Genome-wide analysis of copy number variation and common facial variation in African and European samples

Megan Sorenson[1], David Astling[2], Hung-Chun Yu[2], Feyza Yilmaz[2], Joanne Cole[2,3], Stephanie A. Santorico[1,3,4], Richard A. Spritz[2,3], Audrey E. Hendricks[1,3,4] and Tamim H. Shaikh[2,3]

[1]*Mathematical and Statistical Sciences,* [2]*School of Medicine,* [3]*Human Medical Genetics and Genomics Program University of Colorado-Denver,* [4]*Biostatistics and Informatics- Colorado School of Public Health*

Copy number variations (CNVs) are a significant source of genetic variation underlying human variability and disease. Recent studies have led to the association of CNVs to increased risk for several of common diseases, most notably neurodevelopmental diseases like autism and schizophrenia. Additionally, CNVs have been directly implicated in several genetic disorders leading to multiple birth defects, such as the 22q11 deletion syndrome, in which affected individuals exhibit characteristic facial dysmorphic features. We hypothesize that some of the variability in normal facial shape and appearance results from genetic variation mediated by the CNVs affecting the dosage of genes involved in facial development. In two ongoing studies to identify the genetic determinants underlying facial shape variation, 3700 Bantu Africans and 3200 European-derived Caucasians (EUR) were photographed using 3D morphometric cameras to obtain digitized facial scans and genotyped using high-content genotyping microarrays. The precise facial measurements obtained from 3D facial scans, allowed the extraction of quantitative facial distances and shapes which account for the majority of facial shape variance among study subjects. We propose to perform genome-wide CNV association analysis on facial variation using these cohorts.

We have completed calling the CNVs in the Bantu samples using three CNV calling algorithms: PennCNV, DNAcopy, and VanillaICE. We are currently in the process of determining the set of high quality samples and CNVs that can proceed to the analysis of the relationship between CNVs and face shape. Our overall goal is to identify genetic determinants of facial shape and appearance in order to elucidate the biological pathways underlying facial morphogenesis. Our initial analysis will consist of a single variant association analysis for CNVs at a frequency above 1% to identify low frequency or common CNVs that are associated with specific facial characteristics. Gene-based association analysis will then be carried out for both cohorts using CNVs that occur at a frequency of < 5% and are overlapping an annotated gene or a 20 kb flanking region to identify genetic associations driven by several rare or low-frequency CNVs within the same gene-region. Finally, we will complete an enrichment analysis in the observed versus the expected number of CNVs within or overlapping genes within pre-defined gene sets, including a set of genes that have been previously implicated in either craniofacial development or disorders, as well as curated gene sets not known to be associated with facial morphology.