

# 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 mayoria solo es reportado una sola vez. Aqui nos vamos a enfocar en los primeros 10 dominios con la mayor frecuencia.

Las primeras 10 proteinas 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 funcion podria ser la de ver dominios sobrerrepresentados extrapolandolos a una funcion especifica. Sin embargo, esto se podria ver con mas precision en un analisis de expresion diferencial. Podria complementarse esta informacion con la de los dominios, puede que hayan dominios que se comparten entre algunas proteinas ligeramente sobrerrepresentadas que no se verian en la expresion diferencial. Asimismo, podria ser que haya algun caso en el que un dominio indique algun proceso especifico que ocurra en la celula como algun patogeno o respuesta de estres. En todo caso si hubiera algo que no concuerda podria servir como un control negativo, que los genes que mas se expresen no coincidan con su motivo mayormente expresado.

## Frecuencia de los dominios mas comunes

