**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Reference Number** | NT265 |
| **Submission Date** | 12/01/2017 |
| **Project Title** | Shared and distinct genetics of childhood asthma and adult obstructive lung disease |
| **Tentative Lead Investigator (first author)** | Su H. Chu |
| **Tentative Senior Author (last author)** | Jessica Lasky-Su, Scott Weiss, Elizabeth Karlson |
| **All other authors** | Rachel Knevel, Damien Croteau-Chonka, Emily Wan, Michael Cho |
| **Sites Involved** | We welcome participation of all sites. |
| **Background / Significance** | Chronic obstructive pulmonary disease (COPD) is a heritable respiratory disease with high prevalence and well-established environmental and genetic risk factors. There is increasing evidence that asthma and reduced lung function in childhood are significant risk factors for COPD, including more severe COPD-subtypes with asthma known as asthma/COPD overlap (ACO). However, the specific mechanisms by which childhood asthma might affect risk for COPD, or ACO, later in life remain poorly understood.  The objective of this proposal is to identify actionable genetic variants that influence asthma risk in children and might be linked to risk for COPD and ACO later in life. The aims of this study are as follows:   1. Harvard site to develop phenotyping algorithms for COPD and for ACO leveraging electronic medical records. Marshfield to perform secondary validation. eMERGE sites to implement algorithms to identify COPD cases and controls, and ACO cases and controls. 2. Identify shared and distinct sets of common and rare genetic variants associated with risk for asthma in childhood, COPD in adulthood, and ACO in adulthood.   Although many studies have explored the genetics of asthma and COPD separately, few have been conducted within the same source population, and fewer still have directly studied patients with symptoms of both. To our knowledge, the proposed study will be the first to directly compare the genetics of childhood asthma with the genetics of COPD and ACO in adulthood using the same source population. It will also be among the largest GWAS studies conducted in these respiratory phenotypes to date. |
| **Outline of Project** | The project will undertake several analyses:   1. **Development of COPD and ACO phenotyping algorithms.**    1. Will first be developed at Harvard, validated internally, then at Marshfield, and then requested to be applied to full eMERGE network. 2. **Common variant analysis**    1. Childhood asthma GWAS1       * N.B.: Asthma GWAS analyses were recently conducted and published in CHOP pediatric participants recruited at their site, and adult eMERGE data (Almoguera et al 2017). It is our intention to conduct complementary but slightly different analyses by using all the pediatric subjects from eMERGE in Phase III, but the distinct end goal of our proposal is to identify potential risk-conferral from the loci identified in our proposed childhood asthma GWAS for risk of adulthood COPD/ACO.    2. Adult COPD GWAS    3. Adult ACO GWAS    4. Joint Adult COPD + ACO GWAS 3. **Rare variant analysis**    1. Combined common+rare variant association tests for top genes    2. Rare variant testing for top genes 4. **Enrichment analysis**    1. Pairwise enrichment for common genetic loci from childhood asthma to adult dx    2. Functional enrichment of annotation |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | \*Genotypes:   * Imputed genome-wide data for full eMERGE network Phase III cohort.   \*Phenotypes:   * ICD and CPT codes, age, race/ethnicity, sex, smoking status, BMI, case/control status for asthma, COPD, and ACO, spirometry if available |
| **Desired data** | Genotypes from eMERGE imputed dataset.  ICD codes (presence/absence and frequency), CPT codes, case/control status for asthma, COPD, and ACO.  Common variables for demographics, smoking and BMI. |
| **Planned Statistical Analyses** | 1. **Phenotyping:**    1. Phenotype algorithms developed based on ICD code frequencies, and coded electronic medical record date in Partners Biobank for COPD and ACO.    2. Apply algorithm, validate in secondary eMERGE site, apply to full eMERGE network. 2. **Common Variant Analysis:**    1. Four sets of initial GWAS:       1. Childhood asthma       2. Adult COPD       3. Adult ACO       4. Pooled adult COPD + ACO    2. Obtain top hits from each GWAS, and using childhood asthma SNPs as referent loci, identify overlapping sites. Replicate in pairwise pooled analysis combining childhood asthma cases and controls with each of the adult respiratory GWAS samples.    3. Replicate top overlapping pooled findings from 1b. in UK Biobank.2 3. **Rare Variant Analysis**    1. Genes associated with overlapping sites will be further assessed for rare variant analysis testing.    2. A combined, gene-based analysis of rare and common variants using the Combined Multivariate and Collapsing (CMC) method for case-control data among the top overlapping genes identified above.3    3. Rare variants will be assessed using the SKAT framework (RC-SKAT).4 4. **Enrichment Analysis**    1. Gene set enrichment analysis techniques will be used to analyze pairwise overrepresentation of the top sets of genes implicated by childhood asthma risk loci in each of the COPD, ACO, and pooled COPD and ACO discovery GWAS results.    2. Pathway enrichment analysis will be applied using the Genomic Regions Enrichment of Annotations Tool (GREAT).5 |
| **Ethical considerations** | There are no additional risks involved than the known risk of eMERGE data collection. Informed consent is obtained on all patients to provide blood for DNA analyses. The phenotypic and genetic data will be stored at a secured location in the data storage system at Partners Enterprise Research Information System (ERIS). No data will be shared with unauthorized third parties. Patient’s 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 impact: Nature Genetics / AJRCC/JACI) |
| **Milestones\*\*** | Total Duration of the study: 12 months  Phenotyping algorithm: 6 months   * + Development: 4 months   + Phenotype collection: 2 months   Common Variant Analyses: 1 month  Rare Variant Analyses: 3 months  Enrichment Analysis: 2 months  Draft of manuscript to authors: Dec 2018  First submission: Jan 2019 |

**REFERENCES**

1. Almoguera, B. *et al.* Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am J Respir Crit Care Med* **195,** 456–463 (2017).

2. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 1–36 (2017). doi:10.1101/166298

3. Li, B. & Leal, S. M. Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet* **83,** 311–321 (2008).

4. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D. & Lin, X. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. *Am J Hum Genet* **92,** 841–853 (2013).

5. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28,** 495–501 (2010).