ILLUMINA NGS BEST PRACTICES AND RECOMMENDATIONS

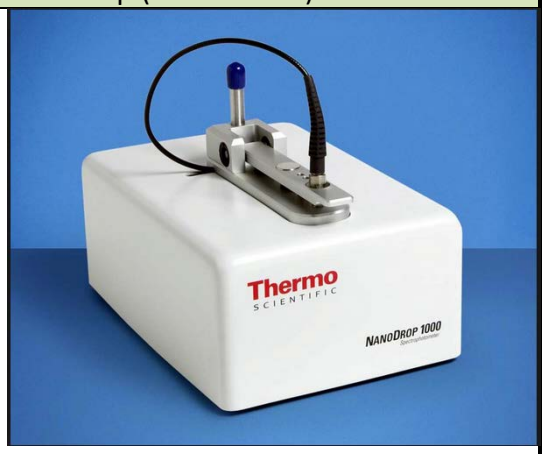
DNA/RNA Isolation

1. Use protocols that generate the purest possible preparations. Many library construction procedures (for RNAseq or DNAseq) are sensitive to inhibitors (of reverse-transcription, ligation, amplification, etc.)
2. Some sources may require additional post-isolation clean-ups to minimize the presence of inhibiting impurities such as: polyphenolics, polysaccharides, humic acid, etc.
3. For plants, in particular, use young tissue whenever possible. If experimental goals allow it, place plant in the dark for 2-3 days before removing tissue for isolation.
4. Treat DNA preps with RNases and proteases. Threat RNA preps with DNases and proteases.
5. Assess purity by absorbance. A260/280 ratio of pure preps should be 1.8-2.0. The A260/230 ratio should be >1.5.
6. When requesting library preparation services, provide as much information as possible regarding your isolation procedure.
7. Commercial kits for DNA or RNA isolation can simplify and speed up your procedure. However, these protocols may need to be supplemented by additional steps to generate sample preparations of sufficient purity and quality.
8. For long term, store DNA/RNA in TE (pH 7.5). For RNA, in particular, avoid storage solutions containing any divalent cations. For short-term storage or working solutions, dissolve in 10 mM Tris pH 7.5

Fig. 1: Nanodrop vs QUBIT: DNA quantification

Quantification of DNA samples by the Nanodrop and QUBIT

Nanodrop (Absorbance)



QUBIT 3.0 fluorometer

Recently submitted DNA samples to our core facility

By Nanodrop

Sample	ng/ul
A	32.3
B	148.9
C	46.1
D	70.6
E	141.3

By QUBIT

Sample	ng/ul
A	32.3
B	167.2
C	19.2
D	35.7
E	5.3

Samples A and B appeared to be very pure preps (see Fig. 2). The Nanodrop and QUBIT values are in good agreement. Samples behaved well during library construction and sequencing.

Samples D and E (see also Fig. 2) were very problematic and required additional purification. Initial attemps for library construction failed. Yield after extensive purification was low but samples were eventually sequenced successfully.

Quantification and Sizing

1. Assess DNA/RNA and library concentration by fluorometry. Nanodrop and other absorbance-based methods, while useful for assessing purity, are not adequate for quantification due to the fact that many substances that are found in biological sources may contribute to absorbance at or near 260 nm.

2. High MW DNA is difficult to quantify due to solution non-homogeneity and viscosity. For critical applications, re-quantitate following fragmentation.
3. Quantitate libraries by fluorometry and qPCR. Some dyes (e.g., those in the PicoGreen and QUBIT assays) effectively distinguish between dsDNA and dNTPs, RNA or ss-oligos.
4. Certain types of libraries are best quantitated by qPCR (e.g., PCR-free, amplicon libraries).
5. Make sure your library inserts are of the appropriate size for the desired run configuration (e.g., it's not a good idea to attempt sequencing in a 2x300 cycles format, a library whose average insert size is 150 bp).
6. The library size range should be as tight as possible. This facilitates the crucial step of sizing and molarity determination for setting clustering and sequencing conditions.
7. Over-amplified libraries can be difficult to resolve on the Bioanalyzer (Agilent) due to the appearance of anomalous peaks or fragments with larger MW than expected. Other sizing methods, including the capillary-based Fragment Analyzer (Advanced Analytical), appear to be less susceptible to electrophoretic mobility aberrations and can be helpful in assessing the size of these libraries more accurately.

Fig. 2: Nanodrop DNA Absorbance Scans Data

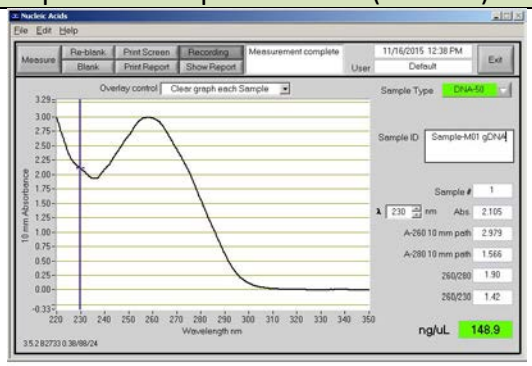
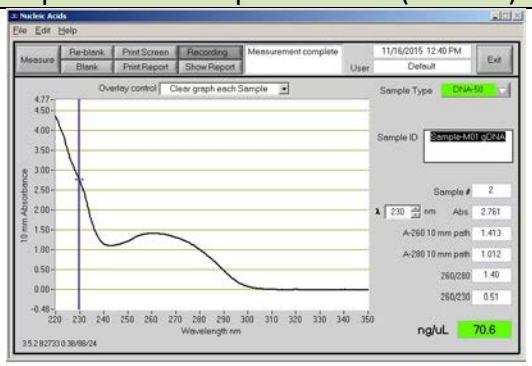
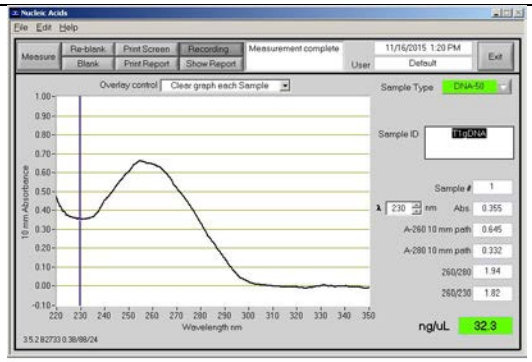
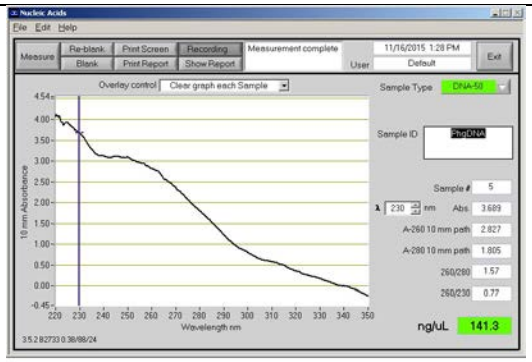
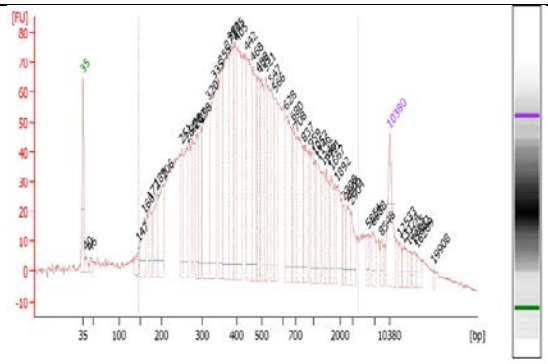
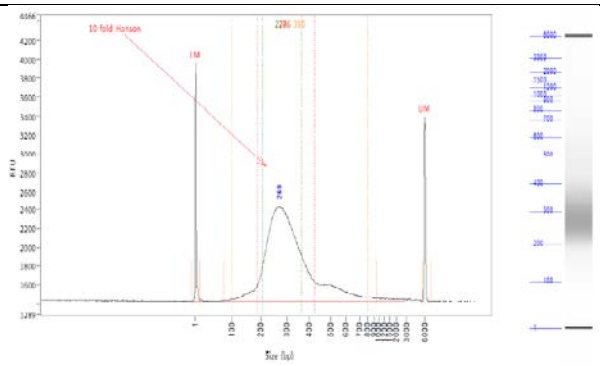
Nanodrop DNA Absorbance Scans and Data	
Samples that sequenced well (A and B)	Samples that were problematic (D and E)
	
	

FIG: BA vs Fragment Analyzer

Determining library size using the BioAnalyzer (BA) and the Fragment Analyzer (FA)	
BIOANALYZER (Agilent)	FRAGMENT ANALYZER (Advanced Analytical)
	
Amplicon library X trace in the BA	Same Amp lib X trace in the FA
	

Standard Libraries

1. Consult with your sequencing facility about the latest sequencing chemistry updates and requirements for current instrumentation.
2. Beware of library designs and protocols in the literature, as they may be outdated and no longer be supported.
3. When requesting standard library construction services from your Core facility, pay close attention to sample quality and quantity requirements.

4. Request an independent library evaluation service when submitting pre-made libraries.
5. You may save money by using third-party library construction reagents from non-Illumina companies (yet reputable) sources. Check with your core facility for suggestions.
6. Commercially available reagents and protocols vary depending on the amount of input material and the specific application. Kits' performance may vary depending on the application and the provider.
7. Fully remove adaptors and primers (i.e., they should be no more than 5% the molar concentration of the library).
8. Always keep in mind the principle of library template diversity. Sequence capture, amplicon-generated, bisulfite-converted, in-line barcoded, and other types of libraries may not sequence well without a balancer-library (spike-in up to 50%).
9. Avoid over-amplification to minimize introduction of biases, PCR artifacts, sequencing read duplicates, and other potential phenomena that may adversely affect the processing and interpretation of your data.
10. Communicate as much information as possible to your sequencing facility about the way in which your libraries were constructed. This would help in making the best recommendations and decisions for setting up a successful sequencing run.

Typical Illumina library: Structure Details

Typical Illumina Library: Structure Details

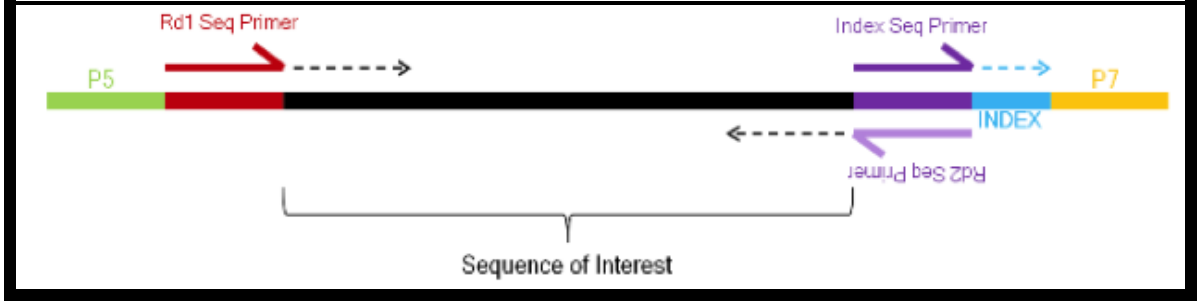
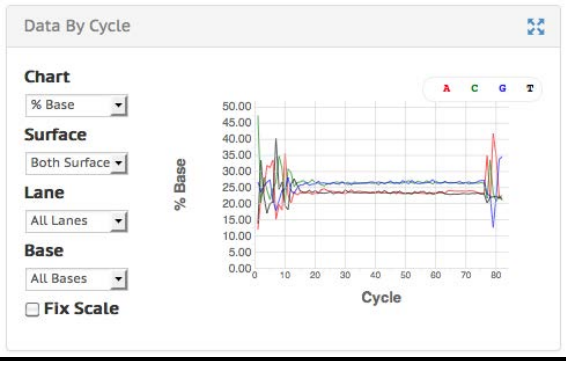
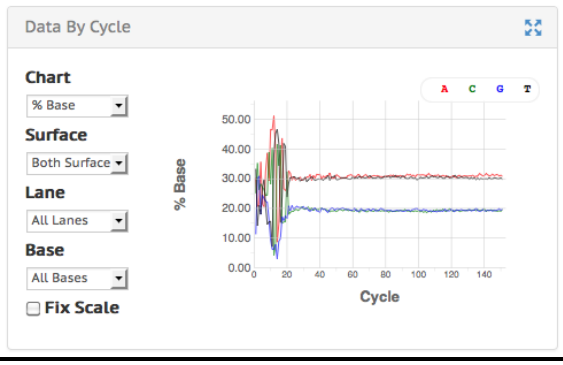
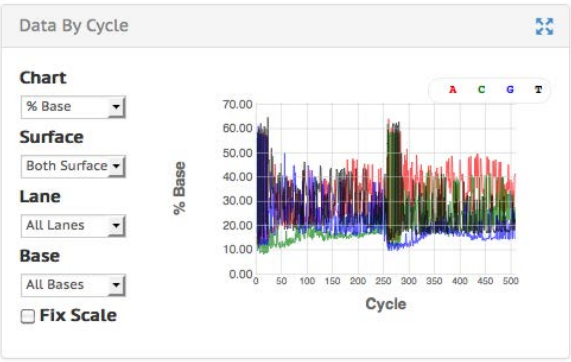
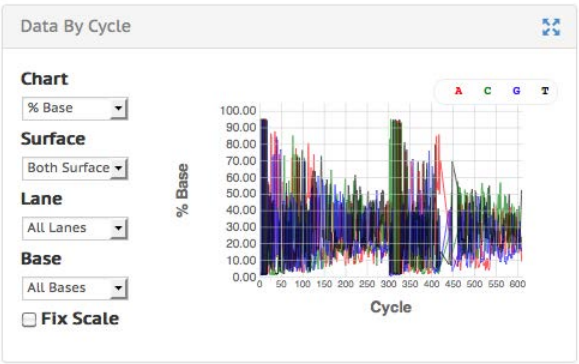


FIG: Effect of base diversity on the sequencing quality of Illumina libraries

<table>
<tr><th colspan="2">Sequencing Quality and Base Diversity: Percentage Base per cycle</th></tr>
<tr><td>Well-balanced library. Sequenced well in the NextSeq</td><td>In-line barcoded, well-balanced library. Sequenced well in the MiSeq</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Amplicon library with PhiX spike-in (15%). Sequenced well in the MiSeq</td><td>Amp library, highly imbalanced (not enough spike-in). Run failed</td></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>

Custom libraries Considerations

1. All libraries to be sequenced on the Paired-End format (MiSeq or NextSeq500) must ultimately contain full-length P5 and P7 adapter sequences.
    - **P5: 5' AAT GAT ACG GCG ACC ACC GA 3'**
    - **P7: 5' CAA GCA GAA GAC GGC ATA CGA GAT 3'**
2. There are instances when an experiment can greatly benefit from the use of a custom primer.

3. In Illumina sequencing chemistry the P5 end is sequenced first. Sequencing always occur "from the top, down"
4. The P5 end (R1) is always sequenced even if only Read2 is wanted. This is required to set cluster location coordinates, matrix and phasing information.
5. Avoid constructing libraries that begin with a linker, barcode, or other "non-random" sequence. They will not "perform" well in the MiSeq or NextSeq500, unless they are base-balanced.
6. If base diversity is limited for the first 6 positions of Read1, then the library must be mixed with a balancer DNA library (e.g., PhiX or a known random fragment library, up to 50%).

Custom Primers Design

1. Custom sequencing and index primers can be used for Read1 (side 1) and Read2 (side 2) on either the MiSeq or the NextSeq500.
2. When ordering primers, the bottom adapter should be 5'-phosphorylated in order to promote ligation. The top adapter should have a phosphorothioate bond (*) before the terminal T to ensure that exonucleases cannot digest the T overhang that pairs to the A-tail added to library fragments.
3. Custom primers must:
    - Span any initial constant regions on the insert being sequenced.
    - Be positioned and anneal so that 5'-->3' extension will occur using the sequence of interest as the template.
    - Have physical-chemical properties which match closely those of Illumina's standard primers: Tm = 66°C, 33bp, 52% GC.
    - Not form any stable secondary structures (i.e. won't stick to itself, form loops, etc.)
    - Be PAGE- or HPLC-purified. Standard desalting does not remove incomplete sequences that occur during oligo synthesis.
    - Be submitted to the NGS core lab along with custom library in a 100 uM, 10 ul solution (10 mM Tris, pH 7.5-8.0 with 0.01 Tween-20).
4. QC your custom primers by testing in Sanger sequencing reaction, and by PCR in the following reaction sets using your library as template:
    - custom primer 1 and Illumina standard P7 primer
    - custom primer 2 and Illumina standard P5 primer

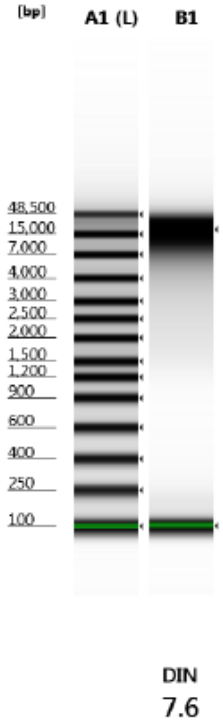
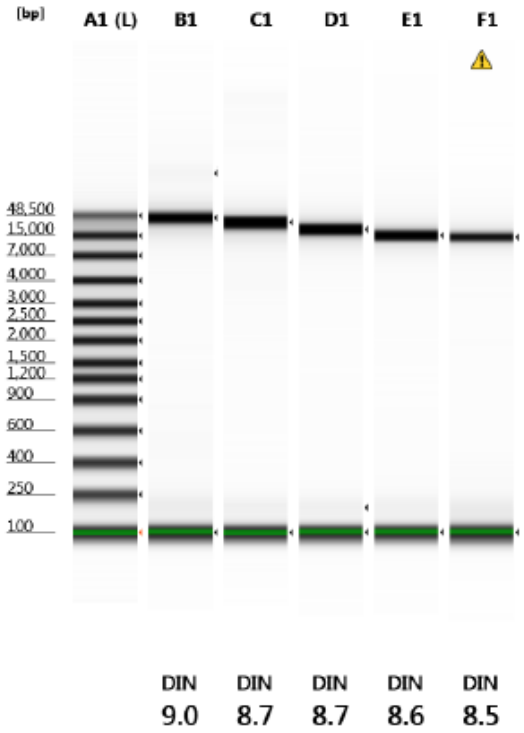
PACBIO SEQUENCING BEST PRACTICES AND RECOMMENDATIONS

DNA/RNA Isolation

1. PacBio sequencing has very stringent sample requirement. This is a consequence of PacBio sequencing chemistry and of the fact that library prep is done without any amplification. DNA or RNA must be highly pure and intact. Use protocols that generate the purest possible preparations
2. PacBio library construction and sequencing are very sensitive to contaminants. Beware of "quick and easy" isolation protocols. Always add extra clean-up steps (MOBIO, ZYMO Research, Clontech, QIAGEN clean-up reagents, etc.) to get rid of polysaccharides, polyphenolics, and any other potentially inhibiting substances.
3. PacBio requires more material than other sequencing systems. Consult with your core facility for required amounts based on your application. Required sample amounts vary (0.25-30 µg) depending on the desired insert size and the application.
4. Since the goal of IsoSeq is sequencing full-length transcript, RNA must be highly pure and intact (RIN>8.0). Always perform a DNase digestion. A minimum amount of 1 ug of fluorometrically (RiboGreen) quantitated RNA is required.
5. Also follow recommendations 3-8 as for Illumina.

Quantification and Sizing

1. Follow recommendations 1-3 as for Illumina.
2. The library size range should be as tight as possible. This facilitates the crucial step of sizing. Tightly sized libraries also result in a better distribution of insert read length (in between adaptors) in the desired size.
3. Use gels or the Electrophoretic Lateral Fractionation system (ELF) to select library fragments of the desired size.
4. Large-insert libraries can be accurately sized using the Pippin Pulse system (Sage Science). However this method requires a large amount of library material, and it's very time consuming. Two other instruments require minute library quantities, are fast and adequately accurate: a) The Agilent TapeStation for fragments up to 25 Kb; b) The Fragment Analyzer (Advanced Analytical) for fragments up to 40 Kb.

Using ELF (SAGE SCIENCES) for Library Fragment Selection		
Single PacBio Library (B1): 5-60 kb, average 18k (TapeStation).	ELF →	Five narrow MW library fractions (B1-F1): 36, 26, 20, 15, and 14 kb
	ELF →	
Notice the lower DIN number of this non-fraction selected library		Fraction C1 (26 kb) was sequenced with high quality, optimum read length results

Standard Libraries

1. Follow recommendations 1-3, as for Illumina libraries.
2. The PacBio system main applications are: small genome sequencing, transcriptomes (IsoSeq) and Amplicons (up to 10 Kb insert size). Consult with your core staff for other less common applications.
3. At this point reagents for PacBio library construction are only available from PacBio. We don't recommend submitting pre-constructed libraries for sequencing.
4. Barcoding is only supported for Amplicon libraries. Two barcoding methods are available: a) Universal amplification primer, and b) Barcoded adaptor ligation. Consult with your facility to determine which method is best suited for your experiment.

5. PacBio library construction protocols are extremely sensitive to DNA quality. Avoid any steps during isolation that may cause RNA or DNA damage such as exposure to UV light, ethidium bromide, excessive shearing, long exposure to temperatures above 65^{o}C, repeated freeze-thawing cycles, etc.
6. PacBio library construction and sequencing are very sensitive to contaminants. DNA or RNA must be highly pure. Beware of “quick and easy” isolation protocols. Always add extra clean-up steps (See DNA/RNA Isolation).
7. Use the SMRTer PCR cDNA synthesis system by Clontech for generating full-length transcript. You can submit intact/pure RNA or full-length cDNA as the point of entry for IsoSeq.
8. Always communicate as much information as possible to your sequencing facility about the way in which your libraries were constructed. This would allow them to make the best recommendations and decisions for setting up the sequencing run.