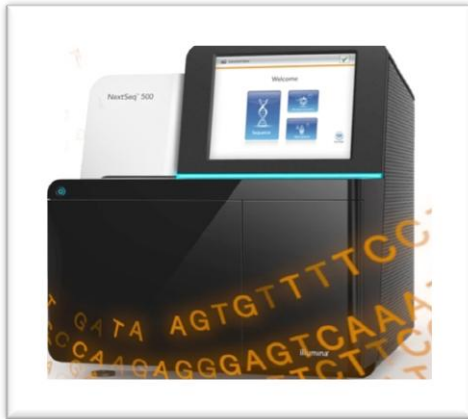


Guía de usuario de SECUENCIACIÓN MASIVA (NGS)
  
mediante la Plataforma NextSeq500 de ILLUMINA

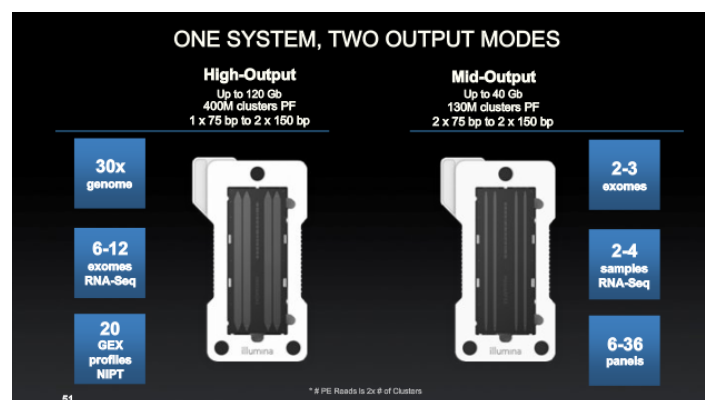


La Unidad de Genómica de CABIMER ofrece **Servicio de Secuenciación Masiva** o (NGS) utilizando la plataforma [NextSeq500](#) de Illumina. Esta tecnología está basada en secuenciación por síntesis (SBS) con nucleótidos marcados con fluorocromos y terminadores reversibles, permitiendo así una secuenciación masiva y paralela en millones de fragmentos de ADN. Específicamente, este sistema resuelve con eficiencia los homopolímeros.

Es apto para diferentes aplicaciones, pudiendo secuenciar desde paneles pequeños de genes a exomas, transcriptomas y genomas de humanos en un solo experimento.

El sistema presenta una **Flexibilidad** muy útil que permite trabajar con distintas aplicaciones y número de muestras. Ello es debido a:

- Hay disponibles **2 tipos de matrices (flowcells)** con diferente rendimiento para poderse adaptar a las necesidades de cobertura (profundidad de lectura) según la aplicación o la preferencia del investigador: **HIGH-Output** y **MID-Output** o de rendimiento alto o medio respectivamente.



- La secuenciación se puede realizar desde un único extremo de las moléculas de ADN (**single-end**) o desde ambos (**paired-end**).
- **Longitud de lectura:** la secuenciación se puede hacer hasta **75bp** o hasta **150bp de longitud**.

- Todas las muestras están marcadas con un “código de barras o *index*”, pudiendo introducir distintas muestras en la misma carrera de secuenciación, amortizando así la reacción y el experimento en general (**multiplexar**).

Con todo ello, la combinación estas características se traducen en un abanico amplio de opciones. La combinación de la capacidad de la matriz y la lectura desde un extremo o los dos, permite obtener hasta **400M lecturas en single-end** y **800M en paired-end** en HIGH-Output. Y hasta **130M lecturas** (M=millones) en **single-end** y **260M en paired-end** en MID-Output.

**NextSeq 500 Sequencing System Performance Parameters**

NEXTSEQ 500 HIGH OUTPUT KIT *			NEXTSEQ 500 MID OUTPUT KIT *		
READ LENGTH	TOTAL TIME†	OUTPUT	READ LENGTH	TOTAL TIME†	OUTPUT
2 × 150 bp	~29 hrs	100-120 Gb	2 × 150 bp	26 hrs	32.5-39 Gb
2 × 75 bp	18 hrs	50-60 Gb	2 × 75 bp	15 hrs	16.25-19.5 Gb
1 × 75 bp	11 hrs	25-30 Gb			

**Reads Passing Filter**

NEXTSEQ 500 HIGH OUTPUT KIT		NEXTSEQ 500 MID OUTPUT KIT	
Single Reads	Up to 400 Million	Single Reads	Up to 130 Million
Paired-End Reads	Up to 800 million	Paired-End Reads	Up to 260 Million

**Quality Scores<sup>††</sup>**

NEXTSEQ 500 HIGH OUTPUT KIT	NEXTSEQ 500 MID OUTPUT KIT
> 75% bases higher than Q30 at 2 x 150 bp	> 75% bases higher than Q30 at 2 x 150 bp
> 80% bases higher than Q30 at 2 x 75 bp	> 80% bases higher than Q30 at 2 x 75 bp
> 80% bases higher than Q30 at 1 x 75 bp	

\* Install specifications based on Illumina PhiX control library at supported cluster densities (between 129 and 165 k/mm<sup>2</sup> clusters passing filter). Actual performance parameters may vary based on sample type, sample quality, and clusters passing filter. All NextSeq 500 kits are paired-end capable.  
 † Total time includes cluster generation, sequencing, and base calling on a NextSeq 500 System enabled with dual-surface scanning.  
 †† A quality score (Q-score) is a prediction of the probability of an error in base calling. The percentage of bases > Q30 is averaged across the entire run.

La plataforma de Illumina ofrece una amplia gama de **aplicaciones** para todo tipo de organismos procariontas y eucariotas superiores:

- **Total RNA Seq:** Estudios de transcriptómica completo, es decir, de transcritos codificantes y no codificantes > 170nt (mRNA, lncRNA, pre-mRNAs, lncRNAs, snoRNAs, pri-miRNAs)
- Perfiles de **expresión génica** de mRNAs (**mRNA-Seq**) y microRNAs a escala global (**Small RNA-Seq**).

- Estudios de Exomas
- Chip-Seq, MNse-Seq, DRIP-Seq, ATAC-Seq...
- Secuenciación De Novo
- Re-secuenciación en busca de **variantes genéticas** de todo tipo
- Estudios de **epigenómica y metilación**
- Paneles a medida

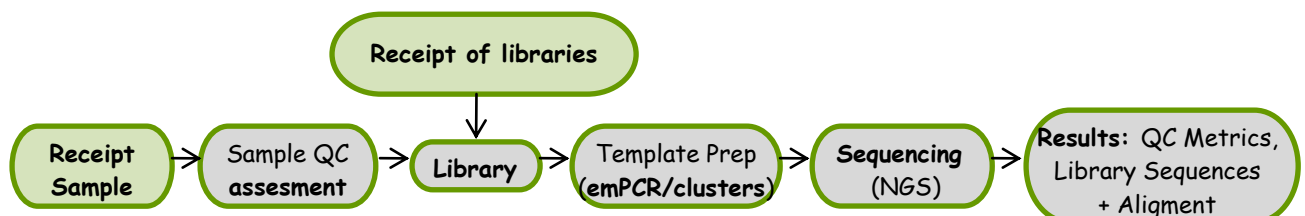
Los datos de los experimentos de Secuenciación se pueden ejecutar en una amplia gama de canales de código abierto, o bien se pueden transferir, analizar y almacenar de manera segura en [BaseSpace®](#) en la nube. BaseSpace (BS) es el entorno informático de genómica de Illumina y también incluye aplicaciones de análisis de datos de Illumina para Exomas, transcriptomas, genomas completos y llamadas de variantes somáticas.

Para ampliar la información, consulte en la web de [Illumina](#).

➤ **EI SERVICIO INCLUYE**

Para una mayor flexibilidad, el Servicio de NGS de la Unidad, ofrece dos puntos de entrada en el flujo de trabajo de Secuenciación:

1. El usuario entrega la muestra y la Unidad se encarga de preparar la librería (esto dependerá de la aplicación solicitada).
2. El usuario entrega en la Unidad la librería ya preparada, cuantificada (Fluorimetría), con rango de tamaño adecuado (Bionalyzer o similar) y que sea compatible con el Sistema de Secuenciación. Una vez entregada, la Unidad realizará los controles pertinentes para comprobar la eficiencia y calidad de la misma (qPCR, Qubit, etc..)



**El Servicio incluye:**

- **Asesoramiento** respecto al diseño experimental (Cobertura, Tipo de Librería necesaria, Preparación de la Librería, Chips, etc...).
- **Control de calidad** de la muestra (ADN, ARN o Librería) mediante BioAnalyzer, Fluorimetría (Qubit), Espectrofotometría, qPCR, etc, según determine el servicio por aplicación.
- **Preparación, Control de Calidad y Cuantificación** de las **Librerías**
- **Matrices y reactivos** para la **Secuenciación** en el NextSeq500 Illumina.
- **Análisis primario de los Resultados** que consistirá en los siguientes archivos:
  - Informe \*.pdf con los parámetros del Control de Calidad de la Carrera y los Datos/Secuencias.
  - Secuencias de la/s Librería/s (archivos FASTQ) en las que ya se ha realizado el Filtrado y *Trimming* para hacer entrega de secuencias que superan ya uno parámetros de Calidad
  - Alineamiento de la/s Librería/s (archivos BAM y BAI) si el usuario seleccionó o proporcionó un genoma de referencia cuando solicitó el Servicio.
  - Si se usan códigos de Barras o índices, los resultados se entregarán de forma individual para cada uno de ellos.
- En algunos tipos de Aplicaciones, la Unidad realiza un análisis de Datos en mayor profundidad (Consultad)

➤ **REQUISITOS DE LAS MUESTRAS DE ILLUMINA**

Si el usuario va a entregar la Librería ya preparada y compatible con Illumina, contactar con la Unidad.

Si va entregar la muestra para que la Unidad prepare la Librería debe seguir las siguientes recomendaciones.

**La calidad del Material de partida es MUY IMPORTANTE para lograr Datos Significativos.** La **cantidad** de partida **depende del experimento** y del kit que en ese momento se haya optimizado en el laboratorio. El siguiente listado es orientativo, PREGUNTE PARA SU APLICACIÓN EN CONCRETO.

**1. Muestras de ADN:**

- Proporcionar la cantidad requerida **resuspendida en agua libre de nucleasas o lowTE.**
- **ADN Bicatenario, no degradado y Sin contaminantes** (RNA, EDTA, Mg<sup>2+</sup>, etc.).

- Realizar Preferiblemente la **purificación utilizando columnas** (Qiagen o similares). Si se escoge otro método, eliminar completamente las trazas de cualquier contaminante (Etanol, fenol, etc...) ya que interfieren en los pasos enzimáticos de preparación de las librerías (**muy sensibles al EDTA e integridad de la muestra**).
- **Cuantificar las muestras de manera precisa.** Es mejor utilizar un **método fluorimétrico** por ejemplo Quant-IT (Qubit, Invitrogen) para medir con precisión la cantidad de ADN de doble cadena. En el caso de DNA-Seq, se puede utilizar la cuantificación por NanoDRop, las Ratios deben ser 260/280 entre 1.8-2.0 y 260/230 entre 2.0-2.2.
- Siempre que sea posible, realizar una electroforesis en gel de agarosa de las muestras para **confirmar la integridad de las mismas** (Proporcionar la imagen). El DNA genómico e íntegro debe observarse como una única banda >10-12 kb.
- Si la muestra es cDNA, indicar el método utilizado para su síntesis.

Tipo de Muestra	Cantidad Total	Dilución mínima	Cantidad mínima	Volumen	Cuantificación
<b>DNAg</b>	100-500 ng	4 ng/μl	100 ng + 3.5 μl QC <sup>2</sup>	2-30 μl	NanoDrop
	1-100 ng	0.04 ng/μl	1 ng + 3.5 μl QC	2-30 μl	Qubit
<b>ChiP enriched DNA (sonicado)</b> <sup>1</sup>	5-10 ng	100 pg/μl	5 ng + 3.5 μl QC	2-50 μl	Qubit
	50pg-50ng	5 pg/μl	50 pg + 3.5 μl QC	2-10 μl	Qubit
<b>MNase-Seq</b>	5-10 ng	100 pg/μl	5 ng + 3.5 μl QC	2-50 μl	Qubit
<b>DRIP-Seq</b>	10-100 ng	1.5 pg/μl	10 ng + 3.5 μl QC	<8.5 μl	Qubit

1. ChiP-enriched, fragmented DNA (200-800bp). Library molecules with shorter inserts (**200–300 bp**) tend to cluster and amplify more efficiently on the Illumina *flow cell*.

2. QC: Quality Control

Una vez recibidas las muestras y según el tipo, la Unidad evaluará la Calidad/Cantidad de las mismas con distinta metodología. Si éstas no cumplen los requisitos de control de calidad iniciales, se le informará al usuario por correo electrónico.

## **2. Muestras de RNA:**


- **Eluir** o resuspender el RNA en **agua libre de RNasas**.
- Las muestras deben estar **libres** de **cualquier** tipo de **contaminante**. Realizar la purificación utilizando columnas (Qiagen RNeasy, Invitrogen Ribominus o similares), o precipitación con etanol. En este último caso eliminar completamente las trazas de etanol

para que no interfiera en los pasos enzimáticos de preparación de las librerías. Es necesario eliminar las trazas de gDNA.

- **Cuantificar** las muestras de **manera precisa**: Espectrofotometro (NanoDrop®), Fulorímetro (Qubit) o Bioanalyzer 2100 de Agilent, según cantidades a determinar.
- Se recomienda utilizar muestras de RNA total con un valor de **RIN>7** y 28S rRNA . 18 rRNA = 2:1 ratio, para proceder a la eliminación de rRNA o el enriquecimiento en mRNA.

Tipo de Muestra	Cantidad Total	Dilución mínima	Cantidad mínima	Volumen	Método de Enriquecimiento
<b>RNA<sub>t</sub>.mRNASeq</b>	100-4000 ng	2 ng/μl	100 ng + 3.5 μl QC <sup>2</sup>	2-50 μl	Poly-A
<b>RNA<sub>t</sub>.Total RNA-Seq</b>	100-1000 ng	10 ng/μl	100 ng + 3.5 μl QC <sup>2</sup>	2-10 μl	Depleción rRNA
<b>Poly-A</b>	10-100 ng	2 pg/μl	10 ng + 3.5 μl QC	<5 μl	-----
<b>rRNA-depleted total RNA</b>	10-100 ng	1.5 pg/μl	10 ng + 3.5 μl QC	<8.5 μl	-----
<b>Small RNA</b>	1 μg total RNA	2 ng/μl	1000 ng + 3.5 μl QC <sup>2</sup>	<5 μl	Selección Tamaños
<b>RIP-Seq</b>	10-100 ng	1.5 pg/μl	10 ng + 3.5 μl QC	<8.5 μl	-----
<b>ChiRP-Seq</b>	50pg-50ng	5 pg/μl	50 pg + 3.5 μl QC	2-10 μl	-----

➤ **ENVÍO DE MUESTRAS**

Las muestras de RNA/DNA, una vez purificadas y en las concentraciones indicadas en los apartados anteriores se enviarán congeladas en nieve carbónica **en tubos Eppendorf® de 1,5 mL** en **Agua libre de RNAsas** ó Low TE, según proceda. Debe acompañarlas una copia original impresa del **Formulario de Solicitud** cumplimentado correctamente. 

Si por algún motivo ajeno a la Unidad, una vez realizada la solicitud, el usuario cancelara la misma, el material fungible correspondiente le será facturado.

La dirección a la cual pueden ser remitidas es:

Unidad de Genómica  
 Centro Andaluz de Biología Molecular y  
 Medicina Regenerativa (Cabimer)  
 Avda. Américo Vespucio s/n

Parque Tecnológico Cartuja´93  
41.092 SEVILLA

Horario de recepción de muestras: 9:00 h a 17:00 h (9:00h a 14:00h Julio y Agosto)

El envío de las muestras **debe ser comunicado previamente** por teléfono (954-467828) o por e-mail (eloisa.andujar@cabimer.es/monica.perez@cabimer.es).

Por su parte, la Unidad de Genómica comunicará vía e-mail la recepción de las mismas así como cualquier problema que pueda surgir tras los controles de calidad realizados sobre éstas.

### ➤ **ENTREGA DE LOS RESULTADOS**

Una vez terminado el experimento y obtenidos los datos de su análisis, los resultados se le enviarán al usuario vía ftp junto con un informe final (\*.pdf) explicando el procesamiento de muestras y datos.

Como se indicaba anteriormente, se entregan los Datos de secuencia en los siguientes Formatos:

- FASTQ (raw sequence+quality scores), que contiene las lecturas que ya han pasado un Filtrado de Calidad.
- BAM (“zipped” SAM; Binary Aligment Map) que representa las secuencias alineadas respecto al genoma de referencia o contigs en el caso de secuenciación *de novo*.

Esta información se ofrece en un **plazo máximo de 15 días desde su comienzo** (salvo excepciones o imprevistos ajenos al protocolo) desde el comienzo del análisis, el cual será comunicado al usuario vía email.

### ➤ **TARIFAS**

Las tarifas que se pueden aplicar dependen mucho de la cantidad de muestras que se usen y se puedan secuenciar en el mismo experimento de secuenciación. Consultar con la Unidad cada proyecto.

## **INFORMACIÓN DE INTERÉS**

### **ESTIMATING SEQUENCING RUNS**

#### **Coverage Equation**

The Lander/Waterman equation is a method for computing coverage<sup>1</sup>. The general equation is:

$$C = LN / G$$

- C stands for coverage
- G is the haploid genome length
- L is the read length
- N is the number of reads

So, if we take one lane of single read human sequence with v3 chemistry, we get

$$C = (100 \text{ bp}) * (189 \times 10^6) / (3 \times 10^9 \text{ bp}) = 6.3$$

This tells us that each base in the genome will be sequenced between six and seven times on average.

#### **Coverage Calculator**

Illumina provides an online coverage calculator that calculates the reagents and sequencing runs needed to arrive at the desired coverage for your experiment, based on the Lander/Waterman equation. The calculator can be found here:

<http://www.illumina.com/CoverageCalculator>

Perform the following steps to run the calculator:

1. Enter the input parameters:
  - The target genome or region size, for example, input 3000 Mb (3 Gb) for human genome.
  - The coverage you want.
    - The total number of cycles. For example, if you want to perform 100 bp paired-end runs (2x100), enter 200.
2. Select the instruments you want to perform the calculation for.
3. Click Submit.

The calculator now writes tables containing the total output required, output per lane or flow cell, and number of lanes or flow cells you need to use for the desired coverage. You can also download the results in a comma-separated values file, so you can share data or use the tables in Excel.

Note that the calculator uses an estimate of reads passing filter commonly found for balanced genomes (such as PhiX or the human genome). If you plan to sequence an unbalanced genome, you may have a lower number of reads passing filter, and consequently a lower output per lane. If you plan a targeted resequencing or enrichment experiment, make sure to read the technical note *Optimizing Coverage for Targeted Resequencing*.

### **NGS FAQs**

#### **1) What types of sequencing kits are available for the NextSeq?**

*There are five different kinds of NextSeq sequencing reagent kits: 75, 150 or 300 cycles in MID or HIGH read outputs as follows:*



<b>NextSeq's available sequencing kits</b>	75 cycles	150 cycles	300 cycles
<i>MID output</i>	-	Y	Y
<i>HIGH output</i>	Y	Y	Y

## 2) What does “x cycles” mean for the NextSeq sequencing kits?

*The number of cycles refer to the maximum number of bases that can be sequenced with the particular kit.*

*The cycles required to read the barcode indices do not count against this number.*

*E.g. 300 cycles implies that a maximum of 300 bases in 1x300bp or 2x150bp format plus 2x8bp barcodes can be sequenced.*

*It is also important to note that you don't have to use all available cycles. E.g. one can run a 1x50bp experiment with the 75 cycles, HIGH output kit.*

## 3) How many reads can be obtained with the different output kits?

*Specs for the kits indicate that one can expect to get a maximum of 130 million and 400 million reads for the MID and HIGH output kits respectively.*

*Do note that this is the maximum optimal outputs that are available from the respective kits. Typically, one should expect to get less than maximum but it is common to get more reads than specs too. But read quality can suffers when the number of reads are way higher than specs.*

## 4) Why is it not possible to always get maximum number of reads on the NextSeq?

*Prepped libraries can be very different due to insert lengths, library prep kit manufacturers or even person-to-person variation, so it is very difficult to get consistent maximum number of reads for different libraries. In fact, we tend to be more conservative with new samples (new users/new organisms/new prep kit manufacturer), usually aiming for 100/300 million reads for MID/HIGH output respectively.*

*This is not a problem specific to the NextSeq; all Illumina models need to face this problem. A way to minimize variation is to be consistent with library quantification methods. E.g. Always using Qubit and Bioanalyzer to quantify library concentration.*

## 5) Single reads vs paired-end reads

*In single reads, only one end of the nucleic acid fragments is sequenced. In paired-end reads, the nucleic acid fragments are first sequenced from one end, and then from the other end. This allows generation of high quality, alignable sequence data. Depending on the application, paired-end reads usually do not overlap.*

*According to Illumina's website, "paired-end sequencing allows detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts".*

## 6) What gets sequenced on the NextSeq?

*Only the nucleic acid inserts/fragments are sequenced on the NextSeq. The P5 and P7 adapters added during library prep for hybridization to the flow cell does not get sequenced. The barcode sequences, needed for multiplexing, do get sequenced but separate from the inserts.*

### **7) What is multiplexing?**

Due to the large number of reads available from a single run of the NextSeq, it is often economical to combine multiple samples (multiplex) in a run to make full use of all available reads.

In such cases, there must be a way to distinguish the different samples from each other. This is by way of adding barcode indices to the different samples during library prep. Barcode indices are 6 or 8 bp DNA sequences that are added to the samples during adapter ligations. After sequencing, NextSeq output reads are automatically sorted into different bins based on their barcode indices.

### **8) Can any kind of nucleic samples/application be multiplexed?**

Any samples can be multiplexed but to get the best results, the following aspects need to be considered during multiplexing:

- i) Fragment lengths from the different samples must match.
  - Different length libraries will hybridize to the flowcell at different efficiency so unmatched fragment lengths can lead to uneven coverage of samples
- ii) Same kind of barcode strategy must be used
  - All must be single indices or all must be dual indices
- iii) Barcode from the same type of kits must be used
  - E.g. TruSeq barcodes cannot be combined with Nextera barcodes
- iv) Illumina-equivalent indices must be known if using non-Illumina library prep kits. The alternative is to know the sequence of the indices and use them as “custom barcode”

### **9) How do I submit a sample to run on the NextSeq?**

- a) First, you will fill out the scheduling form on the BNL NGS website or by clicking this [link](#).
- b) Next, you will fill out the [Sample Barcode Template](#) with the sequences of barcode used.
- c) Within 3 days, BNL will then assign you a date for sequencing.
- d) Finally, you will submit the sample to BNL at least two days before your scheduled date.

### **10) How do I get my data?**

Sequencing data from the NextSeq is automatically uploaded to Illumina’s cloud-based platform called BaseSpace.

Once a run is started, you will receive two links to BaseSpace: one will be a link to the Project folder which will contain the output data files once the run is completed. The other will be a link to the current run which can be used to check on the progress of the run as well as other metrics associated with the run.

### **11) Are DNA-Seq, RNA-Seq the only NGS techniques that can be ran on the NextSeq?**

No, absolutely not. The NextSeq is “blind” to what NGS techniques/applications you are trying to do. As long as the library is prepped properly so that it will hybridize correctly to the flow-cell and there are correct binding sites for the enzymes, the NextSeq will sequence the library.