LEWIS UNIVERSITY  
IMPUTING EQTL AND GWAS SUMMARY STATISTICS TO EXAMINE MULTI-ANCESTRY AD GENETIC RISK FACTORS  
RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF SCIENCE, DATA SCIENCE

BY: HALLEIGH KELCHEN

DIRECTOR: DAWN GRAUNKE  
ROMEOVILLE, IL 29: JUNE, 24

ABSTRACT

Atopic Dermatitis (AD) is an inflammatory skin condition that effects up to 15 to 20 percent of the population. A wide variety of genetic studies have been used to understand the genetic risk factors and etiology of AD, however known variants only account for a small portion of the heritability of this disease. Previous research has focused mainly on those of European descent, leaving many other ancestries unstudied. This study uses GWASLab and the S-PrediXcan TWAS method to examine genetic risk factors for AD across major superpopulations. Analysis with GWASLab identified 12 lead variants (p<5e-8), six of which were not identified in the original GWAS: *PRPF3, STAT5B, LCE5A, PRRT1, TRIB1*, and *CXCR5*. S-PrediXcan analysis using MASHR GTEx v8 cultured fibroblasts and skin tissues identified four lead variants: *LINC00302, LCE5A, OVOLI1,* and *IL-13.* All lead variants identified by S-PrediXcan except *LINC00302* were also identified by GWASLab. A total of seven lead variants not implicated in the original GWAS were identified by this study: *PRPF3, STAT5B, LCE5A, PRRT1, TRIB1*, *LINC00302*, and *CXCR5*. Future analysis is needed to perform fine-mapping and colocalization in the regions of these genes to determine specific ancestries associated with each lead variant.

TABLE OF CONTENTS

[CHAPTER I – INTRODUCTION 1](#_Toc169534874)

[TWAS and Their Uses for Disease Research 6](#_Toc169534875)

[Benefits of TWAS Compared to GWAS 7](#_Toc169534876)

[Overview of Recent Literature 5](#_Toc169534877)

[Objective of Study 7](#_Toc169534878)

[CHAPTER II – METHODS 8](#_Toc169534879)

[Description of Dataset 8](#_Toc169534880)

[Preprocessing and Post-GWAS Analysis Using GWASLab 8](#_Toc169534881)

[TWAS 11](#_Toc169534882)

[S-PrediXcan 11](#_Toc169534883)

[CHAPTER III – RESULTS 14](#_Toc169534884)

[GWASLab 14](#_Toc169534885)

[S-PrediXcan 19](#_Toc169534886)

[CHAPTER IV – CONCLUSIONS 29](#_Toc169534887)

[LITERATURE CITED 35](#_Toc169534888)

LIST OF TABLES

Table 1. Lead SNPs Extracted Using GWASLab (p<5e-8)…………………………………….18

LIST OF FIGURES

Figure 1. Summary of Effects of Skin Barrier on the Pathogenesis of AD……………………….3

Figure 2. Effects of Cytokines on Pathogenesis of Acute and Chronic AD………………………4

Figure 3. Graphic Summary of Methods………………………………………………………...10

Figure 4. Algorithm for S-PrediXcan……………………………………………………………13

Figure 5. Power Plot of Minor Allele Frequencies (MAF) per Samples………………………...15

Figure 6. Manhattan and Quartile-Quartile Plot of GWAS Summary Statistics from GWASLab………………………………………………………………………………………..16

Figure 7. Regional Plot of *FLG* Gene Location on CHR 1q21………………………………….17

Figure 8. Expected vs Observed P-values for the GTEx v8 Tissues and the GWAS Summary Statistics………………………………………………………………………………………….20

Figure 9. Top 10 Significant Gene Associations for Cultured Fibroblasts………………………21

Figure 10. Top 10 Significant Gene Associations for Whole Blood……………………………..22

Figure 11. Top 10 Significant Gene Associations for Sun-Exposed Skin………………………..23

Figure 12. Top 10 Significant Gene Associations for Sun-Free Skin……………………………24

Figure 13. Top 10 Significant Gene Associations for Cultured Fibroblasts by Z-score………….25

Figure 14. Top 10 Significant Gene Associations for Whole Blood by Z-score…………………26

Figure 15. Top 10 Significant Gene Associations for Sun-Exposed Skin by Z-score……………27

Figure 16. Top 10 Significant Gene Associations for Sun-Free Skin by Z-score…………………28

Figure 17. Effects of *IL-13* on Skin in AD (Napolitano, et al., 2023)…………………………...33

Figure 18. Pathogenesis of AD with *IL-13* and *OVOL1*…………………………………………34

LIST OF ABBREVIATIONS

AD Atopic Dermatitis

CHR Chromosome

eQTL. Expression Quantitative Trait Loci

FLG Filaggrin

GWAS Genome-Wide Association Study

ILCs Innate Lymphoid Cells

MAF Minor Allele Frequency

MASHR Multivariate Adaptive Shrinkage

SNP Single Nucleotide Polymorphism

TWAS Transcriptome-Wide Association Study

# CHAPTER I – INTRODUCTION

Atopic dermatitis (AD), also known as atopic eczema, is a common inflammatory skin condition that results in recurrent dry, itchy lesions, skin infections, and blisters. This heterogeneous disease affects up to 15-20% of the population and interactions between susceptibility genes, environmental factors, skin integrity and microbiota, immune dysregulation, and diet play a role in the onset and presentation of this disorder (Martin, et al., 2020). Using twin and family studies, the heritability of AD is estimated to be over 80% (Wu, et al., 2023). A wide variety of genetic studies, such as genome-wide association studies (GWAS), whole-exome sequencing, and next-generation sequencing (NGS) have been used to understand the genetic risk factors and etiology of AD. The strongest known risk factors are null mutations on the filaggrin (*FLG*) gene, a key protein in the formation of the skin barrier and epidermis (Martin, et al., 2020). Filaggrin changes to the composition of keratinocytes and the granular cell layer, alters the pH and hydration of the skin, increases Th2 inflammation, and results in higher penetration of allergens into the skin (Peng & Novak, 2012). A summary of how the pathogenesis of AD is impacted by skin barrier dysfunction and genetic and immunological factors is shown in Figure 1 (Peng & Novak, 2012). Cytokines also play an important role in the pathogenesis of AD. Environmental triggers and damage to the skin barrier cause keratinocytes to release interleukin (IL) 25, IL-33, and chemokines which activate dendritic cells (Kim, Kim, & Leung, 2019). The activated dendritic cells stimulate T- helper 2 cells to produce a variety of cytokines, such as IL-4, IL-13, IL-31, etc. (Kim, et al., 2019). This causes decreased antimicrobial peptide production, itch symptoms, and barrier dysfunction as shown in Figure 2 (Kim, et al., 2019). While part of the factors affecting the pathogenesis of AD have been identified, known genetic variants associated with AD only account for a small portion of the heritability of AD. This suggests there is still a plethora of genetic variants left to discover.

A diagram of a cell block

Description automatically generated

Figure 1. Summary of Effects of Skin Barrier on the Pathogenesis of AD (Peng & Novak, 2012).

A diagram of a chronic stage

Description automatically generated

Figure 2. Effects of Cytokines on Pathogenesis of Acute and Chronic AD (Kim, et al., 2019).

## Overview of Recent Literature

In recent years, genome-wide association studies (GWAS) have been used to determine genetic risk loci for AD. Most of the genetic variants that have been identified by previous AD GWAS map to noncoding regions of the genome and are likely to function by regulating gene expression, which is not uncommon for complex diseases (Wu, et al., 2023). Due to this, transcriptome-wide association studies (TWAS) may be better suited to prioritize the disease-risk genes, by predicting gene expression in large GWAS datasets (Wu, et al., 2023). A recent study by Wu et al. (2023), conducted a joint-tissue TWAS using gene expression weight from four different types of AD-related tissues and two large-scale meta-analyses of AD. This TWAS identified 51 genes associated with AD and used mRNA expression profiles and Mouse Genome Informatics databases to support their findings. Of the 51 genes identified, 19 were thought to be potential causal genes. Seven of the 19 casual genes identified were not implicated by previous GWASs, thus indicating that joint tissue TWAS could be more beneficial in revealing potential candidate genes than traditional GWAS. The seven newly identified genes were *ADCY3, AQP3, CRAT, DOLPP1, NEU4, PDCDC1,* AND *ZSCAN9*. Further research is needed to validate these findings and explain the potential roles of these and other genes in the regulation of AD. Additionally, all the individuals in this study were of European descent, meaning these findings might be relevant to other ethnicities. More research is needed to evaluate multi-ethnic causal genes for AD. (Wu, et al., 2023)

The heritability of eczema in those of European descent is estimated to be up to 90 percent, which is why a majority of GWAS focus on this population (Paternoster, et al., 2015). However, a few studies have been done including Japanese and Chinese populations. A multi-ethnic meta-analysis by Paternoster et al. (2015), was able to verify five previously reported risk loci for Japanese populations, two of which appeared to be Japanese-specific and three that were also associated with those of European descent (Paternoster, et al., 2015). There was also one locus from a Chinese GWAS that showed an association with Europeans (Paternoster, et al., 2015). Another multi-ancestry GWAS was done by Budu-Aggrey et al. (2023), which identified 89 loci associated with AD (Budu-Aggrey, et al., 2023). 75 of the loci were not independent of the European-only analysis, but five of the loci found were not associated with European populations, as opposed to other populations (Budu-Aggrey, et al., 2023).

## TWAS and Their Uses for Disease Research

GWAS have been used successfully to identify genetic variants linked to complex traits and diseases. GWAS reports genomic risk loci, which are chunks of interrelated single nucleotide polymorphisms (SNPs) that have a significant association with selected traits (Uffelmann, et al., 2021). The genomic risk loci can be used for a variety of applications, such as disease risk prediction and understanding genetic architecture. GWAS genotype individuals using DNA and phenotypic information and then go through quality control, an imputation process, and then a variety of statistical tests are implemented, and the results are interpreted using post-GWAS analysis. One of these post-GWAS analysis methods is transcriptome-wide association study (TWAS). TWAS incorporates information from GWAS and expression quantitation trait loci (eQTL) to test for gene-trait associations (Xie, Shan, Zhao, & Hou, 2021). TWAS uses a reference panel to train a genetic regulatory model. Regulatory weights are used to impute gene expression for individuals in GWAS cohorts. Associations between traits and predictive gene expression are then computed to determine the correlation between genes and traits (Mai, et al., 2023). TWAS incorporates transcriptomic and genetic data to detect gene-trait associations that can be used to study the pathology of many complex diseases.

## Benefits of TWAS Compared to GWAS

Unlike GWAS which faces a multiple testing burden, TWAS reduces this burden by performing association tests only for genes that are significantly regulated by genetic variations (Mai, et al., 2023). TWAS can be used to distinguish susceptible risk genes for multifaceted traits. While GWAS can be used to identify many significant variants, many are hard to interpret as they are in non-coding regions (Mai, et al., 2023). Instead, TWAS detects gene-trait associations that make understanding the genetic mechanisms of phenotypes easier to interpret (Mai, et al., 2023). TWAS imputation of eQTL data allows for genetically related genes associated with traits to be detected, whether they are close or far from the variants (Mai, et al., 2023). Additionally, TWAS has a higher gene-based interpretability and lower cost and complexity compared to GWAS (Mai, et al., 2023).

## Objective of Study

A transcriptome-wide association study (TWAS) of AD in multi-ethnic populations has yet to be done. This type of study would allow for more potential candidate genes to be identified than the traditional multi-ethnic GWAS. The objective of this study is to perform a multi-tissue multi-ancestry TWAS using the S-PrediXcan method proposed by Barbeira et al. (2018), expanding the data to include additional populations from the Paternoster et al. (2015) study. This study will identify AD-related genes and loci prominent across different populations, using summary statistics from Paternoster et al. (2015) GWAS.

# CHAPTER II – METHODS

## Description of Dataset

Publically available GWAS summary statistics for the Paternoster et al., (2015) study (GCST003184) were retrieved from the NHGRI-EBI GWAS Catalog (Sollis, et al., 2022). This dataset contains data from multi-ethnic cohorts from 26 studies. There are a total of 18,900 cases and 84,166 controls of European descent from a combination of 22 studies. Additionally, the data for the multi-ancestry analysis contains information from 1 Japanese study containing 1,472 cases of AD and 7,966 controls, one African American study containing 422 cases and 844 controls, one study of Latin-Americans containing 300 cases and 1,592 controls and one study of mixed non-European individuals containing 305 cases and 896 controls. The dataset also includes data from 18 studies, including 30,588 European cases and 226,537 controls, 459 African American cases and 729 controls, 1012 Chinese cases, and 1,362 controls used for replication. (Paternoster et al., 2015)

## Preprocessing and Post-GWAS Analysis Using GWASLab

All data analysis for this study was performed using Python Jupyter Notebooks (Kluyver, et al., 2016). A graphic summary of the methods used in this study is shown in Figure 3. GWASLab, a Python toolkit developed for working with summary statistics, was used to process and visualize the GWAS summary statistics before the TWAS (He, Koido, Shimmori, & Kamatani, 2023). Columns from the original summary statistics were used to calculate the Z score, minor allele frequency (MAF), and MLOG10P. Variants from the major histocompatibility complex region (HLA variants) were removed, along with those with invalid rsID numbers, outliers, and variants with bad statistics resulting in a total of 10,334,839 rows after cleaning. The clean data was extracted to use in the TWAS. A summary of the dataset was generated, showing 752 variants with significant p values and 27,421 variants with rare MAFs. The human assembly GRCh37 was used as a reference panel to extract lead variants from the dataset (Herrero, et al.,2016). Lead variants are defined in this study as variants having a significant p-value of 5e-8at a given locus. GWASLab was used to create a Manhattan and quantile-quantile plot for the data and a regional plot of the genomic location of the *FLG* gene. A liftover was performed on the clean data to transfer the genomic build from hg19 to the newer hg38 build.

Figure 3. Graphic Summary of Methods

## TWAS

TWAS is a post-analysis method used to determine significant gene-trait associations regulated by significant variants, mainly single nucleotide polymorphisms (SNPs) identified by GWAS (Mai, et al., 2023). Association tests are conducted for gene-trait pairs using expression quantitative trait loci (eQTL) effects and GWAS associations (Mai, et al., 2023). The TWAS pipeline can be divided into three steps: training, imputation, and association stages (Mai, et al., 2023). The training stage is used to determine the regulatory effect size of multiple SNPs on the gene expression level using a multivariate SNP model made from a panel containing both genotype and expression information (Mai, et al., 2023). The imputation stage is used to determine the predicted gene expression level of GWAS individuals (Mai, et al., 2023). The association stage implements hypothesis tests between target traits and the predicted gene expression with different association models to calculate effect sizes of significant trait-associated genes (Mai, et al., 2023). These steps can also be combined, as the imputation will also produce association data.

## S-PrediXcan

Summary-PrediXcan (S-PrediXcan), a modification of PrediXcan for GWAS summary statistics, is a gene-level association approach that tests the mediating effects of gene expression on phenotypes is the TWAS method that was implemented in this study (Barbeira, et al., 2018). S-PrediXcan uses the Wald statistic to evaluate the association between the phenotype and the predicted gene expression (Figure 4). Tissue-specific multivariate adaptive shrinkage (MASHR) genotype-tissue expression (GTEx V8) of AD-relevant tissues were used to train the prediction model as both pre-computed weights and the reference panel (Barbeira, et al., 2021). Similar to Wu et. al., (2023), the tissues included were skin, cultured fibroblasts, Epstein-Barr virus (EBV) transformed lymphocytes, and whole blood. The top 10 significant p-values and z-scores for each tissue were graphed.

A diagram of a mathematical model

Description automatically generated

Figure 4. Algorithm for S-PrediXcan (Barbeira, et al., 2018).

# CHAPTER III – RESULTS

## GWASLab

GWASLab was used to evaluate the GWAS summary statistics from Paternoster et al., (2015). GWASLab identified 27,421 variants with rare MAF values and 752 variants with significant P values. A power plot examining the MAFs of the data is shown in Figure 5. The summary statistics were used to create an MQQ plot to identify strongly associated risk loci and examine the skewness of the data as shown in Figure 6. Figure 6 shows a strong association between several positions on chromosome (CHR) 1 in the epidermal differentiation complex region of CHR 1q21. The *FLG* gene, commonly associated with AD is in this region. A regional plot of CHR 1q21 is shown in Figure 7 with the location of the *FLG* gene flagged. GWASLab was also used to extract 12 lead SNPs (p<5e-8), six of which were not identified by the Paternoster et al., (2015) GWAS (Table 1). The lead variants identified by the study that were not identified by the GWAS are *PRPF3, STAT5B, LCE5A, PRRT1, TRIB1*, and *CXCR5*.

A graph of a number of samples

Description automatically generated

Figure 5. Power Plot of Minor Allele Frequencies (MAF) per Samples

A graph of a number of blue and black colored lines

Description automatically generated with medium confidence

Figure 6. Manhattan and Quartile-Quartile Plot of GWAS Summary Statistics from GWASLab

A graph with blue and red dots

Description automatically generated

Figure 7. Regional Plot of *FLG* Gene Location on CHR 1q21

Table 1. 12 Lead SNPs Extracted Using GWASLab (p<5e-8)

A screenshot of a computer

Description automatically generated

## S-PrediXcan

The TWAS performed using S-PrediXcan was used to test the mediating effects of gene expression levels in AD-relevant tissues. Five tissues were used in the study: EBV-transformed lymphocytes, cultured fibroblasts, whole blood, sun-exposed skin, and sun-free skin. No significant associations were found between GTEx v8 EBV-transformed lymphocytes and the GWAS summary statistics. Associations were found between the other four tissues and the summary statistics. A quartile-quartile plot examining the p-values for the four remaining tissues is shown in Figure 8; the red line indicates the expected p-value distribution under the null hypothesis with a uniform distribution. A list of the top ten significant gene associations for cultured fibroblasts, whole blood, sun-exposed skin, and sun-free skin are shown in Figures 9, 10, 11, and 12 respectively. No lead variants were found using GTEx whole blood tissue for S-PrediXcan. Four genes from the cultured fibroblasts and skin tissues were identified as lead variants (p<5e-8): *OVOL1*, *LINC00302*, *LCE5A*, and *IL-13*. Three of the lead variants identified by S-PrediXcan were also identified using GWASLab: *OVOL1*, *LCE5A*, and *IL-13*. Only one lead variant identified using S-PrediXcan was not found in the original GWAS: *LCE5A*. The top 10 significant z-scores for cultured fibroblasts, whole blood, sun-exposed skin, and sun-free skin are shown in Figures 13, 14, 15, and 16, respectively. Lead variants *OVOL1* and *LCE5A* are shown with z-scores greater than 5.45 in the cultured fibroblast and sun-free skin, which is consistent with their p-values in Figures 9 and 12. Although the other lead variants identified using S-PrediXcan are not pictured in the top 10 for each tissue, they all have z-scores greater than 2.576, which is consistent with a 99% confidence level.

A graph of a line

Description automatically generated with medium confidence

Figure 8. Expected vs Observed P-values for the GTEx v8 Tissues and the GWAS Summary Statistics

A graph of different colored bars

Description automatically generated

Figure 9. Top 10 Significant Gene Associations for Cultured Fibroblasts

A graph of different colored bars

Description automatically generated

Figure 10. Top 10 Significant Gene Associations for Whole Blood

A graph with different colored bars

Description automatically generated

Figure 11. Top 10 Significant Gene Associations for Sun-Exposed Skin

A graph with different colored bars

Description automatically generated

Figure 12. Top 10 Significant Gene Associations for Sun-Free Skin

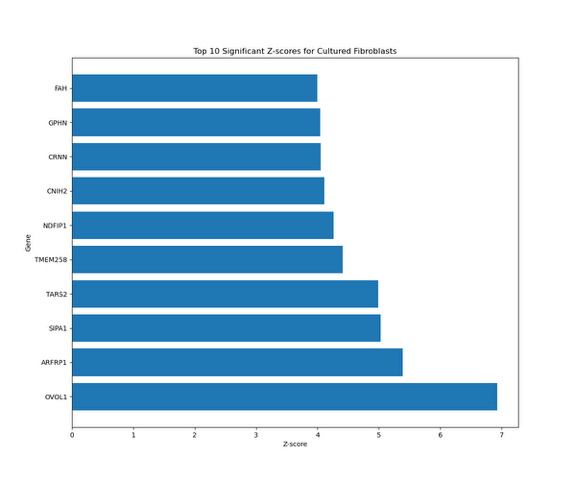


Figure 13. Top 10 Significant Gene Associations for Cultured Fibroblasts by Z-score

A graph of a bar chart

Description automatically generated with medium confidence

Figure 14. Top 10 Significant Gene Associations for Whole Blood by Z-score

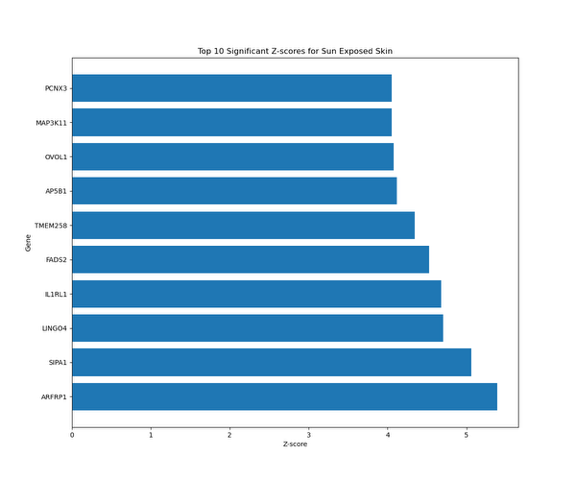


Figure 15. Top 10 Significant Gene Associations for Sun-Exposed Skin by Z-score

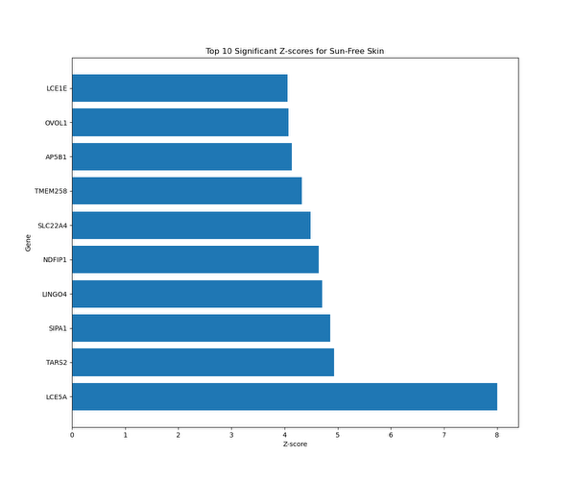


Figure 16. Top 10 Significant Gene Associations for Sun-Free Skin by Z-score

# CHAPTER IV – CONCLUSIONS

AD is the most common allergic inflammatory skin disease. To examine the relationship between gene expression in multiple ancestries and AD, analysis with GWASLab and a multi-tissue TWAS were performed on GWAS summary statistics using S-PrediXcan. Significantly associated alleles, also known as lead variants (p<5e-8) were identified using these methods. GWASLab identified 12 lead variants on genes *PRPF3, LCE5A, IL18RAP, IL-13, PRRT1, TRIB1, OVOLI1, C11ORF30, CXCR5, STAT5B, ACTL9*, and *RTEL1* (Table 1). Six of the lead variants identified were not identified by the original GWAS: *PRPF3, STAT5B, LCE5A, PRRT1, TRIB1*, and *CXCR5*. S-PrediXcan of cultured fibroblasts and sun-exposed skin GTEx tissues identified one lead variant, *OVOL1* and *LINC00302* respectively(Figures 9 and 11). While the associations with whole blood GTEx tissue identified many significant variants, none meet the threshold to be lead variants (Figure 10). S-PrediXcan of sun-free skin GTEx tissue identified two lead variants: *LCE5A* and *IL-13* (Figure 12). Three of the four lead variants identified by S-PrediXcan were also identified by GWASLab: *LCE5A, OVOLI1,* and *IL-13.* This study identified a total of seven lead variants not implicated in the original GWAS: *PRPF3, STAT5B, LCE5A, PRRT1, TRIB1*, *LINC00302*, and *CXCR5*. The 13 lead variants identified by GWASLab all have a 99% confidence level or greater (z-score > 2.576), with the strongest associations being with lead variants *LCE5A* and *OVOLI1* (z-score> 5.45) (Figures 13 and 16).While patients from multiple ancestries were included in the study and contributed to the overall TWAS signal, no ancestry-specific loci were identified. Since the publication of the Paternoster et al., (2015) GWAS, all lead variants identified in this studyhave been implicated in the risk for AD and are known loci (Sollis, et al., 2022).

The pathogenesis of AD involves a combination of genetic, environmental, and immunological factors. The existing pathogenic model is focused on the activation of a variety of type 2 immune cells, such as T helper cells, innate lymphoid cells (ILCs), eosinophils, mast cells, and involvement from other pathways (Chiricozzi, Maurelli, Calabrese, Peris, & Girolomoni, 2023). The current model can explain most of the identified lead variants involvement in the pathogenesis of AD. Signal transducer and activator of transcription 5B (*STAT5B*) is significantly associated with the risk of AD in Europeans (Ando, et al., 2014). Recruitment of *IL-3* and other cytokines causes suppression of *STAT5* activity (Ando, et al., 2014). Silencing of *STAT5B* expression induces apoptosis in human mast cells (Ando, et al., 2014). This may explain part of the pathogenic mechanism for AD. *TRIB1* is a vital regulator of eosinophil differentiation, which are important effector cells in allergic inflammation and allergic diseases (Danger, Feseha, & Brouard, 2022). It is also highly expressed in CD4 T cells, which are found in increased numbers in AD lesions (Danger, et al., 2022). *CXCR5* are chemokine receptors expressed by lymphoid tissue-inducing cells, a form of ILCs (Jia, Wan, & Zhang, 2023). ILCs accumulate in the skin of AD patients and produce a variety of cytokines associated with T helper cells, which causes immune and inflammatory responses (Jia, et al., 2023). Recent studies have shown an abnormal ILC activation is strongly associated with the pathophysiology, onset, and progression of AD (Jia, et al., 2023).

Additionally, *FLG* mutations are the strongest genetic risk factor for AD. *LCE5A* is one of many loss-of-function (LoF) variants of filaggrin, which results in a reduction of skin hydration and allows allergens to penetrate the skin easier (Chong, Visitsunthorn, & Ong, 2022).  *LCE5A* is responsible for keratinization and is located in the epidermal differentiation complex (EDC) CHR 1q21 region with *FLG* (Martin, et al., 2020). *FLG* LoF mutations are detected in up to 50% of European and 27% of Asian AD patients, while no associations have been found between LoF mutations and African populations (Chirocozzi, et al., 2023). Additionally, patients with *FLG* LOF mutations exhibit a more severe form of AD involving persistent lesions, increased skin infections and allergic sensitization, and impairment of the immune system (Chirocozzi, et al., 2023). *LINC00302* is a poorly studies long intergenic nonprotein coding (linc) RNA that is also located in the EDC region (Sahlen, et al., 2021). *LINC00302* is expressed almost entirely in the skin and is downregulated in AD lesions; it is involved in keratinocyte differentiation and affects the expression of *FLG* (Sahlen, et al., 2021). Future research is needed to establish its specific role is keratinocyte differentiation and AD pathogenesis, as little research has been done on this specific variant.

Following the onset of AD, many patients experience the atopic march, which is the progression of other types of allergies, such as food allergy, asthma, conjunctivitis, and hay fever (Chirocozzi, et al., 2023). Two of the more understudied lead variants identified by this study are associated with some of these comorbidities. *C11ORF30*, a transcriptional repressor, is associated with eczema and eczema-associated asthma, a known comorbidity of AD (Martin, et al., 2020)*.* A moderate association has been found between *C11ORF30* and AD in Japanese and Chinese populations(Kaufman, Guttman-Yassky, & Alexis, 2018). *PRPF3* is linked to allergy in general, in addition to eczema and retinitis pigmentosa (Martin, et al., 2020)*.* Its specific role in allergy and AD is unknown; it is known that *PRPF3* is involved in pre-mRNA splicing, but other than that further investigation is needed to identify how it plays into AD pathogenesis. More research is needed on both the role of *C11ORF30* and *PRPF3* specific roles in AD.

In addition to FLG, *IL-13* and *OVOL1* play a pivotal role in the pathogenesis of AD; the three genes form the *IL-13-OVOL1-FLG* axis. Previous studies suggest that *IL-13* is a crucial cytokine that plays a key role in AD pathogenesis; aside from *FLG*, this is the most studied AD-related gene (Napolitano, di Vico, Ruggiero, Fabbrocini, & Patruno, 2023). *IL-13* is overexpressed in the skin lesions of AD patients and polymorphisms of *IL-13* have been associated with an AD predisposition in Chinese, Korean, and Japanese populations (Chirocozzi, et al., 2023). *IL-13* recruits inflammatory cells, changes the skin microbiome, and alters the skin barrier (Napolitano, et al., 2023). It also activates the sensory nerve that causes the itch sensation commonly experienced by AD patients (Napolitano, et al., 2023). Figure 17 shows the effect of *IL-13* on AD pathogenesis (Napolitano, et al., 2023). The function of *OVOL1* is closely related to *IL-13*, as it is an upstream transcription factor that regulates *FLG* expression (Furue, et al., 2019). As shown in Figure 18, high levels of *IL-13* cause *OVOL1* inactivation and up-regulates the periostin-IL-24 axis (Furue, et al., 2019). This results in the downregulation of *FLG* and barrier dysfunction consistent with the pathogenesis of AD. The specific mutation of *OVOL1* identified using GWASLab is strongly associated with the Chinese population (Kaufman, et al., 2018).

A diagram of a cell structure

Description automatically generated with medium confidence

Figure 17. Effects of *IL-13* on Skin in AD (Napolitano, et al., 2023).

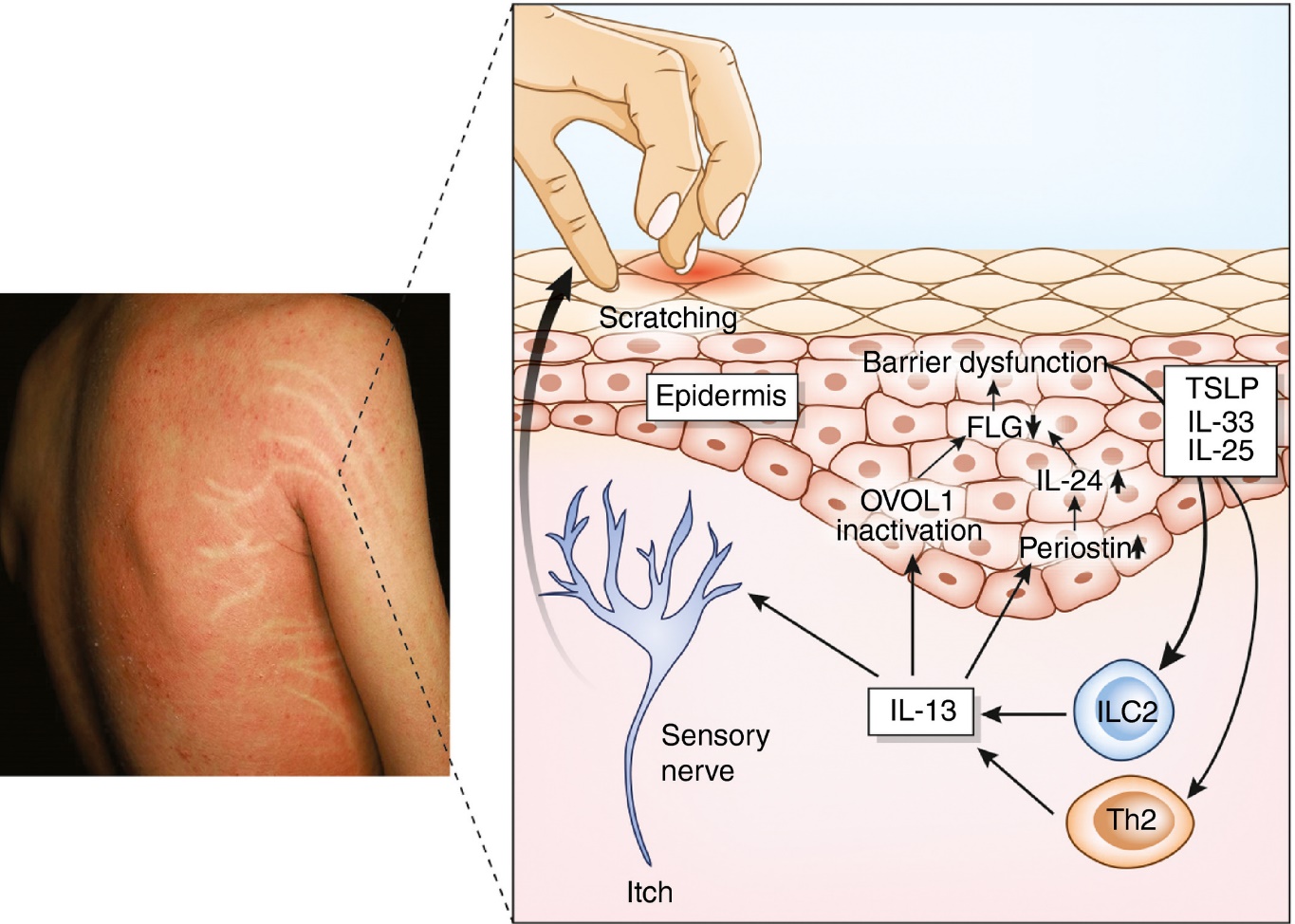


Figure 18. Pathogenesis of AD with *IL-13* and *OVOL1*

This study was able to identify 13 lead variants, seven of which were not associated with the original GWAS. While many ethnic ancestries are included in the GWAS summary statistics, no ancestry-specific associations were able to be identified by this study. Fine mapping at regions containing the lead variants using ancestry specific reference data, such as the MA-FOCUS method proposed in Lu et al.,(2022) is needed to identify ancestry-specific loci. Additionally, the populations used in this study were disproportionate, which may have affected the linkage disequilibrium. Colocalization analysis of the TWAS results is needed to identify causal pathways, prioritize causal variants of AD, and identify any false positives cause by the linkage disequilibrium. Additional data in AD from other populations is also needed to help understand the different risk factors between Europeans and other major populations.

In this study, two powerful approaches, GWASLab and S-PrediXcan were used to identify AD-related genes and loci prominent among different ancestry populations. A total of 13 lead variants associated with AD risk and pathogenesis were identified, seven of which were not implicated in the original GWAS. All AD-related genes identified in this study have been identified in other genetic studies, thus validating the results of this study (Sollis, et al., 2022). Continued research is needed to identify ancestry-specificity of the lead variants identified in this study.

# LITERATURE CITED

Ando, T., Xiao, W., Gao, P., Namiranian, S., Matsumoto, K., Tomimori, Y., Hong, H., Yamashita, H., Kimura, M., Kashiwakura, J., Hata, T. R., Izuhara, K., Gurish, M. F., Roers, A., Rafaels, N. M., Barnes, K. C., Jamora, C., Kawakami, Y., & Kawakami, T. (2014). Critical role for mast cell STAT5 activity in skin inflammation. *Cell Reports*, *6*(2), 366–376. <https://doi.org/10.1016/j.celrep.2013.12.029>

Barbeira, A. N., Dickinson, S. P., Bonazzola, R., Zheng, J., Wheeler, H. E., Torres, J. M., Torstenson, E. S., Shah, K. P., Garcia, T., Edwards, T. L., Stahl, E. A., Huckins, L. M., Aguet, F., Ardlie, K. G., Cummings, B. B., Gelfand, E. T., Getz, G., Hadley, K., Handsaker, R. E., … Im, H. K. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications*, *9*(1). <https://doi.org/10.1038/s41467-018-03621-1>

Barbeira, A. N., Bonazzola, R., Gamazon, E. R., Liang, Y., Park, Y., Kim-Hellmuth, S., Wang, G., Jiang, Z., Zhou, D., Hormozdiari, F., Liu, B., Rao, A., Hamel, A. R., Pividori, M. D., Aguet, F., Bastarache, L., Jordan, D. M., Verbanck, M., Do, R., … Im, H. K. (2021). Exploiting the GTEX resources to decipher the mechanisms at GWAS loci. *Genome Biology*, *22*(1). <https://doi.org/10.1186/s13059-020-02252-4>

Budu-Aggrey, A., Kilanowski, A., Sobczyk, M. K., Shringarpure, S. S., Mitchell, R., Reis, K., Reigo, A., Mägi, R., Nelis, M., Tanaka, N., Brumpton, B. M., Thomas, L. F., Sole-Navais, P., Flatley, C., Espuela-Ortiz, A., Herrera-Luis, E., Lominchar, J. V., Bork-Jensen, J., Marenholz, I., … Paternoster, L. (2023). European and multi-ancestry genome-wide association meta-analysis of atopic dermatitis highlights the importance of systemic immune regulation. *Nature Communications*, *14*(1). <https://doi.org/10.1038/s41467-023-41180-2>

Chiricozzi, A., Maurelli, M., Calabrese, L., Peris, K., & Girolomoni, G. (2023). Overview of Atopic Dermatitis in Different Ethnic Groups. Journal of clinical medicine, 12(7), 2701. <https://doi.org/10.3390/jcm12072701>

Danger, R., Feseha, Y., & Brouard, S. (2022). The Pseudokinase TRIB1 in Immune Cells and Associated Disorders. *Cancers*, *14*(4), 1011. <https://doi.org/10.3390/cancers14041011>

Furue, K., Ito, T., Tsuji, G., Ulzii, D., Vu, Y. H., Kido‐Nakahara, M., Nakahara, T., & Furue, M. (2019). The IL‐13–OVOL1–FLG axis in atopic dermatitis. Immunology, 158(4), 281286. <https://doi.org/10.1111/imm.13120>

He, Y., Koido, M., Shimmori, Y., Kamatani, Y. (2023). GWASLab: a Python package for processing and visualizing GWAS summary statistics. Preprint at Jxiv, 2023-5. <https://doi.org/10.51094/jxiv.370>

Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M. J. Searle, Ridwan Amode, Simon Brent, William Spooner, Eugene Kulesha, Andrew Yates, Paul Flicek, Ensembl comparative genomics resources, Database, Volume 2016, 2016 bav096, <https://doi.org/10.1093/database/bav096>

Jia, H., Wan, H., & Zhang, D. (2023). Innate lymphoid cells: a new key player in atopic dermatitis. *Frontiers in immunology*, *14*, 1277120. <https://doi.org/10.3389/fimmu.2023.1277120>

Kaufman, B. P., Guttman‐Yassky, E., & Alexis, A. F. (2018a). Atopic dermatitis in diverse racial and ethnic groups—variations in epidemiology, genetics, clinical presentation and treatment. *Experimental Dermatology*, *27*(4), 340–357. <https://doi.org/10.1111/exd.13514>

Kim, J., Kim, B. E., & Leung, D. Y. M. (2019, March 1). Pathophysiology of atopic dermatitis: Clinical implications. Allergy and asthma proceedings. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6399565/>

Kluyver, T., Ragan-Kelley, B., Fernando P&#x27;erez, Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90).

Lu, Z., Gopalan, S., Yuan, D., Conti, D. V., Pasaniuc, B., Gusev, A., & Mancuso, N. (2022). Multi-ancestry fine-mapping improves precision in identifying causal genes in transcriptome-wide association studies. *The American Journal of Human Genetics*, *109*(8), 1388–1404. <https://doi.org/10.1016/j.ajhg.2022.07.002>

Mai, J., Lu, M., Gao, Q., Zeng, J., & Xiao, J. (2023). Transcriptome-wide association studies: Recent advances in methods, applications and available databases. *Communications Biology*, *6*(1). <https://doi.org/10.1038/s42003-023-05279-y>

Martin, M. J., Estravís, M., García-Sánchez, A., Dávila, I., Isidoro-García, M., & Sanz, C. (2020). Genetics and epigenetics of atopic dermatitis: An updated systematic review. *Genes*, *11*(4), 442. <https://doi.org/10.3390/genes11040442>

Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D. P., Curtin, J. A., Bønnelykke, K., Tian, C., Takahashi, A., Esparza-Gordillo, J., Alves, A. C., Thyssen, J. P., den Dekker, H. T., Ferreira, M. A., Altmaier, E., Sleiman, P. M., Xiao, F. L., Gonzalez, J. R., Marenholz, I., … Weidinger, S. (2015). Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature genetics*, *47*(12), 1449–1456. <https://doi.org/10.1038/ng.3424>

Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J, Ibrahim A, Ji Y, John S, Lewis E, MacArthur JAL, McMahon A, Osumi-Sutherland D, Panoutsopoulou K, Pendlington Z, Ramachandran S, Stefancsik R, Stewart J, Whetzel P, Wilson R, Hindorff L, Cunningham F, Lambert SA, Inouye M, Parkinson H, Harris LW. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2022 Nov 9:gkac1010. doi: 10.1093/nar/gkac1010. Epub ahead of print. PMID: 36350656.

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). <https://doi.org/10.1038/s43586-021-00056-9>

Wu, H., Ke, X., Huang, W., Shi, W., Yao, S., Duan, Y.-Y., Tian, W., Dong, S.-S., Xue, H.-Z., & Guo, Y. (2023). Multitissue integrative analysis identifies susceptibility genes for atopic dermatitis. *Journal of Investigative Dermatology*, *143*(4). <https://doi.org/10.1016/j.jid.2022.09.006>

Xie, Y., Shan, N., Zhao, H., & Hou, L. (2021). Transcriptome Wide Association Studies: General framework and methods. *Quantitative Biology*, *9*(2), 141–150. <https://doi.org/10.15302/j-qb-020-0228>