Contrastive feature representation for enhanced spatial transcriptomics analysis

Muhammad Aminu[1,#], Bo Zhu[2,#], Natalie Vokes[2,#], Hong Chen[2], Lingzhi Hong[2], Jianrong Li[4], Alissa Poteete[2], Monique B Nilsson[2], Xiuning Le[2], Cascone Tina[2], David Jaffray[3], Nick Navin, Lauren A Byers[2], Don Gibbons[2], John Heymach[2], Ken Chen, Chao Cheng[4], Jianjun Zhang[2,*], Jia Wu[1,2,*]

Authors' affiliations:

[1]Department of Imaging Physics, [2]Department of Thoracic/Head and Neck Medical Oncology, [3]Office of the Chief Technology and Digital Officer, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [4]Department of Medicine, Institution of Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA.

[#] contributed equally

* co-senior authors

Abstract

We developed a novel spatial transcriptomics (ST) feature representation method utilizing graph contrastive learning, designed to analyze multiple ST datasets while effectively distinguishing unique spatial structures within a target dataset. This approach minimizes the influence of common high-variance structures, enabling more precise identification of biologically relevant spatial domains. Initially, we applied our method to lung tissue samples, using a normal lung tissue sample as the background and an abnormal lung tissue sample as the target. The resulting contrastive feature representations successfully distinguished various spatial structures within the target tissue. These identified structures corresponded closely with pathologist-annotated regions, validating the method's accuracy.

When applied to a 10x Visium dataset from mouse brain (including anterior and posterior slices) from the Allen Institute, our approach demonstrated robust spatial domain identification, with the detected domains showing strong concordance with the Allen Brain Atlas. Further examination of the top five contrastive components revealed distinct spatial patterns linked to major anatomical regions, such as the cerebral cortex and choroid plexus. The top genes associated with each component exhibited clear spatial localization within specific brain regions. This highlights the potential of graph contrastive learning in detecting complex tissue architectures in spatial transcriptomics data, offering significant insights into tissue organization and function.