Adhere to the following instructions for successful completion of your project. This document will provide guidelines on DNA sample handling and quality requirements for PacBio HiFi sequencing at CNAG.

Contents

STEPS:

1. Project/subproject Creation – by CNAG Project Management (experimental design, sequencing protocol, number of samples, analysis (if required), invoicing matters, prioritization, deadlines…)
2. Barcoded Tubes shipment – by CNAG Biorepository (shipment logistics, Support in selecting DNA/RNA extraction methods, …)
3. Samples shipment to CNAG - by the Collaborator.
4. Samples reception by CNAG Biorepository.
5. Samples Quality Control and Report - by CNAG Biorepository (QC results, replacement issues…)
6. Samples selection approval – by the Collaborator.
7. Libraries preparation and Sequencing – by CNAG Long-Read Sequencing Team.
8. Data QC and Transfer – by CNAG Production Bioinformatics team

1. General considerations

- After your project has been reviewed and approved by CNAG Project Management and Long-Read Sequencing Manager, CNAG Biorepository will contact you to provide the materials and a URL link to the submission site for sample data collection.
- Refer to INS-110 for specific information about sample submission and shipment to CNAG.
- Use only the material provided by CNAG for sample shipment.
- Questions related to DNA and RNA sample requirements and shipment details should be directed to the CNAG Biorepository (Lídia Agueda, Biorepository Laboratory Manager, lidia.agueda@cnag.eu or Ana González, ana.gonzalez@cnag.eu) or to the Long-Read Sequencing Manager (laura.aguilera@cnag.eu). Other questions regarding experimental design, quotations, change in number of samples, etc… should be addressed to CNAG Project Management team (projectmanager@cnag.eu).

2. Guidelines for extracting and handling High Molecular Weight gDNA (HMW gDNA)

The library preparation process for PacBio WGS using standard DNA input, does not involve any amplification steps. The resulting library contains native DNA molecules, which are used directly as templates for sequencing. Therefore, the quality of the DNA starting material has a direct impact on the sequencing results. Any irreversible DNA damage or impurities present in the input material, such as single-stranded nicks, crosslinks, polysaccharides, etc., will impair the performance of the sequencing run. To achieve optimal sequencing performance and insert size, it is essential to use high-quality, high-molecular-weight genomic DNA (HMW gDNA).

In general, the following precautions need to be taken when extracting and handling DNA:

- Use wide bore pippette tips and avoid vortexing. The gDNA is physically fragile and shears by pipetting and vortexing.
- If precipitating the gDNA is needed through the extraction process, avoid overdrying. Allow the DNA to air-dry.
- Avoid eluting in RNAse-free H_2O or unbuffered solutions. DNA should be eluted and stored in a neutral, buffered solution such as low TE (10 mM Tris, pH 8, 0.1mM EDTA) or TE (10 mM Tris, pH 8, 1mM EDTA). High concentrations of EDTA must be removed (accepted up to 1mM EDTA).

- To help to resuspend the gDNA, carefully invert the tube several times after adding buffer and/or tap the tube gently. Alternatively, allow the DNA to stand in buffer overnight at 25°C to resuspend.
- After gDNA extraction, inactivate DNAse as recommended by the vendor kit if necessary. Avoid heat incativation when possible since overheating can introduce DNA damage. An alternative is AMPure beads purification.
- In case you are working with gel extracted DNA (i.e. when working with amplicons), it can result in lower sequencing performance due to the damage inherently caused by intercalating dyes such as ethidium bromide and exposure to UV radiation. Sequencing amplicons stained with SYBR dyes is untested, and therefore cannot be recommended. If working with a gel-extracted product that has been stained with a dye, it is recommended to bring it through additional rounds of amplification to remove damage and/or dyes prior to library prep and sequencing.
- DNA storage conditions: 4°C (short-term); –20°C / –80°C (long-term).
- The DNA should be shipped to CNAG on 4ºC.

To maximize read length and quality, it is essential that the DNA sample:

- is double-stranded; single-stranded DNA cannot be used to generate the sequencing template.
- has not undergone multiple freeze-thaw cycles as they can lead to DNA damage.
- has not been exposed to high temperatures (e.g. > 65ºC for 1 hour) or pH extremes (< 6 or > 9).
- has an OD260/OD280 ratio of 1.8 to 1.9.
- has an OD260/OD230 ratio of 2.0-2.2.
- has an OD260/OD270 ratio between 1.2-1.3. (Important for the extractions from plants)
- does not contain insoluble material.
- does not contain RNA contamination.
- does not contain denaturants (e.g., guanidinium salts or phenol) or detergents (e.g., SDS or Triton-X100).
- does not contain carryover contamination from the original organism/tissue (e.g., heme, humic acid, polyphenols, etc.)

3. HMW DNA sample quality assessment

A thorough gDNA quality check is required prior to submitting gDNA. The following recommendations to ascertain gDNA integrity, purity, and concentration are recommended:

1) Gel images of your gDNA sample:

gDNA integrity can be initially assessed by agarose gel electrophoresis; however, this method provides limited resolution for HMW gDNA. For accurate fragment size distribution and integrity assessment, automated capillary electrophoresis systems are recommended. The preferred method is the Femto Pulse

System (Agilent), which allows high-resolution analysis of HMW DNA. Alternatively, the Fragment Analyzer or TapeStation systems (Agilent) using Genomic DNA ScreenTape assays can be used.
For high-quality gDNA, the expected average fragment size is typically >48 kb.
These are the optimal fragment size distributions for WG PacBio sequencing:

- **70% of the DNA ≥10 kb (GQN10kb ≥7.0)**
- **50% of the DNA ≥30 kb (GQN30kb ≥5.0)**

To achieve the best results, gDNA samples should not show any signs of degradation, which can be observed through smeared DNA or even bands. The presence of one predominant band showing high-molecular-weight DNA with no degradation is optimal. When running agarose gels, it is good practice to include relevant marker sizes and indicate the amount of sample and ladder loaded.

2) Purity of your gDNA sample:

The gDNA purity should be determined by using the NanoDrop® instrument or equivalent. Readings of both A260/A280, A260/A230 and optionally A260/A270 ratios need to be obtained:

260/280: The ratio of absorbance at 260 nm and 280 nm is used to assess the purity of DNA. A ratio of ~**1.8** is generally accepted as "pure" for gDNA, but is dependent on the nucleotide composition of the submitted sample. A low A260/A280 ratio may indicate the presence of protein, phenol, or other contaminants that absorb strongly at or near 280 nm. Sometimes it may be caused by a very low concentration of nucleic acid. High 260/280 ratios are not indicative of an issue, value higher than 1.9 usually indicates repetition of an RNase digestion.
Ensure the measurements are conducted in a buffered environment such as (TE or Tris HCl, pH8).
Measurements are sensitive to small changes in the pH of the solution which will cause the 260/280 ratio to vary. Acidic solutions will skew the 260/280 ratio lower, while basic solutions will skew the ratio higher.

260/230: The 260/230 ratio provides a secondary measurement of gDNA purity to make inferences about the quality of sample extraction. Readings to determine purity are often higher than the respective 260/280 values. Expected 260/230 values are commonly in the range of **2.0-2.2**. Abnormal 260/230 values may indicate a problem with the sample extraction procedure.
The Protein LoBind tubes will improve UV 260/230 ratios by up to 0.1 – 0.4 by preventing carryover of contaminants stuck to the tube surfaces.

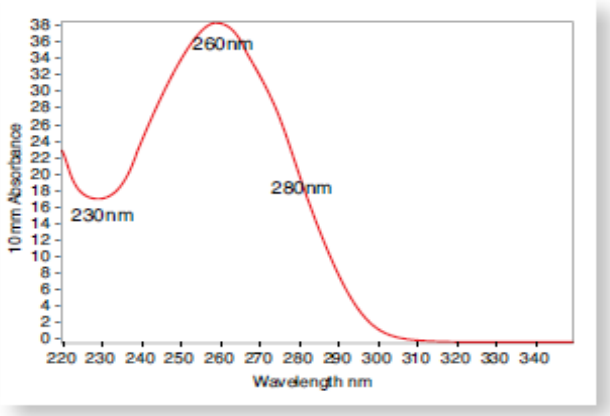
260/270: The 260/270 ratio provides information about plant gDNA sample contamination with polysaccharides and/or polyphenolics. The 260/270 ratio of gDNA sample free of these contaminants should be between **1.2 and 1.3.**

A low A260/A230 ratio may be the result of:
• Carbohydrate carryover (often a problem with plants).
• Residual phenol from nucleic acid extraction.
• Residual guanidine (often used in column-based kits).
• Glycogen used for precipitation.

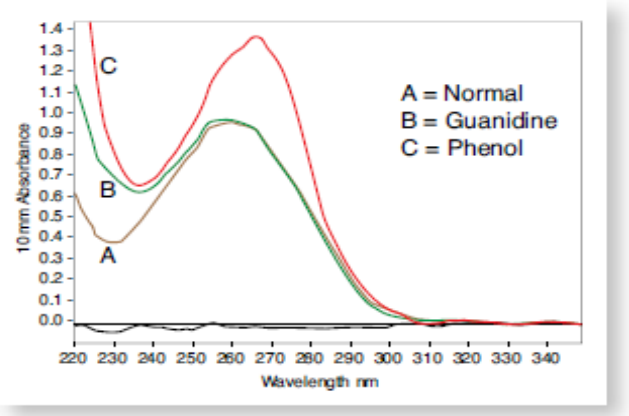
A high A260/A230 ratio may be the result of:
• Making a blank measurement on a dirty pedestal of a Nanodrop instrument.
• Using an inappropriate solution for the blank measurement.
The blank solution should be the same pH and of a similar ionic strength as the sample solution.

In addition to evaluating the Nanodrop absorbance ratios, it is critical that the gDNA concentration measured by Nanodrop does not exceed the concentration measured by fluorometric based methods by more than 150% (ratio Nanodrop concentration/fluorescence concentration ≤1.5). This ratio indicates that the majority of nucleic acids present are double-stranded and that significant amounts of other 260 nm–absorbing contaminants (e.g., RNA) are not present. If the NanoDrop/fluorescence ratio is greater than 1.5, additional purification steps should be performed before proceeding.

It is useful to evaluate the Nanodrop absorption spectra from 220 to 350 nm that reveals the presence of contaminants.



Typical nucleic acid sample spectrum



Comparison of nucleic acid sample spectra with and without 2 common contaminants

Thermo Scientific NanoDrop Spectrophotometers, Rev 11/2010

Note: <u>It is necessary to ship an aliquot of the DNA elution buffer, that will be used for Nanodrop blank.</u>

3) <u>Concentration of your gDNA sample:</u>

Traditional spectrophotometric assays cannot determine DNA concentrations <5 ng/µl. More importantly, almost all spectrophotometric assays do not distinguish between different types of nucleic acids (e.g., double stranded DNA, RNA, dNTPs, and single-stranded DNA). Misleading quantification can result in underestimation of the available gDNA and may impair library preparation.
Therefore, it is highly recommended to use a fluorescence-based assay as the PicoGreen® assay or a Qubit® fluorimeter for quantitation purposes.

Please closely follow the recommended guidelines provided by the respective vendors when carrying out quantitation assays. Perform replicate readings of concentration and use the average of replicates as the final concentration whenever possible.

4. Sample requirements

- The **minimum** HMW gDNA input required for one SMRTcell PacBio WGS run, as quantified by fluorescence-based mehotds, is **5 µg (concentration range 100-200ng/ul).**
- Higher input amounts may be required if the gDNA is fragmented or not HMW, due to the increased risk of DNA loss during library preparation.
- The gDNA can be dissolved in Tris buffer (e.g., 10 mM Tris, pH 7.0 – pH 8.0). Do not use nuclease-free water as this is insufficient for long-term gDNA stabilization. For long-term storage of HMW gDNA we recommend the use of TE buffer or low TE buffer. High concentrations of EDTA must be removed prior to library preparation as this is not compatible with some enzymatic reactions (Mg++ dependent). EDTA concentration accepted up to 1mM.

Note: Input requirements depend on gDNA quality, particularly fragment size distribution. High molecular weight gDNA is strongly recommended, as degraded or shorter fragments may require adjusted input amounts. The gDNA purity is also critical, as contaminants (e.g., proteins, salts, or residual solvents) may interfere with library preparation and sequencing performance.

The following table summarizes the minimal sample requirements for WGS using PacBio technology:

	Quality requirements
HMW gDNA input	5ug per flow cell (quantified by fluoresecence-based method) 100-200ng/ul
Fragment size distribution	70% of the DNA ≥10 kb (GQN10kb ≥7.0) 50% of the DNA ≥30 kb (GQN30kb ≥5.0)
Nanodrop ratios	OD260/OD280 = 1.8-1.9. OD260/OD230 = 2.0-2.2. Nanodrop concentration/fluorescence concentration ≤1.5
Buffers accepted	Tris buffer (e.g., 10 mM Tris, pH 7.0 – pH 8.0) EDTA accepted up to 1mM