**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 10/4/2016 |
| **Project Title** | **“Pulm\_eMERGE” Common and rare variant association of respiratory phenotypes derived from ICD9&10 codes using the network-wide eMERGE cohorts** |
| **Tentative Lead Investigator (first author)** | Jessica Lasky-Su |
| **Tentative Senior Author (last author)** | Scott Weiss |
| **All other authors**  | Damien Chonka-Croteau, Peter Castaldi, Michael Cho, Beth Karlson, Jordan Smoller, Shawn Murphy.*Any additional eMERGE members interested in participating* |
| **Sites Involved** | We propose a network-wide study of respiratory disease phenotypes derived from the ICD 9&10 codes (all sites invited to participate).  |
| **Background / Significance** | Genetic factors heavily influence an individual’s liability for respiratory diseases such as asthma and COPD. While genome-wide association studies (GWAS) have confirmed a number of loci, little is known about how these variants impact disease and the effect of each individual genetic variant is small. While there is widespread belief that rare genetic effects have a larger impact on disease, the assessment of rare genetic variants for respiratory diseases is not comprehensive. Many of the current respiratory genes are highly polygenic and the identification of specific variants within these genes would advance the field tremendously. Therefore, our objective is to identify both common and rare genetic variants for respiratory diseases, both individually and globally. The goal of this study is to systematically characterize pleiotropic effects of these variants using an EMR-based PheWAS approach. Of note, a prior eMERGE publication (Almoguera, AJRCCM, 2016) used a GWAS approach to study associations with asthma using a validated algorithm for defining asthma cases in eMERGE 2. The proposed analyses use ICD-9/ICD-10 codes for asthma, COPD and other respiratory outcomes, not validated asthma cases.An important step in this approach is to combine the various respiratory diseases in a PheWAS to identify both the common and unique genetic associations underlying respiratory diseases. This goal is made feasible through the use the EMR, ICD-9 codes, and curated phenotypes that are linked to genetic data. |
| **Outline of Project** | The project will involve the following steps:1. Define outcomes: Request ICD-9/ICD-10 codes from E1, E2, E3 for asthma, COPD, and respiratory phenotypes. Define respiratory PheWAS groups.
2. Request imputed genotypes from E1, E2, E3.
3. Perform PheWAS on respiratory phenotypes using eMERGE network-wide imputed GWAS for associations with respiratory PheWAS groups.
4. Perform rare variant aggregate analysis using eMERGE-Seq on 4 relevant candidate genes (IL33, VDR, TNF, SERPINA1) for association with respiratory PheWAS groups.
5. We will perform an association analysis with the same respiratory phenotypes using the relevant candidate SNPs: C5, CARD9, NOD2, CD14, HLA-DRB1, IFNG, IL10, IL1A, IL1B, IL1RL1, IL22RA1, IL33, IL4, IL6, VDR, TNF, CTLA4, TNFAIP3, TYK2, AOAH, SERPINA1, SFTPA2, SFTPA1, SFTPC, SFTPB, ABCA3, NKX2-1
6. PheWAS categories will be generated using the following ICD-9 codes: 460-<520. We will also use the corresponding ICD10 codes.
7. Identify major common and disease-specific respiratory variants in terms of effect size and population attributable risk
8. Manuscript preparation and submission
 |
| **Desired****Variables** **(essential for analysis****indicated by \*)** | This study involves PheWAS analysis of ICD and procedure codes for repiratory disease among all EMERGE participants. The required variables include:* ICD-9/ICD-10 codes for PheWAS network-wide (E1, E2, E3)\*, specifically for phenotypes relevant for respiratory diseases
* Imputed GWAS data (E1, E2, E3)
* eMERGESeq data on 109 candidate genes
* adjustment covariates including the following: age, sex, race/ethnicity, cohort/site\*
 |
| **Desired Data** | * Imputed genome-wide genotypes (E1, E2, E3)
* Raw genome-wide genotypes
* Sequence data on 4 genes
* Emerge SNPs panel from specified relevant genes
* eMERGESeq data on 109 candidate genes
* Genetic ancestry information
* Age, sex, race/ethnicity
 |
| **Planned Statistical Analyses** | We will focus on ICD-9/ICD-10 diagnoses related to asthma, COPD, and respiratory diseases using the PheWAS data. We will also obtain the relevant covariates for those patients using the EHR. We will assess the genetic associations with respiratory phenotypes using a multi-faceted approach:1) We will perform a PheWAS using eMERGE3 imputed network-wide genotypes and respiratory phenotypes defined by PheWAS groups. 2) We will perform aggregate association tests using the respiratory phenotypes defined by PheWAS groups for rare common variants in the candidate genes and sequence data specified above, while adjusting for the appropriate covariates.3) We will describe their genetic effect in terms of effect size and population attributable risk, and examine shared phenotypic associations with the various respiratory-relevant phenotypes.  |
| **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 at Brigham and Women’s hospital. 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** | To be determined |
| **Milestones\*\*** | 1. January 2017 Asthma GWAS analysis for asthma, COPD and related phenotypes defined by ICD9/10 codes
2. Rare variant analysis of asthma and COPD candidate genes using NGS and imputed data
3. Jan 2017: Further assessment of asthma genes via genetic effect size and population attributable risk measurement.
4. March 2017: Manuscript preparation
5. June 2017: Draft of first submission
 |

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