Spatial Transcriptomics Workflow
for Integrated Analysis and Precise Cell-type Annotations

Jyotsana Negi, Raj Deep Jha, Aishwaryaa Nallasivan, Ashish Kumar Choudhary, Vijay Dwivedi, Nihesh Rathod, Arivusudar Everad John, Shrutee Jakhanwal, Shanmukh Katragadda, Vamsi Veeramachaneni, Pankaj Kumar, Radhakrishna Bettadapura

Strand Life Sciences, Bangalore, India

VISIT OUR **WEBSITE**

strand
WE SOLVE PROBLEMS IN GENOMICS.

Conference Name
American Spatial Biology Congress-
Dates - May 9 - May 10, 2024
Place - San Diego, USA

Contact
Shrutee Jakhanwal
Product Manager
+91 9650040413
shrutee.jakhanwal@strandls.com

Introduction

1. Spatial transcriptomics (ST) has emerged as a powerful tool for understanding the complexity of tissues by providing a spatial context to gene expression data. However, identifying and accurately localizing cell types remains a significant challenge in this field.
2. This study introduces a custom R-based workflow tailored for analyzing 10X Visium spatial transcriptomics data. The comprehensive computational pipeline covers all stages of data analysis, including raw data processing, quality control, normalization, batch effect correction, identification of spatially variable genes, and cell type annotation, providing a thorough end-to-end solution.
3. To demonstrate the effectiveness of the pipeline, the data analysis is performed on the in-house breast cancer datasets.

HIGHLIGHTS

1. This pipeline is primarily composed of three main modules : **Quality control (QC) module, single-sample analysis module, and combined-sample analysis module**
2. An **automated** approach is integrated into the workflow for filtering outlier spots using robust statistical calculations.
3. In multi-slice data integration analysis the batch effects arise from technical variation across samples. To address this, the pipeline incorporates several specialized integration tools like **CCA, Harmony, and Liger** to effectively correct batch effects.
4. A key challenge in spatial transcriptomics is identifying cell types and accurately locating them on tissue slides. To tackle this, the pipeline incorporates various annotation algorithms, including **ScType, SingleR, Seurat label transfer,** and **robust cell-type decomposition (RCTD),** to enhance cell-type identification and localization precision.

Spatial Transcriptomics Workflow

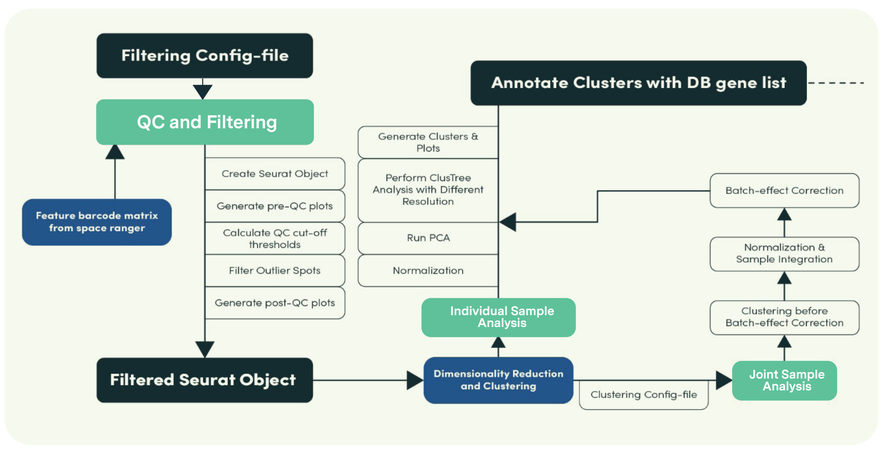


Fig 1 - Workflow of in-house developed spatial transcriptomics pipeline for 10X Visium data analysis

METHODS

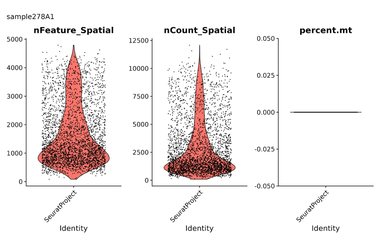
Raw Data Processing

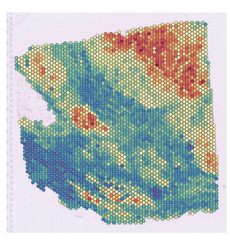
The workflow includes the Space Ranger pipeline to generate gene expression matrices from raw sequencing data produced by the barcode-based 10X Visium technology. Additionally, the workflow offers the capability to directly process publicly available matrices and images.

Quality Control and Filtering

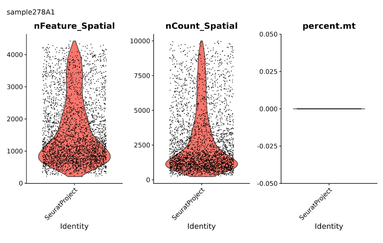
The QC module calculate QC metrics such as library size, expressed features, mitochondrial genes mapping, to identify and eliminate outliers and ensure that high quality spots were included for downstream analysis.

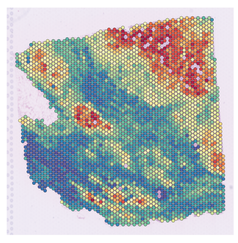
(a) Pre QC





(B) Post QC





(c) QC statistics

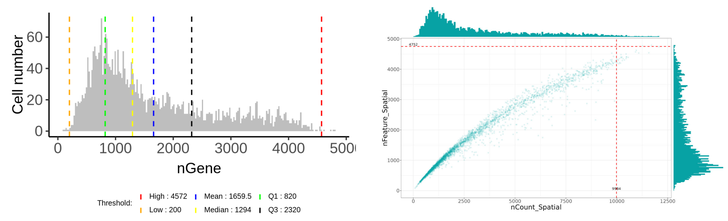


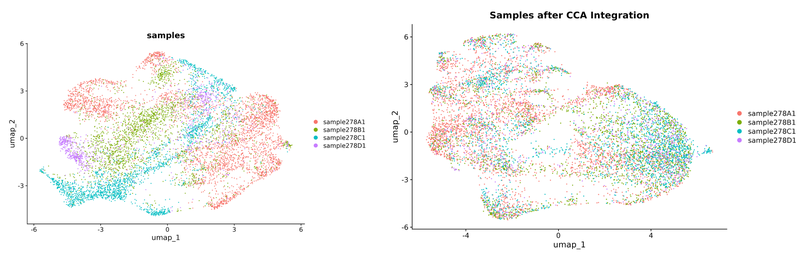
Fig 2 - Visualisation generated by QC module. (a) Pre QC - Violin plot showing the distribution of calculated QC metrics and Spatial plot for demonstrating spatial heterogeneity. (b) Post QC - Violin plot and Spatial plot. (c) Example of plots displaying calculated threshold values. Left - Distribution plot for nGene marked with different threshold values. Right - Scatter plot between nCount and nFeature marked with selected cut off value.

Data Normalization

Both SCTransform and LogNormalization based data normalization methods are included in the pipeline.

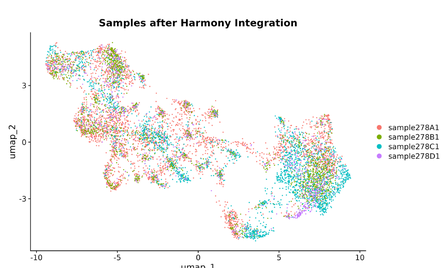
Batch Effect Correction

The pipeline offers the flexibility to include or exclude any sample for integration and comparative analysis across multiple slices. It also provides various options, multiple integration tools for correcting batch effects in the analysis.



(a) UMAP for Samples without integration

(b) UMAP for Samples with CCA Integration

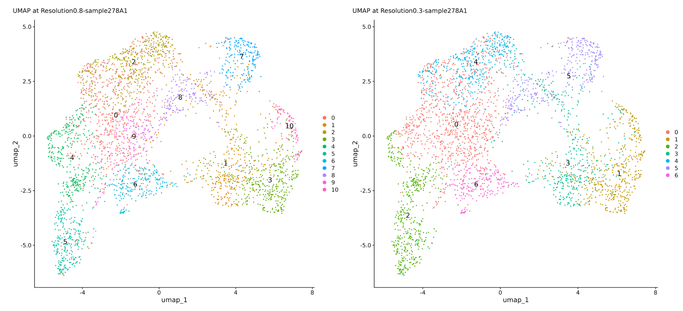


(c) UMAP for Samples with Harmony Integration

Fig 3 - Plots demonstrating different batch effect correction methods for Integration analysis

Dimension Reduction and Clustering

PCA and UMAP are employed for dimensionality reduction, while traditional machine learninf clustering techniques like K-nearest nighbors and Leiden clustering are utilized for clustering purposes. Furthermore, the workflow not only provides clusters at custom resolution but also generates plots for multiple resolutions.



(a) UMAP at resolution 0.3 (b) UMAP at resolution 0.8

Fig 4 - Example of UMAP generated at multiple resolution

Finding Markers

The evaluation of identified clusters is carried out by identifying differentially expressed genes (DEGs) both within and across all clusters, while also marking spatially variable genes (SVGs) using metrics like Moran's I statistic and marker-variogram. The pipeline generates visualizations such as heatmaps, violin plots, and spatial plots to display the identified markers.

Cell type Annotation

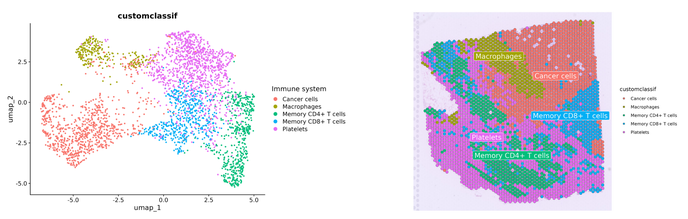
The pipeline incorporates marker gene-based and reference-based annotation as the main methods for cell annotation. Since Visium ST data lacks deep single-cell-level resolution, the applied algorithm also enables integration of scRNA-seq data with ST data, enhancing the understanding of cell-type distribution architecture.

SCType, marker gene-based annotation utilise the manually curated cell-type specific genes from database such as PanglaoDB, CellMarker 2.0, MsigDB or ACT.

SingleR, reference based annotation method leverages publicly available datasets from portal such as NCBI, Blueprint, Encode, Human Primary Cell Atlas, Database of Immune Cell Expression (DICE), Immunologic Genome Project (ImmGen).

Seurat label transfer is a **mapping approach** used to assign cells from scRNA-seq data to spatial locations in histological sections.

RCTD is a **deconvolution** algorithms which estimate the composition of cells within each physical spot by transferring cell-type signatures defined by scRNA-seq.



(a) UMAP (b) Spatial Plot

Fig 5 - Visualization for SCType annotation

Concordance Test

For validating the pipeline's results, an automated module has been created that overlays the annotations and regions of interest (ROIs) from Whole Slide Imaging (WSI) by pathologists and cell type annotations produced by the in-house ST pipeline.

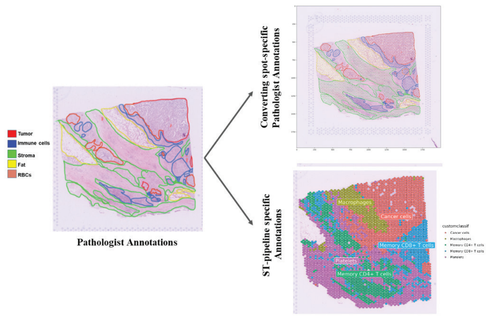


Fig 6 - Illustrating the concordance test overlay of Visium ST annotated data results from the pipeline compared to pathology WSI data