

Chung-2017 metadata

```
globus_metadata <- readRDS("/mnt/nmorais-nfs/marta/pC_myefinfobank/metadata/globus/processed/GSE75688_cl
globus_metadata

## Loading required package: SeuratObject
## Loading required package: sp
## An object of class Seurat
## 82104 features across 555 samples within 1 assay
## Active assay: RNA (82104 features, 0 variable features)
colnames(globus_metadata@meta.data)

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

##
##      GSE75688_BC01      GSE75688_BC02      GSE75688_BC03      GSE75688_BC03LN
##           28           55           38           56
##      GSE75688_BC04      GSE75688_BC05      GSE75688_BC06      GSE75688_BC07
##           58           77           25           52
##      GSE75688_BC07LN      GSE75688_BC08      GSE75688_BC09      GSE75688_BC09_Re
##           54           23           30           30
##      GSE75688_BC10      GSE75688_BC11
##           17           12
table(globus_metadata$batch)

##
## GSE75688
##      555
table(globus_metadata$condition)

##
## ER-HER2+      ER+      HER2+      TNBC
##      94      83      160      218
table(globus_metadata$sex)

##
## female
##      555
table(globus_metadata$tissue)

##
## breast
##      555
```

```

colnames(globus_metadata)[1:5]

## [1] "SRR2973272" "SRR2973273" "SRR2973274" "SRR2973275" "SRR2973276"

metadata <- data.frame(Cell_ID = colnames(globus_metadata),
                       Patient = globus_metadata$patient,
                       Subtype = globus_metadata$condition)

rownames(metadata) <- NULL

table(metadata$Subtype)

##
## ER-HER2+      ER+      HER2+      TNBC
##      94      83      160      218

metadata$Subtype <- gsub("ER-HER2\\+", "HER2+/ER+", metadata$Subtype)
table(metadata$Subtype)

##
##      ER+      HER2+ HER2+/ER+      TNBC
##      83      160      94      218

metadata$Sample <- gsub("GSE75688_", "", metadata$Patient)
metadata$Patient <- gsub("GSE75688_", "", metadata$Patient)
metadata$Patient <- gsub("BC09_Re", "BC09", metadata$Patient)
metadata$Patient <- gsub("BC03LN", "BC03", metadata$Patient)
metadata$Patient <- gsub("BC07LN", "BC07", metadata$Patient)

table(metadata$Patient, metadata$Sample)

##
##      BC01 BC02 BC03 BC03LN BC04 BC05 BC06 BC07 BC07LN BC08 BC09 BC09_Re BC10
## BC01      28      0      0      0      0      0      0      0      0      0      0      0      0
## BC02      0     55      0      0      0      0      0      0      0      0      0      0      0
## BC03      0      0     38     56      0      0      0      0      0      0      0      0      0
## BC04      0      0      0      0     58      0      0      0      0      0      0      0      0
## BC05      0      0      0      0      0     77      0      0      0      0      0      0      0
## BC06      0      0      0      0      0      0     25      0      0      0      0      0      0
## BC07      0      0      0      0      0      0      0     52     54      0      0      0      0
## BC08      0      0      0      0      0      0      0      0      0     23      0      0      0
## BC09      0      0      0      0      0      0      0      0      0      0     30     30      0
## BC10      0      0      0      0      0      0      0      0      0      0      0      0     17
## BC11      0      0      0      0      0      0      0      0      0      0      0      0      0
##
##      BC11
## BC01      0
## BC02      0
## BC03      0
## BC04      0
## BC05      0
## BC06      0
## BC07      0
## BC08      0
## BC09      0
## BC10      0
## BC11     12

```

```

metadata$Gender <- "Female"
metadata$Cancer_Type <- "IDC"

```

```

metadata$Tissue <- "Tumour"
metadata$Tissue[metadata$Sample == "BC07LN"] <- "Lymph-node"
metadata$Tissue[metadata$Sample == "BC03LN"] <- "Lymph-node"

```

```

table(metadata$Tissue, metadata$Sample)

```

```

##
##           BC01 BC02 BC03 BC03LN BC04 BC05 BC06 BC07 BC07LN BC08 BC09 BC09_Re
## Lymph-node    0    0    0     56    0    0    0    0     54    0    0        0
## Tumour        28   55   38      0   58   77   25   52      0   23   30       30
##
##           BC10 BC11
## Lymph-node    0    0
## Tumour        17   12

```

```

metadata$Treatment_Status <- "Naive"
metadata$Treatment_Status[metadata$Patient == "BC05"] <- "Treated"
metadata$Treatment_Type <- "None"
metadata$Treatment_Type[metadata$Patient == "BC05"] <- "Chemotherapy, Herceptin"

```

```

table(metadata$Patient, metadata$Treatment_Type)

```

```

##
##           Chemotherapy, Herceptin None
## BC01                0   28
## BC02                0   55
## BC03                0   94
## BC04                0   58
## BC05               77    0
## BC06                0   25
## BC07                0  106
## BC08                0   23
## BC09                0   60
## BC10                0   17
## BC11                0   12

```

```

table(metadata$Treatment_Status, metadata$Treatment_Type)

```

```

##
##           Chemotherapy, Herceptin None
## Naive                0  478
## Treated              77    0

```

```

metadata$Dissociation <- "Mechanical,Enzymatic"
metadata$Technology <- "Fluidigm"
metadata$Protocol <- "SMARTer Ultra Low RNA Kit"
metadata$Sample_Type <- "Fresh"
metadata$CellorNucleus <- "Cell"
metadata$Sequencing_Machine <- "Illumina_HiSeq_2500"
metadata$Preprocessing <- "Kallisto"
metadata$Reference_Genome <- "hg38"

```

```

table(metadata$Patient, metadata$Subtype)

```

```
##
##      ER+ HER2+ HER2+/ER+ TNBC
## BC01  28    0         0    0
## BC02  55    0         0    0
## BC03   0    0        94    0
## BC04   0   58         0    0
## BC05   0   77         0    0
## BC06   0   25         0    0
## BC07   0    0         0  106
## BC08   0    0         0   23
## BC09   0    0         0   60
## BC10   0    0         0   17
## BC11   0    0         0   12
```

```
metadata$Age <- metadata$Patient
metadata$Age[metadata$Patient == "BC01"] <- 66
metadata$Age[metadata$Patient == "BC02"] <- 72
metadata$Age[metadata$Patient == "BC03"] <- 72
metadata$Age[metadata$Patient == "BC04"] <- 67
metadata$Age[metadata$Patient == "BC05"] <- 46
metadata$Age[metadata$Patient == "BC06"] <- 67
metadata$Age[metadata$Patient == "BC07"] <- 71
metadata$Age[metadata$Patient == "BC08"] <- 67
metadata$Age[metadata$Patient == "BC09"] <- 53
metadata$Age[metadata$Patient == "BC10"] <- 82
metadata$Age[metadata$Patient == "BC11"] <- 47
```

```
metadata$Stage <- metadata$Patient
metadata$Stage[metadata$Patient == "BC01"] <- "IA"
metadata$Stage[metadata$Patient == "BC02"] <- "IIIA"
metadata$Stage[metadata$Patient == "BC03"] <- "IIB"
metadata$Stage[metadata$Patient == "BC04"] <- "IIA"
metadata$Stage[metadata$Patient == "BC05"] <- "IB"
metadata$Stage[metadata$Patient == "BC06"] <- "IIB"
metadata$Stage[metadata$Patient == "BC07"] <- "IIIC"
metadata$Stage[metadata$Patient == "BC08"] <- "IIA"
metadata$Stage[metadata$Patient == "BC09"] <- "IIA"
metadata$Stage[metadata$Patient == "BC10"] <- "IIIA"
metadata$Stage[metadata$Patient == "BC11"] <- "IIA"
```

```
metadata$Race <- NA
metadata$Menopause <- NA
metadata$Parity <- NA
```

```
metadata[1:5,]
```

```
##      Cell_ID Patient  Subtype Sample Gender Cancer_Type  Tissue
## 1 SRR2973272  BC01      ER+   BC01 Female      IDC    Tumour
## 2 SRR2973273  BC01      ER+   BC01 Female      IDC    Tumour
## 3 SRR2973274  BC04    HER2+   BC04 Female      IDC    Tumour
## 4 SRR2973275  BC03 HER2+/ER+ BC03 Female      IDC    Tumour
## 5 SRR2973276  BC03 HER2+/ER+ BC03LN Female      IDC    Lymph-node
## Treatment_Status Treatment_Type      Dissociation Technology
## 1      Naive      None Mechanical,Enzymatic  Fluidigm
## 2      Naive      None Mechanical,Enzymatic  Fluidigm
```

```

## 3      Naive      None Mechanical,Enzymatic  Fluidigm
## 4      Naive      None Mechanical,Enzymatic  Fluidigm
## 5      Naive      None Mechanical,Enzymatic  Fluidigm
##
##          Protocol Sample_Type CellorNucleus Sequencing_Machine
## 1 SMARTer Ultra Low RNA Kit      Fresh      Cell Illumina_HiSeq_2500
## 2 SMARTer Ultra Low RNA Kit      Fresh      Cell Illumina_HiSeq_2500
## 3 SMARTer Ultra Low RNA Kit      Fresh      Cell Illumina_HiSeq_2500
## 4 SMARTer Ultra Low RNA Kit      Fresh      Cell Illumina_HiSeq_2500
## 5 SMARTer Ultra Low RNA Kit      Fresh      Cell Illumina_HiSeq_2500
## Preprocessing Reference_Genome Age Stage Race Menopause Parity
## 1      Kallisto      hg38 66   IA   NA      NA      NA
## 2      Kallisto      hg38 66   IA   NA      NA      NA
## 3      Kallisto      hg38 67   IIA  NA      NA      NA
## 4      Kallisto      hg38 72   IIB  NA      NA      NA
## 5      Kallisto      hg38 72   IIB  NA      NA      NA

```

```

#saveRDS(metadata, "/mnt/nmoraais-nfs/marta/pC_myeinfobank/metadata/chung-2017/chung-2017-clinical-metad

```