128

Genome-wide association study of Hidradenitis Suppurativa in a Multi-ethnic Cohort

Atlas Khan[1], M Geoffrey Hayes[2], John Connolly[3,4], Frank Mentch[3,4], Patrick Sleiman[3,4], Hakon Hakonarson[3,4], Joshua Denny[5], Chunhua Wang[1], George Hripcsak[1], Krzysztof Kiryluk[1], Lynn Petukhova[1]

1. Columbia University, NY, NY, United States; 2. Northwestern University Feinberg School of Medicine, Chicago, IL, United States; 3. Children's Hospital of Philadelphia, Philadelphia, PA, United States; 4. Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States; 5. Vanderbilt University, Nashville, TN, United States.

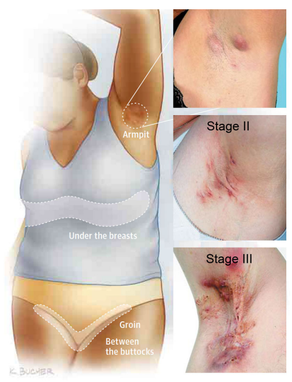
Columbia | Columbia University Department of Dermatology | emerge network Electronic Medical Records & Genomics | NIH National Center for Advancing Translational Sciences | CTSA Clinical & Translational Science Awards Program | CTSA Clinical & Translational Science Awards Program | NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases | epiCURE Columbia University Skin Disease Resource-Based Center

NHGRI U01HG008680 | NCATS KL2 TR001874 | NCATS UL1 TR001873 | NCATS UL1 TR001873 | NIAMS P30 AR069632 | Pilot & Feasibilty

Abstract

Limited knowledge about hidradenitis suppurativa (HS) pathogenesis and a lack of effective therapies contribute to high unmet needs. Unlike other common inflammatory skin diseases, there has never been a genome-wide association study (GWAS) conducted for HS. Translational genetic studies can have an immediate impact on patient care when results provide a rationale for drug repurposing, as our group helped to demonstrate with alopecia areata. Here, we performed a GWAS for HS using data obtained from an electronic biorepository (eMERGE project NT227). We identified a cohort of 600 HS cases and 82,611 controls with comparable multi-ethnic ancestry (λ=1.005). This HS cohort recapitulated HS race and gender predilictions with genetically African female participants accounting for 35% of cases, but only 10% of controls. Genotype data for 40 million variants was tested for association. No locus exceeded our threshold for statistical significance ($p<5x10-8$) although several were close ($p<10-7$), suggesting that a moderate expansion in cohort size may provide adequate power to detect associations. There was no evidence for HLA association. Interestingly, the lead SNP at one locus (rs11075745; $p=8x10-7$) is an eQTL for NFAT5, a mediator of NOTCH signaling whose expression is downregulated in HS lesional skin relative to patient-matched nonlesional skin. This preliminary finding awaits replication. Our group is constructing multi-ethnic replication cohorts that will allow us to expand this study in the near future.

Clinical Presentation



Alikhan, JAMA Derm, 2016 | Jemec et al., NEJM, 2012

Figure 1. Clinical presentation of HS[1,2]. Lesions occur in distinct anatomical regions[1]. Advanced stages of disease are marked by large abscesses, scarring, and extensive dermal sinus tracts that drain purulent exudate[2].

Analytic plan

1. Use diagnosis codes to identify cases and controls in a biorepository that links electronic health records to genome-wide SNP data.
   - Cases have ≥1 ICD acne diagnosis codes (706.1, L73.2)
   - Controls have no HS diagnosis code
2. Data quality control
   - SNP HWE, missingness
   - Sample missingness, relatedness (PCA), and racial comparability (admixture mapping)
3. Association Testing
   - Test 40 million genetic variants for association with acne, controlling for gender, ancestry, and genotyping platform

Figure 2. Study design. This study used data from an electronic biorepository assembled by the eMERGE consortium which links electronic health records to genetic data. Cohorts are assembled from diagnosis codes and standard analytic methods are used to conduct GWAS.

Results I. Cohort description

Disease	Genetically African Male	Genetically African Female	Genetically Asian Male	Genetically Asian Female	Genetically European Male	Genetically European Female	Total
Control	5,538	8,973	879	1,190	30,649	35,382	82,611
Case	51	211	3	11	87	237	600
Total	**5,589**	**9,184**	**882**	**1,201**	**30,736**	**35,619**	**83,211**

Table 1. eMERGE HS cohort. Genetic ancestry was determined with admixture analysis. The majority of our cohort is genetically European (80%). African ancestry and female gender both increase risk of HS[3]. In our data there appears to be an interaction of ancestry and gender, with a female to male ratio of 1.4 across all controls, 2.7 among genetically European cases and 4.1 among genetically African cases. Because gender and race demographics are different between cases and control we control for these variables in our association analysis.

Results II.genetic comparability

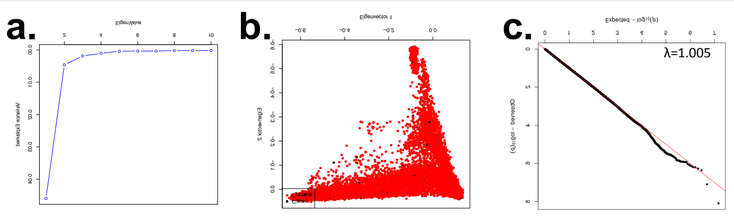


Figure 3. Cases and controls are genetically comparable. (a) Principal component analysis (PCA) with ancestry informative markers indicate that 5 PCs will control for ancestry. (b) A scatter plot of the first two PCs for cases (black dots) and controls (red dots) shows that they cluster. (c)The Q-Q plot shows strong concordance between observed and expected values (λ=1.005), indicating that population stratification is adequately controlled.

Results III. Manhattan Plot

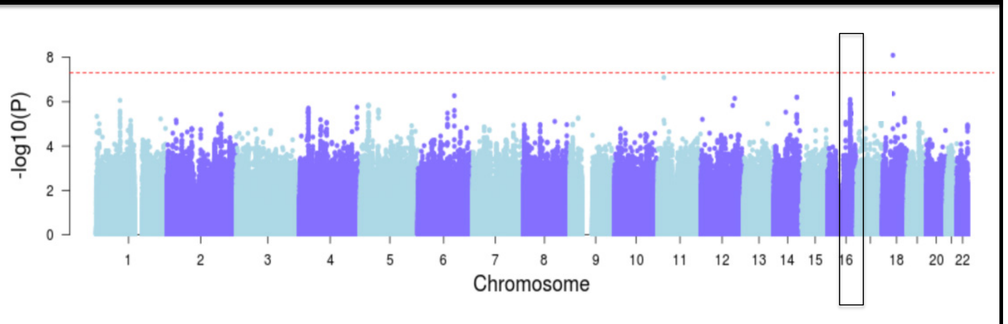


Figure 4. HS GWAS. We tested 40 million genetic variants. No region exceeded our threshold for statistical significance ($p=5x10^{-8}$), including the HLA. However several regions have pvalues and odds ratios that suggest a modest expansion in cohort size will provide adequate power to detect association. We are constructing additional cohorts. The region on chromosome 16q22.1 contains a compelling candidate gene (black box).

Results IV. region 16q22.1

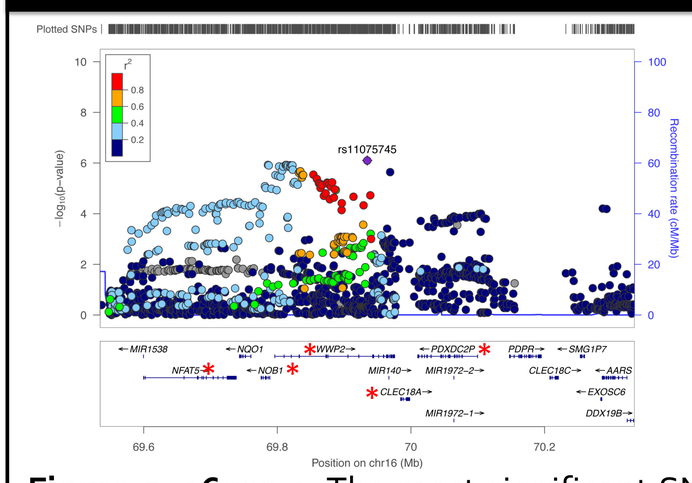
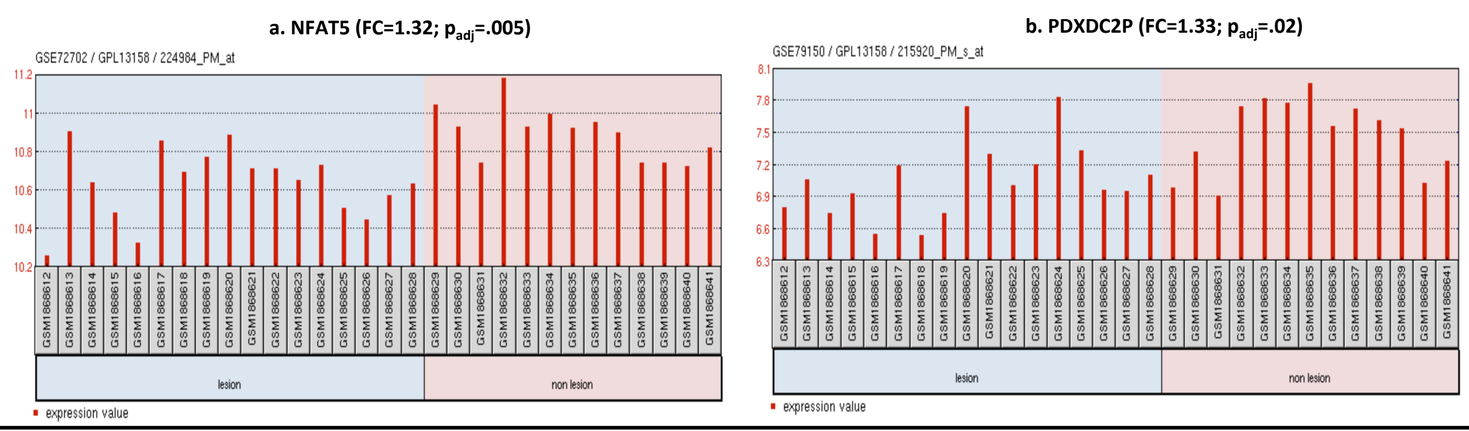


Figure 5. 16q22.1. The most significant SNP in this region (rs11075745; $p=8x10^{-7}$) is located within an intron of WWP2 and is an eQTL (https://gtexportal.org/) for several genes in this region (*).

Results: V. Differential gene expression in HS lesions of two transcripts at 16q22.1

Figure 6. Gene expression in HS patient-matched skin. We used publically available gene expression data (GSE72702)[4] to see if any transcripts reported to be influenced by rs11075745 have altered gene expression in HS skin. NFAT5 and PDXDC2P showed evidence for reduced expression in HS lesions compared to patient-matched non-lesional skin. Analysis was conducted in GEO using their GEO2R algorithm (https://www.ncbi.nlm.nih.gov/geo/geo2r). The threshold for statistical significance is adjusted to account for five tests (p=.01). Only NFAT5 is significant.



Conclusion

We conducted an HS GWAS in an electronic biorepository that links EHR to genome-wide genotype data. We assembled a multiethnic cohort as assessed by genetic ancestry. While no loci exceeded our threshold for statistical significance, our results demonstrate that a moderate expansion in the size of our cohort will allow us to discover HS associated loci. We found no association evidence at the HLA locus supporting the classification of HS as an inflammatory disease, rather than an autoimmune. One of the most significant loci implicates NFAT5, a mediator of Notch signaling.[5]

References

1. Alikhan A. JAMA dermatology 2016;152:736.
2. Jemec GB. The New England Journal of Medicine 2012;366:158-64.
3. Shlyankevich J et al. Journal of the American Academy of Dermatology 2014;71:1144-50.
4. Blok JL et al. The British journal of dermatology 2016;174:1392-4.
5. Tellechea M et al. J Immunol 2018;200:305-15.

Acknowledgements

We would like to thank all the eMERGE consortium research participants for donating their data. This work was additionally made possible by the generosity of members of the Columbia eMERGE project (NHGRI 5U01HG008680). Funding support also comes from NCATS KL2 TR001874, NCATS UL1 TR001873, and NIAMS P30AR069632 Columbia University Skin Disease Resource-Based Center (epiCURE).