TECHNICAL NOTE

Maximizing performance on the NovaSeq™ X Series

Library loading optimization steps to ensure run success

- Determine optimal library loading concentration for NovaSeq X Series flow cells
- Consider library compatibility when pooling different library types

illumina®

Introduction

The NovaSeq X Series is the latest evolution in production-scale sequencing systems from Illumina, delivering multiple advances in high-throughput sequencing workflows.[1] Improvements in scale, flexibility, sustainability, and lab operations are enabled by key innovations, including an updated clustering chemistry on NovaSeq X flow cells. Compared with clustering chemistry used on previous sequencing systems, the NovaSeq X clustering chemistry allows for deeper sequencing applications with minimal sample input.

Whether transitioning projects to the NovaSeq X Series from another sequencing system, or transitioning current NovaSeq X Series projects to a new flow cell or recipe configuration, optimizing library loading can help maximize data yield and quality. This technical note provides recommendations to help achieve optimal results on the NovaSeq X Series, including guidance on library quality, library quantification, loading concentration, nucleotide diversity, and library pooling considerations.

Library quality

Short inserts and contaminants introduced during library preparation, including adapter dimers, primer dimers, and partial library constructs, can negatively impact clustering on the NovaSeq X Series. These short inserts and contaminants are typically removed during cleanup or size-selection steps. If necessary, short inserts and contaminants can be more effectively removed by adding an optional bead purification step to the library preparation protocol.[2] Once library preparation is complete and before sequencing, users should verify the quality and purity of all libraries. Use an Agilent Bioanalyzer or Fragment Analyzer system to check for library integrity, average insert size, and contaminants.

Optimal variant calling with Illumina DNA PCR-Free Prep on the NovaSeq X Series technical note

Library quantification

Having an accurate measure of library quantity is needed to load the correct concentration and provide optimal results on the NovaSeq X Series. Library quantification by size-normalized qPCR is recommended due to its consistency and accuracy. Unlike a fluorometer, which measures all DNA species in a library (including primer dimers and library fragments), qPCR performed with primers designed against Illumina adapters measures only functional library fragments.

Best practices for library quantification

Loading concentration

It is crucial to determine the optimal loading concentration for distinct library types. Loading concentration refers to the final concentration of a library loaded onto an instrument for sequencing. After libraries are prepared, they are diluted to the loading concentration appropriate for the library type, sequencing system, and reagent kit.

Loading libraries at a concentration that is too high or too low may lead to lower sequencing quality and yield, and, possibly, run failures in extreme conditions. Underloading may result in a low percentage of nanowell occupancy (% occupied) and higher duplicate reads, which then requires more reads to achieve target coverage. In contrast, overloading may result in a low percentage of clusters passing filter (% PF).

Moreover, for libraries with a wide insert size distribution, such as those that employ enzymatic fragmentation, loading a higher concentration may result in shorter mean insert size representation (Figure 1). Optimizing library loading concentration (ie, finding a balance between underloading and overloading) is highly recommended when running any new library type or when using a new flow cell type or recipe.

Titrate the loading concentration using the approach in the following example experiment to assess primary and secondary metrics to determine the ideal loading conditions.

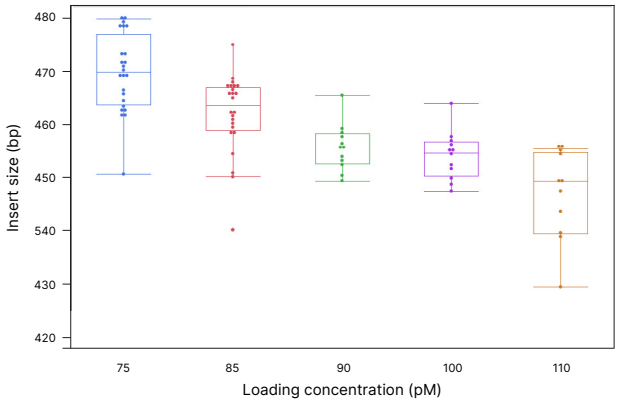

Figure 1: Determining optimal loading concentration for TruSeq PCR-Free libraries on a NovaSeq X 10B flow cell—Lower loading concentrations typically result in clusters with larger mean insert sizes. Data were generated with NovaSeq X Control Software v1.2.

Example: Determining optimal library loading concentration

It is critical to test a wide range of concentrations, including underloaded and overloaded concentrations. Use primary metrics like % PF and % occupied in conjunction with secondary metrics like duplicates, insert size, and coverage to identify the optimal loading concentration.

Step 1: Design titration experiment

For transitioning projects from the NovaSeq 6000 System to the NovaSeq X Series, center titrations at ~30% of the NovaSeq 6000 System loading concentration for the standard onboard clustering workflow.

In this example, TruSeq™ DNA PCR-Free libraries were typically loaded at 350 pM for the NovaSeq 6000 S4 flow cell standard workflow. The titration was performed with 40, 60, 70, 80, 90, 100, 120, and 160 pM in individual lanes of the NovaSeq X 10B flow cell.

Step 2: Assess nanowell occupancy and clusters passing filter

Plot the % PF vs % occupied metrics per lane for each loading concentration to determine which concentrations resulted in underloading, overloading, or balanced loading (Figure 2). These metrics are reported for each run in Sequencing Analysis Viewer software.

Plotting % Occupied by % PF to optimize loading for the NovaSeq X/X Plus, NovaSeq 6000, and iSeq 100

In this example, the positive slope pattern for the 40-pM library indicates an underloaded run (Figure 2A), and the slightly negative, near-vertical slope pattern for the 160-pM library indicates an overloaded run (Figure 2B). The libraries tested at 80–100 pM all displayed optimal loading shape in the % PF vs % occupied plot (Figure 2C).

Step 3: Assess duplicates

Narrow the target concentration range by analyzing the percent of duplicates. In this example, the optimal range was reduced as follows: Concentrations of 80–120 pM have duplicates of ≤ 15%. We focused on 90 pM, 100 pM, and 120 pM where the duplicates were lower (Figure 3).

Step 4: Review average coverage

Analyze coverage across all concentrations. Coverage is application-dependent and influenced by the number of libraries loaded in the same lane. In this example, the 30× coverage values increased confidence in 90 pM and 100 pM as the optimal concentrations (Figure 3).

Step 5: Analyze insert size

Review the insert sizes. Insert size may decrease as loading concentration increases for some library types. The optimal range for your library and application may vary depending on your workflow requirements.

In this example, 90–120-pM samples have similar mean insert sizes (~421–427 bp) (Figure 3). Thus, both 90 pM and 100 pM are optimal loading concentrations for NovaSeq X 10B flow cells for the TruSeq DNA PCR-Free whole-genome sequencing libraries.

NovaSeq X Control Software v1.3 offers improved tolerance to higher loading concentrations, allowing for reduced duplicates and increased useable reads. It is important to note that the software does not directly reduce duplicates, but duplicates are generally observed to decrease with higher loading concentrations. Decrease in insert size and increase in pooling coefficienct of variation (CV) should be closely monitored when the loading concentration is increased. NovaSeq X Control Software v1.3 does not require re-titrating the loading concentration when transitioning from earlier versions of control software; however, it can be re-optimized if desired.

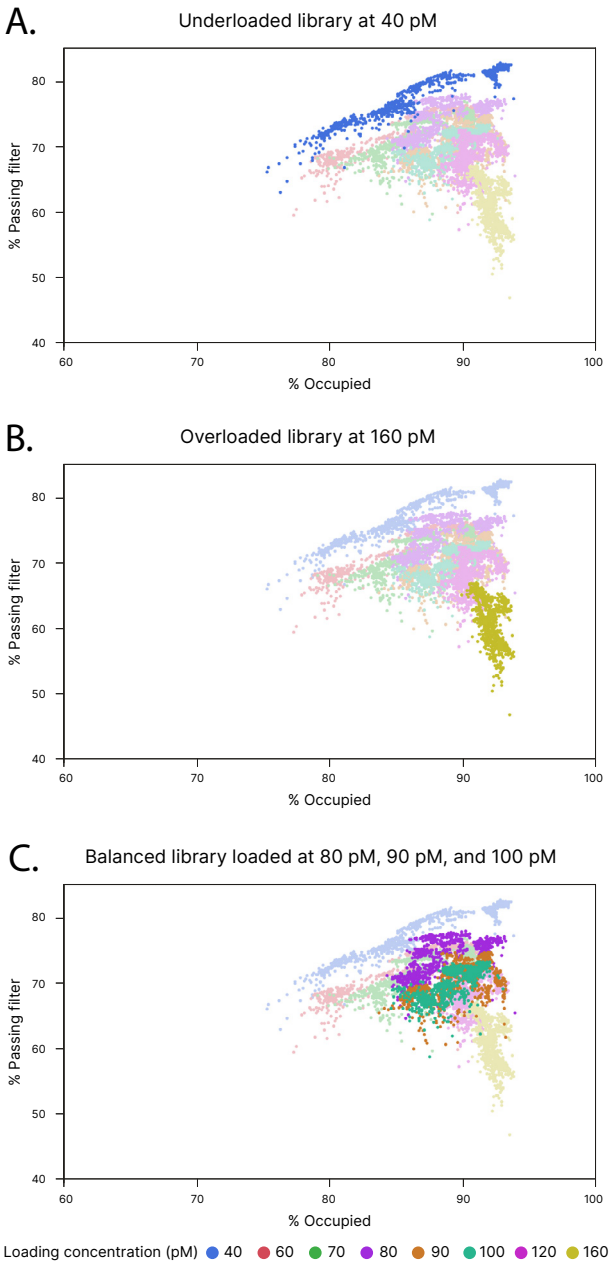

Figure 2: Assessing nanowell occupancy and clusters passing filter to optimize whole-genome sequencing run on a NovaSeq X 10B flow cell—Example titration experiment plotting clusters passing filter (% pass filter) vs nanowell occupancy (% occupied) across different library loading concentrations. Dots represent each field of view of the camera, averaged over two flow cell runs. (A) Underloaded runs with 40 pM, (B) overloaded runs with 160 pM, and (C) balanced loading of runs for 80 pM, 90 pM, and 100 pM. Data were generated with NovaSeq X Control Software v1.2.

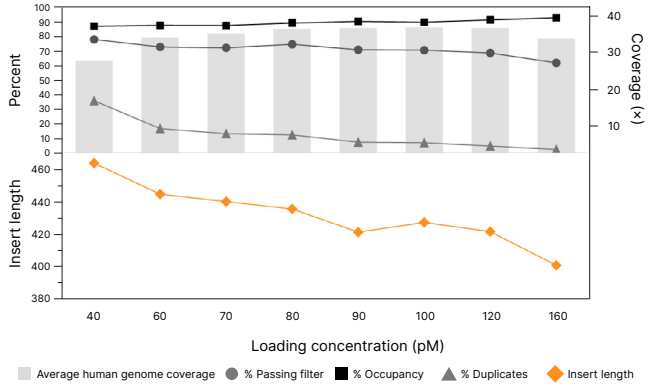

Figure 3: Optimizing whole-genome sequencing run on a NovaSeq X 10B flow cell—Example titration experiment, looking at duplicates, average coverage, and insert size. Data were generated with NovaSeq X Control Software v1.2.

Refer to the dilute and denature section of the NovaSeq X Series user guide for additional examples of optimized loading concentrations.

Nucleotide diversity

Nucleotide diversity means the relative proportion of each base (A, C, G, or T) present in every cycle of the run. Nucleotide balance is important for color matrix correction and intensity normalization by the sequencing system. To increase nucleotide diversity and provide a balanced signal, spike in PhiX, starting with a higher spike-in percentage and reduce the amount based on run performance. Alternatively, pool low-diversity libraries with another high-diversity library.

What is nucleotide diversity and why is it important?

How much PhiX spike in is recommended when sequencing low diversity libraries on Illumina platforms?

Library pooling

The NovaSeq X Series offers a high level of flexibility for sample multiplexing with individually addressable lanes. Follow these best practices when pooling different libraries into a shared lane or onto the same flow cell for optimal results.

Load libraries with widely different insert sizes (eg, whole-genome libraries vs microRNA libraries) into separate lanes to avoid coverage issues due to differences in clustering efficiency. Alternatively, when mixing libraries with significantly different insert lengths, spike in more of the longer insert library. The exact spike-in needs to be determined empirically for each specific library type.

When combining libraries with different read lengths or index cycles in a single run, use the longer cycles for sequencing and specify the shorter cycle number in the sample sheet for trimming during BCL convert.

Exercise caution when combining libraries that require a dark cycle recipe applied to the entire flow cell.[3] Only load other library types on the run that are compatible with a dark cycle (Table 1). Note that libraries with a unique molecular identifier (UMI) sequence starting at cycle 1 (eg, Illumina cfDNA Prep with Enrichment) are not compatible with a dark cycle. NovaSeq X Control Software v1.3 allows Illumina mRNA and Total RNA stranded libraries to be run without dark cycles, provided PhiX is spiked in at 5% or higher.

Also, when pooling for low-plexity (two-plex through eight-plex libraries), follow index adapter pooling strategies for XLEAP-SBS™ chemistry to achieve optimal color balance.

Index Adapters Pooling Guide

Table 1: Compatibility of library types for sequencing with dark cycle

Library prep	Dark cycle compatibility	
	Yes, can tolerate dark cycle recipe	No, run on separate flow cell with no dark cycle
Illumina Stranded Total RNA Prep	X	
Illumina Stranded mRNA Prep	X	
Illumina RNA Prep with Enrichment	X	
TruSeq RNA preps	X	
Illumina DNA Prep	X	
Illumina DNA PCR-Free Prep	X	
Illumina DNA Prep with Enrichment	X	
Illumina cfDNA Prep with Enrichment		X
TruSeq DNA preps	X	
TruSight™ Oncology 500 tissue and ctDNA		X
Most single-cell library preps		X

Summary

The innovative clustering chemistry on the NovaSeq X Series enables higher throughput, more data-intensive applications, and deeper sequencing with minimal sample input. Follow the best practices outlined in this technical note to assess library quality, optimize loading concentration, and pool libraries to maximize performance on the NovaSeq X Series.

Learn more

NovaSeq X and NovaSeq X Plus Sequencing Systems

High-accuracy next-generation sequencing with the NovaSeq X Series technical note

Optimizing cluster density on Illumina sequencing systems technical note

Cluster optimization technical overview

Dark cycle sequencing

References

1. Illumina. NovaSeq X and NovaSeq X Plus Sequencing Systems specification sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/novaseq-x-series-spec-sheet-m-us-00197/novaseq-x-series-specification-sheet-m-us-00197.pdf. Published 2022. Updated 2024. Accessed April 17, 2024.
2. Illumina. Optimal variant calling with Illumina DNA PCR-Free Prep on the NovaSeq X Series technical note. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/ilmn-dna-pcr-free-prep-novaseq-x-tech-note-m-gl-02388/ilmn-dna-pcr-free-prep-novaseq-x-tech-note-m-gl-02388.pdf. Published 2023. Accessed April 17, 2024.
3. Illumina. Custom recipes for Illumina Stranded libraries on NovaSeq X Series. knowledge.illumina.com/instrumentation/novaseq-x-x-plus/instrumentation-novaseq-x-x-plus-faq-list/000008394. Published 2023. Updated 2024. Accessed May 31, 2024.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02678 v2.0