

Tercer Taller

Estadística genómica

Juan David Henao Sánchez

28 de septiembre de 2015

Sobre datos del GEO del NCBI de su elección (que comparen dos condiciones biológicas con al menos 5 réplicas) realice los siguientes pasos luego de normalizar:

1. Realice un MA-plot

- Carga de librerías

```
> library(GEOquery)
> library(vsn)
```

- Extrayendo datos

```
> gds <- getGEO("GDS3750")
> eset <- GDS2eSet(gds, do.log2 = TRUE)
> class(eset)
```

```
[1] "ExpressionSet"
attr(,"package")
[1] "Biobase"
```

```
> eset
```

```
ExpressionSet (storageMode: lockedEnvironment)
assayData: 22277 features, 8 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: GSM430339 GSM430340 ... GSM430346 (8 total)
  varLabels: sample genotype/variation description
  varMetadata: labelDescription
featureData
  featureNames: 1007_s_at 1053_at ... AFFX-TrpnX-M_at (22277 total)
  fvarLabels: ID Gene title ... GO:Component ID (21 total)
  fvarMetadata: Column labelDescription
experimentData: use 'experimentData(object)'
pubMedIds: 20395301
Annotation:
```

```
> dim(eset)
```

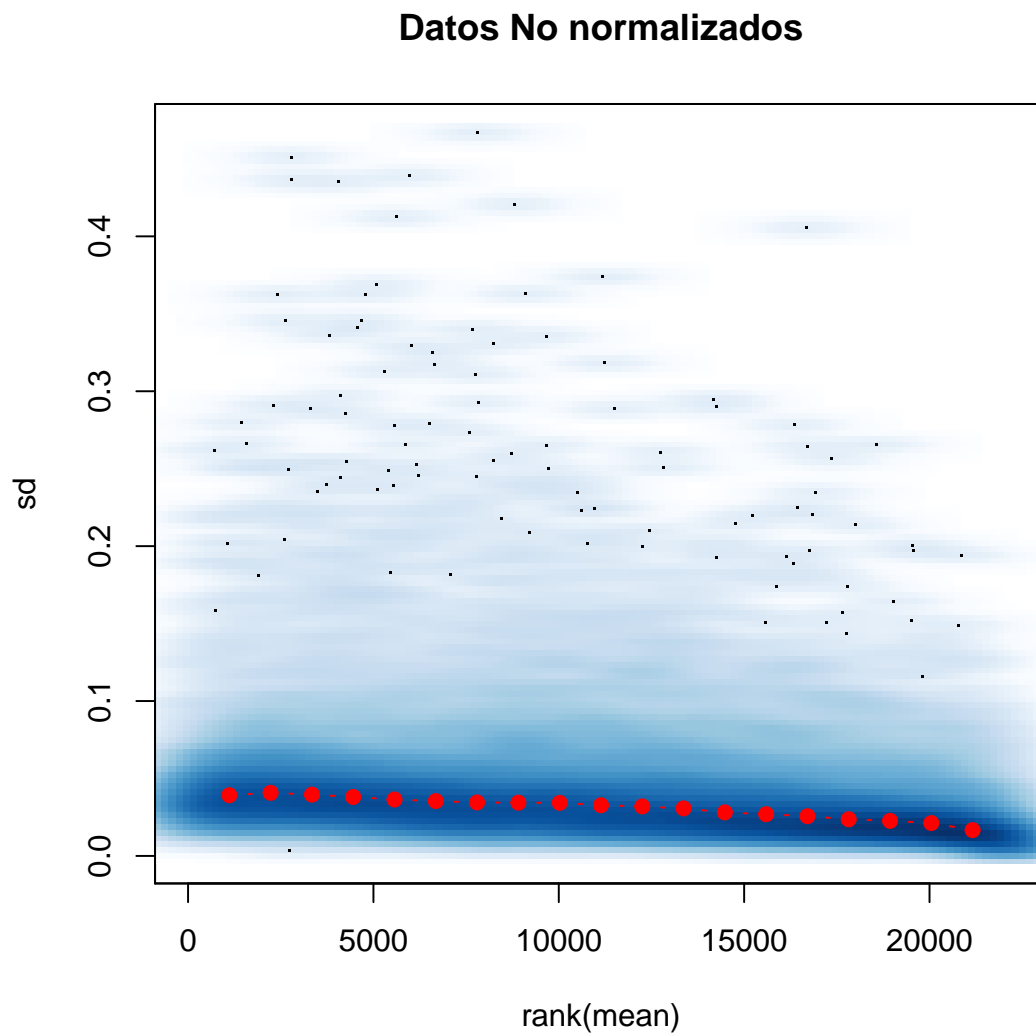
```
Features  Samples  
  22277      8
```

- Normalización

```
> nml <- justvsn(eset)
```

- Grafica de datos No normalizados

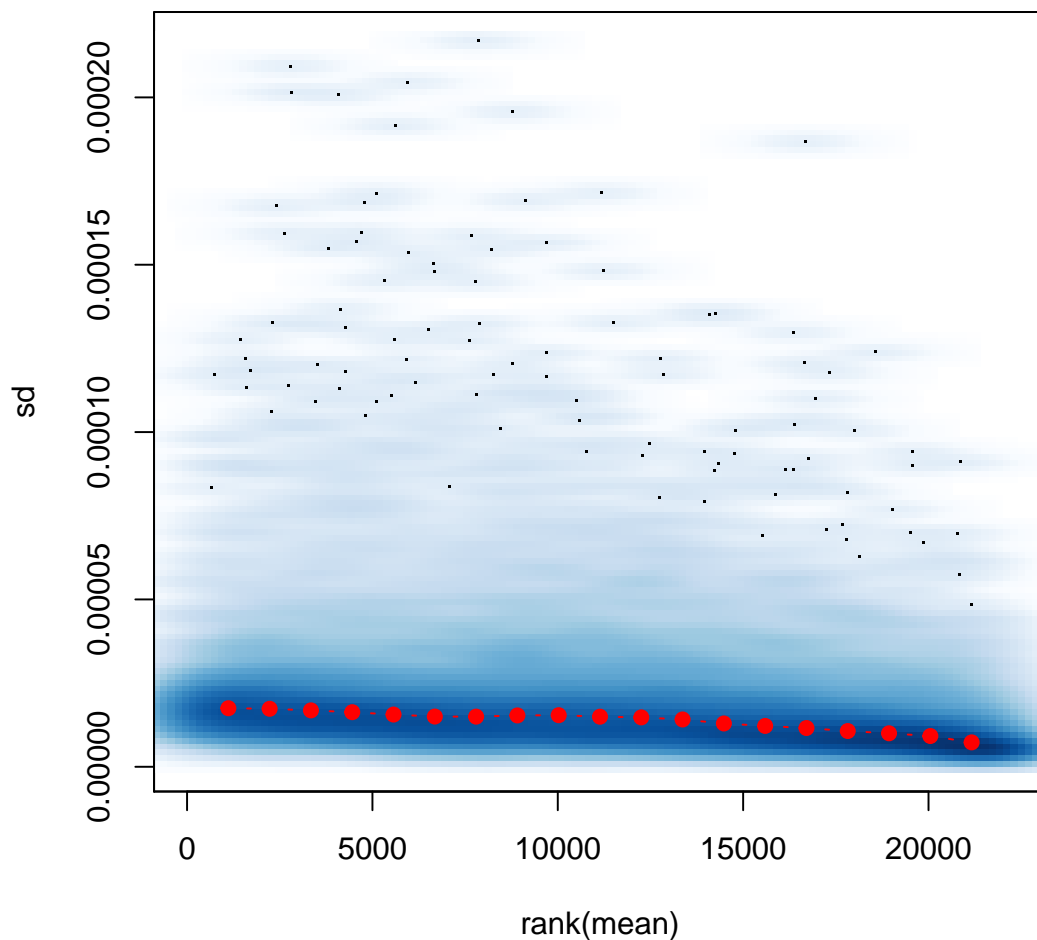
```
> meanSdPlot(eset)  
> title(main='Datos No normalizados',font=2)
```



- Grafica de datos normalizados

```
> meanSdPlot(nml)  
> title(main='Datos normalizados',font=2)
```

Datos normalizados



MA PLOT

■ Calculando variables

```
> iref = seq(1, 7, by=2)
> ismp = seq(2, 8, by=2)
> M= exprs(nml)[,ismp]-exprs(nml)[,iref]
> A=(exprs(nml)[,ismp]+exprs(nml)[,iref])/2
> dim(M)
```

```
[1] 22277      4
```

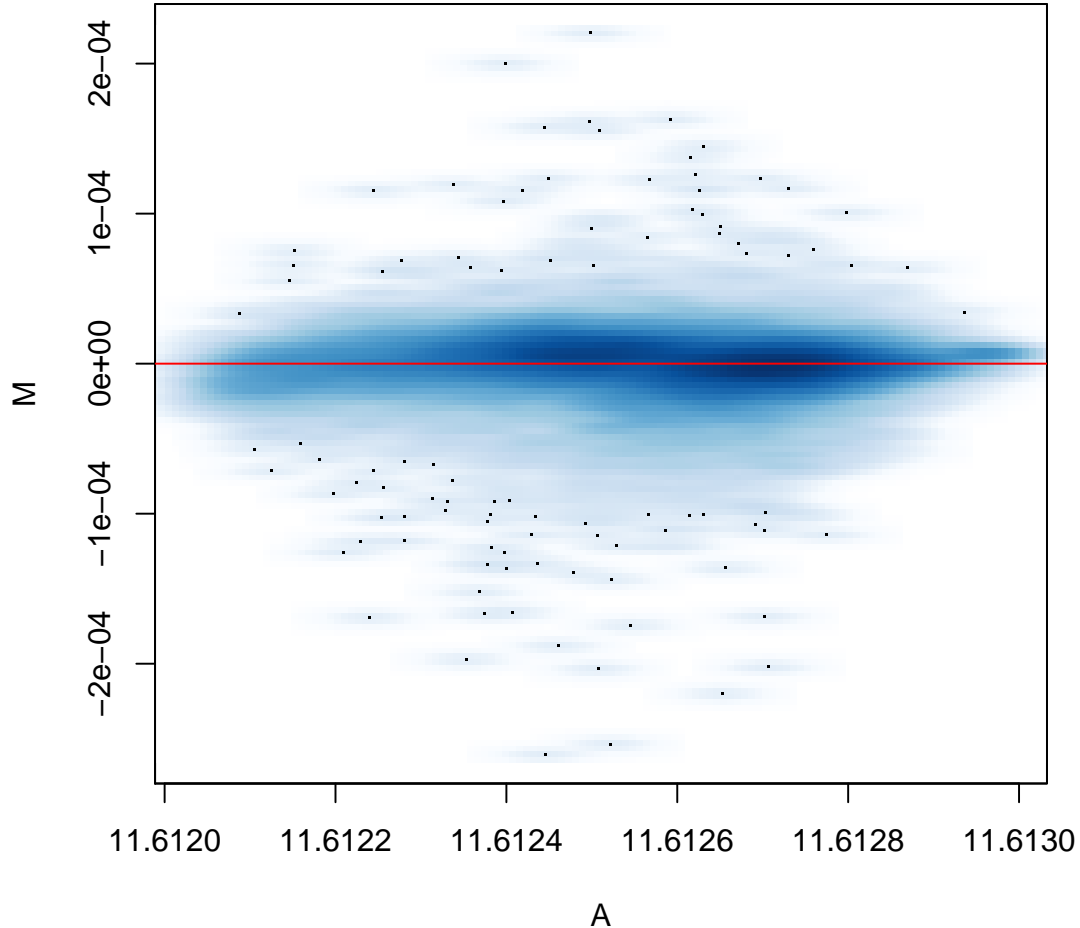
```
> dim(A)
```

```
[1] 22277      4
```

■ Generando la grafica

```
> smoothScatter(rowMeans(A), rowMeans(M), main=" ", xlab="A", ylab="M")
> title(main='MA PLOT',font=2)
> abline(h=0, col="red")
```

MA PLOT



2. Identifique los genes diferencialmente expresados con pruebas T múltiples (rowttest) y resáltelos en el MAplot

- Cargando la libreria necesaria

```
> library(genefilter)
```

- Eliminando los datos N/A

```
> Expresados <- exprs(nml)[complete.cases(exprs(nml)),]
> dim(Expresados)
```

```
[1] 22277      8
```

- Aplicando la prueba T múltiple

```
> PruebaT<-rowttests(Expresados[,c(1:8)],factor(c(0,0,0,0,1,1,1,1)))$p.value<0.05
```

- Resaltando los genes diferencialmente expresados

```

> smoothScatter(rowMeans(A), rowMeans(M), main=" ", xlab="A", ylab="M")
> title(main='MA PLOT',font=2)
> abline(h=0, col="red")
> points(rowMeans(A[PruebaT,]),rowMeans(M[PruebaT,]),col="red",cex=0.1)

```

