Fall 2020 Lab Presentation

Samantha Ho

With Madison: Categorizing variants by ACMG standards

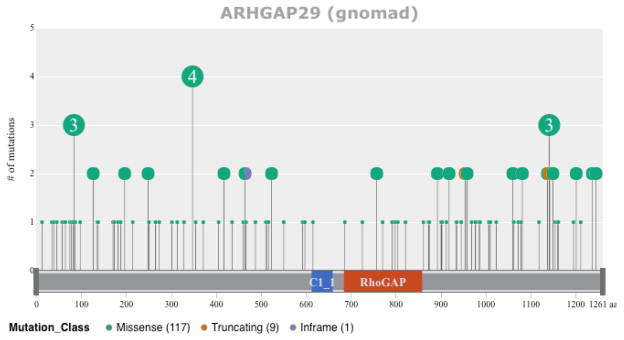
- File: GMKF_ALLEthnicities_annotate_hg38_multianno_LoF_Tier3Missense_Sam
- ACMG classification of LoF and Tier 3 missense variants in all ethnic trios (Colombian, Taiwanese, European)
- Used InterVar and manual research to check off criteria.
  - "XX" = manual
- PVS1 criteria columns included and determined by "Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6185798/)
  - Used Ensemble Variant Effect Predictor (VEP) and UCSC genome browser.
- Results showed 24 pathogenic and 20 likely pathogenic variants

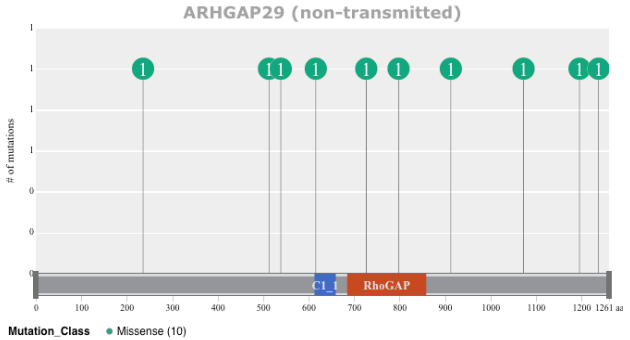
ACMG	PVS1	Predicted to undergo NMD	Biologically relevant transcript	PS1	PS2	PS3	PS4	PS5	PM1	PM2	PM3	PM4	PM5	PM6
likely pathoge	X	X								X				
pathogenic	X									X				
pathogenic	X	X							X	X				
pathogenic	X									X				
likely pathoge	X	X								X				
likely pathoge	X								X	X				
pathogenic	X									X				
pathogenic	X									X				
pathogenic	X	X								X				
pathogenic	X	X								X				
VUS										X				
Pathogenic	X	X	X							X				
Pathogenic	X	X	X							X				
likely pathoge	X	X	X							X				
VUS		X								X				
Pathogenic	X	X	X							X				
pathogenic	X	X	X							X				
pathogenic	X	X	X							X				
VUS		X								X				
VUS										X				
pathogenic	X	X	X							X				
pathogenic	X	X	X							X				
pathogenic	X	X	X							X				
VUS		X								X				
VUS		X								X				
VUS		X								X				

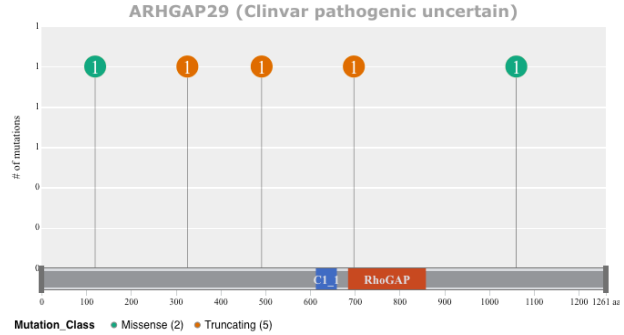
With Courtney: IRF6 gene plots

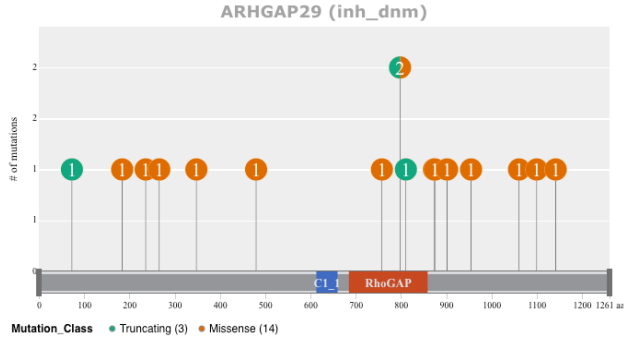
- How can we visualize transcriptional gene targets of IRF6
- Looked at ARHGAP29, ESRP1, ESRP2 and filtered for rare coding pathogenic variants
  - Rare in gnomAD (<0.1%)
  - Exonic or splice variants
  - Protein-altering variants
- Divided by sample: gnomAD, ClinVar, non-transmitted, inherited/denovo
- Compare programs for visualizing the variants
  1. g3viz R package
  2. ggplot
  3. Lollipops Go package

G3viz plots


ARHGAP29 (gnomad)
# of mutations
C1_1
RhoGAP
Mutation_Class
Missense (117)
Truncating (9)
Inframe (1)


ARHGAP29 (non-transmitted)
# of mutations
C1_1
RhoGAP
Mutation_Class
Missense (10)


ARHGAP29 (Clinvar pathogenic uncertain)
# of mutations
C1_1
RhoGAP
Mutation_Class
Missense (2)
Truncating (5)


ARHGAP29 (inh_dnm)
# of mutations
C1_1
RhoGAP
Mutation_Class
Truncating (3)
Missense (14)

Pros:

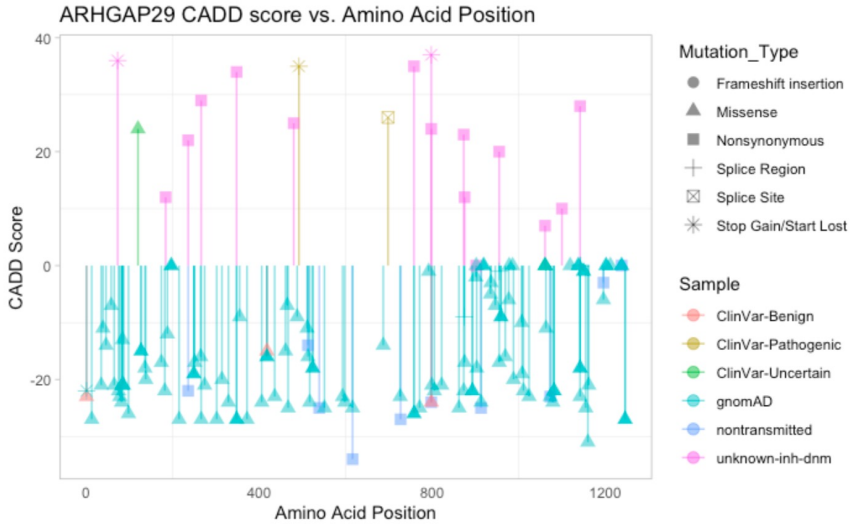
- Interactive – can zoom in and out, roll over points for more information
- Automatic – easy to construct

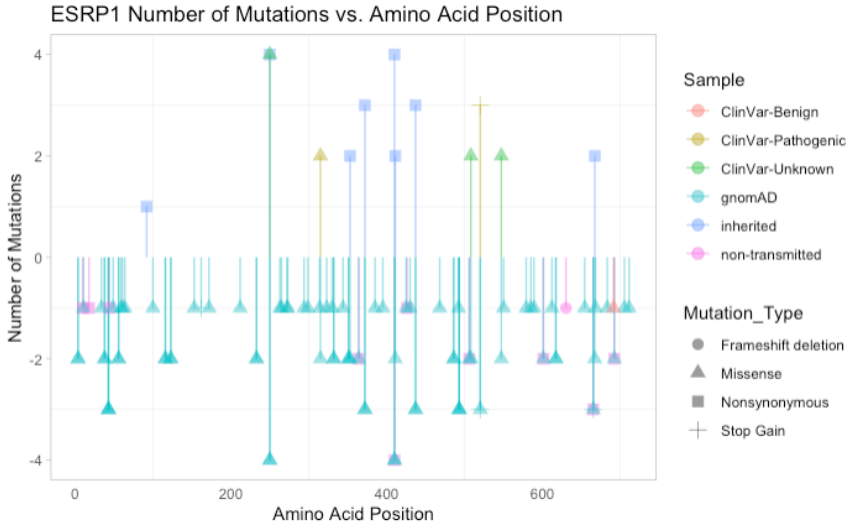
Cons :

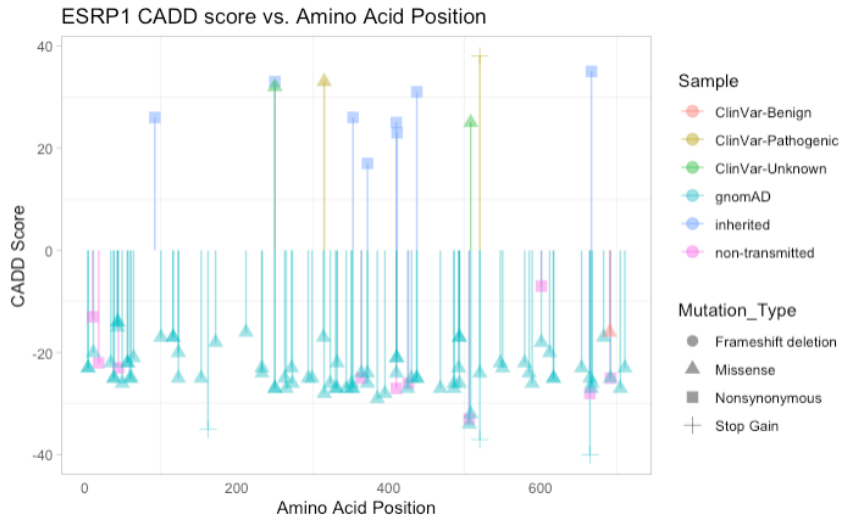
- Can't change the axes
- Doesn't support ESRP1 or ESRP2

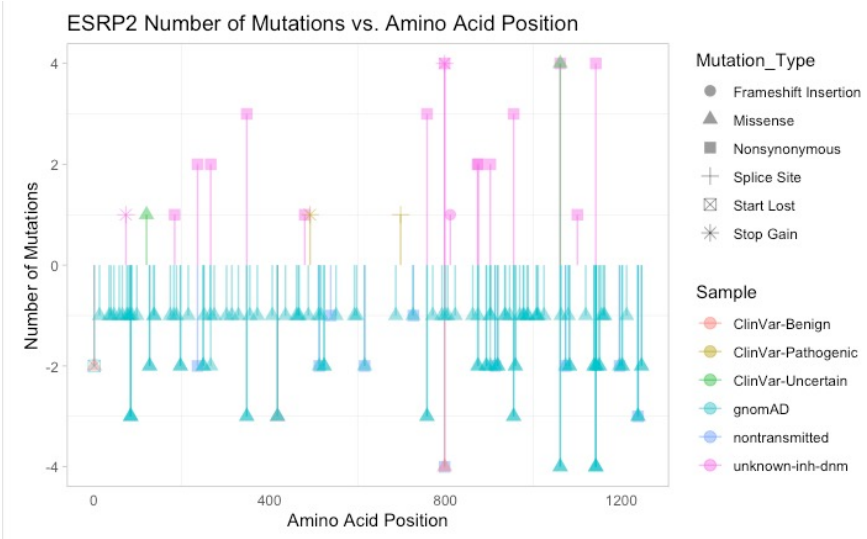
ggplot graphs


ARHGAP29 Number of Mutations vs. Amino Acid Position
Number of Mutations
4
2
0
-2
-4
0
400
800
1200
Amino Acid Position
Mutation_Type
Frameshift insertion
Missense
Nonsynonymous
Splice Region
Splice Site
Stop Gain/Start Lost
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Uncertain
gnomAD
nontransmitted
unknown-inh-dnm


ARHGAP29 CADD score vs. Amino Acid Position
CADD Score
40
20
0
-20
0
400
800
1200
Amino Acid Position
Mutation_Type
Frameshift insertion
Missense
Nonsynonymous
Splice Region
Splice Site
Stop Gain/Start Lost
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Uncertain
gnomAD
nontransmitted
unknown-inh-dnm


ESRP1 Number of Mutations vs. Amino Acid Position
Number of Mutations
4
2
0
-2
-4
0
200
400
600
Amino Acid Position
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Unknown
gnomAD
inherited
non-transmitted
Mutation_Type
Frameshift deletion
Missense
Nonsynonymous
Stop Gain


ESRP1 CADD score vs. Amino Acid Position
CADD Score
40
20
0
-20
-40
0
200
400
600
Amino Acid Position
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Unknown
gnomAD
inherited
non-transmitted
Mutation_Type
Frameshift deletion
Missense
Nonsynonymous
Stop Gain


ESRP2 Number of Mutations vs. Amino Acid Position
Number of Mutations
4
2
0
-2
-4
0
400
800
1200
Amino Acid Position
Mutation_Type
Frameshift Insertion
Missense
Nonsynonymous
Splice Site
Start Lost
Stop Gain
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Uncertain
gnomAD
nontransmitted
unknown-inh-dnm


ESRP2 CADD score vs. Amino Acid Position
CADD Score
40
20
0
-20
0
400
800
1200
Amino Acid Position
Mutation_Type
Frameshift Insertion
Missense
Nonsynonymous
Splice Site
Start Lost
Stop Gain
Sample
ClinVar-Benign
ClinVar-Pathogenic
ClinVar-Uncertain
gnomAD
nontransmitted
unknown-inh-dnm

Pros

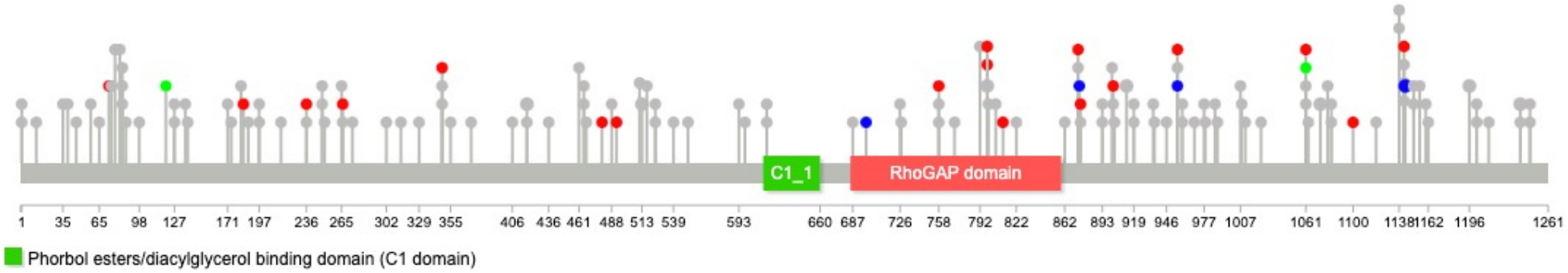
- Lots of options for customization (shape and color for sample and functional consequence)
- Can make the y-axis negative so easier to see the groups of interest

Cons

- Much more time-consuming to create
- Have to be very precise with nomenclature
- Not interactive
- No additional genetic information

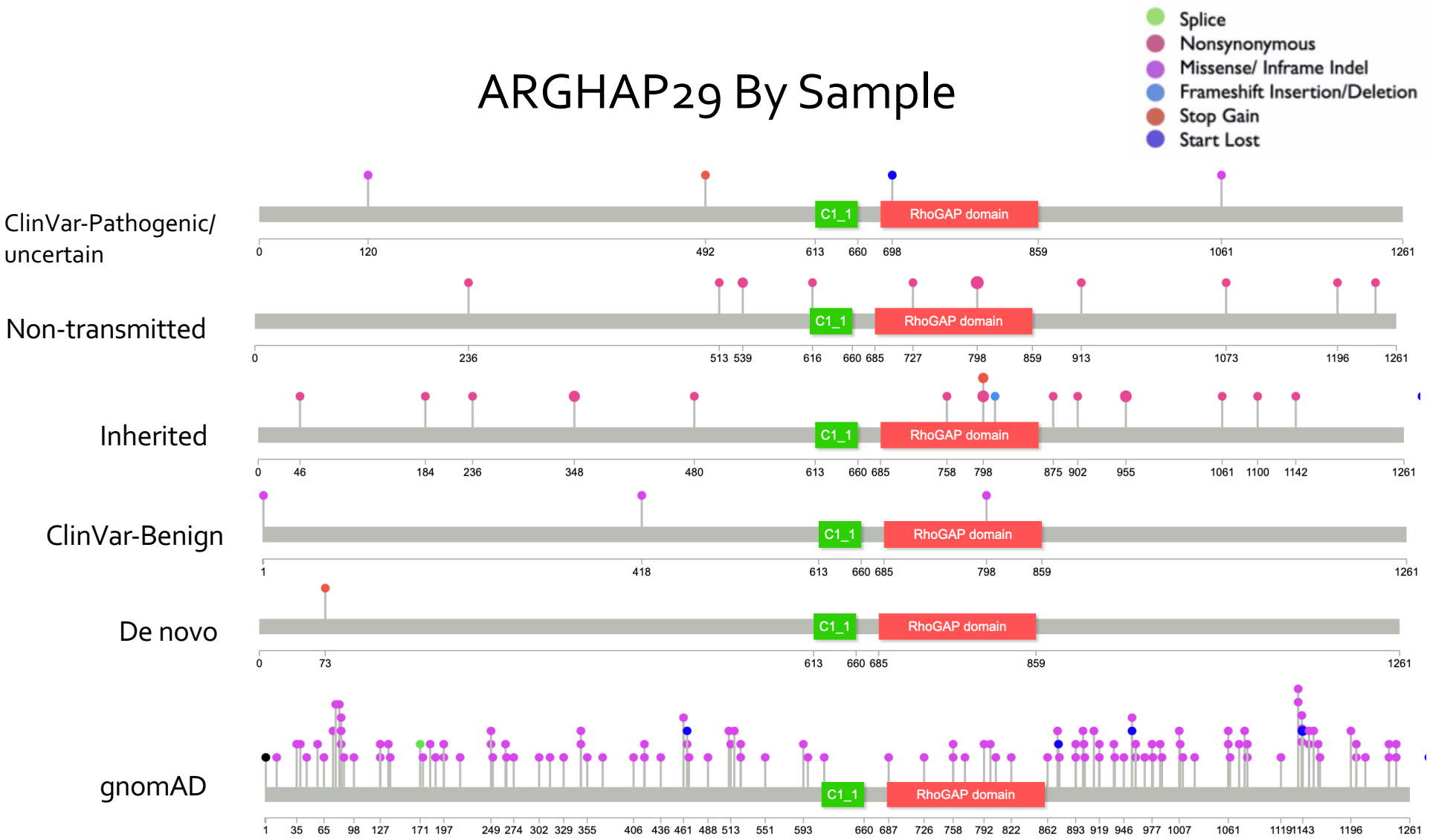
Lollipops Go package

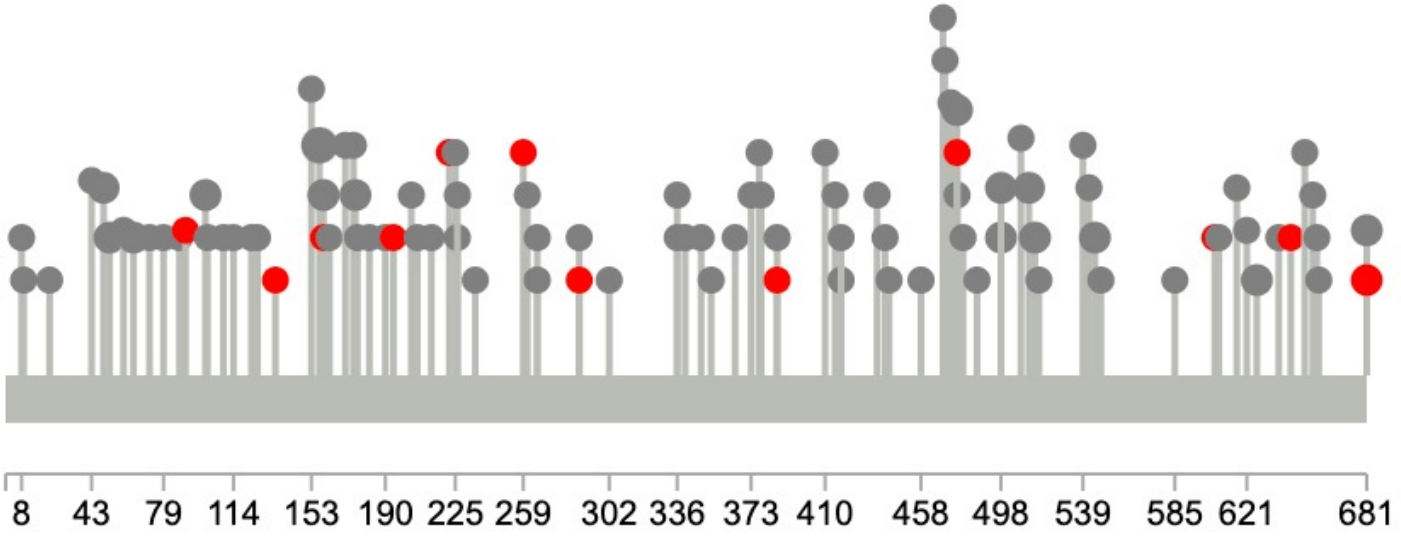
ARHGAP29



- ClinVar-Pathogenic/denovo/inherited
- ClinVar-Benign, gnomAD, nontransmitted
- ClinVar-Uncertain

ARGHAP29 By Sample

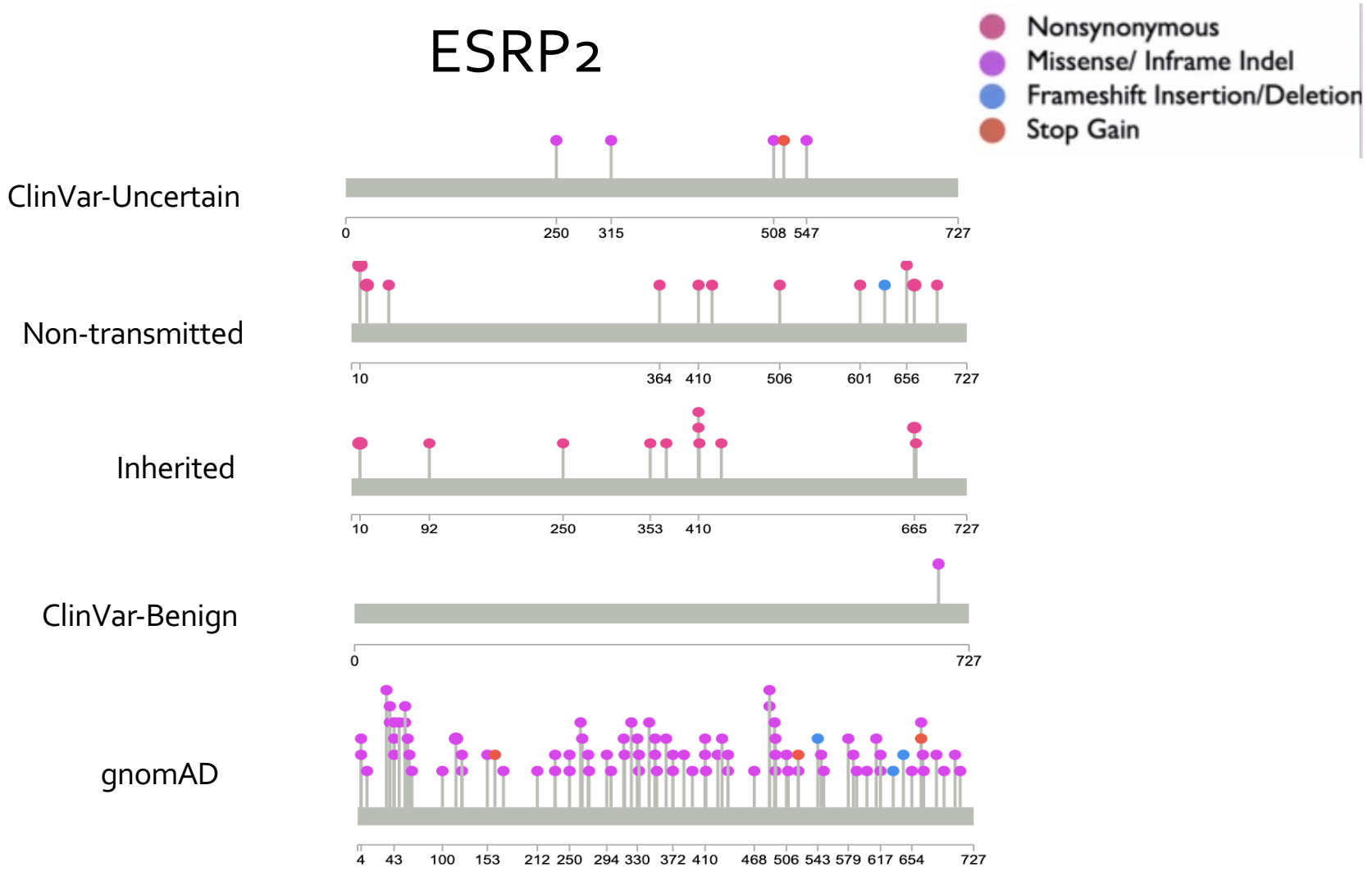




ESRP1

ClinVar-Pathogenic/inherited
ClinVar-Benign, gnomAD, nontransmitted

ESRP2


Nonsynonymous
Missense/ Inframe Indel
Frameshift Insertion/Deletion
Stop Gain
ClinVar-Uncertain
0 250 315 508 547 727
Non-transmitted
10 364 410 506 601 656 727
Inherited
10 92 250 353 410 665 727
ClinVar-Benign
0 727
gnomAD
4 43 100 153 212 250 294 330 372 410 468 506 543 579 617 654 727

Pros

- Convenient, automatic
- Has options to include legend, gene names, color
- Separation of groups shows some overlap

Cons

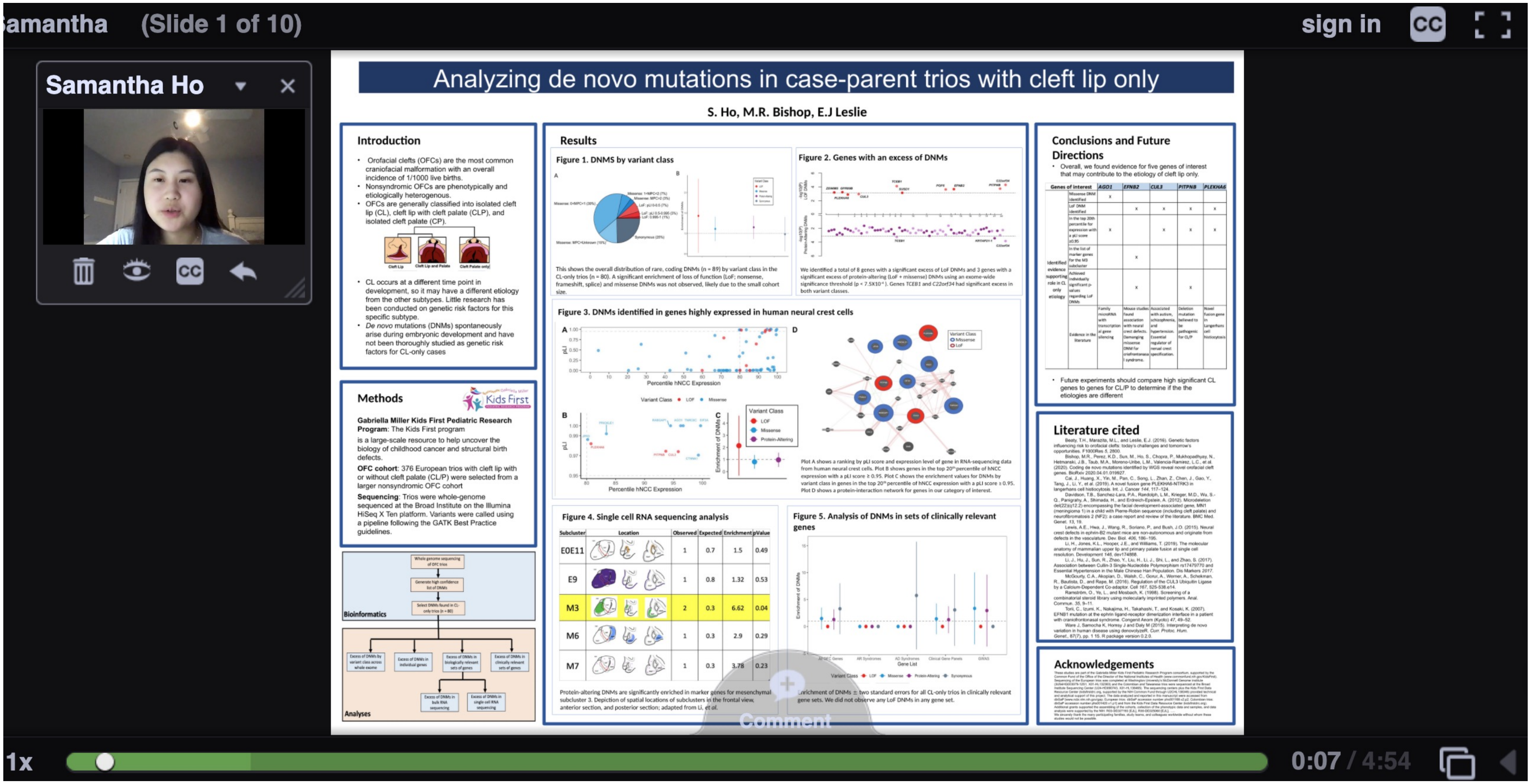
- Not interactive
- ESRP1 and ESRP2 do not have well-characterized domains, so not visually interesting
- Need to make text file list of all variants you want to graph
- https certificate expired

Conclusions and future directions

- Couldn't define any regions within these genes that have a strong association with OFC based off visualization
- Can't definitively say that rare pathogenic variants are associated with CL/P
- Future
  - Some groupings in areas suggest that a statistical test would give us greater insight if these regions are significant or interesting

Fall Undergraduate Virtual Research Symposium

- Participated in the Fall Undergraduate Virtual Research Symposium in November
- Had to upload an abstract and video recording of presentation, where people could leave questions or comments
- Not exactly the same as in-person, but was still fun to record and prepare!


amantha (Slide 1 of 10)
sign in
CC
Samantha Ho
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
Kids First
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.
Bioinformatics
Analyses
Results
Figure 1. DNMS by variant class
Figure 2. Genes with an excess of DNMs
Figure 3. DNMs identified in genes highly expressed in human neural crest cells
Percentile hNCC Expression
Figure 4. Single cell RNA sequencing analysis
Figure 5. Analysis of DNMs in sets of clinically relevant genes
Conclusions and Future Directions
Overall, we found evidence for five genes of interest that may contribute to the etiology of cleft lip only.
Future experiments should compare high significant CL genes to genes for CL/P to determine if the the etiologies are different
Literature cited
Acknowledgements
Comment
1x
0:07 / 4:54

Thank you for listening!

Any questions?