

Explorando el Transcriptoma con Datos de Expresión Genética

Datos de Expresión Genética

Yered Pita-Juárez

6/1/2015

ALL

- 128 pacientes con leucemia linfoide aguda (acute lymphoid leukemia)
- Afecta los los linfocitos en la médula ósea
- Linfocitos: células del sistema inmune
- Puede afectar
 - ▶ linfocitos B
 - ▶ linfocitos T
- Paquetes y datos

```
library("Biobase")  
library("ALL")  
library("genefilter")  
data("ALL")
```

ALL

- 95 muestras con células B
- 33 muestras con células T

```
ALL$BT
```

- Seleccionar muestras con células B

```
bcell = grep("^B", as.character(ALL$BT))
```

- Dos tipos de muestras en células B: mutación BCR/ABL
- Translocación: segmento del cromosoma 22 intercambiado con un segmento del cromosoma 9

```
types = c("NEG", "BCR/ABL")
```

```
moltyp = which(as.character(ALL$mol.biol) %in% types)
```

```
ALL_bcrneg = ALL[, intersect(bcell, moltyp)]
```

```
ALL_bcrneg$mol.biol = factor(ALL_bcrneg$mol.biol)
```

¿Qué genes están siendo expresados de manera diferente entre los tipos de leucemia de las células B?

- Buena parte de las sondas en el microarreglo no es muy informativa
 - ▶ el gen no está expresado
 - ▶ la sonda no lo pudo detectar
- Pre seleccionar genes basados en la desviación estándar
- Poca variabilidad: no es posible distinguir si los genes estan siendo expresados de manera diferente

Filtrado

- `shorth`: promedio del intervalo más pequeño que contiene la mitad de las observaciones

```
library("genefilter")  
sds = rowSds(exprs(ALL_bcrneg))  
sh = shorth(sds)  
sh
```

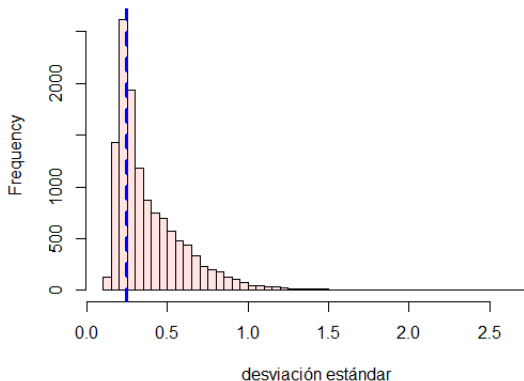
- Seleccionar las sondas con niveles de expresión por arriba de `sh`

```
ALLsfilt = ALL_bcrneg[sds>=sh, ]  
dim(exprs(ALLsfilt))  
dim(exprs(ALL_bcrneg))
```

Filtrado

- Gráfica

```
hist(sds, breaks=50, col="mistyrose",  
     xlab="desviación estándar",main="")  
abline(v=sh, col="blue", lwd=3, lty=2)
```



Prueba de t

- Prueba de t para cada gen comparando los dos tipos de células B

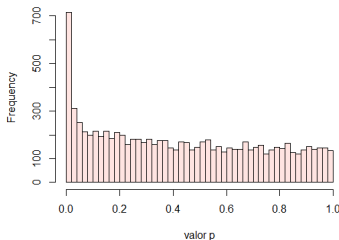
```
table(ALLsfilt$mol.biol)
```

```
tt = rowttests(ALLsfilt, "mol.biol")
```

- Valores p

```
hist(tt$p.value, breaks=50, col="mistyrose",  
      xlab="valor p",main="")
```

- Algunas sondas con valores p pequeños: genes diferencialmente expresados



Comparaciones Múltiples

- Con un valor α de 0.05, vamos a rechazar la hipótesis nula en 441 casos (5%) al azar
- Ajustar el valor alpha

```
library("multtest")  
mt = mt.rawp2adjp(tt$p.value, proc="BH")
```

- Los genes más significativos de acuerdo al valor p ajustado

```
g = featureNames(ALLsfilt)[mt$index[1:10]]
```

- Nombres

```
library("hgu95av2.db")  
links(hgu95av2SYMBOL[g])
```


Ejercicio

Trata de interpretar los genes que estan diferencialmente expresados.

```
> links(hgu95av2SYMBOL[g])
```

	probe_id	symbol
1	1635_at	ABL1
2	1636_g_at	ABL1
3	1674_at	YES1
4	32434_at	MARCKS
5	37015_at	ALDH1A1
6	37027_at	AHNAK
7	39730_at	ABL1
8	39837_s_at	ZNF467
9	40202_at	KLF9
10	40504_at	PON2

Comparaciones Múltiples

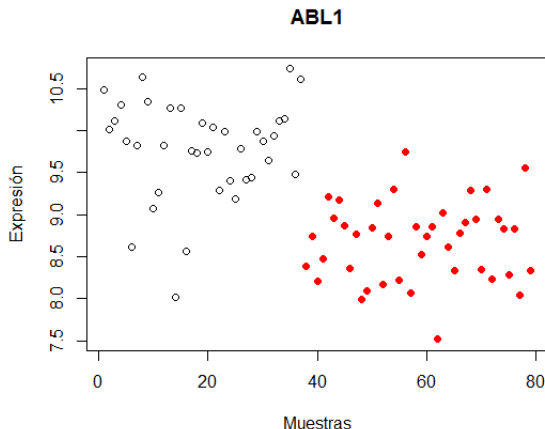
- Gráfica

```
mb = ALLsfilt$mol.biol
y = exprs(ALLsfilt)[g[1],]
ord = order(mb)
plot(y[ord], pch=c(1,16)[mb[ord]],
      col=c("black", "red")[mb[ord]],
      main=hgu95av2SYMBOL[g[1]], ylab="Expresión",
      xlab="Muestras")
```

- Negro: Mutación

- Rojo: Sin mutación

Comparaciones Múltiples



- Abelson murine leukemia viral oncogene homolog 1 (ABL1)
- Proteína codificada por el gen ABL1
- Gen localizado en el cromosoma 9