|  |  |
| --- | --- |
| eMERGE Network: Proposal for Analysis  Project/Manuscript Concept Sheet | |
| Reference Number | NT334 |
| Submission Date | February 25, 2019 |
| Project Title | EHR And Genetic EoE Risk (EAGER): Integrating EHR and genetics to identify EoE risk and define subphenotypes |
| Tentative Lead Investigator  *(first author)* | Benjamin Kuhn, Geisinger (brkuhn@geisinger.edu) |
| Tentative Senior Authors  *(last author)* | Anne Justice, Geisinger (aejustice1@geisinger.edu) |
| All Other Authors | Geetha Chittoor, Nephi Walton, Darren Johnson, Marc Williams, Jennifer E. Below |
| Sites Involved | All Interested eMERGE sites |
| Background / Significance | * Eosinophilic esophagitis (EoE) is a debilitating disease of the gastrointestinal (GI) tract. * In young infants and children, it manifests as vomiting, trouble feeding, and failure to thrive. As the children age, their difficulties with eating food continue, and they often experience substantial physical and psychological symptoms that impair their daily activities. * As adults, patients with untreated EoE often experience food dysphagia, requiring   endoscopy to remove food stuck in the patient’s inflamed esophagus.   * EoE is considered a rare disease, it occurs in 4-15/10,000 people and is the leading cause for chronic dysphagia in adults. * To date, genome-wide association studies (GWAS) have identified 10 genome- wide significant single nucleotide polymorphisms (SNPs) in 7 loci that increase a patient's risk of developing EoE. Two well-replicated loci highlight increased EoE risk through the genotype-dependent expression of *TSLP* and *CAPN14*. * However, none of the previous studies considered structural variants, such as copy number variants (CNVs), took advantage of high-density imputation or sequence data to narrow in on causal variants within these loci, or examined genetic risk in non-European populations. |
| Outline of Project | * We will conduct a meta-analysis of Geisinger’s DiscovEHR data with participating eMERGE sites (excluding Vanderbilt separately from BioVU to not overlap samples in our discovery and replication data sets; VUMC participant is welcomed) to identify genetic susceptibility variants and genomic regions for EoE, including SNPs, single nucleotide variants (SNVs), CNVs, and perform aggregate tests of association (i.e. burden-type, gene-wise tests), while explicitly accounting for sex- specific effects. * We will perform phenome-wide association study (PheWAS) for EoE susceptibility variants in the DiscovEHR and broader eMERGE consortia. * We will replicate significant findings in BioVU. |

|  |  |
| --- | --- |
|  | * This investigation expands further than previous EoE GWAS efforts in eMERGE and other published studies, as it includes 1) replication of significant findings; 2) includes low-frequency/rare variants (incl. gene-based); 3) CNVs; 4) X and Y- chromsome; 5) sex-interaction; 6) and PheWAS in secondary analyses. |
| Desired Variables *(essential for analysis indicated by \*)* | * Age (at each encounter) * Sex * Medications (Yes/No flag) and date of prescription   + “GERD” medications:     - H-2 ANTAGONISTS     - PROTON PUMP INHIBITORS   + “Asthma” medications     - ASTHMA AND BRONCHODILATOR AGENT COMBINATIONS     - BRONCHODILATORS – ANTICHOLINERGICS     - STEROID INHALANTS   + “Allergy” medications     - ANTIHISTAMINES - NON-SEDATING     - NASAL STEROIDS     - LEUKOTRIENE MODULATORS Asthma   + ICD-10-CM K20.0, ICD-9-CM 530.13 - Eosinophilic esophagitis   + Age at diagnosis (e.g. age at 1st code for ICD-10-CM K20.0, ICD-9-CM 530.13 - Eosinophilic esophagitis)   + All ICD-9/10 codes   + Flag for relevant procedures     - Esophagogastroduodenoscopy (EGD), flexible, transoral; with biopsy = CPT 43239     - Esophagoscopy, flexible, transoral; with biopsy = CPT 43202     - EGD with Foreign Body Removal = CPT 43247     - Esophagoscopy with Foreign Body Removal = CPT 43215 |
| Desired Data | Emerge I-III imputed data Targeted exome sequencing CNV calls |
| Planned Statistical Analyses | Quality control for the imputed data, including imputation score cutoff and effective N>30. We will drop out any eMERGE samples that are in the current DiscovEHR data we are using at Geisinger.  For each association model relating to a diagnosis, we will use age, sex, and ancestry matching to randomly select controls at a minimum ratio of 1:5, cases:controls. We will give preference to a minimal model, adjusting each model for sex (in combined analysis), and account for relatedness as a random group factor. We will also perform sensitivity analyses adjusting for other covariates that may be relevant across the PheWAS spectrum (e.g. median age). |
| Ethical Considerations | All data will be de-identified, and only summary data will be shared in resultant manuscripts |
| Target Journal | PLOS Genetics, American (or European) Journal of Human Genetics |

|  |  |
| --- | --- |
| Milestones\*\* | March 2019 for concept sheet approval; January 2020 for completion of meta-analyses, May 2020 manuscript draft. August 2020 for second draft of paper, October 2020 for submission of paper. |

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