|  |
| --- |
| **eMERGE Network: Manuscript Concept Sheet** |
| **Reference Number** *(to be assigned by CC)* | NT464  |
| **Submission Date** | 11/8/2022 |
| **Project Title** | Investigating the shared genetic background between Parkinson’s disease and malignant melanoma |
| **Tentative Lead Investigator** *(first author)* | Bernabe Bustos |
| **Tentative Lead Investigator Email Address** | bernabe.bustos@northwestern.edu |
| **Tentative Senior Author** *(last author)* | Steven Lubbe (steven.lubbe@northwestern.edu) |
| **All Other Authors**  | Dimitri Krainc (Northwestern)Rachel Lewandowski (Northwestern)Alejandro Hernandez (Northwestern)Jing Hu (Northwestern)Talia Krainc (Princeton/Northwestern)Northwestern University Parkinson’s disease and Movement Disorder Center BiorepositoryBruno Benitez (WUSTL/Harvard)Carlos Cruchaga (WUSTL)Zin Gan-Orr (McGill University, Canada)Huw Morris (UCL, UK)Raquel Real (UCL, UK)Matthew Law (Leeds University, UK)Julia Newton-Bishop (Leeds University, UK)Megan Puckelwartz (Northwestern)Elizabeth McNally (Northwestern)Jennifer Pacheco (northwestern University) |
| **Sites Participating** | Northwestern University, WUSTL, Harvard, McGIll University Canada, UCL UK, Leeds University UK.All proposed analyses will be performed locally at Northwestern University |
| **Background / Significance** | Epidemiological studies have shown that individuals with Parkinson’s disease (PD) are at an increased risk for malignant melanoma (MM), and that the reverse is true. These reciprocal increased risks are also seen for first- and second-degree relatives of people with PD or MM. Several additional features are shared between the diseases including the involvement of pigmented cells (substantia nigra neurons in PD and melanocytes in MM); both cell types originate from the neural crest during embryogenesis; and a relationship between hair/eye/skin colour and disease risk (both are increased in fairer individuals). Together these data suggest that these diseases may have a shared genetic background. Previous genetic studies looking at GWAS associated common variants (MAF>1%) have found limited evidence to support this suggesting that this shared genetic risk may be mediated by rare variants (MAF<1%). However, rare variant studies have been limited, have mostly been candidate studies and have sometimes produced conflicting results. An unbiased assessment of the contribution of rare variants in mediating the shared genetics of PD and MM is lacking. We undertook an exome sequencing study assessing rare variants under various inheritance models in about 100 individuals with both PD and MM, or individuals with one disease but with a strong family history of the other disease. Candidate variants and genes with increased rare variant burden were then assessed in several replication PD-MM, PD only or MM only case-control cohorts under observed inheritance models. |
| **Outline of Project** | Here we propose to utilize the eMERGE data to identify individuals with (i) PD and MM, (ii) PD only, (iii) MM only, and (iv) controls (no evidence of either PD or MM) to perform additional rare variant- and gene burden-based replication studies under different inheritance patterns and to undertake expanded meta-analyses of all replication cohorts. We aim to identify rare variant signatures that predispose carriers to both PD and MM.  |
| **Desired Data - Common Variables\*** *(Available from the CC)* | [x] Demographics [x] ICD9/10 codes[x] CPT codes[x] Phecodes[x] BMI | [ ] Common Variable Labs[ ] Common Variable Meds[ ]  Geocoding 2015 ACS variables’[ ] 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)* If sites have resources available to additionally identify these subsets: (i) PD with strong family history (first-degree relatives) of MM, and (ii) MM with family history of PD (first-degree relatives) |
| **Desired Genetic Data** | [ ] eMERGE I-III Merged set (HRC imputed, GWAS)[ ] eMERGE PGx/PGRNseq data set [ ] eMERGEseq data set (Phase III)[x] eMERGE Whole Genome sequencing data set[ ] eMERGE Exome chip data set[x] 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** | Sample and variant level quality controls, including: identify by descent to detect cryptic relationships; Principal component analyses to detect population outliers; Heterozygosity and missingness analysis to detect low quality sequenced individuals; variant frequency filtering to obtain rare variants (frequency < 1% in the general population); variants with depth of coverage > 10, and genotype quality > 20. We will perform variant annotation and prioritization, to select rare damaging protein altering variants, including splicing and loss of function variants. Then we will select cases and controls in independent files, and we will perform variant and gene-wise burden of the selected variants. We will select significant variants after multiple testing correction (FDR < 0.05) for follow-up meta-analysis based replications. |
| **Ethical Considerations** | All genetic and clinical information should be de-identified and removed of HIPAA classified identifiers. No attempt to re-identify participants will be undertaken.  |
| **Target Journal** | Human Molecular Genetics, Neurobiology of Aging, Brain |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | We estimate that after obtaining the data, the project will be completed in 3-4 months. A breakdown of timelines for completion of each processing step:Download data and sample-variant level QCs: 1 month.Variant annotation and prioritization: 2 weeks.Gene-wise burden of rare damaging variants in PD and MM independent case-control cohorts: 2 weeks. Draft preparation: 1 month. Author reviews and submission: 1 month.  |

**\*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