Community Detection and Phi Co-occurrence

None of the communities gene – HPO pairs appeared in the original notable phi-coefficient analysis. In this analysis I combine the HPO terms in interesting communities to be single super-HPO nodes and then re-calculate phi for each super-HPO terms and the genes also found in its community.

The HPO terms in the pairs below represent the super-HPO term for their community, despite having the name of a single term they carry the meaning and value of all HPO terms in their community. Using the threshold of phi > 0.30 for showing signs of correlation, community 8 (dealing with polydactyl) is the only one to show with suggesting a correlation.

Community: 6

TMEM39B : 0.15049870640330867

HP:0002803 DNAH7 : 0.15049870640330867

HP:0002803 MYLK4 : 0.15049870640330867

HP:0002803 MYH2 : 0.15049870640330867

Community: 8

HP:0009556 BGN : **0.31613179478426595**

Community: 9

HP:0002671 SAE1 : 0.22344496439434378

HP:0002671 TBCB : 0.22344496439434378

HP:0002671 PLK3 : 0.22344496439434378

HP:0002671 UHRF1BP1L : 0.22344496439434378

Community: 10

HP:0030127 ERCC3 : 0.2885605676771093

HP:0030127 UVRAG : 0.2885605676771093

A screenshot of a social media post

Description automatically generated

Distributions of the phi coefficients (Matthew’s Correlation Coefficient) calculated by 10,000 bootstraps of the data. In each boot strap a gene was selected at random, then 3-6 of the gene’s neighbor HPO terms were chosen at random to be combine into a super-node. Phi was then calculated for this gene and super-node pair. The blue dotted line is the phi value of community 8 (phi = 0.316), the only community with a phi coefficient suggesting a correlation. The red dashed line is the threshold for significance, the 99.54th percentile. This threshold was calculated using the Bonferroni correction to control for the multiple tests of recalculating phi for 11 gene-HPO pairs in the communities.