Standard Operating Procedure (SOP) for R based analysis of scRNA-seq Data

**Prerequisites**

* R environment with the following libraries installed: Seurat, future, reshape2, clustree, scales, ggplot2.
* Input data: Processed Seurat objects for each sample stored as RDS files.

**Functions Overview**

1. integrate\_samples: Integrates multiple Seurat objects, aligning and integrating data from different samples.
2. find\_clusters: Identifies clusters within the integrated Seurat object.
3. makeplots: Generates various plots for data visualization, including UMAP plots, clustree plots, and bar plots for cell counts and fractions.
4. DEfiles: Performs differential expression analysis between specified groups using the MAST method and saves the results.
5. fractable: Generates a table of normalized cell fractions for further analysis.
6. aggregatePvals: Aggregates p-values from multiple differential expression analyses.
7. DEexp: Finds differentially expressed genes, calculates average expression, and writes results to files.

**Step-by-Step Usage**

Data Preparation

Ensure your scRNA-seq data are pre-processed and stored as Seurat object RDS files named according to your sample identifiers.

Integration

Use integrate\_samples(samples, ref = 1) to integrate your Seurat objects. The samples argument is a vector of sample identifiers (names of your RDS files without the .rds extension). ref specifies the reference sample for integration by its position in the samples vector.

Clustering

Post-integration, apply find\_clusters on the integrated object to identify clusters at various resolutions. This function also filters out results with more than 15 clusters.

Visualization

Generate comprehensive visualizations using makeplots. Specify the integrated object, a base name for output files, and optionally the number of replicates and clusters for detailed plots.

Differential Expression Analysis

Conduct differential expression analysis with DEfiles, specifying the integrated object, a base name for output files, treatments, controls, and superclusters for comparison.

Fraction Table Generation

Create a table of normalized cell fractions using fractable, providing it the integrated Seurat object.

P-value Aggregation

Aggregate p-values from different analyses using aggregatePvals, specifying the method and p-value adjustment method.

DE Genes and Expression

For detailed differential expression analysis and average expression calculation, use DEexp, specifying comparisons and object names.

Notes

* Parallel Processing: The script makes use of the future package for parallel processing to speed up computations. Adjust the workers parameter in plan("multiprocess", workers = X) calls as per your system's capabilities.
* Memory Management: Large datasets may require adjusting the future.globals.maxSize option to increase the memory limit.
* Customization: Depending on the analysis depth and specific requirements, some function parameters (e.g., resolution for clustering, number of PCs for dimensionality reduction) may need adjustment.

**Running the Script**

1. Load the script into your R environment.
2. Import the RDS file you built during pre-processing into your R environment
3. Call the functions sequentially as per your analysis plan, adjusting parameters as necessary.
4. Check the generated plots and data files in your working directory for analysis results.