $\epsilon$  is the normally distributed error terms. GWAS was performed using PLINK 2.<sup>37</sup> By default, PLINK 2 performs logistic regression analysis with "firth-fallout" option, which allows firth regression if the logistic regression fails to converge. Variants were filtered for MAF < 0.01, missing rate >0.05, and HWE <  $1E-5$  in controls. GWAS significance threshold of  $5E-8$  was used to assess the significant association of genotype with AD.

### Replication of known AD GWAS signals

Variants associated with AD in prior GWASs were accessed from the NHGRI-EBI GWAS Catalog.<sup>49</sup> With  $p < 1E-5$ , 258 variants associated with AD were identified. Using LDlinkR,<sup>50</sup> LD variants within  $\pm 100$  kb distance with  $r^2 \geq 0.3$  in the AFR population from the 1000 Genomes Project III were extracted and overlapped with the GWAS results. For each catalog SNP, the GWAS SNP with strongest  $p$  value was identified. Catalog SNPs were pruned when multiple variants were tagged with the same GWAS SNP. The pruning resulted in 173 AD-associated loci with at least one GWAS SNP within  $\pm 100$  kb and  $r^2 \geq 0.3$ . Replication significance was assessed at  $p < 0.05/173 = 2.9E-4$ .

### Joint association and admixture analysis

Association-admixture joint analysis has potential to capture the complementary signals from the genotype-based association mapping and ancestry-based admixture mapping and hence detect signals that are only marginally significant in individual analysis alone. We performed the association-ancestry joint analysis on the childhood AA samples using the Bayesian mixed SNP and admixture (BMIX) testing approach.<sup>32</sup> The BMIX approach combines genotype and ancestry information. BMIX uses the posterior probabilities from the AM as a prior for the association and computes the posterior probability (PP) for the genotype test. The test relies on the user-specified information on the multiple testing burden of admixture and association mapping. Following the analytical approaches of BMIX, we estimated  $n = 1240.5$  for the AM. For the association mapping, we used the  $n = 345450.3$  as the number of testing burden as suggested in the BMIX. The BMIX analysis was conducted for the genotype markers after the QC filtering- MAF < 0.05, HWE  $p \leq 1E-5$  in control and  $\leq 1E-8$  in case, and genotype missing rate of  $\geq 0.05$ . SNPs with posterior probability of joint BMIX test greater than 0.5 are considered significant.

### Gene expression analysis

Validation of the risk loci were performed using the publicly available gene expression data from Gene expression Omnibus (GEO: GSE32924).<sup>51</sup> Data consisted of gene expression profiles from

**Table 1. Demographic information**

Characteristics	Total ( <i>n</i> = 1725)	Cases ( <i>n</i> = 726)	Control ( <i>n</i> = 999)	<i>p</i> value
Female, count (%)	757 (43.88%)	338 (46.56%)	419 (41.94%)	0.063
Mean age in years ( $\pm$ SD)	9.08 (5.77)	8.01 (4.84)	9.85 (6.25)	<0.001
Mean African ancestry ( $\pm$ SD)	0.757 (0.16)	0.761 (0.16)	0.753 (0.17)	0.328
Comorbidities:				
Asthma	1216(70.5%)	516 (71.07%)	700 (70.07%)	0.69
AR	863 (50%)	398 (54.78%)	465 (45.81%)	<0.01
FA	181 (10.5%)	109 (15.35%)	72 (7.09%)	<0.001

AR, allergic rhinitis; FA, food allergy.

skin biopsy samples of lesional and non-lesional skins from 12 AD-affected individuals and eight control individuals. The expression was accessed using the Affymetrix human U133Aplus2 arrays (Affymetrix, Santa Clara, CA). The AD samples consisted of moderate-to-severe patients aged 16–81 years. Details on sample preparation and QCs are described elsewhere.<sup>51</sup> Differential gene expression analyses comparing lesional vs. control and non-lesional vs. controls were performed with NetworkAnalyst 3.0 using the limma algorithm.<sup>52</sup> Differential gene expression results for the targeted prioritized genes were assessed for significance at false discovery rate (FDR)-adjusted *p* (*q* value) <0.05 and an absolute fold change  $\geq 1.5$ .

### qPCR analysis

Validation of the GWAS genes was conducted using real-time qPCR technique. Target genes for the validation were identified using GWAS and the joint ancestry-genotype analysis. Total blood collected from the Mechanisms of Progression of Atopic Dermatitis to Asthma in Children (MPAACH) participants that were diagnosed with AD (*n* = 6) and from control individuals (*n* = 3) was used to extract RNA. The MPAACH study was approved by the CCHMC Institutional Review Board (IRB) under protocol number 2016-5842, informed consent was provided by all subjects, and 800  $\mu$ L of total blood was used for RNA extraction using the RiboPure RNA Blood Extraction Protocol (catalog no. AM1928, Thermo Fisher Scientific, Waltham, MA).

Extracted RNA was checked for quality and quantity using Thermo Scientific Nanodrop OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific) and Agilent TapeStation 4150 System (Agilent Technologies, Santa Clara, CA), respectively. RNA integrity number (RIN) values for the RNA samples were 8.00–9.00. A higher RIN value indicates a higher degree of integrity. Then 250 ng of total RNA was reverse transcribed using SuperScript IV VILO Master Mix (catalog no. 11756500, Thermo Fisher Scientific). For qPCR analysis, the resulting cDNA was run on the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific) using TaqMan gene expression assays and TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific) per manufacturer's instructions. Comparative critical cycle (Ct) values for analysis were obtained from Applied Biosystems Analysis Software and processed in the Thermo Fisher Connect Software (Thermo Fisher Scientific). The fold change expression levels of the target genes were normalized to the 18s and the relative abundance was calculated by the  $2^{-\Delta\Delta CT}$  method.<sup>53</sup>

Experimental values are expressed as means  $\pm$  SEM of three independent RNA samples derived from healthy control individual

and six from AD patients. GraphPad Prism version 9.5 (Boston, MA) was used to determine the statistical differences between healthy control individuals and AD patients. Data analysis was performed using non-parametric Mann-Whitney tests. A *p* value < 0.05 was considered statistically significant when comparing the gene expression between healthy control individuals with AD patients.

### Functional pathway and network analysis

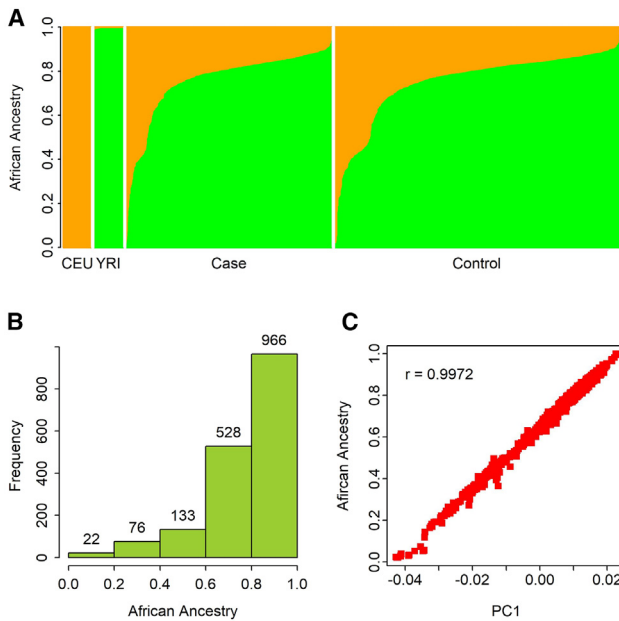
To understand the biological role of the AR-associated genes, over-representation enrichment analysis for functional pathways was conducted using the ConsensusPathDB (<http://cpdb.molgen.mpg.de/>).<sup>54</sup> AD-associated genes identified from the AM, GWAS, and joint analysis were used for the analysis. Significant pathways and gene ontology terms were assessed at FDR-adjusted *p* < 0.05.

## Results

### Demographic characteristics

In total 1,725 samples (726 AD-affected individuals and 999 non-AD control individuals) passed all QC criteria and global African-ancestry proportion  $\geq 0.02$  and were selected for the study. AD-affected individuals were significantly younger than the control individuals (average ages: AD-affected individuals, 8.04 years vs. control individuals, 9.81, *p* < 0.001). Females represented 46.46% of cases compared to 41.97% among control individuals, but the difference was not significant. There was no difference in the proportion of asthma status among AD-affected individuals and control individuals; however, allergic rhinitis (AR) and food allergy (FA) were observed in higher proportion among AD-affected individuals than control individuals (Table 1).

The genome of AA individuals consisted of European and African ancestry with varied composition. Figure 1A shows the variation of the global African-ancestry proportion among AD-affected and control individuals in comparison to the reference individuals of CEU and YRI panels from the 1000 Genomes project. The overall distribution of the global African ancestry was skewed left with over 86% of the samples (1,494/1,725 = 86.6%) having greater than 60% of their genomes derived from the African ancestry (Figure 1B). The average global African ancestry



**Figure 1. Ancestry estimation**

(A) Distribution of global ancestry for two reference populations (CEU and YRI) and samples in case and control cohorts (orange, European ancestry; green, African ancestry).

(B) Histogram showing the distribution of African ancestry in sample.

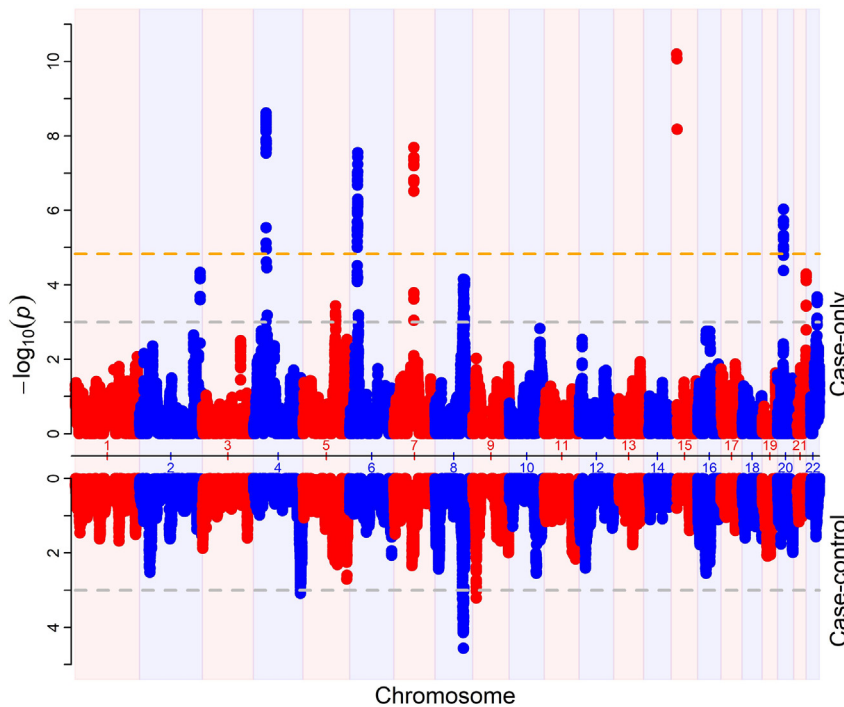
(C) Correlation between African ancestry and PC1.

of the whole cohort was 0.757. The global African ancestry among cases was marginally higher than that for the control individuals, but the difference was not statistically significant ( $p = 0.328$ ; Table 1). The first PC (PC1) explained

most of the global African-ancestry variation with correlation coefficient ( $r$ ) = 0.997 (Figure 1C).

**AM**

To identify the ancestry-specific loci associated with AD in AA, we performed AM using case-only and case-control approaches. The Miami plot showed the results from both analysis (Figure S1). The quantile-quantile (QQ) plot showed both AM results were inflated with genomic inflation factors of 1.22 and 1.56 for case-control and case-only analysis, respectively (Figure S2). Both AM results were adjusted for the genomic inflation factor as described in the “materials and methods” section. Prior to the adjustment of genomic inflations, case-only analysis identified 10 significant loci, whereas case-control analysis identified one significant locus with  $p < 1.47E-5$  (Figure S1). After the adjustment, five of the 10 AM loci remained significant (Figure 2; Table 2A). The top AM peak was observed at chromosome 15q11.2 (case-only adjusted  $p = 6.24E-10$ ). The locus is a gene-rich region including multiple genes from the Immunoglobulin Heavy Diversity (IGHD) family and POTE Ankyrin Domain family. All five significant loci were enriched for European ancestry as suggested by the negative Z score. AM locus at chromosome 7q11.23 was also marginally significant at  $p < 0.05$  in case-control analysis. No locus reached the significant  $p$  value threshold in case-control analysis, with the strongest association observed at locus 8q23.1-2 (case-control adjusted  $p = 2.75E-5$ ; Table 2A). The locus 8q23.1-2 spanned about a 1.3-Mb region and was enriched for the European ancestry. Two additional loci, 4q34.3 and 9q23, reached suggestive



**Figure 2. Miami plot of AM result**

The upper section shows results from case-only analysis. The lower section shows results from case-control analysis. The x axis represents chromosome position. The y axis represents  $p$  value. Orange horizontal line represent genome-wide significance at  $p < 1.47E-5$  with adjustment for genomic inflation. Gray dashed horizontal lines represent suggestive association at  $p$ -value  $< 1E-3$ .

**Table 2. AM loci associated with AD in AA samples**

Locus	START	END	Z <sub>cc</sub>	P <sub>cc</sub>	Z <sub>c</sub>	P <sub>c</sub>	P <sub>c_adj</sub>	P <sub>cc_adj</sub>
<b>A. Case-only AM results</b>								
4p12-q12	48226267	49003382	-0.88	0.38	-7.44	9.77E-14	2.40E-09	0.43
6p22.1-21.33	29696849	31319586	-0.06	0.95	-6.93	4.33E-12	2.81E-08	0.96
7q11.23	76910730	77044434	-2.73	6.32E-3	-6.99	2.66E-12	2.05E-08	0.014
15q11.2	20083584	22560396	1.54	0.123	-8.154	3.51E-16	6.24E-11	0.163
20p11.21	22487052	22672462	-0.138	0.89	-6.119	9.42E-10	9.31E-07	0.91
<b>B. Case-control AM results</b>								
4q34.3	181489069	181576044	3.698	2.17E-04	2.64	8.36E-03	3.45E-02	8.24E-04
8q23.1-2	109871639	111194968	-4.636	3.55E-06	-4.958	7.12E-07	7.04E-05	2.75E-05
9p23	12925921	12942772	-3.784	1.54E-04	-3.24	1.19E-03	9.37E-04	6.21E-03

A. Significant AM loci in case-only analysis with the inflation-adjusted case-only  $p$  value ( $P_{c\_adj}$ )  $< 1.47E-5$ , the genome-wide significance threshold for AM. B. Results from case-control analysis with inflation-adjusted  $p$  value  $< 0.001$ . START, starting base pair position; END, ending base pair position; Z<sub>cc</sub>, case-control test statistics; Z<sub>c</sub>, case-only test statistics; P<sub>cc</sub>, case-control  $p$  value; P<sub>cc\_adj</sub>, inflation-adjusted case-control  $p$  value; P<sub>c</sub>, case-only  $p$  value; P<sub>c\_adj</sub>, inflation-adjusted case-only  $p$  value.

significance level of adjusted  $p < 0.001$  in case-control analysis (Table 2B). The locus 4q34.3 was enriched for African ancestry. All three loci reached marginal association of adjusted  $p < 0.05$  in case-only analysis. AM analyses identified potential AD loci with European ancestry effect. However, we found high inflation factor and further research would be required to validate the findings.

#### Fine mapping of AM loci

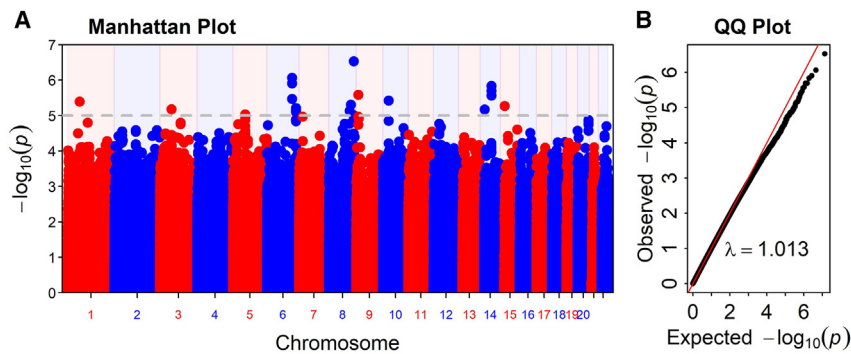
SNP variants mapped to the AM loci were tested for the allelic association with AD. All SNPs that passed the

imputation  $R_{sq} \geq 0.3$  and QC criteria ( $MAF \geq 0.01$ ,  $HWE \geq 1E-5$  in control, missing rate  $< 0.05$ ) were tested for allelic association using the logistic regression test, adjusted for age, sex, and top five PCs. Table 3A shows the strongest association within each AM locus. Multiple testing adjustment was performed for each locus independently using the number of LD-pruned SNPs tested in each locus. No significant association was observed at the five significant AM loci, but SNPs with marginal association at  $p < 0.01$  were identified at each AM locus.

**Table 3. Fine mapping and variant prioritization of AM loci associated with AD in AA samples**

<b>A. Fine mapping of AM loci</b>								
SNP	CHR	POS	OR	P <sub>assoc</sub>	P <sub>thrs</sub>	P <sub>cc_adj</sub>	P <sub>c_adj</sub>	
rs74864956	4	48432376	0.557	3.92E-3	2.35E-5	0.43	2.40E-09	
rs3130569	6	31102955	1.268	3.24E-3	3.24E-5	0.96	2.81E-08	
rs148783322	7	76919259	1.474	2.33E-3	4.81E-4	0.014	2.05E-08	
rs563250505	15	22357139	1.723	3.86E-3	8.47E-4	0.163	6.24E-11	
rs79548864	20	22654011	1.463	7.56E-3	4.46E-4	0.91	9.31E-07	
<b>B. Prioritization of AM loci</b>								
SNP	CHR	POS	AFR_AF	EUR_AF	DELTA	CADD	RDB	GENE
rs375259	4	48343550	0.056	0.3439	0.2879	16.65	2b	SLAIN2
rs3807033	6	30043955	0.4425	0.0825	0.36	18.8	1f	RNF39
rs77580993	6	30059592	0.5068	0.1889	0.3179	17.42	2b	-
rs7765810	6	30063496	0.5091	0.1779	0.3312	17.96	1a	-
rs2277764	20	22566124	0.351	0.0606	0.2904	17.77	2b	FOXA2

A. Fine mapping association test. CHR, chromosome; POS, base-pair position; OR, odds ratio; P<sub>assoc</sub>, association  $p$  value; P<sub>thrs</sub>,  $p$  value threshold for significant association after multiple testing adjustment; P<sub>cc\_adj</sub>, inflation-adjusted case-control  $p$  value; P<sub>c\_adj</sub>, inflation-adjusted case-only  $p$  value. B. Prioritized variants in AM loci. AFR\_AF and EUR\_AF, Minor allele frequency in African and European populations, respectively; RDB, RegulomeDB score; CADD, CADD score; DELTA, absolute difference in AFR\_AF and EUR\_AF.



**Figure 3. AD GWAS results**

(A) Manhattan plot of AD GWAS. Gray horizontal lines represent significance signals at  $p < 1E-5$ .

(B) QQ plot.  $\lambda = 1.013$  is the genomic inflation factor.

### Functional prioritization of variants in AM loci

To identify the putative genetic variants underlying the AM regions, we examined all variants mapped to each significant AM region using the 1000 Genomes Project. Variants were annotated and prioritized using CADD score, allele frequency difference ( $\Delta$ ), and RegulomeDB score (as detailed in the “materials and methods” section). The approach identified five variants with the prioritization criteria  $CADD \geq 14$ ,  $\Delta \geq 0.25$ , and  $RegulomeDB \leq 3$  (Table 3B). On locus 6p22.1-21.33, three variants were identified in the intergenic region upstream of gene *RNF39* (MIM: 607524). Similarly, a 5' UTR variant in gene *SLAIN2* (MIM: 610492) and a regulatory variant in a promoter region near gene *FOXA2* (MIM: 600288) were prioritized at chromosome 4 and chromosome 20, respectively. The multi-omics functional prioritization approach identified 3 AM loci associated with AD and highlights loci encoding the genes *SLAIN2*, *RNF39*, and *FOXA2* as candidate genes associated with AD in AA.

### Genome-wide association analysis

GWAS was performed on 6 million variants with  $MAF > 0.01$  using logistic regression, adjusted for age, sex, and PC1–PC5 (PC1 is highly correlated with global ancestry). No locus reached genome-wide significance at  $p \leq 5E-8$  (Figure 3). At the suggestive significance level of  $1E-6$ , the analysis identified two variants associated with AD in AA (Table 4). The strongest signals were observed at intronic variant rs16904552 ( $p = 2.94E-7$ ,  $OR = 1.744$ , 95% confidence interval [CI] = (1.41, 2.157)) in the gene *EFR3A* (MIM: 611798) on chromosome 8. The second signal was observed at intronic variant rs113357522 ( $p = 8.61E-7$ ,  $OR = 2.717$ , 95% CI = (1.825, 4.047)) in the gene *SGK1* (MIM: 602958) on chromosome 6. GWAS results showed 11 additional loci with suggestive evidence of association at the  $p < 1E-5$  (Table 4).

### Joint analysis of ancestry and genotype

A joint ancestry-genotype analysis was performed using the BMIX approach. The joint analysis accounts for association signals condition upon the local ancestry at each marker. The posterior probability at each genotyped marker is shown in Figure S3. The analysis identified two

novel risk variants for AD susceptibility in AA (Figure 4; Table 5). The first significant association was observed at the intron variant rs2195989 in gene *ANGPT1* (MIM: 601667) on chromosome 8 with the posterior probability (PP) = 0.932. The GWAS analysis showed the locus has moderate evidence of association with AD ( $p = 5.71E-4$ ,  $OR = 1.314$ ) and AM analysis showed the region is also moderately associated with AD with European ancestry as a risk factor ( $P_{cc} = 5.76E-5$ ) (Figure 4). Gene *ANGPT1* encodes angiopoietin-1, which plays roles in angiogenesis, inflammation, and tumor growth.<sup>55</sup> Using the GTEx portal, the variant is found to be expression quantitative trait loci (eQTL) with the target gene *ANGPT1* in muscle-skeletal tissue. The second joint association was found at rs62538818 (PP = 0.614) in the intergenic region between genes *LURAP1L* (MIM: 616130) and *MPDZ* (MIM: 603785). The locus overlaps with a promoter flanking region that is active in the dermal fibroblast, lung, and multiple other cell/tissue types. Dermal fibroblasts are part of skin connective tissue and play an essential role during cutaneous wound healing.<sup>56</sup> Both GWAS and AM analyses showed the variant was moderately associated with AD (GWAS,  $p = 1.22E-5$ ,  $OR = 1.774$ ; AM case-control,  $p = 2.58E-4$ ). Local ancestry distribution at both loci showed the higher African-ancestry proportion among control individuals than among affected individuals (Figure 4).

### Replication of known AD GWAS signals

Variants associated with AD in prior GWASs were accessed from the NHGRI-EBI GWAS Catalog. With  $p < 1E-5$ , 258 variants were identified to be associated with AD. Variants were pruned when multiple variants were identified within  $\pm 100$  kb distance and the SNP with strongest  $p$  value was assigned as the tag SNP for the locus. The pruning resulted in 173 AD-associated GWAS loci. Replication significance was assessed at  $p < 0.05/173 = 2.90E-4$ . Four variants were replicated in AM analysis (Figure 5). Three variants (rs200291258, rs28752924, and rs9368677) are in the locus 6p21.33 ( $P_{cc} = 5.69E-11$ ). The variants are in pairwise moderate correlation in YRI population based on the 1000 Genomes Project III. The fourth variant rs2050190 (gene *TSBP1* (MIM: 618151)) is in the locus 6p21.32 ( $P_{cc} = 0.00028$ ). Both AM loci, 6p21.33 and 6p21.32, showed enriched European ancestry proportions. SNP rs28752924 was found to be associated with AD in European populations,<sup>57</sup> while rs200291258, rs9368677, and rs2050190 were found to be associated with AD in East Asian



**Table 4. GWAS results of AD in AA samples**

SNP	CHR	POS	A1	OR	OR CI (95%)	<i>p</i>	GENE
rs111541801	1	63559620	C	1.858	(1.428, 2.418)	4.04E-06	–
rs141216295	3	61773755	T	2.826	(1.798, 4.441)	6.67E-06	<i>PTPRG</i>
rs11950228	5	64680454	C	0.62	(0.502, 0.766)	9.27E-06	<i>ADAMTS6</i>
rs113357522	6	134505244	C	2.717	(1.825, 4.047)	8.61E-07	<i>SGK1</i>
rs681243	6	154427741	C	1.376	(1.198, 1.58)	6.41E-06	<i>OPRM1</i>
rs77076745	8	110683128	T	0.631	(0.516, 0.772)	7.36E-06	<i>SYBU</i>
rs6998067	8	117951521	T	1.422	(1.223, 1.654)	4.92E-06	<i>AARD</i>
rs16904552	8	132942296	G	1.744	(1.41, 2.157)	2.94E-07	<i>EFR3A</i>
rs16925550	9	10227095	A	0.408	(0.281, 0.593)	2.61E-06	<i>PTPRD</i>
rs60796647	10	31099687	G	2.626	(1.744, 3.953)	3.78E-06	–
rs911912	14	23314675	C	1.761	(1.376, 2.253)	6.76E-06	<i>MMP14</i>
rs177731	14	59379946	T	1.906	(1.466, 2.479)	1.45E-06	–
rs59722874	15	22941055	A	0.431	(0.3, 0.619)	5.35E-6	<i>CYFIP1</i>

A1, risk allele; CI, confidence interval .

populations.<sup>58–60</sup> At  $p < 0.05$ , 28 loci showed ancestry effects based on the case-only AM analysis (Figure 5). The strongest GWAS replication was identified at the intronic rs10519584 in gene *RNF150* ( $p = 5.11E-05$ , OR = 1.638). The variant rs10519584 was also found to be associated with AD in a study of 240 AA samples.<sup>61</sup> At  $p = 0.05$ , nominal replication was detected at 41 additional loci (Figure 5).

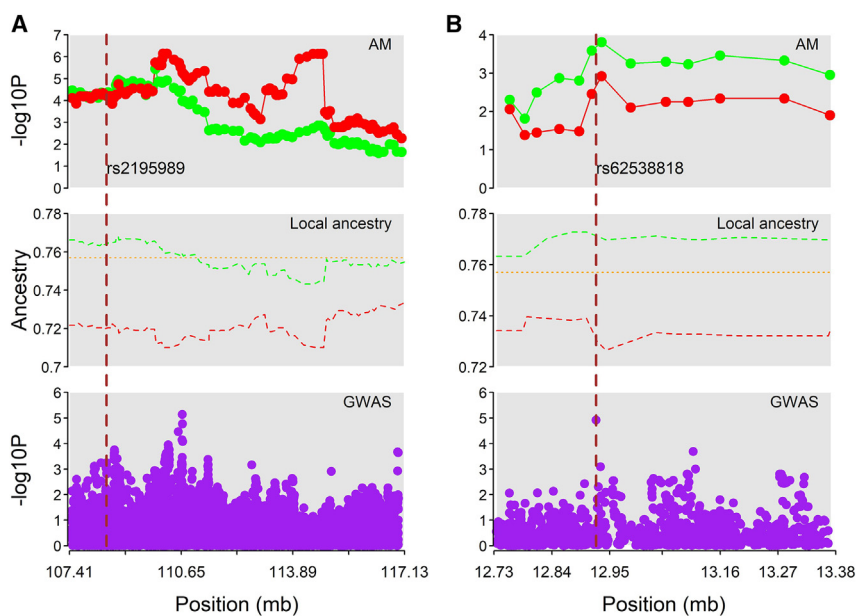
#### Validation of AD target genes using gene expression analysis

Annotations of the variants from the GWAS, AM, and joint analysis using the VEP (<http://grch37.ensembl.org/Tools/VEP>) identified 14 target genes associated with AD

(Table S1). The expression of the genes on 12 samples of lesional AD skins were compared with skin samples from eight control individuals. At FDR < 0.05 and fold change  $\geq 1.5$ , four genes (*SYBU*, *SGK1*, *SLAIN2*, *RNF39*, and *MMP14* [MIM: 600754]) were identified to be significantly differentially expressed in lesional AD skin (Table 6) Four of these (*SGK1*, *SLAIN2*, *RNF39*, and *MMP14*) were also significantly differentially expressed in non-lesional skin against controls.

#### qPCR validation analysis

To further validate the genes associated with AD, qPCR analysis was performed using mRNA expression on



**Figure 4. Regional plot of region with significant BMIX SNPs**

(A) SNP rs2195989 at chromosome 8q23.1. (B) SNP rs62538818 at chromosome 9p23. The upper part shows the results from AM analysis (red dots, case-only AM result; green dot, case-control AM result). The middle section shows local ancestry distribution in cases (red), control (green), and global ancestry (orange). The lower section shows GWAS results. Vertical dashed line marks the BMIX SNP.

**Table 5. Joint analysis of ancestry and genotype effects on AD in AA samples**

CHR	POS	SNP	PP	Z_cc	P_cc	Z_c	P_c	P_gwas	OR	GENE
8	108485912	rs2195989	0.932	-4.026	5.76E-5	-4.02	5.81E-5	5.71E-04	1.314	ANGPT1
9	12924211	rs62538818	0.614	-3.654	2.58E-4	-2.922	3.48E-3	1.22E-05	1.774	-

\_cc, case-control; \_c, case only; P\_gwas, *p*-value from GWAS; PP, posterior probability

control individuals ( $n = 3$ ) and moderately AD-affected individuals ( $n = 6$ ). Table S2 provides the AD severity, sex, and race of the MPAACH participants used for this study. For the analysis, two GWAS genes, *SGK1* and *EFR3A*, and two genes from the joint analysis, *ANGPT1* and *LURAP1L*, were selected. Table S3 provides the TaqMan probes and the assay ID for the selected genes and housekeeping gene used for this study. Analysis showed that all four genes were significantly upregulated in AD patients compared to control individuals at  $p = 0.05$  (Figure 6).

### Functional enrichment analysis

Fourteen genes identified from AM, GWAS, and joint analysis were used in the functional enrichment analysis (Table S1). Over-representation analysis identified two enriched pathways including immune-related pathways PI3K-Akt signaling pathway (Table 7). Enriched Gene Ontology (GO) terms are shown in Table S4. AD-associated genes were enriched for several biological processes related to the nervous system (such as neurogenesis, nervous system development, and generation of neurons), system development, and cell differentiation.

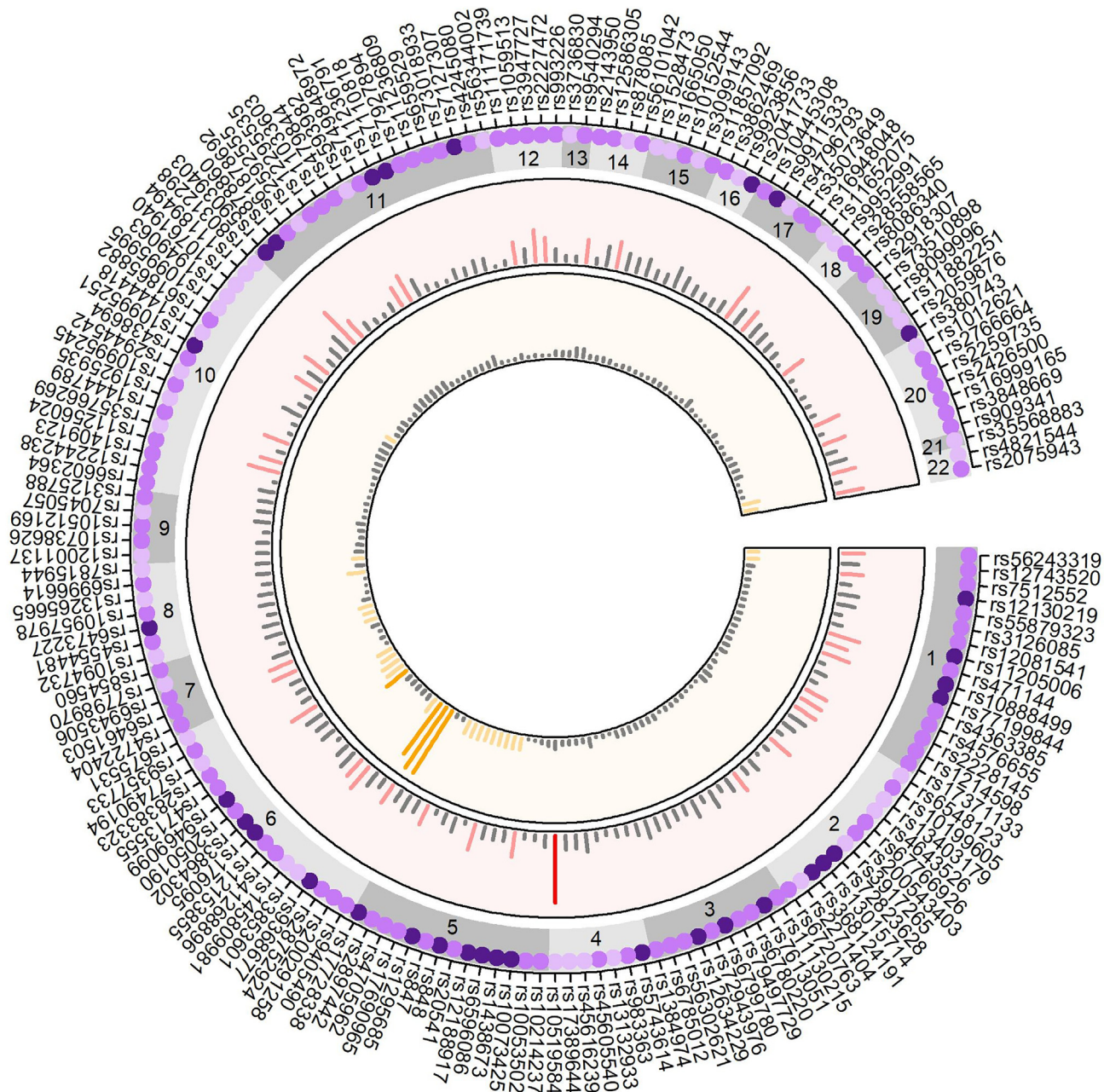
### Discussion

In this study, we performed AM, GWAS, and joint analysis on the AA population using 1,725 samples (726 AD cases and 999 controls). AM, GWAS, and the joint analysis revealed multiple novel risk loci associated with AD in AA population that were not identified in the prior studies with samples of European ancestry. The main finding of the study are the two novel loci associated with AD in AA that are discovered using the ancestry-genotype joint analysis. The two loci are an intronic variant rs2195989 in gene *ANGPT1* on chromosome 8q23.1 and an intergenic regulatory region variant rs62538818 in chromosome 9p23. Additionally, GWAS results identified 13 loci associated with AD in AA at  $p < 1E-5$ . AM identified five loci where local genetic ancestry was associated with AD, all of which showed European ancestry as a disease risk factor. Further validation of the AD loci using a gene expression analysis showed four loci encoding genes *SYBU*, *SGK1*, *RNF39*, *SLAIN2*, and *MMP14* were differentially expressed in lesional skin of individuals with AD. GWAS replicated the locus *RNF150*, which was identified in the study with AA samples.<sup>61</sup> AD loci were enriched for two

pathways, including PI3K-Akt signaling pathway and multiple gene ontology terms.

The joint analysis combines the effects from both genotype and ancestry at SNP level in admixed populations and allows us to detect loci that may not be detected with either approach when performed separately. Both variants exhibited moderate evidence of association with AD when the genotype and the ancestry effects were considered separately using GWAS and AM, respectively. The strongest joint signal was observed at the intronic variant rs2195989 in gene *ANGPT1* on chromosome 8q23.1 with posterior probability 0.764. Both genetic and ancestry effects are moderate at the locus with admixture signal showing European ancestry effect with  $p < 5.76E-05$  and genetic effect with GWAS  $p < 5.71E-04$ , respectively. The gene *ANGPT1* encodes a secreted glycoprotein involved in angiogenesis and tumor growth<sup>55</sup> and hereditary angioedema.<sup>62</sup> The SNP rs2195989 is found to be eQTL with the gene in muscle-skeletal tissue in the GTEx portal (<https://gtexportal.org/home/>). Signals on chromosome 8 have a strong signal buildup in case-only analysis. Similar signal buildup trends were observed for case-control analysis (Figure 2). Such signal buildup in both approaches is less likely to be false positive and could indicate functionality. Further validation using more sample size will be relevant. The next variant, SNP rs62538818 in 9p23, showed moderate admixture effect with  $p = 2.58E-04$  and genetic effect ( $p = 1.22E-06$ ). The SNP overlaps with promoter flanking region ENSR00001447046, which is active in multiple cell/tissue types, including dermal fibroblast, lung, spleen, and stomach. SNP rs2195989 is more common in the African population than the European population (risk allele C, allele frequency: 0.28 [AFR] vs. 0.18 [EUR]) while the SNP rs62538818 is more common in the European population than the African population (risk allele G, allele frequency: 0.02 [AFR] vs. 0.31 [EUR]). Ancestry analysis showed the regions were enriched for European ancestry, so the observed genetic effects at the loci were likely driven by the European-ancestry-specific risk factors.

The AM analysis showed loci with a higher proportion of European ancestry contributed to the AD-susceptibility risk in AA individuals. African-ancestry-specific effect was observed at 6p21.32-21.31, but the locus did not reach significance after the adjustment for genomic inflation. The study also showed that the average proportion of African ancestry across the genome (global ancestry) was not significantly associated with AD ( $p = 0.328$ ), which is consistent with the literature.<sup>27</sup> However, the genome of



Catalog P: ●  $5E-8 < P < 1E-5$  ●  $1E-15 < P < 5E-8$  ●  $P < 1E-15$   
 GWAS P: —  $P > 0.05$  —  $2.9E-4 < P < 0.05$  —  $P < 2.9E-4$   
 AM P: —  $P > 0.05$  —  $2.9E-4 < P < 0.05$  —  $P < 2.9E-4$

**Figure 5. Replication of existing AD-risk variants reported in GWAS Catalog**

The outer track shows the AD-risk SNPs from the GWAS Catalog and the purple dots with varying shade represent the  $p$  value. Second layer showed the GWAS  $p$  value of the SNPs with dark red segments showing the significant replication at  $p < 2.9E-4$  and lighter red segments showing marginal replication at  $p < 0.05$ . The innermost layer shows the AM case-only  $p$  value with dark orange segments representing significant replication.

admixed individuals such as AA consisted of the mosaic of chromosomal segments of ancestry (local ancestry) from parental populations and the test of global ancestry alone would be insufficient to account for the variation in the local ancestry that may harbor the ancestry-specific disease susceptibility risk. This study highlights the importance of

conducting local ancestry analysis using AM approach to discover the ancestry-specific association in the admixed populations.

GWAS results showed that the genomic inflation is minimal ( $\lambda = 1.013$ , Figure 3B), which suggested that the population structure that may have existed in the genotype

**Table 6. Differential gene expression analysis of top genes associated with AD in AA samples**

Gene	FC_L	P_L	Q_L	FC_NL	P_NL	Q_NL
<i>SYBU</i>	2.187	1.90E-04	1.65E-03	1.34	5.41E-03	2.24E-02
<i>SGK1</i>	-1.996	6.55E-05	7.73E-04	-1.52	1.56E-02	4.80E-02
<i>RNF39</i>	1.829	9.32E-03	2.90E-02	2.202	2.15E-03	1.18E-02
<i>SLAIN2</i>	1.677	1.72E-04	1.55E-03	1.706	1.31E-04	1.94E-03
<i>MMP14</i>	1.56	1.75E-02	4.70E-02	1.718	4.18E-03	1.88E-02
<i>EFR3A</i>	1.359	2.32E-03	1.01E-02	1.395	1.28E-03	8.34E-03
<i>ANGPT1</i>	1.303	2.78E-01	3.96E-01	-1.26	1.89E-01	3.05E-01
<i>PTPRG</i>	1.128	5.72E-01	6.78E-01	1.204	3.38E-01	4.71E-01
<i>CYFIP1</i>	1.054	5.16E-01	6.28E-01	1.094	1.62E-01	2.72E-01

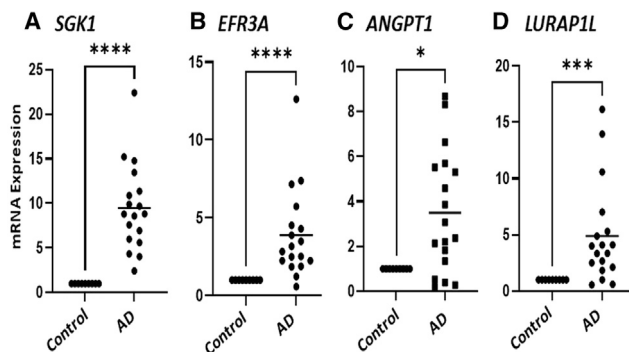
FC, Fold change; P, *p* value; Q, false discovery rate adjusted *p* value; L, lesional skin; NL, non-lesional skin.

data is corrected or controlled during the QC steps and GWAS analysis by using the top five PCs. However, inflations were observed in the AM results with genomic inflation factors of 1.22 and 1.56 for case-control and case-only analysis, respectively (Figure S2). To investigate the effect of covariates and PCs on the inflation of AM result, we performed case-control AM analysis under the logistic regression framework with copies of African-origin allele as a predictor and sex, age, and PCs as covariates. AM results from the analysis did not materialistically change from our initial analysis and the genomic inflation factor remained same (data not shown). This suggests that inclusion of PCs has no effect on the model. We further investigated the extent of the admixture LD among the local ancestry loci using PLINK. Using admixture LD cutoff = 0.8, only 1,308 variants remained and, with cutoff = 0.3, only 241 variants remained. This showed that the local ancestry loci were highly correlated. The observed inflation in the AM analysis could potentially be driven by the high level of local admixture LD. LD analysis on SNP data is very common and reference LD panels for ancestry-matched sample data can be obtained, and methods such as

LDSC<sup>63</sup> have been developed to quantify the LD-induced inflation. However, similar adjustment of AM results would require a suitable reference panel for the admixture LD for the AA population.

We performed fine mapping analysis of the AM region using association test and functional annotation of variants mapped to the regions. The association test allowed us to identify the variants with genetic effects potentially contributing to the ancestry effects. Association test of SNPs showed some evidence of allelic effects of variants in the regions at  $p < 0.01$ . The moderate association may suggest the presence of ancestry-specific allelic effects that the data may have limited power to detect in the association analysis. SNP annotation using multi-omics functional tools such as CADD and RegulomeDB identified variants with strong evidence for potential regulatory functions and substantial difference in the allele frequency differences among the ancestral (African and European) populations. Three prioritized variants in the region 6p22.1-21.33 near gene *RNF39* were found to be eQTL with genes in the region including *RNF39* in whole blood, *HLA-V* in skin, whole blood, and lung. Prioritized variant rs375259 in chromosome 4 near gene *SLAIN2* was found to be eQTL to the gene in multiple tissues including colon, esophagus, and lung.

GWAS identified two loci associated to AD in AA at  $p < 1E-6$ . The strongest association was observed at the variant rs113357522, intronic to gene *SGK1* in chromosome 6q23. The lead variant rs113357522 on gene *SGK1* is a low-frequency African-specific variant, potentially exhibiting African-ancestry effects. The gene is ubiquitously expressed and encodes a membrane protein that affects cell wall composition and plays an important role in cellular stress response.<sup>64</sup> *SGK1* expression is modulated by hormones such as gluco- and mineralocorticoids and has been linked to psychiatric disorders, including major depression.<sup>65</sup> Variants in the gene are found to be associated with psychiatric disorder, protein level measurement, and electrocardiography. The second variant discovered from the GWAS is an intronic variant in the gene *EFR3A*



**Figure 6. Expression levels of select AD-associated genes** (A) *SGK1*, (B) *EFR3A*, (C) *ANGPT1*, and (D) *LURAP1L*. All genes were upregulated in AD patients when compared to healthy control individuals. Experiments were done on RNA isolated from blood using three healthy control individuals and six participants diagnosed with AD. \* $p < 0.05$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

**Table 7. Functional enrichment analysis**

Pathway	Overlapped genes	<i>p</i> value	q value
Transcriptional regulation by MECP2	<i>SGK1</i> ; <i>OPRM1</i>	6.14E−4	4.91E−3
PI3K-Akt signaling pathway	<i>SGK1</i> ; <i>ANGPT1</i>	1.74E−2	4.63E−2
Signaling by receptor tyrosine kinases	<i>CYFIP1</i> ; <i>SGK1</i>	2.90E−2	5.81E−2
Signal transduction	<i>OPRM1</i> ; <i>ANGPT1</i> ; <i>CYFIP1</i> ; <i>SGK1</i>	4.04E−2	6.46E−2

Table shows the pathways enriched for the AD-associated genes.

at chromosome 8. The gene encodes a peripheral membrane protein and regulates localization of phosphatidylinositol 4-kinase (*PI4K*) to the plasma membrane.<sup>66,67</sup> The lead variant at this locus rs16904552 is common in African and East Asian populations (MAF: AFR = 0.17, EAS = 0.17) but rare in the European population (MAF < 0.01). This suggests that the study identified GWAS loci with potentially African-ancestry-specific effects. Additional loci associated with AD in AA were identified at  $p < 1E-5$ , including *MMP14*. An animal model showed lack of *MMP14* in dermal fibroblast was linked to skin homeostasis and wound repair.<sup>68</sup> The lead variant rs911912 is common in the African population (AFR MAF = 0.11) but rare in the European populations. GWAS based on the European/East Asian samples may have failed to detect the loci due to the low frequency or due to African-ancestry-specific effect.

Gene expression profiling identified *SYBU*, *RNF39*, *SLAIN2*, *MMP14*, and *SGK1* genes with significantly differentially expressed in lesional skin of AD compared with non-AD control individuals with adjusted  $p < 0.05$  and fold change  $\geq 1.5$ . Expression of *SYBU*, *SLAIN2*, *RNF39*, and *MMP14* were downregulated in lesional skin of AD-affected individuals. Exploring the GTEx portal for the potential eQTL and sQTL effects, rs375259 showed sQTL effects on the expression of *SLAIN2* in several tissues, including in the skin, and eQTL effects in lung, esophagus, and adipose. Similarly, rs3807033 showed eQTL effects on the expression of *RNF39* in lung and whole blood. *MMP14* gene was also found to be significantly downregulated in lesional skin of AD-affected individuals. *MMP14* is associated to skin homeostasis and repair<sup>68</sup> and also relevant to lipoedema, an estrogen-dependent disorder of adipose tissue.<sup>69</sup> The GWAS gene *SGK1* is significantly upregulated in lesional skin.

The present study has notable strengths and limitations. This is the first AA genetic mapping AD study using the joint analysis of AM and GWAS. The joint analysis combines the ancestry and genetic effects at loci to detect disease-associated risk variants that are not detected when studied separately. To capture the African-ancestry-specific variants, samples were genotyped using MEGA and genotype imputation was carried out using the CAAPA reference panel. The ancestry-specific genotyping and imputation improve the power for disease-gene mapping under both AM and GWAS approaches to identify ancestry-specific AD-risk variants in AA. Multi-omic features including *in silico* regulatory annotations and tran-

scriptomic analysis prioritized the potential functional variants associated with AD and would provide a list of risk variants for further study of pathogenesis of AD. First, this study did not include socioenvironmental or geographic attributes that might influence or confound the ancestry in disease prevalence of AD. Second, sample size is not large enough to detect the variants with smaller effects as our strongest association at rs16904552 did not reach the GWAS cutoff  $p$  value of  $5E-8$ , even with an odds ratio of 1.744. Third, even though the control group consisted of samples with no history of AD, it may have other comorbid atopic disorders such as asthma, FA, and AR that share the risk genes.<sup>19,70</sup> Fourth, the source of the genomic inflation in the AM results was not fully explored in the study, and further research would be required to account for residual inflation that could have arisen due to stratification within ancestral populations and to validate the findings. It should be also noted that the accuracy of local ancestry inference for AM depends on reference panels (e.g., sample size, representativeness) and software (e.g., RFMix v2). Nevertheless, this study presents comprehensive disease-gene mapping in an AA cohort using GWAS, AM, and joint analysis and highlights the novel discovery of ancestry-specific AD-risk variants.

In summary, we have identified multiple target variants and genes associated with AD in the AA population. The joint analysis of AM and GWAS identified novel genetic risk loci of AD, an intronic variant rs2195989 in gene *ANGPT1* and a regulatory variants rs62538818 in chromosome 9, in the AA population. The *ANGPT1* gene is known to be involved in angiogenesis, tumor growth, and hereditary angioedema. AM analysis showed that both African and European ancestry contributed to the genetic risks of AD in AA. AM loci 6p21.32-21.31 was enriched for African ancestry while remaining loci were enriched for the European ancestry. Two GWAS loci, rs113357522 (*SGK1*) and rs16904552 (*EFR3A*), are found to be common in the African population but rare in the European population. AD-susceptibility loci *RNF39*, *SLAIN2*, *MMP14*, and *SGK1* were differentially expressed in lesional skin from AD-affected individuals. Further biological validation on the role of the discovered variants in AD using independent AA samples is needed. The current euro-centric genomics studies only provide incomplete genetic architecture of human diseases including AD and, hence, hinder the translation of genetic research into clinical practices and therapeutic discovery.<sup>71,72</sup> Hence, there is an urgent need to

increase representation of diverse ancestry in genomic research and to conduct ancestry-specific assessment of pathogenic variants. Our study highlights the need for population-specific genomic resources and creation of multi-ancestry cohorts for future studies.

### Data and code availability

Access to the data is available for collaboration with the lead contact with appropriate institutional approval. Data from the study cohort are not publicly available as they contain information that could compromise participant consent and confidentiality.

### Acknowledgments

This work was supported by the National Institutes of Health (NIH) grants R01 HG011411 (to T.B.M.) and U19AI070235 (to G.K.K.H.).

### Declaration of interests

The authors declare no competing interests.

### Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.xhgg.2024.100350>.

### Web resources

PLINK 2, <https://www.cog-genomics.org/plink/2.0/>

RFMix v2, <https://github.com/slowkoni/rfmix>.

Beagle, <https://faculty.washington.edu/browning/beagle/beagle.html>.

Conform-gt, <https://faculty.washington.edu/browning/conform-gt.html>.

Ensembl, <http://grch37.ensembl.org/index.html>.

VEP, <http://grch37.ensembl.org/info/docs/tools/vep/index.html>.

GWAS Catalog, <https://www.ebi.ac.uk/gwas/home>.

LDlinkR, <https://cran.r-project.org/web/packages/LDlinkR/index.html>.

RegulomeDB, <https://www.regulomedb.org/regulome-search/>

GTEEx, <https://gtexportal.org/home/>

ConsensusPathDB, <http://cpdb.molgen.mpg.de/>

Received: October 17, 2023

Accepted: September 3, 2024

### References

- Boothe, W.D., Tarbox, J.A., and Tarbox, M.B. (2024). Atopic Dermatitis: Pathophysiology. *Adv. Exp. Med. Biol.* *1447*, 21–35.
- Drucker, A.M., Wang, A.R., Li, W.Q., Severson, E., Block, J.K., and Qureshi, A.A. (2017). The Burden of Atopic Dermatitis: Summary of a Report for the National Eczema Association. *J. Invest. Dermatol.* *137*, 26–30.
- Gupta, J., Johansson, E., Bernstein, J.A., Chakraborty, R., Khurana Hershey, G.K., Rothenberg, M.E., and Mersha, T.B. (2016). Resolving the etiology of atopic disorders by using genetic analysis of racial ancestry. *J. Allergy Clin. Immunol.* *138*, 676–699.
- Odhiambo, J.A., Williams, H.C., Clayton, T.O., Robertson, C.F., Asher, M.I.; and ISAAC Phase Three Study Group (2009). Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three. *J. Allergy Clin. Immunol.* *124*, 1251–1258.e23.
- Wadonda-Kabondo, N., Sterne, J.A.C., Golding, J., Kennedy, C.T.C., Archer, C.B., Dunnill, M.G.S.; and ALSPAC Study Team (2004). Association of parental eczema, hayfever, and asthma with atopic dermatitis in infancy: birth cohort study. *Arch. Dis. Child.* *89*, 917–921.
- Larsen, F.S., Holm, N.V., and Henningsen, K. (1986). Atopic dermatitis. A genetic-epidemiologic study in a population-based twin sample. *J. Am. Acad. Dermatol.* *15*, 487–494.
- Schultz Larsen, F. (1993). Atopic dermatitis: a genetic-epidemiologic study in a population-based twin sample. *J. Am. Acad. Dermatol.* *28*, 719–723.
- Brown, S.J., Elias, M.S., and Bradley, M. (2020). Genetics in Atopic Dermatitis: Historical Perspective and Future Prospects. *Acta Derm. Venereol.* *100*, adv00163.
- Elmose, C., and Thomsen, S.F. (2015). Twin Studies of Atopic Dermatitis: Interpretations and Applications in the Filaggrin Era. *J. Allergy* *2015*, 902359.
- Palmer, C.N.A., Irvine, A.D., Terron-Kwiatkowski, A., Zhao, Y., Liao, H., Lee, S.P., Goudie, D.R., Sandilands, A., Campbell, L.E., Smith, F.J.D., 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.
- Kim, B.E., and Leung, D.Y.M. (2018). Significance of Skin Barrier Dysfunction in Atopic Dermatitis. *Allergy Asthma Immunol. Res.* *10*, 207–215.
- Debinska, A., Danielewicz, H., Boznański, A., Matusiak, Ł., and Szepietowski, J.C. (2022). A small proline-rich protein (SPRR) gene variant contributes to atopic eczema and eczema-associated asthma susceptibility. *Postepy Dermatol Alergol* *39*, 965–971.
- Almoguera, B., Vazquez, L., Mentch, F., March, M.E., Connolly, J.J., Peissig, P.L., Linneman, J.G., Plaza-Serón, M.D.C., Pino-Yanes, M., Burchard, E.G., et al. (2019). Novel locus for atopic dermatitis in African Americans and replication in European Americans. *J. Allergy Clin. Immunol.* *143*, 1229–1231.
- Gao, P.S., Rafaels, N.M., Mu, D., Hand, T., Murray, T., Boguniewicz, M., Hata, T., Schneider, L., Hanifin, J.M., Gallo, R.L., et al. (2010). Genetic variants in thymic stromal lymphopoietin are associated with atopic dermatitis and eczema herpeticum. *J. Allergy Clin. Immunol.* *125*, 1403–1407.e4.
- Ellinghaus, D., Baurecht, H., Esparza-Gordillo, J., Rodríguez, E., Matanovic, A., Marenholz, I., Hübner, N., Schaarschmidt, H., Novak, N., Michel, S., et al. (2013). High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat. Genet.* *45*, 808–812.
- Grosche, S., Marenholz, I., Esparza-Gordillo, J., Arnau-Soler, A., Pairo-Castineira, E., Rüschenhoff, F., Ahluwalia, T.S., Almquist, C., Arnold, A., Australian Asthma Genetics Consortium AAGC, et al. (2021). Rare variant analysis in eczema identifies exonic variants in DUSP1, NOTCH4 and SLC9A4. *Nat. Commun.* *12*, 6618.
- Margaritte-Jeannin, P., Budu-Aggrey, A., Ege, M., Madore, A.M., Linhard, C., Mohamdi, H., von Mutius, E., Granell, R., Demenais, F., Laprise, C., et al. (2022). Identification of OCA2 as a novel locus for the co-morbidity of asthma-plus-eczema. *Clin. Exp. Allergy* *52*, 70–81.

18. Johansson, Å., Rask-Andersen, M., Karlsson, T., and Ek, W.E. (2019). Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum. Mol. Genet.* *28*, 4022–4041.
19. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemer, V., van Dongen, J., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* *49*, 1752–1757.
20. Mucha, S., Baurecht, H., Novak, N., Rodríguez, E., Bej, S., Mayr, G., Emmert, H., Stözl, D., Gerdes, S., Jung, E.S., et al. (2020). Protein-coding variants contribute to the risk of atopic dermatitis and skin-specific gene expression. *J. Allergy Clin. Immunol.* *145*, 1208–1218.
21. Brown, S.J. (2021). What Have We Learned from GWAS for Atopic Dermatitis? *J. Invest. Dermatol.* *141*, 19–22.
22. Shaw, T.E., Currie, G.P., Koudelka, C.W., and Simpson, E.L. (2011). Eczema prevalence in the United States: data from the 2003 National Survey of Children's Health. *J. Invest. Dermatol.* *131*, 67–73.
23. Williams, H.C., Pembroke, A.C., Forsdyke, H., Boodoo, G., Hay, R.J., and Burney, P.G. (1995). London-born black Caribbean children are at increased risk of atopic dermatitis. *J. Am. Acad. Dermatol.* *32*, 212–217.
24. Child, F.J., Fuller, L.C., Higgins, E.M., and Du Vivier, A.W. (1999). A study of the spectrum of skin disease occurring in a black population in south-east London. *Br. J. Dermatol.* *141*, 512–517.
25. Ait-Khaled, N., Odhiambo, J., Pearce, N., Adjoh, K.S., Mae-sano, I.A., Benhabyles, B., Bouhayad, Z., Bahati, E., Camara, L., Cateau, C., et al. (2007). Prevalence of symptoms of asthma, rhinitis and eczema in 13- to 14-year-old children in Africa: the International Study of Asthma and Allergies in Childhood Phase III. *Allergy* *62*, 247–258.
26. Janumpally, S.R., Feldman, S.R., Gupta, A.K., and Fleischer, A.B., Jr. (2002). In the United States, blacks and Asian/Pacific Islanders are more likely than whites to seek medical care for atopic dermatitis. *Arch. Dermatol.* *138*, 634–637.
27. Abuabara, K., You, Y., Margolis, D.J., Hoffmann, T.J., Risch, N., and Jorgenson, E. (2020). Genetic ancestry does not explain increased atopic dermatitis susceptibility or worse disease control among African American subjects in 2 large US cohorts. *J. Allergy Clin. Immunol.* *145*, 192–198.e11.
28. Sanyal, R.D., Pavel, A.B., Glickman, J., Chan, T.C., Zheng, X., Zhang, N., Cueto, I., Peng, X., Estrada, Y., Fuentes-Duculan, J., et al. (2019). Atopic dermatitis in African American patients is T(H)2/T(H)22-skewed with T(H)1/T(H)17 attenuation. *Ann. Allergy Asthma Immunol.* *122*, 99–110.e6.
29. Brunner, P.M., and Guttman-Yassky, E. (2019). Racial differences in atopic dermatitis. *Ann. Allergy Asthma Immunol.* *122*, 449–455.
30. Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., Curtin, J.A., Bønnelykke, K., Tian, C., Takahashi, A., et al. (2015). Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* *47*, 1449–1456.
31. Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K.E., Hafler, D.A., Oksenberg, J.R., Hauser, S.L., Smith, M.W., O'Brien, S.J., Altshuler, D., et al. (2004). Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* *74*, 979–1000.
32. Shriner, D., Adeyemo, A., and Rotimi, C.N. (2011). Joint ancestry and association testing in admixed individuals. *PLoS Comput. Biol.* *7*, e1002325.
33. Butsch Kovacic, M., Biagini Myers, J.M., Lindsey, M., Patterson, T., Sauter, S., Ericksen, M.B., Ryan, P., Assa'ad, A., Lierl, M., Fischer, T., et al. (2012). The Greater Cincinnati Pediatric Clinic Repository: A Novel Framework for Childhood Asthma and Allergy Research. *Pediatr. Allergy Immunol. Pulmonol.* *25*, 104–113.
34. Martin, L.J., Murrison, L.B., and Butsch Kovacic, M. (2020). Building a Population Representative Pediatric Biobank: Lessons Learned From the Greater Cincinnati Childhood Cohort. *Front. Public Health* *8*, 535116.
35. DeWan, A.T., Egan, K.B., Hellenbrand, K., Sorrentino, K., Pizzoferrato, N., Walsh, K.M., and Bracken, M.B. (2012). Whole-exome sequencing of a pedigree segregating asthma. *BMC Med. Genet.* *13*, 95.
36. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nat. Protoc.* *5*, 1564–1573.
37. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* *4*, 7.
38. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* *48*, 1284–1287.
39. Mathias, R.A., Taub, M.A., Gignoux, C.R., Fu, W., Musharoff, S., O'Connor, T.D., Vergara, C., Torgerson, D.G., Pino-Yanes, M., Shringarpure, S.S., et al. (2016). A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat. Commun.* *7*, 12522.
40. Chen, G., Shriner, D., Zhou, J., Doumatey, A., Huang, H., Gerry, N.P., Herbert, A., Christman, M.F., Chen, Y., Dunston, G.M., et al. (2010). Development of admixture mapping panels for African Americans from commercial high-density SNP arrays. *BMC Genom.* *11*, 417.
41. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* *93*, 278–288.
42. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* *108*, 1880–1890.
43. Gautam, Y., Altaye, M., Xie, C., and Mersha, T.B. (2017). AdmixPower: Statistical Power and Sample Size Estimation for Mapping Genetic Loci in Admixed Populations. *Genetics* *207*, 873–882.
44. Gautam, Y., Ghandikota, S., Chen, S., and Mersha, T.B. (2019). PAMAM: Power analysis in multiancestry admixture mapping. *Genet. Epidemiol.* *43*, 831–843.
45. Grinde, K.E., Brown, L.A., Reiner, A.P., Thornton, T.A., and Browning, S.R. (2019). Genome-wide Significance Thresholds for Admixture Mapping Studies. *Am. J. Hum. Genet.* *104*, 454–465.
46. Devlin, B., Roeder, K., and Wasserman, L. (2001). Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* *60*, 155–166.
47. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of

- variants throughout the human genome. *Nucleic Acids Res.* *47*, D886–D894.
48. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* *22*, 1790–1797.
  49. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* *51*, D977–D985.
  50. Myers, T.A., Chanock, S.J., and Machiela, M.J. (2020). LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front. Genet.* *11*, 157.
  51. Suarez-Farinas, M., Tintle, S.J., Shemer, A., Chiricozzi, A., Nograles, K., Cardinale, I., Duan, S., Bowcock, A.M., Krueger, J.G., and Guttman-Yassky, E. (2011). Nonlesional atopic dermatitis skin is characterized by broad terminal differentiation defects and variable immune abnormalities. *J. Allergy Clin. Immunol.* *127*, 954–964.e1-4.
  52. Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* *47*, W234–W241.
  53. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* *25*, 402–408.
  54. Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.* *11*, 1889–1907.
  55. Poto, R., Cristinziano, L., Modestino, L., de Paulis, A., Marone, G., Loffredo, S., Galdiero, M.R., and Varricchi, G. (2022). Neutrophil Extracellular Traps, Angiogenesis and Cancer. *Bio-medicines* *10*, 431.
  56. Kisiel, M.A., and Klar, A.S. (2019). Isolation and Culture of Human Dermal Fibroblasts. *Methods Mol. Biol.* *1993*, 71–78.
  57. Sliz, E., Huilaja, L., Pasanen, A., Laisk, T., Reimann, E., Mägi, R., FinnGen; and Estonian Biobank Research Team, Hanula-Jouppi, K., Peltonen, S., et al. (2022). Uniting biobank resources reveals novel genetic pathways modulating susceptibility for atopic dermatitis. *J. Allergy Clin. Immunol.* *149*, 1105–1112.e9.
  58. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* *53*, 1415–1424.
  59. Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Sakashita, M., Yamada, T., Fujieda, S., Tanaka, S., Doi, S., et al. (2012). Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* *44*, 1222–1226.
  60. Tanaka, N., Koido, M., Suzuki, A., Otomo, N., Suetsugu, H., Kochi, Y., Tomizuka, K., Momozawa, Y., Kamatani, Y., Biobank Japan Project, et al. (2021). Eight novel susceptibility loci and putative causal variants in atopic dermatitis. *J. Allergy Clin. Immunol.* *148*, 1293–1306.
  61. DeVore, S.B., Stevens, M.L., He, H., Biagini, J.M., Kroner, J.W., Martin, L.J., and Khurana Hershey, G.K. (2022). Novel role for caspase recruitment domain family member 14 and its genetic variant rs11652075 in skin filaggrin homeostasis. *J. Allergy Clin. Immunol.* *149*, 708–717.
  62. Lopez Lera, A. (2021). Pathophysiology and underlying mechanisms in hereditary angioedema. *Balkan Med. J.* *38*, 82–88.
  63. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
  64. Tomishige, N., Noda, Y., Adachi, H., Shimoi, H., and Yoda, K. (2005). SKG1, a suppressor gene of synthetic lethality of kex2-Deltagas1Delta mutations, encodes a novel membrane protein that affects cell wall composition. *Yeast* *22*, 141–155.
  65. Dattilo, V., Amato, R., Perrotti, N., and Gennarelli, M. (2020). The Emerging Role of SGK1 (Serum- and Glucocorticoid-Regulated Kinase 1) in Major Depressive Disorder: Hypothesis and Mechanisms. *Front. Genet.* *11*, 826.
  66. Chung, J., Nakatsu, F., Baskin, J.M., and De Camilli, P. (2015). Plasticity of PI4KIIIalpha interactions at the plasma membrane. *EMBO Rep.* *16*, 312–320.
  67. Bojjireddy, N., Guzman-Hernandez, M.L., Reinhard, N.R., Jovic, M., and Balla, T. (2015). EFR3s are palmitoylated plasma membrane proteins that control responsiveness to G-protein-coupled receptors. *J. Cell Sci.* *128*, 118–128.
  68. Kumper, M., Zamek, J., Steinkamp, J., Pach, E., Mauch, C., and Zigrino, P. (2022). Role of MMP3 and fibroblast-MMP14 in skin homeostasis and repair. *Eur. J. Cell Biol.* *101*, 151276.
  69. Kruglikov, I.L., Joffin, N., and Scherer, P.E. (2020). The MMP14-caveolin axis and its potential relevance for lipoe-dema. *Nat. Rev. Endocrinol.* *16*, 669–674.
  70. Guo, H., An, J., and Yu, Z. (2020). Identifying Shared Risk Genes for Asthma, Hay Fever, and Eczema by Multi-Trait and Multiomic Association Analyses. *Front. Genet.* *11*, 270.
  71. Graham, S.E., Clarke, S.L., Wu, K.H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* *600*, 675–679.
  72. Duncan, L., Shen, H., Gelaye, B., Meijssen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* *10*, 3328.