BLCA_Juric2025_process

December 25, 2025

1 load data

[ ]:
```
obj.BLCA <- readRDS("/project/sex_cancer/data/BLCA_Juric2025/
↪GSE269877_dta_cancer.submission.rds") ## sex information not provided in the␣
↪normal sample object, making it not included in SexTumorDB
obj.BLCA
obj.BLCA@meta.data %>% head(n = 2)
```

[ ]:
```
obj.BLCA@meta.data %>%
dplyr::select(c("orig.ident", "Tumor.Category", "Groups", "Sex", "Age")) %>%
.[!duplicated(.$orig.ident),] %>% subset(Groups != "Recurrent") %$%
table(.$Sex)
```

[ ]:
```
obj.BLCA <- RenameAssays(obj.BLCA, RNACleaned = "RNA")
obj.BLCA@assays$RNA@data <- obj.BLCA@assays$RNA@counts
```

2 filter samples

[ ]:
```
obj.BLCA <- obj.BLCA %>%
            subset(Groups %in% c("Naive_wo", "Naive_w")) ## drop samples␣
↪collected after BCG treatment (treated samples)
length(unique(obj.BLCA$orig.ident))
```

3 modify meta.data

[ ]:
```
rownames(obj.BLCA) %>% .[grepl("^MT-", .)] ## MT genes are removed in the␣
↪original dataset, seems to be set <30
```

[ ]:
```
obj.BLCA@meta.data <- obj.BLCA@meta.data %>%
                      transform(barcode = rownames(.)) %>%
                      dplyr::select(-c("nCount_RNACleaned",␣
↪"nFeature_RNACleaned")) %>%
                      dplyr::rename(c("DonorID" = "Patient")) %>%
                      transform(SampleID = orig.ident) %>%
```

```
                        transform(Cohort = "BLCA_Juric2025", Chemistry = "10x 3'␣
 ↪v3", Tissue = "Bladder", SampleType = "tumor") %>%
                        dplyr::select(-c("Exome.seq", "cell.names.CD6.status",␣
 ↪"percent.rps", "percent.rpl", "Tumor.Category", "Stage", "Grade")) %>%
                        .[colnames(obj.BLCA),]
obj.BLCA@meta.data %>% head(n = 2)
```

[ ]:
```
options(repr.plot.height = 4, repr.plot.width = 20)
DimPlot_scCustom(obj.BLCA, pt.size = 1, reduction = "umap", group.by = "Sex",␣
 ↪label = F, label.size = 8, colors_use = pal_igv("default")(51), raster = F)|
DimPlot_scCustom(obj.BLCA, pt.size = 1, reduction = "umap", group.by =␣
 ↪"low_res_compartment", label = F, label.size = 8, colors_use =␣
 ↪pal_igv("default")(51), raster = F)|
DimPlot_scCustom(obj.BLCA, pt.size = 1, reduction = "umap", group.by = "cell.
 ↪names", label = F, label.size = 8, colors_use = pal_igv("default")(51),␣
 ↪raster = F)
```

4 cell type annotation

4.1 assign oCT

[ ]:
```
obj.BLCA@meta.data <- obj.BLCA@meta.data %>%
                        dplyr::rename(c("oCT" = "cell.names"))
```

4.2 check annotation

check cell type annotation provided in the original research via COSG

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

obj <- obj.BLCA
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))
})
```

4.3 assign mCT

[ ]:
```
obj.BLCA@meta.data <- obj.BLCA@meta.data %>%
                      mutate(mCT = case_when(oCT %in% c("Proliferating␣
 ↪uroepithelial", "Intermediate A", "Intermediate B", "Intermediate C",␣
 ↪"Intermediate D", "Intermediate E", "Umbrella", "Basal") ~ "Epi",
                                             oCT %in% c("Endothelial arterial",␣
 ↪"Enodothelial venous", "Activated endothelial") ~ "Endo",
                                             oCT %in% c("WNThi periurothelial␣
 ↪fibroblast", "Myofibroblast","Periurothelial fibroblast", "Interstitial␣
 ↪fibroblast") ~ "Fibro",
                                             oCT %in% c("B") ~ "B",
                                             oCT %in% c("Plasma") ~ "Plasma",
                                             oCT %in% c("Treg") ~ "Treg",
                                             oCT %in% c("CD4 T central memory␣
 ↪1", "CD4 T central memory 2", "CD4 T effector memory", "CD4 T exhausted") ~␣
 ↪"CD4T",
                                             oCT %in% c("CD8 T 2", "CD8 T␣
 ↪effector", "CD8 T resident memory") ~ "CD8T",
                                             oCT %in% c("Proliferating␣
 ↪lymphocyte") ~ "T_proliferation",
                                             oCT %in% c("NK CD56 bright", "NK␣
 ↪CD56 dim") ~ "NK",
                                             oCT %in% c("MMDSC") ~ "Mono",
```

```
                                                oCT %in% c("CCL17+ dendritic",␣
 ↪"Immature dendritic", "Type 2 conventional dendritic", "pDC") ~ "DC",
                                                oCT %in% c("Macrophage") ~ "Mph",
                                                oCT %in% c("Mast") ~ "Mast",
                                                TRUE ~ 'Others'))

obj.BLCA@meta.data %>% subset(mCT == "Others") %$% table(.$oCT, useNA= "ifany")␣
 ↪## all removed.during.subclustering
```

[ ]:
```
## filter out removed cells
obj.BLCA <- obj.BLCA %>% subset(mCT != "Others") ## discard 9766 cells␣
 ↪annotated as "removed.during.subclustering" by the original research
obj.BLCA
```

4.4 assign gCT

[ ]:
```
unique(obj.BLCA$mCT)
```

[ ]:
```
obj.BLCA@meta.data <- obj.BLCA@meta.data %>%
                      mutate(gCT = case_when(mCT %in% c("Epi") ~ "Tumor",
                                                mCT %in% c("DC", "CD8T", "Treg",␣
 ↪"Mph", "T_proliferation", "CD4T", "NK", "Plasma", "Mono", "B", "Mast") ~␣
 ↪"Immune",
                                                mCT %in% c("Fibro", "Endo") ~␣
 ↪"Stromal",
                                                TRUE ~ 'Others'))
table(obj.BLCA$gCT)
```

5 save

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