**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Reference Number** | NT264 |
| **Submission Date** | 11/30/2017 |
| **Project Title** | A PheWAS approach to identify clinical correlates of rare variation within genes linked to Hereditary Spastic Paraplesia |
| **Tentative Lead Investigator (first author)** | Adam Gordon |
| **Tentative Senior Author (last author)** | Gail Jarvik |
| **All other authors** |  |
| **Sites Involved** | All |
| **Background / Significance** | Hereditary Spastic Paraplesia (HSP) is a debilitating adult-onset neurogenetic disease that affects roughly 1-10 per 100,000 patients of European descent. The gene *SPG7* contains multiple causative alleles that collectively represent the most common neurogenetic disease-causing alleles among Europeans. Initially mapped to 16q24.3 via linkage analysis of a family with recessive HSP, *SPG7* encodes paraplegin, a protease critical to mitochondrial misfolded protein response.  Although several alleles within this gene have been linked to HSP, p.Ala510Val (rs61755320) in particular is thought to be the most common driver of HSP among European patients. Despite this strong association, penetrance of this allele in a larger patient population needs further investigation due to the disparity between expected homozygote prevalence and HSP incidence (20-40/100,000 and 1-10/100,000 respectively).  Though *SPG7* alleles may be the most common driver of this disease, over 41 genes have been associated with HSP, and the specific disease phenotypes observed in HSP patients are often driven by low frequency variation within these genes. Analysis of rare alleles within *SPG7* and other HSP-linked genes within the eMERGE cohort can not only clarify previous associations derived from family studies, but can also be especially valuable in unraveling the genetic underpinnings of HSP phenotype heterogeneity. |
| **Outline of Project** | 1. Acquire eI-III imputed array dataset (already collected) 2. Assign PheWAS groups 3. Assess phenotype association with variation within HSP genes 4. Write manuscript |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | **Phenotypes:**   1. PheWAS data outlined by Denny algorithm   **Covariates:**  Demographic: Height, weight (at visit), sex, self-identified race and decade of birth. (already collected) |
| **Desired data** | eI-III imputed array dataset, PheWAS codes, Covariates, demographic information |
| **Planned Statistical Analyses** | **Association:**  Association tests will be performed using logistic regression (case control of PheWAS group) in R with the PheWAS package, adjusted for decade of birth, self-reported sex, study site, and the first three principal components. We will also compare results from these standard PheWAS methods to SPA, a newly developed parallel methodology optimized for phenotypes with low case:control ratios (PMID: 28602423). |
| **Ethical considerations** | There are no physical risks involved. |
| **Target Journal** | ? |
| **Milestones\*\*** | Project duration: five months |