Identifying putative protein protein interactions in Quantitative MS data

By

Kalaivani Muralidhara

Under Supervision

Dr.Lars Malmstroem

# Introduction:

The dataset contains Peptide, Protein, Intensity, retention time, and Peptide score.In addition it has PG column which is the uncorrected “p”value or score for the whole peptide.So there are 12655 rows and 197 columns of which 64 Intensity columns are our focus for now So the matrix we have now is 12655\*64 data points.

# Methods:

# Results:

Figure 1: Heatmap

Figure 2: U map

Figure 3: Table on Eucleadean Distance matrix

Figure 4: HDBSCAN

Figure 5: Table consisting of proteins behaving similarly.

Conclusions and Discussions:

# Image

Heatmap of Intensity verses Peptide:Image

Heatmap of Intensity(Ranging from normalised 0.4 to 0.0 values) verses Peptide:

12655\*65=809920 is the total intensities values out of which the range of 0 to 0.1 is 490041 followed by 0.1 to 0.2 are 265280 and so on.So it is very evident that the Intensities values are mostly spread between 0.0 to 0.4 and the values from 0.5 to 1.0 can be replaced by 0.4 by the snippet below and when we now generate a heatmap.

| Intensities | Count |
| --- | --- |
| 0.0 to 0.1 | 490041 |
| 0.1 to 0.2 | 265280 |
| 0.2 to 0.3 | 38503 |
| 0.3 to 0.4 | 8311 |
| 0.4 to 0.5 | 3443 |
| 0.5 to 0.6 | 2089 |
| 0.6 to 0.7 | 1064 |
| 0.7 to 0.8 | 541 |
| 0.8 to 0.9 | 253 |
| 0.9 to 1.0 | 395 |

Image

# Umap:

# Image