|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT379 | |
| **Submission Date** | 2-19-2020 | |
| **Project Title** | Assessing the frequencies of individual ACMG evidence criteria usage in clinical variant interpretation across self-reported race and disease domains | |
| **Tentative Lead Investigator** *(first author)* | Meghan McKenna, Adam Gordon | |
| **Tentative Senior Author**  *(last author)* | Adam Gordon | |
| **All Other Authors** |  | |
| **Sites Participating** |  | |
| **Background / Significance** | Despite the ease and relative cheapness of genetic sequencing, there remains a dearth of data regarding the genomic landscape of people with ancestral origins outside of Europe (Manrai et al, 2016, Caswell-Jin 2016, Landry et al. 2018, Sirisena, N. D., & Dissanayake, V. H. 2017). This lack of information creates challenges for classifying mutations and there is a higher incidence of detection of VUSs in understudied populations (Casell-Jin, 2016). Within the healthcare system this could translate into large groups of people being systemically underserved by genomic medicine.  A particular repercussion of missing data is the misclassification of variants during clinical variant interpretation. A 2016 study found that black Americans had been told that they carried pathogenic mutations which increased their risk for hypertrophic cardiomyopathy (Manrai et al. 2016) that were later reclassified as benign. These false positive results invalidated risk assessments made for the patients and their families. Health management plans were created from these incorrect diagnoses that later had to be changed. The combination of presumably high rates of VUSs and little data on misclassified variants are red flags that people in historically understudied minorities are at risk of being under or over treated for genetic conditions (Manrai et al, 2016, Landry et al 2018). This incorrect management could mean increases health inequities by producing healthcare outcomes that are avoidable and unfair (Senier 2018).  It has been suggested that these inequities can be mitigated through increased sequencing and study of understudied minorities. Ancestry-matched controls are also recommended for interpreting variants. Furthermore, increasing data sharing has the potential to reduce these disparities (Manrai et al, 2016). While these recommendations are useful, they are not perfect. Sequencing and classifying new populations is time intensive. Additionally, waiting for ancestry-matched controls may delay reporting of pathogenic variants. Another possible solution to improper variant classification is an assessment of the underlying ACMG variant interpretation standards and guidelines. Though improving through data sharing initiatives, testing laboratories can lack inter-laboratory interpretation consensus (Amendola, L. M et al 2016, Harrison, S. M., et al. 2017). While these evidence codes apply to each variant and ancestry group equally, they retain the potential to propagate disparate impacts; it is possible that a lack of underlying data from the literature in variants unique to individuals of a variety of ancestral origins may cause ACMG evidence codes to apply to different populations disproportionately. This could be a driver of disparities in genetic medicine.  The goal of this study is to investigate whether there are disparities in clinical variant interpretation due to varying applicability of ACMG variant interpretation standards in different self-reported races, disease domains, and pathogenicity. Understanding different rates of criteria use has the potential to identify any structural biases in the design and implementation of ACMG evidence rules to improve their applicability in future revisions. | |
| **Outline of Project** | * collect/re-engineer ACMG criteria used for all reported variants * harmonize criteria, stratify by disease domain and self-reported race * per-criteria statistical analysis comparing usage frequencies by disease domain across races   + down-sample findings in Whites to match numbers among non-whites, bootstrap | |
| **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)*** | ‘Evidence summary’ text for all reported (P/LP) variants in eMERGE-III  Criteria used for each variant interpretation, where available | |
| **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  No | |
| **Planned Statistical Analyses** | Chi-square, to measure statistical significance for comparisons of stratified frequencies of criteria usage | |
| **Ethical Considerations** | n/a | |
| **Target Journal** | Genetics in Medicine | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Collect/re-engineer criteria = March/April 2020  Statistical analysis = April 2020  Manuscript prep = May 2020 | |

**\*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