|  |  |
| --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | |
| **Reference Number**  *(to be assigned by CC)* | NT312 |
| **Submission Date** | October 23, 2018 |
| **Project Title** | Genetic variants associated with Osteoarthritis (Pan-OA analysis) |
| **Tentative Lead Investigator** *(first author)* | Yanfei Zhang |
| **Tentative Senior Author**  *(last author)* | Ming Ta Michael Lee |
| **All Other Authors** | Steven A. Lietman,  Thomas R. Bowen,  Manu Shivakumar,  Vida Abedi, |
| **Sites Participating** | All sites |
| **Background / Significance** | Osteoarthritis (OA) is the most common form of arthritis and is a leading cause of disability. Its prevalence and severity are increasing due to the aging population. It is also a major economic burden with an estimated medical expenditure of more than $62 billion annually in US alone. Currently, the treatment for OA is mainly pain management or joint replacement surgery. There is no treatment to stop the disease progression. Several genetics study on OA were published from arcOGEN, deCODE, and UKBB. Novel loci have been identified and need to be replicated. This proposal will be a large OA genetics study in US. We expect to firstly validate the reported loci, secondly to identify novel loci, thirdly we expand the study of hand OA. We also want to identify rare variants with larger effect size to OA. |
| **Outline of Project** | Five groups will be analyzed, knee OA, hip OA, knee/hip OA, hand OA, and Pan-OA.  ICD-codes based diagnosis of OA will be used to identify cases.  Controls should not have any codes of any OA, RA, and other related skeletal disorders.  The result will be meta-analyzed with Geisinger MyCode data. |
| **Desired Data - Common Variables\***  *(Available from the CC)* | * Demographics * BMI * ICD codes |
| **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) * 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** | We will focus on five groups of analysis,   1. Knee OA 2. Hip OA 3. Knee and/or Hip OA 4. Hand OA 5. Pan-OA     We will perform a genome-wide single-variant association analysis using logistic regression modeling. Analyses will be performed for all common variants with a minor allele frequency greater than 1% that pass relevant quality criteria (e.g., imputation quality). Variants will be evaluated under an additive genetic model adjusting for age, sex, BMI, study site, ancestry-informative principal components (PCs) capturing population substructure. If appropriate, sex, race/ethnicity-stratified analyses will be conducted and combined via meta-analysis. We will also perform gene-level aggregate testing of rare variants using the sequence kernel association test (SKAT) using MAF-based weighting and subsetting to variants with predicted functional impact via available annotation resources. |
| **Ethical Considerations** | None |
| **Target Journal** |  |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Analyses can begin immediately as data is made available. We anticipate a first draft of the manuscript by Q1 2019. |