Background In routine cancer diagnosis, pathologists manually characterize tumor cells based on hematoxylin and eosin (H&E)-stained images. On the other hand, transcriptomic-based tumor molecular subtypes were shown to be associated with important clinical features including tumorigenesis and prognosis. Leveraging recent development of spatial transcriptomics (ST) which allows in-situ transcriptomic profiling of tissues, we aim to develop a first-of-its-kind machine learning (ML)-enabled integrated morphology-transcriptome tumor single-cell phenotyping approach.

Methods Two tissue sections each from tumor and adjacent-normal areas collected from a hepatocellular carcinoma (HCC) patient were profiled using 10× Visium ST platform. Using the companion H&E image, individual epithelial cells were segmented (StarDist algorithm) with 53 morphological and staining features extracted (QuPath v0.3.2). These cells were unsupervisedly clustered using encoder-based ensemble method where the optimal clustering solution was determined based on a consensus score of three clustering metrics. Phenotypic gene signatures of the cell clusters were determined through deconvoluting the ST data. Gene ontology (GO) analysis was done using single sample gene set enrichment, based on the molecular signatures database.

Results At the optimal clustering setting, 4 epithelial cell clusters, characterized by differential nuclear size, were detected individually in each HCC tissue (figure 1). Manual inspection by a pathologist (YZ) confirmed that the tumor epithelial cells demonstrated different nuclear sizes and revealed that the two smaller cell clusters looked relatively more well-differentiated, and ~1% found outside the tumor nest, suggesting potential epithelial to mesenchymal transition (EMT) activity. Whereas the two larger clusters were moderately-differentiated and demonstrated hyperchromatic nuclei and pleomorphism. GO analysis confirmed the upregulation of EMT in the smallest cluster, in both tumor tissues. While epithelial cells in the two normal-adjacent tissues appeared morphologically non-cancerous, the corresponding cell clusters contributed to similar cell fractions as that of the tumor tissues; two smaller clusters contributed to ~70% of the total cells across all tissues (figure 2). Cell clusters with similar nuclear size shared 30%-65% of the top 20 pathways across tissues, indicating inter-tissue phenotypic consistency. Cells were found near cell-type of its own followed by cell-type of similar size, suggesting preference cell clustering of similar phenotypes (figure 3).

Conclusions Our ML approach revealed four morphologically-transcriptomically distinct tumor cell subsets in the HCC tissues, with the smallest cells appeared EMT-like. We revealed intra-patient tumor cell heterogeneity yet phenotypic consistency across tissue sampling sites. Altogether, our proposed approach would enable more refined tumor cell phenotyping, advancing our understanding of tumor biology.

Acknowledgements I would like to thank NTU Undergraduate Research Experience on Campus (URECA) program for giving me opportunity to work on this project for the past year.
Abstract 1288 Figure 3  Distribution of the nearest neighbor distance of cells within a cluster to all 4 clusters, respectively in (A) tumor section 1, (B) tumor section 2, (C) adjacent-normal section 1, and (D) adjacent-normal section 2.