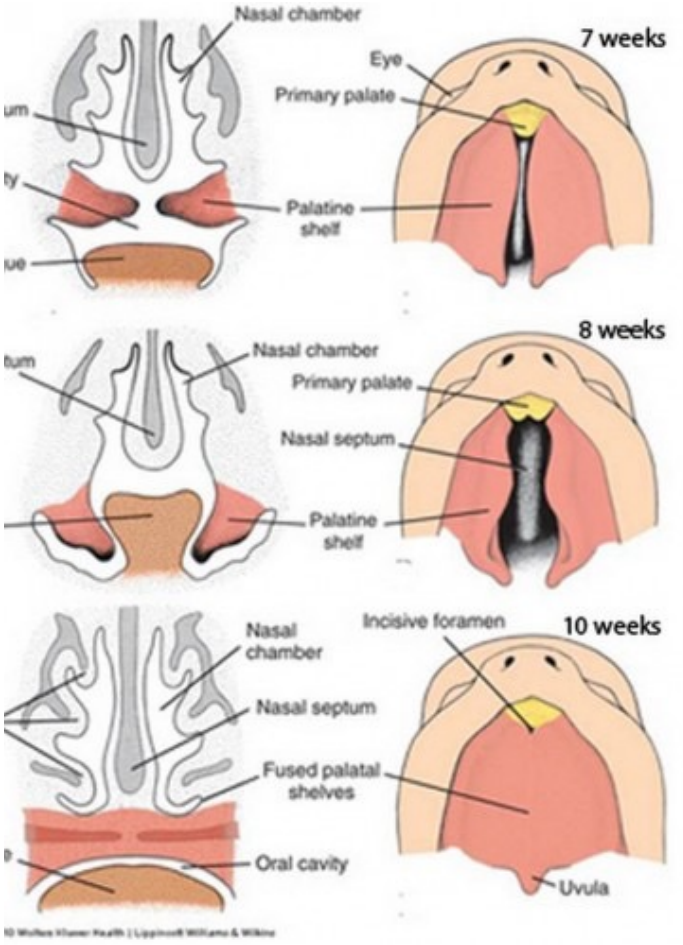
Analysis of CL and CP in Single Cell and Bulk RNA Sequencing

Samantha Ho

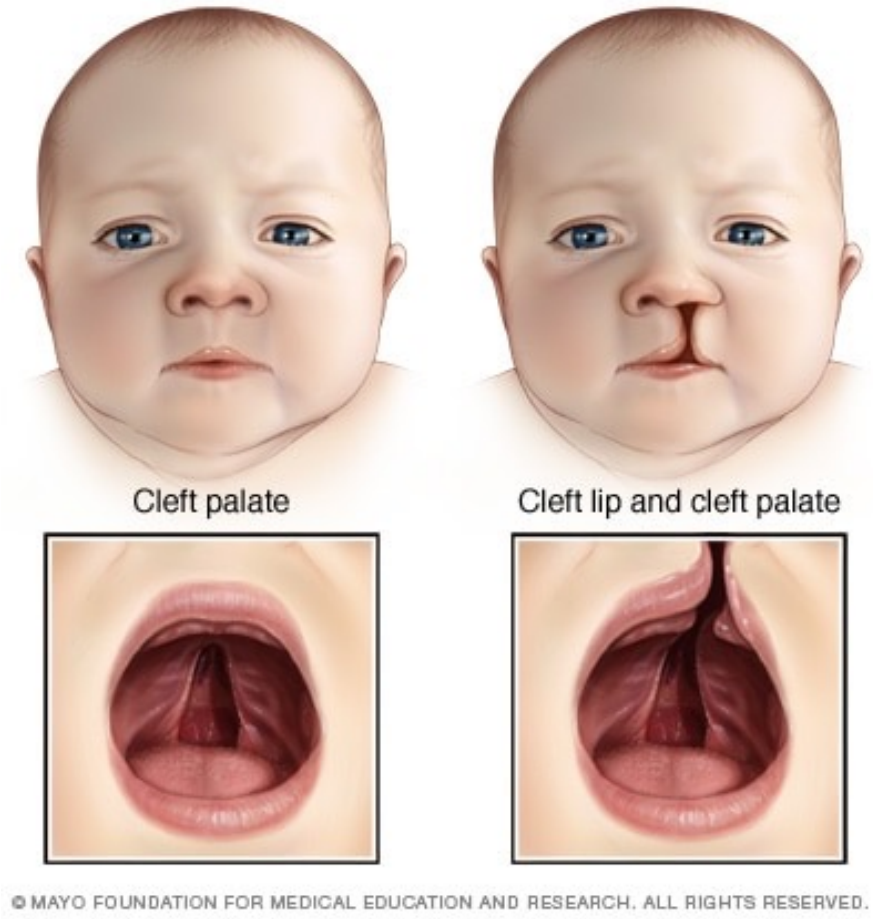
Craniofacial Development

- The development of the craniofacial region is a complex process involving massive movement of cells and tissues, including neural tube closure, midline patterning, neural crest generation, differentiation of facial primordia, etc.
- Requires several signaling pathways and gene expression networks controlled by regulatory mechanisms
- Alterations in craniofacial development leads to craniofacial congenital defects, including orofacial clefts



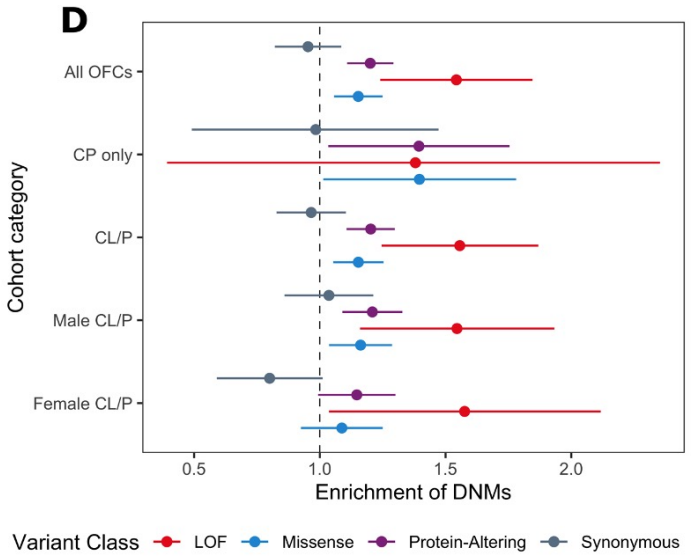
Orofacial Clefts (OFCs)

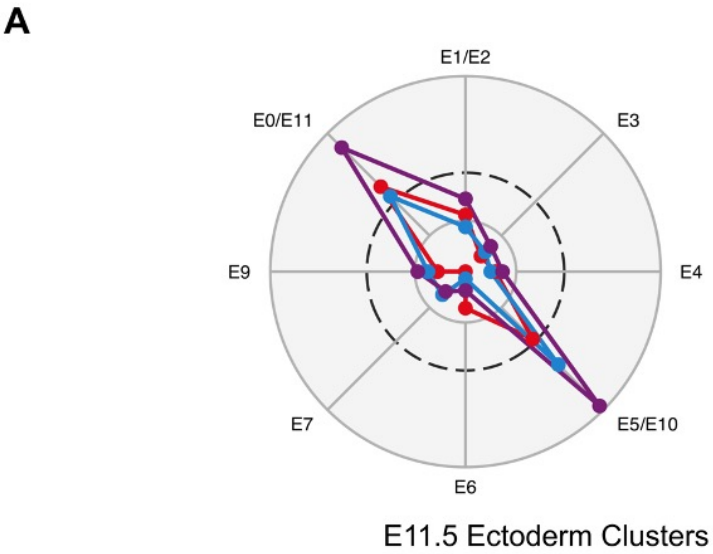
- Common birth condition
- Can occur as a part of genetic syndrome (syndromic) or isolated (nonsyndromic)
  - Majority are isolated
- Nonsyndromic OFCs are phenotypically and etiologically heterogenous
  - Possible combination of genetic and environmental factors
- Three general categories: cleft lip only (CL), cleft palate only (CP) or cleft lip with palate (CL/P)



De Novo Mutations in OFCs

- Possible role of de novo mutations (DNMs) since only 15% of OFC cases have a family history
- Madison found genome wide enrichment of DNMs by variant class for probands with CL, CP, and CL/P in Figure 1D
  - She also found significant excess of protein-altering DNMs in ectodermal cell sub-clusters E5/E10 and E0/E11 (Figure 2A)
  - We are doing something similar in craniofacial expressed genes, but subset for CL only and CP only





Objectives and Hypotheses

- To examine genetic risk factors through DNMs in CL only and CP only
  - Using epithelial clusters in single cell and bulk RNA sequencing
- **Hypotheses:**
  - There will be an enrichment of DNMs in epithelial clusters in CP and CL
    - For CL, clusters that are directly related to lip formation will have significant enrichment
  - This enrichment will differ between CP and CL

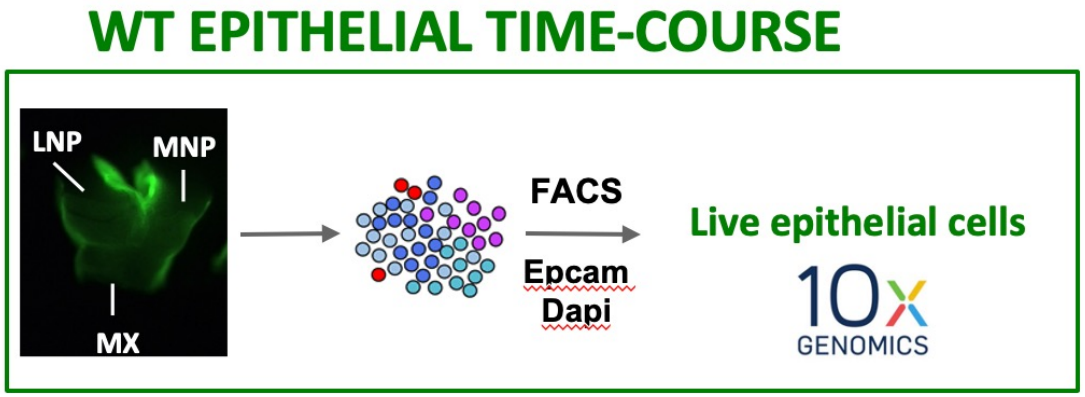
Methodology

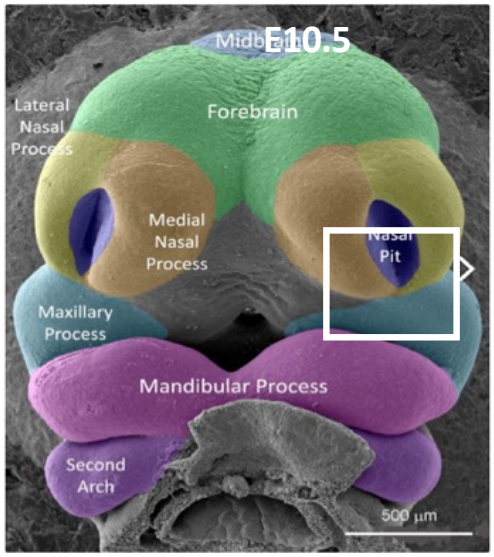
- We looked at de novo mutations in trios with cleft lip only and cleft palate only
- Used data from Gabriella Miller Kids First research

Orofacial cleft	Number of trios
Cleft lip	80
Cleft palate	58

Single Cell RNA Sequencing Pipeline

- Conducted RNA sequencing analysis in main epithelial mouse clusters

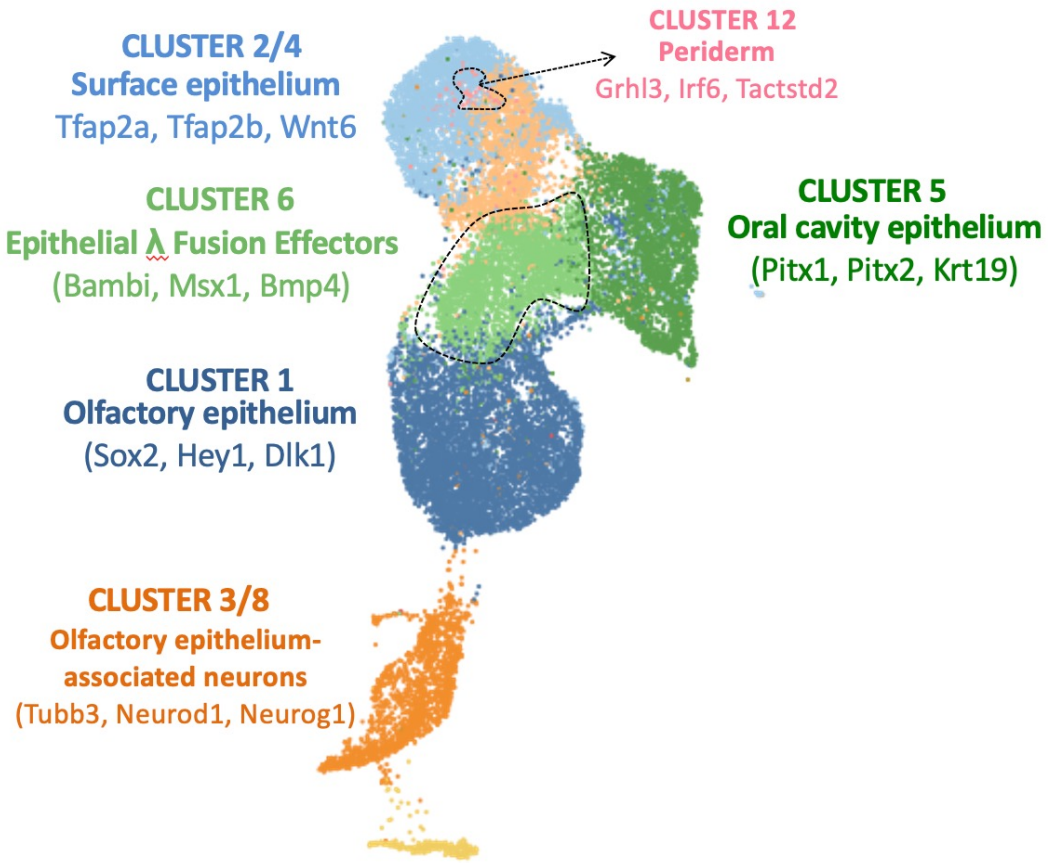




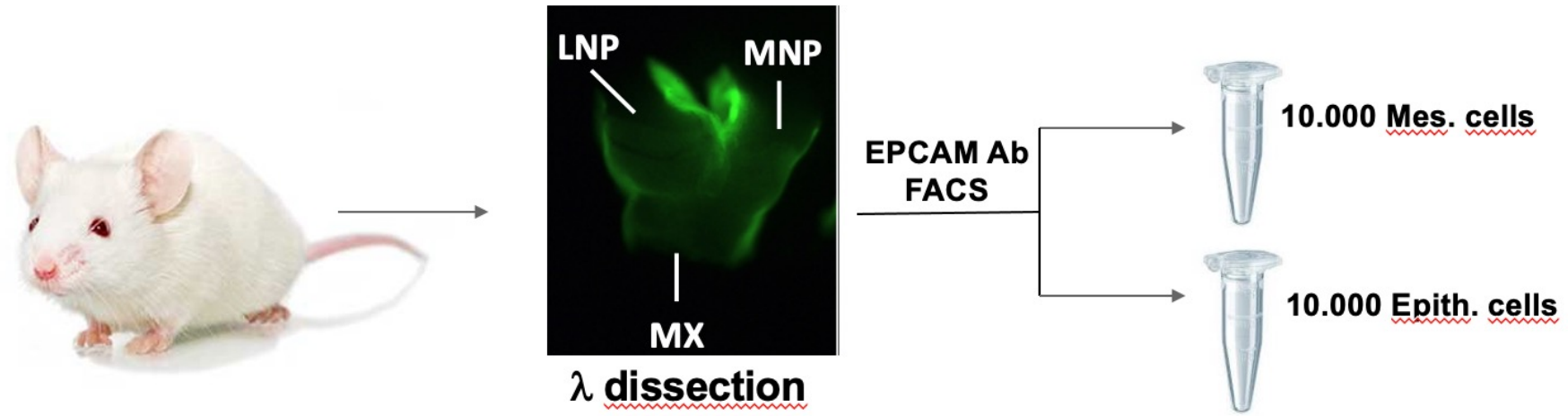
Selleri Lab

E10.5 Epithelial Clusters

- Separate analysis by single cell clusters in craniofacial regions
- All clusters contain top 150 differentially expressed genes in each subpopulation
- We analyzed DNMs in these genes in CL and CP separately



- In addition to single cell RNA sequencing, we also have bulk RNA data available
- Using E11.5 mice
- Mesenchymal and epithelial cells compared
- All differentially expressed genes that came from analysis, including those that were not significant



Selleri Lab

Bulk RNAseq

Converted Ensembl mouseIDs to hgnc using R package bioMart

- Filtered for padjusted ≤ 0.05 (4090 genes)

Used USC TableBrowser to annotate variants

- 13 missing identifiers (4077 genes)

Filtered annotations for exonic variants

- Removed monomorphic variants
- 23 variants in 23 genes for CL
- 14 variants in 14 genes for CP

Conducted R Denovolyzer byGene and byClass

Single Cell RNAseq

Analyzed each cluster for CP and CL

For CL, only one variant

- In Cluster 12
- Conducted byGene and byClass

For CP, 3 variants in 3 genes

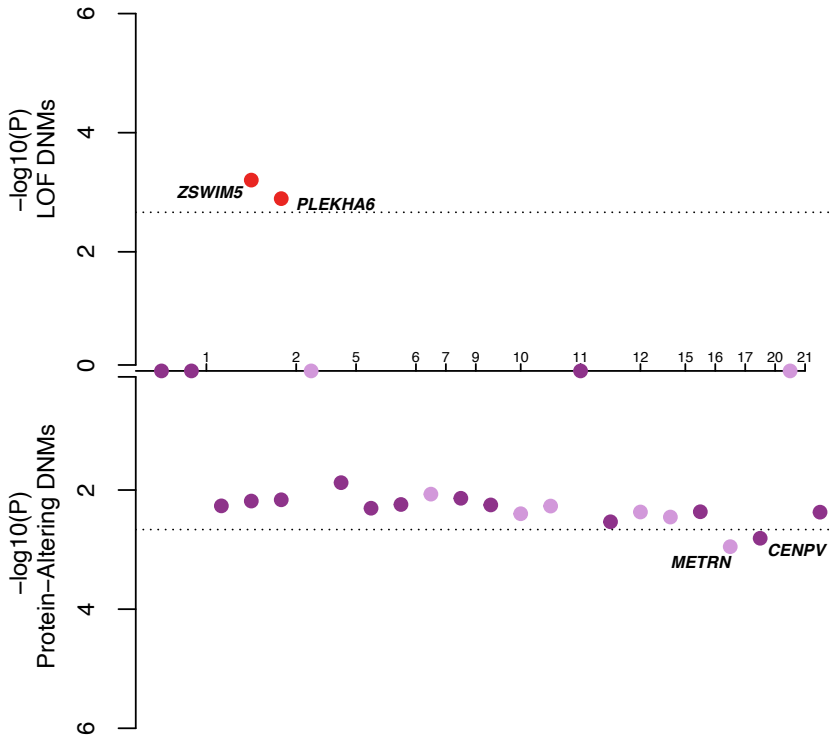
- In Clusters 1, 2_4, 5, and 12
- Conducted byGene and byClass

Results

Cleft Lip Only

CL BulkRNA byGene

- Bonferroni Pvalue: -log(0.00217)
- 23 total variants
- *ZSWIM5* (frameshift) and *PLEKHA6* (nonsense)
- *METRN* and *CENPV* (mis)



CL BulkRNA byClass

class	observed	expected	enrichment	pValue
syn	5	0	103.47	2.11E-09
mis	16	0.1	150.51	1.15E-29
lof	2	0	133.57	0.00011099
prot	18	0.1	148.41	4.49E-33
all	23	0.2	135.61	6.22E-41

CL SingleCell RNAseq byClass and byGene

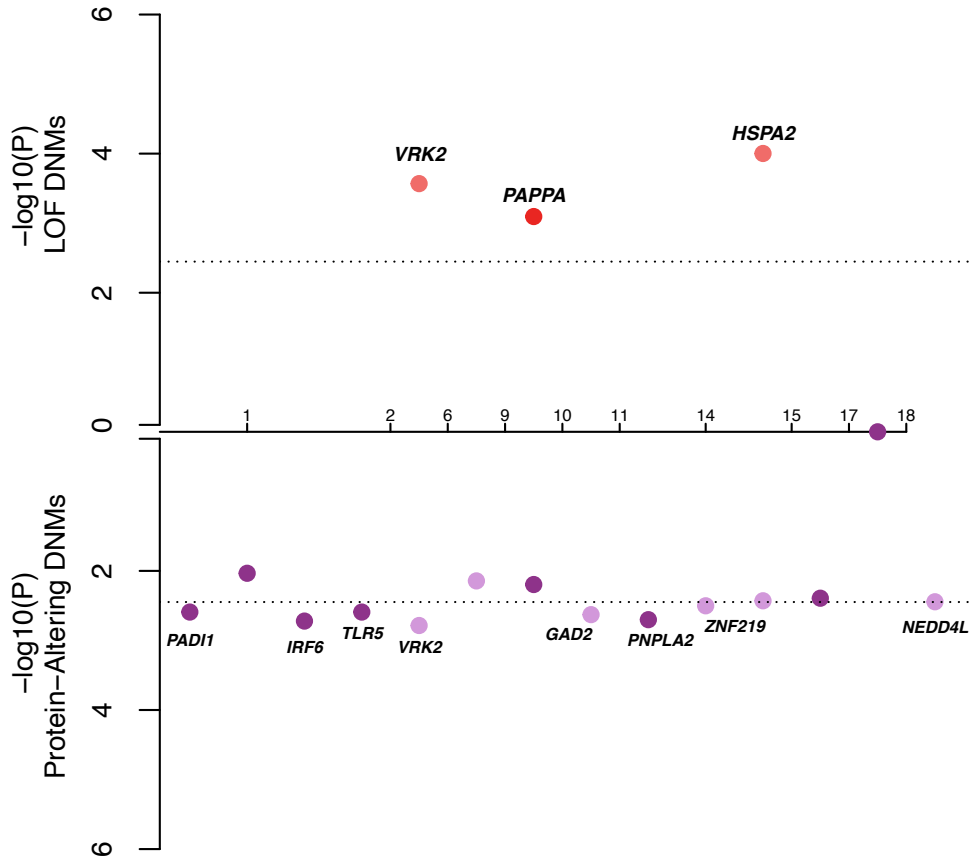
Gene	Class	prot_observed	prot_expected	enrichment	prot_pValue	Negative log pvalue
MPRIP	mis	1	0	148	0.00673	2.171984936

class	observed	expected	enrichment	pValue
mis	1	0	159.54	0.0062483
prot	1	0	148.12	0.0067287
all	1	0	105.33	0.009449

Cleft Palate Only

CP BulkRNA byGene

- Bonferroni Pvalue: -log(0.00357)
- 14 total variants
- 3 significant LOF variants
  - *VRK2*: splicing
  - *PAPPA*: stopgain
  - *HSPA2*: frameshift deletion
- 8 significant protein-altering variants
  - 1 in LOF (*VRK2*)

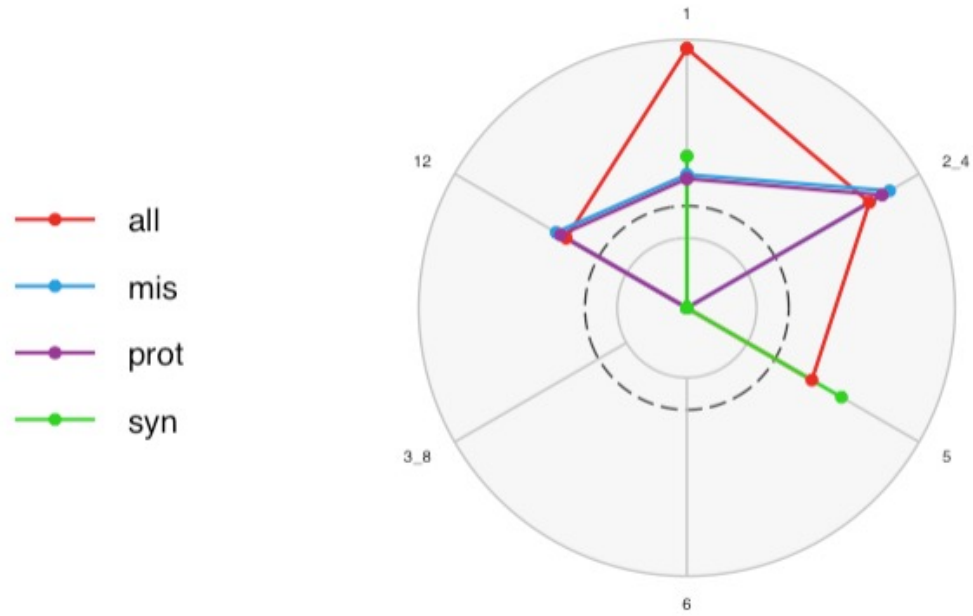


CP Bulk byClass

class	observed	expected	enrichment	pValue
syn	1	0	45.689	0.021649
mis	10	0	206.72	1.85E-20
lof	3	0	500.53	3.57E-08
prot	13	0.1	239.11	5.54E-27
all	14	0.1	183.59	2.40E-27

CP SingleCell RNA seq byClass

- No CP cases had any variants in any gene in Clusters 3_8 and 6
- Outer bound: -log($1x10^{-5}$)
- Dashed bound: -log(0.05/4 clusters)
- Inner bound: -log(0.05)



CP SingleCell RNA pValue Table byClass

Class	Cluster 1	Cluster 2_4	Cluster 5	Cluster 6	Cluster 3_8	Cluster 12
syn	0.0014802	NA	0.00046868	NA	NA	NA
mis	0.0032773	4.31E-05	NA	NA	NA	0.0015325
prot	0.0039109	6.22E-05	NA	NA	NA	0.0018969
all	1.45E-05	0.00011722	0.002033	NA	NA	0.0025157

CP SingleCell RNA seq byGene

Distribution of variants across the 6 clusters with the DenovolyzerbyGene results

Gene	Cluster 1	Cluster 2_4	Cluster 5	Cluster 6	Cluster 3_8	Cluster 12	Class	prot_observed	prot_expected	prot_pValue
GAD2	X	X					mis	1	0	0.00235
MEST	X	X	X				syn	0	0	1
IRF6						X	mis	1	0	0.0019

Conclusions

- We tested the enrichment of DNMs in genes in epithelial clusters from single cell and bulk RNA sequencing experiments
- We identified an enrichment for several genes in the single cell clusters including…
  - CL: *MPRIP*
  - CP*: IRF6, MEST, GAD2*
- We identified an enrichment for several genes in the bulk RNA, including…
  - CL: *ZSWIM5, PLEKHA6, METRN, CENPV*
  - CP*: VRK2, PAPPA, HSPA2*
- Identified different genes enriched with DNMs in CP compared to CL

Limitations

- Results are difficult to interpret given:
  - Large pValues
  - Small sample sizes
  - Could not test all clusters

Thank you!