|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* |  | |
| **Submission Date** |  | |
| **Project Title** | Exploring the polygenic components of congenital cardiac disease | |
| **Tentative Lead Investigator** *(first author)* | David Wu | |
| **Tentative Lead Investigator Email Address** | David.h.wu@vanderbilt.edu | |
| **Tentative Senior Author**  *(last author)* | Nancy Cox  nancy.j.cox@vumc.org | |
| **All Other Authors** | Dan Zhou, Tyne Miller-Fleming, Wei-qi Wei, Henry Ong, Antonis Hatzopoulos | |
| **Sites Participating** | VUMC/Vanderbilt | |
| **Background / Significance** | Congenital heart disease (CHD) is the most common birth defect and most frequent cause of death in newborns, carrying a lifetime burden of mortality and morbidity. Clinically, it is defined by structural defects in the heart or thoracic great vessels, occurring in about 1% of all births, and can occur either nonsyndromically as isolated events or syndromically accompanied by defects in other tissues and organs. Such problems result from perturbations to the tightly coordinated molecular processes of embryogenesis, and presentations can range widely from undiagnosed and undetected cases to severe disease requiring prompt surgical correction. Given the immense disease burden on lives of both individuals and care providers, there has been a great deal of work investigating the complex biology underlying CHD, but our understanding still remains incomplete. Our current understanding posits important roles for both environmental and genetic factors in the disease etiology. Recent work has established a strong genetic basis for disease and uncovered insights into biological mechanisms. However, even with the abundance of Mendelian genes discovered, an overwhelming majority - around 70-80% - of CHD cases still lack an established identifiable cause. On top of this enormous knowledge gap, no genetic study has confidently determined whether the underlying cause of the disease in an individual is due to a single variant with high effect size or the aggregation of many small effects of functionally consequential variants across many genes. In particular, the latter hypothesis may explain an important portion of unknown cases especially when clinical tests primarily operate under the presumptions of the former, but such an idea has never been successfully validated. Our work here helps to address this crucial gap while also rectifying the problems of functional interpretation that typically afflict genetic studies. | |
| **Outline of Project** | The aim of the project is to extract relevant genomic information relevant to expression of known mendelian congenital heart disease genes to aggregate these effects in a risk score so that we can explore whether this aggregated effect is significantly associated with the disease. Moreover, we aim to subset genes based on molecular biological function and explore the phenome associated with aggregated effects of genetic variation across different functional subsets of these genes. Taken together, our study aims to establish the polygenic basis for congenital heart disease and determine whether this genetic architecture is shared with other disease | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Geocoding 2015 ACS variables  Other: **Case/Control status for clinical diagnoses are needed** |
| **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** | GREx will be generated with available genotype data using best performing model for each gene (JTI/UTMOST/PrediXcan). Functional subsets of gene GREx values will be used to construct risks scores thst will then be created and run in a phenome association study with covariates including but not limited to PC-based ancestry, age, and sex. Lab values and medications will also be explored depending on phenotypes | |
| **Ethical Considerations** | De-identified data should allow our study to fall under non-human subjects. | |
| **Target Journal** | Genome Medicine, American Journal of Human Genetics, or Nature Medicine | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Upon approval and receipt of data, an estimated 6 weeks will be needed to perform analyses and generate results for further evaluation. This will be compiled with the positive results we have already generated in an independent cohort. Afterwards, a month will be required to prepare a draft before submission. Tentatively a goal of October to November 2022 for journal submission is set. | |

**\*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