|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT331 | |
| **Submission Date** | 2/22/2019 | |
| **Project Title** | GWAS for Lupus Identified with a Classification Criteria-Based Phenotyping Algorithm | |
| **Tentative Lead Investigator** *(first author)* | Theresa Walunas | |
| **Tentative Senior Author**  *(last author)* | Laura Rasmussen-Torvik | |
| **All Other Authors** | Anika Ghosh, Jennifer Pacheco, Abel Kho, Rosalind Ramsey-Goldman, Maureen Smith and others from interested eMERGE sites | |
| **Sites Participating** | Northwestern University and other interested eMERGE sites | |
| **Background / Significance** | Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease that has diverse manifestations that can occur over a long period of time. The complexity of the disease and its varied presentation makes identification of patients with SLE difficult. A better understanding of disease presentation based on clinical aspects of disease could support earlier identification of disease, personalized treatment, improve identification of patients for clinical trials, and support research into the genetic and environmental mechanisms of SLE. We developed an rules-based algorithm for the detection of SLE based on the Systemic Lupus International Collaborating Clinics Classification Criteria for SLE [1] 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. We propose to use this classification criteria based algorithm to identify patients with SLE and to determine if it is possible to identify subpopulations based on clinical classification criteria and genetic information.  [1] 1. Petri M, Orbai AM, Alarcon GS, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis and rheumatism.* 2012;64(8):2677-2686. | |
| **Outline of Project** | We are developing three algorithms to identify people with SLE and classify their disease based on the SLICC clinical classification criteria set. The first algorithm is the “uncomplicated” algorithm with no NLP, the second is the uncomplicated algorithm with NLP and the final is the algorithm ported to the OMOP framework. We will determine which algorithm has the highest PPV and select that one to conduct the GWAS. We will conduct a GWAS on patients with SLE and develop sub-phenotypes based on gene analyses. | |
| **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)  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** | We will conduct a cross sectional study of SLE (definite SLE and probable SLE vs not SLE) based on the SLICC classification criteria. Definite SLE is defined as having one criteria in each of the clinical and immunologic domains and meeting 4 or more criteria over all. Probable SLE is defined as having one criteria in clinical and immunologic domains and 3 criteria overall Candidate gene analyses (using significant hits from our GWAS) will be completed using definite SLE and possible SLE vs. non-SLE. Additional supervised machine learning based sub-phenotyping based on clinical classification criteria and significant GWAS hits will be explored. | |
| **Ethical Considerations** | None | |
| **Target Journal** | Arthritis Care and Research | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | July 31, 2019: Clean data delivered from all participating sites for NLP based algorithm  September 30, 2019: GWAS Analysis Completed  November 30, 2019: Draft Manuscript Circulated to Co Authors  December 30, 2019: Second Draft Circulated to Co-Authors  January 30, 2020: Article submitted for publication. | |

**\*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