Summary of Fall 2021 VUS Analysis

Written on April 8, 2022 by Samantha Ho

**Goal**: To identify if there is a significant relationship between variants of unknown significant (VUS) in genes with pathogenic variants. If there is a pathogenic gene with a burden of VUS, this may indicate that some of the VUS are pathogenic.

**Methodology and Summary:**

I used data from the Gabriella Miller Kids First Pediatric Research Program, which had 837 case-parent trios with a total of 1040 VUS in 220 unique genes. I first obtained descriptive statistics by looking at the distribution of VUS in trios by population and cleft status.

Chart, bar chart

Description automatically generated

Next, I looked for frequency of VUS in genes on 4-5 orofacial gene lists (DDG2P, Gene Panel, NHS, Prevention Genetics, OMIM) then took the top five genes to conduct chi-squared and Fisher testing. Chart

Description automatically generated

Through statistical testing, I found that CLP cases had significantly more VUS than controls, which is supported by the initial cleft status descriptive percentages. However, I cannot make further conclusions with this information since there is not a significant difference between each of the overall percentages and the majority of cases in any of the groups have VUS. I did not find any statistically significant genes at p=0.05 when conducting a Fisher test with VUS or pathogenic variants in cases vs. controls.

Next, I looked at VUS and pathogenic variants by function and consequence in genes with pathogenic variants. This showed that many of the genes with the highest number of pathogenic variants also had an excess of VUS. The following graph shows this distribution ordered by gene transcript length.

Chart, bar chart

Description automatically generated

Then, I checked to see if there was a significant difference between percentage of loss of function and missense VUS in cases vs. controls, but they had very similar percentages as shown below.

Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

I also made lollipop graphs of the three genes with highest number of pathogenic variants (ARHGAP29, COL2A1, and CTNND1) to examine the location of VUS and pathogenic variants on the gene by mutation type.

Chart, box and whisker chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

Lastly, I analyzed the distribution of pathogenic and benign scores for missense variants in genes with VUS, pathogenic, and benign variants. These graphs showed nearly opposite distributions, which indicates that the scores are consistent for each variant and are likely accurate. The VUS variants that had a pathogenic score of 8-9 and benign score <5 may be pathogenic, so this subset of variants should be examined in future analysis to determine pathogenicity.

Chart

Description automatically generated

Chart

Description automatically generated

Copies of the analysis files and Rcode can be found in folder “VUS Analysis” in Labbies.

In terms of cleft lip (CL) only trios, there are 3 pathogenic variants in CL only samples (in 3 unique cases). This is compared to a total of 78 cases of any cleft type with pathogenic variants and 4 controls with pathogenic variants. There are 101 total VUS in CL only samples in 70 unique genes. The genes and their VUS counts are shown here: