|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT411 |
| **Submission Date** | 11/17/2020 |
| **Project Title** | Identifying disease phenotypes within hidradenitis suppurativa |
| **Tentative Lead Investigator** *(first author)* | Eric Mukherjee, VUMC |
| **Lead Investigator Email** | eric.mukherjee@vumc.org |
| **Tentative Senior Author** *(last author)* | Lee Wheless, VUMC |
| **eMERGE Site Sponsor & Contact** | Dan Roden (VUMC) |
| **All Other Authors**  | LaVar Edwards, VUMCPauleatha Diggs, VUMCLisa Bastarache, VUMCLynn Petukhova, ColumbiaChris Sayed, UNC Others TBD |
| **Sites Participating** | VUMCOthers TBD |
| **Background / Significance** | Hidradenitis suppurativa (HS) is a chronic and relapsing inflammatory skin disorder of the hair follicle characterized by deep-seated painful nodules and abscesses most often located in skin regions with apocrine glands. Diagnosis of HS is often delayed with previous studies showing a delay of around 7 years and as late as 12 years in one study. The clinical presentation of HS displays profound heterogeneity, complicating HS diagnosis. HS varies in disease location--most often HS affects the axillae, groin, inframammary and intermammary folds, perineal region, or buttocks, but it may manifest in other areas such as the face and neck--and disease severity. The Hurley staging system or the Sartorius severity score are often employed to categorize disease severity.Recently, more attempts have been made to categorize the phenotypes of HS in addition to classifying disease severity. The clinical heterogeneity of HS and resulting diagnostic challenges complicate the ability to perform population-based studies that explore associated conditions and health outcomes based upon disease severity. Presently, there are no clinical indicators or biomarkers to aid in prognosticating the disease course for patients. Given the various trajectories for disease severity and co-morbid conditions, this is an area of need that has been previously noted. A large cohort of HS patients with both phenotype and genotype data available is needed to facilitate this development.  |
| **Outline of Project** | 1. Develop phenotype algorithm to identify cases, Hurley stage from EHR
2. Identify and compare subgroups based on demographics and Phecodes that correlate with disease severity, treatment response
3. Identify and compare subgroups based on genetic variants that correlate with clinical phenotypes
4. Develop genotype and/or phenotype risk scores for prediction of disease and early diagnosis and treatment
 |
| **Desired Data - Common Variables\*** *(Available from the CC)* | xDemographics xICD9/10 codesxCPT codesxPhecodes* BMI
 | * Common Variable Labs

xCommon Variable Meds* Other: Case/Control status on Phase I and Phase II phenotypes
 |
| **Other Desired Data *(Available from participating sites)*** | *NLP for keywords “Hidradenitis”, “hydradenitis”, “Hurley Stage”*  |
| **Desired Genetic Data** | X eMERGE I-III Merged set (HRC imputed, GWAS)* eMERGE PGx/PGRNseq data set
* eMERGEseq data set (Phase III)
* eMERGE Whole Genome sequencing data set

x eMERGE Exome chip data set* eMERGE Whole Exome sequencing data set
* Other (not listed above):
 |
| **Does project pertain to an existing eMERGE Phenotype?** | * Yes, if so please list

X No. We are actively developing a phenotype |
| **Planned Statistical Analyses** | AUC, F-1 score maximization for algorithms; PheWAS based on demographics, common and rare genetic variants; cluster analysis of PheCodes;  |
| **Ethical Considerations** | Ethical considerations around the return of genomic research results in patients with hidradenitis suppurativa |
| **Available Funding or Resources** | VUMC VICTR funding |
| **Target Journal** | Nature Genetics (genetic clustering paper); others Journal of Investigative Dermatology |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Submit manuscript on phenotype definition by March 2021Have phecodes data available for initial PheWAS, clustering by April 2021First draft of PheWAS manuscript(s) completed by June 2021Submit first PheWAS manuscript by July 2021Genetic data available for PheWAS by August 2021Draft of genetic PheWAS, clustering paper ready by October 2021Submit genetic paper by November 2021 |
|  |  |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants