Gene enrichment analysis of WGS patients to identify rare and deleterious genetic determinants of eczema herpeticum

Claire Malley[1], Meher Boorgula[2], Sameer Chavan[2], Nicholas Rafaels[2], Malaika Mathias[1], Kathleen Barnes[2], and Rasika A. Mathias[1]

On behalf of the Atopic Dermatitis Research Network

1. Johns Hopkins University, Division of Allergy & Asthma, Baltimore, Maryland, USA.
2. Translational Informatics and Computational Resource (TICR), Colorado Center for Personalized Medicine (CCPM), Denver, CO, USA.

ABSTRACT

❖ Current understanding of the genetic loci and pathways implicated in the occurrence **atopic dermatitis (AD)** is limited, and particularly so for the rarer condition of recurrent **eczema herpeticum (EH)** in AD.

❖Through the NIAID sponsored Atopic Dermatitis Research Network (ADRN), we used high-throughput whole genome sequencing data to bridge these gaps in knowledge.

❖We identified SNPs and indels with predicted damaging sequence annotations.

STUDY AIM

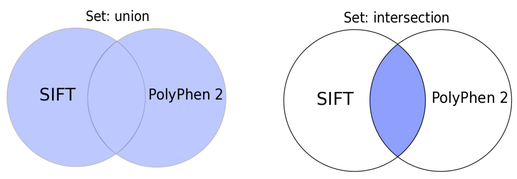
To identify candidate gene mutations and pathways implicated in atopic dermatitis and eczema herpeticum

METHODS

❖The ADRN project sampled the clinical traits and sequencing data of individuals with atopic dermatitis and from a non-atopic control group (NA, see Table 1).

❖Genotyping of all samples was performed on Illumina's OMNI 2.5 million SNP array, and whole-genome sequencing was performed on the HiSeq platform.

❖A custom pipeline screened for exonic SNPs with functional annotations that were frameshift, nonsynonymous, missense, stopgain, or stoploss. We required a damaging prediction from either (union) or both (intersection) of the SIFT and PolyPhen 2 algorithms.



❖We compared carrier status for all damaging variants versus those with less than or equal to 5% global frequency, or missing, in the 1000 Genomes Project database (uncommon).

❖Carrier status for uncommon indels with exonic annotations was separately analyzed and compared between the groups.

❖We ran Platypus variant caller on the aligned BAM files for EH individuals and found no additional long indels missed by the Illumina Isaac variant caller.

❖The remaining subset lists of variants were input to the Gene Ontology PANTHER database to prioritize genes by statistical overrepresentation.

RESULTS

Table 1. Sample Characteristics

Trait	EH	AD	NA	AD vs. NA 2 sample t-test	AD vs. EH 2 sample t-test
N	49	492	239	--	--
Male identified, N, % of total	23 (47%)	222 (45%)	74 (31%)	3.5E-4	0.8106
Age, mean, SD	21.0 (15.7)	28.0 (18.4)	37.5 (15.7)	3.6E-14	4.7E-3
Total IgE, geometric mean, 95% CI	1033 (1927-4905)	267 (220-325)	12.2 (10.7-13.9)	< 2.2E-16*	1.8E-5*
Eosinophils, geometric mean, 95% CI	345 (387-598)	20.5 (17.3-24.2)	99 (91-109)	< 2.2E-16*	1.1E-2*
Phadiotop, geometric mean, 95% CI	42.2 (41-405)	20.5 (17.3-24.2)	--	--	2.0E-3*
EASI, mean, range	15.4 (0-55)	14.5 (0.3-66.8)	--	--	0.6803
Raika-Langeland score, mean, SD	6.9 (1.7)	6.9 (1.5)	--	--	0.9717

* log10 transformed values

Figure 1. Variation in SNP annotations in EH

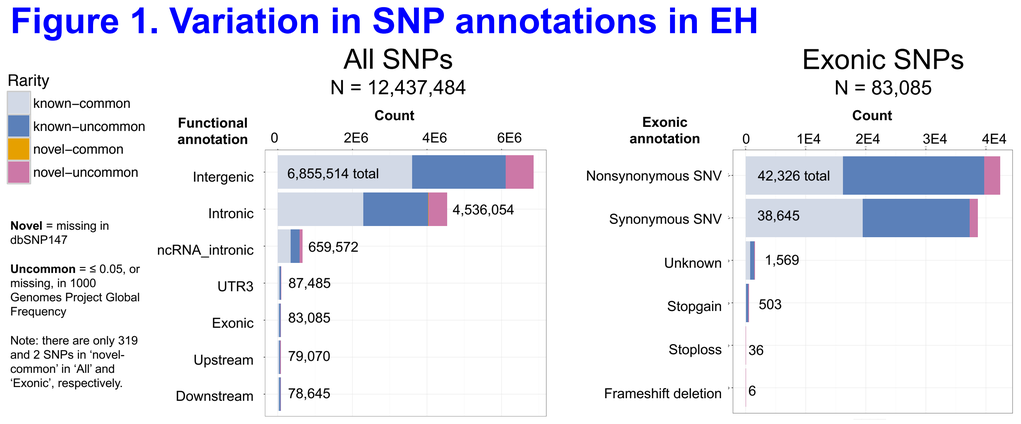
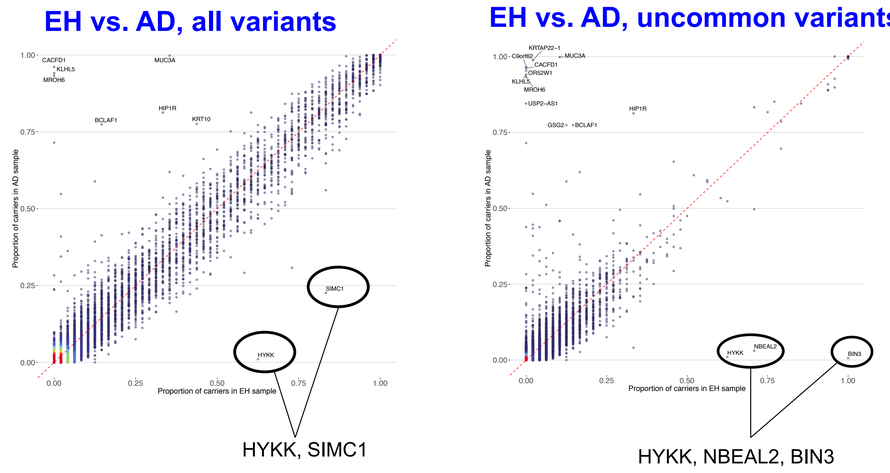


Figure 2. Carrier status of probably damaging SNPs compared between EH and AD

A. Union of SIFT and PolyPhen 2 annotations

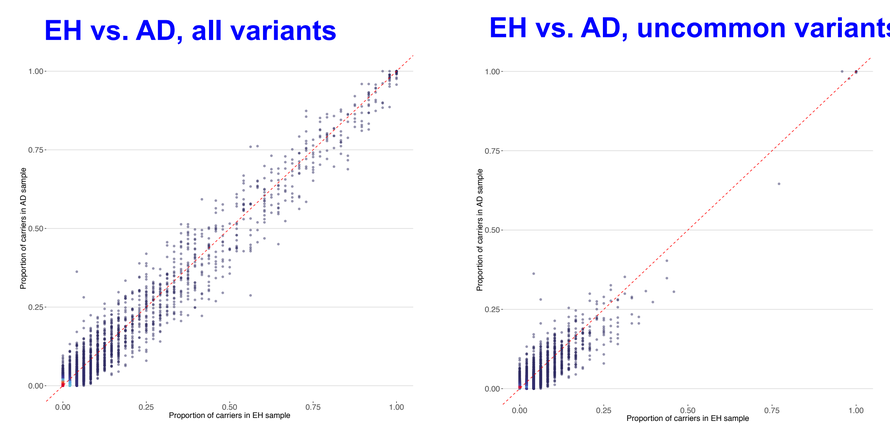


HYKK, SIMC1

HYKK, NBEAL2, BIN3

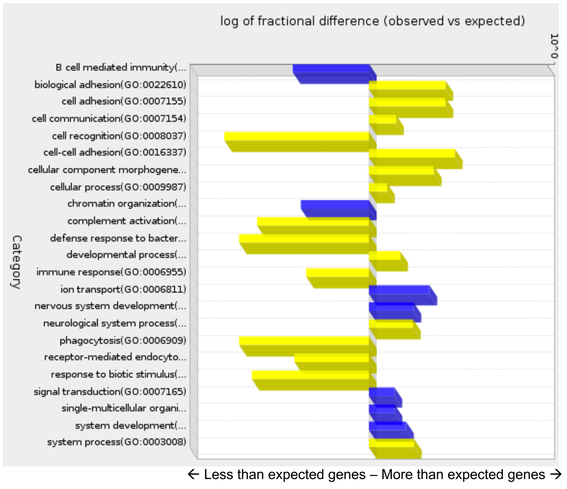
Shown: 91,837 SNPs in 15,236 genes

B. Intersection of SIFT and PolyPhen 2 annotations



Shown: 38,726 SNPs in 11,790 genes

C. Gene Ontology PANTHER database



Input: Intersection set of 11,790 genes

Test: Statistical overrepresentation

Highlighted in yellow: *p* value ≤ 0.005

Significantly overrepresented pathways include (+ more than expected / - less than expected, *p* value): biological adhesion (+, 3.37E-6), cell adhesion (+, 3.37E-6), and cell recognition (-, 1.26E-9).

See http://www.pantherdb.org/ for more information about PANTHER.

Summary of Results

❖EH vs. AD (Figures 2.A-B) suggest HYKK, SIMC1, NBEAL2, and BIN3 are potentially enriched for mutations in EH. HYKK protein, or hydroxylysine kinase, phosphorylates lysine during collagen catabolism.

❖Cell-cell adhesion genes are enriched in the EH, AD samples compared to baseline *Homo sapiens* in Gene Ontology pathways.

❖AD vs. NA shows no variants held in higher proportion by either group (not shown).

❖Indel variant calling did not find any additional genes of interest.

❖Sequencing artifacts likely account for the variants with high relative proportion of carriers in AD but with few carriers in EH (i.e. MUC3A). Alternatively, the 'damaging' annotation for these variants could be protective in AD.

CONCLUSIONS/IMPLICATIONS

❖Rather than a single SNP or indel, there are likely a suite of genetic mutations in skin barrier pathways that together contribute to the recurrent eczema phenotype.

❖Next steps will be to validate the carrier status of the mutations *in-silico* and with targeted sequencing of patients.

❖There are 2 individuals in the AD group with eczema vacciniatum, a rare complication of the smallpox vaccine. We will investigate if any novel mutations in their WGS data are also seen in EH.

❖ Funded by National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no: HHSN272201000020C