|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT436 |
| **Submission Date** | 10/15/2021 |
| **Project Title** | Genome-wide association study of all-cause heart failure |
| **Tentative Lead Investigator** *(first author)* | Michael Levin |
| **Tentative Senior Author** *(last author)* | Ben Voight and Scott Damrauer |
| **eMERGE Site Sponsor & Contact** | Marylyn Ritchie  |
| **All Other Authors**  | Marylyn Ritchie, Pankhuri Singhal, Nosheen Reza, Yuki Bradford and other eMERGE AuthorsNoah Tsao, Tiffany Bellomo, William Bone, Krishna Aragam, Yifan Yang, Michael Morley, Megan Burke, Renae Judy, Zoltan Arany, Thomas Cappola, Sharlene Day, Patrick Ellinor, Kenneth Margulies |
| **Sites Participating** | All |
| **Background / Significance** | Heart failure is a leading cause of cardiovascular morbidity and mortality. While family-based studies have identified rare, large-effect variants that contribute to heart failure/cardiomyopathy, the common genetic basis of heart failure remains largely uncharacterized. The largest GWAS of heart failure to date included only individuals of European ancestry and has identified only 11 common-variant risk loci. Further understanding the polygenic nature of heart failure may help reveal novel genes and biological pathways, facilitate downstream causal evaluations of risk factors/drug reprioritization via Mendelian randomization, and improve risk stratification via polygenic risk score analyses. |
| **Outline of Project** | We have performed a multi-ancestry GWAS meta-analysis of all-cause heart failure, using publicly-available GWAS summary data. In a complimentary analysis we have applied a multi-trait GWAS method to integrate heart failure with additional genome wide association studies of related cardiac imaging traits. We performed extensive downstream bioinformatic analyses (genetic correlation, colocalization, transcriptome-wide association studies, gene expression profiling, and Mendelian randomization) to prioritize genes, tissues, cell-types, pathways, and drug-repurposing opportunities. A manuscript based on this data has received favorable reviews at *Nature Communications*, and in the revisions stage we are planning to further expand the scope of our heart failure meta-analysis to include additional cases, particularly of diverse ancestry. Data contributed by eMERGE will be integrated in our multi-ancestry GWAS meta-analysis. |
| **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)* none |
| **Desired Genetic Data** |  **eMERGE I-III Merged set (HRC imputed, GWAS)** |
| **Does project pertain to an existing eMERGE Phenotype?** | **We will be defining heart failure case status with ICD 9/10 codes as described below.**  |
| **Planned Statistical Analyses** | ·Phenotype: all-cause heart failure* We recommend including: ICD10: I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.8, I42.9, I50.0, I50.1, I50.9 and ICD9: 4254, 4280, 4281, 4289 (from Aragam et al Circulation 2019: <https://doi.org/10.1161/CIRCULATIONAHA.118.035774>).
* We recommend defining cases as individuals with one or more of these codes on two or more dates, with the remaining individuals considered controls

Method: Case/control analysis stratified by ancestry (if possible), adjusting for population structure using GWAS pipeline of choice (eg. Plink/Saige/Regenie preferred)* Imputation: at least 1000 Genomes phase 3 or other contemporary imputation panel (eg. HRC, TOPMed)
* Genome build: ideally hg19/GRCh37, although we can use liftOver if data is provided on hg38

GWAS output from pipeline of choice (separate files per ancestry) should include columns for chromosome, position, effect allele, non-effect allele, effect allele frequency, effect size, standard error, p-value, n\_cases, n\_controls, n\_total, rsid (if possible), imputation quality (eg. INFO; if possible) |
| **Ethical Considerations** | None |
| **Available Funding or Resources** |  Dr. Ritchie start-up funding. |
| **Target Journal** | *Nature Communications* |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 1. Perform analyses in October/November 2021
2. Submit results to Penn in November/December 2021
3. Submit revision to Nature Communications before end of 2021
 |

**\*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