



Curso Teórico-Práctico

EPIDEMIOLOGÍA GENÓMICA



UNA HERRAMIENTA PARA
FORTALECER LA VIGILANCIA DE
AGENTES INFECCIOSOS



Día 2: NGS epidemiología aplicada, consenso y calidad
Bogotá / Septiembre 24 del 2024



Control de calidad de datos de secuenciación masiva y plataformas de análisis (CZ-ID)

Alexander Chivatá Ávila
M.Sc. Ciencias-Bioquímica
Grupo de Investigación en Genómica de
Microorganismos Emergentes
Instituto Nacional de Salud

Objetivos

- ¿Qué es el control de calidad?
- ¿Por qué realizar el control de calidad?
- ¿Qué observamos cuando realizamos el control de calidad?
- ¿Qué herramientas podemos utilizar para el control de calidad y qué observamos?
- ¿Cuáles son los problemas más comunes?
- ¿Cuáles son las mejores prácticas para garantizar la calidad?
- Preguntas

Control de calidad CQ

QC: Análisis esencial para garantizar el éxito de la secuenciación y la confiabilidad de los datos obtenidos.

¿Por qué es importante realizar CQ?

Problemas técnicos:

- Errores en la secuenciación
- Contaminación de las muestras
- Fragmentación inadecuada
- Adaptadores no eliminados
- Lecturas de baja calidad
- Etc.

- Los problemas técnicos no causan fallas en los pipelines de análisis.
- Los problemas técnicos no impiden la generación de hits.
- Los errores técnicos a menudo se ven biológicamente reales.
- Encontrar problemas en el futuro con el trabajo de seguimiento es lento y costoso.
- El control de calidad de los datos de secuenciación sin procesar ahorra tiempo, esfuerzo y dinero.

¿Cuáles son las métricas de calidad?

Densidad de grupos

- Representa el número de clusters de ADN generados por cada superficie de secuenciación.
- Cada cluster corresponde a una secuencia de ADN individual que será leída durante el proceso.

Si la densidad es baja



No se está utilizando todo el potencial del flujo de secuenciación

Si la densidad es muy alta



Los clusters se superponen y la calidad de las lecturas disminuye

¿Cuáles son las métricas de calidad?

Q score (Phred Quality score)

- Indica la probabilidad de que una base llamada sea incorrecta.
- Un Q score más alto significa mayor confianza en la llamada de esa base.

| Puntaje de calidad | Probabilidad de error | Precisión de la inferencia |
|--------------------|-----------------------|----------------------------|
| 10 (Q10) | 1 in 10 = 10% | 90% |
| 20 (Q20) | 1 in 100 = 1% | 99% |
| 30 (Q30) | 1 in 1000 = 0.1% | 99.9% |

Se considera que un **Q30** es ideal para lecturas de alta calidad

¿Cuáles son las métricas de calidad?

Porcentaje de filtros de lecturas aprobados

- Cuántas lecturas pasaron los filtros de calidad iniciales del sistema de secuenciación.

↓
Ej: Longitud de lecturas
Calidad de la imagen

- Las lecturas que no cumplen con ciertos umbrales son descartadas para evitar la inclusión de datos de baja calidad en el análisis final.
- Un alto porcentaje de lecturas que pasaron el filtro (>80%) es un indicador de que la corrida fue exitosa.

¿Cuáles son las métricas de calidad?

Densidad de grupos

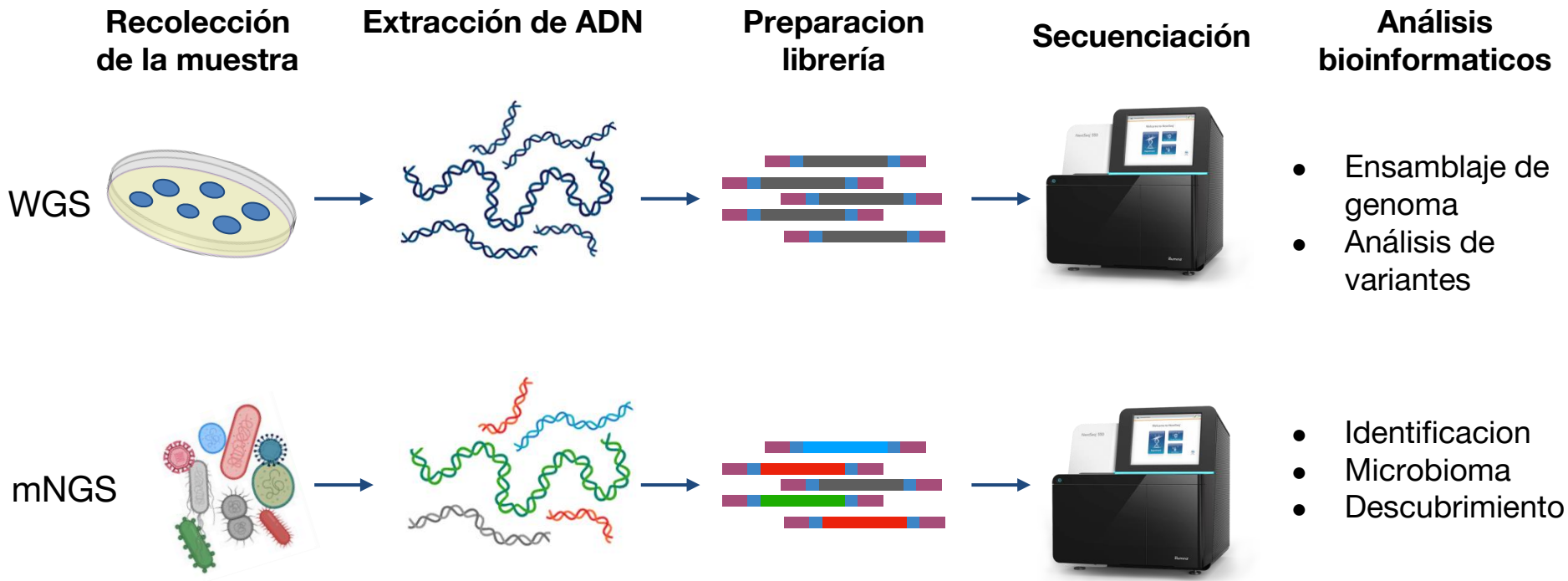
Q Score

Porcentaje de filtros
de lecturas aprobados

¿De dónde provienen estas métricas?



WGS & mNGS resumen del flujo de trabajo



Formato de secuencia: FASTA vs FASTQ



FASTQ


- Secuencias sin procesar
- Formato más común de salida.
- Incluye las secuencias de nucleótidos y la información de calidad de cada base.

Cada lectura está representada en cuatro líneas:

- **Línea 1:** Un identificador de secuencia que comienza con @ y puede incluir detalles sobre la posición del cluster o la lectura.
- **Línea 2:** La secuencia de nucleótidos (A, T, G, C).
- **Línea 3:** Siempre comienza con + y a veces repite el identificador de la primera línea, pero también puede estar vacío.
- **Línea 4:** Los **scores de calidad** ([formato ASCII](#)) que corresponden a cada base de la secuencia en la Línea 2. Estos scores indican la confianza en la identificación de cada base (**Q score**)

Quality Score Encoding

In FASTQ files, quality scores are encoded into a compact form, which uses only 1 byte per this encoding, the quality score is represented as the character with an ASCII code equal to the quality score plus 33. The following table demonstrates the relationship between the encoding character, its ASCII code, and the quality score represented.

 When Q-score binning is in use, the subset of Q-scores applied by the bins is displayed.

| Symbol | ASCII Code | Q-Score |
|--------|------------|---------|
| ! | 33 | 0 |
| " | 34 | 1 |
| # | 35 | 2 |
| \$ | 36 | 3 |
| % | 37 | 4 |
| ° | 38 | 5 |

Formato FASTQ : Secuencias sin procesar

Ejemplo de una secuencia:

```
@NB501961:14:HM7TLBGX2:1:11102:3233:17234 1:N:0:GATCACCA+GATCACCA
CATTCGGCTGGGTTTCGTCACCCTGCGGGAAGATGCGGGTCCAGGCGATAGAGGTGCGGAAGCATTGAAGCCCA
+
AAAAAEEEEEEEEEEEE/EEEEEEEEEEEEEEEEEEEEEEEE<EEEEEEEEEE//EAEEEEEEEAA/EEE/EEAAEEEE<6/
```

1. Identificador con información del experimento de secuenciación
2. Bases A, C, T, G y N.
3. Un separador (+).
4. Puntaje Phred del llamado de cada base ([formato ASCII](#))

| Puntaje de calidad | Probabilidad de error | Precisión de la inferencia |
|--------------------|-----------------------|----------------------------|
| 10 (Q10) | 1 in 10 = 10% | 90% |
| 20 (Q20) | 1 in 100 = 1% | 99% |
| 30 (Q30) | 1 in 1000 = 0.1% | 99.9% |

$$Q = -10\log_{10}(e)$$

Formato de secuencia: FASTA vs FASTQ



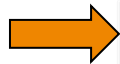
Contiene la secuencia de nucleótidos (o aminoácidos) y un identificador de secuencia.

Formato:

- **Línea 1:** Comienza con > seguido del nombre o identificador de la secuencia.
- **Línea 2 y posteriores:** Contienen la secuencia en sí, representada en nucleótidos.

Identificador

Secuencia



>Nucleotide Sequence Name

ATTGGCTAATTGGCTAATTGGCTAATTGGCTAATTGGCTAATTGG
CTAATTGGCTAATTGGCTAATTGGCTAATTGGCTAATTGGCTAATT
GGCTAATTGGCTAATTGGCTAATTGGCTAATTGGCTAATTGGCT

Cuántas secuencias hay?

File: Sample_R1.fastq

@NB501961:14:HM7TLBGX2:1:11102:3233:17234 1:N:0:GATCACCA+GATCACCA

CATTCGGCTGGGTTTCGTCAACCCTGCGGGAAGATGCGGGTCCAGGCGATAGAGGTGCGGAAGCATTTGAAGCCCATCTCGGCGATCAGTTTGATGTCTTCTTT
GTAGCGACCGTAGAAGTCGACGGCTTCGTGGTTCGGGTAGTATTTTN

+

AAAAEEEEEEEEEE/EEEEEEEEEEEEEEEEEE<EEEEEEEEEE//EAEEEEEAA/EEE/EEAEEEE<6/<AE<E<A//EEEEAE<EA/<AA//EAAEAA<E/EAE/
//E<A/<AA//A<A<<A6<A//AA/E///6A#

@NB501961:14:HM7TLBGX2:1:11103:13923:13163 1:N:0:GATCACCA+GATCACCA

GAAGTGGAGCTTTAGCCCCGGTGATCCTGTGCCTCATTTGTATGGCTACCGCCATGGTGATGTACCTGCCATTTCTCAAAGCTTACGAAAAACAGCTGCTGGCCC
AGGAGCGTGAAAAACGCCGTGCGCCAGGCTGACAACGCGGCGCAGAN

+

AAAAAEAAAAAAAAAEEEEEEEEEEEEE/EEEEAAEAEEAEEAEEEEEEEEEEEAEEEEEEEEEEEA<EEEEEEAE<A
EAEAAA<AAEEAEEAE/6/A/EEEE/EEE<AE/#

@NB501961:14:HM7TLBGX2:1:11104:15318:11334 1:N:0:GATCACCA+GATCACCA

GATAGTGCTGTTGCGGAGCGATCGCCAGTACGCGCTCGCGCTCTCTGCGCCACGCGCAGCTCTTTACCACGACCGTTGAGTACCAGGCTGGCGGTGCGTTT
GGAGACGCCCCGCCAGTTTCGGCGATATCTTTGATAGTTACGCGTTTGG

+

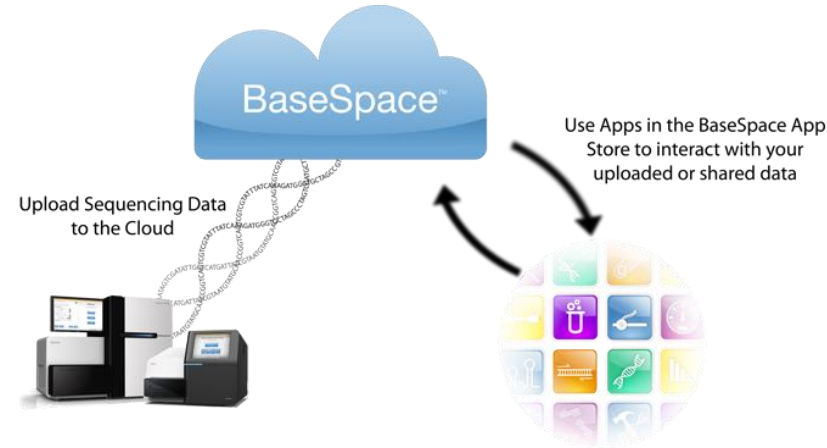
AAAAAEFFFFFFFAEEAEFFFFFFEEAEAE<EAEE/EE/EEE<AEFFEE/EEEEEEEEEEEE/EEE//AE<E/EEEEAEAA/
E<<AEFFEE/<66A///<</AA<//6AE<<6/<6

Configuración y control de calidad de una ejecución de secuenciación (Illumina)

1. Crear hoja de muestra
 - a. Utilice el Administrador de experimentos de Illumina
2. Configuración Ejecutar secuenciación y cargar hoja de muestra ([MiSeq](#) & [Iseq100](#) instrucciones)
3. Ejecución de secuenciación de control de calidad
 - a. BaseSpace (solo si el secuenciador tiene conexión a Internet)
 - b. Visor de análisis de secuencias (se ejecuta localmente, no requiere Internet, solo Windows)

Cómo controlar la calidad de una secuenciación en BaseSpace

- BaseSpace es un software para:
 - Evaluar las métricas de control de calidad de la ejecución de secuenciación
 - Almacenar datos
 - Analizar datos
 - Compartir datos
- Regístrate para obtener un estándar gratis Cuenta BaseSpace [aquí](#)
 - Almacenamiento de datos limitado
 - Monitoreo de ejecución de secuenciación en línea y en tiempo real
 - Control de calidad de ejecución de secuenciación en línea

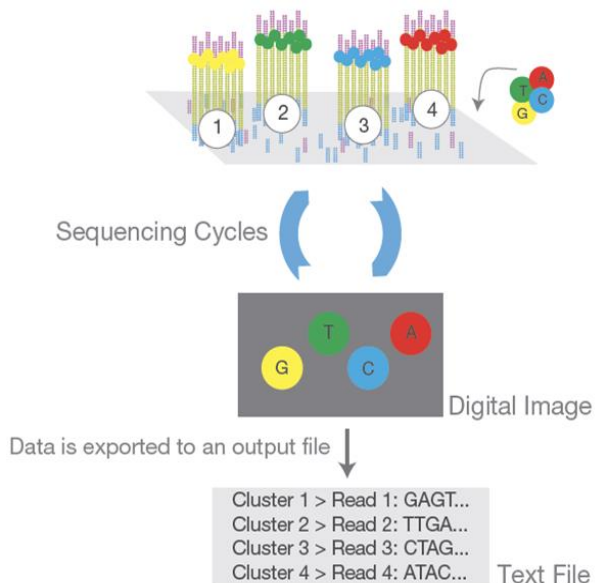


Cómo controlar la calidad de una secuenciación en BaseSpace

- ¿Los datos de secuenciación son de buena calidad? (%PF, avg%Q30)
- ¿La corrida estuvo sobreagrupada o subagrupada? (%Ocupado)
- ¿La ejecución funcionó como se esperaba? (% base por ciclo, % PhiX, % tasa de error)
- ¿La Librería que cargué era de buena calidad?
 - Dímeros adaptadores (% base por ciclo)
 - Longitud (%base por ciclo,%PhiX)
 - Agrupamiento (% de lecturas identificadas por muestra)
 - Concentración (%Ocupado)
- ¿Cuántas lecturas generó la ejecución de secuenciación? (Lecturas PF)
- ¿Se realizó correctamente la demultiplexación? (% indeterminado, % de lecturas identificadas por muestra)

¿Qué es el sobreagrupamiento/subagrupamiento?

Un grupo es un grupo clonal de un fragmento de biblioteca generado en el primer paso de la secuenciación. Lo ideal sería que cada grupo genere una lectura (de extremo emparejado).

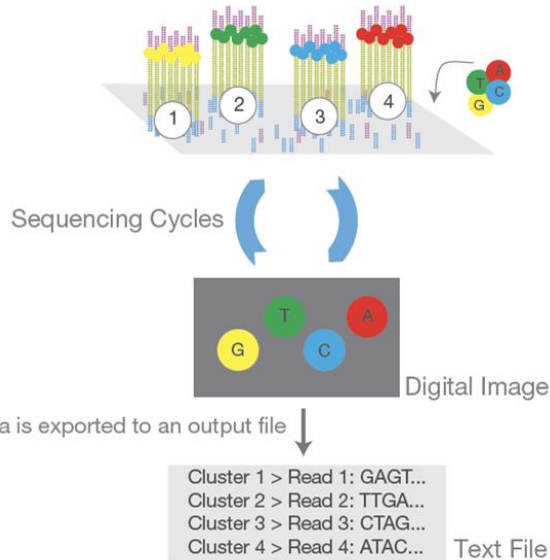


Subagrupamiento

Se cargó una cantidad baja de material en la celda de flujo.

- Esto reduce la cantidad de agrupamientos y, por lo tanto, también las lecturas que generará la ejecución de secuenciación.
- El subagrupamiento generalmente no afecta a Q30.

¿Qué es el sobreagrupamiento/subagrupamiento?



Sobreagrupamiento

se cargó demasiado material en la celda de flujo, lo que hace que los grupos resultantes estén demasiado cerca entre sí.

- Esta alta señal de fondo reducirá la calidad de la llamada de base.
- El secuenciador no puede distinguir correctamente un grupo de otro, lo que genera señales mixtas y muchas lecturas que no superan el filtro QC inicial.

¡La sobreagrupación es peor que la subagrupación!

Cómo controlar la calidad de una secuenciación en BaseSpace

- **PROMEDIO %Q30**= porcentaje de lecturas que tienen bases con un puntaje de calidad promedio >Q30
 - Lo ideal es un AVG%Q30 >90%, pero >70% también es bueno
 - Si es <70%: evalúa por qué la calidad es inferior a la esperada. Posibles motivos:
 - Sobreagrupamiento: se cargó demasiado material en la celda de flujo
 - Dímeros adaptadores: estos producirán lecturas con solo ~70 pb de buena señal y el resto, "nada" de baja calidad.
- **%PF**= porcentaje de lecturas de filtro aprobadas
 - Lo ideal es >70%, pero >50% también es bueno
 - Si es <50%, verifique si hay indicios de agrupamiento excesivo o insuficiente.

Cómo controlar la calidad de una secuenciación en BaseSpace

Run: COVID-19 ARTIC_Batch-10: Charts

SUMMARY BIOSAMPLES **CHARTS** METRICS INDEXING QC SAMPLE SHEET FILES



Flow Cell: BPL20321-2824 Extracted: 318 Called: 318 Scored: 318

Flow Cell Chart

Chart

Intensity

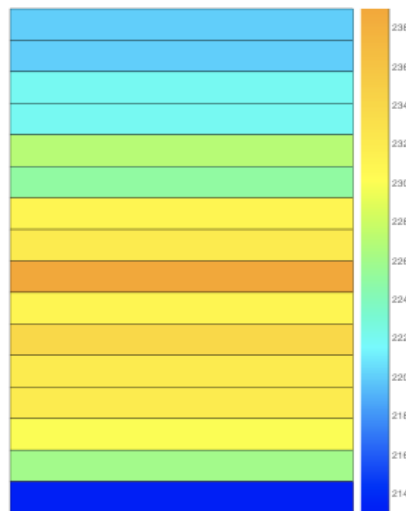
Cycle

Cycle 36

Channel

1

☐ Fix Scale



Data By Cycle

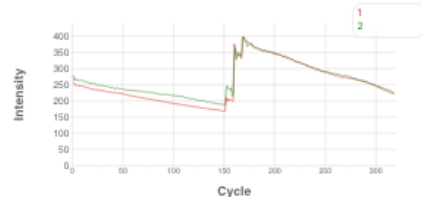
Chart

Intensity

Channel

All Channels

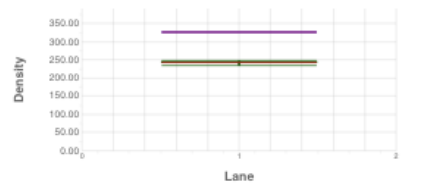
☐ Fix Scale



Data By Lane

Show

Density



QScore Distribution

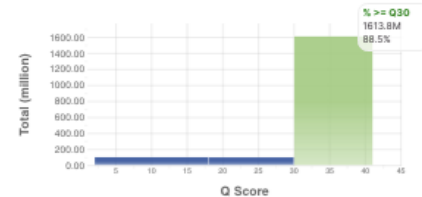
Read

All Reads

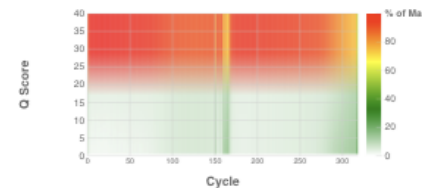
Cycle

All Cycles

☐ Fix Scale



QScore Heatmap



Cómo controlar la calidad de una secuenciación en BaseSpace

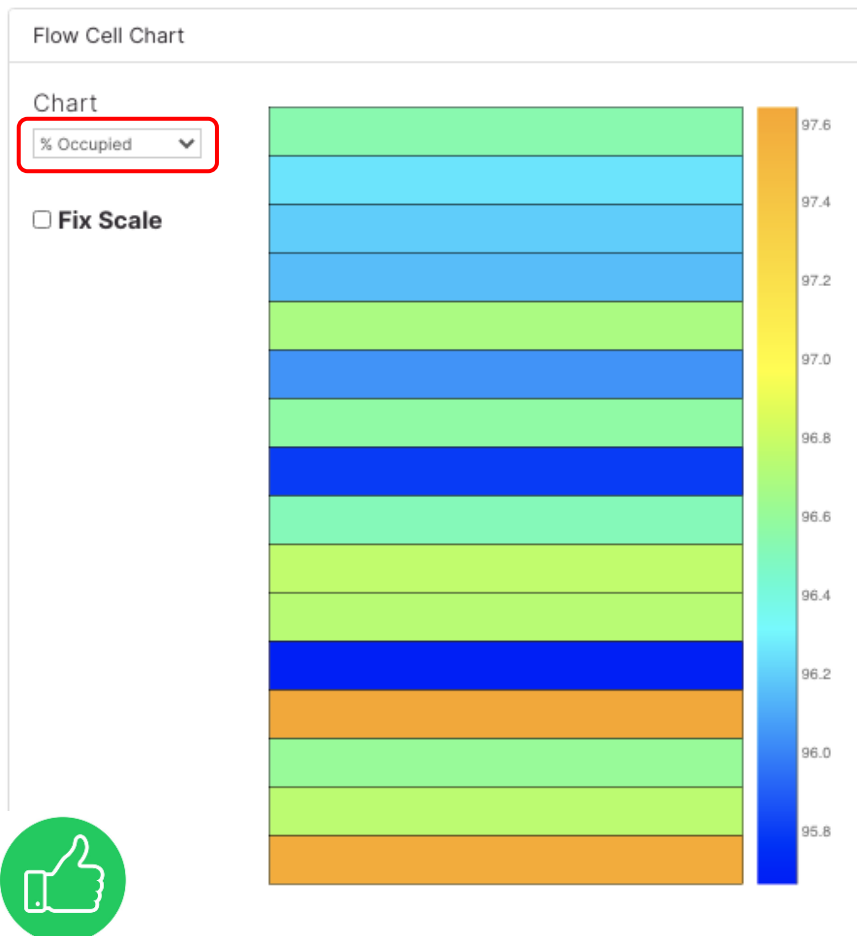


Gráfico de celdas de flujo: seleccione % ocupado en 'Gráfico'

-> ¿Signos de sobreagrupamiento o subagrupamiento?

- El mejor rango de valores es 90-98%.
- >98 %: posible sobreagrupamiento, lo que reducirá los puntajes %PF y AVG%Q30.



Revise los datos de cuantificación de la biblioteca para verificar si hay errores o considere reducir la cantidad de carga para las próximas ejecuciones de secuenciación.

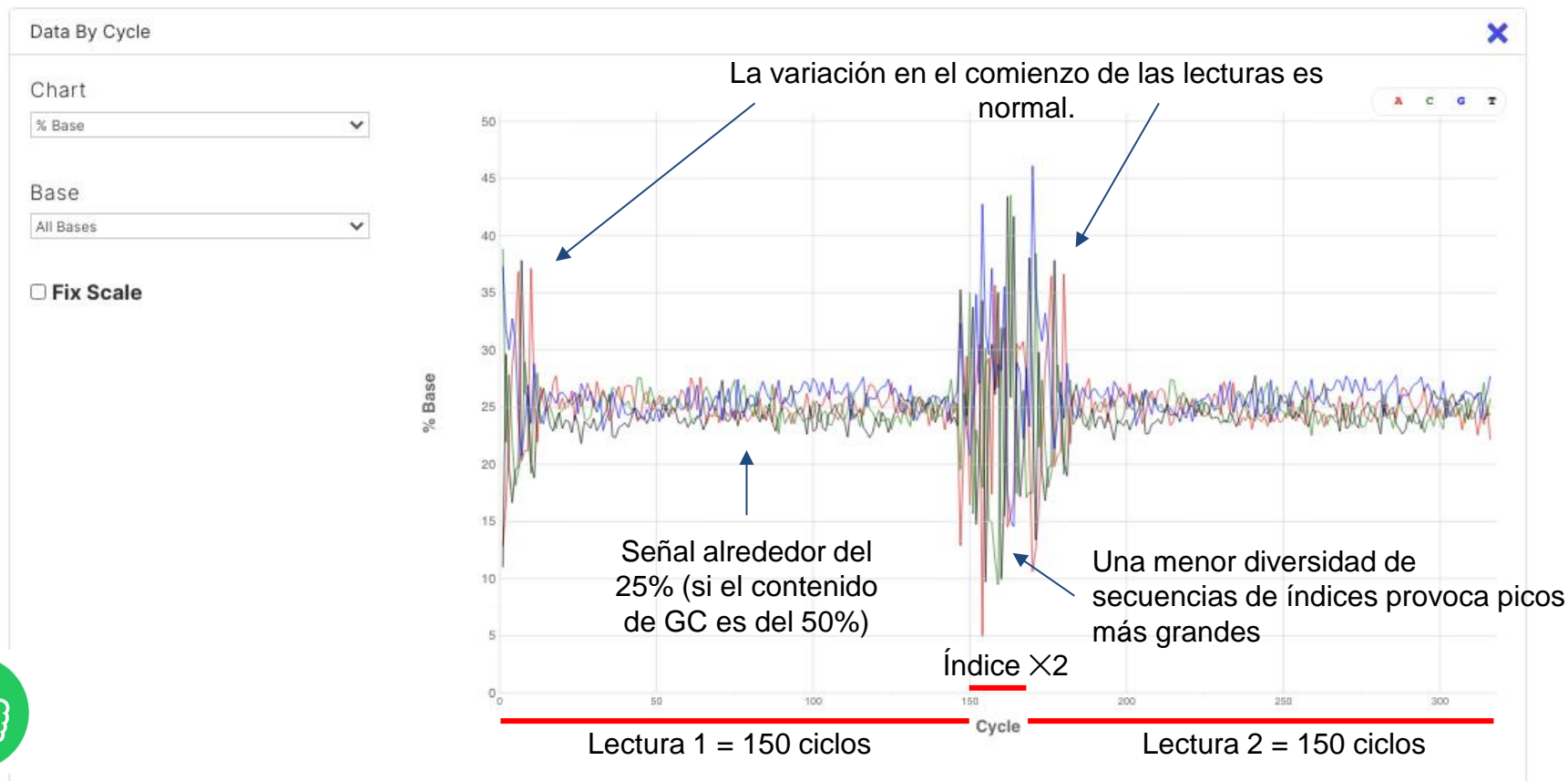
- <90%: posible subagrupamiento



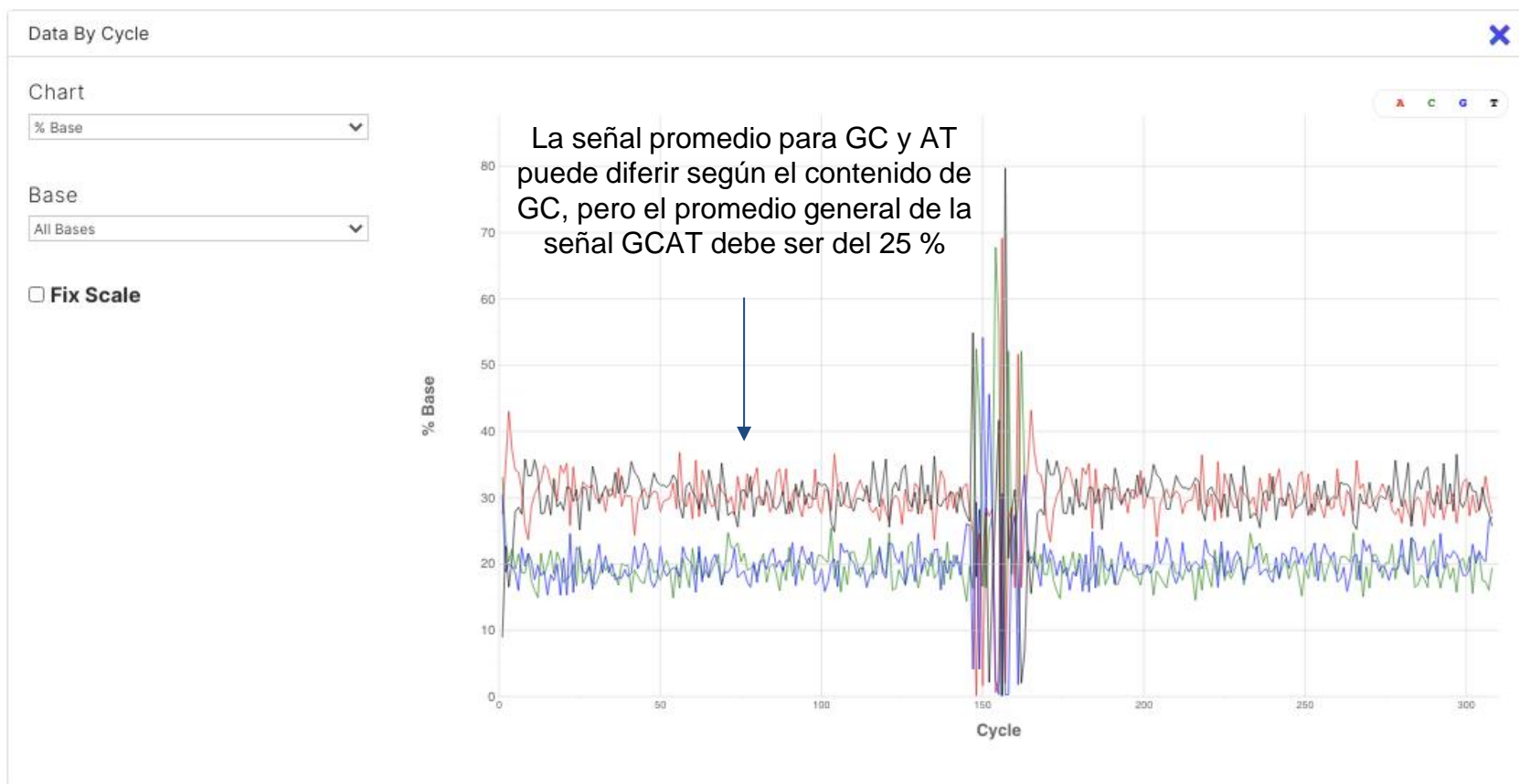
Revise los datos de cuantificación de la biblioteca para verificar si hay errores o considere aumentar la cantidad de carga para las próximas ejecuciones de secuenciación para obtener más lecturas de una ejecución.

Cómo controlar la calidad de una secuenciación en BaseSpace

Gráfico de datos por ciclo: seleccione % Base en "Gráfico" -> ¿Dímeros adaptadores? ¿Fragmentos cortos?



Cómo controlar la calidad de una secuenciación en BaseSpace



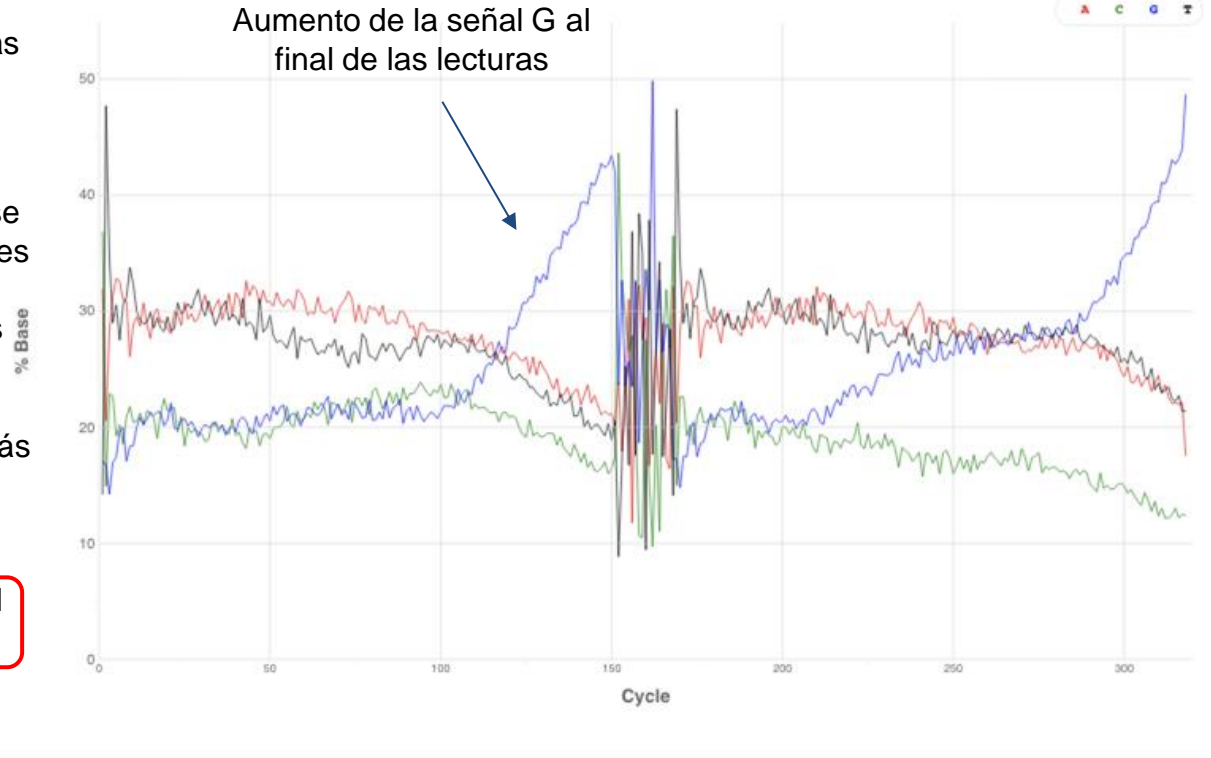
Cómo controlar la calidad de una secuenciación en BaseSpace

Número 1 - Lecturas cortas

Aumento de la señal G al final de las lecturas

- G es la “base oscura” en la química iSeq/MiSeq
 - ¿Insertos más cortos que 150 pb?
 - ¿Problema con el alargamiento de las hebras durante el agrupamiento?
- Si el tamaño de inserción fue de hecho más corto de lo habitual (verifique el control de calidad de preparación de la Librería)

¡reduzca el tiempo de fragmentación para el futuro!



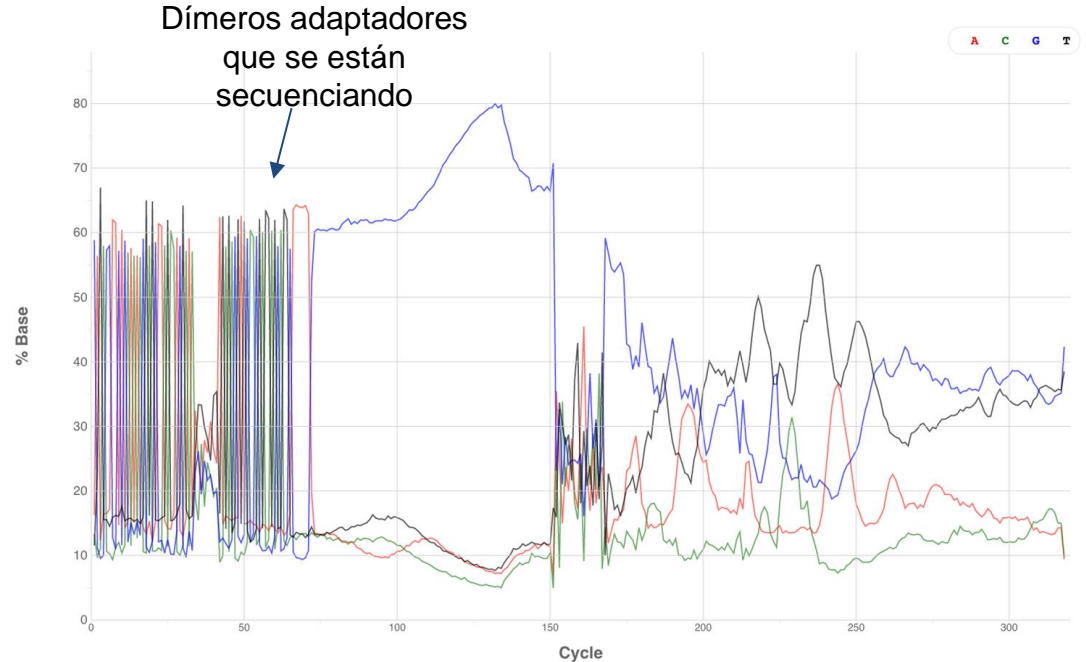
Cómo controlar la calidad de una secuenciación en BaseSpace

Número 2: dímeros de adaptadores

Picos grandes durante los primeros 70 ciclos y luego un aumento en la señal G

- Los dímeros adaptadores producirán lecturas con baja variación de secuencia (picos grandes) que tienen una longitud de aproximadamente 70 pb.
- Los dímeros adaptadores son fragmentos cortos y, por lo tanto, se secuencian preferentemente, ocupando un espacio de secuenciación valioso.

¡Es crucial eliminar TODOS los dímeros adaptadores durante el control de calidad de preparación de la biblioteca!



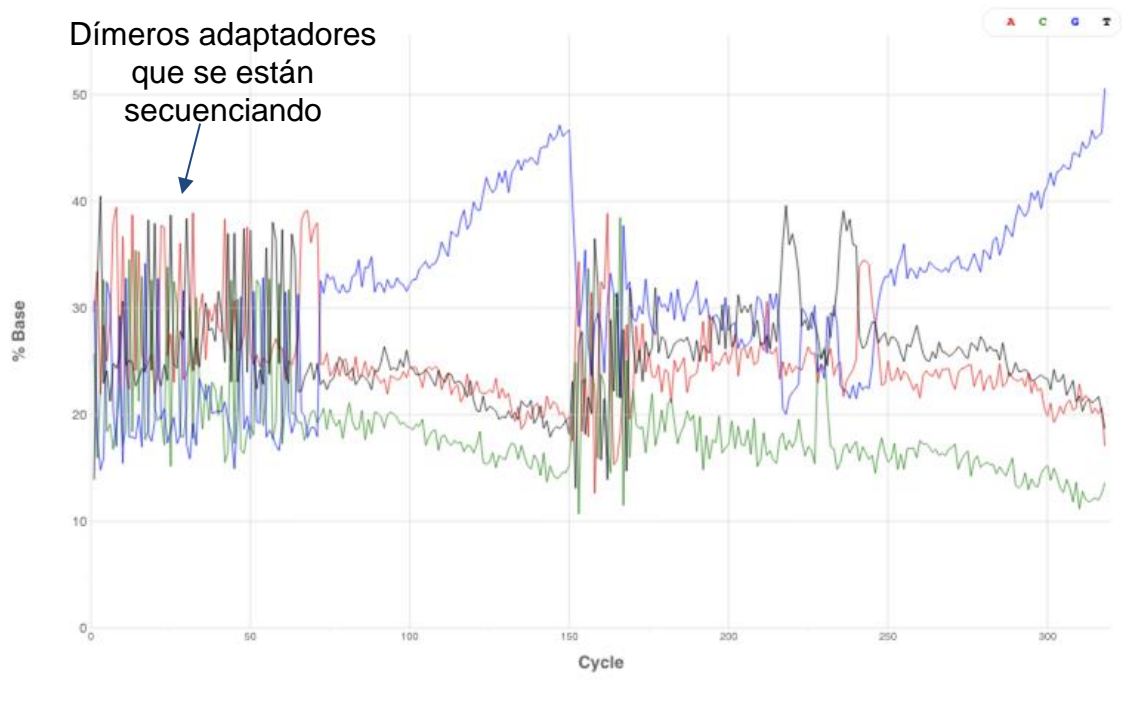
Cómo controlar la calidad de una secuenciación en BaseSpace

Número 3: Dímeros adaptadores y lecturas cortas

Picos grandes durante los primeros 70 ciclos, seguidos de una señal más estable y luego un aumento constante en la señal G

- Los dímeros adaptadores producirán lecturas con baja variación de secuencia (picos grandes) que tienen una longitud de aproximadamente 70 pb.
- Los dímeros adaptadores son fragmentos cortos y, por lo tanto, se secuencian preferentemente, ocupando un espacio de secuenciación valioso.

¡Es crucial eliminar TODOS los dímeros adaptadores durante el control de calidad de preparación de la biblioteca!



Cómo controlar la calidad de una secuenciación en BaseSpace

En la página 'Métricas'(si usa PhiX) → Control de calidad (bacteriofago Phix174)

- Métricas por lectura:
 - a. **% alineado**(PhiX): debe estar cerca del % de PhiX que agregaste.
 - i. Si es mayor, PhiX superó a su librería , lo que significa que colocó menos librería de lo que pensaba (verifique dos veces sus cálculos de agrupamiento y cuantificación).
 - ii. Si es menor de lo esperado, es posible que fragmentos pequeños como dímeros adaptadores hayan superado a PhiX Y/O usted subestimó la cantidad de material en su muestra (agregó mucho más de lo que pensaba, verifique dos veces sus datos de cuantificación).
 - b. **Tasa de error** %(PhiX): debe ser menor al 1%.

Cómo controlar la calidad de una secuenciación en BaseSpace

COVID-19 ARTIC_Batch-10

SUMMARY BIOSAMPLES CHARTS **METRICS** INDEXING QC SAMPLE SHEET FILES



Per Read Metrics

| READ | CYCLES | YIELD | PROJECTED YIELD | ALIGNED (%) | ERROR RATE (%) | INTENSITY CYCLE 1 | %>Q30 |
|-----------------------|--------|------------|-----------------|-------------|----------------|-------------------|-------|
| Read 1 | 151 | 860.56 Mbp | 860.56 Mbp | 2.72 | 0.45 | 259.25 | 90.79 |
| Read 2 (I) | 8 | 40.16 Mbp | 40.16 Mbp | 0.00 | 0.00 | 210.56 | 87.55 |
| Read 3 (I) | 8 | 40.16 Mbp | 40.16 Mbp | 0.00 | 0.00 | 376.44 | 74.09 |
| Read 4 | 151 | 860.56 Mbp | 860.56 Mbp | 2.43 | 1.09 | 390.75 | 87.34 |
| Non-index Reads Total | 302 | 1.72 Gbp | 1.72 Gbp | 2.58 | 0.77 | 325.00 | 89.06 |
| Total | 318 | 1.80 Gbp | 1.80 Gbp | 2.58 | 0.77 | 309.25 | 88.69 |

Métricas por carril:

a. Lecturas PF: número de lecturas de filtro aprobadas que se generaron en esta ejecución.

Este número debe ser cercano (o superior) al número mínimo de lecturas garantizadas que se anuncia que genera su secuenciador.

Si es menor, probablemente haya realizado una agrupación insuficiente o excesiva. Verifique sus cálculos de cuantificación y ajuste la concentración de carga si es necesario.

| Secuenciador | Lecturas máximas (millones) | Longitud máxima de lectura (pb) |
|-----------------------------|-----------------------------------|---------------------------------------|
| MiSeq | 22 - 25 | 2 x 300 |
| Microsecuenciación MiSeq | 4 | 2x150 |
| iSeq 100 | 4 | 2x150 |
| HiSeq 4000 | 250 - 400 | 2x150 |
| NovaSeq 6000 SP | 325 - 400 | 2x250 |
| NovaSeq 6000 S1 | 750 - 800 | 2x150 |
| NovaSeq 6000 S2 | 1650 - 2050 | 2x150 |
| NovaSeq 6000 S4 | 2000 - 2500 | 2x150 |

COVID-19 ARTIC_Batch-10

SUMMARY BIOSAMPLES CHARTS **METRICS** INDEXING QC SAMPLE SHEET FILES



Desplácese hacia abajo hasta 'Métricas por carril'

Número de lecturas de extremos
emparejados que se generaron

Per Lane Metrics

| LANE | STATUS | READ | CLUSTER PF(%) | %≥ Q30 | YIELD | ERROR RATE(%) | READS PF | DENSITY | TILES | LEGACY PHAS / PREPHAS(%) | COMMENTS | INTENSITY |
|----------------------------|------------------|------------|---------------|--------|----------|-------------------|------------------|---------|-------|--------------------------|----------|-----------|
| <input type="checkbox"/> 1 | <u>QC Passed</u> | Read 1 | 74.62±1.18 | 90.79 | 0.86 Gbp | <u>0.45 ±0.04</u> | <u>5,737,044</u> | 326 ±0 | 16 | 0.365 / 0.227 | | 259±8 |
| | | Read 2 (I) | | 87.55 | 0.04 Gbp | <u>0.00 ±0.00</u> | | | | 0.000 / 0.000 | | 211±6 |
| | | Read 3 (I) | | 74.09 | 0.04 Gbp | <u>0.00 ±0.00</u> | | | | 0.000 / 0.000 | | 376±18 |
| | | Read 4 | | 87.34 | 0.86 Gbp | <u>1.09 ±0.06</u> | | | | 0.269 / 0.127 | | 391±19 |

Cómo controlar la calidad de una secuenciación en BaseSpace

En la página 'Control de calidad de indexación':

- a. **% lecturas identificadas**= lecturas con códigos de barras identificados: **debe ser >80 %** (cuanto más alto, mejor).

Si este porcentaje es bajo, debe verificar que su hoja de muestra haya asignado las secuencias de índice/código de barras correctas para cada muestra.

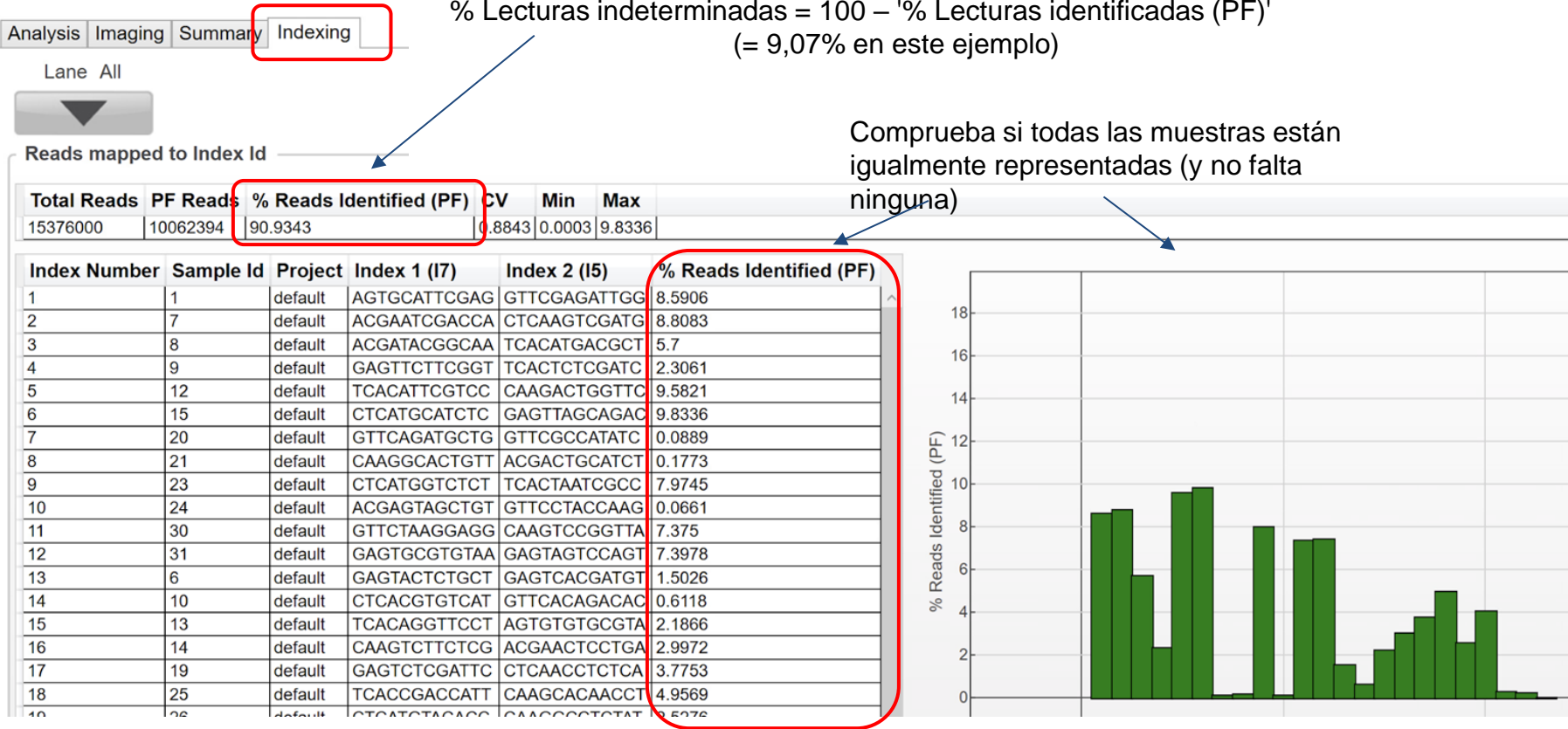
- a. **% lecturas indeterminadas**= lecturas sin un código de barras conocido asociado.

Si este porcentaje es alto, debe verificar que la hoja de muestra sea correcta.

- a. **PF lectura** : este número detalla la cantidad de lecturas de extremo único (no emparejadas). Este número siempre es el doble de la cantidad real de lecturas de extremo emparejado que se muestra en la página "Métricas".

Cómo controlar la calidad de una ejecución de secuenciación en Sequence Analysis Viewer

Haga clic en la pestaña "Indexación" para ver las métricas de demultiplexación.



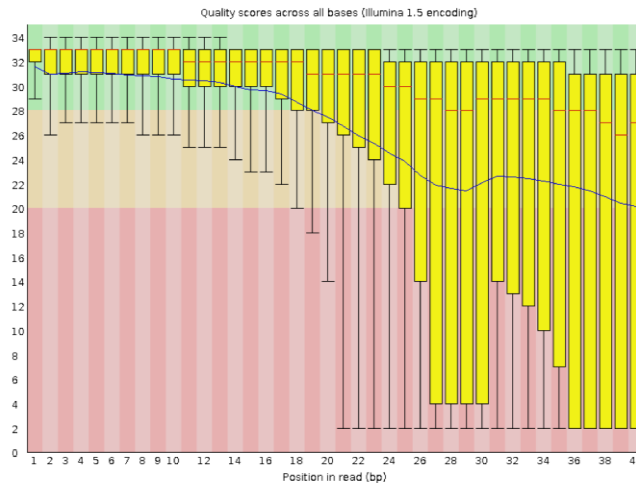
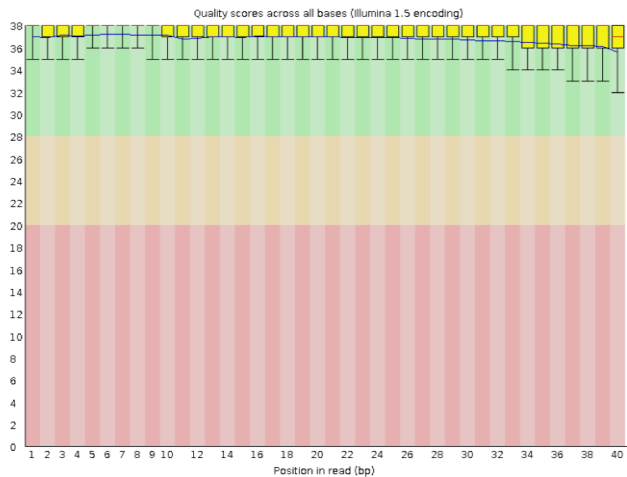
Analisis de calidad por FASTQC



Lee archivos fastq sin procesar

Reporte HTML

Calidad del llamado de bases (Phred score)



Trimming
(trimomatic,
cutadapt,etc)

Limpieza de
adaptadores

Lineas de comando/
Lenguaje de programación

- la calidad debería ser buena para la mayoría de las lecturas a lo largo de toda la secuencia
- Si la calidad se deteriora, debemos entender cómo y por qué
- Buena calidad es generalmente Phred > 28 - 30
- Calidad preocupante es Phred < 20

CZID



Gratis



Open source



En la nube



Amigable para
quienes no manejan
línea de comando

Identificación de
patógenos en datos de
metagenómica

Ejemplo: Determinar la causa de un
brote febril sin explicación etiológica

Generacion de
genomas consenso

Ejemplo: Secuenciamiento de amplicones
de COVID-19

CZID para control de calidad

Total Reads: Secuencias totales subidas al pipeline.

Passed QC: Secuencias que han superado los filtros de calidad (fastp).

Passed Filters: Secuencias utilizadas para el análisis (que han superado el control de calidad, el filtrado de hospedero/humano y el submuestreo).

Duplicate Compression Ratio (DCR): Relación entre las secuencias presentes antes de la eliminación de duplicados y después de la eliminación de duplicados.

ERCC Reads: Número total de lecturas que corresponden a controles internos ERCC.

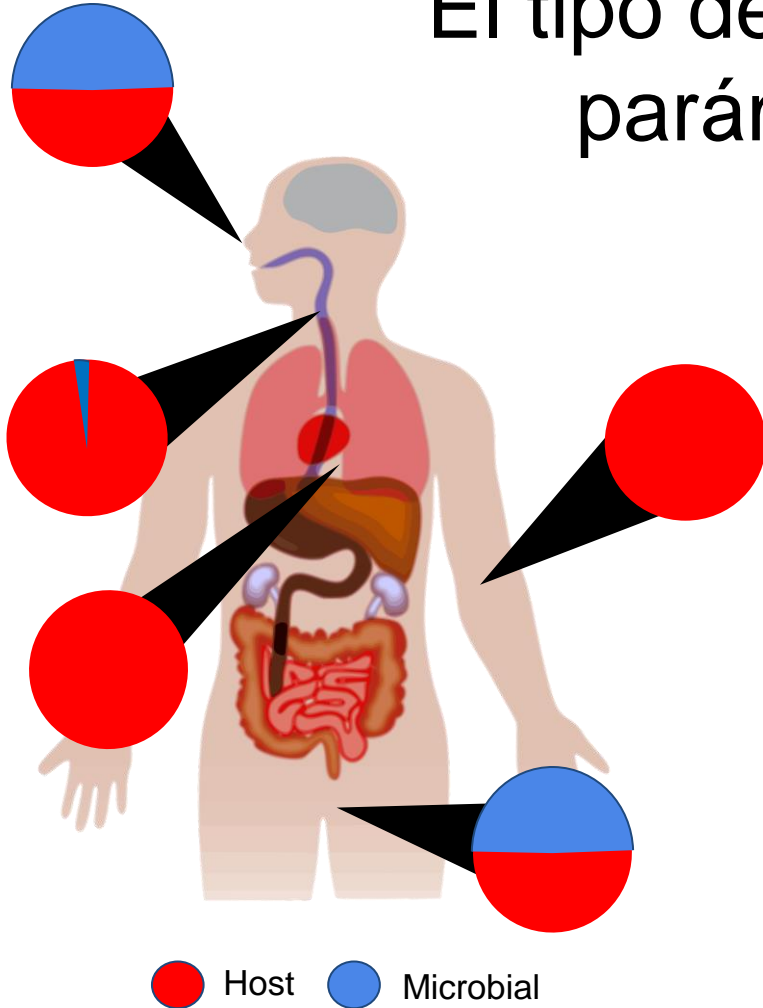
Mean insert size: Longitud media de la secuencia de nucleótidos existente entre los adaptadores .

Pipeline para mNGS utilizado por CZID para secuencias Illumina

| Function | Upgraded tool |
|-----------------------------|-----------------|
| Adapter trimming | fastp |
| Low quality read removal | PriceSeq |
| Low complexity read removal | LZW |
| ERCC read removal | Bowtie2 |
| Gather host gene counts | Kallisto |
| Host read removal | Bowtie2, HISAT2 |
| Human read removal | Bowtie2, HISAT2 |



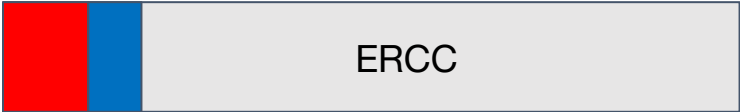
CZID: cuenta con Pipeline para nanopore

El tipo de muestra impacta los parámetros de calidad



| Tipo de muestra | Cantidad de huesped | Cantidad de microorganismos |
|-------------------------|---------------------|-----------------------------|
| Líquido ceforraquideo | > 99% | << 5% with infection |
| Aspirado bronquial | 95% | 5% |
| Isopo nasofaríngeo | 50% | 50% |
| Sangre | > 99% | << 1% (with infection) |
| Isopo Rectal or Vaginal | 50% | 50% |

Tendencias de calidad según la muestra

| Tipo de muestra | Distribución típica  | Tendencia en el control de calidad |
|--|--|---|
| <p>Isopos, muestras ambientales, orina</p> <p>Gran diversidad y material microbiano</p> |  | <ul style="list-style-type: none"> • Más secuencias de microorganismos • Menos DCR |
| <p>Suero, plasma, liquido cefalorraquideo</p> <p>Menor diversidad y contenido microbiano</p> |  | <ul style="list-style-type: none"> • Más secuencias de microorganismos • DCR más altos • Mas “ruido” |

ERCC Reads: Número total de lecturas que corresponden a controles internos ERCC.

Ejercicio Práctico

¿Hay suficientes lecturas totales?

¿Mis muestras tienen suficientes lecturas de alta calidad?

¿Mis muestras tienen suficiente diversidad de secuenciación?

¿Mis muestras tienen longitudes de inserción suficientes?

¿Cómo se procesaron mis muestras a través del pipeline?

Gracias

Contacto:

jchivata@ins.gov.co

jchivata@unal.edu.co



Algunas de las diapositivas usadas en esta presentación hacen parte del material disponible en Help Center de CZID

<https://chanzuckerberg.zendesk.com/hc/en-us>

QC Icon

ic

ID

My Data

Public

Upload

User

Medical Detectives v2

Public project 6 members

Share

Add Data

Search Public...

Samples 17

Taxon Filters

TAXON

ANNOTATION

Metadata Filters

LOCATION

TIMEFRAME

VISIBILITY

HOST

SAMPLE TYPE

Metagenomics 17

Metagenomics - Nanopore 0

Consensus Genomes 2

Antimicrobial Resistance 17

il

0 Selected

17 samples

| <input type="checkbox"/> | Sample | Created On | Host | Passed Filters | Passed QC | Total Reads | DCR | Water Contr... |
|--------------------------|--|----------------------------|-------|--------------------|-----------|-------------|------|----------------|
| <input type="checkbox"/> | <div><div>Patient 016 (CSF) COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 3,652,560 2.44% | 53.18% | 150,000,000 | 1.95 | No |
| <input type="checkbox"/> | <div><div>H2O_1 COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 1,670 43.02% | 43.64% | 3,882 | 1.05 | Yes |
| <input type="checkbox"/> | <div><div>H2O_2 COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 4,894 15.92% | 16.64% | 30,742 | 1.04 | Yes |
| <input type="checkbox"/> | <div><div>Patient 015 (CSF) COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 51,794 0.26% | 89.45% | 19,642,962 | 4.72 | No |
| <input type="checkbox"/> | <div><div>Patient 014 (Eye) COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 53,376 1.58% | 81.94% | 3,368,440 | 2.08 | No |
| <input type="checkbox"/> | <div><div>Patient 013 (CSF) COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 1,378,654 2.51% | 66.27% | 54,869,610 | 3.41 | No |
| <input type="checkbox"/> | <div><div>Patient 012 (CSF) COMPLETE</div><div>User Medical Detectives v2</div></div> | 2024-02-06 4 months ago | Human | 1,004,490 1.43% | 76.79% | 70,250,230 | 4.17 | No |

DESCRIPTION

EDIT

No description.

OVERALL

Samples 17

Projects 1

Avg. reads per sample 38,907,162

Avg. reads passing filters per sample 707,266

DATE CREATED

2024-02-06

METADATA

Host

Human 17

Sample Type

Cerebrospinal 14

Medical Detectives v2

Public project

6 members

Share

Add Data



Search My Data...

Samples 17



Taxon Filters

TAXON

ANNOTATION

Metadata Filters

LOCATION

TIMEFRAME

VISIBILITY

HOST

SAMPLE TYPE

Metagenomics 17

Metagenomics - Nanopore 0

Consensus Genomes 2

Antimicrobial Resistance 17



0 Selected

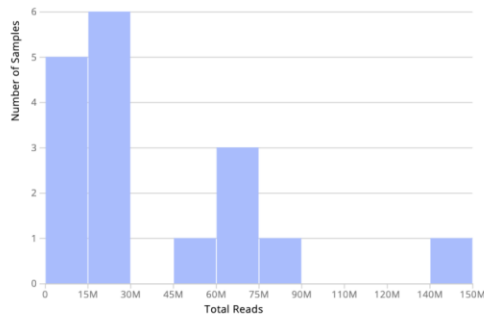


17 samples

Showing 17 of 17 samples.

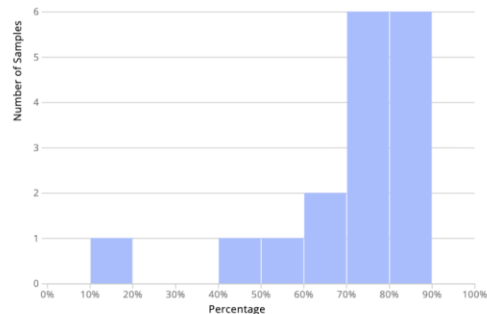
Do my samples have enough total reads?

Total Reads



Do my samples have enough quality reads?

Passed QC



Summary Samples 17

OVERALL

Samples 17

Projects 1

Avg. reads per sample 38,907,162

Avg. reads passing filters per sample 707,266

DATE CREATED

2024-02-06

METADATA

Host

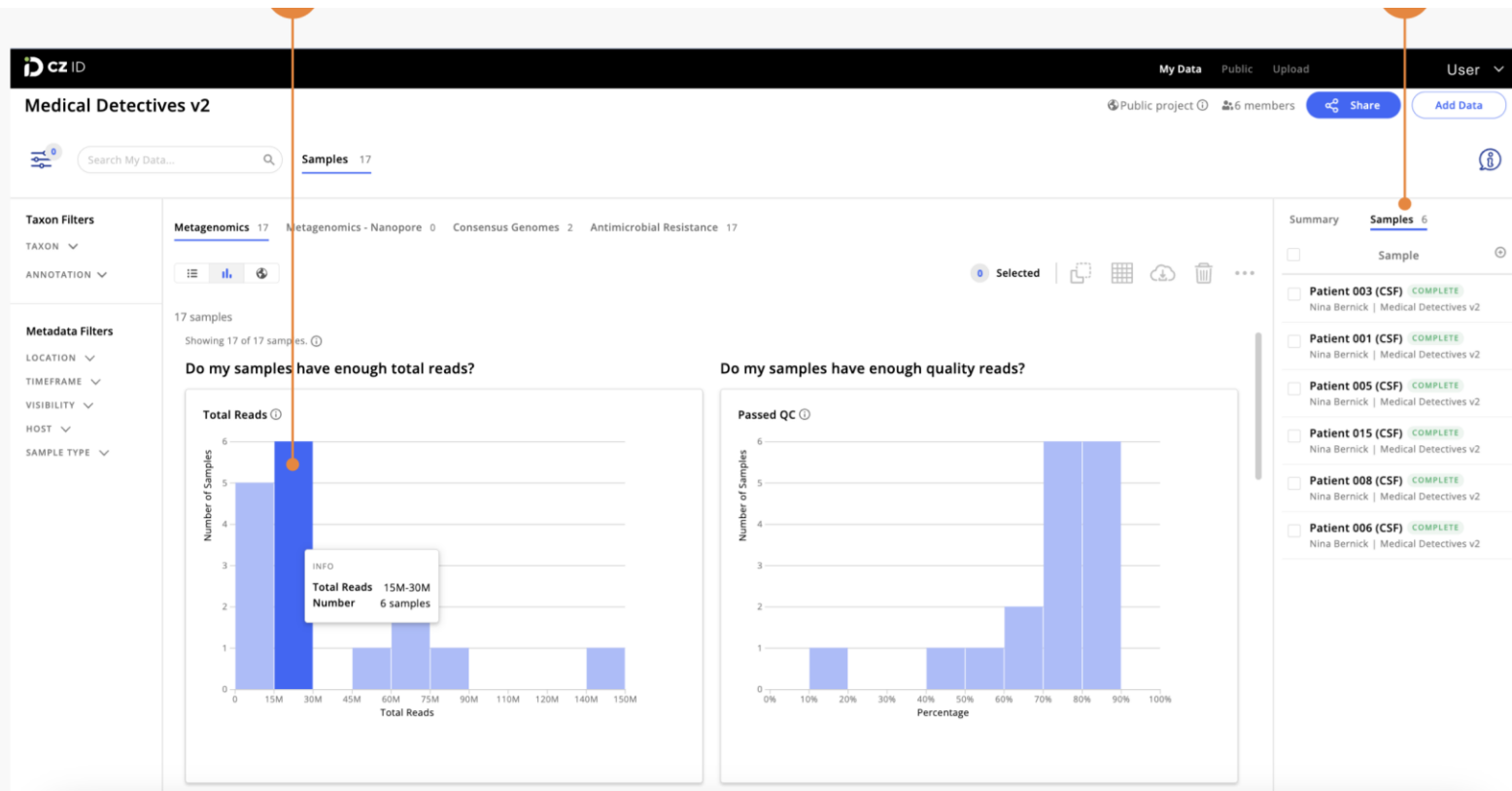
Human 17

Sample Type

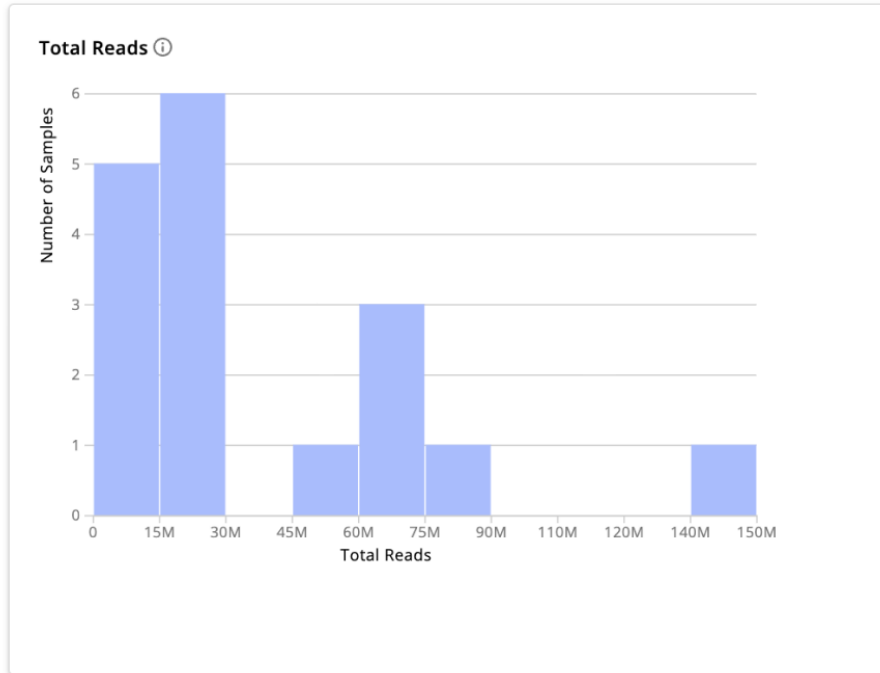
Cerebrospinal Fluid 14
Water 2
Eye 1

Location

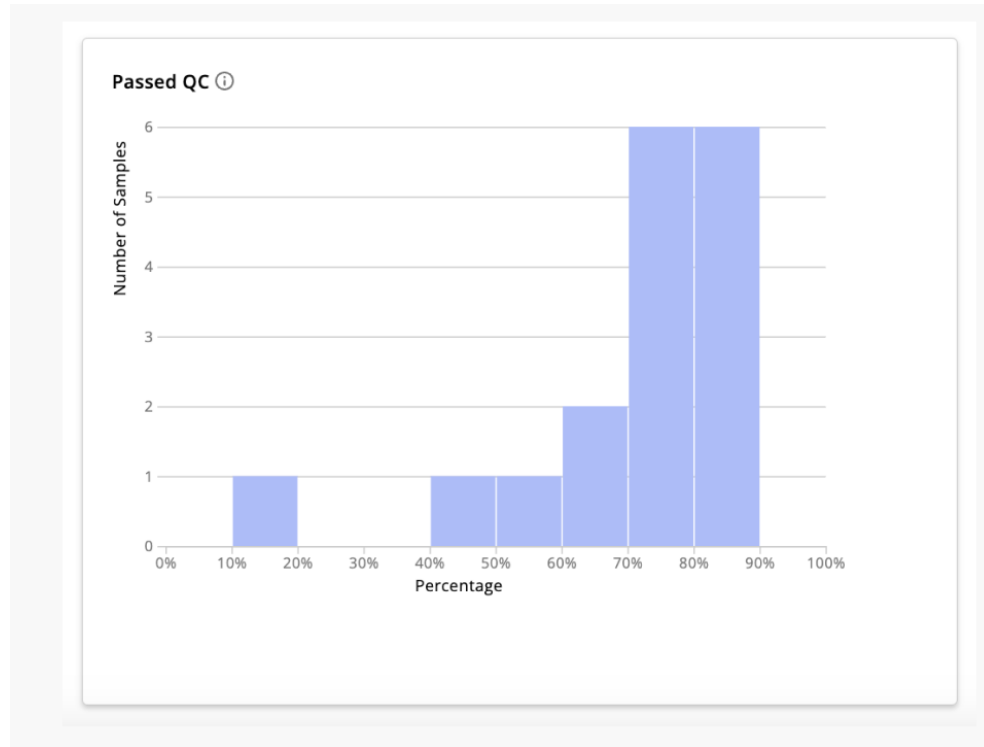
- Muestras representadas en cada barra del histograma



- Lecturas Totales

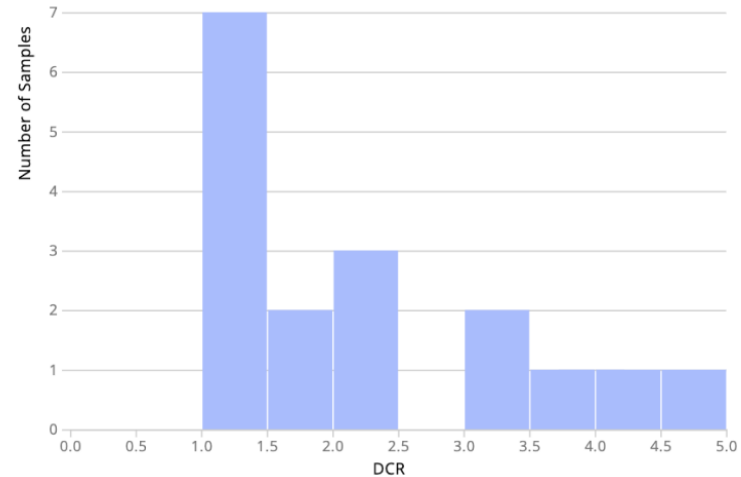


¿Mis muestras tienen suficientes lecturas de alta calidad?



¿Hay den

Duplicate Compression Ratio ⓘ



● <https://chanzuckerberg.zendesk.com/hc/en-us/articles/360053758913-Sample-QC>

