**­eMERGE Network Proposal for Analysis**

­Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Reference Number** | NT258 |
| **Submission Date** | September 7th, 2017 |
| **Project Title** | Phenotype risk scores to estimate phenotypic effects of rare variants |
| **Tentative Lead Investigator (first author)** | Lisa Bastarache |
| **Tentative Senior Author (last author)** | Josh Denny |
| **All other authors** | Dan Roden, Robert Carroll, Jake Hughey, Michael Temple, Melissa Basford, and any eMERGE III authors interested in participating. |
| **Sites Involved** | All eMERGE III sites (uses eMERGE-Seq data) |
| **Background / Significance** | Mendelian diseases are often defined by multiple clinical phenotypes. We have developed a method to aggregate EHR phenotypes using Mendelian disease patterns, as captured in OMIM clinical descriptions. Using weighted phenotype groups, we are able to assign a phenotype risk score (PRS) to an individual based on their similarity to a Mendelian disease pattern. By applying this method to eMERGEseq data, we can estimate the pathogenicity of exonic variants with respect to the Mendelian diseases they cause. Because of the sensitivity achieved by aggregating phenotypes, this method may detect pathogenic variants as well as variants that have partial penetrance or reduced expressivity. In so doing, we may find more nuanced effects of rare variants beyond what is captured in current resources designed to facilitate variant interpretation.  Clinical descriptions in OMIM have been mapped to the Human Phenotype Ontology (HPO), and we have mapped phecodes (aggregated ICD codes) to HPO terms. We are currently developing methods to leverage lab values to inform phenotype risk scores.  Initial validation within VUMC data finds symptomatic diagnosed Mendelian diseases in the EHR (p<1e-40). A search for high PRS in 21,701 genotyped individuals uncovered 18 rare variants associated with phenotypes consistent with their Mendelian diseases. These findings validate the method as a tool to detect pathogenicity. The study also provides early demonstrate that Mendelian disease variants may underlie common diseases, some with severe outcomes such as transplants and cancers. A paper describing this is in review. |
| **Outline of Project** | 1. Map ICD codes to phecodes (will use data from CC with permission) 2. Score individuals for target Mendelian disease based on clinical descriptions in OMIM 3. Use linear regression to detect association between rare variants and PRS 4. Plan to make the results available publicly via an eMERGE website |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | **Covariates**  Age,Sex,Genetic ancestry\*  eMERGEseq data for all genes and variants with Mendelian phenotypes\*  **Phenotype information**  ICD9 and ICD10 codes\*  Ability to review charts for subset of individuals found to have severe manifestations of disease (optional) |
| **Desired data** | Sequence data for genes associated with Mendelian diseases that can be profiled using PRS |
| **Planned Statistical Analyses** | Linear regression to model relationship between PRS and coding variants in genes. |
| **Ethical considerations** | There are no physical risks involved. |
| **Target Journal** | Nature Genetics, AJHG, PLoS Genetics |
| **Milestones\*\*** | 1 year for completion |

**\*\*** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.