**Clase del 28-10**

Sigue con la explicación de dividir en ventana a la secuencia de los cromosomas para aplicar análisis con varias funciones. Dividir en ventanas.

Hago un cociente para determinar proporciones entre las ventanas. Por ejemplo, hacer un cociente C/G cuando supere el 1 hay mas C cuando sea mas chico hay mas G. Me guarda un número por ventana. Luego habría que hacer un gráfico.

Siempre que hacemos algo reiteradamente hago una función. Agarro una ventana y aplico la función.

Para crear una secuencia al azar:

usamos el comando **sample**, hay que definir las letras a usar y el tamaño y poner replace TRUE para que sea con reposición, pueda utilizar mas de una vez las bases sino no tiene sentido.

Función para este caso de C/G. **Creamos la función**

Nombre\_Funcion = **Function** (secuencia) {Argumento de la función, en este caso usamos count, antes habíamos definido el count2 y 3 que corresponde a C y G y divido **return**(Nombre\_Funcion) }

contar = count(seq,wordsize = 1,by = 1)

contar[2]

F = function(seq){

fraccion = contar[2] / contar[3]

return(fraccion)}

F(seq)

Luego tenemos que crear las ventanas. Esto es parecido al script de la clase del 25-10 pero lo toqueteamos un poco y explicamos.

1. Definimos tamaño de ventanas
2. Al dividir el largo de la secuencia por el largo definido de la cantidad e ventanas, obtenemos la cantidad de ventanas. Intentar que nos queden aprox 100 ventanas
3. Tenemos que crear el loop para indicarle a cada base dentro de la secuencia, que número de ventana le corresponde. A esto se le llama factor al parecer.

Ejemplo: actagctagctag…………..actag(B100)ctagctagctagc

11111111111…………..11111(B100)2222222222

Código usado:

length(seq2)

tamaño\_ventanas=100

cantidad\_ventanas = length(seq2)/tamaño\_ventanas

**f=list()**

**for (i in 1: round(length(seq2)/tamaño\_ventanas)) {**

**f[[i]] = rep(i,tamaño\_ventanas)**

**}**

1. Luego con lo obtenido por el factor, debemos unir todo, porque teníamos una lista. Para esto usamos la función unlist

Luego lo hay que hacer, es dividir esas ventanas en base al código de números que habíamos puesto, para esto usamos el código **Split** que crea una lista donde pone las 100 bases con referencia al nuemro de ventana determiando.

ventanas = split(x = seq2, f = f)

Al final, aplicamos la función a cada una de esas ventanas por medio de la función **sapply** y creamos el plot

C\_sobre\_G = sapply(ventanas, Fun)

plot(C\_sobre\_G)

Generalmente para hacer la comparación usamos g-c/g+c (diferencia relativa) o c-g (diferencia absoluta)Esto aplicando sobre los cds genómicos como ventanas.

**Interpretación**

Al final con los resultados podemos analizar los gráficos y determinar cosas en base a que fragmentos poseen tanto contenido g o c en la región. Podemos identificar zonas promotoras, reguladoras, eventos evolutivos, etc.