|  |  |
| --- | --- |
| **eMERGE Network Analysis**  Project/Manuscript Concept Sheet | |
| **Reference Number** | NT291 |
| **Submission Date** | 6/4/2018 |
| **Tentative Lead Investigator** *(first author with contact information and affiliation)* | Shefali Setia Verma, [shefali.setiaverma@pennmedicine.upenn.edu](mailto:shefali.setiaverma@pennmedicine.upenn.edu) |
| **Tentative Senior Author**  *(last author)* | Marylyn D. Ritchie, mdritchie@geisinger.edu |
| **eMERGE Site Sponsor & Contact** | University of Pennsylvania, Marylyn Ritchie |
| **Project Title** | How genomic patterns are linked to phenotypic patterns? |
| **All Other Authors** | TBD |
| **Other eMERGE Sites Involved** | All eMERGE network |
| **Background / Significance** | LD patterns are influenced by several factors such as genetic drift, population structure, recombination and epistatic selection. Identifying associations of epistatic variants (non-linear effect) with phenotypes is of great interest but it comes with several challenges. One of the biggest challenge is to identify which models to test. We hypothesize that pairwise variants that are in LD due to epistatic selection might be associated with multiple phenotypes. |
| **Outline of Project** | 1. **Haplotype block identification:** Running PLINK to identify haplotype blocks 2. **Ohta’s Dstat analysis:** For each haplotype block, calculate Ohta dstat analysis using ‘ohtadstats’ R package 3. **Selection of epistatic models:** Following Ohta’s theory, we would like to identify epistaic models based on Dstats 4. **Association testing with ICD-9 codes:** Regression for selected models against all ICD-9 codes using PLATO |
| **Desired Variables**  *(essential for analysis*  *indicated by* ***\*****)* | We seek the following variables for our analyses:   * Primary phenotypes\*: ICD-9 codes * Confounder variables\*: age, year of birth, sex, and race/ethnicity |
| **Desired Data** | eMERGE-III HRC imputed data |
| **Planned Statistical Analyses** | 1. Imputation and quality control. 2. Ohta D stats calculations 3. Post processing of Dstats to select epistatic models 4. SNP-SNP interaction testing on selected models using PLATO with all ICD-9 codes |
| **Ethical Considerations** | Genomics data and phenotypic data will be de-identified to protect confidentiality. |
| **Target Journal** | TBD |
| **Milestones\*\*** | 1. Complete QC in May 2. Run analyses in June 3. Write relevant portions of the manuscript by end of Fall 2018. |

***\*\**** *This section should include the timeline for completion of project, including: approval, project duration, first and second draft of the paper and submission.*

*References:*

1. Ohta T (1982a). Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc Natl Acad Sci USA 79: 1940–1944.
2. Ohta T (1982b). Linkage disequilibrium with the island model. Genetics 101: 139–155.
3. Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM et al. (2003). Evolutionary genomics of inversions in Drosophila pseudoobscura: Evidence for epistasis. Proc Natl Acad Sci USA 100: 8319–8324.
4. TM Beissinger, M Gholami, M Erbe, S Weigend, A Weigend, N de Leon, D Gianola and H Simianer (2016). Using the variability of linkage disequilibrium between subpopulations to infer sweeps and epistatic selectiTBAon in a diverse panel of chickens. Heridity(@016) 116,158-166