

test_dcq.R

diez

Wed Jan 6 12:47:49 2016

```
library(glmnet)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following object is masked from 'package:tidyr':
##
##     expand
##
## Loading required package: foreach
## Loaded glmnet 2.0-2

db <- read.table("data/CellSurfaceMarkersDict.txt", stringsAsFactors = FALSE)
#head(db)
rownames(db) <- db[,2]
#head(db)

x <- read.delim("data/immgenGeneExpressionData.txt", stringsAsFactors = FALSE, check.names = FALSE)
#head(x)
x <- x[x[,1] %in% db[,1], ]
rownames(x) <- x[,1]
#head(x)
x <- as.matrix(x[, -1])
#head(x)
#x[1:10, 1:10]
x <- x[order(rownames(x)), order(colnames(x))]
#x[1:10, 1:10]

y <- read.delim("data/LungDataSet.txt", check.names = FALSE)
#head(y)
library(limma)
y <- avereps(y[, -1], ID = y[, 1])
#head(y)
y <- y[rownames(y) %in% db[, 2], ]
rownames(y) <- db[rownames(y), 1]
#head(y)
y <- y[rownames(x), ] # reorder
#head(y)

#####
# test
# fit2 <- glmnet(x, y[, 1], family = c('gaussian'), alpha = .05, nlambda = 100, lambda.min = .2)
#
# foo <- fit2$beta[, 100]
# barplot(foo[foo != 0])
```

```
#####
```

```
## do all.
res <- lapply(1:ncol(y), function(k) {
  fit2 <- glmnet(x, y[,k], family = c('gaussian'), alpha = .05, nlambda = 100, lambda.min.ratio = .2)
  fit2$beta[,100]
})
res <- do.call(rbind, res)
rownames(res) <- colnames(y)
#head(res)

res <- res[, colSums(res) != 0]

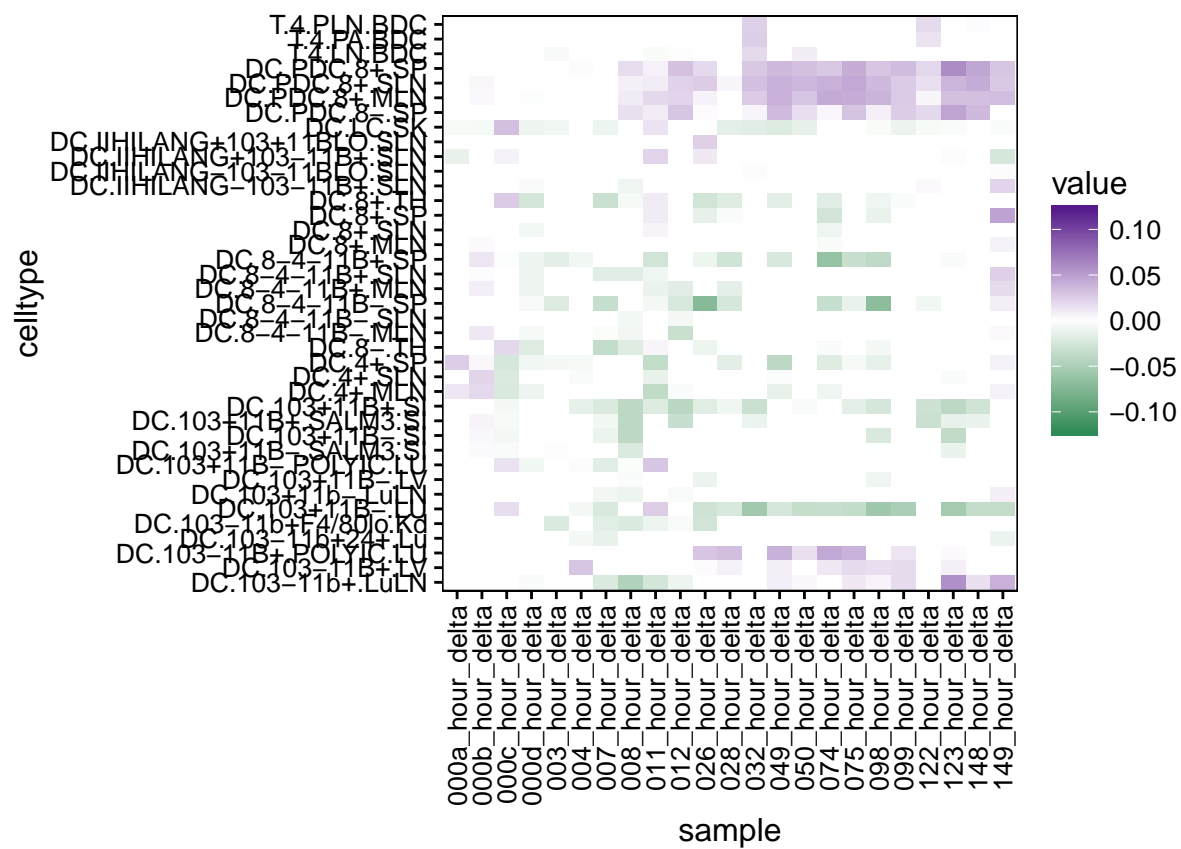
d <- melt(res, varnames = c("sample", "celltype"))
head(d)
```

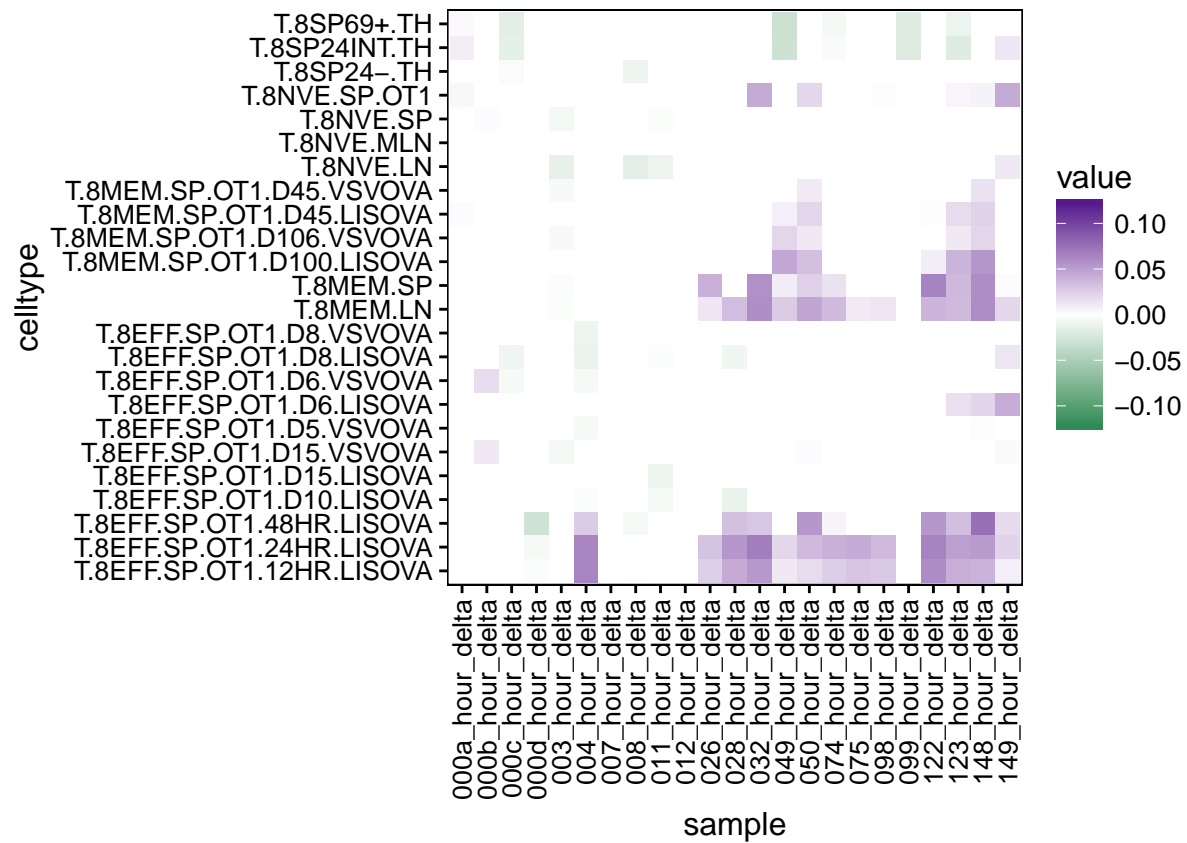
```
##           sample celltype      value
## 1 000a_hour_delta B.FO.LN -0.0152802878
## 2 000b_hour_delta B.FO.LN  0.0000000000
## 3 000c_hour_delta B.FO.LN  0.0000000000
## 4 000d_hour_delta B.FO.LN  0.0000000000
## 5  003_hour_delta B.FO.LN -0.0003130248
## 6  004_hour_delta B.FO.LN  0.0000000000
```

```
#ggplot(d,aes(x = sample, y = celltype,fill = value)) + geom_tile() + scale_fill_gradient2(low = "seagrass", high = "mud", mid = "sand")
```

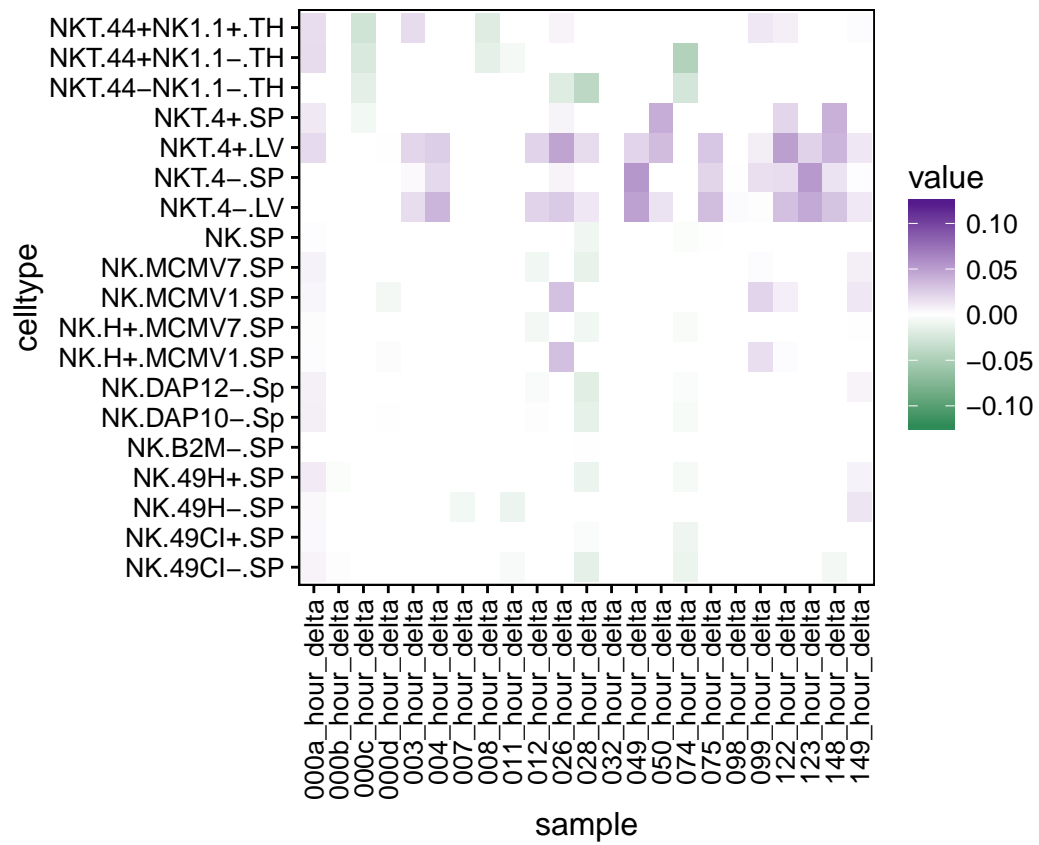
```
# DCs
```

```
ggplot(d %>% filter(grepl("DC", celltype)),aes(x = sample, y = celltype,fill = value)) + geom_tile() + scale_fill_gradient2(low = "seagrass", high = "mud", mid = "sand")
```



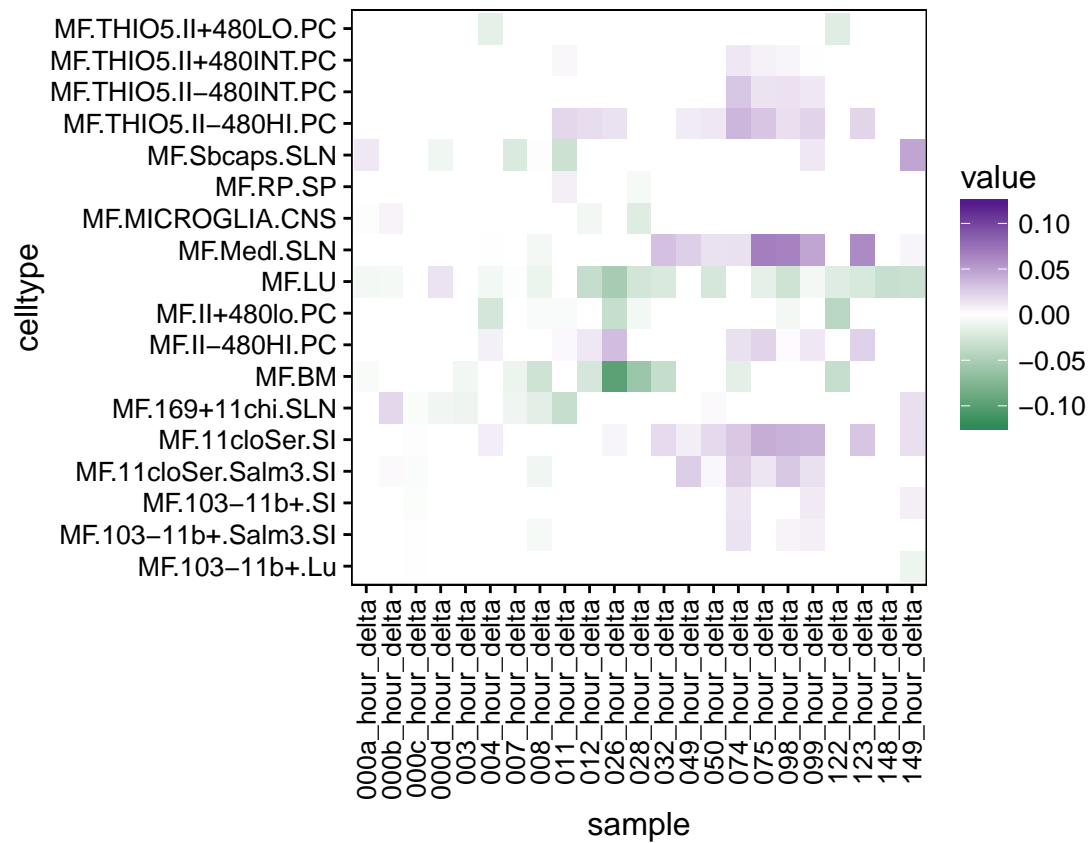


```
# NK(T)
ggplot(d %>% filter(grepl("NK", celltype)), aes(x = sample, y = celltype, fill = value)) + geom_tile() +
```



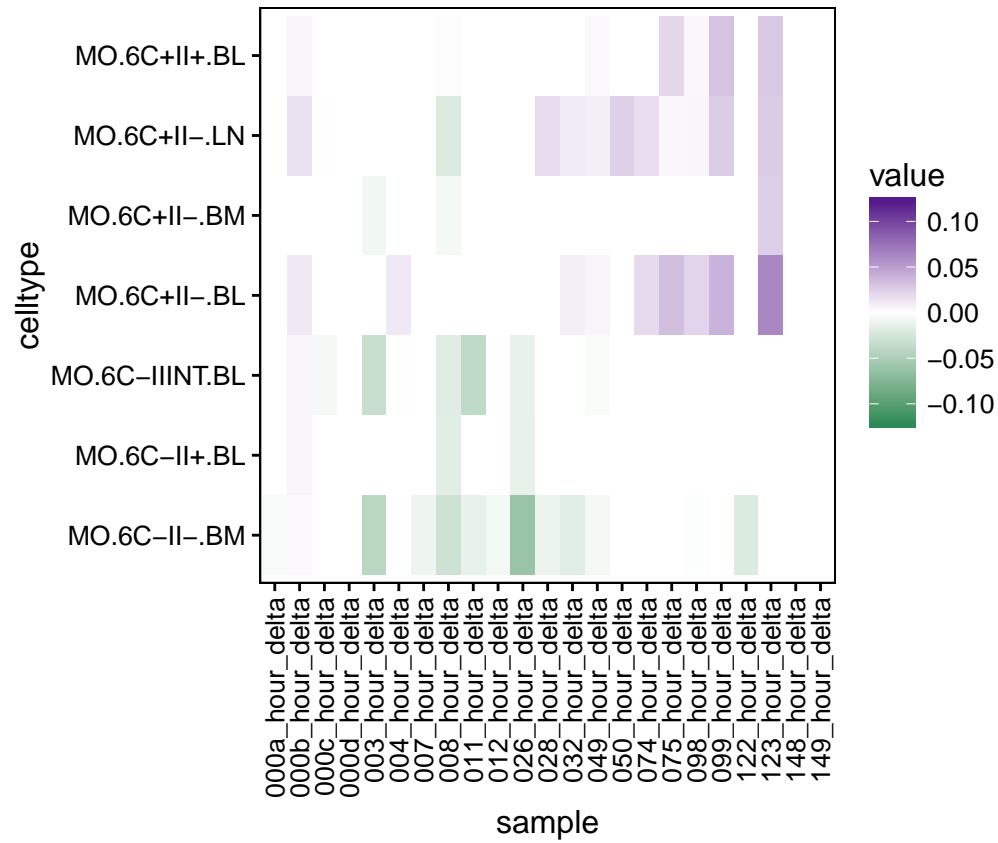
```
# MF
```

```
ggplot(d %>% filter(grepl("MF", celltype)), aes(x = sample, y = celltype, fill = value)) + geom_tile() +
```



MO

```
ggplot(d %>% filter(grepl("MO", celltype)), aes(x = sample, y = celltype, fill = value)) + geom_tile() +
```



```
# SC
```

```
ggplot(d %>% filter(grepl("SC", celltype)), aes(x = sample, y = celltype, fill = value)) + geom_tile() +
```

