**eMERGE Network Proposal for Analysis**

Project/Manuscript Concept Sheet

|  |  |
| --- | --- |
| **Submission Date** | January 23 2017 |
| **Reference Number** | NT209 |
| **Project Title** | 22Q11.2 Deletion Syndrome, Leveraging Copy Number Variation to Examine Health Outcomes |
| **Tentative Lead Investigator (first author)** | P Sleiman |
| **Tentative Senior Author (last author)** | H Hakonarson |
| **All other authors** | B Almoguera, M Harr, F Mentch, L Vazquez, JJ Connolly, Genomics WG and other interested authors |
| **Sites Involved** | All |
| **Background / Significance** | Chromosome 22q11.2 deletion syndrome (22q11.2DS) is the most common chromosomal microdeletion disorder (approximately 0.7–3.0 million base pairs in size resulting in loss of ~90 known or predicted genes including 46 protein-coding genes and 7 microRNAs, 10 non-coding RNAs, and 27 pseudogenes) estimated to result mainly from *de novo* non-homologous meiotic recombination events occurring in approximately 1 in every 1,000 fetuses. About 4% of infants with 22q11.2DS succumb, while cardiac defects, hypocalcemia and airways disease are risk factors for early death, with median age at death of 3–4 months. However, many individuals with 22q11.2DS, survive well into adulthood. In this context, the syndrome has become a model for understanding rare and frequent congenital anomalies such as heart defects, medical conditions including immunodeficiency, allergy, asthma, psychiatric and developmental differences, which may provide a platform into better understanding these phenotypes, while affording opportunities for translational strategies across the lifespan for both patients with 22q11.2DS and for those with these associated features in the general population. The diverse phenotype and outcomes of nearly every organ system make this population valuable for understanding the variables that impact on the manifestations of the deletion, which is relatively consistent from person to person.  The eMERGESeq panel captures six SNPs (five in the *COMT* gene and one flanking the region), which can be used to capture 22q11.2DS, while existing genotype data can be readily used to detect the syndrome.  We propose to use these data to assess the prevalence of 22q11.2DS in these eMERGE cohorts, and to determine a health outcome across multiple organ systems and outcome measures as available.  Sites willing to return results from this study will likely be required to validate findings in a CLIA environment. |
| **Outline of Project** | We will use PennCNV and XHMM to derive CNVs from existing array and new eMERGESeq data. Data will be returned to participating sites for outcome evaluation of relevant phenotypes (e.g heart defects, immunodeficiency, allergy, asthma, psychiatric and developmental differences) and for additional validation if required. |
| **Desired**  **Variables (essential for analysis**  **indicated by \*)** | Existing congenital diagnoses, 22q11.2DS diagnoses. |
| **Desired data** | Signal intensity data from eMERGE SNP arrays (.idat data files are preferred).  PGRNSeq data (namely the *COMT*locus)  The following six SNPs from eMERGESeq are informative on 22q DS: rs740603 (chr22:19945177), rs6269 (chr22:19949952), rs4633 (chr22:19950235), rs4818 (chr22:19951207), rs4680 (chr22:19951271), and rs181362 (chr22:21932068). Any potential off target sequence variants residing within the 22q11.2 region are of added value. |
| **Planned Statistical Analyses** | Analyses of CNVs in array and sequencing data using PennCNV and XHMM. |
| **Ethical considerations** | Sites wishing to return results of analyses to patients will likely require CLIA validation. |
| **Target Journal** | AJHG or higher |
| **Milestones\*\*** | Analyses of PGRNSeq data: 2/31/17  Analyses of Array data: 4/31/17  Analyses of eMERGESeq data: TBD  Draft 1 of manuscript: 6/30/17  Submission of manuscript:8/31/17 |