|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT378 | |
| **Submission Date** | 02/19/2020 | |
| **Project Title** | Drug metabolism phenotypes of patients with rare or unknown VKOR and CYP2C9 variants | |
| **Tentative Lead Investigator** *(first author)* | Melissa Chiasson | |
| **Tentative Senior Author**  *(last author)* | Doug Fowler | |
| **All Other Authors** | Melissa Chiasson, Clara Amorosi, Maitreya Dunham, Doug Fowler, Lisa Bastarache | |
| **Sites Participating** | University of Washington  Vanderbilt University | |
| **Background / Significance** | Genetic variation in VKOR and CYP2C9 can lead to altered metabolism of warfarin and other drugs. Common alleles of these genes such as VKOR -1639G>A, CYP2C9\*2 and CYP2C9\*3 have been functionally characterized and used in genotype-guided dosing efforts, but rare variants lack functional annotations and thus clinical actionability. We have developed massively parallel assays of abundance and activity for both VKOR and CYP2C9 and expect some of these variants to be present in eMERGE datasets. By comparing drug metabolism phenotypes such as warfarin dose/INR to our experimentally-generated functional scores, we can determine the clinical utility of our dataset. | |
| **Outline of Project** | We have developed massively parallel assays of VKOR and CYP2C9 activity and abundance, resulting in functional annotations for 2,695 VKOR and 8,062 CYP2C9 single missense variants. These assays leverage deep sequencing of variant libraries with appropriate functional selections for VKOR and CYP2C9 to produce functional scores for thousands of protein variants en masse. These datasets are a powerful resource for genotype-guided dosing with warfarin, phenytoin, and other drugs. Because our datasets are so comprehensive, there are likely to be many patients with one or more of these VKOR or CYP2C9 variants with existing clinical data in the eMERGE databases. We will examine genotype-based differences in drug metabolism phenotypes and determine the extent to which our functional scores explain differences in patient drug metabolism phenotypes. This will allow us to determine the how our datasets can be a resource for informing future genotype-guided dosing efforts. | |
| **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)*  Nothing beyond common variables. | |
| **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: Warfarin dose/response, PGx medication risk prediction model, Epilepsy/Antiepileptic drug response algorithmWarfarin dose/response No | |
| **Planned Statistical Analyses** | We will use regression analysis to replicate known associations between VKOR/CYP2C9 variants and drug response, and test for association with novel variants. | |
| **Ethical Considerations** | None | |
| **Target Journal** | Nature Genetics | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | 2020-03-02 Get the data  2020-03-30 Check the data and format for analysis. Run the replication analysis. Run the novel variant analysis.  2020-05-04 Finish draft manuscript  2020-05-25 Submit manuscript | |
|  |  | |

**\*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