

sci-Space data and code

Thanks for taking the time to review our paper! If there are any questions or problems please feel free to reach out to us via the editor or directly through email (sanjays@uw.edu).

In this folder you will find all the analyses and data needed to recreate almost all of paper in its entirety.

Data

The data provided here are processed data types, namely count matrices for single cell data, and rescaled images for imaging data. For reference, code used to map and process the raw reads can be accessed online on our [github repository](#).

Processed and raw data can be downloaded from NCBI at GSE166692.

Code

At the header of each script we have tried to include a high-level summary of the script's contents and the goal of the analysis. Some of these scripts rely cluster computing due to the size of the count matrices handled. If this is the case, we have also noted this in the header.

Changing the variable `space_directory` to the downloaded folder titled `Submission` should enable running all of the scripts provided. We have tried to use only relative paths within each script.

Use the code + data

The final cell dataset (CDS) object with spatial, annoatomical and celltype annotations can be downloaded from GEO (Acession #GSE166692) and be used to reconstruct this object in the form of choice --

`GSE166692_sciSpace_cell_metadata.tsv.gz` Contains all the cell metadata such as slide, coordinate, cluster and cell type. Saved as a tsv object.

`GSE166692_sciSpace_count_matrix.mtx.gz` Contains all the gene expression data for the 14 E14 slides presented in the paper. Saved as a matrix marker file.

`GSE166692_sciSpace_gene_metadata.tsv.gz` Contains all the gene metadata -- associated with each row of the count matrix.

`GSE166692_sciSpace_hashTable.tsv.gz` Contains the raw unprocessed labeling counts recovered after sequencing. Spatial and slide metadata for each cell has been extracted from this table and incorporated into the cell metadata.

`GSE166692_sciSpace_imaging.RDS.gz` R data object containing slide hulls, images and polygons associated with each slide. Used in plotting.

Notebooks

The notebooks follow sequentially and track with the paper. Notebooks often build upon the result of prior analyses

`Notebook 0` - Initial notebooks analyse data from control experiments, oligo-layout for spatial grid and code used to preprocess sci-Space imaging data.

`Notebook 1` - Python snippet used to run scrublet

`Notebook 2` - Notebooks mapping single cell transcriptomes to slides and looking at initial QC metrics

`Notebook 3` - Perform dimensionality reduction, clustering and subclustering on the single cell transcriptomes

`Notebook 4` - Use K-nearest neighbor label transfer and Garnett to label cell types recovered

`Notebook 5` - Visualize recovered celltype labels and map cells to spatial positions and refine these spatial calls.

`Notebook 6` - Create organ regions of interest, visualize celltype distribution spatially and render virtual in situ. Make comparisons with other data types.

`Notebook 7` - Explore spatially restricted and cell state restricted gene expression

`Notebook 8` - Investigate the relationship between physical space and transcriptome dissimilarity. Calculate variance explained statistics using spatial position and celltype labels.

`Notebook 9` - Analyse pseudotemporal, spatial trajectories in the developing brain.

`Notebook 10` - Write out files for GEO submission