|  |  |  |
| --- | --- | --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT346 | |
| **Submission Date** | 05/20/2019 | |
| **Project Title** | Genetics of Osteoarthritis Consortium | |
| **Tentative Lead Investigator** *(first author)* | Yanfei Zhang | |
| **Tentative Senior Author**  *(last author)* | Ming Ta Michael Lee | |
| **eMERGE Site Sponsor & Contact** | Geisinger; Ming Ta Michael Lee | |
| **All Other Authors** | Osteoarthritis consortium sites authors can be found at: https://www.genetics-osteoarthritis.com/people/people/index.html | |
| **Sites Participating** | All sites in eMERGE  Osteoarthritis consortium sites participating sites can be found at: https://www.genetics-osteoarthritis.com/contributing-studies/contributing-studies/index.html | |
| **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. The Genetics of Osteoarthritis (GO) consortium is a global collaboration with a focus on progressing our understanding of the genetic underpinning of osteoarthritis and related traits. GO aims to bring together all globally available genetic studies of osteoarthritis in order to make new discoveries possible. More than 15 sites have participated in this consortium with more than half million OA patients and controls available for analysis. Additional sites are being recruited for replication analysis. | |
| **Outline of Project** | Six groups will be analyzed, knee OA, hip OA, knee/hip OA, hand OA, spine 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 GO data. | |
| **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)*** | *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)*  N/A | |
| **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 six groups of analysis,  1. Knee OA  2. Hip OA  3. Knee and/or Hip OA  4. Hand OA  5. Spine 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. Post meta-analysis analysis such as genetic correlation with epidemiologically linked traits, polygenic risk scores, mendelian randomization analyses and PheWAS will also be carried out. Individual data will be analyzed at Geisinger, only summary level data will be shared with the consortium. | |
| **Ethical Considerations** | None | |
| **Available Funding or Resources** | N/A | |
| **Target Journal** | TBD | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Analysis will begin immediately once data is available. We anticipate to have draft manuscript by the end of 2019. | |

**\*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