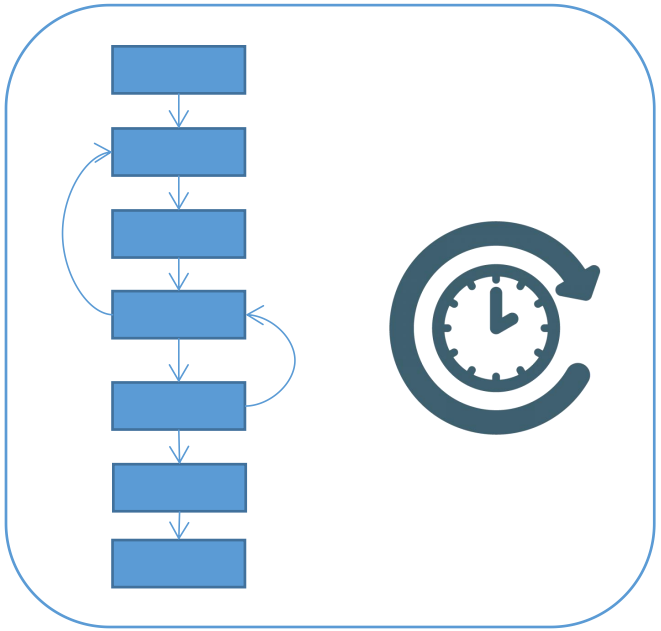
Seurat scRNA pipeline

Mingbo Cheng

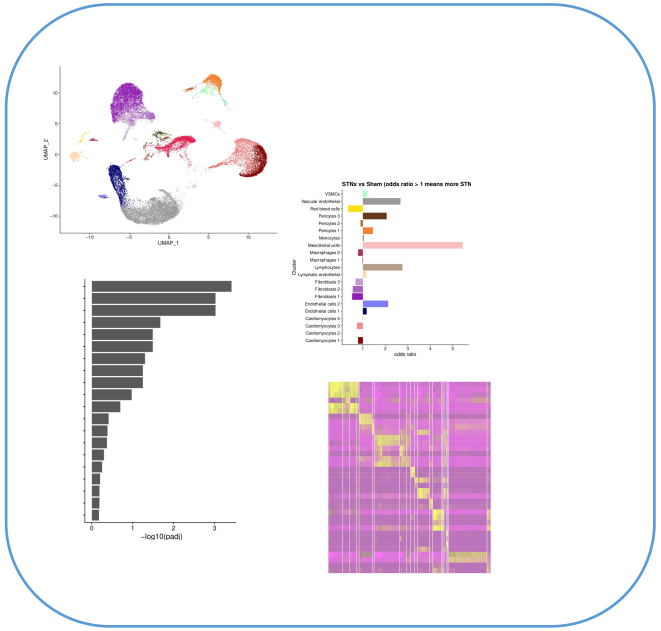
13/12/2020

Idea

Data Producing

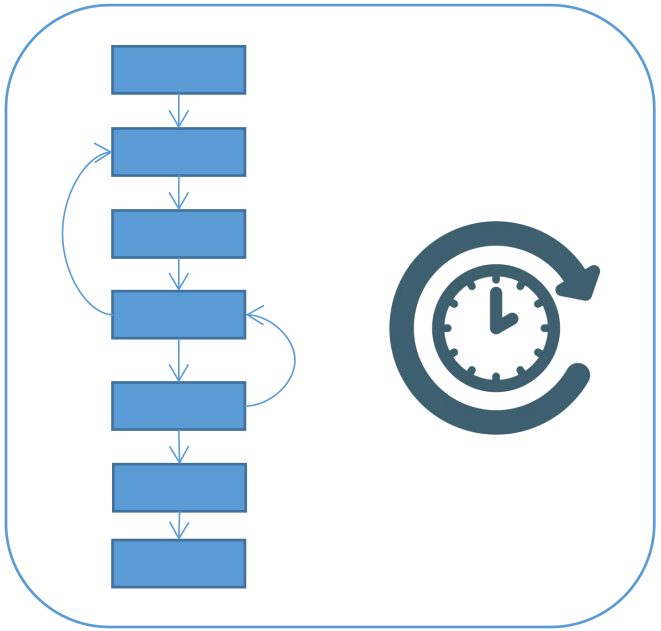


Visualization



How it works

Data Producing



seurat object

meta.data

clustering reusults
annotation
.....

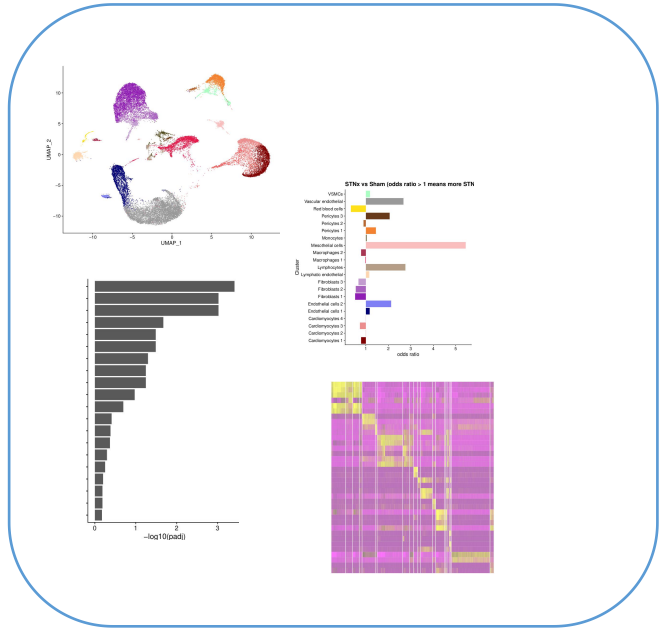
reductions

PCA
UMAP
.....

tools

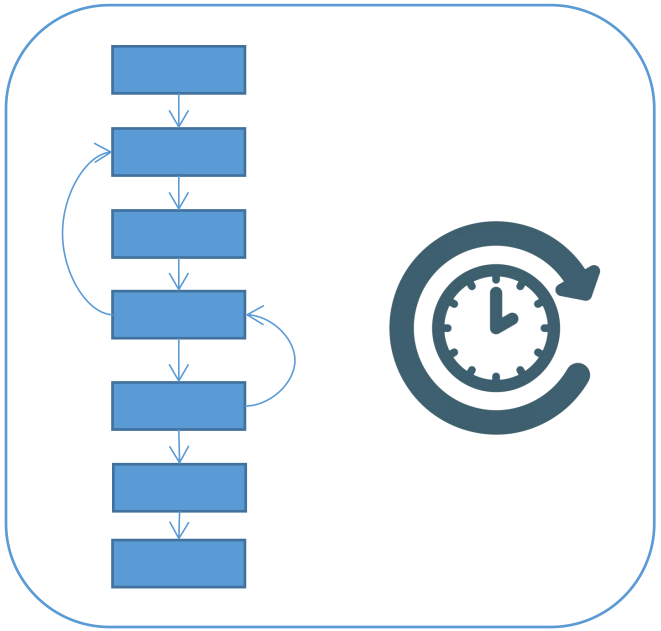
DE analysis
GO analysis
Pathways analysis
.....

Visualization

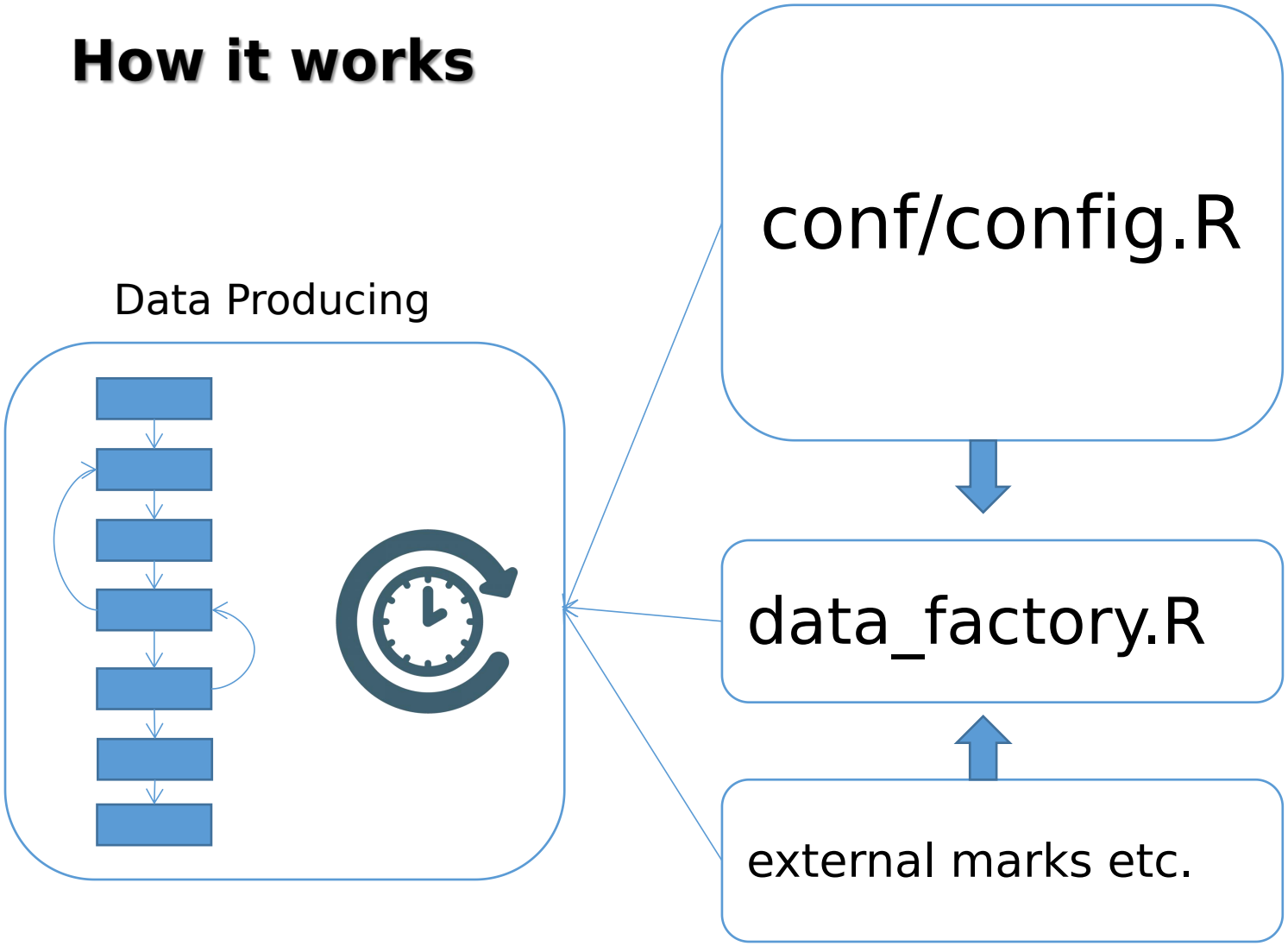


How it works

Data Producing

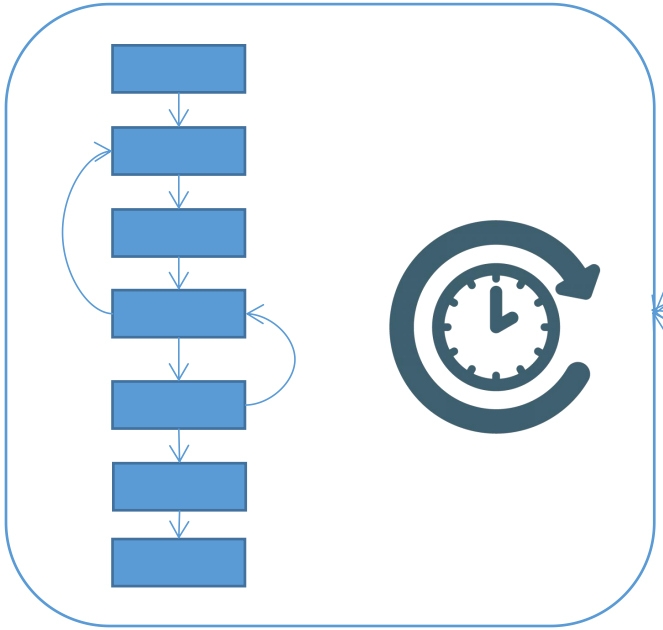


How it works


conf/config.R
Data Producing
data_factory.R
external marks etc.

How it works

Data Producing



conf/config.R

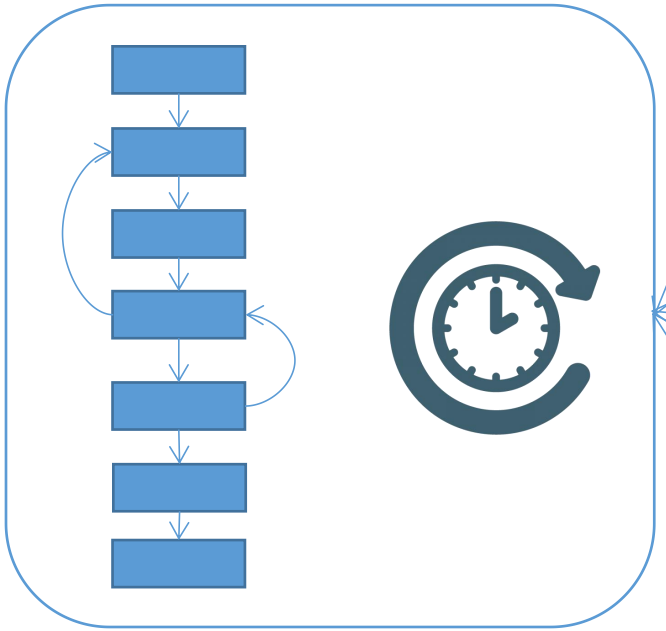
data_factory.R

external marks etc.

All processing methods

How it works

Data Producing



conf/config.R

data_factory.R

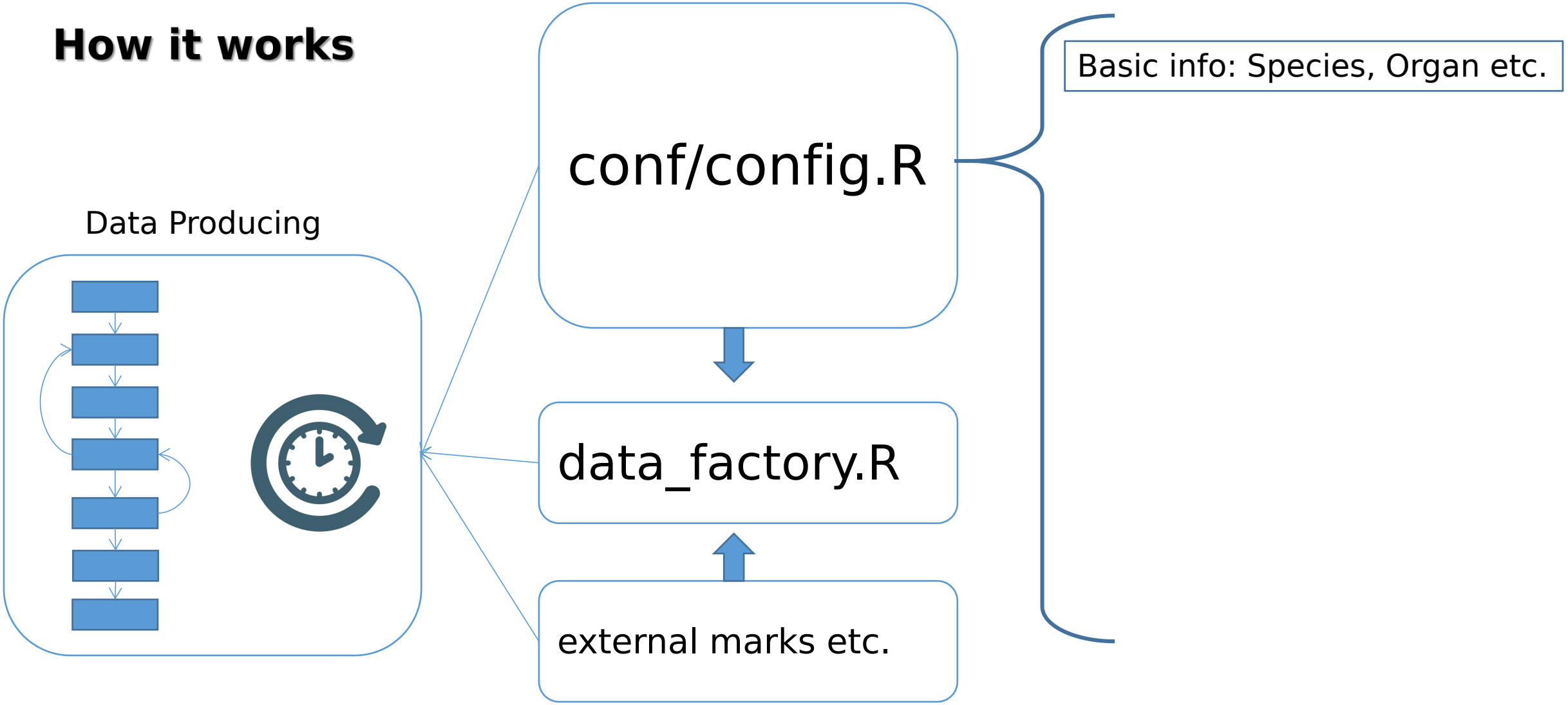
external marks etc.

All processing methods

Parameters

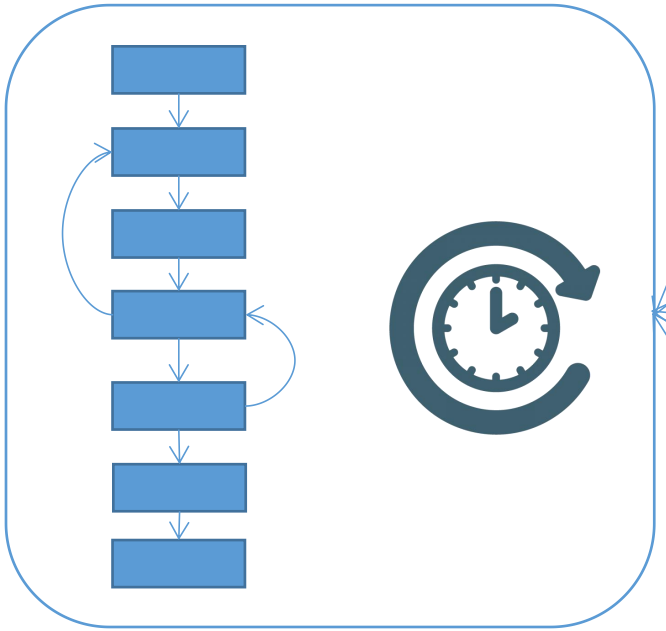
- #cpu cores
- save directory
- config file name
- count matrix format
- default cluster name
- mito/ribo filtering threshold
- etc....

How it works


conf/config.R
Basic info: Species, Organ etc.
Data Producing
data_factory.R
external marks etc.

How it works

Data Producing



conf/config.R

data_factory.R

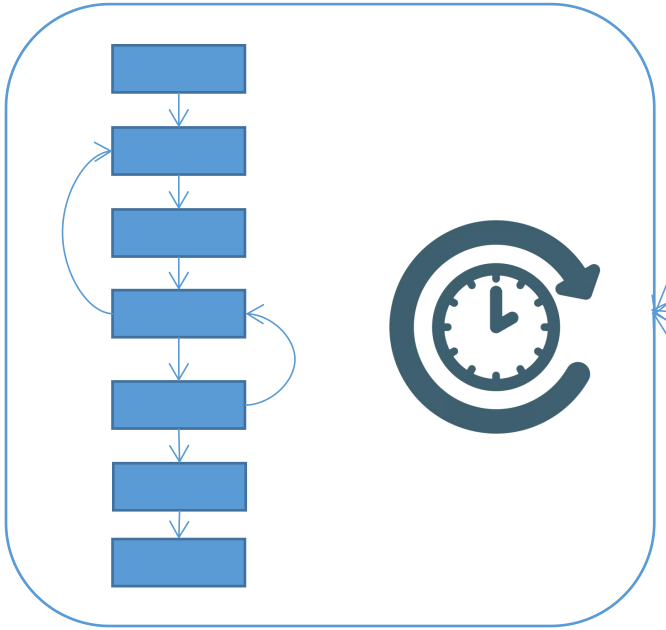
external marks etc.

Basic info: Species, Organ etc.

```
### --------------Initail info----------------------------
PROJECT = "Mouse heart Gli&CD45 project" ## set project name
ORGAN = 'Heart'           #For external annotation. Options: Blood,
SPECIES = "Mouse"         #For external annotation. Options: Human,
MCA_NAME = "Neonatal-Heart" #For MCA annotation.      Options: check

# filtering params when create seurat object
MINCELLS  = 5
MINGENES  = 50

### -------------- Data SRC-----------------------------
ANNOTATION_EXTERNAL_FILE = "external/Human_and_mouse_cell_markers-Ma

data_src = c(

     NK1_Gli1_IRI     =   "data/summed_mtx/sum_NK1_Gli1_IRI/",
     NK2_CD45_IRI     =   "data/summed_mtx/sum_NK2_CD45_IRI/",
     NK3_Gli1_Sham    =   "data/summed_mtx/sum_NK3_Gli1_Sham/",
     NK4_CD45_Sham    =   "data/summed_mtx/sum_NK4_CD45_Sham/"
)


##------------------ SET REPLICATE GROUP -------------
stage_lst = c(
    NK1_Gli1_IRI      = "IRI",
    NK2_CD45_IRI      = "IRI",
    NK3_Gli1_Sham     = "Sham",
    NK4_CD45_Sham     = "Sham"
```

How it works

Data Producing



conf/config.R

Basic info: Species, Organ etc.

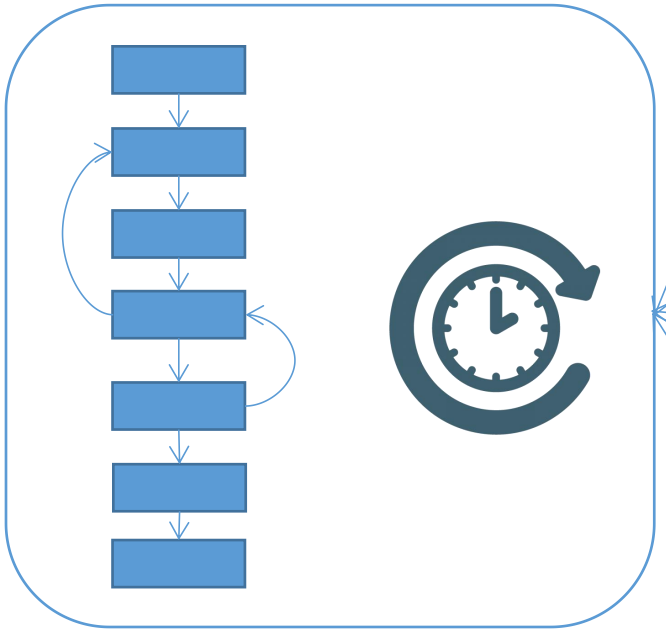
Input location: 10x/h5

data_factory.R

external marks etc.

How it works

Data Producing



conf/config.R

data_factory.R

external marks etc.

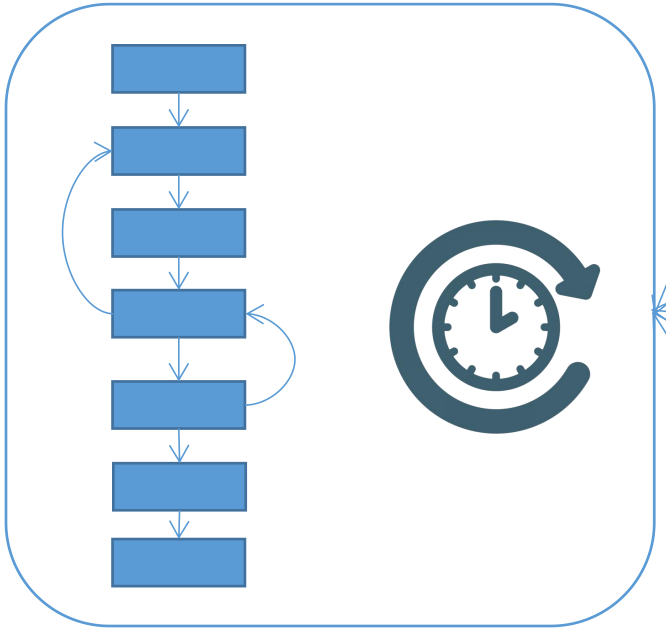
Basic info: Species, Organ etc.

Input location: 10x/h5

```
### --------------Initail Info----------------------------
PROJECT = "Mouse heart Gli&CD45 project" ## set project name
ORGAN = 'Heart'           #For external annotation. Options: Blood, Heart
SPECIES = "Mouse"         #For external annotation. Options: Human, Mouse
MCA_NAME = "Neonatal-Heart" #For MCA annotation.      Options: check http

# filtering params when create seurat object
MINCELLS  = 5
MINGENES  = 50

### -------------- Data SRC-----------------------------
ANNOTATION_EXTERNAL_FILE = "external/Human_and_mouse_cell_markers-Markers

data_src = c(

     NK1_Gli1_IRI     =   "data/summed_mtx/sum_NK1_Gli1_IRI/",
     NK2_CD45_IRI     =   "data/summed_mtx/sum_NK2_CD45_IRI/",
     NK3_Gli1_Sham    =   "data/summed_mtx/sum_NK3_Gli1_Sham/",
     NK4_CD45_Sham    =   "data/summed_mtx/sum_NK4_CD45_Sham/"
)


##------------------ SET REPLICATE GROUP -------------
stage_lst = c(
    NK1_Gli1_IRI      = "IRI",
    NK2_CD45_IRI      = "IRI",
    NK3_Gli1_Sham     = "Sham",
    NK4_CD45_Sham     = "Sham"
)
```

How it works

Data Producing



conf/config.R

Basic info: Species, Organ etc.

Input location: 10x/h5

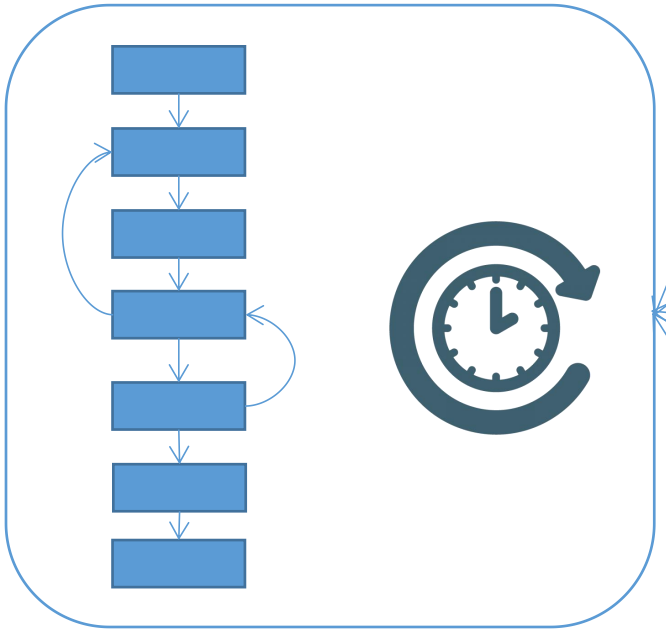
Execution plan: Which steps?

data_factory.R

external marks etc.

How it works

Data Producing



conf/config.R

data_factory.R

external marks etc.

Basic info: Species, Organ etc.

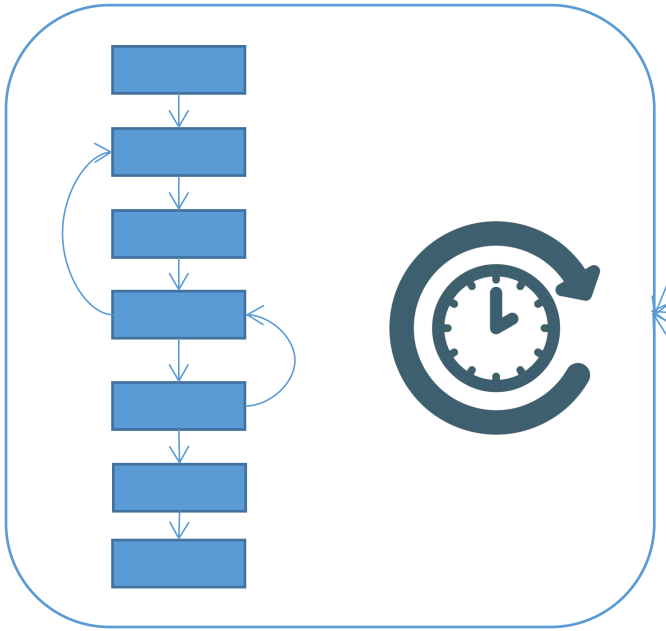
Input location: 10x/h5

Execution plan: Which steps?

- proprecessing
- Integration
- clustering
- DE, GO, Pathways
- etc....

How it works

Data Producing



conf/config.R

data_factory.R

external marks etc.

Basic info: Species, Organ etc.

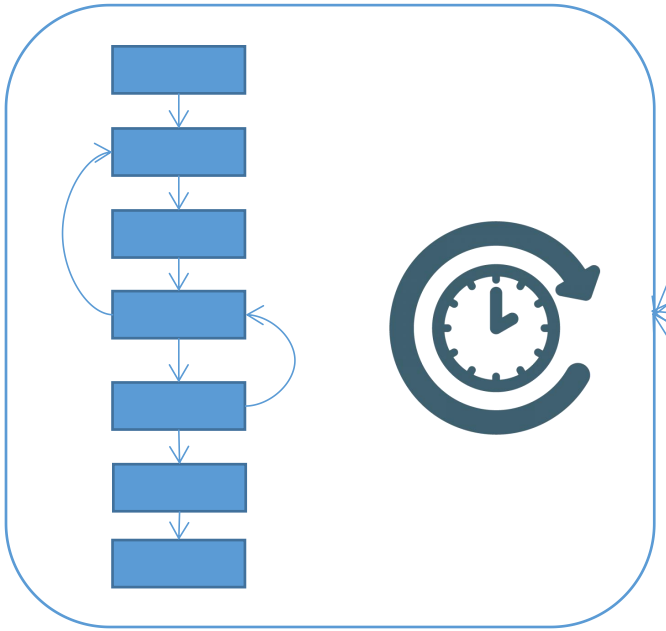
Input location: 10x/h5

Execution plan: Which steps?

```
## 0. omit,     1. calc & save,      2.
conf = c(
       scrna_rawdata                = 1,
       scrna_filter                 = 1,
       scrna_preprocess             = 1,
       scrna_cellcycle              = 1,
       scrna_cycleRegressOut        = 1,
       scrna_CCmitoRegressOut       = 1,
       scrna_mitoRegressOut         = 0,
       scrna_riboRegressOut         = 0,
       scrna_CCriboRegressOut       = 0,
       scrna_mitoRiboRegressOut     = 0,
       scrna_CCmitoRiboRegressOut = 0,
       scrna_integration            = 1,
       scrna_ScaleIntegration       = 1,
       scrna_batchclustering        = 1,
       scrna_batch_markergenes      = 1,
       scrna_clustering             = 1,
       scrna_clusterwise_xcell      = 1,
       scrna_cluster_annotation     = 1,
       scrna_del_mitogenes          = 0,
       scrna_markergenes            = 1,
```

How it works

Data Producing



conf/config.R

data_factory.R

external marks etc.

Basic info: Species, Organ etc.

Input location: 10x/h5

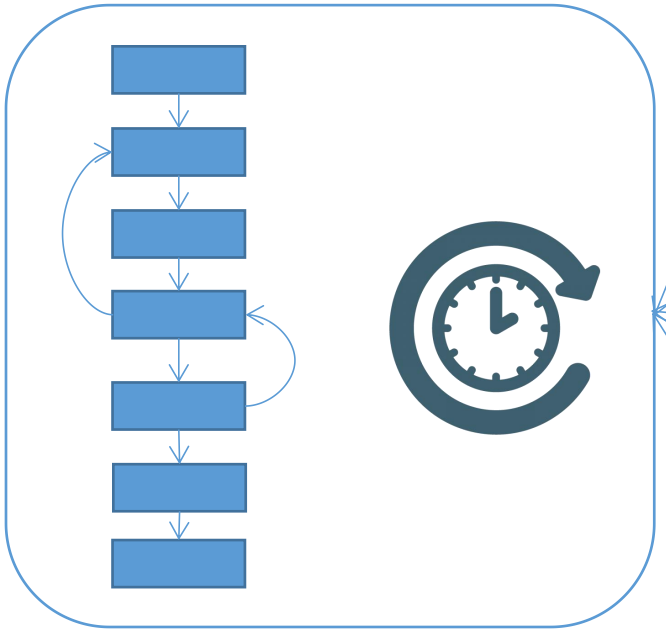
Execution plan: Which steps?

- proprecessing
- Integration
- clustering
- DE, GO, Pathways
- etc....

settings for specific functions

How it works

Data Producing



conf/config.R

data_factory.R

external marks etc.

Basic info: Species, Organ etc.

Input location: 10x/h5

Execution plan: Which steps?

settings for specific functions

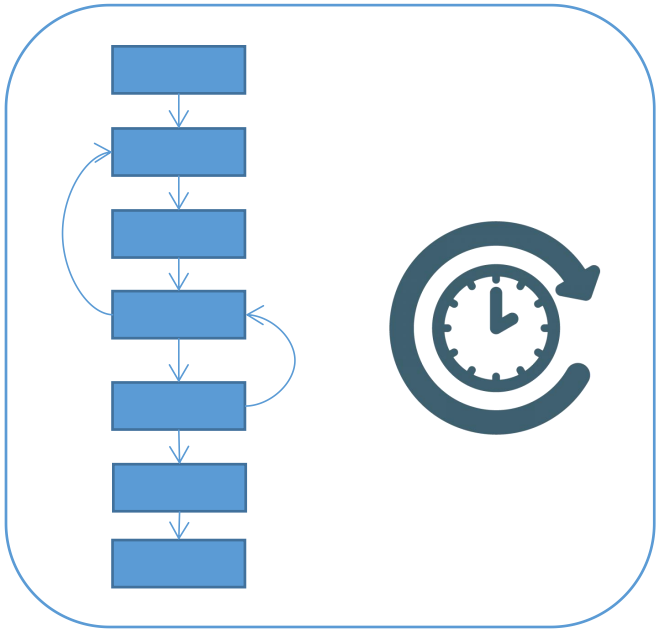
```
## name[your operation name], value[dataframe which clus
scrna_clusterwise_filtercell_settings <- list(
  "mito_cluster0,3,4,5"     =  data.frame(type="mito", r
  "ribo_cluster2,7_filter"  =  data.frame(type="ribo", r
  "mito_cluster6_filter"    =  data.frame(type="mito", r
  "mito_cluster8_filter"    =  data.frame(type="mito", r
)

## name[new_cluster_name], value[which clusters to merge
scrna_merge_clusters = list(
        "1+7" = c(1, 7),
        "2+6" = c(2, 6),
        "10+11+16" = c(10, 11, 16)
)


scrna_remove_clusters = c(1, 3, 6)
scrna_remove_recluster = c(1, 3, 6)


### cluster annotation
from_cluster_slot = "removed_clusters"
cluster_annotation <- c(
```

!!!Know every step

```
## 0. omit,     1. calc & save,      2. load   3.
conf = c(
      scrna_rawdata              = 1, ## read count matrix and merge samples to a Seurat obj
      scrna_filter               = 1, ## filter nFeatureRNA and nCountRNA
      scrna_preprocess           = 1, ## Normalize & FindvariablegFeatures and ScaleData
      scrna_cellcycle            = 1, ## Cell cycle scoring
      scrna_cycleRegressOut      = 1, ## Regress out cell cycle effects
      scrna_CCmitoRegressOut     = 1, ## Regress out mito & cell cycle
      scrna_mitoRegressOut       = 0, ## Regress out mito only
      scrna_riboRegressOut       = 0, ## Regress out ribo only
      scrna_CCriboRegressOut     = 0, ## Regress out ribo & cell cycle
      scrna_mitoRiboRegressOut   = 0, ## Regress out ribo&mito
      scrna_CCmitoRiboRegressOut = 0, ## Regress out ribo&mito  & cell cycle
      scrna_integration          = 1, ## Integrate samples using Seurat 3
      scrna_ScaleIntegration     = 1, ## ScaleDatai&PCA and UMAP
      scrna_batchclustering      = 1, ## clustering with resolution from 0.1 to 0.8
      scrna_batch_markergenes    = 1, ## Marker Genes for clusters with different resolutions
      scrna_clustering           = 1, ## Set seurat_clusters or re-calculate
      scrna_clusterwise_xcell    = 1, ## remove cells of each cluster according distinct criterion
      scrna_cluster_annotation   = 1, ## Annotate clusters according to `cluster_annotation`
      scrna_del_mitogenes        = 0, ## !!!DANGEROUS, once deleted, never recovered!!!
      scrna_markergenes          = 1, ## markergenes for seurat_clusters
      scrna_genesorteR           = 1, ## genesorteR analysis
      scrna_go                   = 1, ## Gene Ontology analysis
      scrna_kegg                 = 1, ## kegg enrichment analysis
      scrna_reactome             = 1, ## reactome enrichment analysis
      scrna_hallmark             = 1, ## hallmark enrichment analysis
      scrna_fishertest_clusters  = 1, ## fisher test for clusters and stages
      scrna_MCAannotate          = 1, ## scMCA annotation celltypes
      scrna_ExternalAnnotation   = 1, ## Annotation from given databases(tsv)
      scrna_dego_name            = 1, ## DE & GO between samples
      scrna_dego_stage           = 1, ## DE & GO between stages
      scrna_dego_stage_vsRest    = 1, ## DE & GO between one stage and all Rest
      scrna_pathway_name         = 1, ## samples comparison KEGG&Reactome&hallmark
      scrna_pathway_stage        = 1, ## stages comparison KEGG&Reactome&hallmark
      scrna_pathway_stage_vsRest = 1, ## stages vsRest comparison KEGG&Reactome&hallmark
      scrna_clusterwise_xcell    = 0, ## keep cells for each cluster according to mito&ribo
      scrna_fishertest_clusters  = 0, ## fisher test for clusters and stages
      scrna_merge_clusters       = 0, ## merge clusters
      scrna_remove_clusters      = 0, ## remove clusters
      scrna_remove_recluster     = 0, ## remove clusters and recluster with defualt resolution
      scrna_markergenes          = 0, ## markergenes for seurat_clusters
      scrna_go                   = 0, ## Gene Ontology analysis
      scrna_dego_name            = 0, ## DE & GO between samples
      scrna_dego_stage           = 0) ## GO down for mark genes
```

How it works

Data Producing



seurat object

meta.data

clustering reusults
annotation
.....

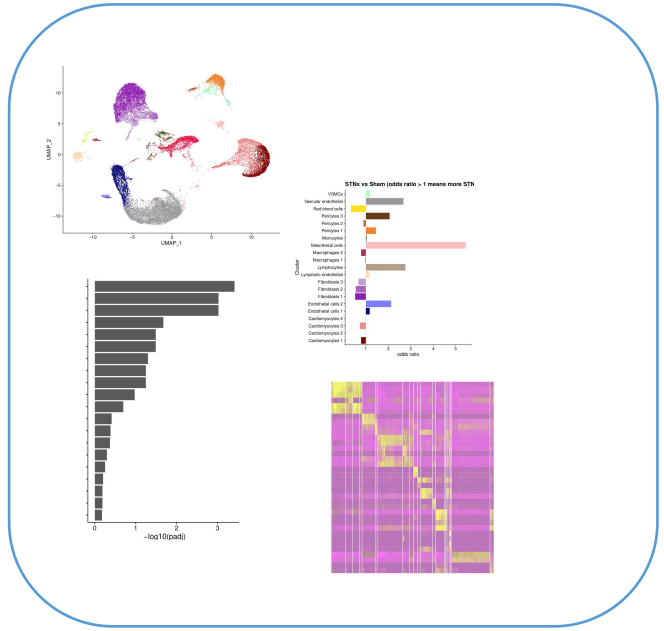
reductions

PCA
UMAP
.....

tools

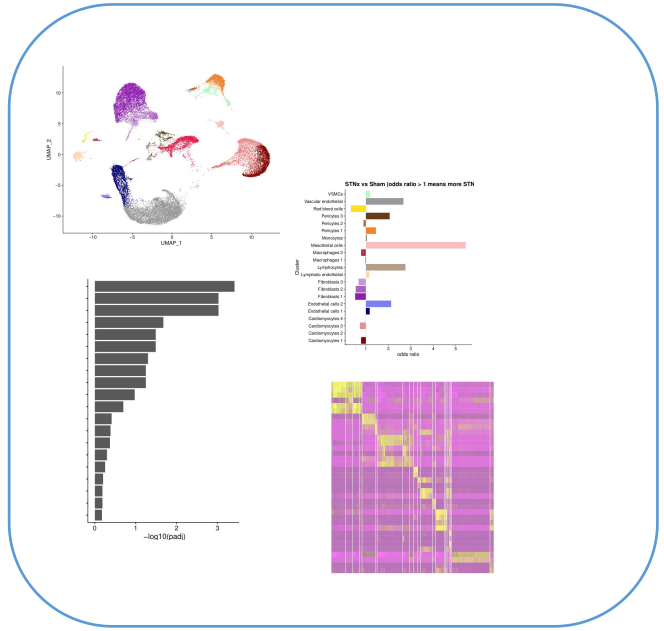
DE analysis
GO analysis
Pathways analysis
.....

Visualization



How it works

Visualization



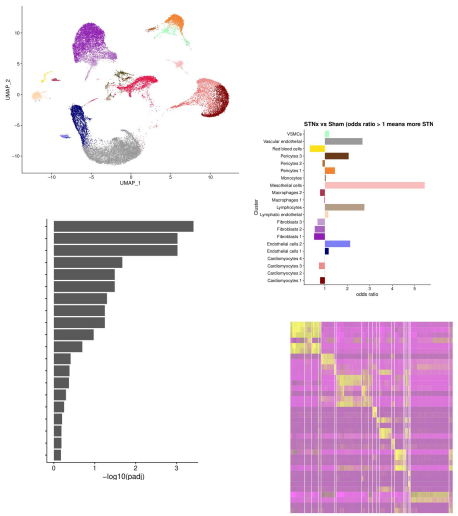
How it works

- conf/config.R
- excels generated
- Seurat Robjects

run.sh

- 1_quality_report.Rmd
- 2_clustering.Rmd
- 2_clusters_DEs.Rmd
- 3_DE_GO-analysis.Rmd
- 3_external_markers.Rmd
- make_report.R
- singleton_clustering.Rmd
- template/
  - DE-GO-1v1.template
  - DE-GO-stagesVS.template
  - index.template

Visualization



How it works

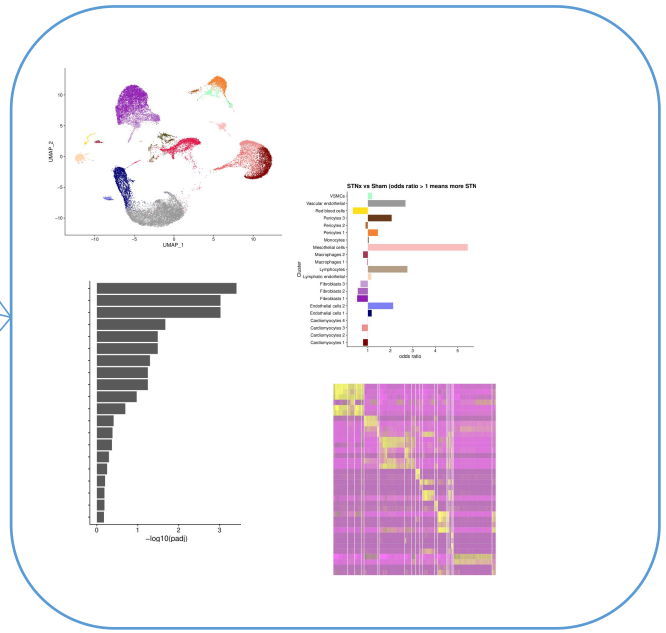
```
#!/bin/bash
RED='\033[0;31m'
NC='\033[0m' # No Color

FUNCS=(
        QC
        DEs
        Clusters
#        DEGO
#        EXT_MARKERS
        DEGO_1v1
        DEGO_stage
)



#!!!!!!!!------clusters to choose------------
# In general, we choose seurat_clusters,
# If you are using removed or merged clusters,
# choose the following:
            # seurat_clusters
            # merged_clusters
            # removed_clusters
            # remove_recluster

#cluster="removed_clusters"
#cluster="remove_recluster"
#cluster="merged_clusters"
#cluster="annotation"
#cluster="singleton"

cluster="seurat_clusters"

#!!!!!!!!!----------------------------------


echo -e "Use cluster slot ${RED} $cluster ${NC}"
mkdir -p report/data
```

- conf/config.R
- excels generated
- Seurat Robjects

run.sh

- 1_quality_report.Rmd
- 2_clustering.Rmd
- 2_clusters_DEs.Rmd
- 3_DE_GO-analysis.Rmd
- 3_external_markers.Rmd
- make_report.R
- singleton_clustering.Rmd
- template/
  - DE-GO-1v1.template
  - DE-GO-stagesVS.template
  - index.template

Visualization



How it works

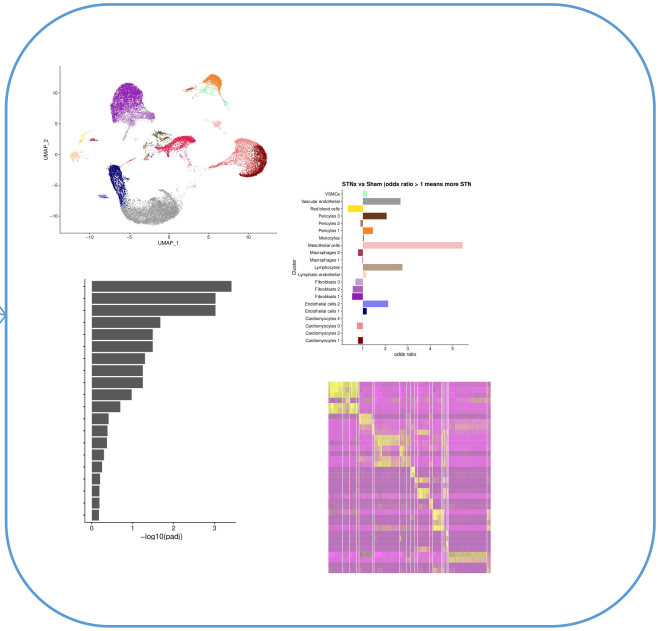
- conf/config.R
- excels generated
- Seurat Robjects

report

run.sh

- 1_quality_report.Rmd
- 2_clustering.Rmd
- 2_clusters_DEs.Rmd
- 3_DE_GO-analysis.Rmd
- 3_external_markers.Rmd
- make_report.R
- singleton_clustering.Rmd
- template/
  - DE-GO-1v1.template
  - DE-GO-stagesVS.template
  - index.template

Visualization

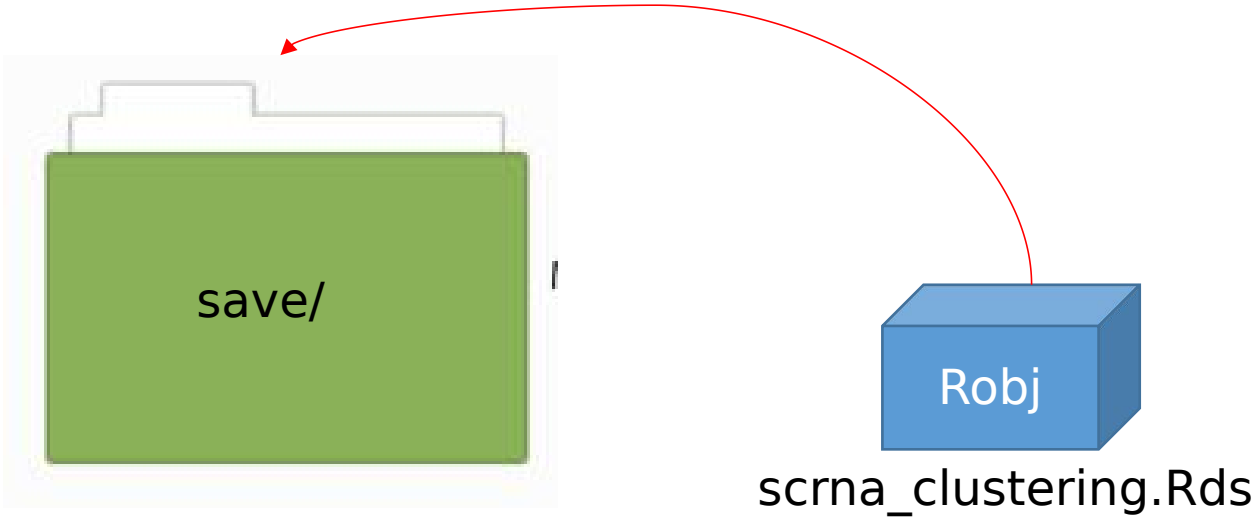


Most Frequent questions

- Is it possible to do further analysis given an existed Seurat Object?

Most Frequent questions

- Is it possible to do further analysis given an existed Seurat Object?




```
## 0. omit,     1. calc & save,      2.
conf = c(
       scrna_rawdata                = 1,
       scrna_filter                 = 1,
       scrna_preprocess             = 1,
       scrna_cellcycle              = 1,
       scrna_cycleRegressOut        = 1,
       scrna_CCmitoRegressOut       = 1,
       scrna_mitoRegressOut         = 0,
       scrna_riboRegressOut         = 0,
       scrna_CCriboRegressOut       = 0,
       scrna_mitoRiboRegressOut     = 0,
       scrna_CCmitoRiboRegressOut   = 0,
       scrna_integration            = 1,
       scrna_ScaleIntegration       = 1,
       scrna_batchclustering        = 1,
       scrna_batch_markergenes      = 1,
       scrna_clustering             = 1,
       scrna_clusterwise_xcell      = 1,
       scrna_cluster_annotation     = 1,
       scrna_del_mitogenes          = 0,
       scrna_markergenes            = 1,
```

Most Frequent questions

- Is it possible to do further analysis given an Seurat Object?

```
## 0. omit,     1. calc & save,      2.
conf = c(
       scrna_rawdata              = 0,
       scrna_filter               = 0,
       scrna_preprocess           = 0,
       scrna_cellcycle            = 0,
       scrna_cycleRegressOut      = 0,
       scrna_CCmitoRegressOut     = 0,
       scrna_mitoRegressOut       = 0,
       scrna_riboRegressOut       = 0,
       scrna_CCriboRegressOut     = 0,
       scrna_mitoRiboRegressOut   = 0,
       scrna_CCmitoRiboRegressOut = 0,
       scrna_integration          = 0,
       scrna_ScaleIntegration     = 0,
       scrna_batchclustering      = 0,
       scrna_batch_markergenes    = 0,
       scrna_clustering           =2,
       scrna_clusterwise_xcell    = 1,
       scrna_cluster_annotation   = 1,
       scrna_del_mitogenes        = 0,
       scrna_markergenes          = 1,
```

Most Frequent questions

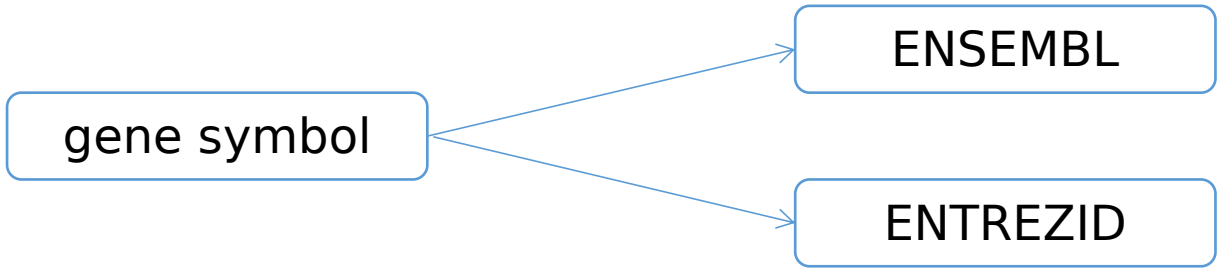
- Is it possible to do further analysis given an Seurat Object?

!!! umap reduction name

```
## 0. omit,     1. calc & save,      2.
conf = c(
       scrna_rawdata              = 0,
       scrna_filter               = 0,
       scrna_preprocess           = 0,
       scrna_cellcycle            = 0,
       scrna_cycleRegressOut      = 0,
       scrna_CCmitoRegressOut     = 0,
       scrna_mitoRegressOut       = 0,
       scrna_riboRegressOut       = 0,
       scrna_CCriboRegressOut     = 0,
       scrna_mitoRiboRegressOut   = 0,
       scrna_CCmitoRiboRegressOut = 0,
       scrna_integration          = 0,
       scrna_ScaleIntegration     = 0,
       scrna_batchclustering      = 0,
       scrna_batch_markergenes    = 0,
       scrna_clustering           =2,
       scrna_clusterwise_xcell    = 1,
       scrna_cluster_annotation   = 1,
       scrna_del_mitogenes        = 0,
       scrna_markergenes          = 1,
```

Most Frequent questions

- Is it possible to do further analysis given an Seurat Object?
- Why there are so many warnings especially for GO & pathway analysis?



Most Frequent questions

- Is it possible to do further analysis given an Seurat Object?
- Why there are so many warnings especially for GO & pathway analysis?
- Failed to run visualization part: cannot find ‘INTE_UMAP’

Most Frequent questions

- Is it possible to do further analysis given an Seurat Object?
- Why there are so many warnings especially for GO & pathway analysis?
- Failed to run visualization part: cannot find ‘INTE_UMAP’
- Why there are so many Robjects generated?

Most Frequent questions

- Is it possible to do further analysis given an Seurat Ob
- Why there are so many warnings especially for GO & p
- Failed to run visualization part: cannot find 'INTE_UMA
- Why there are so many Robjects generated?

```
180M Oct 22 09:21 scrna_RegressOutAll.Rds
121M Nov 10 16:20 scrna_go.Rds
121M Nov 10 16:20 scrna_fishertest_clusters.Rds
121M Nov 10 16:21 scrna_MCAannotate.Rds
121M Nov 10 16:22 scrna_ExternalAnnotation.Rds
198M Nov 10 16:36 scrna_dego_name.Rds
218M Nov 10 16:40 scrna_kegg.Rds
220M Nov 10 16:42 scrna_reactome.Rds
220M Nov 10 16:43 scrna_hallmark.Rds
220M Nov 10 16:47 scrna_dego_stage.Rds
221M Nov 10 16:49 scrna_pathway_stage.Rds
224M Nov 10 16:53 scrna_pathway_name.Rds
2.6M Nov 10 16:54 all_de_list.Rds
262M Nov 10 16:59 scrna_dego_stage_vsRest.Rds
263M Nov 10 17:02 scrna_pathway_stage_vsRest.Rds
740M Nov 11 17:49 scrna_rawdata.Rds
 13M Nov 11 17:50 scrna_filter.Rds
166M Nov 11 17:50 scrna_preprocess.Rds
 20M Nov 11 17:50 scrna_cellcycle.Rds
175M Nov 11 17:50 scrna_cycleRegressOut.Rds
176M Nov 11 17:51 scrna_CCmitoRegressOut.Rds
177M Nov 11 17:51 scrna_mitoRegressOut.Rds
178M Nov 11 17:51 scrna_riboRegressOut.Rds
180M Nov 11 17:52 scrna_CCriboRegressOut.Rds
181M Nov 11 17:52 scrna_mitoRiboRegressOut.Rds
183M Nov 11 17:52 scrna_CCmitoRiboRegressOut.Rds
 28M Nov 11 17:53 scrna_integration.Rds
 45M Nov 11 17:53 scrna_ScaleIntegration.Rds
 45M Nov 11 17:53 scrna_batchclustering.Rds
 47M Nov 11 18:05 scrna_batch_markergenes.Rds
 47M Nov 11 18:05 scrna_clustering.Rds
 38M Nov 11 18:05 scrna_clusterwise_xcell.Rds
 38M Nov 11 18:06 scrna_markergenes.Rds
 46M Nov 30 20:37 scrna_del_mitogenes.Rds
```

TODO

- Merge code from Tiago
- Add harmony integration
- NABA geneset score
- Add KEGG/Reactome/hallmark visualization
- scHCL for human cell annotation
- Integrated with ligand receptor analysis?

Dummy example

- our hpc : 134.130.18.27
  - module add scRNA/1.0.3
  - /data/scRNA/scrna_seurat_pipeline_demo

Thanks

Q&A