

## Azizi 2018 metadata - indrop

```
globus_metadata <- readRDS("/mnt/nmoraais-nfs/marta/pC_myeinfobank/metadata/globus/processed/GSE114725_a
globus_metadata

## Loading required package: SeuratObject
## Loading required package: sp
## An object of class Seurat
## 14854 features across 25793 samples within 1 assay
## Active assay: RNA (14854 features, 0 variable features)
colnames(globus_metadata)[1:5]

## [1] "cell_246" "cell_260" "cell_346" "cell_188" "cell_33"
colnames(globus_metadata@meta.data)

## [1] "patient" "batch" "condition" "sex" "cell_type" "tissue"
table(globus_metadata$patient)

##
## BC1_1 BC1_2 BC1_3 BC1_4 BC2_1 BC2_2 BC2_3 BC2_4 BC3_1 BC3_2 BC3_3 BC3_4 BC3_5
## 1756 582 1205 1927 946 881 1112 223 68 201 618 49 103
## BC4_1 BC4_2 BC4_3 BC4_4 BC4_5 BC4_6 BC5_1 BC5_2 BC5_3 BC5_4 BC6_1 BC6_2 BC6_3
## 786 1349 811 1349 1122 940 355 352 509 563 802 922 1774
## BC7_2 BC7_3 BC8_1 BC8_2 BC8_3
## 891 1000 1181 764 652
table(globus_metadata$batch)

##
## GSE114725
## 25793
table(globus_metadata$condition)

##
## healthy_tumor_adj tumor_unknown
## 4540 21253
table(globus_metadata$sex)

##
## female
## 25793
table(globus_metadata$tissue)

##
## breast
## 25793
```

```

metadata <- data.frame(Cell_ID = colnames(globus_metadata),
                        Sample = globus_metadata$patient)
rownames(metadata) <- NULL

```

```

metadata$Sample_Type <- "Fresh"
metadata$Protocol <- "inDrop_v2"
metadata$Technology <- "inDrop"
metadata$Dissociation <- "Mechanical,Enzymatic"
metadata$CellOrNucleus <- "Cell"
metadata$Sequencing_Machine <- "Illumina_HiSeq_2500"
metadata$Preprocessing <- "SEQC"
metadata$Reference_Genome <- "GRCh38"
metadata$Gender <- "Female"
metadata$Race <- NA
metadata$Parity <- NA
metadata$Treatment_Status <- NA
metadata$Treatment_Type <- NA
metadata$Stage <- NA

```

```

metadata$Patient <- metadata$Sample
metadata$Patient <- gsub("_[1-8]$", "", metadata$Patient)
table(metadata$Patient, metadata$Sample)

```

```

##
##      BC1_1 BC1_2 BC1_3 BC1_4 BC2_1 BC2_2 BC2_3 BC2_4 BC3_1 BC3_2 BC3_3 BC3_4
## BC1  1756   582  1205  1927     0     0     0     0     0     0     0     0
## BC2    0     0     0     0   946   881  1112   223     0     0     0     0
## BC3    0     0     0     0     0     0     0     0    68   201   618   49
## BC4    0     0     0     0     0     0     0     0     0     0     0     0
## BC5    0     0     0     0     0     0     0     0     0     0     0     0
## BC6    0     0     0     0     0     0     0     0     0     0     0     0
## BC7    0     0     0     0     0     0     0     0     0     0     0     0
## BC8    0     0     0     0     0     0     0     0     0     0     0     0
##
##      BC3_5 BC4_1 BC4_2 BC4_3 BC4_4 BC4_5 BC4_6 BC5_1 BC5_2 BC5_3 BC5_4 BC6_1
## BC1     0     0     0     0     0     0     0     0     0     0     0     0
## BC2     0     0     0     0     0     0     0     0     0     0     0     0
## BC3   103     0     0     0     0     0     0     0     0     0     0     0
## BC4     0   786  1349   811  1349  1122   940     0     0     0     0     0
## BC5     0     0     0     0     0     0     0   355   352   509   563     0
## BC6     0     0     0     0     0     0     0     0     0     0     0   802
## BC7     0     0     0     0     0     0     0     0     0     0     0     0
## BC8     0     0     0     0     0     0     0     0     0     0     0     0
##
##      BC6_2 BC6_3 BC7_2 BC7_3 BC8_1 BC8_2 BC8_3
## BC1     0     0     0     0     0     0     0
## BC2     0     0     0     0     0     0     0
## BC3     0     0     0     0     0     0     0
## BC4     0     0     0     0     0     0     0
## BC5     0     0     0     0     0     0     0
## BC6   922  1774     0     0     0     0     0
## BC7     0     0   891  1000     0     0     0
## BC8     0     0     0     0  1181   764   652

```

```

metadata$Age <- metadata$Patient
metadata$Age[metadata$Age == "BC1"] <- 38
metadata$Age[metadata$Age == "BC2"] <- 60
metadata$Age[metadata$Age == "BC3"] <- 43
metadata$Age[metadata$Age == "BC4"] <- 52
metadata$Age[metadata$Age == "BC5"] <- 78
metadata$Age[metadata$Age == "BC6"] <- 58
metadata$Age[metadata$Age == "BC7"] <- 65
metadata$Age[metadata$Age == "BC8"] <- 72

```

```

metadata$Subtype <- metadata$Patient
metadata$Subtype[metadata$Subtype == "BC1"] <- "ER+/PR+"
metadata$Subtype[metadata$Subtype == "BC2"] <- "ER+"
metadata$Subtype[metadata$Subtype == "BC3"] <- "TNBC"
metadata$Subtype[metadata$Subtype == "BC4"] <- "ER+/PR+"
metadata$Subtype[metadata$Subtype == "BC5"] <- "TNBC"
metadata$Subtype[metadata$Subtype == "BC6"] <- "ER+"
metadata$Subtype[metadata$Subtype == "BC7"] <- "HER2+"
metadata$Subtype[metadata$Subtype == "BC8"] <- "TNBC"

```

```

metadata$Cancer_Type <- metadata$Patient
metadata$Cancer_Type[metadata$Cancer_Type == "BC1"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC2"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC3"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC4"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC5"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC6"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC7"] <- "IDC"
metadata$Cancer_Type[metadata$Cancer_Type == "BC8"] <- "IDC"

```

```

metadata$Grade <- metadata$Patient
metadata$Grade[metadata$Grade == "BC1"] <- 1
metadata$Grade[metadata$Grade == "BC2"] <- 2
metadata$Grade[metadata$Grade == "BC3"] <- 3
metadata$Grade[metadata$Grade == "BC4"] <- 1
metadata$Grade[metadata$Grade == "BC5"] <- 3
metadata$Grade[metadata$Grade == "BC6"] <- 2
metadata$Grade[metadata$Grade == "BC7"] <- 3
metadata$Grade[metadata$Grade == "BC8"] <- 2

```

```

metadata$Menopause <- metadata$Patient
metadata$Menopause[metadata$Menopause == "BC1"] <- "Pre"
metadata$Menopause[metadata$Menopause == "BC2"] <- "Post"
metadata$Menopause[metadata$Menopause == "BC3"] <- "Pre"
metadata$Menopause[metadata$Menopause == "BC4"] <- "Pre"
metadata$Menopause[metadata$Menopause == "BC5"] <- "Post"
metadata$Menopause[metadata$Menopause == "BC6"] <- "Post"
metadata$Menopause[metadata$Menopause == "BC7"] <- "Post"
metadata$Menopause[metadata$Menopause == "BC8"] <- "Post"

```

```

metadata$Tissue <- "Tumour"

```

```

metadata[1:5,]

```

```

##      Cell_ID Sample Sample_Type Protocol Technology      Dissociation

```

```

## 1 cell_246 BC5_1      Fresh inDrop_v2      inDrop Mechanical,Enzymatic
## 2 cell_260 BC5_1      Fresh inDrop_v2      inDrop Mechanical,Enzymatic
## 3 cell_346 BC5_1      Fresh inDrop_v2      inDrop Mechanical,Enzymatic
## 4 cell_188 BC5_1      Fresh inDrop_v2      inDrop Mechanical,Enzymatic
## 5 cell_33  BC5_1      Fresh inDrop_v2      inDrop Mechanical,Enzymatic
##      CellOrNucleus Sequencing_Machine Preprocessing Reference_Genome Gender Race
## 1      Cell Illumina_HiSeq_2500      SEQC      GRCh38 Female  NA
## 2      Cell Illumina_HiSeq_2500      SEQC      GRCh38 Female  NA
## 3      Cell Illumina_HiSeq_2500      SEQC      GRCh38 Female  NA
## 4      Cell Illumina_HiSeq_2500      SEQC      GRCh38 Female  NA
## 5      Cell Illumina_HiSeq_2500      SEQC      GRCh38 Female  NA
##      Parity Treatment_Status Treatment_Type Stage Patient Age Subtype Cancer_Type
## 1      NA      NA      NA      NA      BC5 78      TNBC      IDC
## 2      NA      NA      NA      NA      BC5 78      TNBC      IDC
## 3      NA      NA      NA      NA      BC5 78      TNBC      IDC
## 4      NA      NA      NA      NA      BC5 78      TNBC      IDC
## 5      NA      NA      NA      NA      BC5 78      TNBC      IDC
##      Grade Menopause Tissue
## 1      3      Post Tumour
## 2      3      Post Tumour
## 3      3      Post Tumour
## 4      3      Post Tumour
## 5      3      Post Tumour

```

```
#saveRDS(metadata, "azizi-2018-indrop-clinical-metadata.rds")
```