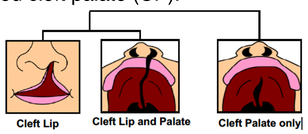
# Analyzing de novo mutations in case-parent trios with cleft lip only

**S. Ho, M.R. Bishop, E.J Leslie**

## Introduction

- Orofacial clefts (OFCs) are the most common craniofacial malformation with an overall incidence of 1/1000 live births.
- Nonsyndromic OFCs are phenotypically and etiologically heterogenous.
- OFCs are generally classified into isolated cleft lip (CL), cleft lip with cleft palate (CLP), and isolated cleft palate (CP).



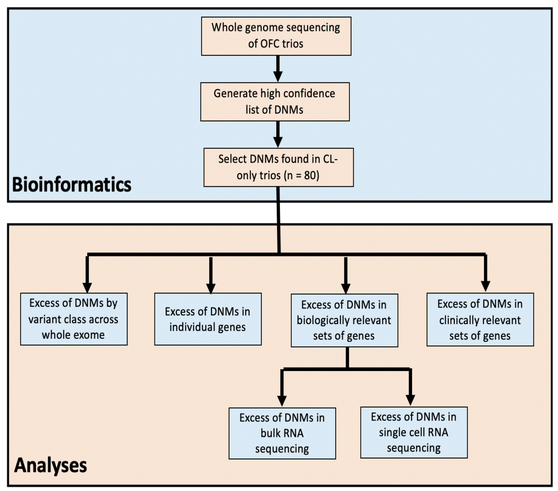

- CL occurs at a different time point in development, so it may have a different etiology from the other subtypes. Little research has been conducted on genetic risk factors for this specific subtype.
- *De novo* mutations (DNMs) spontaneously arise during embryonic development and have not been thoroughly studied as genetic risk factors for CL-only cases

## Methods

**Gabriella Miller Kids First Pediatric Research Program**: The Kids First program is a large-scale resource to help uncover the biology of childhood cancer and structural birth defects.

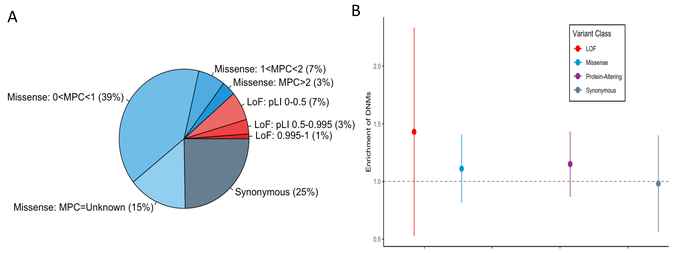
**OFC cohort**: 376 European trios with cleft lip with or without cleft palate (CL/P) were selected from a larger nonsyndromic OFC cohort

**Sequencing**: Trios were whole-genome sequenced at the Broad Institute on the Illumina HiSeq X Ten platform. Variants were called using a pipeline following the GATK Best Practice guidelines.



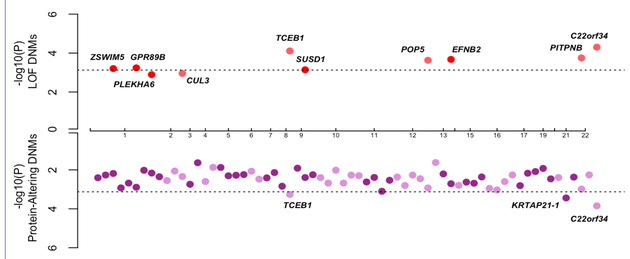

## Results

### Figure 1. DNMS by variant class



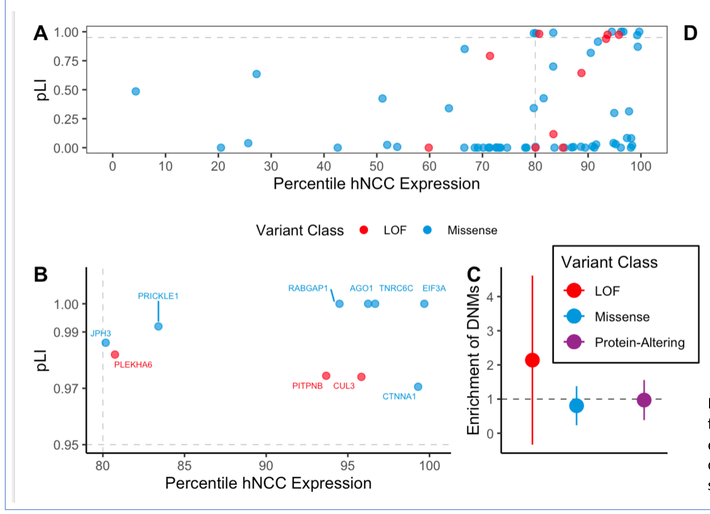

This shows the overall distribution of rare, coding DNMs (n = 89) by variant class in the CL-only trios (n = 80). A significant enrichment of loss of function (LoF; nonsense, frameshift, splice) and missense DNMs was not observed, likely due to the small cohort size.

### Figure 2. Genes with an excess of DNMs



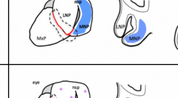
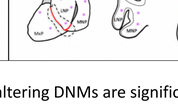

We identified a total of 8 genes with a significant excess of LoF DNMs and 3 genes with a significant excess of protein-altering (LoF + missense) DNMs using an exome-wide significance threshold ($p < 7.5X10^{-4}$ ). Genes *TCEB1* and *C22orf34* had significant excess in both variant classes.

### Figure 3. DNMs identified in genes highly expressed in human neural crest cells



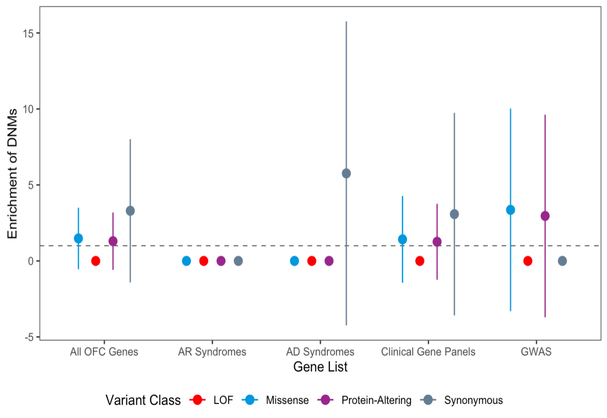

Plot A shows a ranking by pLI score and expression level of gene in RNA-sequencing data from human neural crest cells. Plot B shows genes in the top 20$^{th}$ percentile of hNCC expression with a pLI score ≥ 0.95. Plot C shows the enrichment values for DNMs by variant class in genes in the top 20$^{th}$ percentile of hNCC expression with a pLI score ≥ 0.95. Plot D shows a protein-interaction network for genes in our category of interest.

### Figure 4. Single cell RNA sequencing analysis

| Subcluster | Location | Observed | Expected | Enrichment | pValue |
|---|---|---|---|---|---|
| E0E11 |  | 1 | 0.7 | 1.5 | 0.49 |
| E9 |  | 1 | 0.8 | 1.32 | 0.53 |
| M3 |  | 2 | 0.3 | 6.62 | 0.04 |
| M6 |  | 1 | 0.3 | 2.9 | 0.29 |
| M7 |  | 1 | 0.3 | 3.78 | 0.23 |

Protein-altering DNMs are significantly enriched in marker genes for mesenchymal subcluster 3. Depiction of spatial locations of subclusters in the frontal view, anterior section, and posterior section; adapted from Li, *et al.*

### Figure 5. Analysis of DNMs in sets of clinically relevant genes




Enrichment of DNMs ± two standard errors for all CL-only trios in clinically relevant gene sets. We did not observe any LoF DNMs in any gene set.

## Conclusions and Future Directions

- Overall, we found evidence for five genes of interest that may contribute to the etiology of cleft lip only.

| Genes of interest | | *AGO1* | *EFNB2* | *CUL3* | *PITPNB* | *PLEKHA6* |
|---|---|---|---|---|---|---|
| Identified evidence supporting role in CL only etiology | Missense DNM identified | X | | | | |
| | LoF DNM identified | | X | X | X | X |
| | In the top 20th percentile for expression with a pLI score ≥0.95 | X | | X | X | X |
| | In the list of marker genes for the M3 subcluster | | X | | | |
| | Achieved individually significant p-values regarding LoF DNMs | | X | | X | |
| | Evidence in the literature | Family microRNA with transcriptional gene silencing | Mouse studies found association with neural crest defects. Damanging missense DNM for criofrontonasal syndrome. | Associated with autism, schizophrenia, and hypertension. Essential regulator of nerual crest specification. | Deletion mutation believed to be pathogenic for CL/P | Novel fusion gene in Langerhans cell histiocytosis |

- Future experiments should compare high significant CL genes to genes for CL/P to determine if the the etiologies are different

## Literature cited

Beaty, T.H., Marazita, M.L., and Leslie, E.J. (2016). Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Res *5*, 2800.

Bishop, M.R., Perez, K.D., Sun, M., Ho, S., Chopra, P., Mukhopadhyay, N., Hetmanski, J.B., Taub, M.A., Moreno-Uribe, L.M., Valencia-Ramirez, L.C., et al. (2020). Coding de novo mutations identified by WGS reveal novel orofacial cleft genes. BioRxiv 2020.04.01.019927.

Cai, J., Huang, X., Yin, M., Pan, C., Song, L., Zhan, Z., Chen, J., Gao, Y., Tang, J., Li, Y., et al. (2019). A novel fusion gene PLEKHA6-NTRK3 in langerhans cell histiocytosis. Int. J. Cancer *144*, 117–124.

Davidson, T.B., Sanchez-Lara, P.A., Randolph, L.M., Krieger, M.D., Wu, S.-Q., Panigrahy, A., Shimada, H., and Erdreich-Epstein, A. (2012). Microdeletion del(22)(q12.2) encompassing the facial development-associated gene, MN1 (meningioma 1) in a child with Pierre-Robin sequence (including cleft palate) and neurofibromatosis 2 (NF2): a case report and review of the literature. BMC Med. Genet. *13*, 19.

Lewis, A.E., Hwa, J., Wang, R., Soriano, P., and Bush, J.O. (2015). Neural crest defects in ephrin-B2 mutant mice are non-autonomous and originate from defects in the vasculature. Dev. Biol. *406*, 186–195.

Li, H., Jones, K.L., Hooper, J.E., and Williams, T. (2019). The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development *146*, dev174888.

Li, J., Hu, J., Sun, R., Zhao, Y., Liu, H., Li, J., Shi, L., and Zhao, S. (2017). Association between Cullin-3 Single-Nucleotide Polymorphism rs17479770 and Essential Hypertension in the Male Chinese Han Population. Dis Markers *2017*.

McGourty, C.A., Akopian, D., Walsh, C., Gorur, A., Werner, A., Schekman, R., Bautista, D., and Rape, M. (2016). Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor. Cell *167*, 525-538.e14.

Ramström, O., Ye, L., and Mosbach, K. (1998). Screening of a combinatorial steroid library using molecularly imprinted polymers. Anal. Commun. *35*, 9–11.

Torii, C., Izumi, K., Nakajima, H., Takahashi, T., and Kosaki, K. (2007). EFNB1 mutation at the ephrin ligand-receptor dimerization interface in a patient with craniofrontonasal syndrome. Congenit Anom (Kyoto) *47*, 49–52.

Ware J, Samocha K, Homsy J and Daly M (2015). Interpreting de novo variation in human disease using denovolyzeR. *Curr. Protoc. Hum. Genet.*, 87(7), pp. 1 15. R package version 0.2.0.

## Acknowledgements

These studies are part of the Gabriella Miller Kids First Pediatric Research Program consortium, supported by the Common Fund of the Office of the Director of the National Institutes of Health (www.commonfund.nih.gov/KidsFirst). Sequencing of the European trios was completed at Washington University's McDonnell Genome Institute (3U54HG003079-12S1, X01-HL132363) and the Colombian and Taiwanese trios were sequenced at the Broad Institute Sequencing Center (U24-HD090743, X01-HL136465, X01-??????). The sequencing centers plus the Kids First Data Resource Center (kidsfirstdrc.org, supported by the NIH Common Fund through U2CHL138346) provided technical and analytical support of this project. The data analyzed and reported in this manuscript were accessed from dbGaP [www.ncbi.nlm.nih.gov/gap; European trios: dbGaP accession number phs001168.v2.p2; Colombian trios: dbGaP accession number phs001420.v1.p1; Taiwanese ??????] and from the Kids First Data Resource Center (kidsfirstdrc.org). Additional grants supported the assembling of the cohorts, collection of the phenotypic data and samples, and data analysis were supported by the NIH: R03-DE027193 [EJL], R00-DE025060 [EJL], ….
We sincerely thank the many participating families, study teams, and colleagues worldwide without whom these studies would not be possible.