**APPENDIX 2**

***External Collaborator Proposal* for eMERGE Network Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | January 20, 2017 |
| **Tentative Lead Investigator (first author with contact information and affiliation)** | Reyna L. Gordon, PhD  Assistant Professor, Department of Otolaryngology  Assistant Professor, Department of Psychology  Associate Director, Program for Music, Mind and Society at Vanderbilt  Vanderbilt University Medical Center  1215 S. 21st Ave South, MCE 10267  Nashville, TN 37232  615-322-5550  Email: [reyna.gordon@vanderbilt.edu](mailto:reyna.gordon@vanderbilt.edu) |
| **Tentative Senior Author (last author)** | Nancy Cox, Ph.D.  Director, Vanderbilt Genetics Institute |
| **eMERGE Site Sponsor & Contact** | VGER |
| **Project Title** | New approaches to the genetic basis of developmental language disorder |
| **All other authors** | Xue Zhong (VUMC) and J. Devin McAuley (Michigan State)  We are open to collaborating with other individuals from other sites who contribute data to the project and participate in the development of the study design. |
| **Other eMERGE Sites Involved** | We request involvement from all sites in the eMERGE network that have deposited records/GWAS data into the database that correspond to our phenotypes. |
| **Background / Significance** | Specific language impairment (SLI) is a neurodevelopmental language disorder characterized by language acquisition delay and continued difficulties with grammatical structures and vocabulary in the absence of general slow development, hearing loss, physical abnormality of the speech apparatus, autism spectrum disorder, and acquired brain damage. These communication struggles have a negative impact on long-term emotional/behavioral, academic, social, and economic outcomes. Although SLI continues into adulthood, research focused on the identification and treatment of SLI in elementary school aged-children is essential in order to improve long-term outcomes. Population health approaches to SLI have estimated its prevalence at a staggering 7.4% rate of kindergarteners (Tomblin et al, 1997), yet was previously identified in only 29% of this subset. A recent Institute of Medicine report on the public health relevance of speech and language disorders has called attention to the urgent need to improve identification and availability of services to treat SLI, which is often co-morbid with other disorders such as reading disability and ADHD that also have academic and social consequences.  Similar to many complex neurological disorders, the etiology of SLI appears to have a strong polygenic basis (Newbury et al, 2010). Obtaining a greater understanding of the genetic basis of SLI can help us develop a genetic screening tool for detecting SLI in patients. We plan to conduct a genome-wide association study (GWAS), a non-candidate-driven examination of a genome-wide set of genetic variants in a population aimed at determining if variants are associated with a particular phenotype, e.g. SLI.  The proposed GWAS study is part of a broader line of research aimed at examining a potential shared genetic basis of rhythm and grammar skills in children. |
| **Outline of Project** | A multi-site genome-wide association study (GWAS) will be conducted by compiling data from extant GWAS studies and data from eMERGE and BioVU on Specific Language impairment (SLI) and language phenotypes. Association analyses will be conducted first in a cohort of eMERGE and BioVU data, and then combined in a meta-analysis with extant GWAS on SLI phenotypes. SNPs with the strongest signals differentiating language phenotypes will be utilized to generate polygenic predictor scores of SLI. This line of research is aimed at eventually created a novel clinical method of screening for SLI from individual DNA in patients. We will also utilize PrediXcan (Gamazon et al, 2015) to impute gene expression levels in samples from individuals with SLI and to identify additional candidate genes. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | EMRs with ICD-9/ICD-10 codes for developmental language disorder (~2355 cases)  Control dataset (4710); matched in age group, gender, and race.  Associated individual genotyped GWAS data  Age, gender, race/ethnicity (as covariates) |
| **Desired data** | Data analysis is expected to yield effect sizes for individual SNPs, and will be modeled to construct polygenic predictor scores of SLI (as described below). |
| **Planned Statistical Analyses** | For the eMERGE and BioVU data, we will use case-control analysis to determine which SNPs have the strongest differentiation between individuals with and without SLI. SNPs with the strongest signals will be used to calculate polygenic score or genetic risk score (GRS), which summarize genetic effects of a set of alleles within each subject. Association between a trait and this composite score implies the presence of a genetic signal among the selected markers (SNPs). This approach has been used to establish a common genetic basis for related disorders. Another advantage of using GRS is the gain of power through summing the risk alleles, which is particularly useful to obtain the evidence of genetic effects when no single locus achieves genome-wide significance. At the moment, the utility of GRS mainly lies in testing association rather than predicting complex traits as the useful level of prediction requires the predictors (SNPs) to be estimated from a greater larger sample size than currently available. Certain p-value thresholds (e.g., p<0.05) are used to include a marker in a score. Different selection thresholds for including a marker in the score will be explored. |
| **Ethical considerations** | Once accessed, all data will be stored at secured locations in the data storage systems at VUMC. No data will be shared with unauthorized third parties. Patient identity will remain confidential and will not be compromised by the proposed analysis. eMERGE guidelines will be followed at all times throughout this project. |
| **Available Funding or Resources** | * Planned R01 application to NIDCD (February 2017) * Collaboration with Dr. Nancy Cox at the Vanderbilt Genetics Institute * Vanderbilt University Trans-Institutional Program award (The Program for Music, Mind and Society at Vanderbilt) |
| **Milestones\*\*** | February 2017: project approval and R01 grant submission  April 2017: GWAS data access  December 2017: NIH R01 grant commences, if funded.   Commence GWAS analysis and interpretation  June 2018: draft of manuscript  December 2018: manuscript submission |

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