Project Number: CC021

Title: Validation of the cleft phenotype in the Cleft Collective

PI: Amy Davies

Affiliation: The Cleft Collective, The University of Bristol

Approval Date: January 2020

Scientific Outline:

Previous research has shown different subtypes of orofacial cleft have distinct aetiologies and epigenetic profiles[1]. It is, therefore, important that the phenotype is recorded correctly to ensure research carried out is as accurate as it can be.

The Cleft Collective obtain different levels of phenotype from a range of sources including parental questionnaires, biological transfer forms, surgical questionnaires and directly from medical records. In addition, the Cleft Collective (CC) have obtained a large number of phenotypes through work carried out by The Cleft Multidisciplinary Collaborative (CMC)[2] and further work is underway to obtain phenotype data through CRANE[3] and directly from medical records. It has become clear throughout the process of cleaning these data that there are a small number of cases where the phenotype is discrepant between sources.

The aim of this project is to identify the extent of discrepancies present and to validate the orofacial cleft phenotype recorded in the Cleft Collective. In turn this will improve the robustness of the data.

Phenotype data from all sources (CC parent questionnaires, CC forms accompanying biological samples, CMC, CRANE, medical records) will be collated into one dataset. Where possible, phenotype will be compared using the most detailed classification (LAHSHAL). Where phenotype is collected in a less detailed format to LAHSHAL, the LAHSHAL classification will be broken down to the corresponding phenotype. For example, parents in the CC report their child's cleft classification in terms of type of cleft (Cleft lip ± palate) and the laterality of the cleft (Bilateral or Unilateral; Right/Left) through parent questionnaires, whereas medical records might report the cleft using the LAHSHAL classification. So a classification collected through access to medical records recorded as "LAHSH—" will be recoded into a new variable as a "Right sided Unilateral Cleft Lip and Palate" to enable us to compare the medical record with the parental report.

Descriptive statistics will be performed to report the frequencies of phenotypes per source, the number of matched phenotypes and the number of discrepancies present. Inter-rater validity will be explored using generalised measures of agreement (Fleiss' K and Krippendorff's alpha).

A derived variable will be created for each record to determine final phenotype. Where phenotypes have differed across sources the most frequently recorded phenotype will be used within the derived variable. Where there is no agreement across sources the Cleft Collective will go back to the Cleft Team to determine a final classification.

1- Sharp, G, et al, (2017), 'Distinct DNA methylation profiles in subtypes of orofacial cleft'. Clinical Epigenetics
2- http://craniofacialsociety.co.uk/cleft-multidisciplinary-collaborative
3- https://www.crane-database.org.uk/