|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT366 |
| **Submission Date** | November 7, 2019 |
| **Project Title** | Replication of GWAS signals from UK Biobank Respiratory Infection GWAS  |
| **Tentative Lead Investigator** *(first author)* | Alex Williams |
| **Tentative Senior Author** *(last author)* | Martin Tobin |
| **eMERGE Site Sponsor & Contact** | Marylyn Ritchie |
| **All Other Authors**  | Alex Williams, Catherine John, Nick Shrine, Leicester co-authors Caroline Hayward and colleagues (Generation Scotland, ORCADES, VIKING)Ben Brumpton and colleagues (HUNT)Marylyn Ritchie and other eMERGE AuthorsErik Abner, Tõnu Esko (Estonian Biobank)Arden Moscati, Ruth Loos (BioMe)Teemu Palviainen, Jaakko Kaprio, Aarno Palotie, Mark Daly (FinnGen)Jiangyuan Liu, Su Chu, Michael Cho (Partners Biobank)Nancy Cox and colleagues (BioVU/Vanderbilt)Yanfei Zhang, Ming Ta Lee (DiscovEHR)Traci Bartz, Sina Gharib (Cardiovascular Health Study)Ian Hall, David Michalovich, Louise Wain, Martin Tobin(some author details within each group are still being collected) |
| **Sites Participating** | All |
| **Background / Significance** | Williams, Tobin et al have recently conducted a genome-wide association study of respiratory infection susceptibility, using ICD-10 codes in the full release of the UK Biobank data. They have finalized the list of variants for follow-up, and are now reaching out to cohorts that may be interested in contributing to replication.eMERGE has been invited to participate in the replication |
| **Outline of Project** | GWAS in UK Biobank has already been completed. There are 40 association signals for replication.In eMERGE, our outline is:* Identify target SNPs or LD proxies in the eMERGE-III imputed dataset
* Identify cases and controls based on ICD9/10 inclusion and exclusion criteria
* Conduct association testing for 40 signals
* Report PLINK results to Williams and Tobin
 |
| **Desired Data - Common Variables\*** *(Available from the CC)* | Demographics  ICD9/10 codes |  |
| **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)* Please provide a breakdown of the ICD codes used to define the respiratory infection cases, i.e. a count of how many individuals were defined as cases by each ICD code. |
| **Desired Genetic Data** | eMERGE I-III Merged set (HRC imputed, GWAS) |
| **Does project pertain to an existing eMERGE Phenotype?** | No, not that we are aware of. |
| **Planned Statistical Analyses** | 1. Defining case/control based on following criteria:
2. Extracting chr:pos for SNPs to replicate from UKBB (40 SNPs)
3. GWAS using logistic regression in PLINK
 |
| **Ethical Considerations** | None |
| **Available Funding or Resources** |  Dr. Ritchie start-up funding. |
| **Target Journal** |  |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1. Perform analyses in November 2019
2. Manuscript draft by March 2020
 |

**\*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