

ABSTRACT

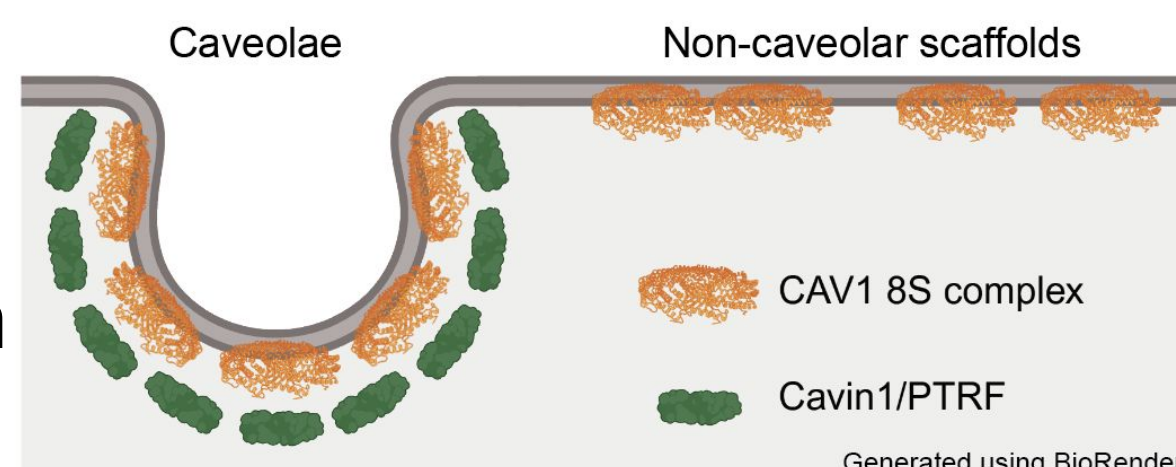
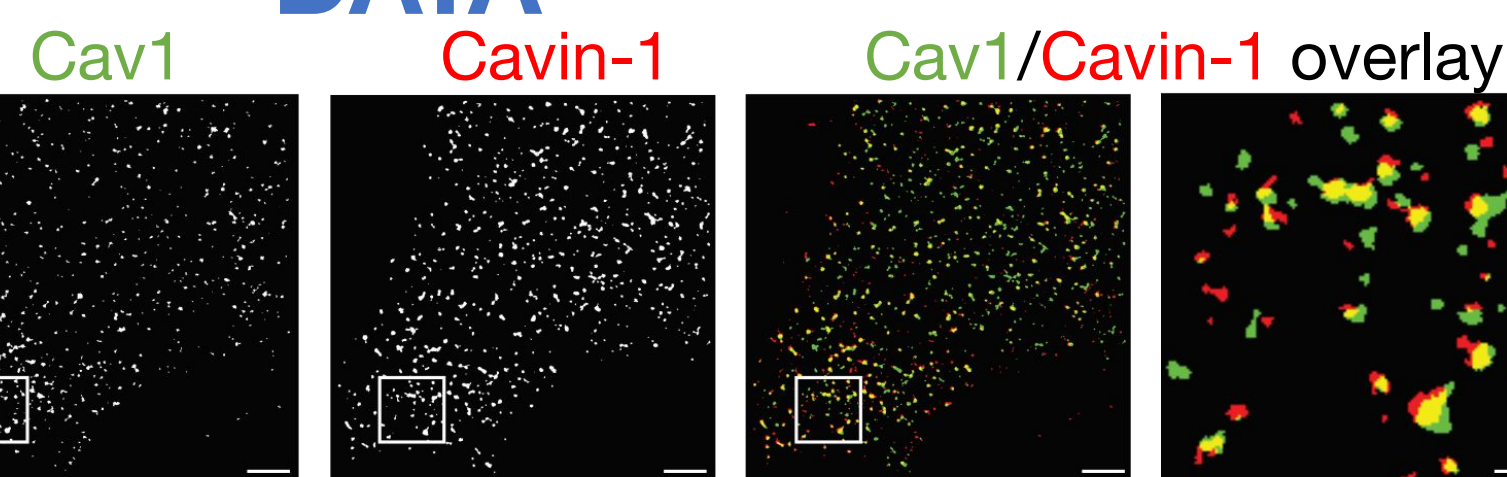
- Single-molecule localization microscopy (SMLM) captures nanoscale 3D point clouds of proteins, but identifying structural variability across different cell conditions is limited by a lack of object-level annotations
- We introduce siMILe, a weakly-supervised multiple instance learning (MIL) algorithm designed to identify condition-specific protein assemblies using only image-level or cell-level labels

CONTRIBUTIONS

- siMILe is the first application of MIL for 3D SMLM point cloud data
- Enhances instance classification through adversarial erasing, iteratively removing identified structures and retraining to ensure the model detects all discriminative structures
- The algorithm replaces standard thresholding with a symmetric 3-center k-means classifier

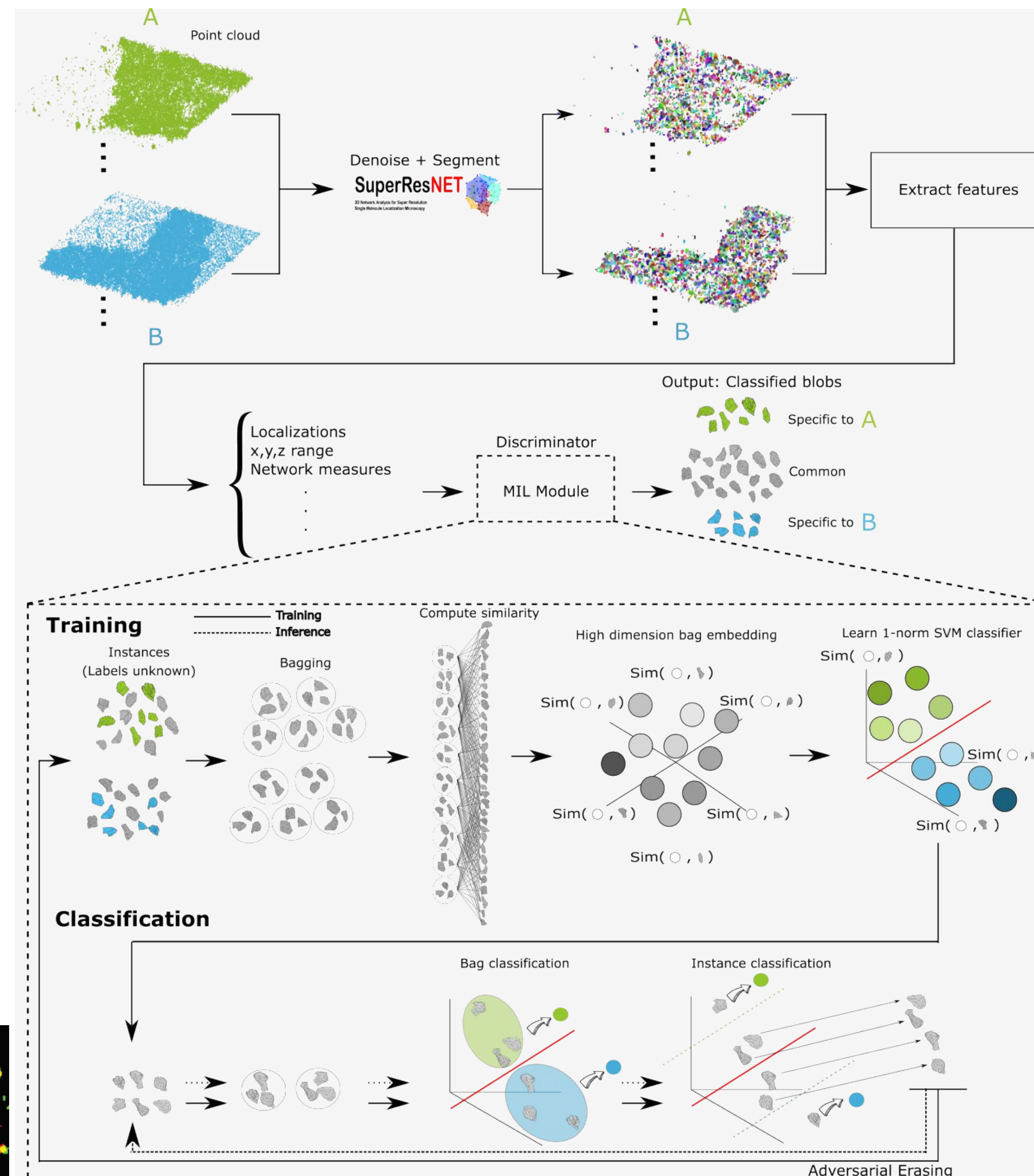
DATA

Cav1 is an integral membrane protein and localize to



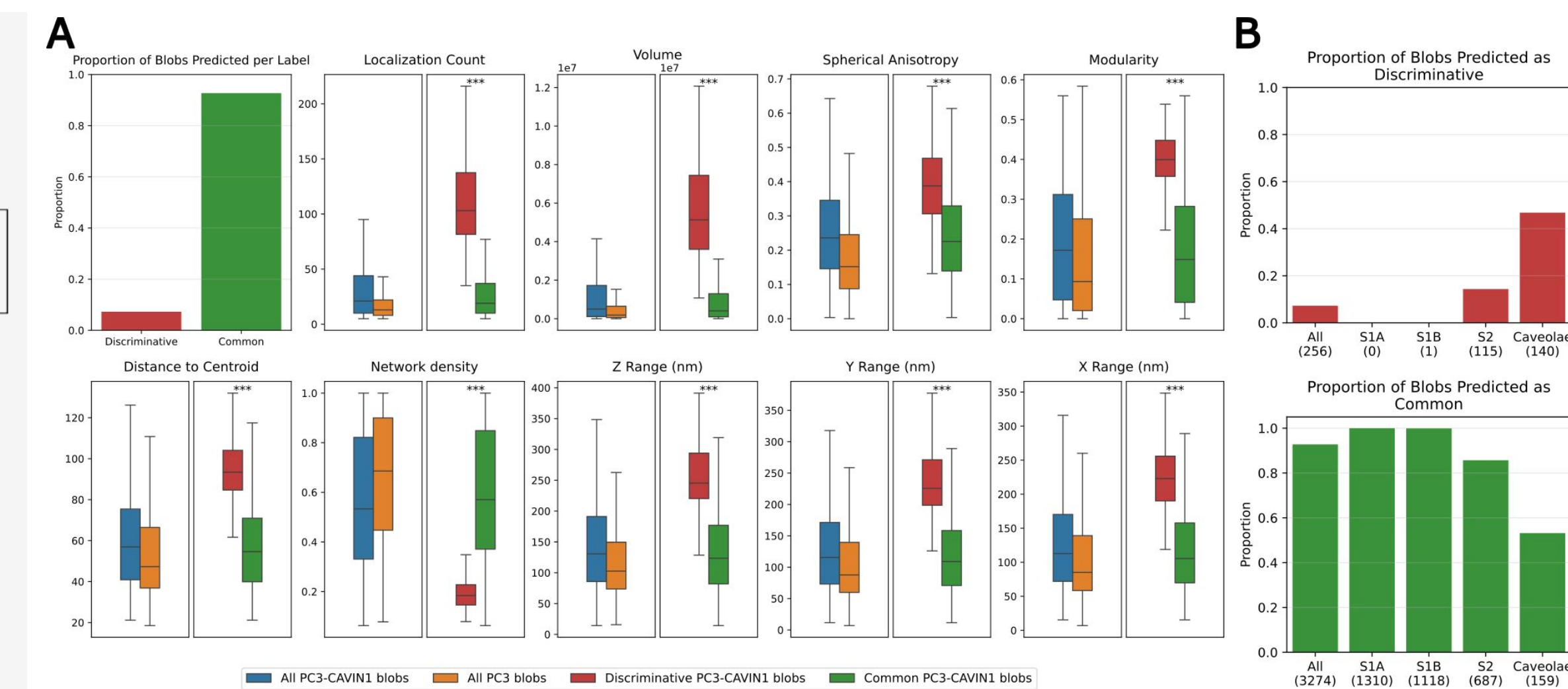
Cell Type	cav1	cavin-1	caveolae	scaffolds
PC3	+	—	—	+
PC3 transfected with cavin-1	+	+	+	+

METHODS

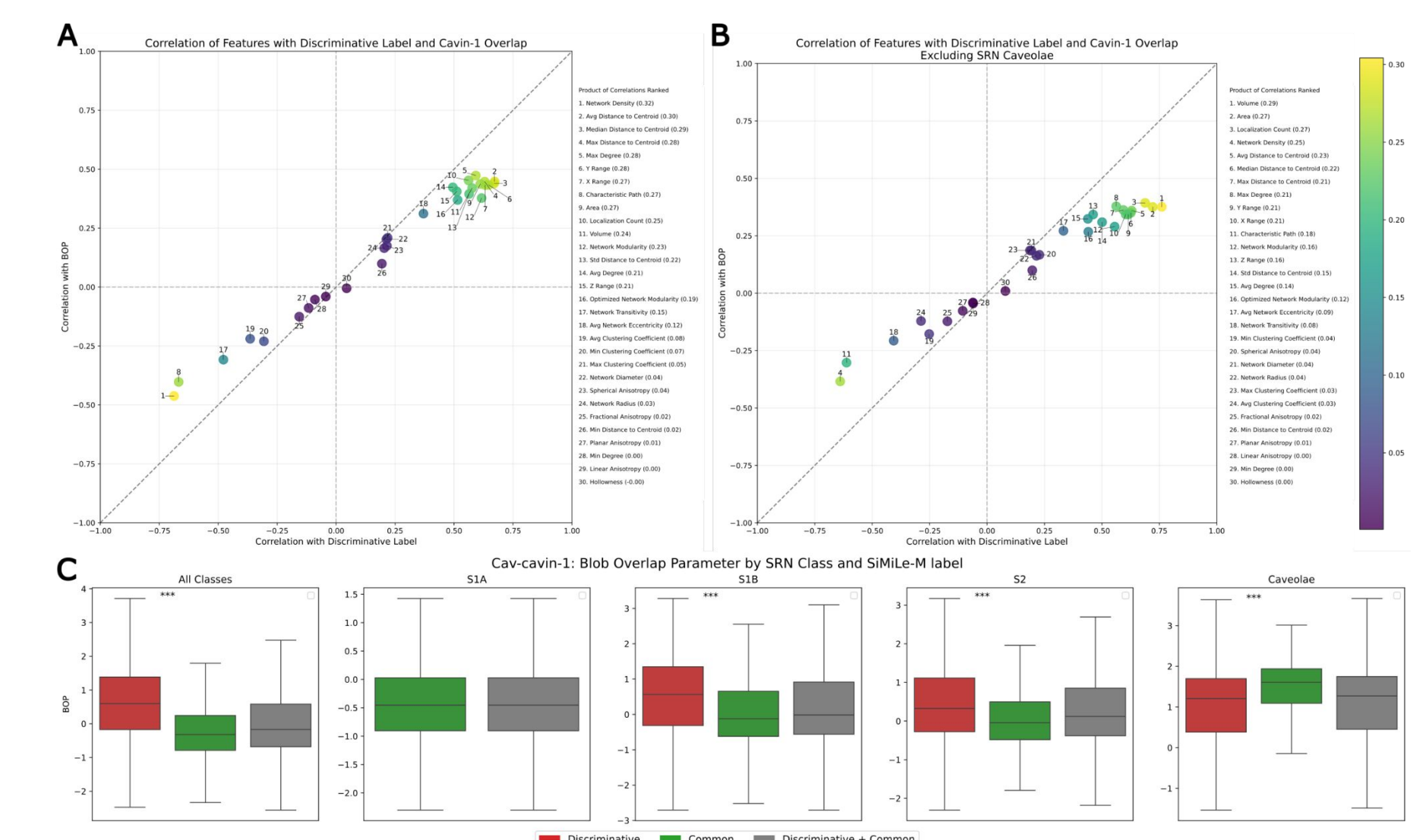


- SuperResNET first segments point localizations into clusters and maps them to a 30-dimensional feature matrix
- Blobs are grouped into labeled bags, and an L1-norm linear SVM learns to separate these bags based on their proximity to specific target concepts in the feature space
- Instances are scored and classified into three distinct clusters. Discriminative instances are erased, and the SVM retrains iteratively until bag classification performance drops below a set threshold

RESULTS



siMILe successfully identified caveolae and higher-order S2 scaffolds as discriminative to PC3 transfected with cavin-1 cells



Strong correlation between cavin-1 interaction and discriminant labels

CONCLUSION

- siMILe provides a weakly-supervised solution for discovering condition-specific changes in molecular architectures
- The tool enables novel biological discoveries by identifying condition-specific structures based purely on image-level labels