

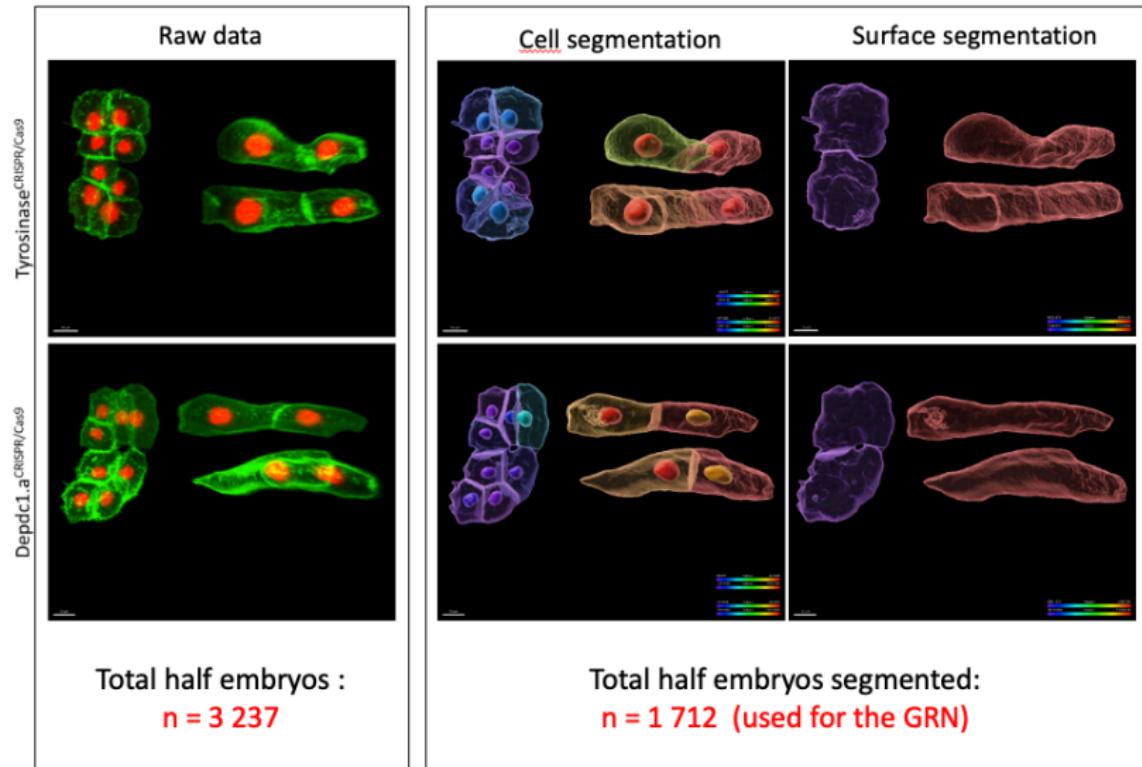
CRISPR Screen

Unsupervised Clustering with Automated Hyperparameter Selection

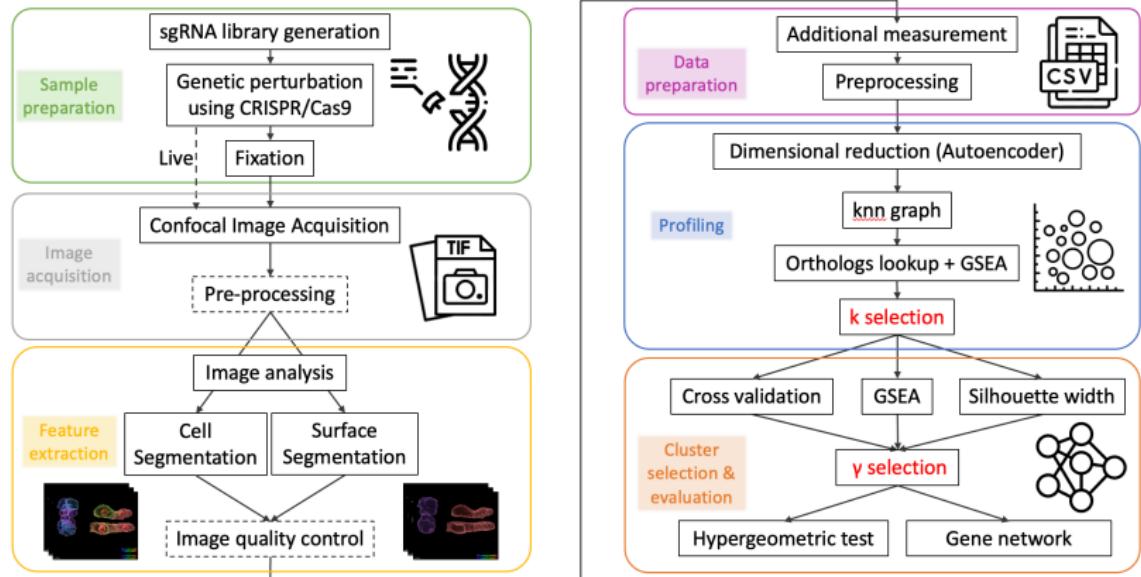
Keira Wiechecki

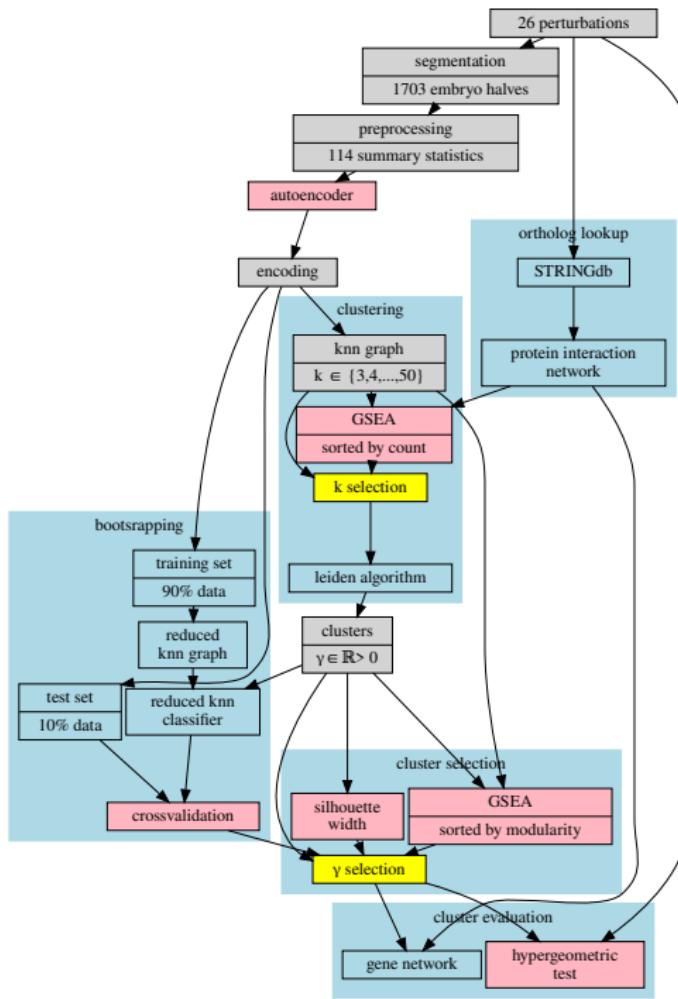
November 15, 2022

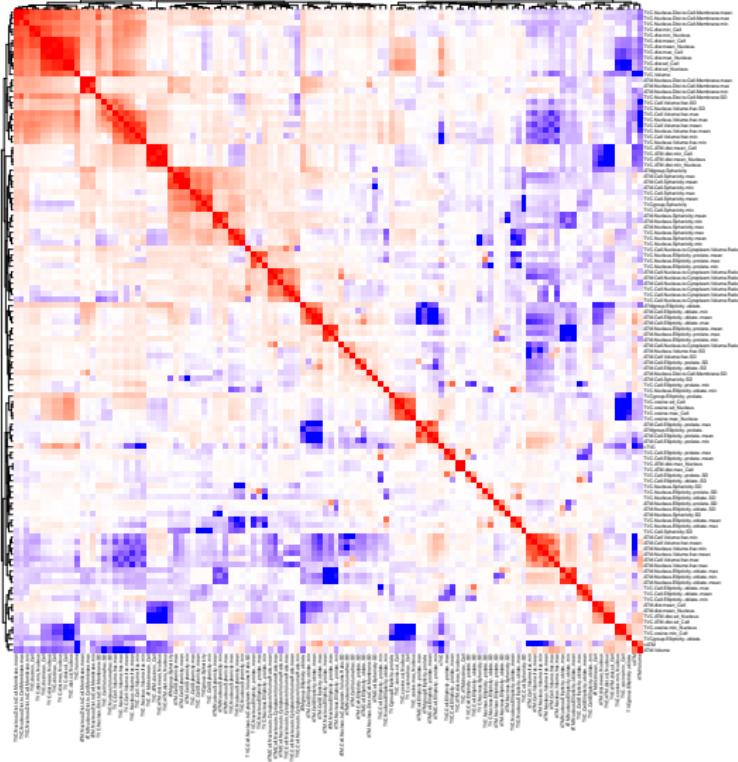
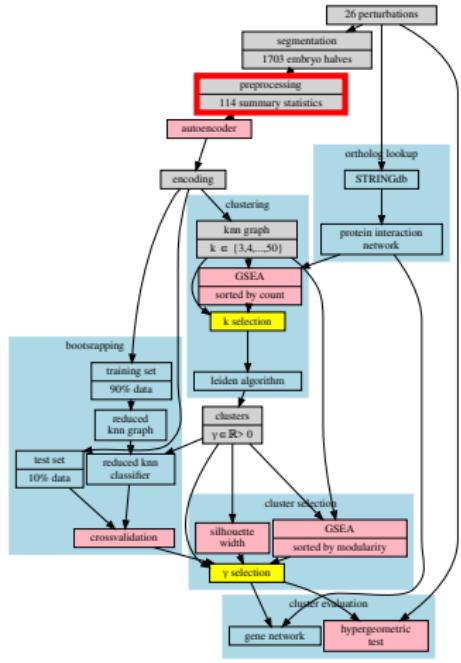
Rationale



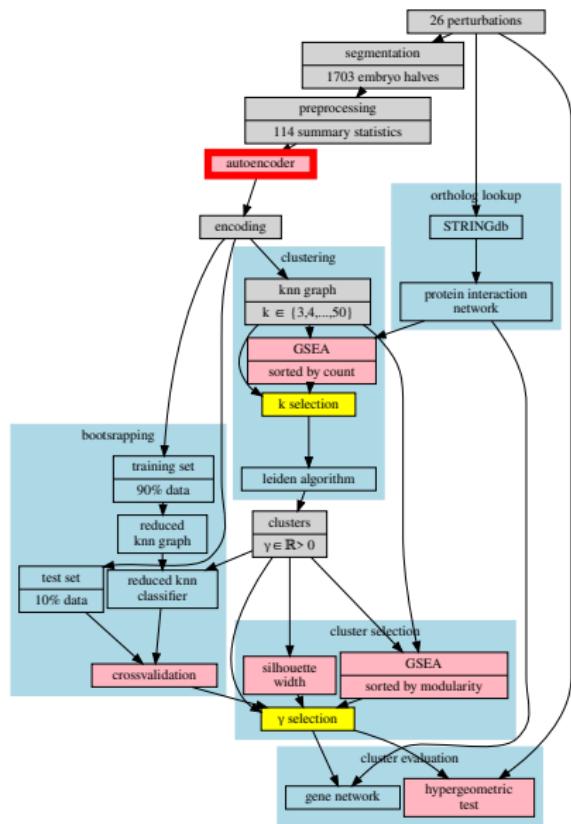
Overview



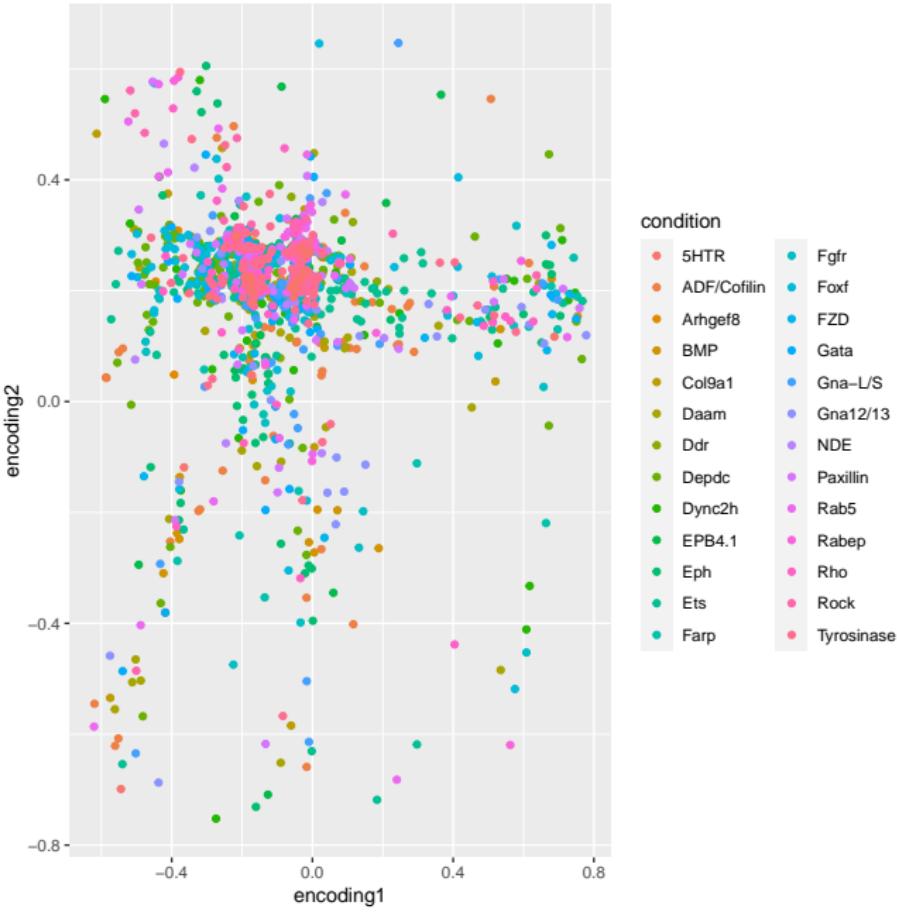
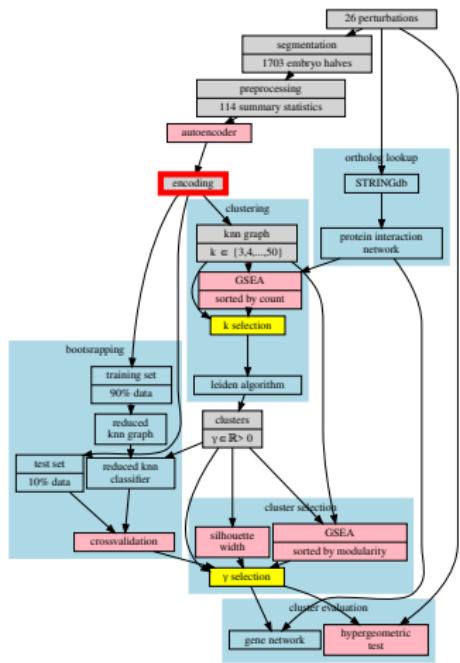




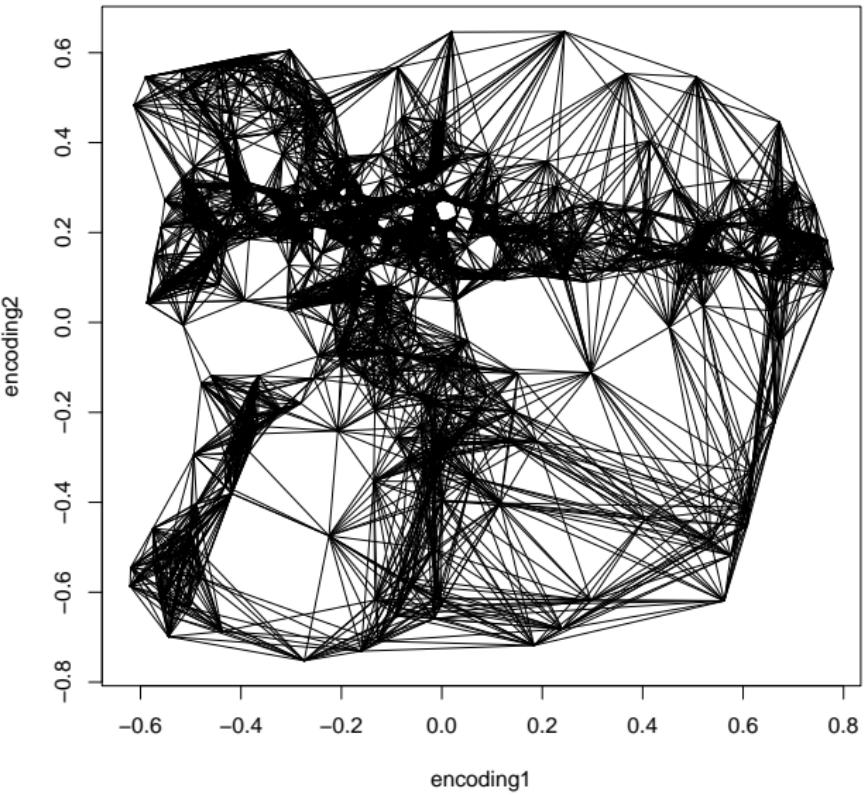
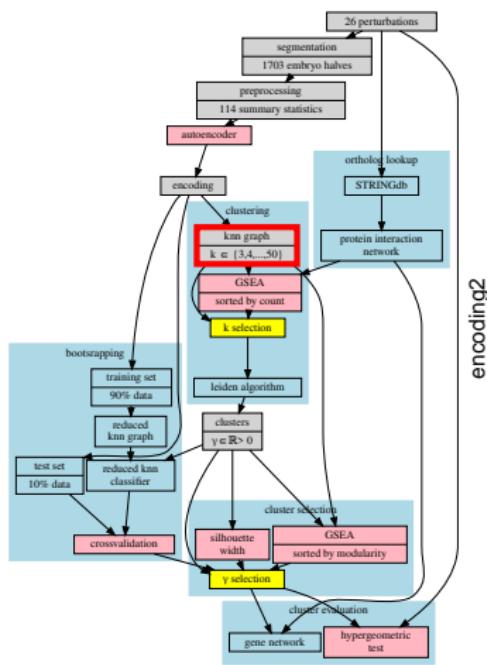
Dimension Reduction



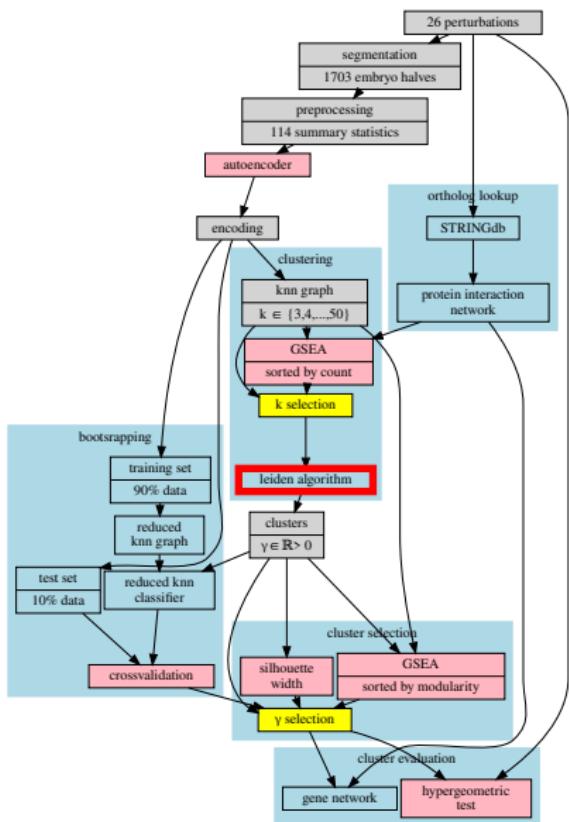
An autoencoder is a method of dimension reduction that uses a neural network to find a lower dimensional encoding which can be decoded to recover the input. This reduces exaggeration of distance due to the number of parameters measured.



Graph Representation



Clustering

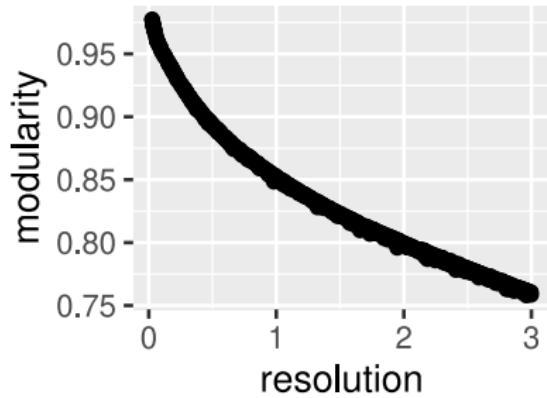
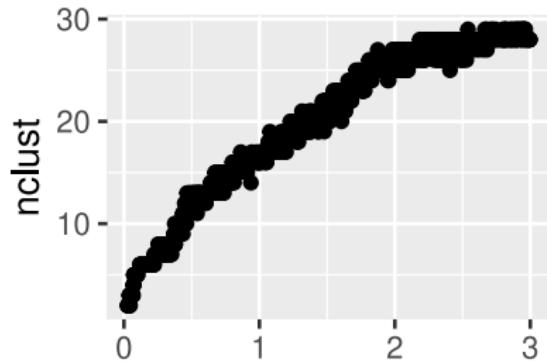
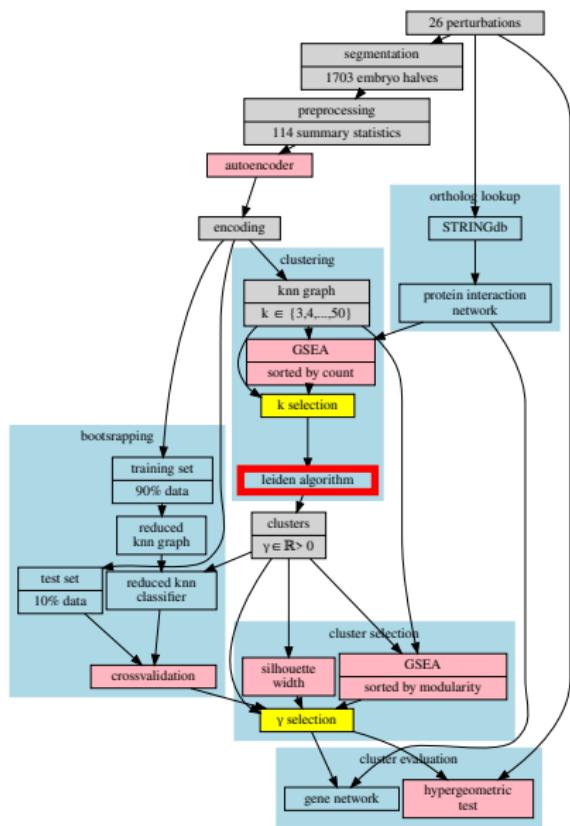


The leiden algorithm attempts to find a clustering that maximizes modularity H for a given graph and resolution γ . Modularity is defined as

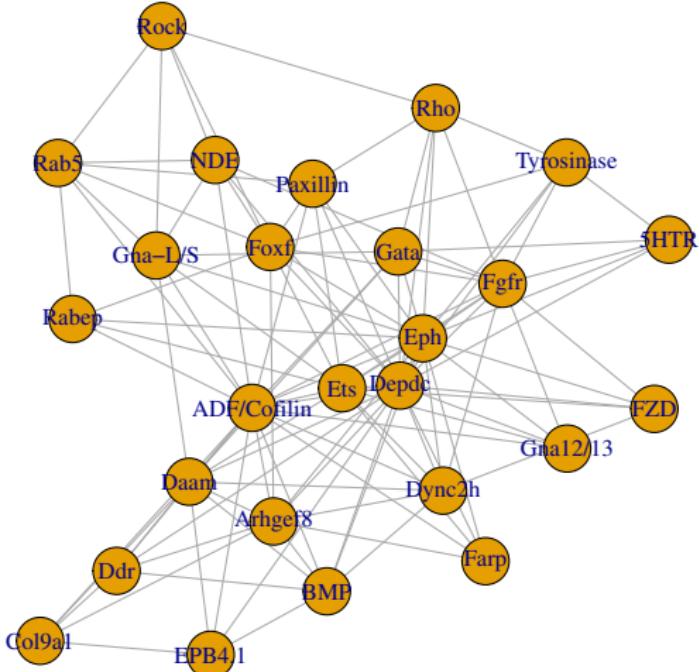
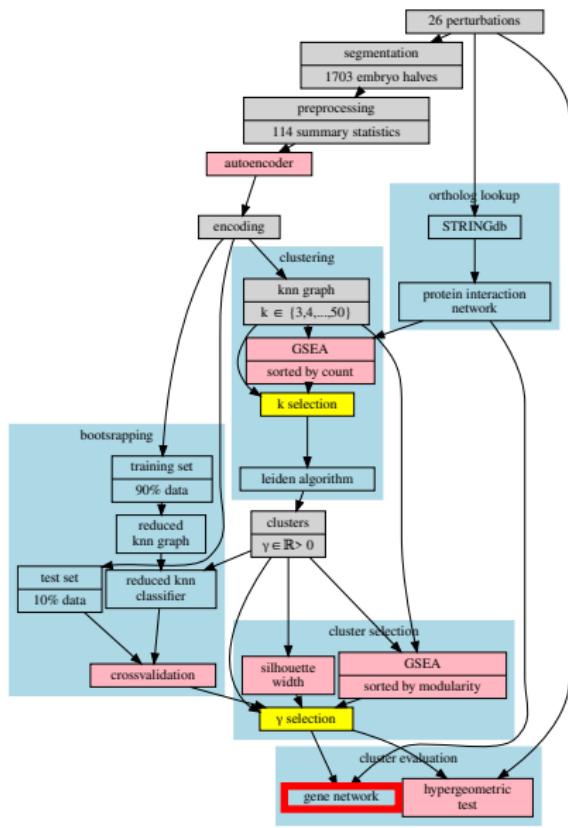
$$H = \frac{1}{2m} \sum_c (e_c - \gamma \frac{K_c^2}{2m})$$

where m is the average degree of the graph, e_c is the number of edges in cluster c , and K_c is the number of nodes in cluster c . This gives a measure of how well-connected clusters compared to expectation based on average degree of the graph and number of nodes in a cluster. A higher γ results in more clusters. ◀ ▶ ⏪ ⏩ ⏴ ⏵

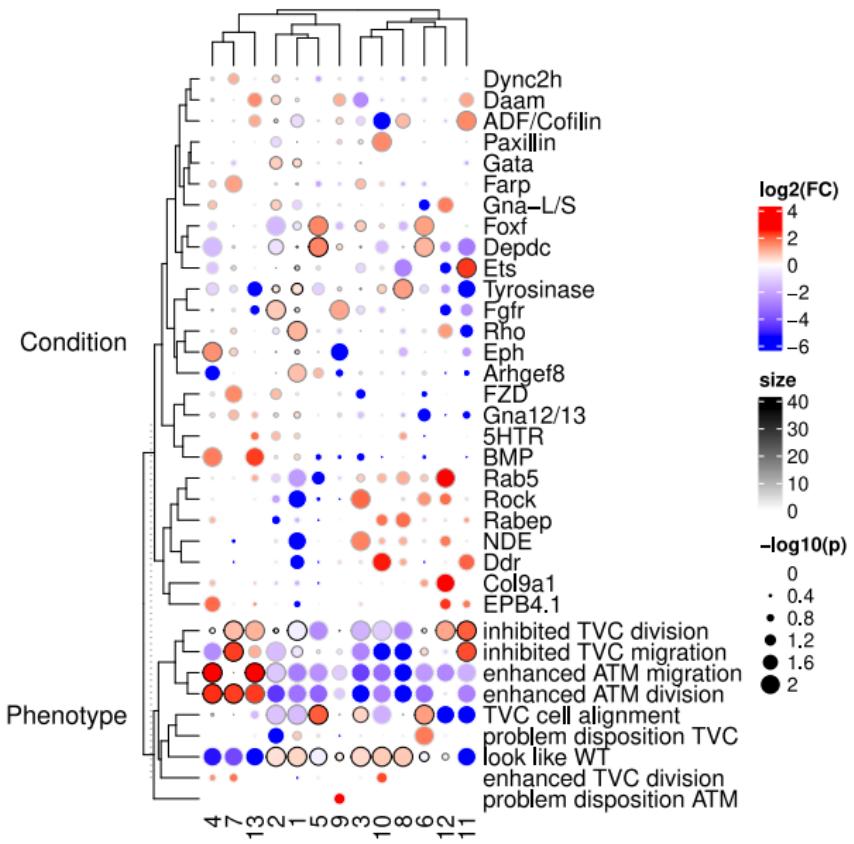
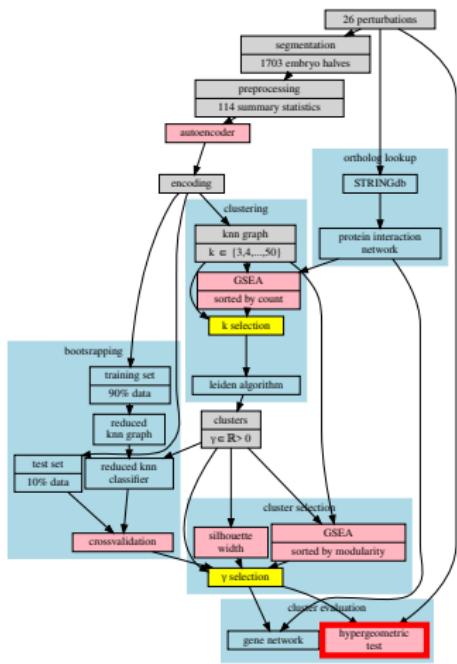
γ Selection



Gene Interaction Network



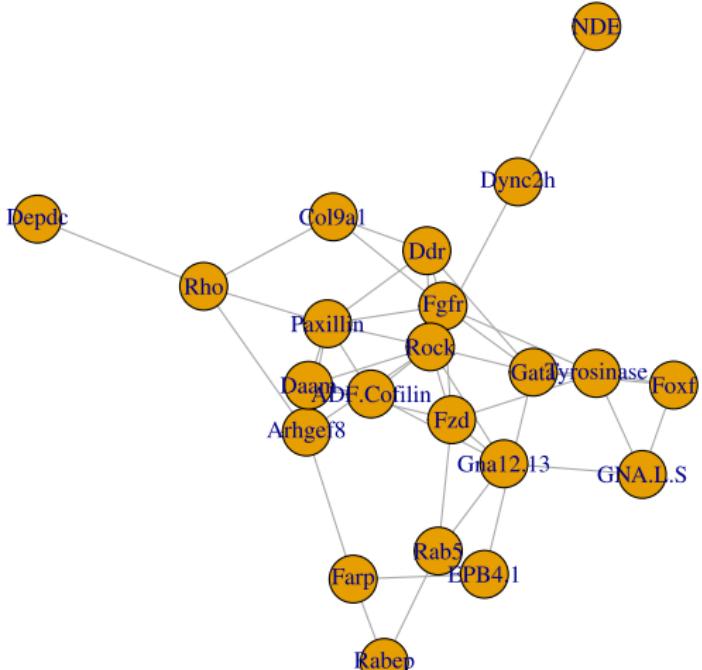
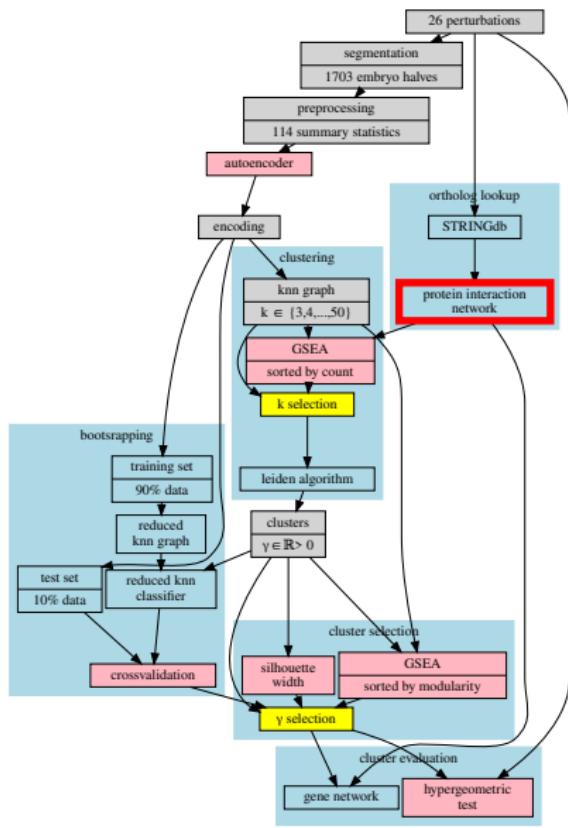
Cluster Analysis



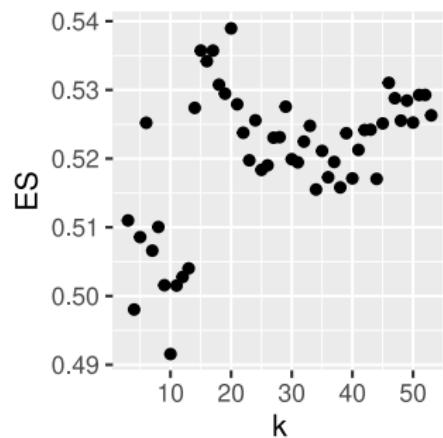
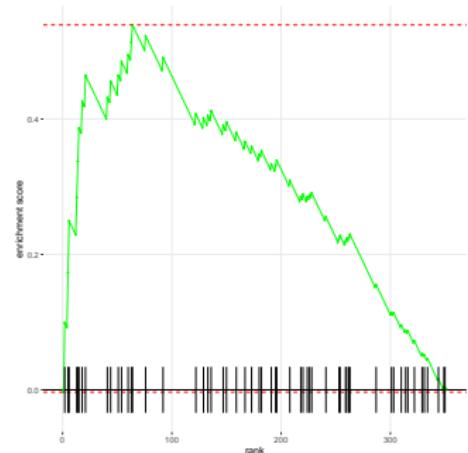
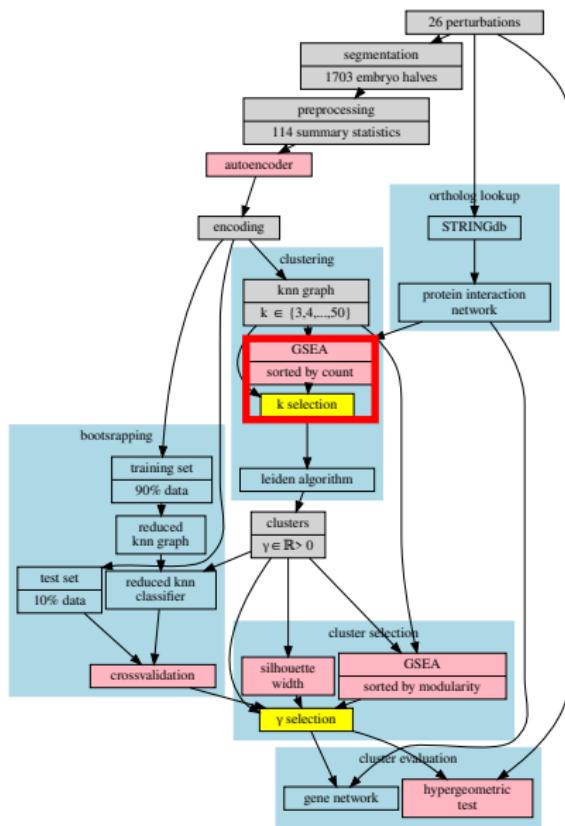
Thank You

Margaux Failla
Andreas Tjärnberg
Yelena Bernadskaya

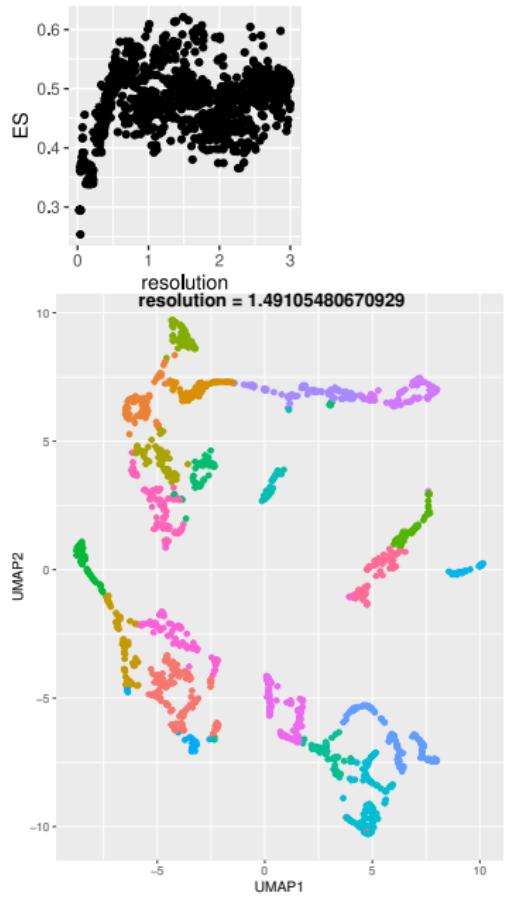
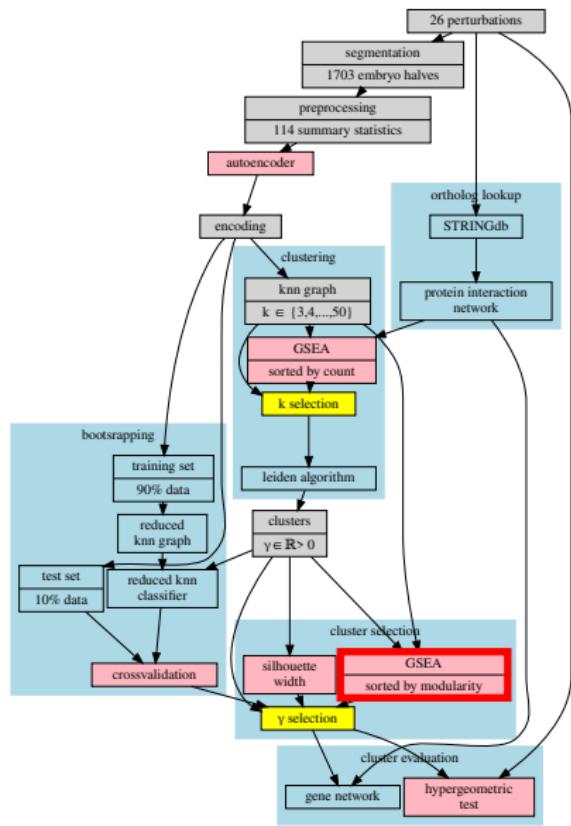
Known Protein Interactions



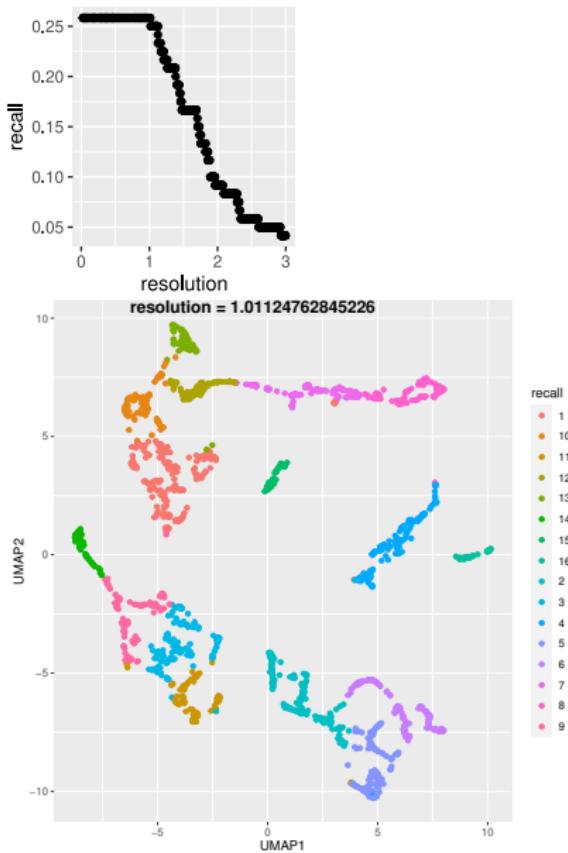
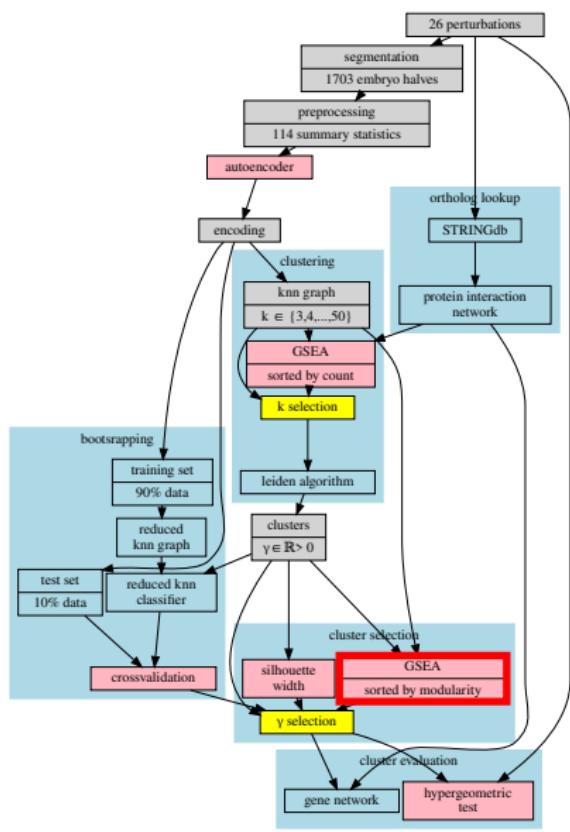
k Selection



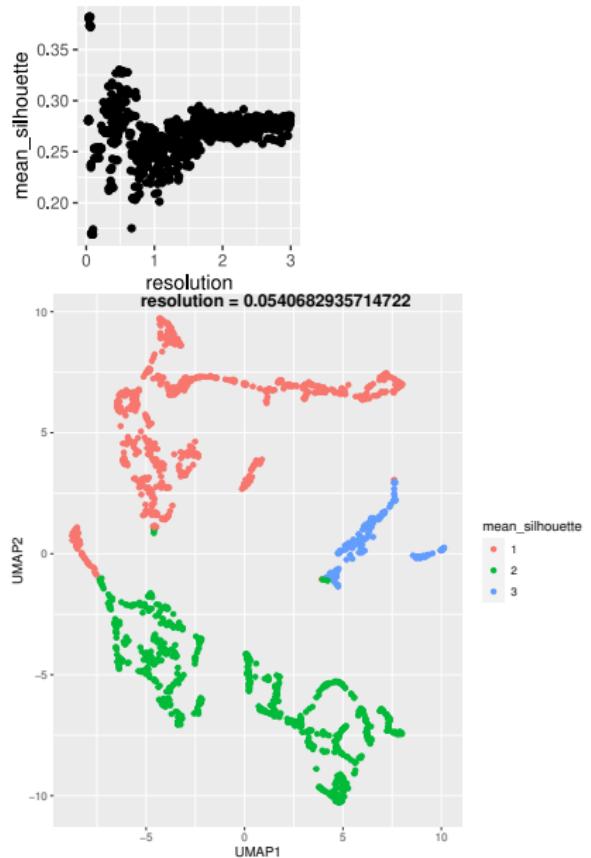
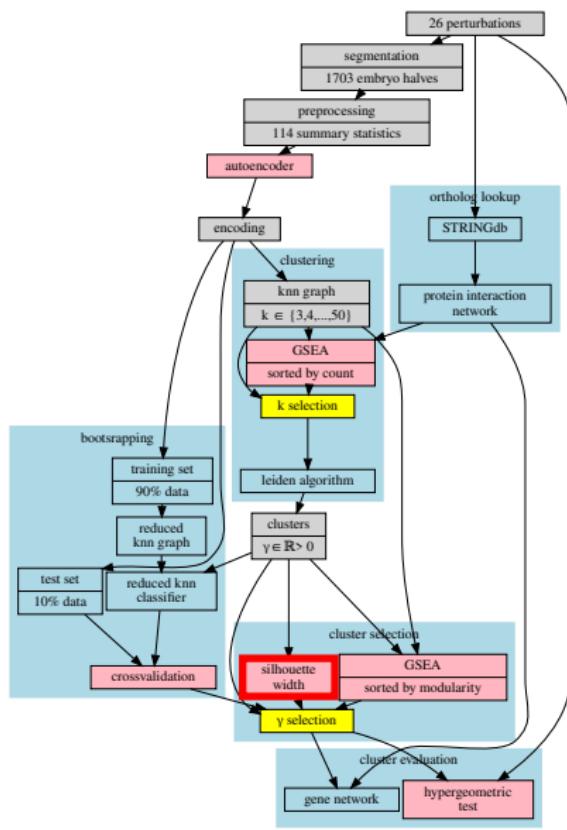
GSEA by H

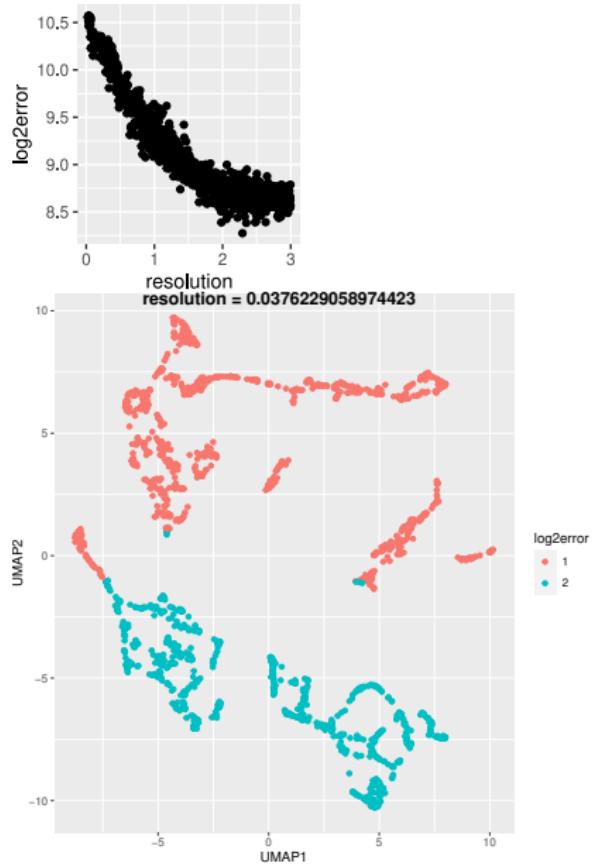
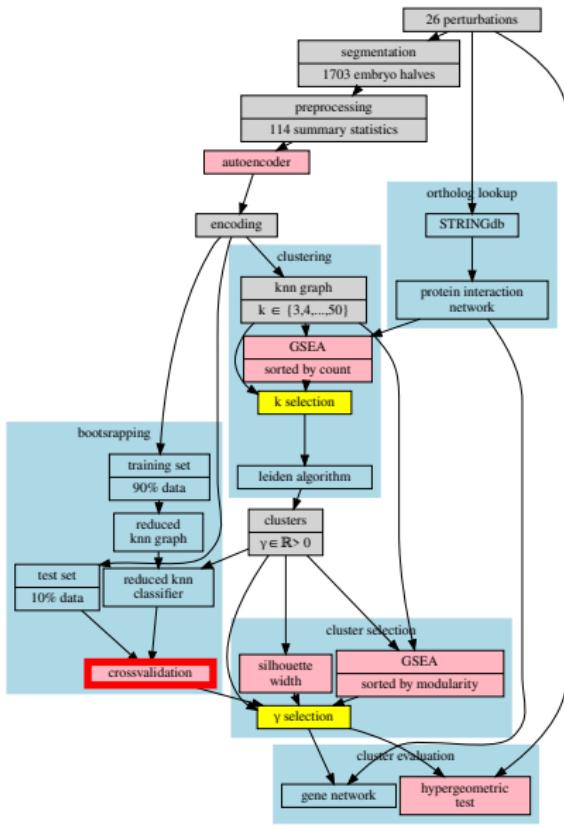


Recall for Known Interactions



Silhouette Width





$ES \times \text{recall} \times \text{silhouette width} \times -\log_2 \text{error}$

