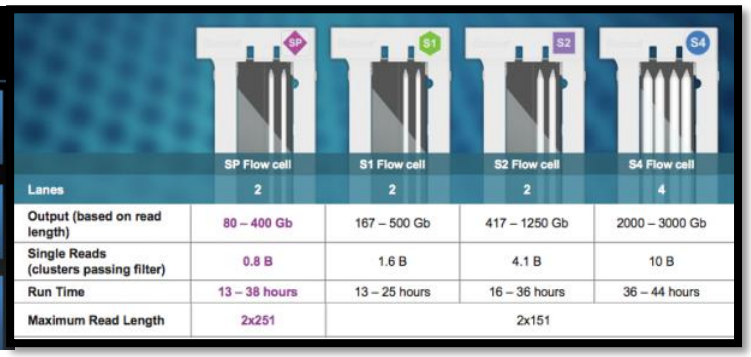
Next Generation Sequencing (NGS) User Guide
ILLUMINA system

CABIMER Genomics Core Facility offers a **Next Generation Sequencing (NGS)** service, working with Illumina iSeq100, NextSeq500 and NovaSeq6000. This technology is based on sequencing by synthesis (SBS), using fluorochrome labelling nucleotides and reversible terminators, allowing for a massive and parallel sequencing of millions of DNA fragments. Specifically, these systems efficiently solve homopolymers.

They are suitable for different applications, from small gene panels sequencing, to exomes, transcriptomes, human genomes and even higher complexity applications. It can be performed in a single experiment (run), as well as working in parallel with different projects to scale up to production-level sequencing with high quality data

One of the main advantages of this systems is its **Flexibility**, allowing working with a broad range of applications and sample number. This is possible due to different factors:

- **Scalable Architecture**: different sequencing **matrix (flowcells)** are available, with different output to adapt to coverage needs (read depth) depending on the application or the user preference. Mix and match six types of flow cells **MID-Output and HIGH-Output** (NextSeq500) and **SP, S1, S2 and S4** (NovaSeq6000).



	SP Flow cell	S1 Flow cell	S2 Flow cell	S4 Flow cell
Lanes	2	2	2	4
Output (based on read length)	80 – 400 Gb	167 – 500 Gb	417 – 1250 Gb	2000 – 3000 Gb
Single Reads (clusters passing filter)	0.8 B	1.6 B	4.1 B	10 B
Run Time	13 – 38 hours	13 – 25 hours	16 – 36 hours	36 – 44 hours
Maximum Read Length	2x251	2x151		

- **Adjustable output:** sequence **one or dual flow cell** system with a broad output, with different read lengths simultaneously, Single-end or Paired-end.

- **Read length**: sequencing can be done up to **35bp, 50bp, 75bp**, **100bp, 150bp, 300bp o 500bp length**, depending on the system and flowcell.

- All samples are tagged with a "barcode or index", allowing the processing of different samples at the same sequencing run, reducing price per gbase.

To sum up, the combination of these features results in a lot of solutions of sequencing possibilities for researchers.

System Configurations	Millons Reads/flow[2]	Gigabases
NXS Mid Output -150 cycles	130M	16-19
NXS Mid Output -300 cycles		32-39
NXS High Output - 75 cycles	400M	25-30
NXS High Output -150 cycles		50-60
NXS High Output - 300 cycles		100-120
NVS SP - 100 cycles	800M	65-80
NVS SP - 200 cycles		134-167
NVS SP - 300 cycles		200-250
NVS SP - 500 cycles		325-400
NVS S1 - 100 cycles	1.600M	134-167
NVS S1 - 200 cycles		266-333
NVS S1 - 300 cycles		400-500
NVS S2 - 100 cycles	4.100M	333-417
NVS S2 - 200 cycles		667-833
NVS S2 - 300 cycles		1.000-1.250
NVS S4 – 35 cycles	10.000M	280-350
NVS S4 - 200 cycles		1.600-2.000
NVS S4 - 300 cycles		2.400-3.000

>**cycles= nucleotides or bp**

Illumina platform offers a wide range of applications, for all kind of prokaryotic and superior eukaryotic organisms:

Popular Applications & Methods	NextSeq 550 Series — Key Application	NovaSeq 6000 — Key Application
Large Whole-Genome Sequencing (human, plant, animal)		●
Small Whole-Genome Sequencing (microbe, virus)	●	●
Exome & Large Panel Sequencing (enrichment-based)	●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)	●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)	●	●
Chromatin Analysis (ATAC-Seq, ChIP-Seq)	●	●
Methylation Sequencing	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)	●	●
Cell-Free Sequencing & Liquid Biopsy Analysis	●	●

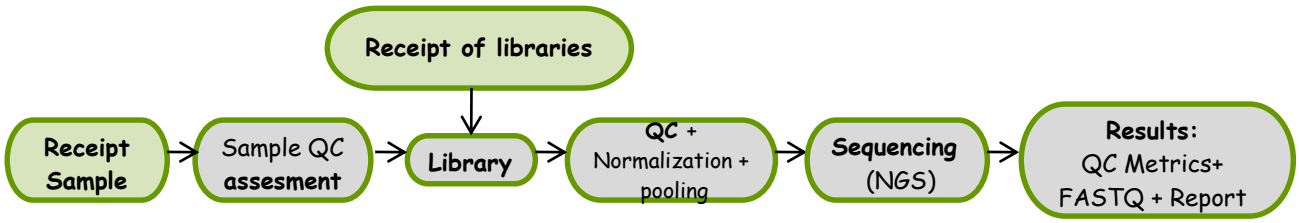
Data obtained at the sequencing experiments can be analyzed in different open-code applications, and it can also be safely stored at the BaseSpace® cloud service. BaseSpace (BS) is an informatics environment for genomic software hosted by Illumina, which also includes data analysis applications for exome, transcriptiome, whole genome, metagenomics, and somatic variants among many others.

For more information, visit the Illumina website.

➢ THE SERVICE INCLUDES:

For a higher flexibility, the Facility offers two entry points at the workflow for NGS services:

1. The user sends the sample and the facility prepares the library (depending on the solicited application).
2. The user sends the library, already quantified (fluorimetry) and with an adequate size range (checked at Bioanalyzer or similar). It must be compatible with the Sequencing Platform. Once the libraries are received, the lab will proceed with the quality analysis to assess the efficiency and quality of these libraries (qPCR, Qubit…etc.).



Another features included:

- **Advice for** experimental design (applications, coverage… etc.).
- **Sample quality control** (DNA, RNA o Library) using BioAnalyzer, Fluorimetry (Qubit), Spectrophotometry, qPCR, etc, depending on the application.
- **Normalization, quality control and quantification (libraries)**
- **Flowcells and reactives** for the **sequencing run** at the Illumina platform.
- **Primary analysis of the** results, consisting on the next files:
  - Report *.pdf with the quality control parameters of the sequencing run and data/sequences.
  - Library sequences (FASTQ files) already filtered and trimmed. These sequences already pass the appropriate quality parameters.
  - If barcodes or indexes are used for library preparation, results are individually sent for each one of them.

➢ ILLUMINA SAMPLE REQUIREMENTS

If the user decides to send the library already prepared and compatible with the Illumina platforms, please contact the Facility.

If samples will be sent for the preparation of the library at Genomics, the following recommendations must be carefully followed:

Starting material quality is HIGHLY IMPORTANT to achieve good results. The quantity of the sample is determining for the experiment and kit, which has been optimized at the lab. The next information is descriptive; ask at the people in charge at the Facility for your specific application:

A. DNA samples input:

- Provide the required sample amount diluted in **DNase free water** or **lowTE**.
- **Double strand DNA, non degraded and without contaminants** (RNA, EDTA, Mg2+,etc.).

- **Purify ideally with columns** (Qiagen or similar). If a different method is used, remove completely all traces of any possible contaminant (ethanol, phenol… etc.) since they could inhibit the enzymatic steps of the library preparation (**very EDTA sensitive**).
- **Accurately quantify samples.** A **fluorometric method**, such as Quant-IT (Qubit, Invitrogen), is more desirable to measure dsDNA quantity with high precision. In the case of DNA-seq, NanoDrop quantification can also be used. Ratios must be between 1.8-2.0 for 260/280, and between 2.0-2.2 for 260/230.
- Whenever it is posible, run an electrophoresis in agarose gels to confirm sampel integrity (and send us the resulting image). Genomic DNA must be visualized as a single 10-12 kb band.
- For cDNA samples, specify always the synthesis method performed.

Illustrative table. CONFIRMATION WITH THE FACILITY . HIGHLY IMPORTANT

Sample type	Total quantity	Minim concentrat (qubit)	Volume
DNAg	100-500 ng	4 ng/µl	>15-30 µl
	1-100 ng	0.04 ng/µl	>15-30 µl
ChIP enriched DNA (sonicated)[1]	5-10 ng	100 pg/µl	>15 µl
	50pg-50ng	5 pg/µl	>15 µl
MNase-Seq	5-10 ng	100 pg/µl	>15-30 µl
DRIP-Seq	10-100 ng	1.5 pg/µl	>8 µl

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

Once received at the lab and depending on sample type, sample quality/quantity will be assesed. If samples **do not meet the initial quality control requisites**, the user will be **notified via email**.

B. RNA samples input:

- **Elute** or resuspend RNA in **RNase free water** (previously **treated with DNase**).
- Samples must be **free** of **any kind of contaminant**. Purification must be done by columns (Qiagen RNeasy, Invitrogen Ribominus or similar), or ethanol precipitation. In the case of ethanol precipitation, all traces must be completely removed in order not to allow interferences with the enzymatic steps of library preparation. It is also necessary to remove **gDNA** traces. DNAse heat inactivation is not recommended since it could affect sample integrity. It is not sufficient with its inactivation, its removal is completely essential.
- **Quantify** samples **precisely**: Spectrophotometer (NanoDrop®), Fluorimeter (Qubit) or Bioanalyzer 2100 de Agilent, depending on quantity range.

- A **RIN>7** value is highly recommended, with a 28S rRNA - 18 rRNA = 2:1 ratio. There are alternatives for samples at low concentrations and/or quailities (FFPE… etc.), for this please ask the facility.

Illustrative table. CONFIRMATION WITH THE FACILITY HIGHLY IMPORTANT

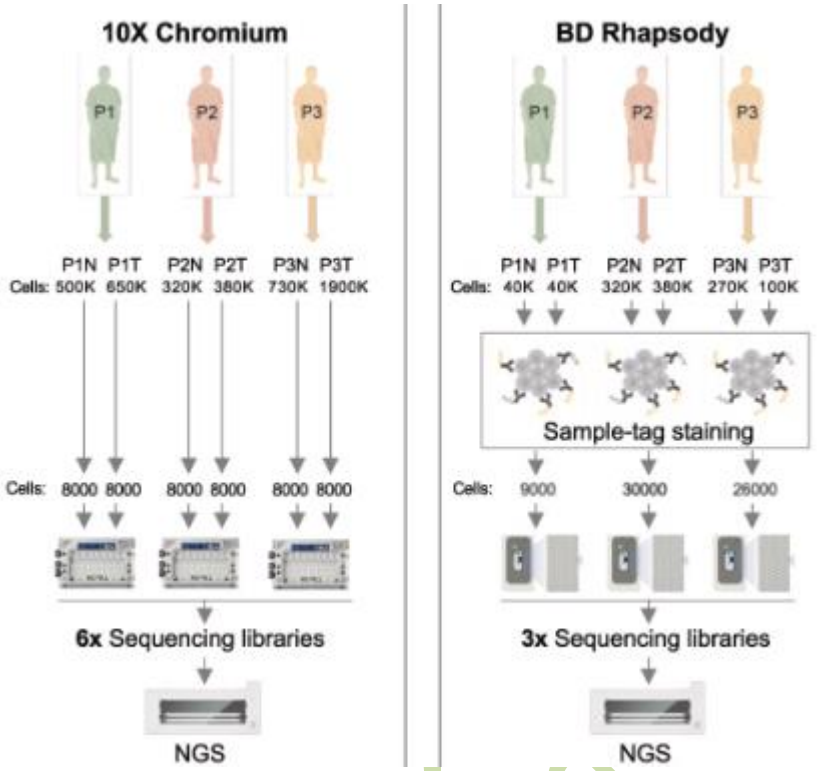
Sample Type	Total quantity	Minim concentrat (qubit)	Volume
RNAtotal "mRNA-Seq"	100-4000 ng	50 ng/µl	>15-50 µl
RNAtotal "Total RNA-Seq"	100-1000 ng	50 ng/µl	>15-50µl
Poly-A	10-100 ng	2 pg/µl	>15-50 µl
rRNA-depleted total RNA	10-100 ng	1.5 pg/µl	>8 µl
Small RNA	1 µg total RNA	50 ng/µl	>15 µl
RIP-Seq	10-100 ng	1.5 pg/µl	>15 µl
ChiRP-Seq	50pg-50ng	5 pg/µl	>15 µl

These numbers are approximate and recommended for standard human experiments. Ask for the specifications for each experiment with the Facility.

C. <u>Single-cell</u>:

For different single-cell applications, the platform is supplemented with two instruments, **10X Genomics Chromium Controller or BD Rhapsody HT Xpress**, that sepa rates the cells and tags the genetic content in them.

For sample requirements for applications, please contact to the facility. The information is regularly updating, as the applications too. Overall, workflow is as shown in the following scheme:



➢ SAMPLES SHIPMENT

RNA/DNA samples, once purified and at the concentrations indicated by the people in charge at the Facility, should be sent frozen in carbonic snow in 1.5 mL Eppendorf tubes in RNAse-free water or Low TE, as appropriate. They must be attended by an original printed copy of the correctly completed Application Form.

For Single cell applications, please ask to the facility.

If for any reason beyond the Genomics Core Facility's control, once the application has been made, the user cancels the application, the corresponding consumables will be invoiced.

The address to which they can be sent is:

Unidad de Genómica
Centro Andaluz de Biología Molecular y
Medicina Regenerativa (Cabimer)
Avda. Américo Vespucio nº 24.
Parque Tecnológico Cartuja´93
41.092 SEVILLA

Sample reception schedule: 9:00 am to 4:00 pm (9:00 am to 2:00 pm July and August).

Samples sent must be notice by telephone (954-467828) or by e-mail (eloisa.andujar@cabimer.es/monica.perez@cabimer.es).

The Genomics Core Facility will communicate via e-mail the reception of the samples, as well as any problem that may arise after the quality controls carried out on them.

- **RESULTS**

Once the experiment is finished and the analysis data obtained, results will be sent to the user via ftp together with a final report (*.pdf) explaining the sample and data processing.
As indicated above, the sequence data is delivered in FASTQ format (raw sequence+quality scores), which contains the readings that have already passed a Quality Filtering.

- **PRICES**

Please ask for the facility for each proyect. Prices for each service are public on the webside: https://www.cabimer.es/unidades-apoyo/genomica/

- **Addicional Infomation**

ESTIMATING SEQUENCING RUNS

Coverage Equation

The Lander/Waterman equation is a method for computing coverage1. 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 bp)*(189×106)/(3×109 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 (2×100), 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.