

## Differential Gene Expression and Functional Enrichment Analysis of Smoothed PDAC Spatial Transcriptomics Slides

Ben Duggan<sup>1</sup>, Yusong Liu<sup>2</sup>, Tongxin Wang<sup>3</sup>, Jie Zhang<sup>4</sup>, Travis S. Johnson<sup>5</sup>, Xiufen Ye<sup>2</sup>, and Kun Huang<sup>5,6</sup>

<sup>1</sup>Indiana University School of Medicine; <sup>2</sup>Harbin Engineering University, Collage of Intelligent Systems Science and Engineering; <sup>3</sup>Indiana University Bloomington, Department of Computer Science; <sup>4</sup>Indiana University School of Medicine, Department of Medical and Molecular Genetics; <sup>5</sup>Indiana University School of Medicine, Department of Biostatistics and Health Data Science; <sup>6</sup>Regenstrief Institute

### Background:

Cancer heterogeneity can impact diagnosis and therapeutics. This has been studied at the cellular level using single cell RNA sequencing (scRNA-seq) but not spatially. Spatial transcriptomics (ST) has recently enabled measuring transcriptomes at hundreds of locations in a tissue, revolutionizing our understanding of tumor heterogeneity. Low RNA quantities in scRNA-seq and ST introduce missed reads and noise. Multiple methods have been developed to impute and smooth scRNA-seq data including MAGIC and SAVER. To date, only our newly developed spatial and pattern combined smoothing (SPCS) method has been developed specifically for ST. We compared the biological interpretability of these smoothing methods to determine which method best informs tumor heterogeneity.

### Methods:

Ten ST slides from six patients with PDAC were computationally smoothed using SAVER, MAGIC, and SPCS. ST spots from unsmoothed and smoothed slide were split into *TM4SF1* high or low expression groups and differentially expressed genes (DEG) found. Significant up- and down-regulated DEGs were used for functional enrichment analysis (EA) to compare the biological insights each method provides.

### Results:

DEGs were found in eight samples, with SPCS finding the most DEGs in six of the eight samples. The number of EA terms generally correlated with the numbers DEGs. SPCS had the most up-regulated enriched terms for every ST slide. Top ten up- and down-regulated EA terms were presented from one ST slide to demonstrate that SPCS gives more biologically interpretable results. Lastly, we present SPCS terms previously been reported in PDAC or involved in the *TM4SF1* cancer pathway.

### Conclusion:

Smoothing ST data using SPCS provided more DEGs and more useful EA terms. The EA terms found using SPCS are more consistent with our current understanding of PDAC. Not smoothing, MAGIC, and SAVER missed many important terms which SPCS found. Compared to other methods, SPCS smoothed data is more interpretable.