

Explorando el Transcriptoma con Datos de Expresión Genética

GSEA

Yered Pita-Juárez

8/1/2015

Gene Set Enrichment Analysis

- Prueba de expresion diferencial usando grupos de genes
- La idea principal es usar grupos de genes pre definidos
- Definir los grupos de genes para ayudar a interpretar los resultados
- Analizar muestras del ALL
- leucemia linfoide aguda en celulas B
- Dos grupos: con o sin la mutación BCR/ABL

```
library("ALL")  
data("ALL")  
bcell = grep("^B", as.character(ALL$BT))  
moltyp = which(as.character(ALL$mol.biol) %in% c("NEG", "BCR/ABL"))  
ALL_bcrneg = ALL[, intersect(bcell, moltyp)]  
ALL_bcrneg$mol.biol = factor(ALL_bcrneg$mol.biol)
```

Gene Set Enrichment Analysis

- Filtrado: seleccionar los genes que estan por arriba del percentil del 50% de la variabilidad

```
library("hgu95av2.db")  
library("genefilter")  
ALLfilt_bcrneg = nsFilter(ALL_bcrneg, var.cutoff=0.5)$set
```

- KEGG: base de datos que contiene genes en rutas biologicas
- Definir los grupos de genes basados en rutas biologicas

```
library("GSEABase")  
gsc = GeneSetCollection(ALLfilt_bcrneg,  
                        setType=KEGGCollection())  
  
Am = incidence(gsc)  
dim(Am)  
nsF = ALLfilt_bcrneg[colnames(Am),]
```

Gene Set Enrichment Analysis

- Solo vamos a considerar las rutas biológicas con mas de 10 genes

```
selectedRows = (rowSums(Am)>10)  
Am2 = Am[selectedRows, ]
```

- Vamos a usar la prueba de GSEA

```
library("Category")  
set.seed(123)  
NPERM = 1000  
pvals = gseattperm(nsF, nsF$mol.biol, Am2, NPERM)  
pvalCut = 0.025  
lowC = names(which(pvals[, 1]<=pvalCut))  
highC = names(which(pvals[, 2]<=pvalCut))
```

Gene Set Enrichment Analysis

- Determinar el nombre de las rutas biológicas que tienen expresión diferencial

```
library("KEGG.db")  
head(getPathNames(lowC))  
head(getPathNames(highC))
```

Ejercicio

Trata de interpretar las rutas biológicas que están diferencialmente expresadas.

"Homologous recombination"

"Ribosome"

"Complement and coagulation cascades"

"Lysosome"

"Axon guidance"

"Pathogenic Escherichia coli infection"

"Shigellosis"

"Viral myocarditis"