NSCLC_Salcher2022_process

December 25, 2025

1 load data

[ ]:
```
obj.NSCLC <- readRDS('/project/sex_cancer/data/NSCLC_Salcher2022/
↪NSCLC_Salcher2022_coreAtlas.rds')
obj.NSCLC <- UpdateSeuratObject(obj.NSCLC)
```

2 modify meta.data

[ ]:
```
obj.NSCLC@meta.data <- obj.NSCLC@meta.data %>%
                       dplyr::rename(c("Sex" = "sex")) %>%
                       transform(barcode = rownames(.))
```

3 filter sample

[ ]:
```
obj.NSCLC <- obj.NSCLC %>%
             subset(Sex %in% c('female', 'male')) %>%
             subset(tissue == 'lung') %>%
             subset(origin != 'nan') %>%
             subset(platform == '10x') %>%
             subset(disease != 'chronic obstructive pulmonary disease')
length(unique(obj.NSCLC$sample))
table(obj.NSCLC$study)
```

[ ]:
```
## de-factor
obj.NSCLC@meta.data <-obj.NSCLC@meta.data %>%
                      mutate_if(~ !is.numeric(.), ~ ext_list(.)) %>%
                      dplyr::rename(c('SampleID' = 'sample', 'Chemistry' =␣
↪'assay', 'DonorID' = 'donor_id')) %>%
                      transform(Sex = ifelse(Sex == 'female', 'F', 'M')) %>%
                      mutate(SampleType = case_when(origin == 'normal' ~␣
↪'normal',
                                                    origin == 'normal_adjacent'␣
↪~ 'normal_adjacent',
                                                    origin == 'tumor_primary' ~␣
↪'tumor',
```

```
                                                    TRUE ~ 'Others')) %>%
                        transform(Cohort = 'NSCLC_Salcher2022')
```

4 trans ENSG to symbol

[ ]:
```
table(rownames(obj.NSCLC@assays$RNA@counts) == rownames(obj.
 ↪NSCLC@assays$RNA@data))
trans <- obj.NSCLC@assays$RNA@meta.features %>%
         dplyr::select(c('feature_type', 'feature_name')) %>%
         .[rownames(obj.NSCLC@assays$RNA@counts),] %>%
         rownames_to_column('feature')
trans %>% head(n = 2)

## rename counts
rownames(obj.NSCLC@assays$RNA@counts) <- trans$feature_name
## rename data
rownames(obj.NSCLC@assays$RNA@data) <- trans$feature_name
## rename meta.feature
obj.NSCLC@assays$RNA@meta.features <- trans %>%␣
 ↪column_to_rownames('feature_name')
```

5 cell type annotation

5.1 assign oCT

[ ]:
```
obj.NSCLC@meta.data  <- obj.NSCLC@meta.data %>%
                        mutate(oCT = cell_type_major)
```

5.2 check annotation

check cell type annotation provided in the original research via COSG

[ ]:
```
## check marker expression
marker_annotation <- readRDS("marker_annotation.rds")

obj <- obj.NSCLC
DefaultAssay(obj) <- "RNA"
obj <- obj %>% NormalizeData(normalization.method = "LogNormalize", scale.
 ↪factor = 10000, verbose = F)
Idents(obj) <- ext_list(obj$oCT)

marker_oCT <- obj %>%
              cosg(groups = "all", assay = "RNA", slot = "data",
              mu = 10, ## The penalty factor to penalize gene expression in␣
 ↪cells not belonging to the cluster of interest
```

```
                n_genes_user = 50, # Number of top ranked genes returned in the␣
 ↪result
                remove_lowly_expressed=T, # If TRUE, genes that express a␣
 ↪percentage of target cells smaller than a specific value (expressed_pct) are␣
 ↪not considered as marker genes for the target cells. The default value is␣
 ↪TRUE.
                expressed_pct=0.1) # If TRUE, genes that express a percentage of␣
 ↪target cells smaller than a specific value (expressed_pct) are not␣
 ↪considered as marker genes for the target cells.
marker_oCT <- cbind(marker_oCT[[1]] %>% melt(id.vars = NULL) %>% dplyr::
 ↪rename(c("oCT" = "variable", "marker" = "value")),
                    marker_oCT[[2]] %>% melt(id.vars = NULL) %>% dplyr::
 ↪select(-"variable") %>% dplyr::rename(c("COSGscore" = "value"))) %>%
                mutate(Cohort = unique(obj$Cohort)) %>% mutate(oCT =␣
 ↪ext_list(oCT))

oCT_marker <- marker_oCT
oCT_list <- unique(oCT_marker$oCT)
lapply(oCT_list, function(x){
        check <- oCT_marker %>% subset(oCT == x & marker %in%␣
 ↪marker_annotation[[x]])
        ifelse(nrow(check) == 0, print(x), return(check))
})
```

5.3 assign mCT

[ ]:
```
obj.NSCLC@meta.data  <- obj.NSCLC@meta.data %>%
                        mutate(mCT = case_when(cell_type_major %in% c('Tumor␣
 ↪cells') ~ 'Tumor',
                                               cell_type_major %in% c('Alveolar␣
 ↪cell type 1', 'Ciliated', 'Club', 'transitional club/AT2', 'Alveolar cell␣
 ↪type 2') ~ 'Epi',
                                               cell_type_major %in% c('T cell␣
 ↪CD8') ~ 'CD8T',
                                               cell_type_major %in% c('T cell␣
 ↪CD4') ~ 'CD4T',
                                               cell_type_major %in% c('T cell␣
 ↪regulatory') ~ 'Treg',
                                               cell_type_major %in% c('NK␣
 ↪cell') ~ 'NK',
                                               cell_type_major %in% c('B cell',␣
 ↪'Plasma cell') ~ 'B',
                                               cell_type_major %in%␣
 ↪c('Monocyte') ~ 'Mono',
```

```
                                                cell_type_major %in%␣
↪c('Macrophage alveolar', 'Macrophage') ~ 'Mph',
                                                cell_type_major %in% c('cDC2',␣
↪'cDC1', 'pDC', 'DC mature') ~ 'DC',
                                                cell_type_major %in%␣
↪c('Neutrophils') ~ 'Neu',
                                                cell_type_major %in% c('Mast␣
↪cell') ~ 'Mast',
                                                cell_type_major %in%␣
↪c('Endothelial cell') ~ 'Endo',
                                                cell_type_tumor %in%␣
↪c('Fibroblast adventitial', 'Fibroblast alveolar', 'Fibroblast␣
↪peribronchial') ~ 'Fibro',
                                                cell_type_tumor %in%␣
↪c('Pericyte') ~ 'Pericyte',
                                                cell_type_tumor %in%␣
↪c('Mesothelial') ~ 'Mesothelial',
                                                cell_type_tumor %in% c('Smooth␣
↪muscle cell') ~ 'SMC',
                                                cell_type_major %in% c('other')␣
↪~ 'Others',
                                            ))
head(obj.NSCLC@meta.data, n  = 2)
```

5.4 assign gCT

[ ]:
```
obj.NSCLC@meta.data  <- obj.NSCLC@meta.data %>%
                        mutate(gCT = case_when(mCT %in% c('Tumor', 'Epi') ~␣
↪'Tumor',
                                               mCT %in% c('Neu', 'Mast',␣
↪'CD8T', 'CD4T', 'NK', 'DC', 'B', 'Treg', 'Mono', 'Mph') ~ 'Immune',
                                               mCT %in% c('Pericyte', 'SMC',␣
↪'Mesothelial', 'Fibro', 'Endo') ~ 'Stromal',
                                               TRUE ~ 'Others',
                                            ))
head(obj.NSCLC@meta.data, n  = 2)
```

5.5 discard unannotated cells

[ ]:
```
obj.NSCLC <- obj.NSCLC %>% subset(gCT != "Others")
obj.NSCLC
```

[ ]:
```
table(obj.NSCLC$oCT, obj.NSCLC$gCT, useNA = "ifany")
```

6 save

[ ]:
```
saveRDS(obj.NSCLC, 'obj.NSCLC.use.rds')
```