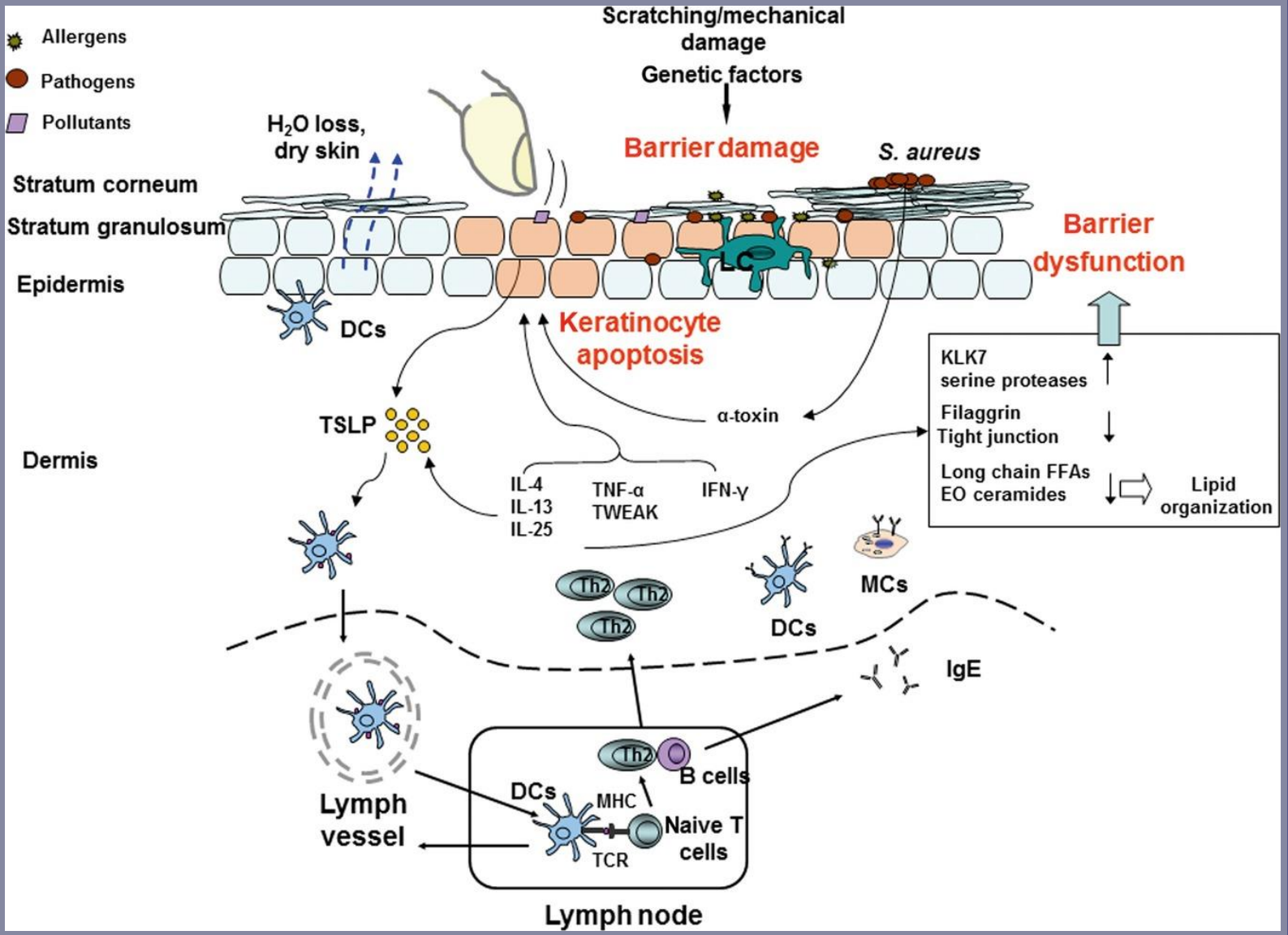
Imputing eQTL and GWAS Summary Statistics to Examine Multi-Ancestry Atopic Dermatis Genetic Risk Factors

By: Halleigh Kelchen



Peng & Novak (2015) *Clin. Exp. Allergy*

What Is Atopic Dermatitis?

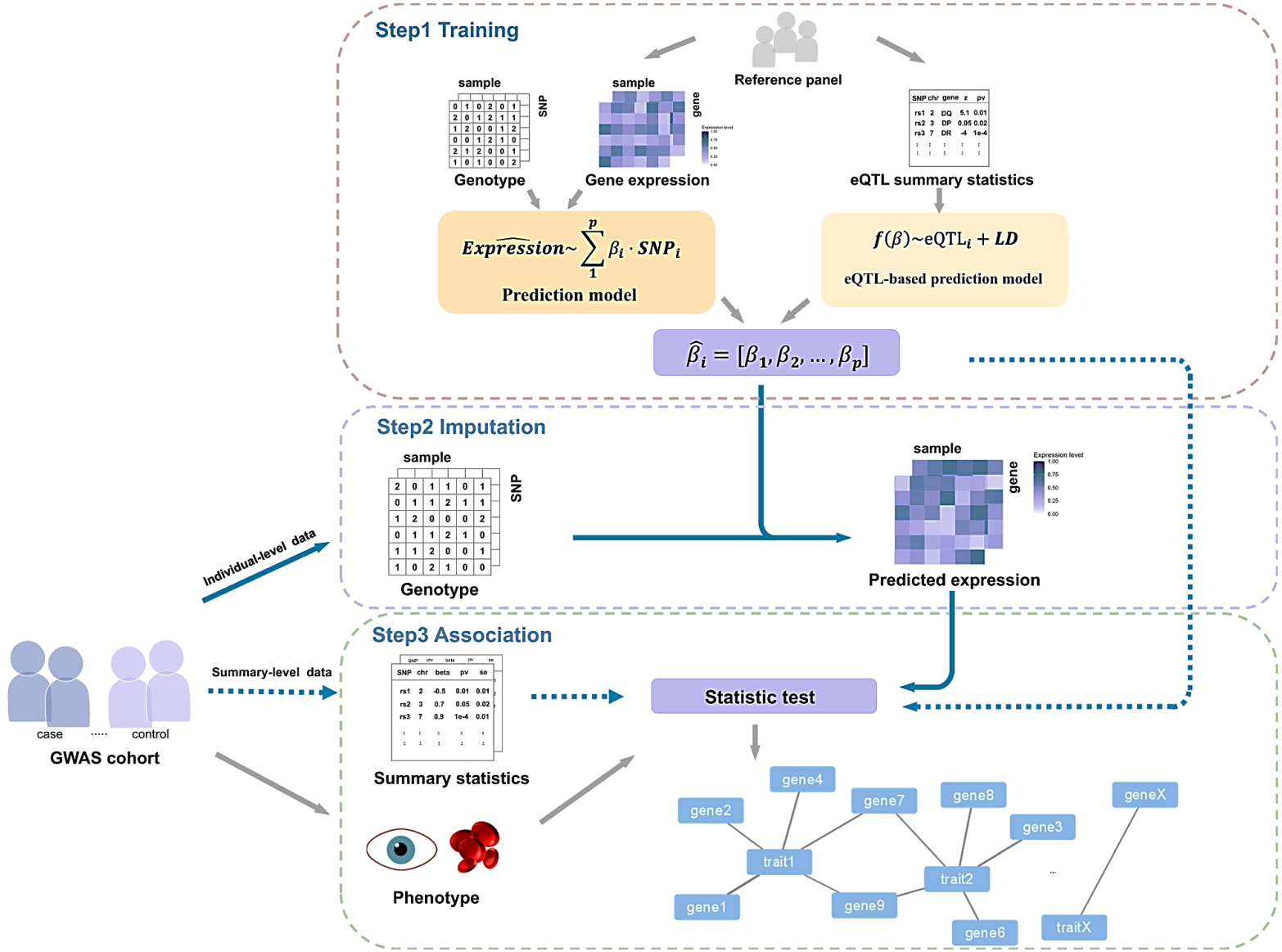
- Also known as atopic eczema
- Common inflammatory skin condition
- Accompanied by recurrent dry, itchy lesions, skin infections, and blisters
- Affects up to 15-20% of the population
- Heritability is estimated to be over 80%

Overview of Transcriptome-Wide Association Studies (TWAS)

- A post-genome-wide association study (GWAS) tool
- Used to evaluate the relationship between gene expression and a phenotype
- Combines expression quantitative loci (eQTL) with GWAS results
- Detects gene-trait associations which can be used to study the pathology of many complex diseases

Mai et al. (2023) *Communications Biology*

Example Workflow of TWAS Model Analysis

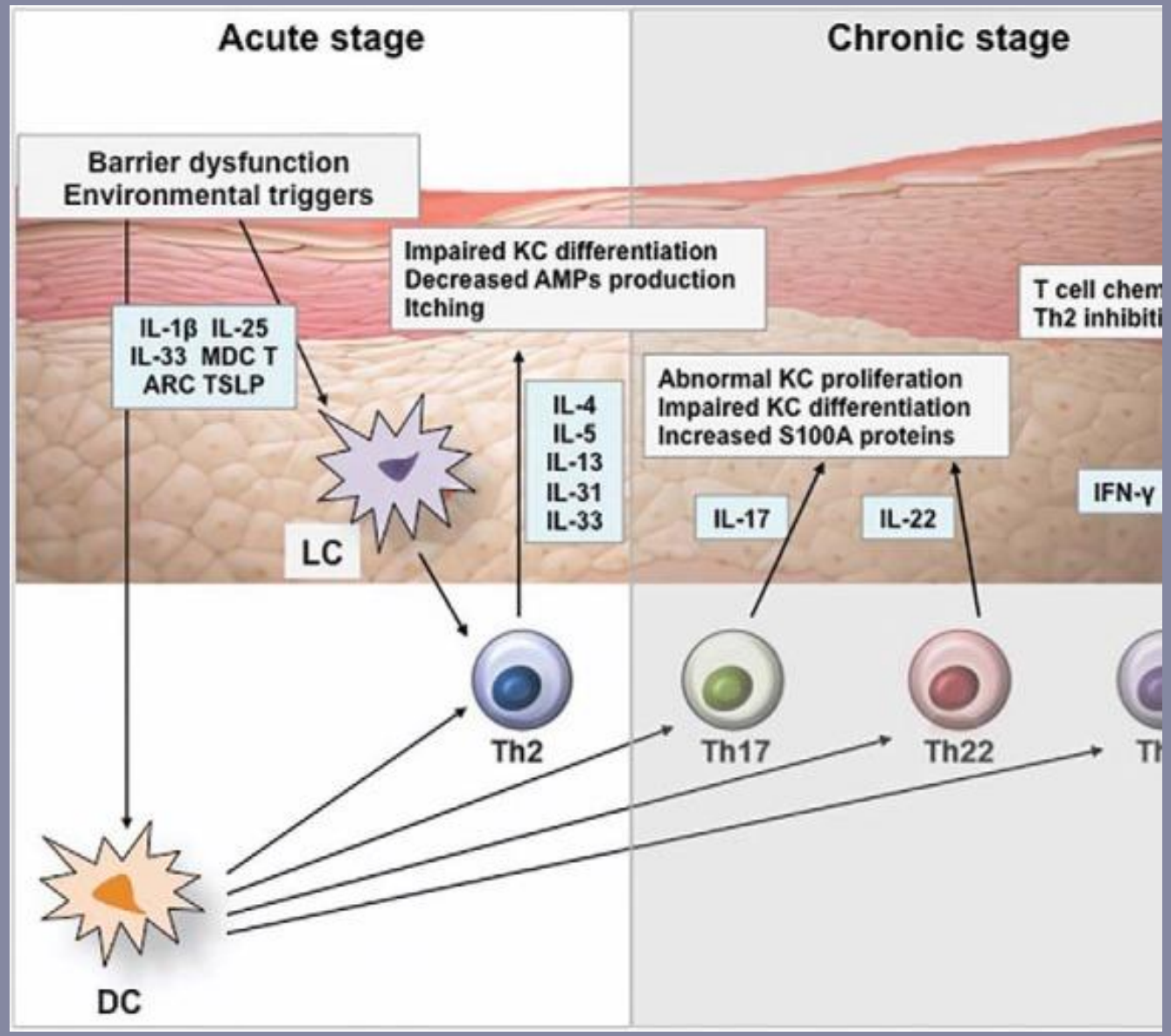


Mai et al. (2023) *Communications Biology*

Recent Literature

- Strongest known risk factor for AD is null mutations to the filaggrin (*FLG*) gene
- Polymorphisms on different immune pathway genes cause alternations in T-helper (th) type 2 signaling pathway
- Other immune-related genes associated with AD risk are interleukin (IL) 4, IL-13, IL-31, IL-33, toll-like receptor 2, etc.
- A strong correlation has been found between AD and other inflammatory disorders

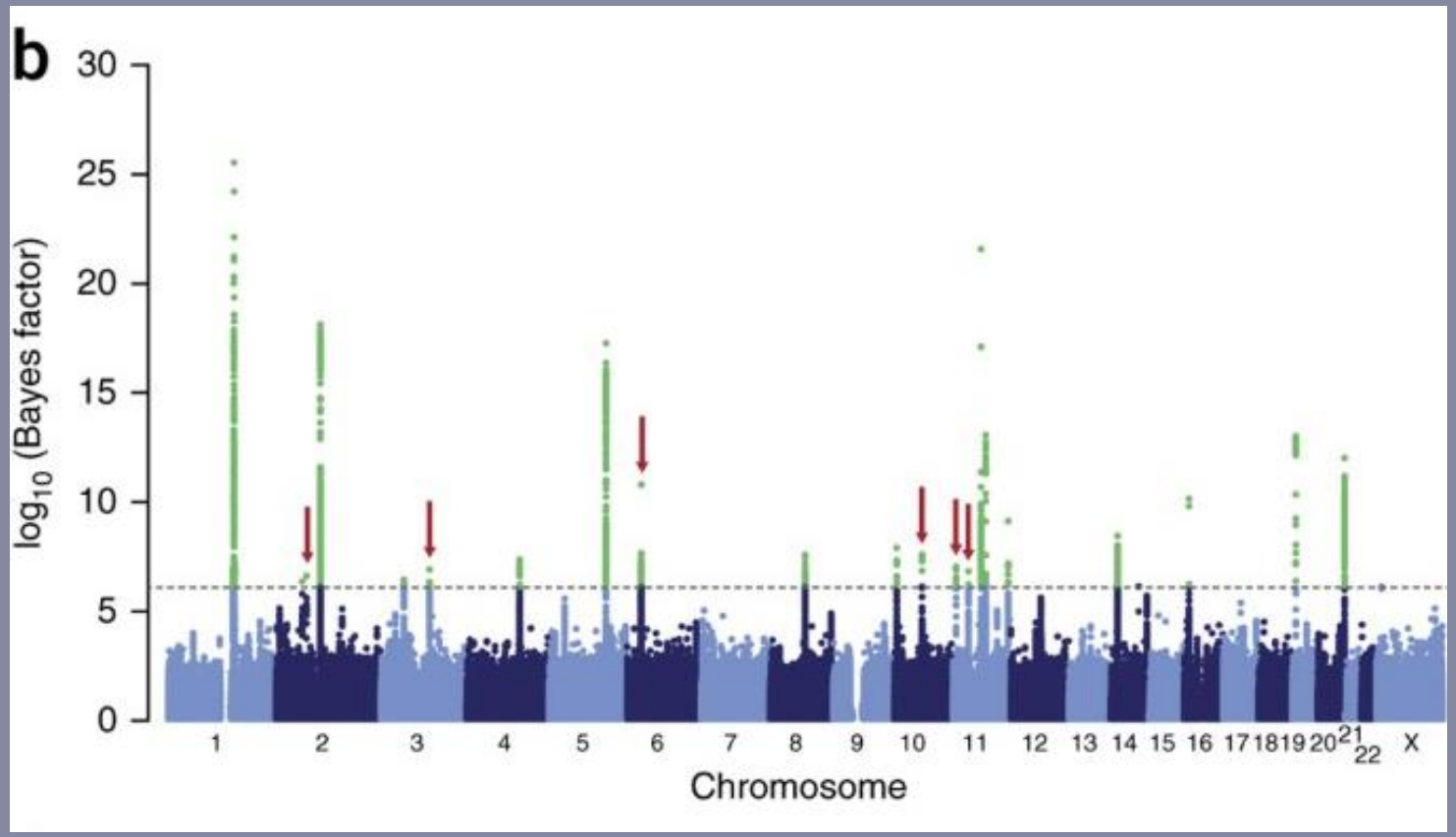
Kim et al. (2019) *Allergy Asthmas Proc*



Kim et al. (2019) *Allergy Asthmas Proc*

Effects of Cytokines on Epidermis in AD

- Disrupted epidermal barrier and environmental triggers stimulate keratinocytes to release IL-1β, IL-25, IL-33, MDC, TARC, and TSLP
- Dendritic cells and Langerhans cells are activated.
- Activated dendritic cells stimulate Th2 cells to produce IL-4, IL-5, IL-13, IL-31, and IL-33
- This causes barrier dysfunction, decreased AMP production, impaired keratinocyte differentiation, and itch symptoms.



Arrows mark variants not associated in the European-only analysis.

Paternoster et al., (2015) GWAS

- Multi-ancestry meta-analysis of 26 AD studies
- Identified 27 loci, 11 of which were new
- Identified 4 loci associated with both AD and allergies and 1 associated with asthma

Aims and Objectives

To identify AD-related genes and loci prominent across different populations, using summary statistics from Paternoster et al.,(2015)

Data Overview

From Paternoster et al., (2015) multi-ancestry GWAS

Contains 21,399 cases of AD and 95,464 controls from 26 individual studies

European: 21,399 cases, 95,464 controls	Japanese: 1,472 cases, 7,966 controls	African-American: 422 cases, 844 controls	Latin-American: 300 cases, 1592 controls	Mixed non-European: 305 cases, 896 controls

Methods Pipeline

Preprocessing and Analysis with GWASLab of Summary Statistics

Liftover of Genome from build hg19 to hg38

S-PrediXcan TWAS with AD Relevant Tissues

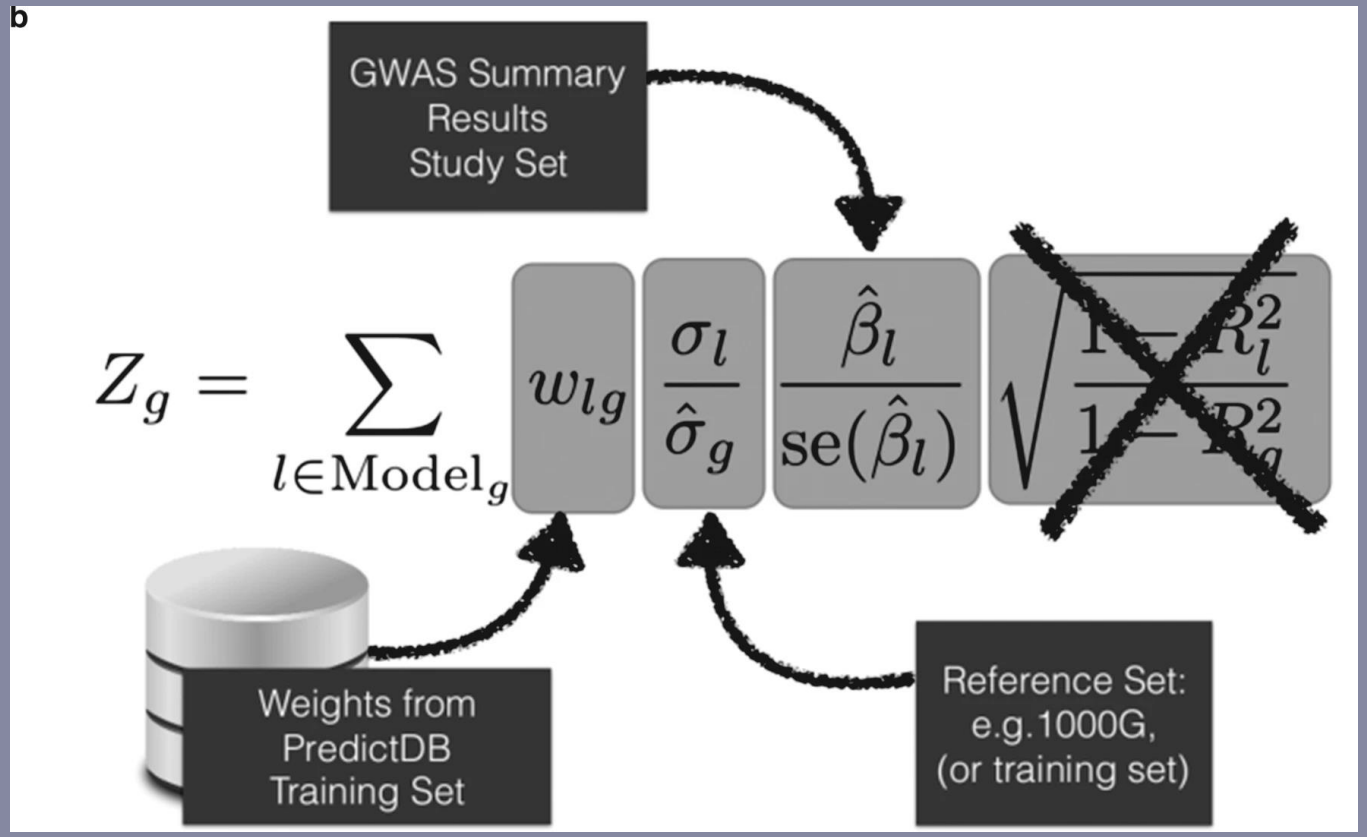
GWASLab

Python toolkit for processing and visualizing GWAS summary statistics

Allows for quality control (QC), standardization, normalization, harmonization, and data visualization

Used in this study to preprocess and clean data prior to S-PrediXcan and to visualize key features the dataset

He et al.,(2023) *Jxiv*



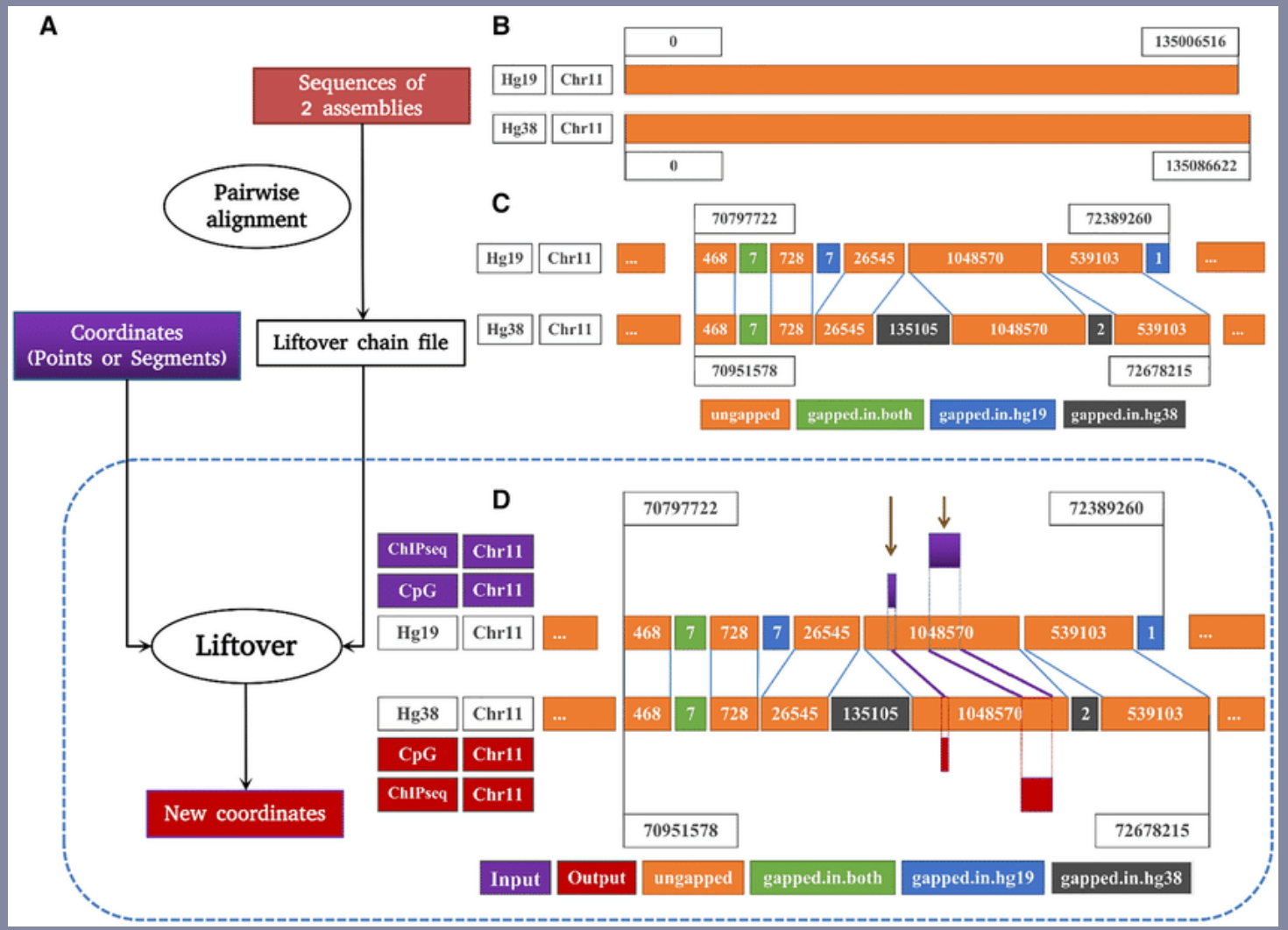
Summary-PrediXcan (S-PrediXcan)

- A version of PrediXcan that's uses GWAS summary statistics instead of individual data
- Provides tissue-specific genotype-expression models
- Tests the mediating effects of gene expression levels on phenotypes
- Uses the Wald statistic to evaluate the association between predicted gene expression and a phenotype

Barbeira et al.,(2018) *Nature Communications*

Liftover and Harmonization of Data

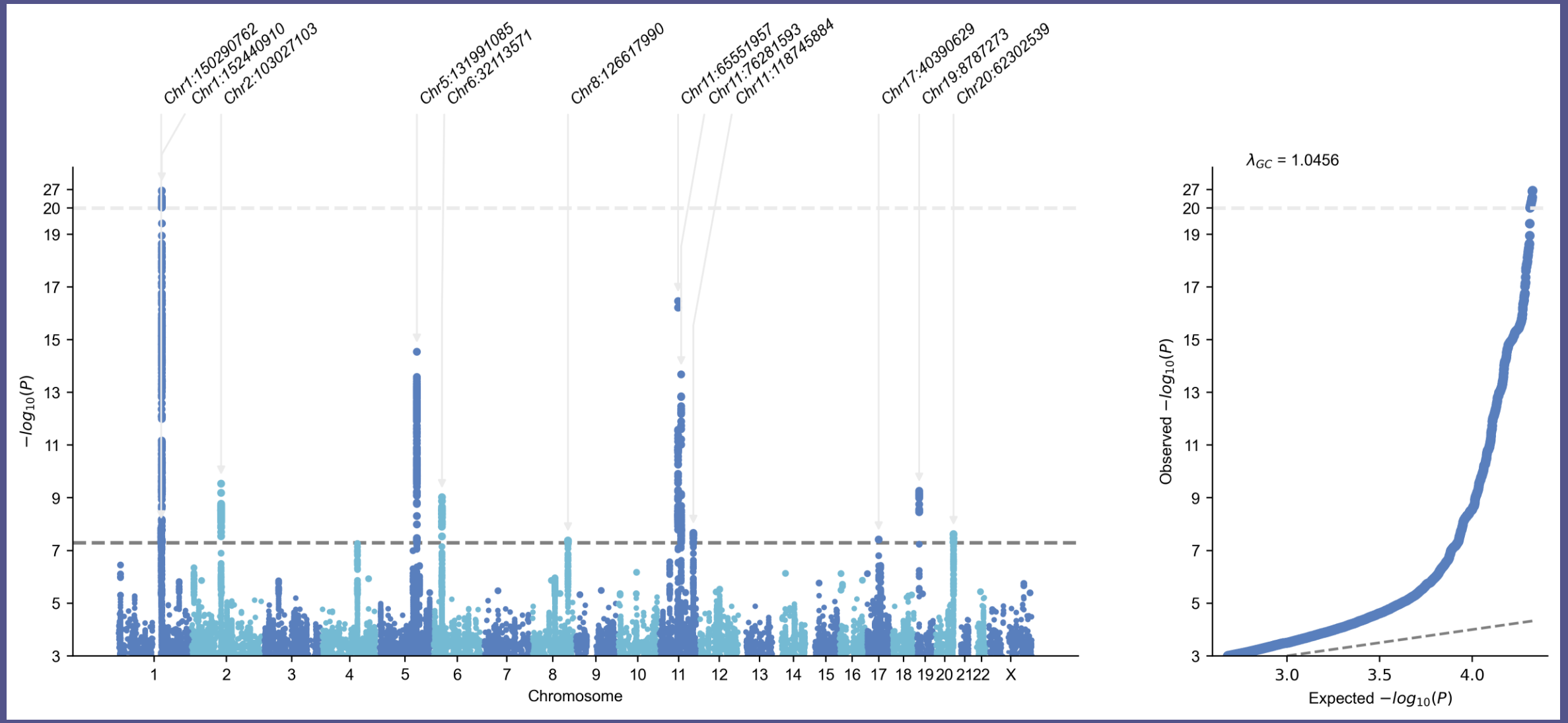
- A liftover from chromosome build hg19 to hg38 was done using GWASLab's built-in liftover function
- The data was harmonized using a hg38 SNP coordinate map to match the formatting of S-PrediXcan



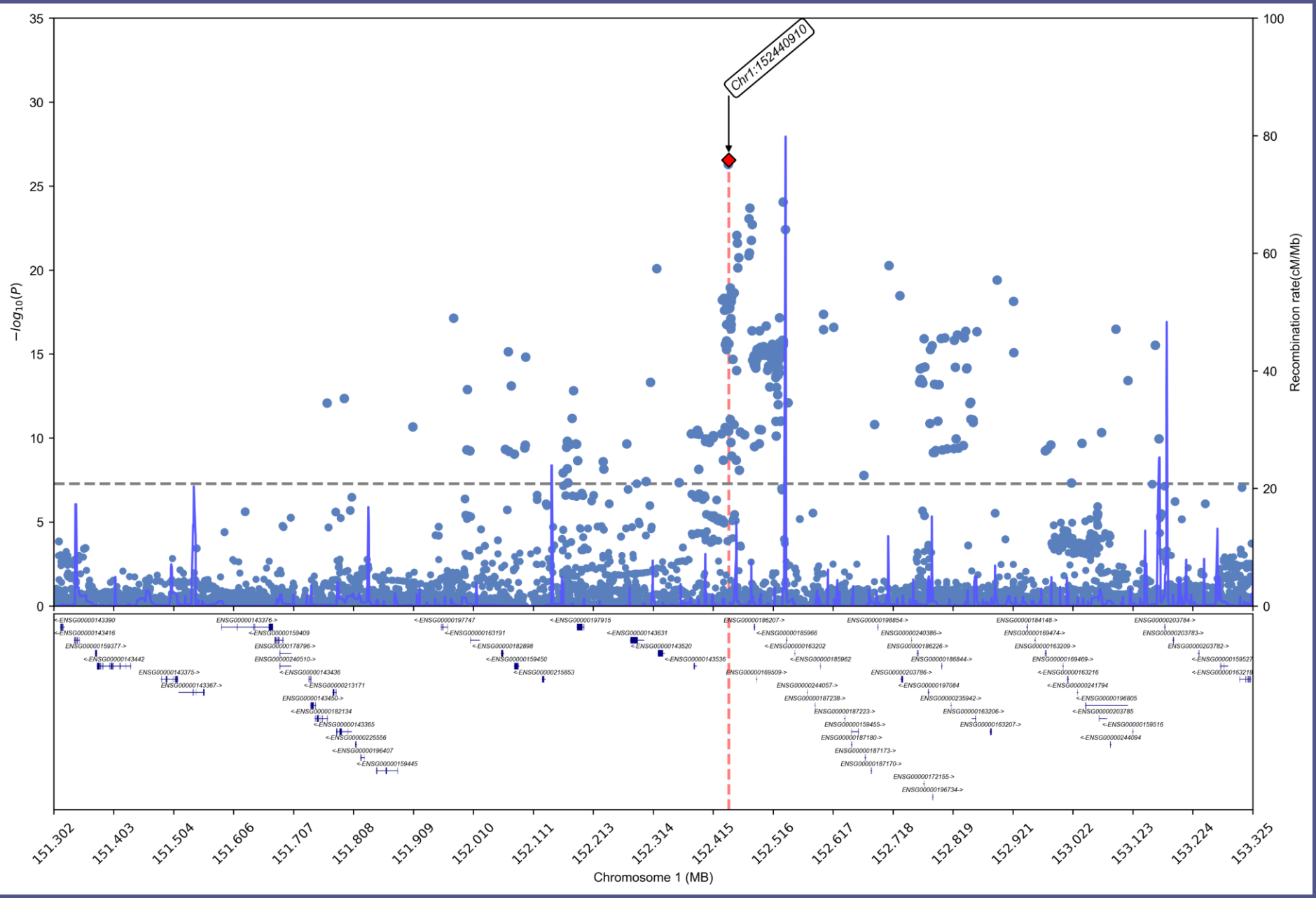
Luu et al.,(2020) *NAR Genomics and Bioinformatics*

Input for S-PrediXcan

- Tissue-specific multivariate adaptive shrinkage (MASHR) genotype-tissue expression (GTEx) V8 weights were used for imputation
- AD-relevant tissues examined were whole blood, Epstein-Barr (EBV) transformed lymphocytes, cultured fibroblasts, sun-exposed skin, and not sun-exposed skin
- The MASHR eQTL model is pre-trained, so no training set was needed
- GWAS summary statistics from Paternoster et al. (2015) were used as the study set

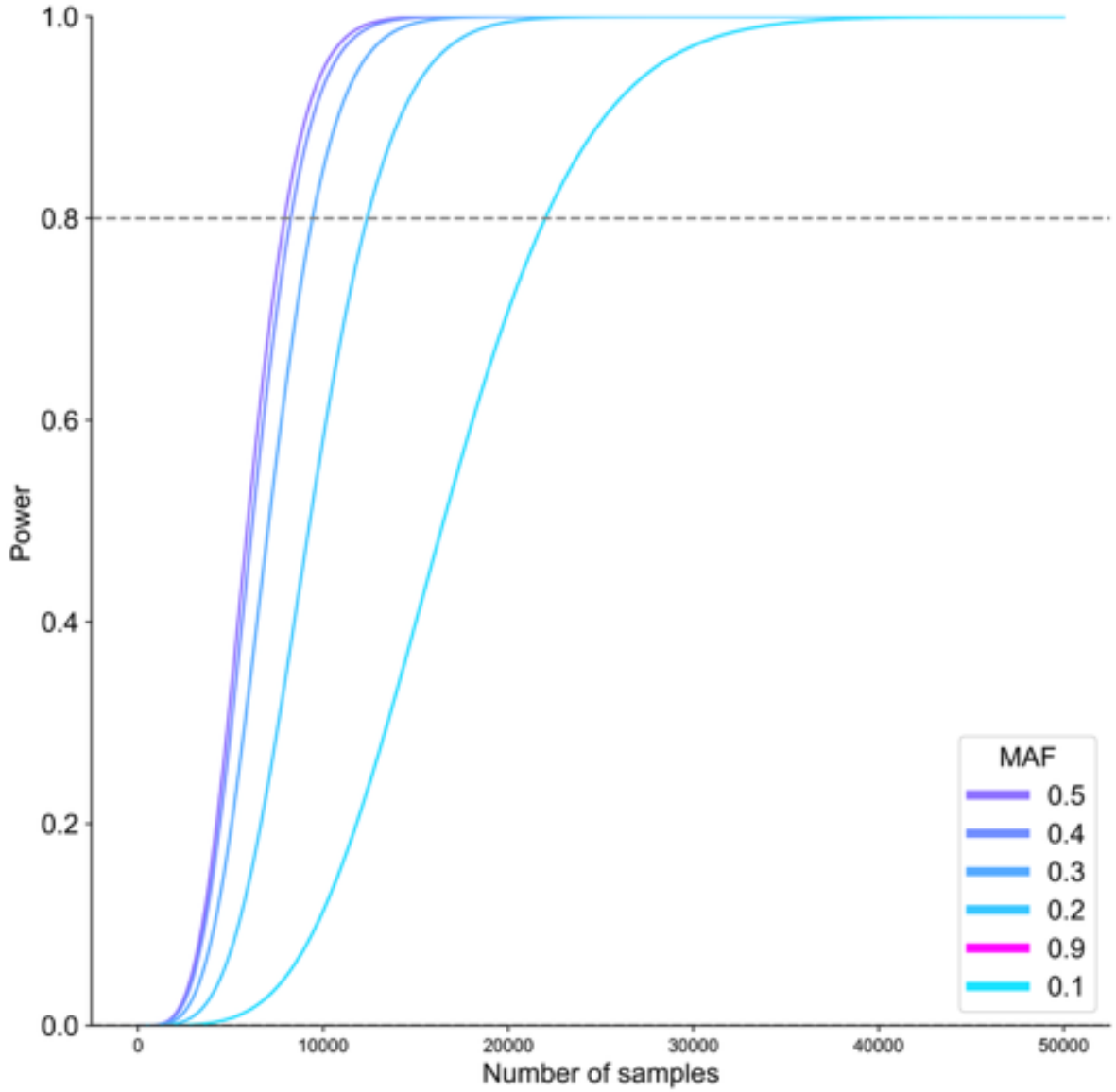


GWASLab Manhattan and Quartile-Quartile Plot



FLG Gene Regional Plot

Power Plot of GWAS Summary Statistics



rsID	CHR	POS	EA	NEA	EAF	MAF	BETA	SE	Z	P	MLOG10P	N	I2	STATUS	LOCATION	GENE
rs2477121	1	150290762	T	A	0.586709976	0.413290000	0.099648030	0.017571215	5.671094876	1.45e-8	7.838631997	54325	0.189415	1950099	3163	PRPF3
rs12144049	1	152440910	T	C	0.725919008	0.274081	-0.20180656	0.018641221	-10.8258229	2.8e-27	26.55284196	54298	0.818214	1950099	42410	LCE5A
rs6419573	2	103027103	C	T	0.759037017	0.240963000	-0.12395001	0.019650617	-6.30769047	2.92e-10	9.534617148	54326	0	1950099	8046	IL18RAP
rs12188917	5	131991085	C	T	0.205033004	0.205033	0.170064241	0.021523774	7.901227515	2.89e-15	14.53910215	54323	0.205918	1950099	870	IL13
rs116089928	6	32113571	T	A	0.206586003	0.206586	-0.14119612	0.023064852	-6.12170065	9.53e-10	9.020907099	44763	0	1950099	2565	PRRT1
rs12334935	8	126617990	A	G	0.477744996	0.477745	0.092611085	0.016877628	5.487209494	4.18e-8	7.378823718	54324	0.003426	1950099	-167343	TRIB1
rs479844	11	65551957	G	A	0.566488027	0.433512	0.143747034	0.017038418	8.436642030	3.45e-17	16.46218090	54328	0.271602	1950099	2536	OVOL1
rs2212434	11	76281593	T	C	0.450704008	0.450704	0.129132607	0.016879171	7.650411399	2.09e-14	13.67985371	54326	0.545882	1950099	-17524	C11orf30
rs10790275	11	118745884	C	G	0.808486998	0.191513000	0.122434423	0.021852897	5.602663132	2.16e-8	7.665546248	54614	0	1950099	8591	CXCR5
rs8066625	17	40390629	A	G	0.107211999	0.107212	0.175584748	0.031911994	5.502155323	3.84e-8	7.415668775	54316	0.241792	1950099	0	STAT5B
rs2918299	19	8787273	T	C	0.166014000	0.166014	0.142564966	0.022955116	6.210596591	5.45e-10	9.263603497	52271	0.126049	1950099	20478	ACTL9
rs6062486	20	62302539	A	G	0.687358975	0.312640999	0.104572605	0.018726006	5.584351533	2.4e-8	7.619788758	54327	0.001764	1950099	0	RTEL1,RTEL1

12 Lead Variants (p<5e-8) Identified by GWASLab

GWASLab Lead Variants I

6 of the variants identified were not identified by the original GWAS

PRPF3 is a protein-coding gene responsible for m-RNA splicing and is associated with retinitis pigmentosa

STAT5B encodes a protein that is stimulated by cytokines and growth factors

GWASLab Lead Variants II

LCE5A is predicted to be involved in keratinization and enables identical protein binding activity

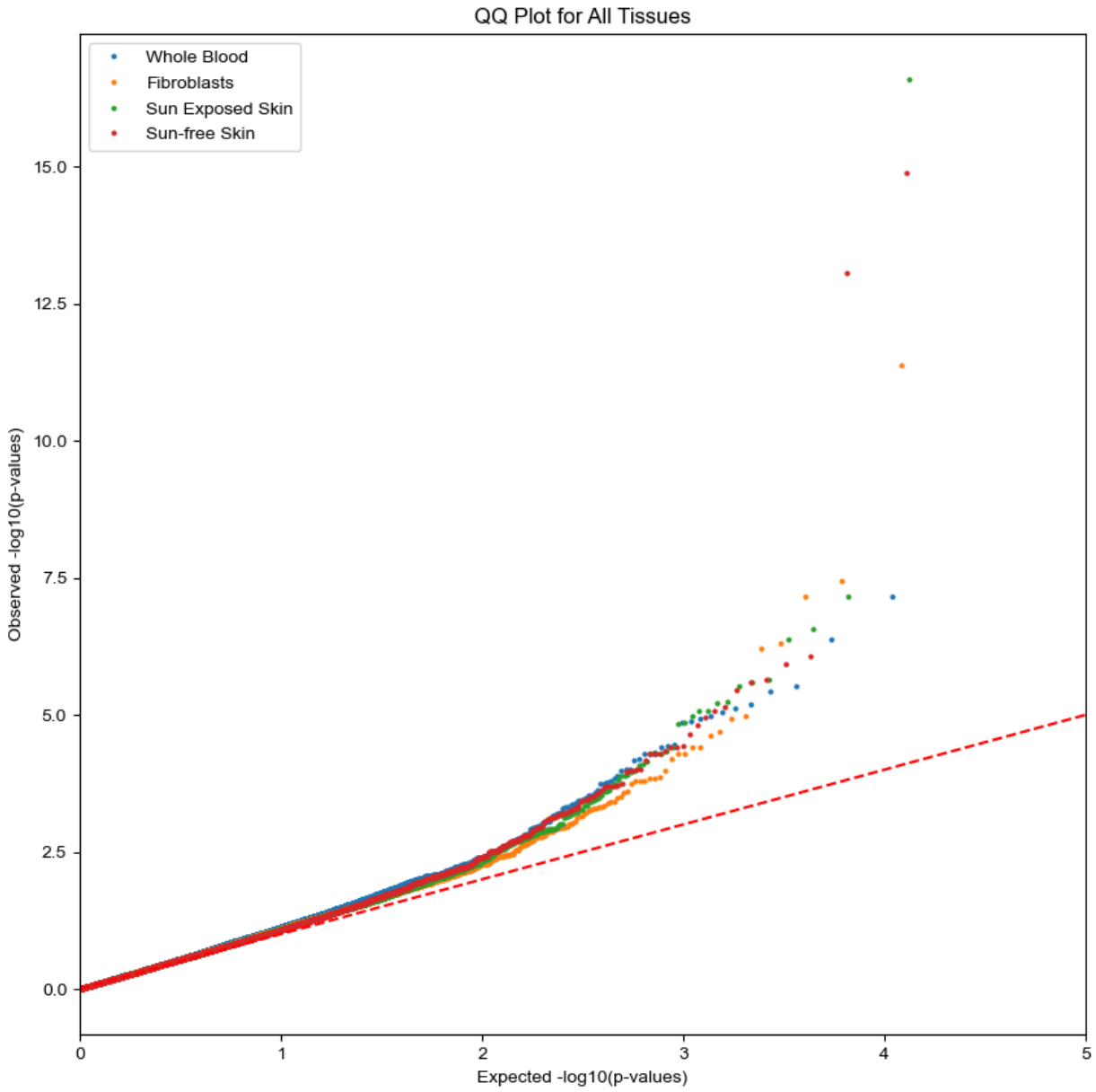
PRRT1 also enables identical protein binding activity

TRIB1 is involved in protein kinase kinase binding activity and inhibition

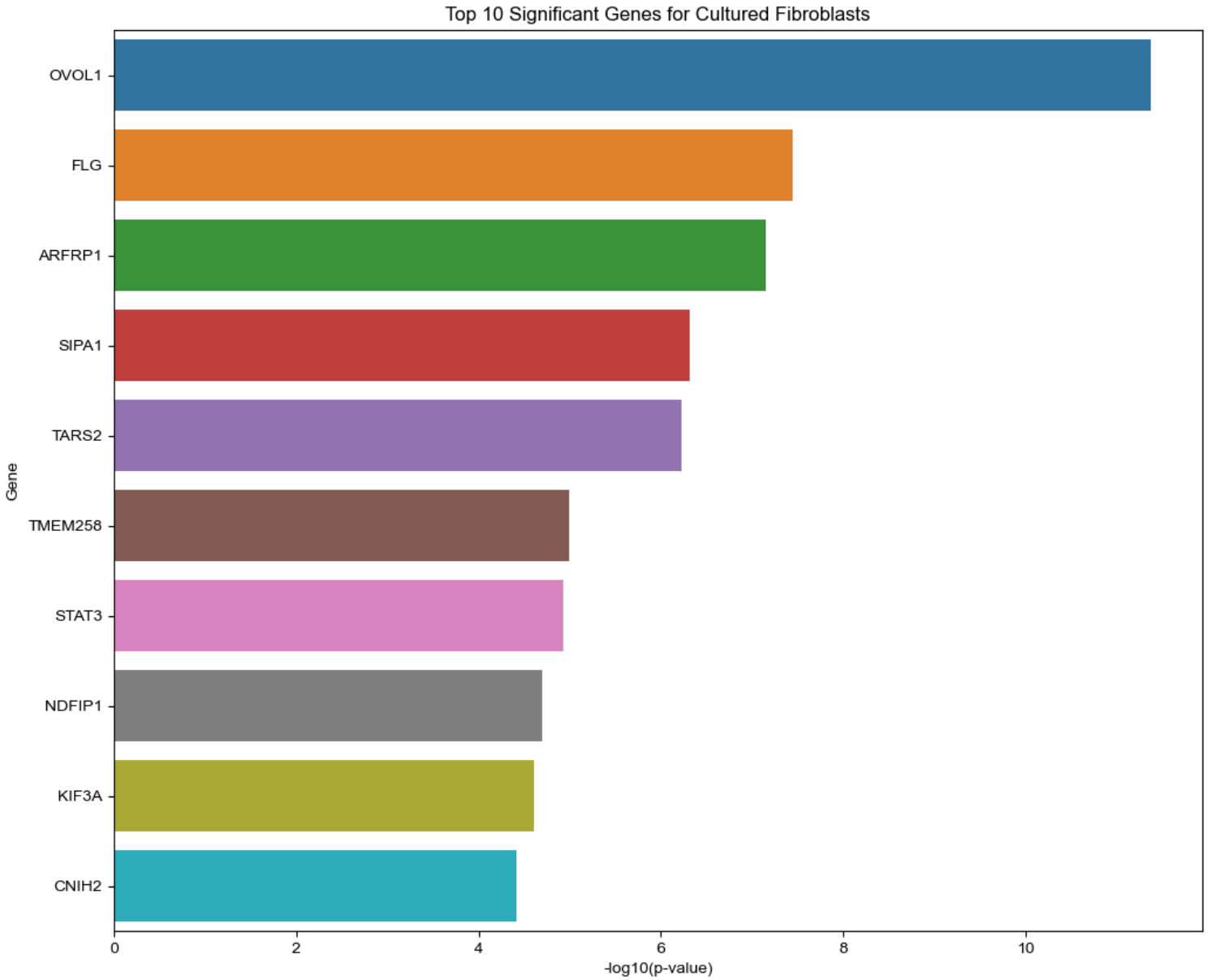
CXCR5 encodes a membrane protein that is a chemokine receptor that is involved in B-cell migration

S-PrediXcan EBV-Transformed Lymphocytes Results

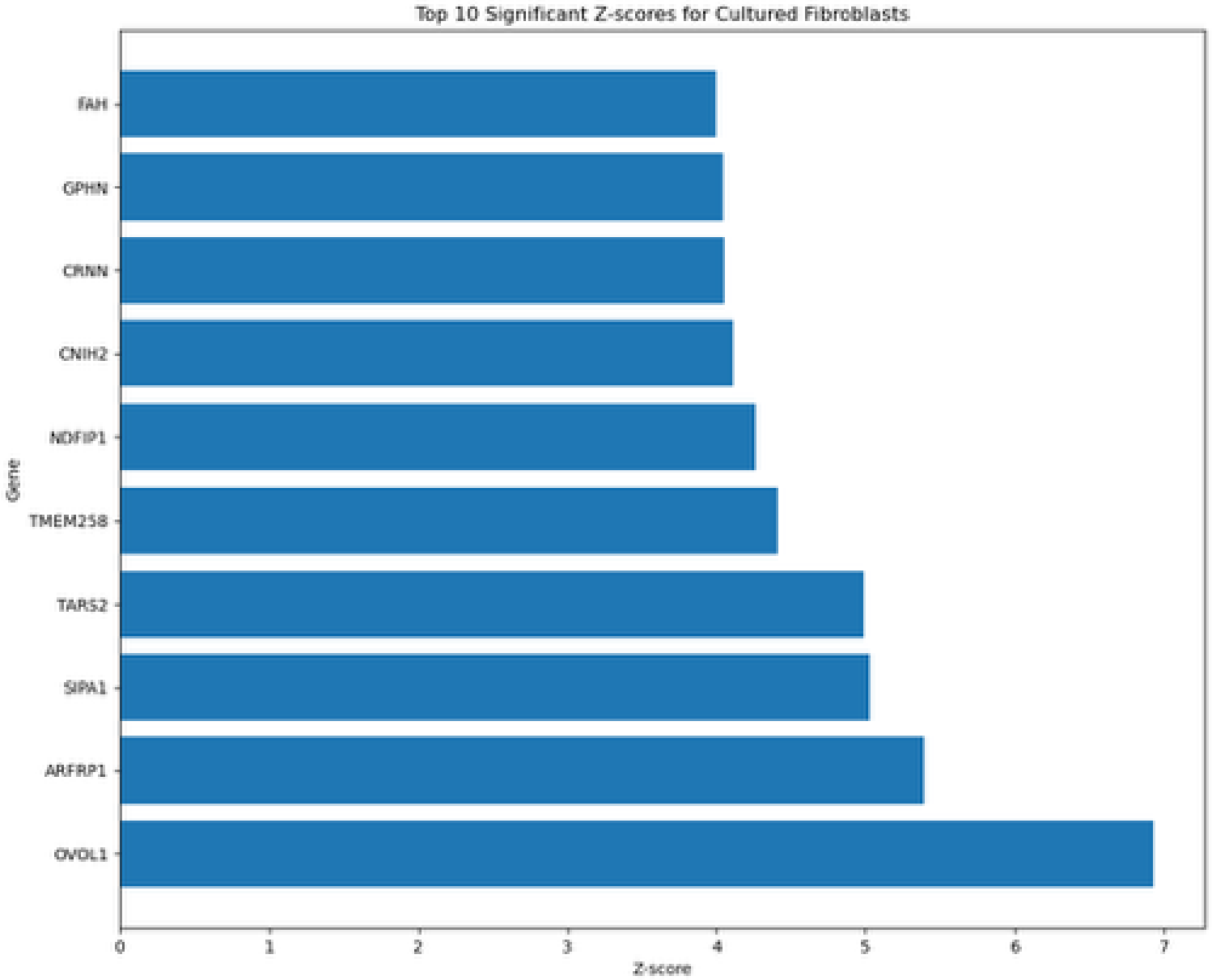
- No significant associations were found between the GTEx v8 EBV-transformed lymphocytes and the GWAS summary statistics



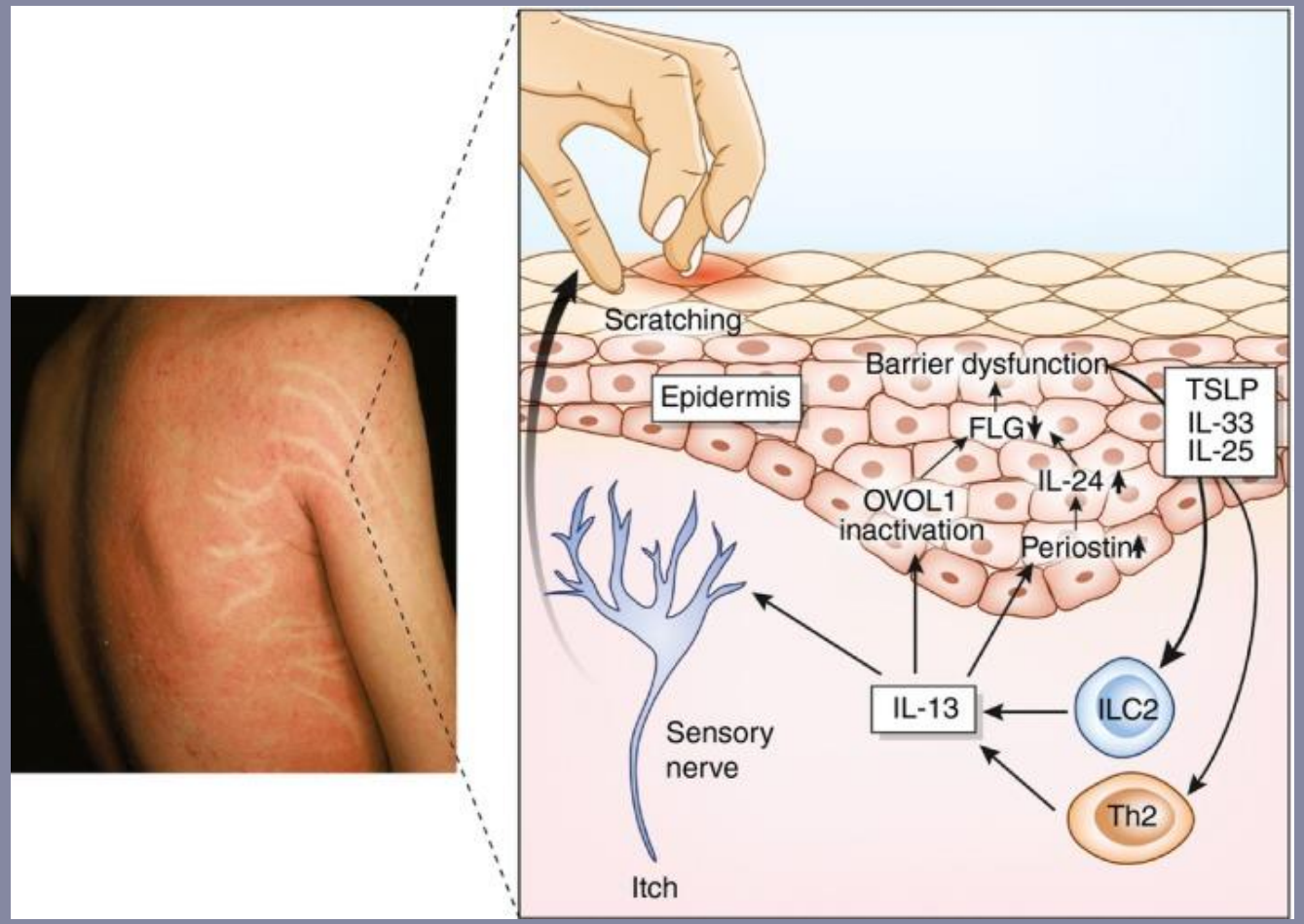
Q-Q Plot for 4 Tissues with Associations



S-PrediXcan Cultured Fibroblasts Results



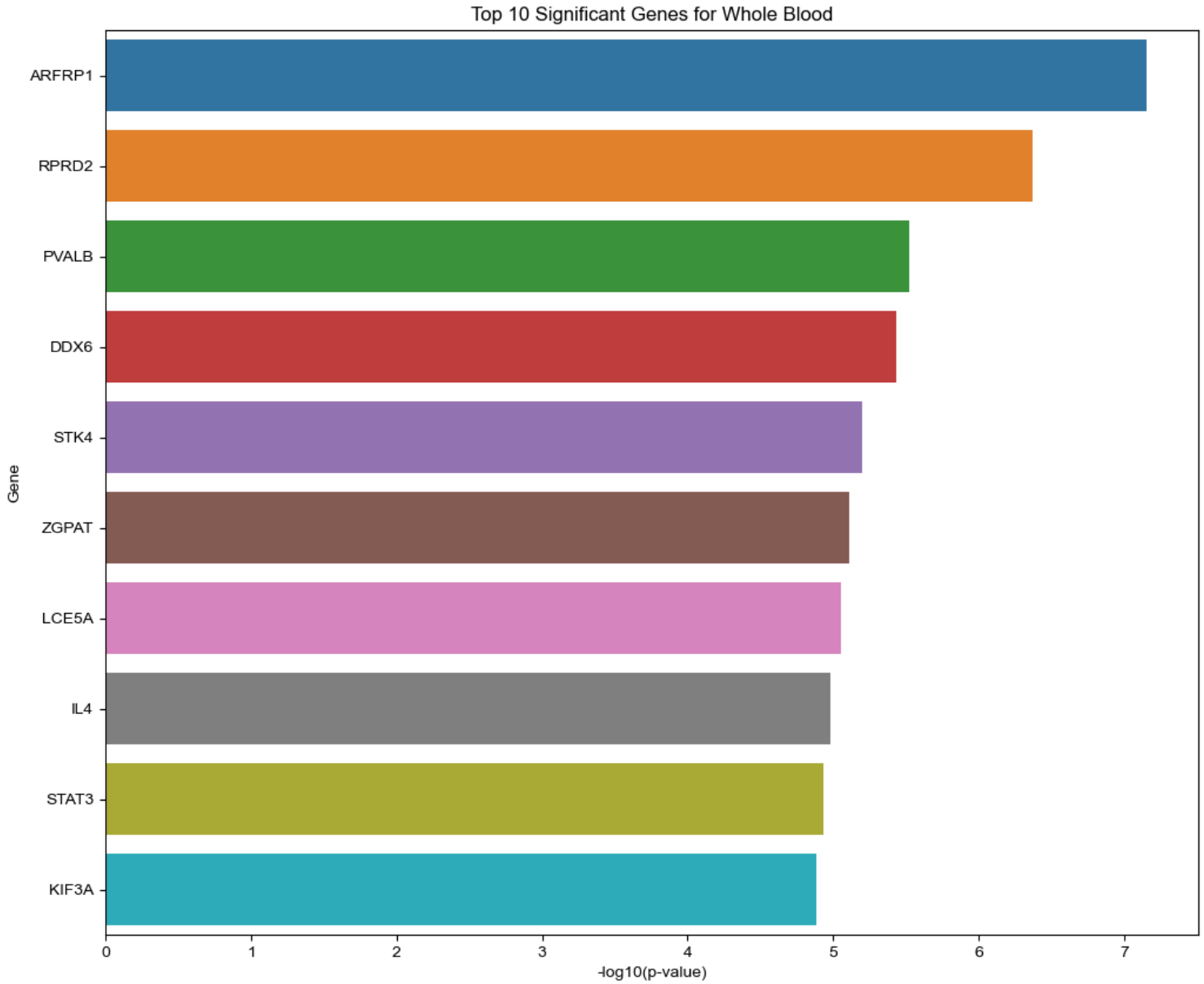
Top 10 Significant Gene Associations for Cultured Fibroblasts by Z-score



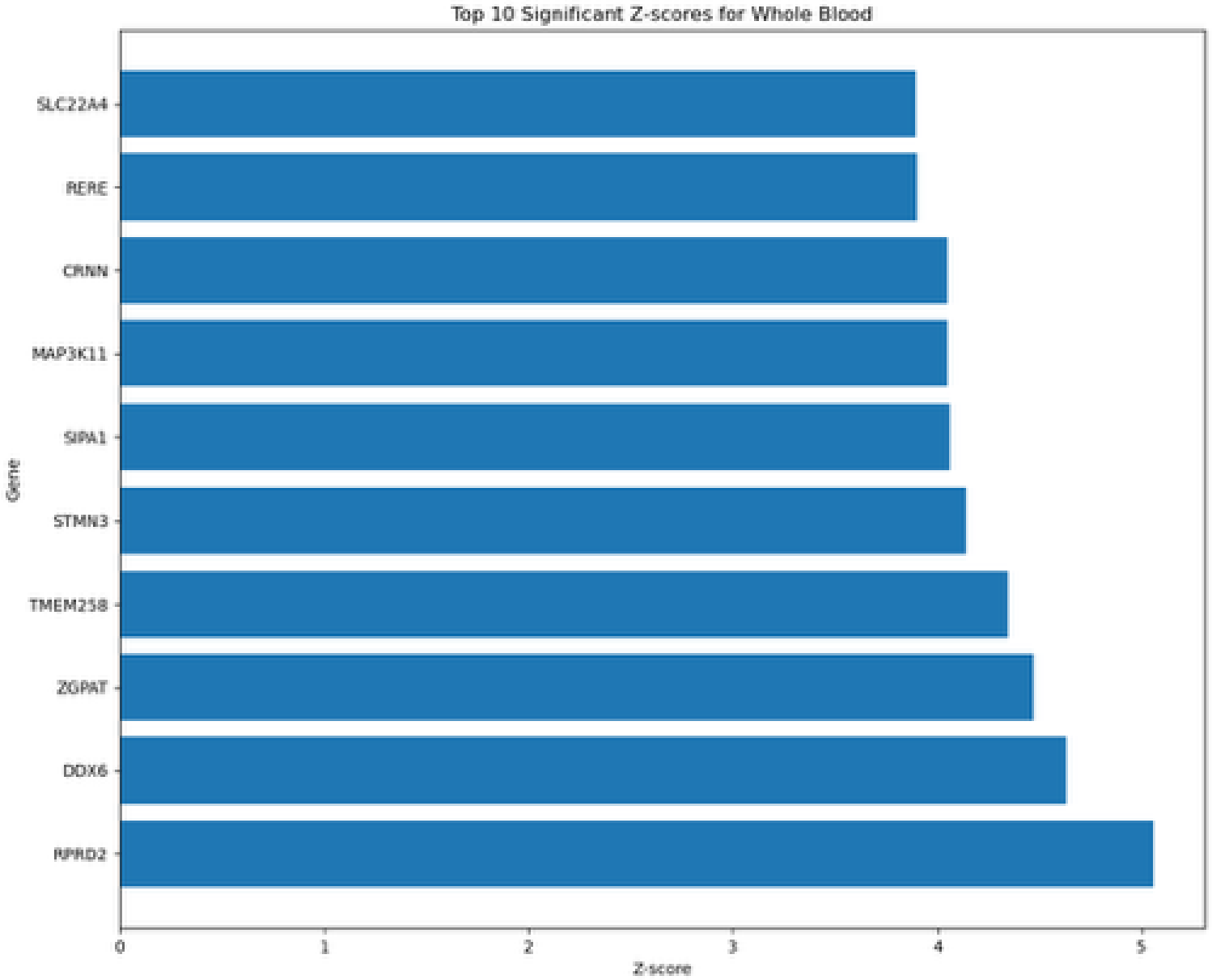
Furue et al.,(2019) *Immunology*

OVOL1

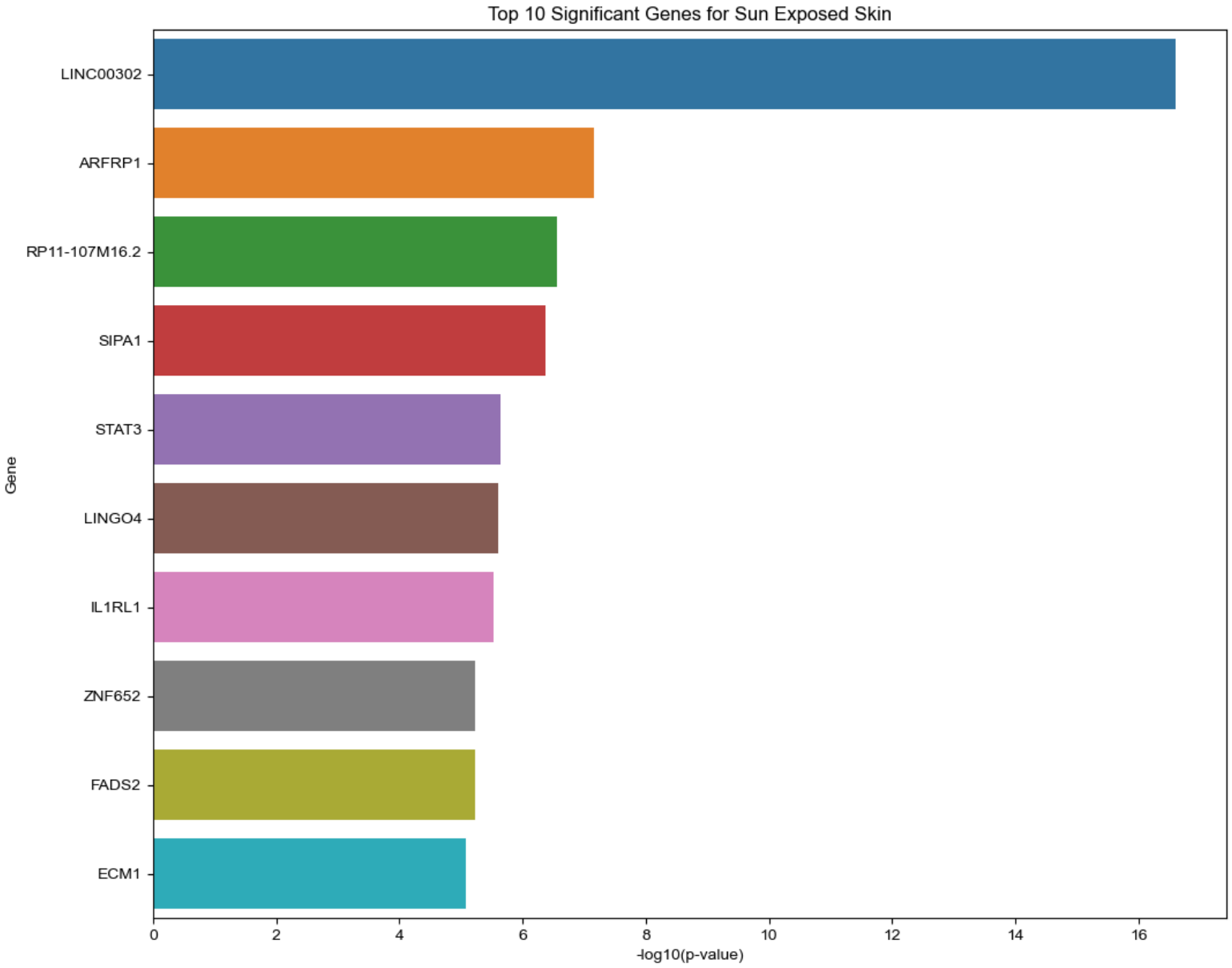
- *R*egulates cell differentiation and proliferation in the skin
- Its role is closely related to *IL-13* and *FLG*
- High amounts of *IL-13* cause *OVOL1* inactivation and upregulation of *IL-24*
- Causes down-regulation of *FLG* and subsequent barrier dysfunction



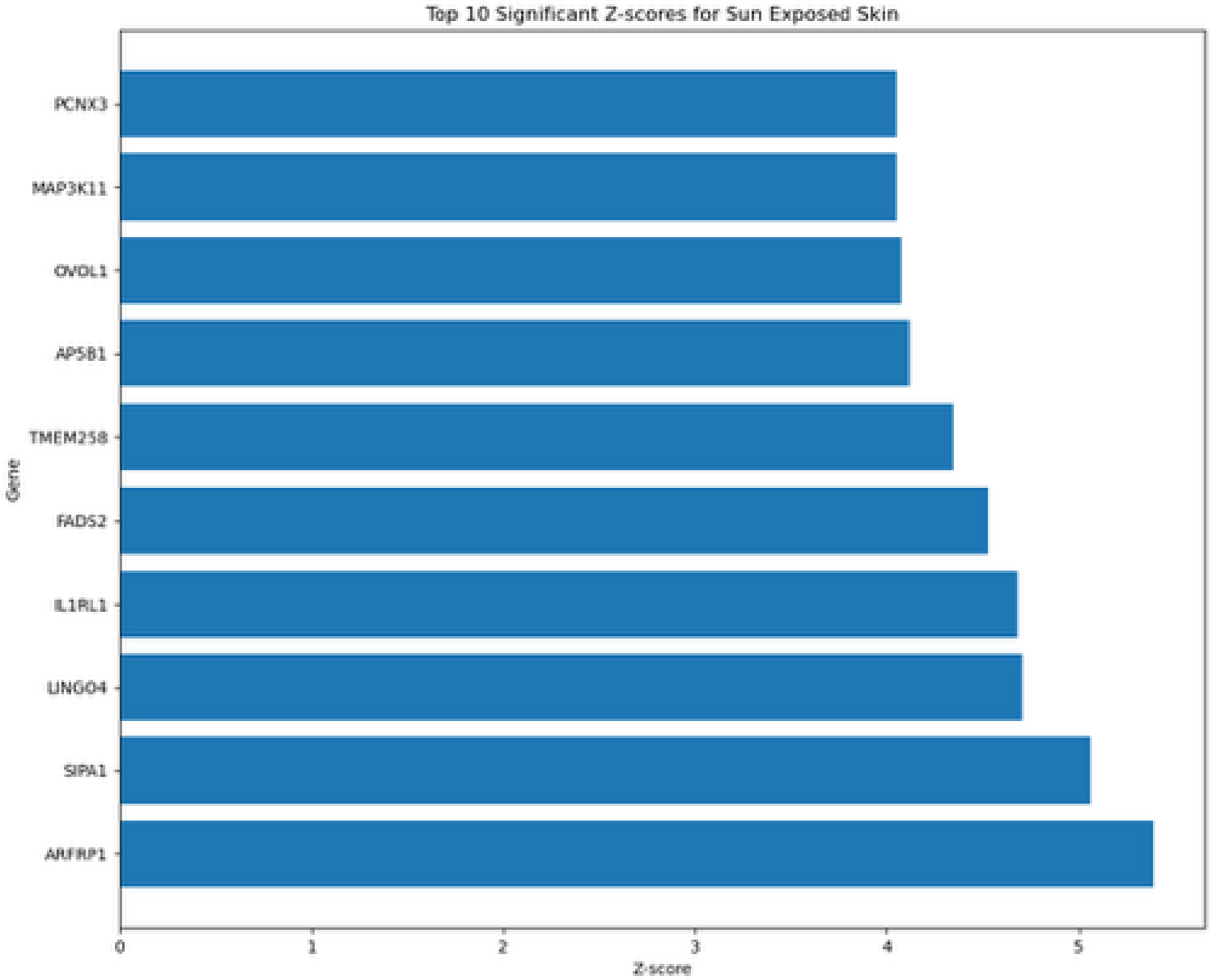
S-PrediXcan Whole Blood Results



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



S-PrediXcan Sun Exposed Skin Results

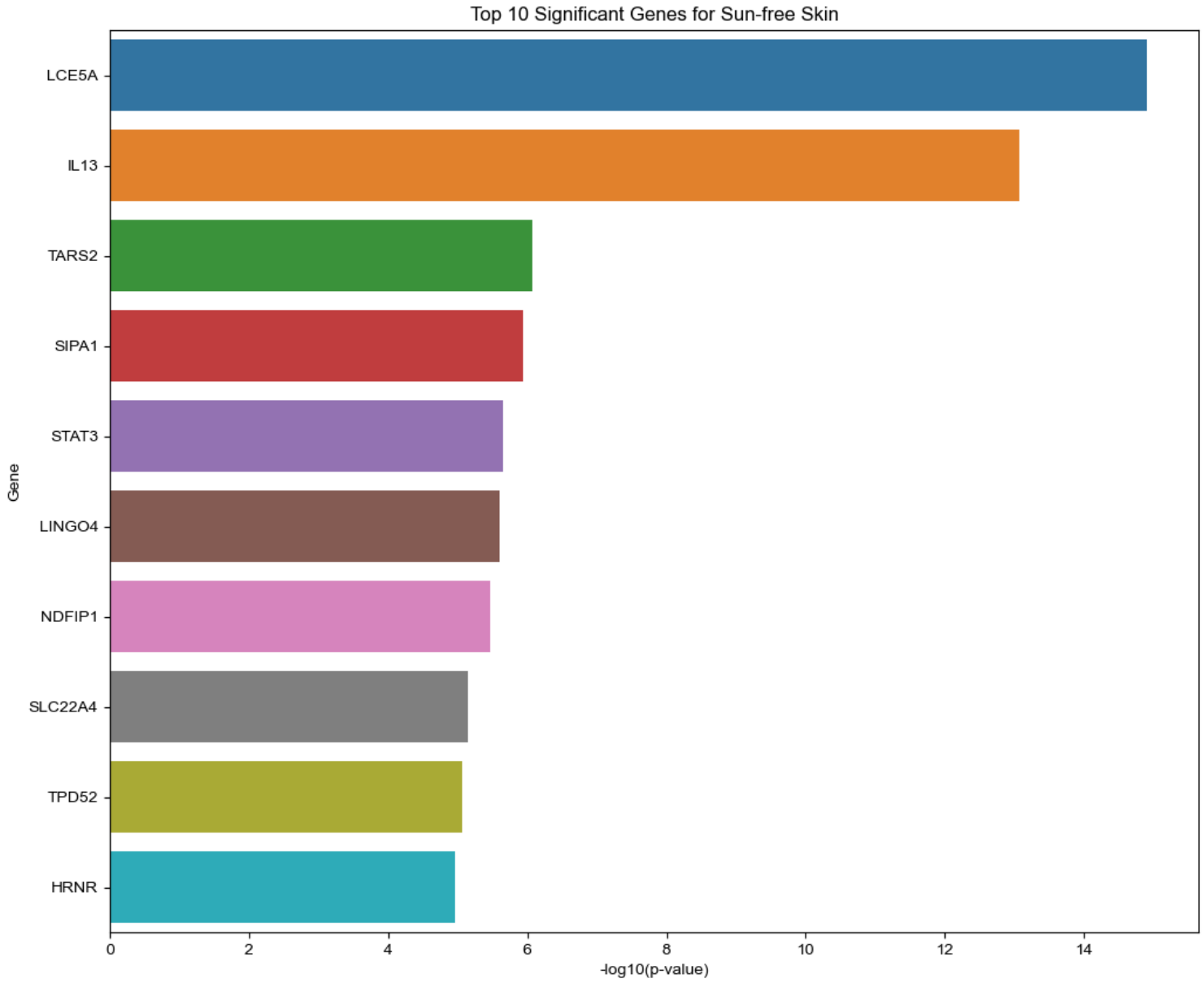


Top 10 Significant Gene Associations for Sun-Exposed Skin by Z-score

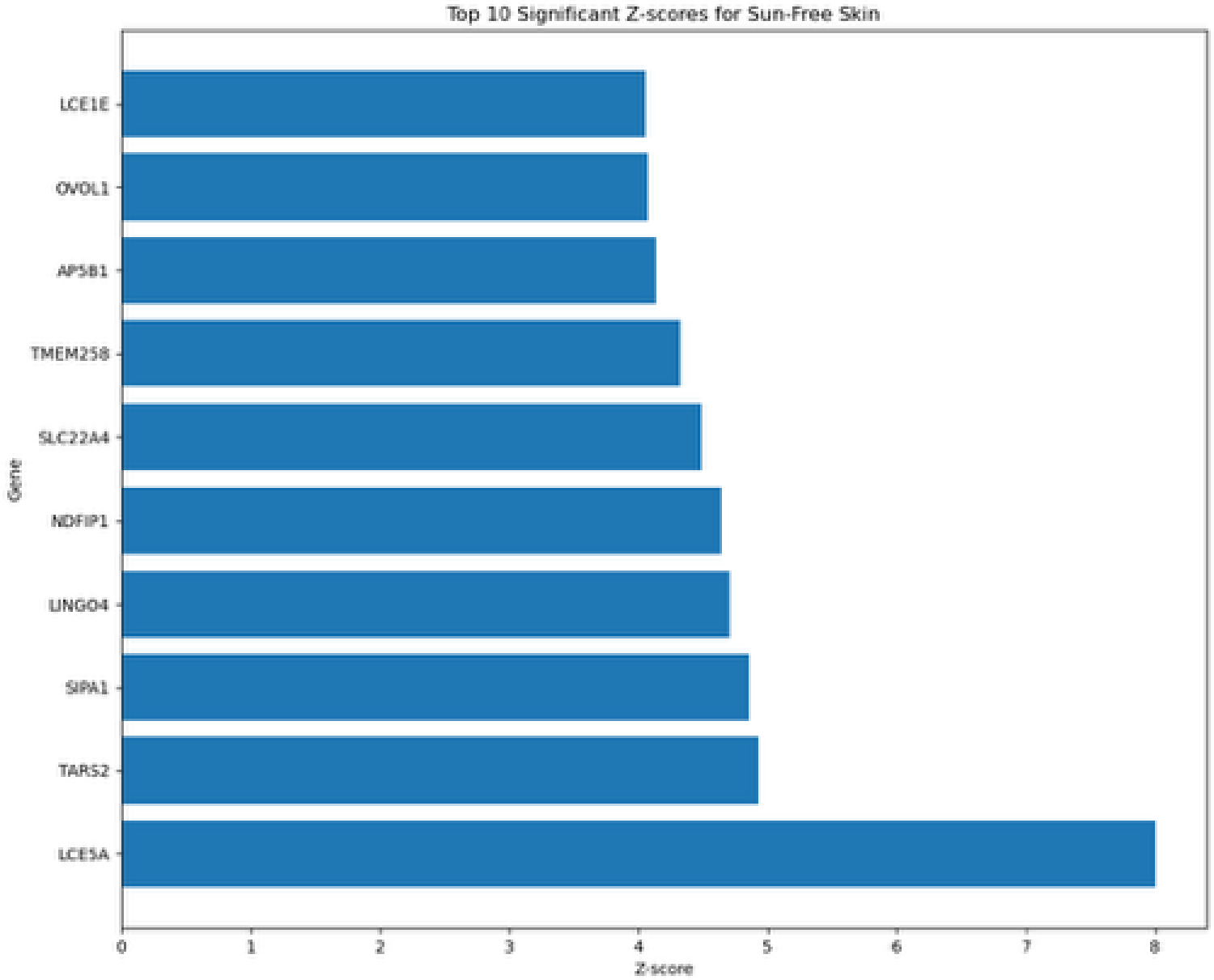
LINC00302

- Non-protein coding RNA that is expressed almost exclusively in the skin
- It is involved in keratinocyte differentiation and pathogenesis
  - The exact role is unknown
- Is down-regulated in AD lesions
- Knockdown in primary keratinocytes affects the expression of *FLG* and other related genes

Sahlen et al.,(2021) *J Allergy Clin Immunol*



S-PrediXcan Sun-free Skin Results



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

LCE5A

LCE5A is involved in the epidermal differentiation complex of the skin barrier

Allows allergens to penetrate deeper skin layers easier

It is a loss of function variant of *FLG*

Reduces skin hydration

Chong et al.,(2022) *J Asthma Allergy*

IL-13

- Crucial cytokine involved in AD
- Overexpressed in AD lesions
- Recruits inflammatory cells, modifies skin microbiome, and alters skin barrier
- Also activates the sensory nerve responsible for the itch sensation
- Aside from *FLG*, this is one of the most studied genes related to AD

Napolitano.,(2023) *Frontiers in Medicine*

Conclusions I

GWASLab identified 12 lead variants	• *PRPF3, LCE5A, IL18RAP, IL13, PRRT1, TRIB1, OVOL1, C11orf30, CXCR5, STAT5B, ACTL9*, and *RTEL1*
3 of the 4 lead variants found using S-PrediXcan were also found by GWASLab	• *OVOL1, LCE5A,* and *IL-13*
LINC00302 (also known as *XP33*) was identified as a lead variant in the sun-exposed skin tissue	• Was not identified by the GWASLab analysis
This study identified 7 genes not identified by the original Paternoster et al.,(2015) GWAS	• *LINC00302, PRPF3, LCE5A, PRRT1, TRIBI, OXCR5*, and *STAT5B*

EBV-transformed lymphocytes were not a relevant tissue for this AD study

All lead variants identified by this study have since been implicated in other AD research

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)

Conclusions II

Limitations and Next Steps

- The multi-ancestry cohorts were made up of predominately European samples, more samples of other ancestries are needed
- Fine-mapping of regions containing lead variants should be done to further evaluate the gene-trait relationship
- The study evaluated AD-risk genes in multiple populations but did not identify ancestry-specific genes
- Multi-ancestry fine-mapping, such as the MA-FOCUS method can be done to determine ancestry-specific genes relevant to AD
- Colocalization can be performed to evaluate causal interference in the study

References

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

Bhattacharya, A., Li, Y., & Love, M. I. (2021). Mostwas: Multi-omic strategies for transcriptome-wide association studies. *PLOS Genetics*, *17*(3) https://doi.org/10.1371/journal.pgen.1009398

Bhattacharya, A., Hirbo, J. B., Zhou, D., Zhou, W., Zheng, J., Kanai, M., Pasaniuc, B., Gamazon, E. R., & Cox, N. J. (2022). Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: Lessons from the global biobank meta-analysis initiative. *Cell Genomics*, *2*(10), 100180. https://doi.org/10.1016/j.xgen.2022.100180

Budu-Aggrey , 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).

References

Chong, A. C., Visitsunthorn, K., & Ong, P. Y. (2022). Genetic/environmental contributions and immune dysregulation in children with atopic dermatitis. *Journal of Asthma and Allergy*, *Volume 15*, 1681–1700. https://doi.org/10.2147/jaa.s293900

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), 281–286.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

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/

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 ranscriptome-wide association studies. *The American Journal of Human Genetics*, *109*(8), 1388 1404.https://doi.org/10.1016/j.ajhg.2022.07.002

References

Luu, P.-L., Ong, P.-T., Dinh, T.-P., & Clark, S. J. (2020). Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. *NAR Genomics and Bioinformatics*, *2*(3). https://doi.org/10.1093/nargab/lqaa054

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

Napolitano, M., di Vico, F., Ruggiero, A., Fabbrocini, G., & Patruno, C. (2023). The hidden sentinel of the skin: An overview on the role of interleukin-13 in atopic dermatitis. *Frontiers in medicine*, *10*, 1165098. https://doi.org/10.3389/fmed.2023.1165098

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

Peng, W., & Novak, N. (2015). Pathogenesis of atopic dermatitis. *Clinical & Experimental Allergy*, *45*(3), 566–574. https://doi.org/10.1111/cea.12495

Sahlén, P., Spalinskas, R., Asad, S., Mahapatra, K. D., Höjer, P., Anil, A., Eisfeldt, J., Srivastava, A., Nikamo, P., Mukherjee, A., Kim, K.-H., Bergman, O., Ståhle, M., Sonkoly, E., Pivarcsi, A., Wahlgren, C. F.,Nordenskjöld, M., Taylan, F., Bradley, M., & Tapia-Páez, I. (2021). Chromatin interactions in differentiating keratinocytes reveal novel atopic dermatitis– and psoriasis-associated genes. *Journal of Allergy and Clinical Immunology*, *147*(5), 1742–1752. https://doi.org/10.1016/j.jaci.2020.09.035

U.S. National Library of Medicine. (2008, April 21). *Home - gene - NCBI*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene

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