ccRCC_Hu2024_process

December 25, 2025

1 load data

[ ]:
```
obj.ccRCC <- readRDS('/project/sex_cancer/data/ccRCC_Hu2024/snRNA.rds')
```

2 modify meta.data

[ ]:
```
info <- read_xlsx('/project/sex_cancer/data/ccRCC_Hu2024/
 ↪ccRCC_Hu2024PatientInfo.xlsx', skip = 1)

names(info)[1] <- 'sample'
info1 <- info[info$sample%in%unique(obj.ccRCC@meta.data$sample),]
info1$Sex=gsub('male','M',info1$Sex)
info1$Sex=gsub('female','F',info1$Sex)

meta <- obj.ccRCC@meta.data %>% transform(barcode = rownames(.))
meta <- merge(meta, info1, by = 'sample',all = TRUE) %>%␣
 ↪column_to_rownames('barcode') %>% .[colnames(obj.ccRCC),]
```

[ ]:
```
# modify meta.data
obj.ccRCC@meta.data <- meta@meta.data %>%
                       transform(barcode = rownames(.), Cohort =␣
 ↪'ccRCC_Hu2024', orig.ident = sample) %>%
                       dplyr::rename(c('SampleID' = 'sample', 'percent.mt' =␣
 ↪'percent.mito')) %>%
                       dplyr::select(-c('nUMI', 'nGene', 'class'))
```

[ ]:
```
## add sample info
sample_info <- openxlsx::read.xlsx('/project/sex_cancer/data/ccRCC_Hu2024/
 ↪41588_2024_1662_MOESM4_ESM.xlsx', startRow = 2, rowNames = TRUE)
sample_info <-  sample_info %>%
                transform(SampleID = rownames(.)) %>%
                subset(SampleID %in% obj.ccRCC$SampleID) %>%
                dplyr::rename(c('Sex' = 'sex'))

obj.ccRCC@meta.data <- obj.ccRCC@meta.data %>%
                       merge(sample_info, by = 'SampleID', all = TRUE) %>%
```

```
                        transform(SampleType = 'tumor', DonorID = SampleID) %>%
                        `rownames<-`(.$barcode)
```

[ ]:
```
options(repr.plot.height = 5, repr.plot.width = 25)
DimPlot_scCustom(obj.ccRCC, pt.size = 1, group.by = "celltype", reduction =␣
 ↪'umap', label = T, label.size = 4, colors_use = pal_igv("default")(51))|
DimPlot_scCustom(obj.ccRCC, pt.size = 1, group.by = "celltype2", reduction =␣
 ↪'umap', label = F, colors_use = pal_igv("default")(51))
```

3 cell type annotation

3.1 assign oCT

[ ]:
```
obj.ccRCC@meta.data <- obj.ccRCC@meta.data %>%
                        transform(oCT = celltype2)
```

3.2 check annotation

check cell type annotation provided in the original research via COSG

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

obj <- obj.ccRCC
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))
})
```

3.3 assign mCT

[ ]:
```
obj.ccRCC@meta.data  <-  obj.ccRCC@meta.data %>%
                         mutate(mCT = case_when(celltype %in% c('Malignant') ~␣
↪'Epi',
                                                celltype2 %in% c('1_B',␣
↪'2_Plasma') ~ 'B',
                                                celltype2 %in% c('1_CD4-Tcm/
↪Th', '1_CD4-Th') ~ 'CD4T',
                                                celltype2 %in% c('1_CD4-Treg')␣
↪~ 'Treg',
                                                celltype2 %in%␣
↪c('2_CD8-cycling', '2_CD8-ITGAE', '2_CD8-Tex') ~ 'CD8T',
                                                celltype2 %in% c('3_MAIT') ~␣
↪'MAIT',
                                                celltype2 %in% c('4_NK-KLRC1',␣
↪'4_Temra/aNK') ~ 'NK',
                                                celltype2 %in% c('4_Macro01',␣
↪'4_Macro02-IGF1', '4_Macro03-LILRB5', '4_Macro04-GPNMB',
                                                                 ␣
↪'4_Macro05-NR4A3', '4_Macro06-C3', '4_Macro07-GBP1', '5_Macro-prolif') ~␣
↪'Mph',
                                                celltype2 %in% c('6_DC-cDC1',␣
↪'6_DC-cDC2', '6_DC-pDC', '6_DC-mDC') ~ 'DC',
                                                celltype2 %in%␣
↪c('3_Mono01-SMIM25', '3_Mono02-VCAN') ~ 'Mono',
                                                celltype2 %in% c('7_Mast') ~␣
↪'Mast',
                                                celltype2 %in% c('peri') ~␣
↪'Pericyte',
                                                celltype2 %in% c('EC_Cap01-F8',␣
↪'EC_artery', 'EC_Cap04-NRP2', 'EC_low UMI', 'EC-cycling', 'EC_vein',
```

```
␣
↪'EC_Cap03-IGFBP5', 'EC_Cap02', 'EC_Cap07-DNASE1L3', 'EC_Cap05-KIT',␣
↪'EC_Cap06-WT1', 'EC_Cap-CD36') ~ 'Endo',
                                                celltype2 %in% c('Fib') ~␣
↪'Fibro',
                                                celltype2 %in% c('SMC') ~ 'SMC',
                                                TRUE ~ 'Others'
                                                ))
```

3.4 assign gCT

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

[ ]:
```
options(repr.plot.height = 5, repr.plot.width = 30)
DimPlot_scCustom(obj.ccRCC, pt.size = 1, group.by = "gCT", reduction = 'umap',␣
 ↪label = T, label.size = 4, colors_use = pal_igv("default")(51))|
DimPlot_scCustom(obj.ccRCC, pt.size = 1, group.by = "mCT", reduction = 'umap',␣
 ↪label = T, label.size = 4, colors_use = pal_igv("default")(51))|
DimPlot_scCustom(obj.ccRCC, pt.size = 1, group.by = "oCT", reduction = 'umap',␣
 ↪label = F, colors_use = pal_igv("default")(51))
```

[ ]:
```
meta_now[,c('SampleType', 'Sex', 'SampleID')] %>% .[!duplicated(.$SampleID),]␣
 ↪%$% table(.$SampleType, .$Sex)
table(meta_now$SampleType)
```

4 save

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