

# Anotacion de transcriptoma

*pgonzale*

*April 8, 2016*

## Distribucion de dominios pfam

Usando su reporte de Trinotate, generen una lista de los dominios de PFAM, ya sea en R o en excel. Generen una gráfica de barras de los 10 dominios más abundantes en la anotación de su transcriptoma. Debajo de la gráfica discutan brevemente (1 párrafo de máximo 10 líneas) cual es la relevancia biológica de conocer estos dominios.

Lo primero fue extraer los Pfams a partir del archivo .xls. Este comando me lo paso Falcon.

```
cut -f8 *xls|sed '1d' | sed '/^.$/d'|sed 's/^\n/g'| sed 's/^\n.*//> PFsOfAnnotation.txt
```

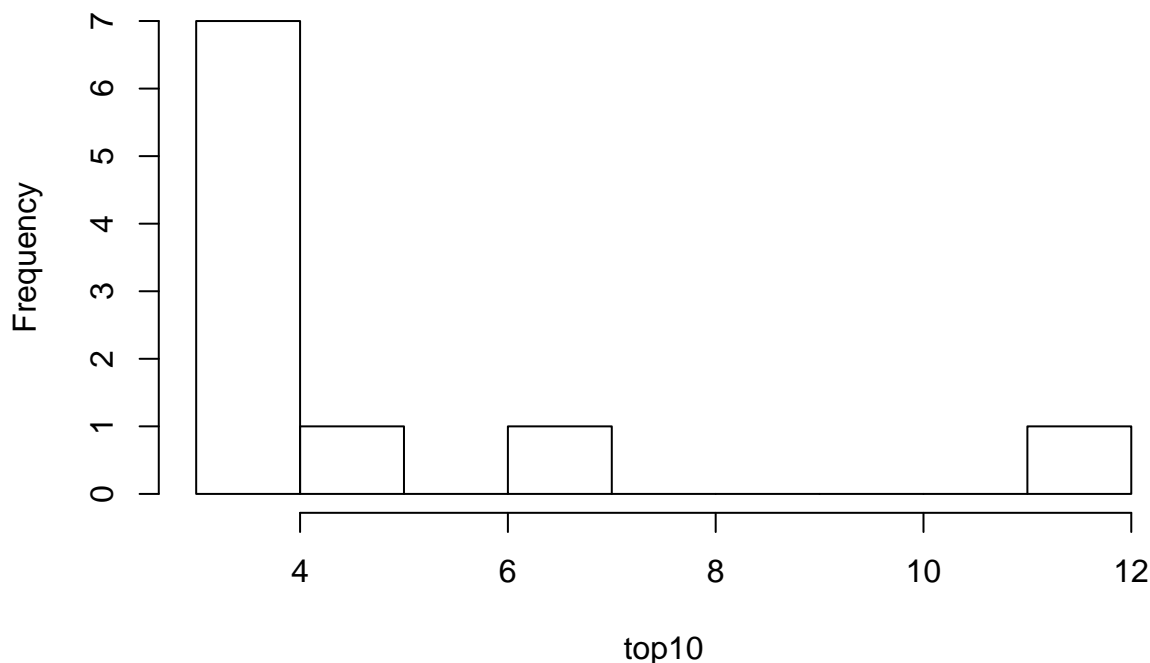
Ahora sí, evaluamos los 10 dominios más representados.

```
PFs <- read.delim("~/Documents/transcriptomic-exercise/output/annotation/PFsOfAnnotation.txt", quote="\n")
head(PFs[, 1])
```

```
## [1] "PF07716.12" "PF00128.21" "PF09260.8" "PF00171.19" "PF04032.13"
## [6] "PF08449.8"
```

```
top10 <- sort(table(PFs[, 1]), decreasing = T)[1:10]
hist(top10, main = "Frecuencia de Pfams", breaks = 12)
```

## Frecuencia de Pfams



```
# Quienes son:  
names(top10)
```

```
## [1] "PF00400.29" "PF00076.19" "PF00004.26" "PF00012.17" "PF00078.24"  
## [6] "PF00133.19" "PF00432.18" "PF12382.5" "PF00069.22" "PF00097.22"
```