SpatialGEE Integration Tutorial

1. Introduction

This tutorial demonstrates how to use **SpatialGEE** for spatial co-profiling integration of ATAC and RNA data. We will:

- Perform differential expression analysis using `run_gee_gst()` and `run_wilcoxon()`.
- Use IMIX for integrative analysis based on P-values.

2. Installation and Loading

To use the **SpatialGEE** package, install it and load it into R session:

```
if(!requireNamespace("devtools", quietly = TRUE))
    install.packages("devtools")
devtools::install_github("yishan03/SpatialGEE", quiet = TRUE)
library(SpatialGEE)
```

3. Example Dataset

The package includes an example dataset, `coprofile_example_data`, based on a subset of the processed spatial epigenome-transcriptome mouse brain co-profiling data (Zhang et al., 2023).

3.1. Dataset Description

- The dataset contains spatial co-profiling data (ATAC & RNA), including metadata and gene-level measurements across 9,215 cells.
- Metadata columns include:
  - `Barcodes`: Cell barcodes.
  - `x`, `y`: Spatial coordinates.
  - `Pathology.Annotations`: Pathology labels (`"non-Corpus callosum"` and `"Corpus callosum"`).
- Gene expression and accessibility data include 100 selected genes: `Gabbr2`, `Pde7b`, ..., `Itga8`.

3.2. Dataset Example

```
data(coprofile_example_data)
atac <- coprofile_example_data$ATAC
rna <- coprofile_example_data$RNA
```

```
head(coprofile_example_data$ATAC) %>%
  dplyr::select(1:7)
#>           Barcodes  x y Pathology.Annotations   Gabbr2    Pde7b    Rims1
#> 1 AAACATCGAACTCACC 31 2   non-Corpus callosum 1.571669 1.268047 1.219150
#> 2 AAACATCGAAGAGATC 30 2   non-Corpus callosum 1.523837 1.406939 1.221459
#> 3 AAACATCGAAGGACAC 29 2   non-Corpus callosum 1.516391 1.580516 1.247479
#> 4 AAACATCGAATCCGTC 28 2   non-Corpus callosum 1.402581 1.591711 1.249481
#> 5 AAACATCGAATGTTGC 27 2   non-Corpus callosum 1.450364 1.572980 1.272308
#> 6 AAACATCGACACAGAA 82 2   non-Corpus callosum 1.542466 1.032041 1.039183
```

```
head(coprofile_example_data$RNA) %>%
  dplyr::select(1:7)
#>           Barcodes  x y Pathology.Annotations Gabbr2 Pde7b Rims1
#> 1 AAACATCGAACTCACC 31 2   non-Corpus callosum      0     0     0
#> 2 AAACATCGAAGAGATC 30 2   non-Corpus callosum      0     0     0
#> 3 AAACATCGAAGGACAC 29 2   non-Corpus callosum      0     0     3
#> 4 AAACATCGAATCCGTC 28 2   non-Corpus callosum      0     0     1
#> 5 AAACATCGAATGTTGC 27 2   non-Corpus callosum      1     0     7
#> 6 AAACATCGACACAGAA 82 2   non-Corpus callosum      0     0     0
```

4. Differential Expression Analysis

4.1. Continious data type

We use `run_wilcoxon()` to contentious data types, such as normalized ATAC.

```
atac_DE_res <- run_wilcoxon(
  atac,
  compare_levels = c("non-Corpus callosum", "Corpus callosum"))
head(atac_DE_res)
#>       gene       p_value
#> 1   Gabbr2  7.429736e-32
#> 2    Pde7b  1.385373e-89
#> 3    Rims1 8.215698e-167
#> 4   Snap25 8.054409e-202
#> 5    Il12a  1.722868e-01
#> 6 Cacna2d3  9.952009e-43
```

4.2. Count data type

We use `run_gee_gst()` to count data types, such as RNA counts.

```
rna_DE_res <- run_gee_gst(
  rna,
  compare_levels = c("non-Corpus callosum", "Corpus callosum"))
```

```
head(rna_DE_res)
#>       gene      p_value
#> 1   Gabbr2 0.0005489216
#> 2    Pde7b 0.0125448984
#> 3    Rims1 0.5848623358
#> 4   Snap25 0.0006652406
#> 5    Il12a 0.0004134127
#> 6 Cacna2d3 0.0020262216
```

4.3. Combine ATAC & RNA P-values

We combine the results from ATAC and RNA.

```
merged_pvalue <- merge(atac_DE_res, rna_DE_res, by = "gene") %>%
  rename_with(~ c("atac_pvalue", "rna_pvalue"), starts_with("p_value")) %>%
  tibble::column_to_rownames("gene") %>%
  na.omit()
head(merged_pvalue)
#>          atac_pvalue   rna_pvalue
#> Acot11  1.381807e-78 4.918471e-01
#> Ak4     5.703768e-04 2.031796e-02
#> Ak5    3.708405e-119 1.695025e-03
#> Alas1   3.372929e-14 8.999031e-01
#> Ank     3.119800e-24 5.372984e-01
#> Apod    2.115761e-23 1.645404e-09
```

5. Run IMIX for ATAC-RNA Integration

We run IMIX for integration and retrieve significant genes with FDR control.

```
integration_res <- IMIX::IMIX(data_input = merged_pvalue)
#> Assign initial values
#> number of iterations= 39
#> number of iterations= 93
#> Start IMIX-ind procedure!
#> Successfully Done!
#> Start IMIX-cor-twostep procedure!
#> Successfully Done!
#> Start IMIX-cor model procedure!
#> Successfully Done!
#> Start IMIX-cor-restrict procedure!
#> Successfully Done!
#> Warning: IMIX_cor_restrict did not converge.
#> Assign IMIX_ind results to IMIX_cor_restrict.
#> The AIC/BIC values for IMIX_cor_restrict are not reliable and should not be used!
#> Start Model Selection
#> Start Label Sorting
#> Start Adaptive FDR Control
#> Finished!
head(integration_res$significant_genes_with_FDRcontrol)
```

```
#>        localFDR class_withoutFDRcontrol class_FDRcontrol
#> Ak5           0                       4                4
#> Arfip1        0                       4                4
#> Atp2b2        0                       4                4
#> Cdc5l         0                       4                4
#> Cpne5         0                       4                4
#> Dgkb          0                       4                4
```

References

Wang, Y., & Wei, P. (2025). A Mixture Model Approach for Integrating Spatial Transcriptomics and Epigenomics. Unpublished manuscript.

Wang, Z., & Wei, P. (2020). IMIX: A multivariate mixture model approach to association analysis through multi-omics data integration. Bioinformatics, 36(22–23), 5439–5447.

Zhang, D., Deng, Y., Kukanja, P., Agirre, E., Bartosovic, M., Dong, M., Ma, C., Ma, S., Su, G., Bao, S., Liu, Y., Xiao, Y., Rosoklija, G. B., Dwork, A. J., Mann, J. J., Leong, K. W., Boldrini, M., ... Fan, R. (2023). Spatial epigenome–transcriptome co-profiling of mammalian tissues. Nature, 616(7955), 113–122.