**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | 2/9/2016 |
| **Project Title** | Pharmacogenetic variation identified via targeted next-generation sequencing among 9000 eMERGE subjects |
| **Tentative Lead Investigator (first author)** | Adam Gordon |
| **Tentative Senior Author (last author)** | David Crosslin |
| **All other authors** | Gail P. Jarvik |
| **Sites Involved** | University of Washington/Group Health, Seattle, WA;  Mayo Clinic, Rochester, MN;  Marshfield Clinic Research Foundation, Marshfield, WI;  Northwestern University, Chicago, IL;  Vanderbilt University, Nashville, TN;  Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; and  Broad Institute of Harvard & MIT; Cambridge, MA.  Essentia Health; Duluth, MN.  Geisinger, Danville, PA.  The Mt. Sinai Hospital; New York, NY. |
| **Background / Significance** | In order to investigate the utility of linking next-generation sequencing data with electronic medical records, the network obtained data from PGRNseq, a sequencing panel targeting 84 pharmacogenes, on 9000 eMERGE subjects from across the network. As this now represents one of the largest available datasets with deep sequencing of pharmacogenes linked to EHR data, this network-wide dataset presents a valuable opportunity to assess the extent both of common alleles with known pharmacogenetic effect as well as rare variation within these genes that may influence a variety of drug phenotypes. |
| **Outline of Project** | 1. Deep annotation of full eMERGE-PGx dataset 2. Compare frequencies of known PGx alleles in our dataset with those reported in other public databases 3. Summary and analysis of novel, rare, potentially deleterious variation across all PGRNseq genes for which CPIC guidelines exist. 4. Generate a phased dataset using the network-wide multisample VCF to quantify the extent of rare variation across known PGx haplotypes. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | none |
| **Desired data** | PGx BAMS and VCFs |
| **Planned Statistical Analyses** | We will perform population-level analyses of the eMERGE-PGx dataset as a whole including the generation of PCA plots and allele frequency spectra. |
| **Ethical considerations** | There are no physical risks involved. |
| **Target Journal** | ? |
| **Milestones\*\*** | Project duration: five months  First draft: June 2016 |