To: All NextSeq and NovaSeq users

From: Hadas Keren-Shaul, Head of the Sandbox Genomics, LSCF & Dena Leshkowitz, Head of Bioinformatics Unit, LSCF

Date: March 25, 2021

Subject: Mixing different library types in the same sequencing run

Pooling different library types in the same sequencing run can be problematic due to differences in library size, sequencing parameters (index length) and cluster efficiency. These problems might not be solved by bioinformatics analysis. We follow Illumina recommendations **NOT** to pool different library types in the same sequencing run. See detailed information in: https://emea.support.illumina.com/bulletins/2020/07/is-it-possible-to-pool-different-library-types-in-the-same-seque.html

One of the reasons people are interested in mixing MARS-seq with other library types is to save on sequencing cost. Please note that the sandbox has 192 available barcodes hence with a bit of planning, one can find a sequencing partner without mixing library types.

Choosing to mix different library types in a sequencing run requires optimization and validation at the user's own risk. If you decide to do so, we highly encourage you to pay special attention to the following issues:

a. **Sequencing parameters** – sequencing parameters (the lengths of rd1, rd2, i7, i5) are unique to each library preparation protocol. It is crucial to maintain the sequencing parameters that are essential to each protocol (i.e. the complete index length, UMI length, reads necessary for proper alignment, etc.).
b. **Nucleotide diversity** – nucleotide diversity is a critical factor for high quality data generation. When mixing different library types, make sure that nucleotide diversity is kept under all conditions. See detailed info in: https://emea.support.illumina.com/bulletins/2016/07/what-is-nucleotide-diversity-and-why-is-it-important.html.
c. **Reading sequencing indexes or library barcodes in a mixed library scenario** - one should take into account that as opposed to standard Illumina libraries, when sequencing MARS-seq libraries i7 and i5 are not being sequenced, since the sample barcode is located on rd2 (read 2). Therefore when mixing them with Illumina libraries the MARS-Seq reads will generate an 'empty' index (= constant sequence) for i7 and i5 which is part of the Illumina adapter. This can harm the quality of the bases in the index cycles depending on the fraction of MARS-Seq reads from the total reads. A proper ratio of the different types of libraries (namely, having a real i5 and i7 versus not using a real i5 and i7) should be optimized. We have seen cases (mostly in NovaSeq runs) in which there was an 'empty' index in some of the samples which resulted in sequencing errors of the 'non-empty' index in the same run - we think this is probably related to the low diversity of the nucleotides in the index which contains an 'empty' (i.e. constant) sequence.
d. **Quality and quantity of libraries** - low quality libraries can compromise the quality of the whole sequencing run. In addition, the actual quantity of the library (the real 'Illumina' molarity) can affect read distribution between the samples. For more details on library quantification and quality control

see in: https://emea.support.illumina.com/bulletins/2016/05/library-quantification-and-quality-control-quick-reference-guide.html.

e. **Library size** – shorter insert libraries tend to cluster better than longer libraries. Hence, cluster efficiency and read distribution can be affected if libraries of different lengths are being mixed. Furthermore, if the number of cycles during a sequencing run exceeds the length of the shorter library, run quality can be negatively impacted.

f. **Bioinformatics demultiplexing and running UTAP** - The bioinformatics automatic demultiplexing pipelines NGS-PIPELINE via Susanc3, supports demultiplexing of a single library type: either MARS-seq (sample barcode in rd2), Illumina (single (i7), dual index (i7+i5)) or 10X genomics. Due to all the explained above, we **do not recommend or support** mixing libraries. If you are demultiplexing on your own, you can use the UTAP MARS-Seq and RNA-Seq pipeline for analysis with the demultiplexed sequences, yet the file structure and the files names should comply to the described in: https://bbcunit.atlassian.net/wiki/spaces/BP/pages/1012760577/UTAP+Transcriptome+RNA+-seq+and+MARS-seq+pipelines+guidelines.