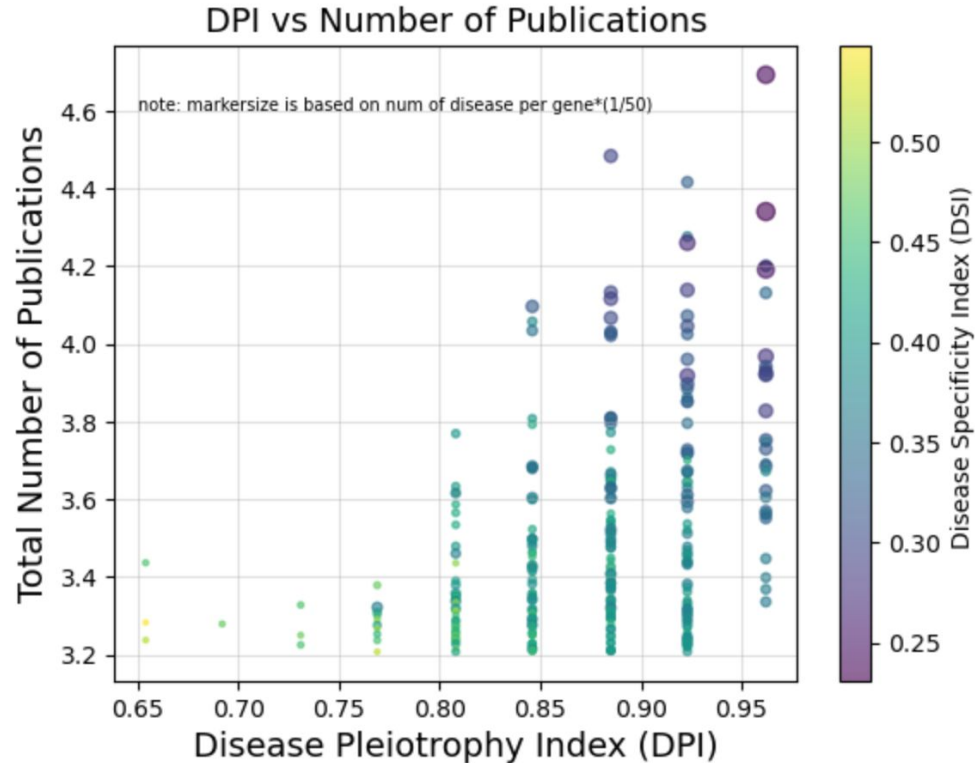


# Vivian Li

**Research Question:** Identify the top 300 genes with the highest number of total publications, and find the relationship between the DPI, DSI, and the number of publications of the gene.



Based on the scatterplot, the gene with the highest DPI and the lowest DSI has more publications than genes with the lowest DSI and DPI. Furthermore, the size of the plot is based on the number of unique diseases that is associated with each gene (scaled by 1/50). It demonstrated that a higher DPI would associate to more diseases related to the gene, which may correlate to an increase in the number of publications.