

## Reporte TumorSec: Control de calidad de los datos generados.

En el siguiente reporte se presentan diversas métricas de calidad correspondiente a la corrida 201219\_TumorSec este informe fue generado el día 22/12/20 por el pipeline bioinformático de TumorSec desarrollado por el laboratorio de Genómica del Cáncer, Universidad de Chile. Estas métricas representan una guía para verificar la calidad de los resultados obtenidos por cada muestra en la secuenciación.

### 2.- Control de calidad de la corrida:

#### 2.1.- Información de la corrida:

Nombre kit	MiSeq Standard Reagent Kit V2
Longitud de Lecturas	2x150 bp
Maximum Output	5.1 Gb
Cantidad de Lecturas	24-30 Millones
Calidad de Lecturas	80% > Q30.
Modo	Paired End
Equipo	MiSeq
N° Serie	M03158
Ciclos	2x150
Kit Preparación Librería	SeqCap EZ HyperCap Roche
Regiones Blanco	326,00
Largo Total Regiones Blanco (pb)	88.277,00

Tabla 1: Información del kit de secuenciación y regiones blanco de TumorSec

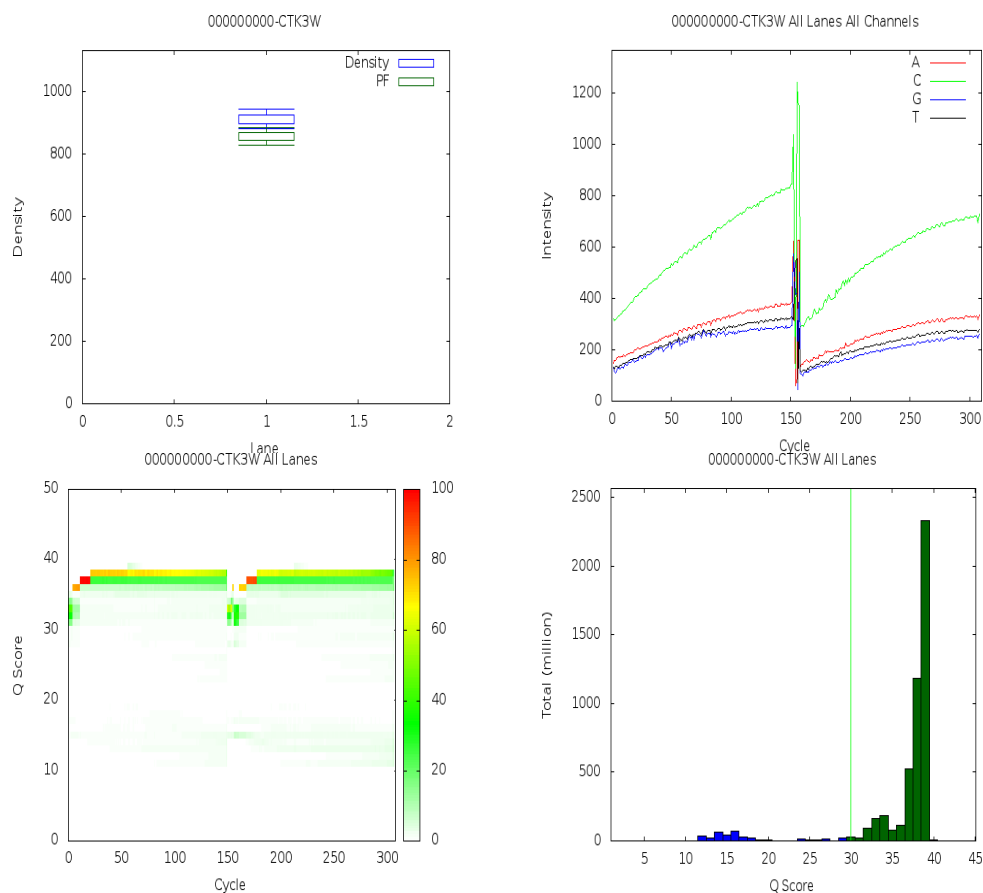


Imagen 1: Métricas de BaseSpace Illumina MiSeq. A: Densidad de clusters (K/mm2) por línea. B: Gráfico de intensidad por ciclo. C: Gráfico de score de calidad por ciclo. D: Diagrama de distribución de QScore, número de bases por score de calidad

## 2.1.- Métricas según Lecturas:

Read	Rendimiento (Gb)	Rend Esperado (Gb)	Intensidad	%>=Q30	Phas/Prephas (%)
Read1	2.47	2.47	147	95.38	0.095/0.135
Read2(l)	0.08	0.08	622	95.39	0.000/0.000
Read3	2.47	2.47	127	90.79	0.087/0.110
Non-indexed	4.94	4.94	137	93.08	nan
Total	5.03	5.03	299	93.12	nan

Tabla 2: Métricas de las lecturas de BaseSpace Illumina

## 2.2.- Métricas según Línea:

Línea	Densidad (K/mm2)	Clúster PF (%)
Línea 1	910+/-17	94.18+/-0.36

## 3.- Control de calidad de los datos generados.

### 3.1.1- Tabla resumen de las métricas de calidad

Sample	R_before trimming	R_after trimming	% of trim	% of dups	Reads On-Target	% On-Target	Mean read coverage	% uniformity
Undetermined_S0	969128.0	752140.0	22.39	25.57	391584.00	68.21	299.85	98.77
253843BTUCH_S1	4446976.0	3440000.0	22.64	65.54	654838.00	52.27	461.71	99.69
NA14638_S2	9171840.0	8803900.0	4.01	34.72	3953573.00	67.19	3068.09	98.47
NA14094_S3	7841518.0	7524674.0	4.04	33.41	3240862.00	63.22	2577.25	98.77
NA14634_S4	10526480.0	10144194.0	3.63	29.98	4970556.00	68.27	3804.38	98.77

Tabla 3: Métricas de calidad. Sample: Muestras. R\_before trimming: Número de lecturas iniciales. R\_after trimming: Número de lecturas despues de la poda .% of trim : Porcentaje de lecturas eliminadas. % of dups: Porcentaje de lecturas duplicadas. Reads On-Target: Número de lecturas en regiones blanco % On-Target: Porcentaje en regiones blanco. Mean read coverage: Profundidad promedio de lecturas en regiones blanco. % Uniformity: Porcentaje de lecturas en regiones blanco, donde su profundidad promedio está entre el rango  $[2 \times \text{mediana}, \text{mediana}/2]$ .

### 3.1.2- Profundidad y amplitud de las coberturas en las regiones blanco:

SAMPLE	MEAN_COVERAGE	MAX_COVERAGE	PCT_100X	PCT_200X	PCT_300X	PCT_400X	PCT_500X
Undetermined_S0	297.08	537.0	99.20	93.35	47.80	7.35	0.12
253843BTUCH_S1	449.81	896.0	99.88	99.22	95.11	67.22	27.67
NA14638_S2	3015.26	4808.0	100.00	100.00	100.00	100.00	100.00
NA14094_S3	2553.74	4015.0	100.00	100.00	100.00	100.00	100.00
NA14634_S4	3753.15	6005.0	100.00	100.00	100.00	99.99	99.99

Tabla 4: Métricas de cobertura en regiones targets

### 3.1.3- Amplitud de regiones blanco con cobertura mínima de 300X

SAMPLE	Prom_pct_amp	Nº < 70 %	Nº < 80 %	Nº < 90 %	Nº < 100 %
Undetermined_S0	48.43	187	209	239	268
253843BTUCH_S1	93.88	19	22	31	58
NA14638_S2	100.00	0	0	0	1
NA14094_S3	100.00	0	0	0	1
NA14634_S4	99.99	0	0	0	2

Tabla 5: Número de regiones blanco que poseen menos del 80% una profundidad mínima de 300X. Estas métricas también se observan para un 70%, 90% y 100% a un mínimo de 300X. Para resultados de mejor calidad, se espera un bajo número de regiones a 80% a un mínimo de 300X. El detalle de las regiones blanco se encuentra en los archivos excel y reporte complementario, generados por el pipeline TumorSec

### 3.2.- Calidad de bases por posición en fastq R1 y R2 sin filtrar: FASTQC

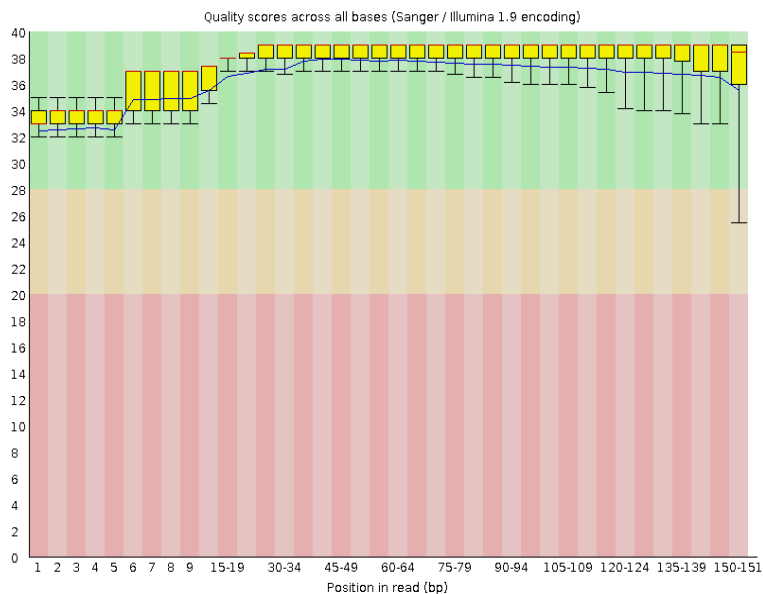


Imagen 2: Descripción general de los valores de calidad para todos los fastq R1. Se observa un gráfico de cajas con la distribución de las calidades por posición en la lectura. Este gráfico permite observar si los datos tienen algún problema antes de realizar cualquier análisis posterior.

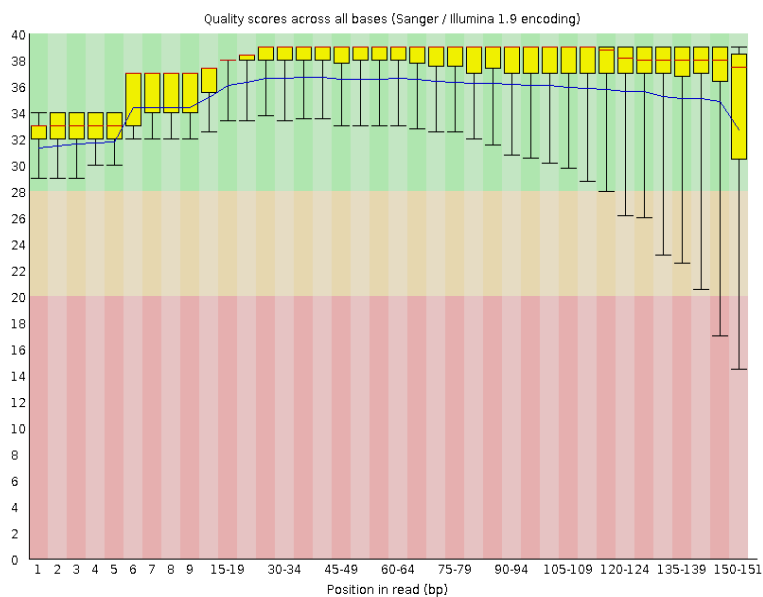


Imagen 3: Descripción general del rango de valores de calidad para todos los fastq R2 en cada posición.

### 3.3.- Cantidad de lecturas antes y después del trimming:

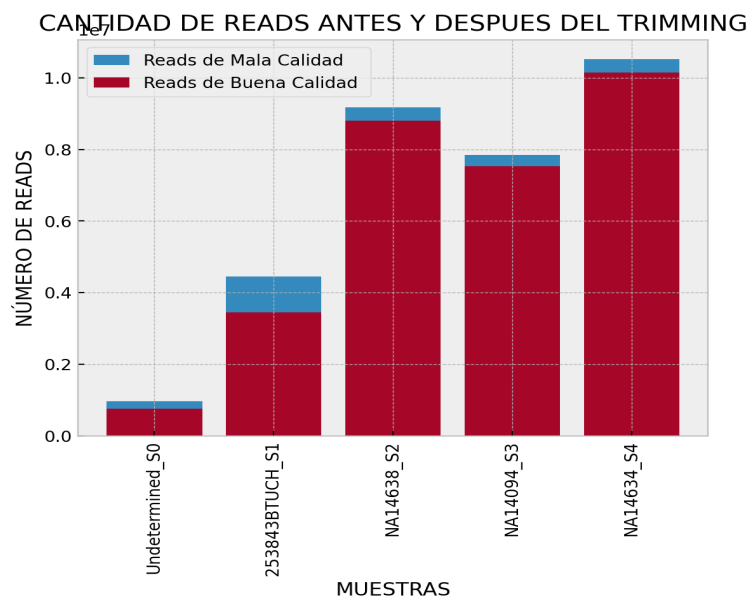


Imagen 4: Representación gráfica del trimming de datos realizado por fastp

### 3.4.- Calidad de lecturas antes y después del trimming:

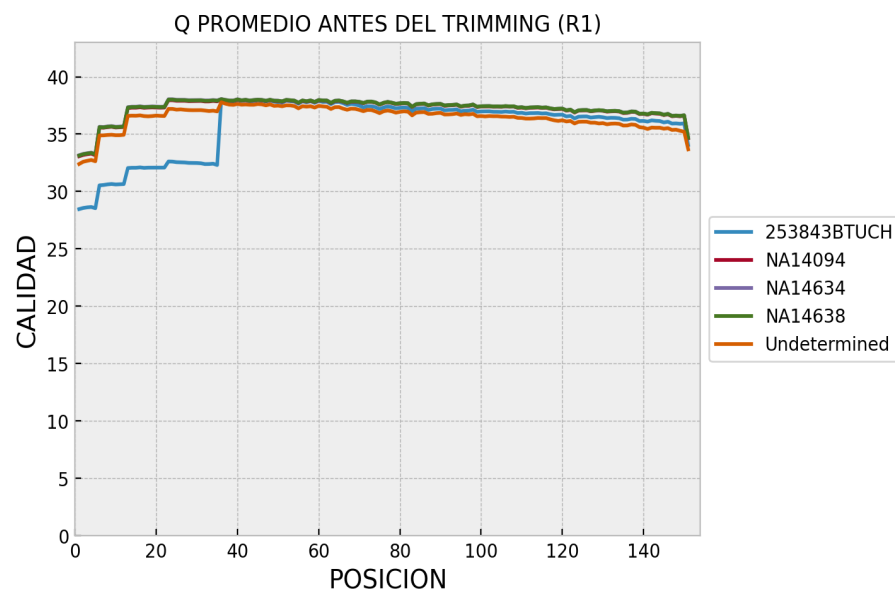


Imagen 5: Calidad promedio de lecturas por posición antes del trimming (R1)

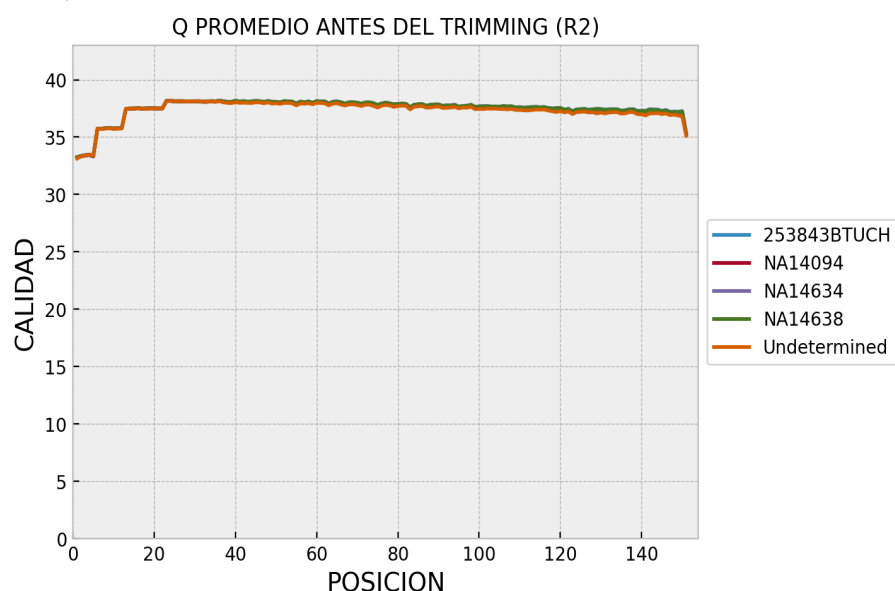


Imagen 6: Calidad promedio de lecturas por posición antes del trimming (R2)

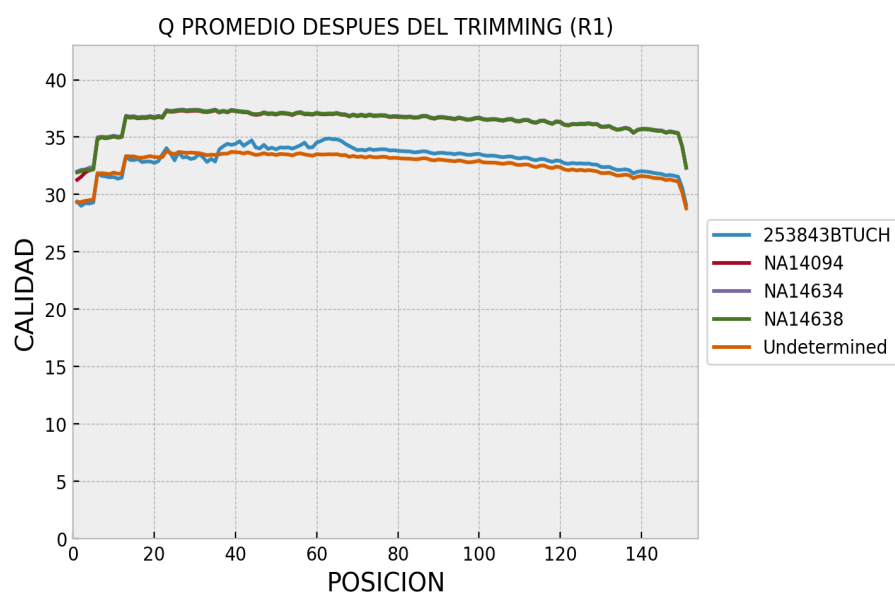


Imagen 7: Calidad promedio de lecturas por posición después del trimming (R1)

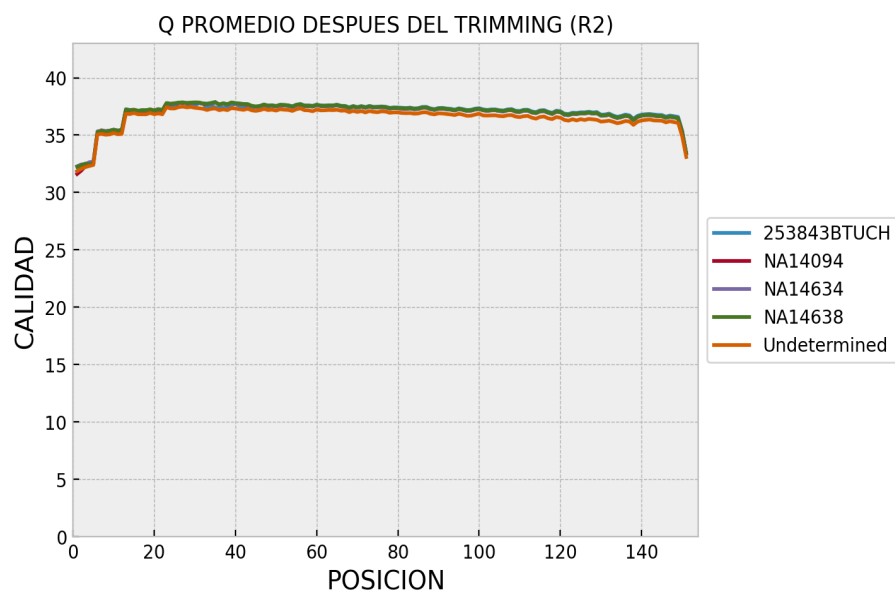


Imagen 8: Calidad promedio de lecturas por posición después del trimming (R2)

### 3.5.- Porcentaje de duplicados y tamaño del inserto

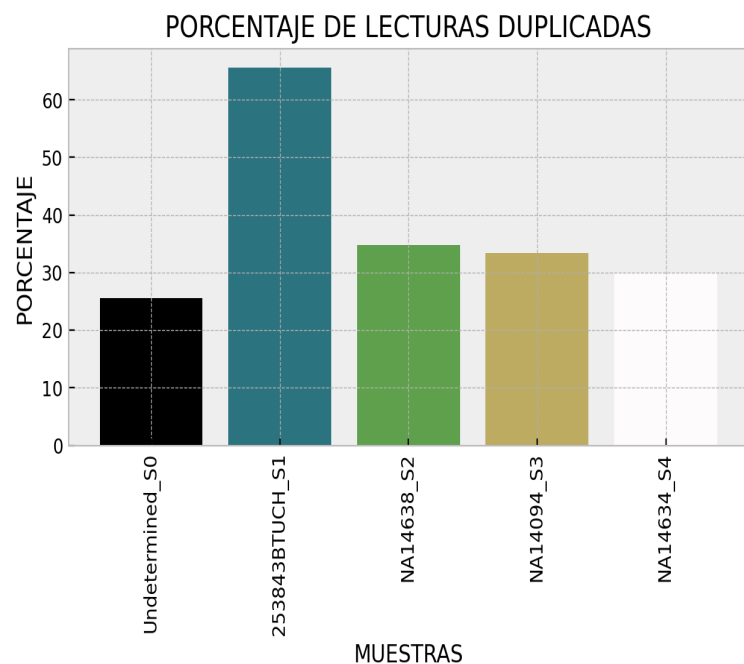


Imagen 9: Porcentaje de lecturas duplicadas, estas son marcadas e ignoradas en el llamado de variantes

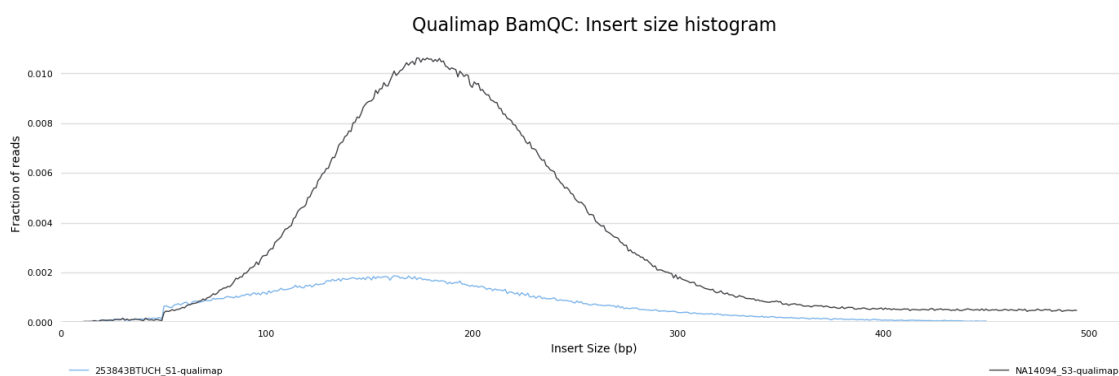


Imagen 10: Histograma de distribución del tamaño del inserto. Métricas obtenidas por qualimap

### 3.6.- Porcentaje de lecturas on-target y profundidad:

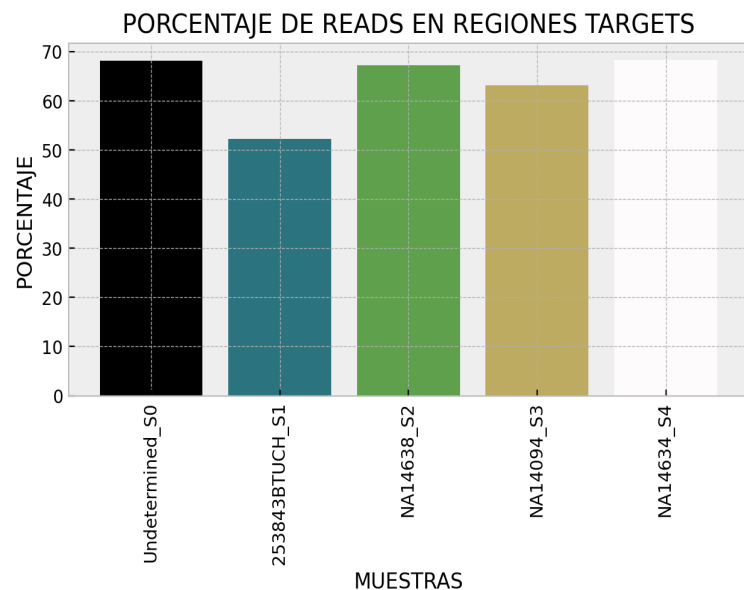


Imagen 11: Porcentaje de lecturas que se encuentran en regiones blancas por muestra

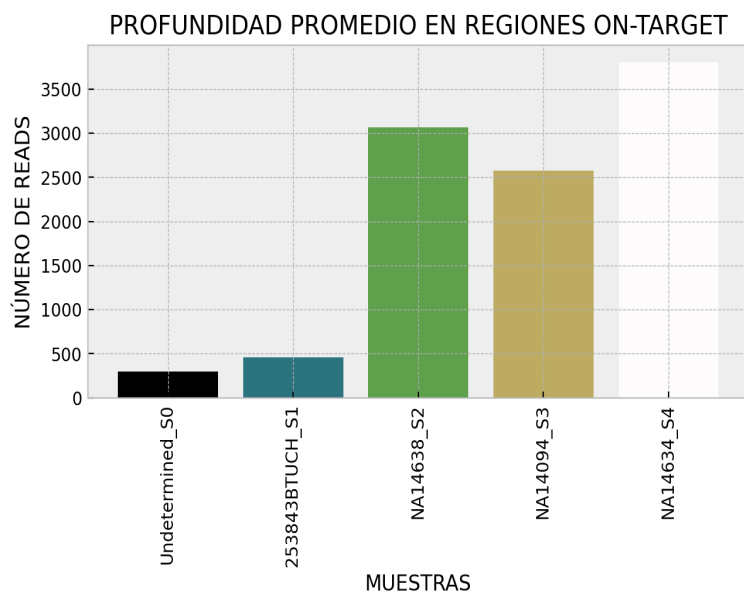


Imagen 12: Promedio de profundidad por muestras en regiones targets



### 3.6.- Uniformidad, intervalos con profundidad definida, bases con profundidad definida:

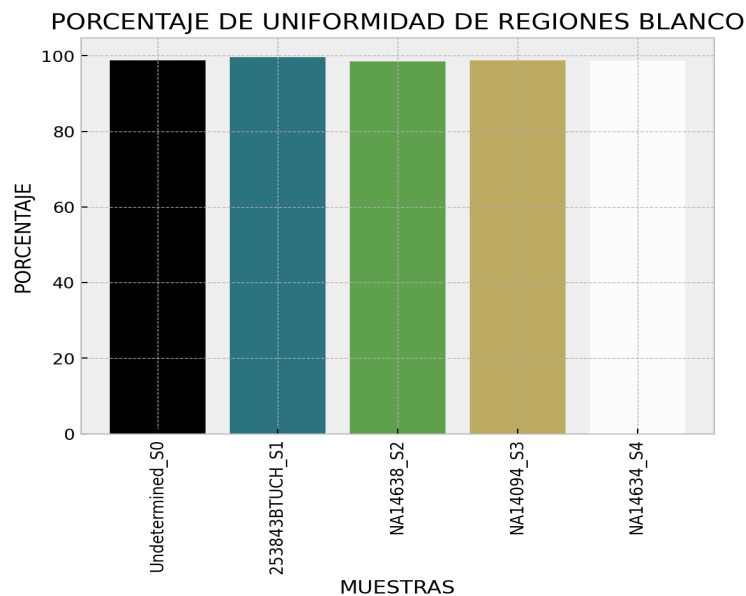


Imagen 13: Uniformidad de cobertura promedio de la corrida

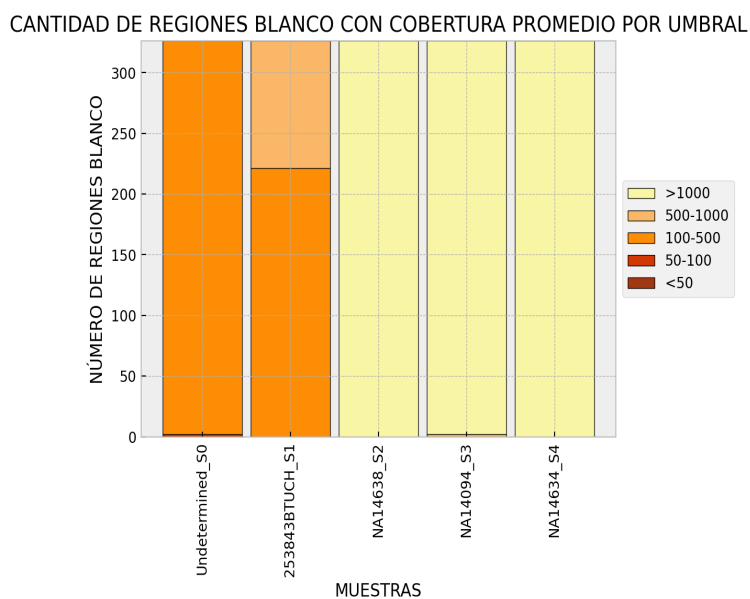


Imagen 14: Cantidad de coberturas promedio por rango para cada muestra

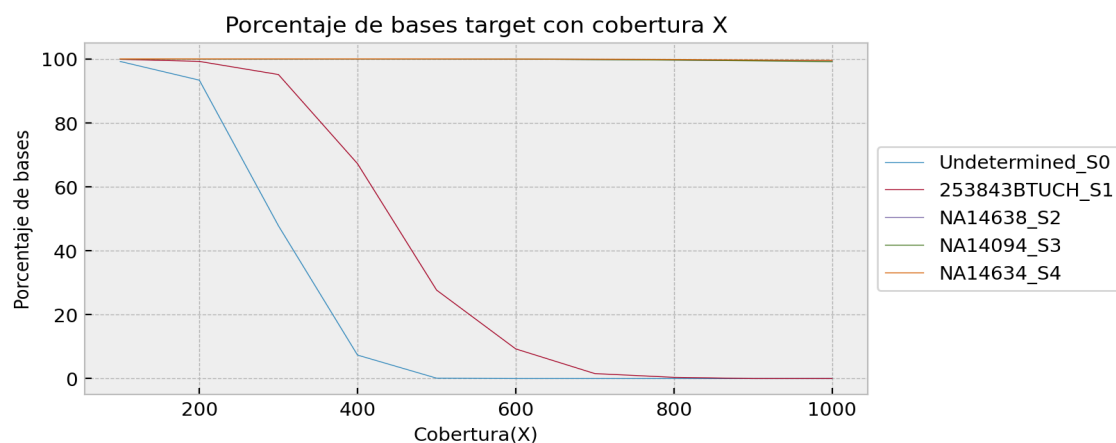


Imagen 15: porcentaje de bases con cobertura en un rango de 100X-1000X

Imagen 17: Porcentaje de bases targets que poseen una cobertura mayor o igual a 300X

