Cancer programs - Barkley et al

```
library(Seurat)
library(plyr)
library(pheatmap)
library(RColorBrewer)


rhps <- read.delim('../DBarkley_RHPs.txt', check.names = F)
head(rhps)
```

```
##       Cycle     Stress Interferon Hypoxia  Oxphos Metal Mesenchymal    pEMT
## 1    ACYP1      ABHD3      ACTN1     ADM   ATOX1 FKBP5        AQP1    ACTB
## 2     ANLN AC016629.8      APOL1   AHNAK   ATP5H  HEY1         BGN   ACTG1
## 3   ANP32B    ADAMTS1      APOL2   ALDOA C14orf2 HTRA1       CALD1   ADIRF
## 4   ANP32E    ANKRD28      APOL3 ANGPTL4   COX14  MT1F      CCDC80  AKAP12
## 5   APITD1    ANKRD37      APOL6    BLNK  MYEOV2  MT1G       COL1A1  AKR1C1
## 6  APOBEC3B        ARC        B2M   BNIP3  NDUFA1  MT1H       COL1A2   ANXA1
##   Alveolar  Basal Squamous Glandular  Ciliated     AC    OPC     NPC
## 1     AGER ALDH3A1  AKR1B10      AGR2 AC007906.2    AGT   ASCL1   ARL4D
## 2    CAPN8   ASPN     AQP3      AQP5       AGR3  ALDOC MAP3K1 CAMK2N1
## 3     CAV1   ASS1  C10orf99     BCAT1      ARMC3   APOE    MMP2    CCNG2
## 4     CAV2  ATP1B3      CA2   C15orf48  C1orf194   AQP4   OLIG1   CELF4
## 5    CLIC5 COL17A1   CALML3       CD47  C20orf85  BBOX1   OLIG2   CELF5
## 6     CST6 COL18A1   CALML5    CEACAM5   C5orf49  CASQ1   SEZ6L   DAAM1
```

```
subdataset <- readRDS('../tmp/crc_smc.malignantcells.Rds')
head(subdataset@meta.data); nrow(subdataset@meta.data) # 17334
```

```
##                                  nCount_RNA nFeature_RNA        Library
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1      35998         4823 PM-PS-0001-T-A1
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1      31383         5252 PM-PS-0001-T-A1
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1       7302         1713 PM-PS-0001-T-A1
## AAACGGGAGGGAAACA-1-PM-PS-0001-T-A1       3759         1233 PM-PS-0001-T-A1
## AAACGGGGTATAGGTA-1-PM-PS-0001-T-A1      23097         3874 PM-PS-0001-T-A1
## AAAGATGAGGCCGAAT-1-PM-PS-0001-T-A1      14860         3282 PM-PS-0001-T-A1
##                                  Patient  Sample Cell_subtype RNA_snn_res.0.3
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               0
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               2
## AAACGGGAGGGAAACA-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               1
## AAACGGGGTATAGGTA-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               1
## AAAGATGAGGCCGAAT-1-PM-PS-0001-T-A1   SMC01 SMC01-T         CMS2               2
```

```
## [1] 17334
```

```r
label <- subdataset@meta.data[, c(3,4,7)]
colnames(label)[3] <- "res_0.3"
label <- subset(label, !res_0.3 %in% c('7'))
label <- droplevels(label)
head(label, n=3); nrow(label) # 17276
```

```
##                                    Library Patient res_0.3
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1 PM-PS-0001-T-A1   SMC01       2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1 PM-PS-0001-T-A1   SMC01       0
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1 PM-PS-0001-T-A1   SMC01       2
```

```
## [1] 17276
```

```r
rhp_scores <- data.frame(matrix(ncol = length(levels(label$res_0.3)), nrow = ncol(rhps)))
colnames(rhp_scores) <- levels(label$res_0.3)
rownames(rhp_scores) <- colnames(rhps)
rhp_scores
```

```
##              0  1  2  3  4  5  6
## Cycle       NA NA NA NA NA NA NA
## Stress      NA NA NA NA NA NA NA
## Interferon  NA NA NA NA NA NA NA
## Hypoxia     NA NA NA NA NA NA NA
## Oxphos      NA NA NA NA NA NA NA
## Metal       NA NA NA NA NA NA NA
## Mesenchymal NA NA NA NA NA NA NA
## pEMT        NA NA NA NA NA NA NA
## Alveolar    NA NA NA NA NA NA NA
## Basal       NA NA NA NA NA NA NA
## Squamous    NA NA NA NA NA NA NA
## Glandular   NA NA NA NA NA NA NA
## Ciliated    NA NA NA NA NA NA NA
## AC          NA NA NA NA NA NA NA
## OPC         NA NA NA NA NA NA NA
## NPC         NA NA NA NA NA NA NA
```

```r
for (clst in colnames(rhp_scores)) {
  tmpObj <- subset(subdataset, idents = clst)
  tmpExpr <- as.matrix(GetAssayData(tmpObj, slot = 'data'))

  tmpExprMean <- data.frame(row.names = rownames(tmpExpr),
                            Symbol = unlist(lapply(rownames(tmpExpr), function (x) unlist(strsplit(as.ch
                            cluster = rowMeans(tmpExpr))
  head(tmpExprMean)

  for (rhp in colnames(rhps)) {
    rhp_genes <- as.character(rhps[, rhp][rhps[, rhp] != ""])
    tmpExprMean_rhp_genes <- subset(tmpExprMean, Symbol %in% rhp_genes)
    head(tmpExprMean_rhp_genes)
    rhp_scores[rhp, clst] <- mean(tmpExprMean_rhp_genes$cluster)

  }
```

```
}

head(rhp_scores, n=3)
```

```
##                    0         1         2         3         4         5
## Cycle     0.2309410 0.1947164 0.4563568 0.1821317 0.2220477 0.1105891
## Stress    0.5990444 0.6048755 0.5514586 0.2197096 0.5904205 0.5945622
## Interferon 0.2783132 0.3567254 0.2862536 0.2165848 0.2779594 0.1658859
##                    6
## Cycle     0.2789568
## Stress    0.5886283
## Interferon 0.3641805
```
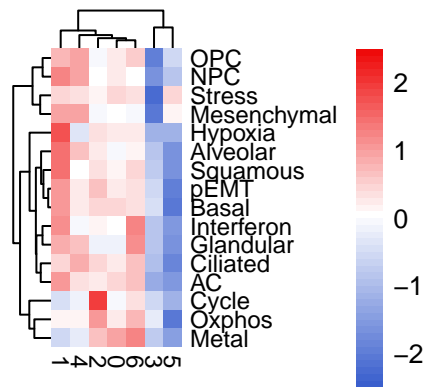
```
min(rhp_scores)
```

```
## [1] 0.003211767
```

```
max(rhp_scores)
```

```
## [1] 0.9750182
```

```
pheatmap(rhp_scores, scale = 'row', #filename = '../res_various/res_0.3/dbarkley_rhps.pdf',
         treeheight_row = 10, treeheight_col = 10,
         cellwidth = 7, cellheight = 7, fontsize_row = 9, fontsize_col = 9, border_color = NA,
         clustering_distance_rows = 'euclidean', clustering_distance_cols = 'euclidean', clustering_met
         breaks = c(-25:25)/10,
         color = colorRampPalette(c("#3a5fcd", 'white', "#ee0000"))(n = 51))
```

```
#dev.off()
```