|  |  |
| --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | |
| **Reference Number**  *(to be assigned by CC)* | NT318 |
| **Submission Date** | 12/18/2018 |
| **Project Title** | Genetic variants associated with C.Difficile Infection (CDI) |
| **Tentative Lead Investigator** *(first author)* | Yanfei Zhang |
| **Tentative Senior Author**  *(last author)* | Ming Ta Michael Lee |
| **All Other Authors** | Manu Shivakumar,  Vida Abedi, |
| **Sites Participating** | All sites |
| **Background / Significance** | *C.Difficile infection* (CDI) is a critical healthcare problem in both hospital and community settings. Host, pathogen, and environmental factors all have important contributions towards the pathogenesis of CDI. Previous studies showed that old age, use of certain antibiotics and medications, GI surgery, and hospitalization are risk factors for CDI. In addition, studies looking at the genomes of the bacteria found large genetic heterogeneity in C.diff between different cases [1]. However, the genetics study on the host is limited. To date, there is only one genetics study of CDI in patients who had Ulcerative Colitis [2]. In this study, only SNPs that confer susceptibility to either Ulcerative Colitis or Crohn’s disease were investigated. No GWAS study of CDI have been reported.  At Geisinger, CDI cases are identified based on lab test results. Patient who were tested positive (either toxin A or toxin B or both or present of antigen) were considered as cases. Controls were identified as patients who had the CDIFF test with negative results. We identified about 1200 cases and 7000 controls and performed GWAS adjusted for age, sex, BMI and first 6 PCs. We found multiple loci with P<10-5, including signal from HLA region. We would like to perform a replication study using eMERGE data given the CDIFF algorithm is already installed. It will add more impact if HLA allele information is available.  References:  [1] Diverse Sources of C. difficile Infection Identified on Whole-Genome Sequencing. David W.Eyer et al. NEJM 2013, Vol. 369, No.13  [2] genetic risk factors for clostridium difficile infection in Ulcerative Colitis. Aliment Pharmacol Ther. 2013 September ; 38(5): 522–530. doi:10.1111/apt.12425. |
| **Outline of Project** | 1. Discovery GWAS using Geisinger MyCode data 2. Replication study using eMERGE data. |
| **Desired Data - Common Variables\***  *(Available from the CC)* | * Demographics * BMI * CDIFF phenotype algorithm case/control status. |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)* |
| **Desired Genetic Data** | * eMERGE I-III Merged set (HRC imputed, GWAS) * Other (not listed above): if HLA type is available |
| **Does project pertain to an existing eMERGE Phenotype?** | Yes, if so please list: CDIFF algorithm |
| **Planned Statistical Analyses** | We will run the logistic model adjust for age, sex, BMI and PCs. |
| **Ethical Considerations** | None |
| **Target Journal** | BMC medicine, PLOS genetics, |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | The draft of GWAS of CDI using Geisinger MyCode data is to be ready by the end of December.  Analyses can begin immediately as data is made available. We can update the current manuscript by including the eMERGE analyzes. We anticipate the manuscript by Q1 2019. |