|  |
| --- |
| **eMERGE Network: External Collaborator Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT308 |
| **Submission Date** | 10/05/2018 |
| **Project Title** | Genome-wide association studies of spinal disorders and pain |
| **Tentative Lead Investigator** *(first author)* | Pradeep Suri, MD, MS (external collaborator, University of Washington) |
| **Tentative Senior Author** *(last author)* | Gail Jarvik, MD, PhD |
| **eMERGE Site Sponsor & Contact** | Gail Jarvik, MD, PhD (Sponsor); Melody Rynerson Palmer (Contact) |
| **All Other Authors**  | Melody Rynerson Palmer, PhD; David Crosslin PhD; other interested eMERGE investigators |
| **Sites Participating** | All |
| **Background / Significance** | Low back pain (LBP) and spinal disorders cause more disability than any other health condition worldwide.1 Our recent work in the CHARGE consortium and UK Biobank has identified various genetic variants associated with LBP and new insights into its genetic architecture.2,3 However, this work indicated the need for more refined LBP phenotyping and the potential value of examining more specific spine-related phenotypes, such as available through cohorts and biobanks linked to electronic medical record (EMR) data. There is already an extensive literature using ICD-9/10 diagnostic codes and other EMR data for defining LBP and spine-related phenotypes.4-7Our aims are to 1) discover genetic variants associated with LBP and more specific spine related phenotypes (e.g. lumbosacral radiculopathy/radicular syndrome [LSRS], symptomatic lumbar spinal stenosis [SLSS]) in the electronic Health Records and Genomics (eMERGE) Network; 2) to develop polygenic risk scores for these LBP and spine-related phenotypes. We will perform genome-wide association studies with genotypes imputed to the Haplotype Reference Consortium, including adults age 18 years and older. We plan to replicate our findings in other datasets, such as the UK Biobank.  |
| **Outline of Project** | For this work, we expect to use eMERGE data that is currently available from the coordinating center.1. Use of ICD9/10 and CPT codes to define LBP and spine phenotypes. 4-7
2. Perform GWAS analysis using ~99,000 imputed merged data set
3. Create candidate polygenic risk scores (PGSs)8
4. Attempt replication of results for #2 and #3 in external datasets

A future goal, beyond the scope of the research proposed in this concept sheet, is to obtain grant funding that will allow us to develop a more accurate phenotyping algorithm using a broader range of EMR data, and run the algorithm at relevant (adult) sites.  |
| **Desired Data - Common Variables\*** *(Available from the CC)* | [x] Demographics [x] ICD9/10 codes[x] CPT codes[ ] Phecodes[x] BMI | [x] Common Variable Labs[x] Common Variable Meds[ ] Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | None other than above |
| **Desired Genetic Data** | [x] 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 [x] No |
| **Planned Statistical Analyses** | * We will use existing ancestry principal components analysis across all samples, and within the largest ancestry subsets
* Phenotypes of interest are as listed below; all will be defined using existing algorithms that use ICD-9 codes.5-7
1. Non-specific LBP
2. ‘LSRS’ or Lumbosacral radiculopathy/radicular syndrome
3. ‘SLSS’ or symptomatic lumbar spinal stenosis
* We will conduct sensitivity analyses excluding LBP cases likely due to rare/specific causes (tumor, infection, malignancy).
* We will conduct secondary analyses examining failure of non-surgical treatment within each subgroup, defined using CPT codes reflecting any lumbosacral spine surgery (phenotype 1), decompression surgery (phenotypes 2 and 3), and subsequent revision surgery.
* Logistic regressions of imputed SNPs with an additive genotype model in PLINK 1.9.
	+ With all samples, including adults age 18 years or older
	+ Within ancestry subsets
	+ Adjusting for sex, age, ancestry principal components
	+ Secondary analyses adjusting for height, weight, BMI
* Derive PGSs using LDPred vs. pruning/thresholding methods in PLINK
* Replication using logistic regression in UK Biobank
 |
| **Ethical Considerations** | Using de-identified data only, there are no ethical concerns. |
| **Available Funding or Resources** | Analyses by Dr. Palmer are funded through a career development award (IK2RX001515) to Dr. Suri. We expect to submit grant applications that will fund future work and involve the team participating in this current proposed project.  |
| **Target Journal** | We expect 1 manuscript, or 2 manuscripts if findings for genome-wide significant variants and PGSs are reported separately. Target journals include Pain, Journal of Pain, PLOS Genetics |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Dates of expected completion 1. Planning/application of phenotype algorithm and data cleaning - October 2018
2. Perform GWAS analysis - November 2018
3. Create candidate polygenic risk scores (PGSs) - February 2019
4. Attempt replication of results for #2 and #3 in external datasets- May 2019
5. Send manuscript to co-authors- September 2019
6. Submit manuscript for publication- October 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