

# Tarea de anotacion

*Juan Esteban*

*March 14, 2016*

Primero cargo la tabla con la frecuencia de cada dominio reportado. Viendo los datos, se puede ver que la mayoría solo es reportado una sola vez. Aquí nos vamos a enfocar en los primeros 10 dominios con la mayor frecuencia.

Las primeras 10 proteínas tienen las siguientes funciones:

- 1.- PF00400.29 (8 veces) WD40 repeat
- 2.- PF00076.19 (6 veces) RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
- 3.- PF12382.5 (4 veces) Retrotransposon peptidase
- 4.- PF00012.17 (4 veces) Hsp70
- 5.- PF13361.3 (3 veces) UvrD-like helicase C-terminal domain
- 6.- PF08585.9 (3 veces) RecQ mediated genome instability protein
- 7.- PF08449.8 (3 veces) UAA transporter family
- 8.- PF06723.10 (3 veces) MreB/Mbl protein
- 9.- PF03981.9 (3 veces) Ubiquinol-cytochrome C chaperone
- 10.- PF03303.10 (3 veces) WTF protein

Posiblemente una función podría ser la de ver dominios sobrerrepresentados extrapolándolos a una función específica. Sin embargo, esto se podría ver con más precisión en un análisis de expresión diferencial. Podría complementarse esta información con la de los dominios, puede que hayan dominios que se comparten entre algunas proteínas ligeramente sobrerrepresentadas que no se verían en la expresión diferencial. Asimismo, podría ser que haya algún caso en el que un dominio indique algún proceso específico que ocurra en la célula como algún patógeno o respuesta de estrés. En todo caso si hubiera algo que no concuerda podría servir como un control negativo, que los genes que más se expresen no coincidan con su motivo mayormente expresado.

```
PF<-read.table("/home/usuario/Documents/Biologia_computacional/Transcriptomica/Anotacion/PFAM_data",sep=";",as.is=T)
PF<-PF[,-174,]
first_10<-PF[228:238,]
barplot(first_10$V1,names.arg = as.vector(first_10$V2), las=2)
```

