**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

This is project is part of the Phenotyping Workgroup’s list of Network phenotypes and the algorithm(s) is expected to be implementation-ready by 06/2016.

|  |  |
| --- | --- |
| **Submission Date** | 1/21/2016 |
| **Project Title4** | **Combined GWAS-PheWAS Approach to Serologic Markers of Autoimmunity & Inflammation** |
| **Tentative Lead Investigator (first author)** | Nikol Mladkova |
| **Tentative Senior Author (last author)** | Krzysztof Kiryluk, Ali Gharavi |
| **All other authors** | George Hripcsak, Chunhua Weng, Ning (Sunny) Shang, Yufeng Shen, Jun Zhang, Lynn Petukhova, Mitch Elkind, others and “The eMERGE Network” plus ***any additional eMERGE authors interested in participating*** |
| **Sites Involved** | We propose a network-wide study (all sites invited to participate). The analyses will be led by Columbia University. |
| **Background / Significance** | Many chronic inflammatory and destructive diseases are autoimmune, including rheumatoid arthritis, Graves' disease, Hashimoto's thyroiditis, or Sjogren's syndrome, which each affect about 1% of the world's population. In addition, autoimmune diseases also comprise less-common conditions, such as systemic lupus erythematosus (SLE), multiple sclerosis (MS), inflammatory bowel disease (IBD), autoimmune hepatitis (AH), primary biliary cirrhosis (PBC), ankylosing spondylitis (AS), and various types of glomerulonephritis (GN). GWAS approach has been incredibly successful for a number of autoimmune phenotypes, shed new light on the shared pathogenic mechanisms across the entire spectrum of autoimmune disorders, and identified a striking overlap of genetic loci between autoimmune and chronic inflammatory diseases.  Serologic markers of autoimmunity and inflammation, such as ANA titers, ANCA (PR3 and MPO) titers, complement (C3/C4) levels, and quantitative immunoglobulin levels (IgG, IgA, IgM), are measured routinely in the clinical evaluation for a variety of inflammatory and immune-mediated diseases, including SLE, glomerulonephritis (GN) or different types of systemic vasculitis (SV). Moreover, a substantial fraction of healthy individuals may also have elevated titers of some of these markers (e.g. ANA) and genetic regulation of these important biomarkers is largely unknown.  Here, we propose to use the serologic markers of autoimmunity and inflammation derived from EHR for a large number of genotyped EMERGE participants to discover novel disease susceptibility loci. To accomplish this, we propose a stepwise GWAS-PheWAS approach that combines GWAS for quantitative endophenotypes (autoimmunity markers) with follow-up PheWAS to define pleiotropic disease associations. We hope that this approach will offer a powerful new method for detection of novel susceptibility loci that are shared across the entire spectrum of autoimmune and inflammatory diseases. |
| **Outline of Project** | The project will be conducted in several stages:   1. Building Columbia EMERGE autoimmunity and inflammation phenotypes and sub-phenotypes (also included as Columbia EMERGE III subphenotypes) 2. Implementation of phenotype algorithms for all individuals with available GWAS datasets network-wide 3. Phenotype quality control analyses 4. Genome-wide association analyses 5. PheWAS to discover pleiotropic associations for significant loci 6. Manuscript preparation and submission |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | Implementation of standardized Columbia EMERGE phenotypes related to autoimmunity and inflammation across all sites with available GWAS datasets:   * ANA serology sub-phenotype\* (quantitative) * ANCA serology sub-phenotype\* (quantitative) * C3 and C4 sub-phenotype\* (quantitative) * Total protein, IgA, IgM, and IgG levels sub-phenotypes\* (quantitative) * Autoimmunity phenotype (Columbia EMERGE III)\* * Age, sex, race/ethnicity, albumin, type of ascertainment (if disease-driven)\* * ICD9/ICD10 codes for PheWAS of significant loci for the phenotypes defined as above |
| **Desired data** | * Standardized autoimmunity/inflammation phenotypes as above * All genotype data from EMERGE sites * Imputed genotype data for all sites. |
| **Planned Statistical Analyses** | Standard case-control GWAS for binary traits and quantitative GWAS for continuous traits (e.g. C3/C4 levels or immunoglobulin levels). Secondary analyses will include conditional analyses, haplotype analyses and rare-variant association scans. Data from individual centers will be metaanalyzed genome-wide using standard approaches. These analyses will be performed in the Kiryluk Lab at Columbia University.  PheWAS analysis will be performed for a small number of genome-wide significant loci to better define their disease associations and potential pleiotropic effects. Depending on site preferences, the PheWAS analyses may be performed by each individual site or centrally at Columbia. The summary statistics will be combined by meta-analysis across all EMERGE sites. |
| **Ethical considerations** | There are no additional risks involved. The EMR and genomic data will be stored at a secured location in the data storage system of Kiryluk and Gharavi labs. No data will be shared with unauthorized third parties. Patient identity will not be compromised by the proposed analysis. We will also abide by the EMERGE guidelines in this regard. |
| **Target Journal** | TBD, depending on the findings |
| **Milestones\*\*** | Total Duration of the study: 3 years  Implementation of phenotyping algorithms: 2016  Implementation of GWAS analyses: 2017  Draft of manuscript to authors: 2018  First submission: 2018 |

**\*\*** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.