|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT372 | |
| **Submission Date** | 01/14/2020 | |
| **Project Title** | Association of PRS for Lupus with Classification Criteria-Based Disease Attributes | |
| **Tentative Lead Investigator** *(first author)* | Theresa Walunas | |
| **Tentative Senior Author**  *(last author)* | Laura Rasmussen-Torvik | |
| **All Other Authors** | Adam Gordon, Yu Deng, Jennifer Pacheco, Yuan Luo, Abel Kho, Vesna Mitrovic, Rosalind Ramsey-Goldman, Maureen Smith and others from interested eMERGE sites | |
| **Sites Participating** | Northwestern University and other interested eMERGE sites | |
| **Background / Significance** | Lupus is a disease of highly variable presentation and severity that affects predominantly women of color and remains difficult to identify given that manifestations occur over time and can be difficult to distinguish from less severe inflammatory and autoimmune conditions. Strong evidence suggests that existing damage is a predictor of long term damage (1) and early intervention and treatment with anti-malarials such as plaquenil can lead to slower disease progression and improved health outcomes.  Unsurprisingly for a highly variable disease, GWAS to explore the genetic components have revealed a large panel of genes in European and Asian cohorts that are likely to contribute to disease development. A recent meta-analysis showed a strong association of genetic risk score for lupus in persons with East Asian and African origins. In addition, groups have explored genetic risk scores to predict and association with subphenotypes using 22 polymorphisms (2) and longitudinal disease activity using 41 polymorphisms (Gianfresco, et al., 2016). The genetic risk score was associated with earlier onset of disease (before 34 years of age after 34 years of age, OR 1.06 [1.03-1.08] and the presence of anti-dsDNA antibodies (OR 1.10 [1.07-1.13]) in European populations but was not found to be associated with disease activity as determined by self-report, though a subset of variants were associated with clinically relevant disease activity scores.  We developed a rules-based algorithm for the detection of SLE based on the Systemic Lupus International Collaborating Clinics Classification Criteria for SLE which is comprised of 17 criteria divided into clinical and immunologic domains. To be classified as having “definite SLE” an individual must have a criteria in each domain and have 4 or more criteria overall. Using a gene set derived from the previous PRS studies and enhanced with information from an eMERGE based GWAS with our SLICC based-phenotype, we propose to use this classification criteria based algorithm to identify patients with SLE and examine the associations of Lupus PRS with these criteria.  1 Bruce, IN, et al. Ann. Rheum Dis. 2015. 74(9): 1706-13  2 Taylor, KE, et al., PLoS Genet. 2011. 7(2): e1001311  3 Gianfresco, MA, et al., Genes and Immunity. 2016. 17(6): 358-62 | |
| **Outline of Project** | 1. Calculate and refine lupus polygenic risk scores (PRS) for overall eMERGE cohort based on two previously developed PRSes (see above). 2. Compare the PRS performance in European ancestry population and non-European ancestry populations based on self-reported demographics and PCAs. 3. Assess the relationship between the PRS and individual classification criteria attributes 4. Compare PRS performance to in subpopulations of patients with common disease presentation as determined by subphenotyping based on groups of clinical classification criteria attributes 5. Optimize the PRS for one eMERGE site and test its performance across different sites and evaluate the transferability of PRS. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  The lupus phenotype is dependent on lab data. We will need labs for the following:  Autoantibodies: (anti-Smith, anti-phospholipid, anti-dsDNA)  Low Complement  Direct Coombs Test  Anti-Nuclear Antibody  WBC (to include leukocytes and thrombocytes)  All tests are part of the SLE phenotype developed within eMERGE. Anti-NA, anti-Smith, anti-dsDNA are also part of the autoimmune disease phenotype | |
| **Desired Genetic Data** | eMERGE I-III Merged set (HRC imputed, GWAS)  eMERGE PGx/PGRNseq data set  eMERGEseq data set (Phase III)  eMERGE Whole Genome sequencing data set  eMERGE Exome chip data set  eMERGE Whole Exome sequencing data set  Other (not listed above):HLA from PGRNSeq and imputed from other sets | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list Phase III, Systemic lupus erythematosus phenotype  No | |
| **Planned Statistical Analyses** | 1. Examine ROC AUC for the PRS, NRI for PRS, OR (lupus) for top and bottom deciles of PRS risk | |
| **Ethical Considerations** | None | |
| **Target Journal** | Arthritis Care and Research or similar | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * 01/2020: Develop initial PRS calculation pipeline for eMERGE; * 04/2020: Evaluate the transferability of lupus PRS across different ancestries; * 07/2020: Assess the association of the PRS with individual classification criteria attributes and subpopulations of lupus patients based on classification criteria; * 8/2020: Manuscript draft completion; * 9/2020: Manuscript submission; | |

**\*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