|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT345 | |
| **Submission Date** | May 15, 2019 | |
| **Project Title** | CCHMC top 6 gene- MC4R; a brief-report—evaluation and prevalence of MC4R mutations among emerge-seq population | |
| **Tentative Lead Investigator** *(first author)* | Bahram Namjou | |
| **Tentative Senior Author**  *(last author)* | John B. Harley… | |
| **All Other Authors** | Ni, Yizhao, Lingren, Todd… | |
| **Sites Participating** | We invite each site to join | |
| **Background / Significance** | Melanocortin 4 Receptor gene (MC4R) mutations are the most common cause of monogenic obesity in children and adult with 2.5-6% prevalence. MC4R is a G protein coupled with a 332-amino acid protein encoded by a single exon localized on chromosome 18q22.It is expressed predominantly in brain and is known to play a key role in energy homeostasis. Both common and rare variants have been implicated in MC4R-link obesity. According to ClinVar variation report, more than 95 mutations in MC4R are considered as pathogenic. The affected individuals are often carrier (heterozygotes) with an autosomal dominant pattern of inheritance and therefore with variable phenotypes. In fact, instances have been reported in which individuals in the general population have these mutations but are not obese. In addition to loss of function mutations, new data indicate the presence of gain of function mutations with protective effect against obesity. MC4R is one of the CCHMC top 6 genes for sequencing in >20K emerge-seq participants in order to better understand the MC4R mutation frequency in different racial group, effect of age on penetrance (for obesity or other related phenotypes such as type 2 Diabetes and coronary artery disease), PheWAS analysis and to discover other potential new variants link to obesity. Finally, common GWAS variants near MC4R also has been linked to BMI and complex subtype of obesity in which we plan to explore in more than 100K emerge-GWAS population. | |
| **Outline of Project** | 1. Receiving final dataset for emerge seq and emerge-gwas from CC 2. Annotate all known pathogenic, likely pathogenic and VUS(es) for MC4R 3. Define the cases and controls for different metabolic diseases using ICD9 and Phecodes and evaluate the penetrance of detected variants by estimating the proportion of positive findings in case versus controls in different age group 4. Perform PheWAS and PheRS and assess pleiotropy 5. Evaluate haplotype structure in different ancestry , perform conditional analyses to detect interaction effects between common and rare variants by merge and re-imputing all collection in this genomic region 6. Functional mapping, gene based analyses and evaluation of potential transcription factor binding sites near MC4R will also be in our agenda. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | Demographics  ICD9/10 codes  CPT codes  Phecodes  BMI | Common Variable Labs  Common Variable Meds  Other: Case/Control status on Phase I and Phase II phenotypes |
| **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  a) obesity, extreme obesity, BMI, T2DM , CAD and other metabolic diseases  No | |
| **Planned Statistical Analyses** | PheWAS, PheRS, Logistic or linear regression, conditional analyses, SKAT, CMC | |
| **Ethical Considerations** | None | |
| **Target Journal** | TBD | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 6/2019-Multi-sample calls will be distributed by CC  6/19-9/19 extensive Analyses mentioned above  10/19-12/19 preparation of manuscript-brief report   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  | |  | | |  |  |  | |  | |  | | |  |  |  | |  | |  | | |  |  |  | |  | |  | | |  | |  | |  | |  | | |  |  |  | |  | |  | | | |

**\*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