Visium HD combined with deep-learning-based cell segmentation on H&E images yield accurate cell annotation at single-cell resolution

Tianyou Luo[1], Mario Rosasco[1], Christine Hoeman[1], Sonal Khare[1], Daniel Gorski[1], Tim Rand[1], Richard Klinghoffer[1], Justin Guinney[1], Chi-Sing Ho[1]

[1]Tempus AI, Inc., Chicago, IL

Abstract Presentation # 8060

INTRODUCTION

Bulk and single-cell next-generation sequencing (NGS) have been valuable tools for characterizing gene expression profiles of tumor samples. However, their utility in investigating tissue architecture and cellular interactions in the tumor microenvironment (TME) is limited by the lack of spatial context. NGS-based Spatial Transcriptomics (ST) technologies have gained increasing attention for their ability to provide spatial context of gene expression, but they have been constrained by low resolutions until the recent launch of 10X Genomics Visium HD platform which achieves whole-transcriptome profiling at 2 μm resolution. However, the default binning of Visium HD at 8 μm resolution overlooks cell morphology and complicates downstream biological interpretations of the data by merging multiple cells in one bin or splitting large cells into multiple bins. Here we aim to solve that challenge by combining Visium HD data with custom deep-learning-based cell segmentations to yield accurate single-cell level data and enable downstream clustering and cell type annotations.

METHODS

- We collected primary tumor samples (one immune-active sample and one immune-inactive sample) from 2 patients with non-small cell lung cancer (NSCLC) and generated ST data with Visium HD platform.
- We trained a cell-segmentation neural network and applied it to H&E images to generate single-cell-level gene counts.
- Unsupervised clustering was performed on the single-cell level data and a large language model (LLM)-based cell type annotator was used to infer the cell types of each cluster.

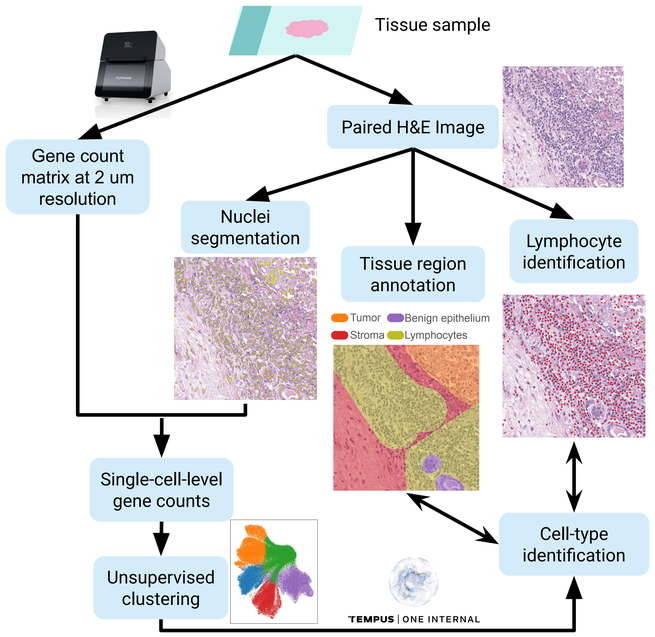

Figure 1. ST data generation and single-cell-level data analysis workflow.

SUMMARY

- We demonstrate the feasibility and advantages of analyzing Visium HD data at single-cell resolution using deep-learning-based segmentation models applied to H&E images.
- Cell clusters and LLM-based cell type annotations derived from single-cell resolution Visium HD data are highly consistent with pathologist annotations and morphology-based cell classification models.
- Analyzing Visium HD data at true single-cell level enables use of standard single-cell RNA-seq analysis toolboxes for downstream biomarker discovery and achieves enhanced biological interpretability compared to default 8 μm pixel resolution.

RESULTS

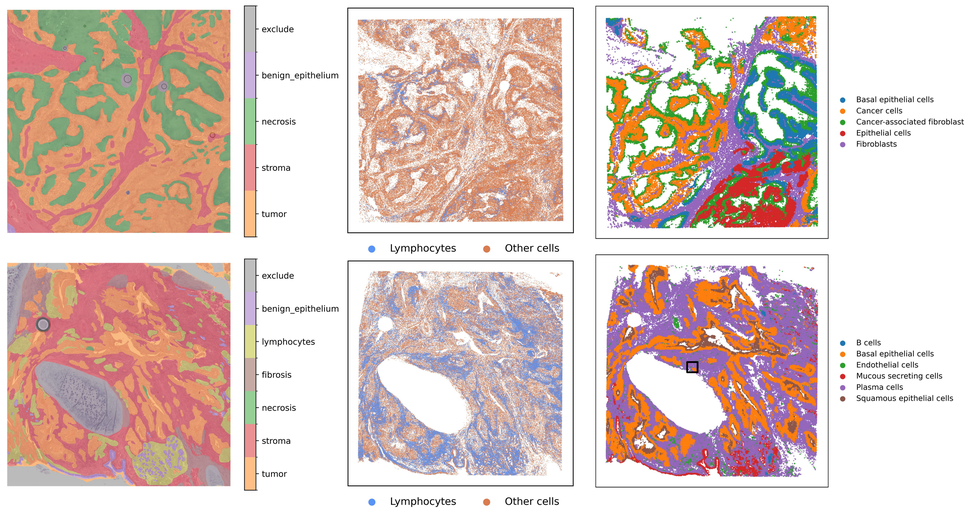

Figure 2. Visual comparison of automated cell type annotations (right panel) with pathologist's annotations (left) and results from H&E-based lymphocyte identification model (center) reveals similar spatial distribution patterns for both the immune-inactive sample (top row) and the immune-active sample (bottom row). The region in the black box of the immune-active sample is further studied in Figure 5.

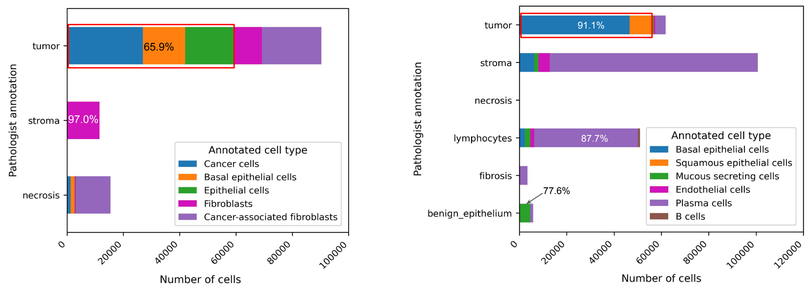

Figure 3. Automated cell type annotations align well with pathologist's annotations. In the immune-inactive sample (left), 65.9% and 97.0% of cells in annotated tumor and stroma regions were classified as cancer/epithelial cells and fibroblasts respectively. In the immune-active sample (right), 87.7% of cells in annotated lymphocyte regions were classified as plasma cells or B cells, 91.1% of cells in annotated tumor regions were classified as basal/squamous epithelial cells, and 77.6% of cells in annotated benign epithelium regions were classified as mucous secreting cells. Also, 76.9% of H&E-identified lymphocytes were classified as plasma cells/B cells.

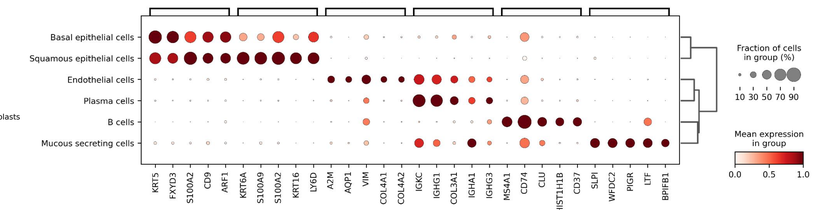

Figure 4. Expression of marker genes identified for each cluster in the immune-active sample, demonstrating the distinct gene expression profiles that differentiate each cell cluster. Cell type assignments (left) are generated by prompting a LLM with each cluster's marker genes.

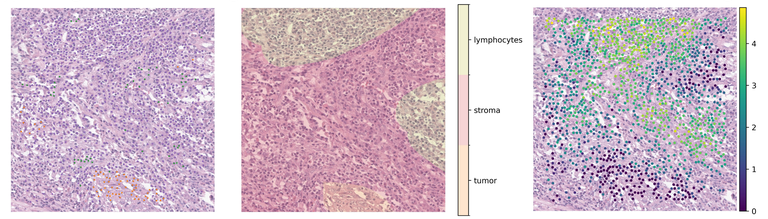

Figure 5. H&E image and cell type annotations (left), pathologist's annotations (center) and log-transformed IGHG1 gene expression (right) for the selected region in Figure 2 for the immune-active sample. Visualizations and marker gene expressions confirmed the annotated stroma region is heavily infiltrated by lymphocytes.

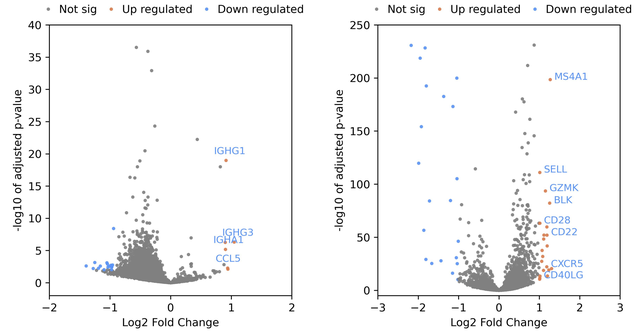

Figure 6. Differential gene expression analysis results between H&E-identified lymphocytes and non-lymphocytes for the immune-inactive sample (left) and the immune-active sample (right). Multiple lymphocyte marker genes were upregulated in H&E-identified lymphocytes, demonstrating consistency between morphology- and transcript- based models. P-values derived from a t-test.

ACKNOWLEDGMENTS

We thank Dana DeSantis from the Tempus Science Communications team for poster development.

Correspondence: tianyou.luo@tempus.com