Tercer Taller
Estadística genómica

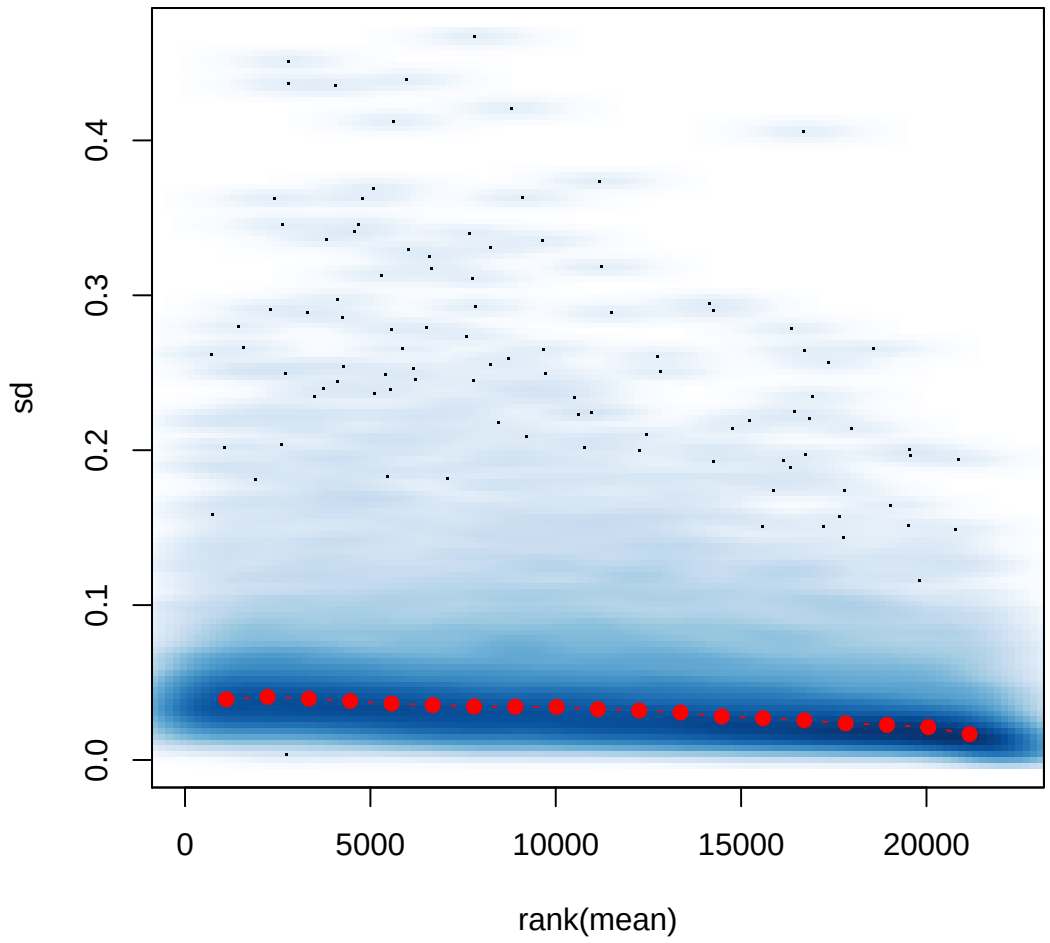
Juan David Henao Sánchez

22 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:

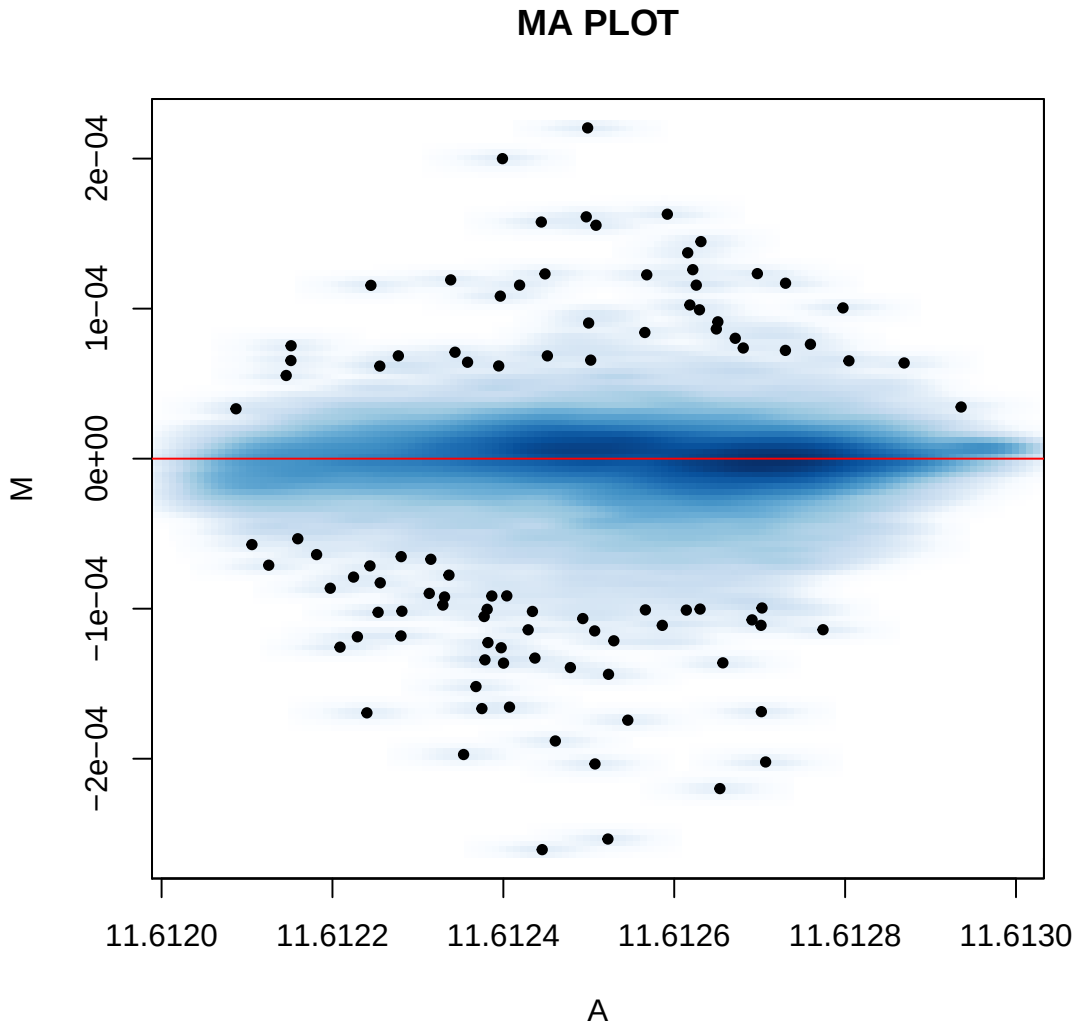
1. Realice un MA-plot

```
> library(GEOquery)
> library(vsn)
> ########################
> gds <- getGEO("GDS3750")
> eset <- GDS2eSet(gds, do.log2 = TRUE)
> 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

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

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



MA PLOT

```
> ################################
> 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

> smoothScatter(rowMeans(A), rowMeans(M), main=" ", xlab="A", ylab="M", pch=20)
> title(main='MA PLOT',font=2)
> abline(h=0, col="red")
```



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

```
> library(genefilter)
> library("GSEABase")
> 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)
> class(ALL_bcrneg)

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

> ####################################
> gsc=GeneSetCollection(ALL_bcrneg, setType=KEGGCollection())
> gsc

GeneSetCollection
  names: 04610, 00232, ..., 00785 (228 total)
```

```
  unique identifiers: 189_s_at, 31825_at, ..., 41859_at (5333 total)
  types in collection:
    geneIdType: AnnotationIdentifier (1 total)
    collectionType: KEGGCollection (1 total)
```

```
> Am= incidence(gsc)
> dim(Am)

[1]  228 5333

> ######################################
> nsF=ALL_bcrneg[colnames(Am),]
> dim(nsF)

Features  Samples
    5333       79
```

PRUEBA T (rowttest)

```
> rtt=rowttests(nsF, "mol.biol")
> rttStat=rtt$statistic
> dim(rtt)

[1] 5333    3

> names(rtt)

[1] "statistic" "dm"        "p.value"

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

[1]  228 5333

> dim(Am2)

[1]  207 5333

> dim(rtt)

[1] 5333    3

> ####################################
> z=0
> for(i in 1:dim(Am2)[1]){
+ z[i]=sum(rttStat[Am2[i,]==1])/sqrt(sum(Am2[i,]))
+ }
> length(z)

[1] 207

> ########################################
> resGSEA <- cbind(rownames(Am2),z)
> resGSEAdown <- resGSEA[resGSEA[,2]<(-1.96),]
> resGSEAup <- resGSEA[resGSEA[,2]>1.96,]
```