**Scripts**

A continuación se presentan los diferentes scripts utilizados en la tesis: “Pipeline Bioinformático para el Análisis de Repertorios de Inmunoglobulinas Secuenciadas por Nanopore”.

Los archivos de configuración de los script se encuentran en: “Archivos configuración y barcoding”

1.- Basecalling:

Los comandos utilizados se encuentran en el documento:

Comando\_basecalling.txt

2.- Nanoplot

Información Nanoplot.txt

3.- FastQC

El comando para utilizar 4 núcleos se encuentra en:

Comando\_fastqc.txt

4.- Estratificación

Para esto se trabaja con los summarys del proceso de basecalling

Se encuentran los script en “Script\_para\_estratificar”, donde se encuentra:

sequencing\_summary\_ejemplo.txt, que es un ejemplo de summary

Unir\_summarys\_2D.R, que se puede utilizar para unir los summarys, generando el archivo datos\_total.RData

El script 2D\_fail\_1.R, obtiene los reads con Qscore menor a 1

El script 2D\_fail\_2.R, obtiene los reads con Qscore mayor o igual a 1 y menor a 2

…

…

…

El script 2D\_fail\_9.R, obtiene los reads con Qscore mayor o igual a 9

5.- Convertir Fastq a Fasta

Comando para unir fastq y luego convertir los fastq a fasta

Convertir\_fastq\_a\_fasta.txt

6.- Script para separar el archivo fasta con todas las secuencias, en un archivo por secuencia

Obtener\_fastas\_individuales.R

7.- Script de demultiplexación de los amplicones

Primero se debe instalar bl2seq e instalada la biblioteca “Biostrings” de R.

En la carpeta “Script para obtener distancias” se tiene el script

El script se corre con el comando Rscript:

Rscript distancia\_alinear\_secuencias.R capeta\_con\_scuencias\_fastas\_individuales N°\_de\_primera\_secuencia N°\_última secuencia

Por ejemplo:

Rscript distancia\_alinear\_secuencias.R Secuencias\_analizar/ 1 10000

El script tiene por defecto la carpeta de salida "./resultados\_distancia\_secuencias\_3/"

Además hay dos archivos de configuración:

configuracion\_barcoding.txt, que contiene los nombres de los archivos fastas que contienen los archivos barcoding

configuracion\_alineamiento.txt, que contiene la ruta donde está instalado el programa /home/roberto/anaconda2/bin/

Se notará que existen dos archivos fastas para el barcoding IGL, esto es debido a que la secuencia:

“CTG AGG AGC TYC AAG CCA ACA”, tiene una Y, que corresponde a una C o T.

Como salida se obtiene una obtiene una matriz que contiene por cada barcoding: la distancia, Query, barcoding y el barcoding en el sentido que alineo.

8.- Obtención de datos forward y reverse

8.1.- Obtener\_data\_forward\_y\_reverse\_c.R

Utilizar el archivo "configuracion\_barcoding.txt" para la configuración de la carpeta “Archivos configuración y barcoding”

8.2.- dividir\_en\_forward\_reverse\_c.R

Archivos de configuración en:

Archivos configuración y barcoding

9.- Se eliminan las 4 primeras bases teóricas de Index y las 4 primeras letras de End Adaptor

eliminar\_bases.R

Se ingresa un archivo salida del script (8.2)

Y la carpeta de la salida, se coloca por ejemplo la carpeta: “sin\_4\_bases”

Ejemplo corriendo el script con Rscript

Rscript eliminar\_bases.R archivo\_forward.csv sin\_4\_bases

10.- Constucto

Construct\_seq.R

En la carpeta “constructo”, ejemplos de constructos generados.

11.- Obtener indels y sustituciones

comparaciones\_sustituciones.R

12. graficar indels y sustituciones

graficos\_sustitucion.R

graficos\_sustitucion\_25\_02\_2019.R

13. se agregan códigos de cobertura

cobertura\_19\_02\_2019\_2.R

cobertura\_20\_02\_2019.R

14.- gráficos de cadenas pacientes e ig

15.- Ejemplos de datos con los que se trabajaron

https://drive.google.com/drive/folders/15i-JgN0MczfVnu8muiGnpzckpbY3Wxks?usp=sharing