|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT333 | |
| **Submission Date** | 2/25/2019 | |
| **Project Title** | Identifying similarities in tissue-specific transcriptomic architecture of lipid traits between humans and mice. | |
| **Tentative Lead Investigator** *(first author)* | Yogasudha Veturi, post-doctoral fellow, Department of Genetics, University of Pennsylvania yveturi@pennmedicine.upenn.edu | |
| **Tentative Senior Author**  *(last author)* | Marylyn Ritchie, marylyn@pennmedicine.upenn.edu | |
| **eMERGE Site Sponsor & Contact** | University of Pennsylvania, Marylyn Ritchie | |
| **All Other Authors** | Yuki Bradford, Ronald Krauss, Marisa Medina, Elizabeth Theusch, Neil Risch and Thomas Hoffman | |
| **Sites Participating** | All emerge network | |
| **Background / Significance** | There is evidence to suggest that genes associated with lipid phenotypes can be replicated in mice. In order to obtain a better idea of the extent of overlap in transcriptomic architecture between the species, our goal is to use variance component analyses to estimate the proportion of variance explained by suggestive genes from human transcriptome-wide association analysis (TWAS) analyses in a diversity outbred mouse dataset that is on a high-fat diet. | |
| **Outline of Project** | • In this study, we propose to:  • (1) Run TWAS on eMERGE III European American adults  • (2) Replicate results in independent GWAS cohorts  • (3) Estimate proportion of variance explained in lipid phenotypes by large/small-effect TWAS genes using gene-expression data (liver) from an independent diversity outbred mouse cohort to help analyze extent of overlap in transcriptomic architectures between the species. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI  •Primary phenotypes: blood lipid levels (high density lipoprotein, low density lipoprotein, serum cholesterol and triglyceride levels in eMERGE III data) | Common Variable Labs  Common Variable Meds  Other: Case/Control status on Phase I and Phase II phenotypes  •Related phenotypes: systolic/diastolic blood pressure, serum urate levels, body-mass-index, smoking status, drinking status, medications taken |
| **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)* | |
| **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): | |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list  No | |
| **Planned Statistical Analyses** | Impute gene expression for eMERGE III European American adults using PrediXcan. We will use tissue-specific weights obtained from the GTEx consortium for this purpose and conduct single-gene regression analyses (TWAS) for lipid phenotypes using PLATO after adjusting for necessary covariates. We will subsequently replicate these results using MetaXcan on summary statistics from two independent GWAS cohorts (Global Lipids Genetics Consortium and Genetic Epidemiology Resource on Adult Health and Aging). Finally, we will apply maximum likelihood and Bayesian approaches to estimate the proportion of variance explained by individual-level liver gene expression data (utilizing the large and small effect TWAS genes obtained from the above three independent cohorts) in lipid phenotypes from 478 diversity outbred mice of both sexes that were fed either standard chow or a high-fat/high-sucrose diet from weaning until approximately 6 months of age. | |
| **Ethical Considerations** | None | |
| **Available Funding or Resources** |  | |
| **Target Journal** | TBD | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1) Submit manuscript by April 2019. | |

**\*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