illumina®

Sequencing Library QC with the iSeq™ System

The iSeq 100 System enables measurement of pooled library quality before a large-scale sequencing study on the NovaSeq™ 6000 System.

Introduction

This application note describes using the iSeq 100 System to sequence libraries as a quality control (QC) step before high-throughput sequencing. To maximize the efficiency of high-throughput sequencing, it is important to know the quality of the starting library. Performing QC using the iSeq 100 System before committing to a full-scale NovaSeq run can save time and money, while leading to more consistent sequencing results.

Using a simple, streamlined workflow, detailed QC parameters can be generated on the iSeq 100 System. These QC parameters can be used to detect sample drop outs, dilute overrepresented samples, rebalance library pools to include more samples per run while maintaining desired coverage per sample, and ensure balanced index representation across samples. Illumina collaborated with the Hartwig Medical Foundation (Amsterdam, Netherlands) to demonstrate the use of the iSeq 100 System in library QC and rebalancing.

Library quantification and rebalancing

Multiplexing enables large numbers of libraries to be pooled and sequenced simultaneously during a single sequencing run, resulting in significant improvements in sample throughput and time to data for large studies. An important consideration when pooling libraries is proper quantification and balancing of libraries within the pool. If libraries are combined in unequal concentrations, it can result in biased representation of certain libraries over others. Under representation can require additional sequencing, while over representation can lead to wasted sequencing capacity. Traditional methods of quantifying libraries for NGS include fluorimetric methods, such as the Qubit Fluorometer (Thermo Fisher Scientific, USA), and qPCR.

Library rebalancing with the iSeq 100 System

Using the iSeq 100 System, prepared and pooled libraries can be sequenced and automatically demultiplexed using the Generate FASTQ Analysis Module in Local Run Manager. The resulting metric, percent reads identified, provides valuable information on the proportion of libraries in the pool, enabling rebalancing before sequencing on the NovaSeq 6000 System. This fast and simple workflow can be completed before any large-scale sequencing study for time- and cost-savings (Figure 1).

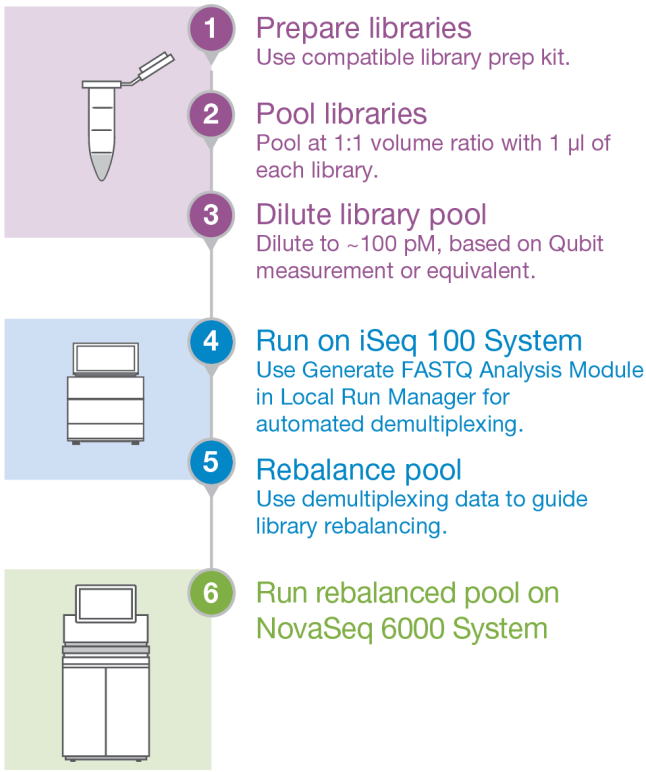

Figure 1: Library rebalancing workflow with iSeq 100 System — The iSeq 100 System provides a fast and simple workflow to rebalance library pools before a NovaSeq run.

Methods and results

NGS libraries of varying plexities were prepared from input DNA or RNA, pooled at 1:1 volume ratios with 1 µl of each library, and diluted to ~100 pM final loading concentration (Table 1). Unbalanced library pools were run on the iSeq 100 System. The percent of total for each library in the pool determined by sequencing on the iSeq 100 System showed high correlation to quantification with Qubit measurement or qPCR, as indicated (Figure 2). Using the iSeq data, libraries were rebalanced in each pool by normalizing to the relative index representation for each sample, and were sequenced again on the iSeq 100 System and on the NovaSeq 6000 System. Results showed that the pools were evenly balanced among each library for both systems (Figure 2A, 2B, 2D). Pipetting noise can affect results if rebalancing is not performed accurately with calibrated pipettes.

Table 1: Library prep for rebalancing

Input	Library prep	Plexity	Figure panel
DNA	TruSeq™ DNA Nano[a]	7-plex	2A
DNA	TruSeq DNA Nano[a]	24-plex	2B
RNA	KAPA RNA HyperPrep Kit[a]	12-plex	2C
DNA	Nextera™ Flex for Enrichment	8, 12-plex pools	2D

a. In collaboration with Illumina, the Hartwig Medical Foundation performed library preparation, sequencing, and data analysis independently.

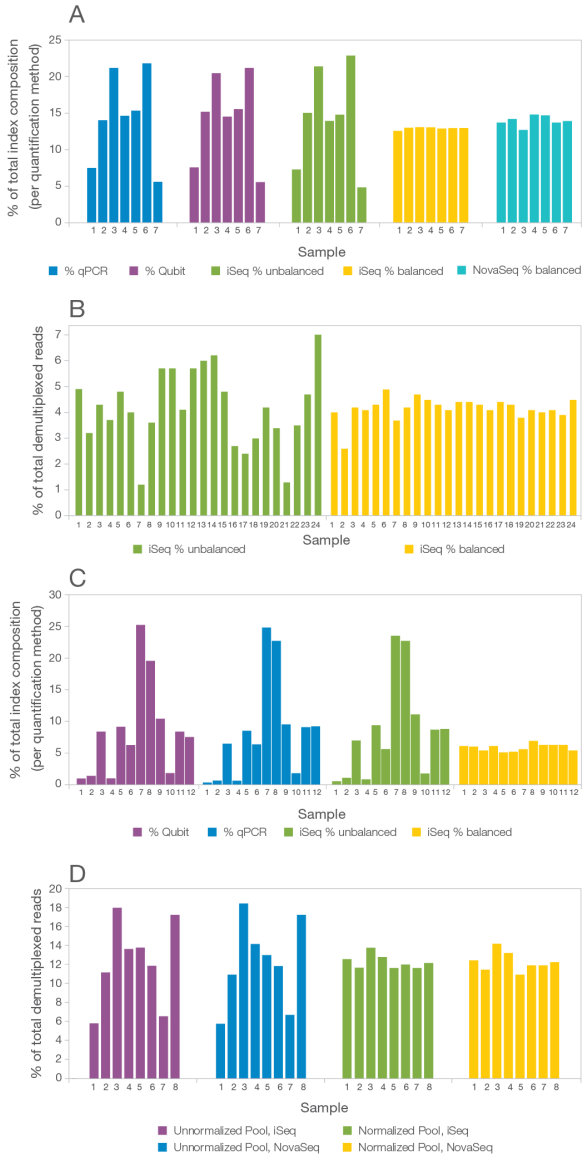

Figure 2: Library rebalancing on the iSeq 100 System— The iSeq 100 System enables rebalancing of (A) 7-plex TruSeq DNA Nano libraries with TruSeq CD indexes, (B) 24-plex TruSeq DNA Nano libraries with TruSeq CD indexes, (C) 12-plex KAPA RNA HyperPrep libraries with TruSeq CD indexes, and (D) 8, 12-plex Nextera Flex for Enrichment library pools with IDT for Illumina-Nextera UD indexes (96 indexes).

Comparison of library rebalancing on the iSeq 100 System to qPCR

qPCR has been a preferred method for NGS library quantification and library pool rebalancing. Typically, qPCR for library QC uses sequencing adapter-specific primers for amplification, resulting in highly accurate quantification of adapter-ligated molecules, as opposed to fluorimetric methods that measure total nucleic acids present. To demonstrate its exceptional performance, library rebalancing using the iSeq 100 System was directly compared to qPCR.

Methods and results

Two separate 24-plex TruSeq DNA PCR-Free libraries with IDT for Illumina-TruSeq UD indexes (24 indexes) were prepared. Unbalanced pools were measured by triplicate qPCR reactions on the Roche LightCycler 480 Instrument II, adjusted to sample insert size using the Fragment Analyzer Automated CE System (Advanced Analytical, USA), or a single run on the iSeq 100 System. The resulting data was used to rebalance each library pool by normalizing to the relative index representation (or qPCR value) for each sample, and then run on the NovaSeq 6000 System. Results show exceptional correlation between the two methods, with as good or better coefficients of variation (CV) for iSeq rebalancing, compared to qPCR for both libraries (Figure 3).

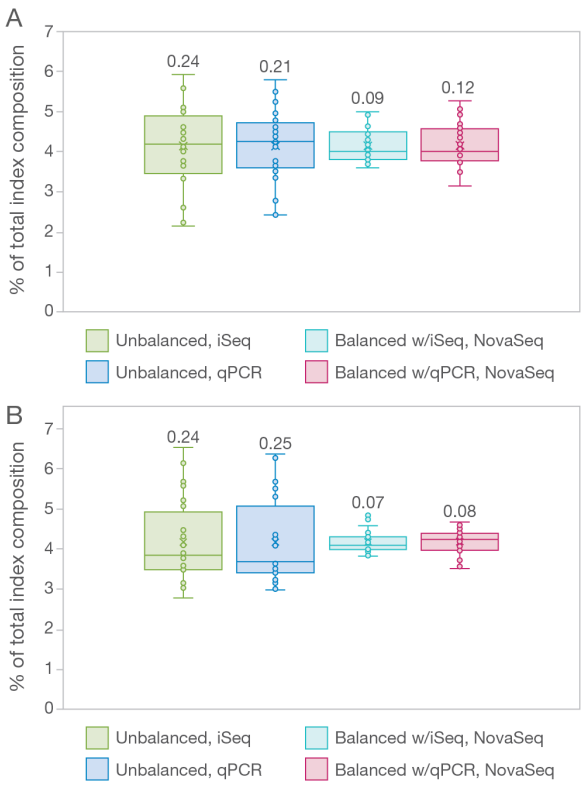

Figure 3: Comparison of library rebalancing with the iSeq 100 System and qPCR— Library rebalancing with the iSeq 100 System shows as good or better CVs (displayed above each data point), as compared to qPCR for two separate libraries (A, B).

Also, to assess measurement variability across the two methods, the same library was sequenced in triplicate on iSeq resulting in a demultiplexing CV of 0.67 as compared to 0.69 with triplicate qPCR measurements of the same sample (data not shown). These results, combined with setup time of approximately five minutes and automated index demulitplexing results generated with Local Run Manager, position the iSeq 100 System as an ideal method for library quantitation and rebalancing.

Index representation correlation between the iSeq 100 and NovaSeq 6000 Systems

One of the key advantages of using the iSeq 100 System for library QC is the use of proven Illumina sequencing by synthesis (SBS) chemistry, which enables data comparison across platforms.

Methods and results

To assess baseline correlation of index representation between the iSeq 100 and NovaSeq 6000 Systems, three independent 96-plex TruSeq DNA Nano libraries with IDT for Illumina TruSeq UD Indexes (96 indexes) were prepared. Each unbalanced library was sequenced on the iSeq 100 and NovaSeq 6000 Systems. Plotting demultiplexed values for each index shows significant correlation between the two systems across all three libraries (Figure 4). This high degree of correlation enables identification of drop-outs and significant outliers from a library pool before a NovaSeq run.

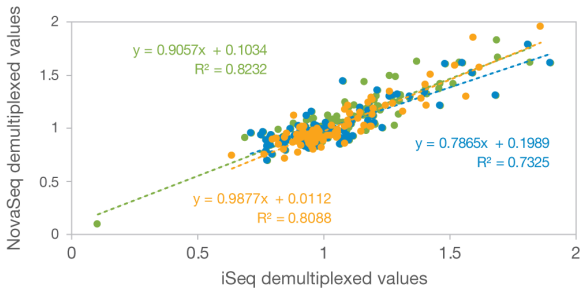

Figure 4: Baseline correlation of index representation— Index representation on the iSeq 100 System shows high correlation with the NovaSeq 6000 System across three 96-plex TruSeq DNA Nano libraries (green, orange, blue) with a mean R^2 of 0.79.

Known sources of variability that do not affect index representation

There is a consistent difference on an index by index basis between the iSeq 100 and NovaSeq 6000 Systems. Before these systematic differences can be characterized, it is important to understand what factors contribute to the variability of index representation.

- **Library input concentration**: Index representation does not depend on loading concentration.
- **Library pool context**: Index representation does not change depending on other indexes in the pool.
- **NovaSeq loading configuration**: The NovaSeq 6000 loading configuration (Xp workflow or standard) does not affect index representation.
- **Linearity**: The absolute representation level of an index does not impact the relative representation of that index.

Methods and results

Each factor was tested for potential impact on index representation with the iSeq 100 and NovaSeq 6000 Systems. Increasing the loading concentration of Nextera DNA Flex with Nextera CD indexes (Figure 5A) or TruSeq DNA Nano with IDT for Illumina-TruSeq UD indexes (Figure 5B) libraries across a broad concentration range had no effect on index representation on either system. To test the effect of library pool context, four 24-plex TruSeq DNA Nano library pools with unique index combinations were prepared. Index representation for a single pool remained unchanged whether it was combined with one or three additional 24-plex pools on both systems (Figure 5C, 5D).

The NovaSeq 6000 System offers two methods for flow cell loading: the NovaSeq Xp workflow or standard workflow. Demultiplexed values from both workflows show exceptional correlation, indicating there is no effect on index representation (Figure 5E). To test for linearity, two 24-plex pools were mixed at a 4:1 ratio and compared to a 1:1 ratio mix, showing no effect on index representation (Figure 5F). These results indicate that none of the factors tested impact index representation.

All libraries in this study were loaded at 100 pM for the iSeq 100 System, 220 pM for the NovaSeq Xp workflow, and 400 pM for the standard NovaSeq workflow. These loading concentrations resulted in acceptable primary metrics. As index representation has been shown to not be affected by input concentration (Figure 5), optimizing the loading concentration for each library was not required.

Using learned scaling factors to improve demultiplexing prediction

Given the significant correlation in sequencing results between the two platforms, and systematic differences on an index to index basis, demultiplexed values from a sequencing run on the iSeq 100 System can be used to predict demultiplexed values for the same library pools before a run on the NovaSeq 6000 System. Being able to accurately predict these values increases the number of samples for which minimum coverage can be guaranteed. This means more samples can be fit into a NovaSeq sequencing run, providing additional cost savings.

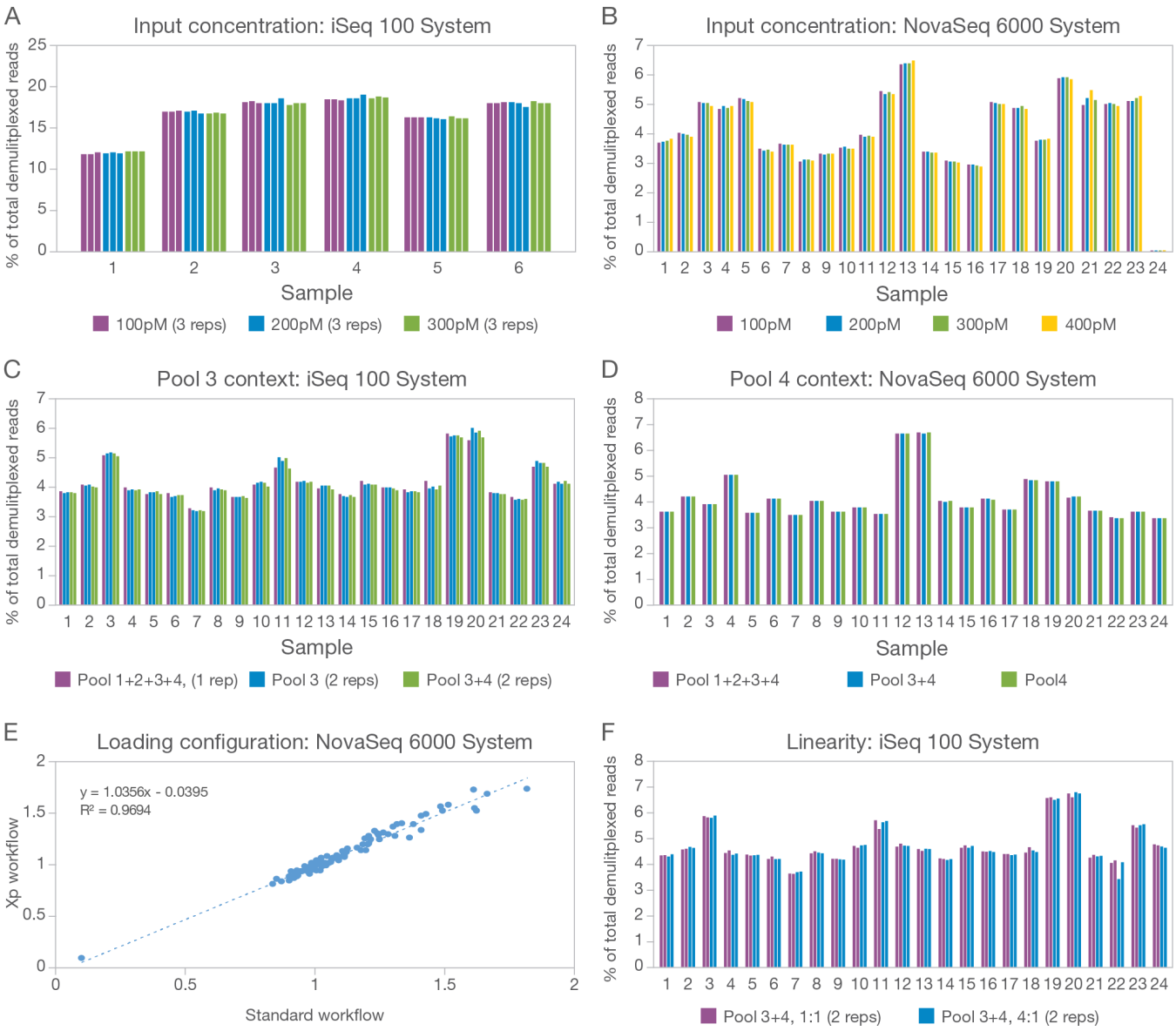

Figure 5: Index representation is unaffected by potential sources of variability— (A, B) Library input concentration, (C, D) library pool context, (E) NovaSeq loading configuration, and (F) linearity do not affect index representation on either the iSeq 100 or NovaSeq 6000 System. The number of replicates (reps) is indicated.

iSeq and NovaSeq demultiplexing values differ slightly, but they differ in a predictable way per index. Sequencing the same pool on both the iSeq 100 and NovaSeq 6000 Systems provides valuable information about the differences in representation between the two platforms. Specifically, an index scaling factor can be calculated by dividing a NovaSeq demultiplexing value for a specific index by the iSeq demultiplexing value for that same index. These index values are obtained by normalizing against the total amount of reads demultiplexed for the sequencing run, providing a more meaningful comparison between sequencing runs with different total demultiplexed values.

Baseline demultiplexed values from the iSeq 100 System can be used to predict the demulitplexed values for a NovaSeq 6000 run. However, multiplying these values by the index scaling factors computed from an initial iteration of a sequencing experiment (specific to sample type, library prep type, and index adapter combination) improves the R^2 correlation with the actual demultiplexed values obtained from sequencing the library on the NovaSeq 6000 System (Table 2).

To illustrate this process, if you consider the following:

r = iteration of sequencing experiment (specific to sample type, library prep type, and index adapter combination)

i = index

I_{ri} = iSeq demultiplex value

N_{ri} = Actual NovaSeq demultiplex value

N'_{ri} = Prediction of the NovaSeq demultiplex value

Then for the baseline procedure the iSeq demultiplex value gives the predicted NovaSeq demulitplex value. However, the iSeq demultiplex value can be multiplied by a scaling factor to predict the NovaSeq demultiplex value, as follows:

$$N'_{ri} = I_{ri} * f_{ri}$$

For the first sequencing experiment, (r=0), nothing is known about the scaling factors, so they are set to 1. After each subsequent sequencing experiment (r=1), the scaling factors can be updated in the following manner:

$$f_{(r+1),i} = (r^* f_{ri} + N_{ri}/I_{ri})/(r+1)$$

For example, determining the ratio of NovaSeq:iSeq demultiplexed values (eg, 1.112%/1.105% = 1.095 for index 29) indicates that iSeq demultiplexed values under- or over-represent indexes relative to the NovaSeq 6000 System (Figure 6). This ratio can be applied as a scaling factor to the demultiplexed values from a subsequent iSeq run of the same library type. Data comparison shows that N' (NovaSeq demultiplexed values predicted from iSeq values using scaling factors) correlates with N (actual NovaSeq demulitplexed values) better than I (iSeq demulitplexed values) (Figure 6).

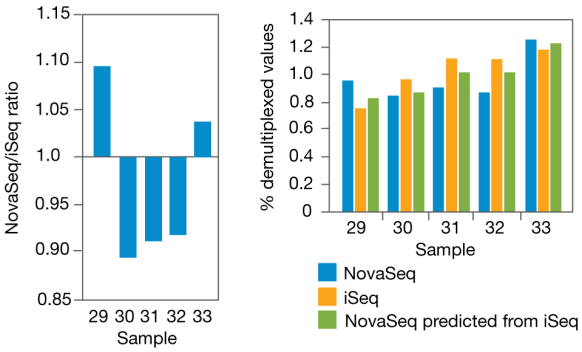

Figure 6: Scaling factors improve predicted NovaSeq demultiplexed values — Applying scaling factors to iSeq demulitplexed values improves prediction of NovaSeq demultiplexed values.

This process can be iterated, such that every subsequent replicate of the same sequencing experiment (same sample type, library prep type, and index adapter combination) provides another set of scaling factors. These factors can be averaged to give even better predictive performance, because noise from run to run variability is eliminated (Table 2 and Figure 7).

Table 2: Iteration of scaling factors improves prediction

	iSeq prediction vs NovaSeq, R^2 correlation		
Iteration	Prep A	Prep B	Prep C
0	0.73	0.82	0.81
1	0.89	0.87	0.91
2	0.93	0.92	0.94

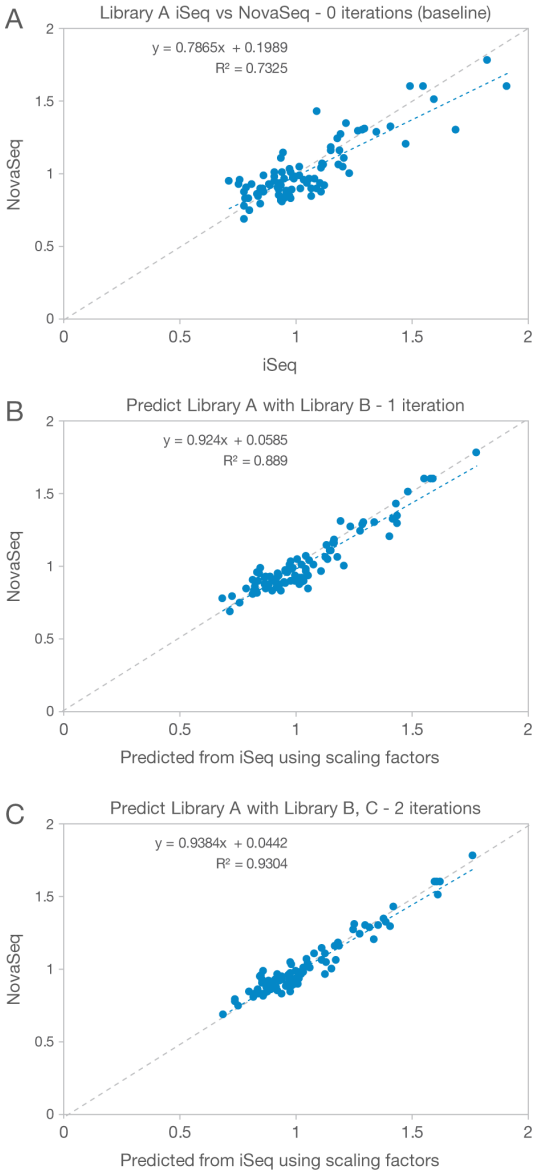

Figure 7: Iterative improvement in demultiplexing prediction — Correlation plots for (A) demultiplexing values from an iSeq run, and demultiplexing values for a NovaSeq run calculated from scaling factors from (B) one iSeq run and (C) averaged scaling factors from two iSeq runs against demultiplexing values from sequencing the same library on the NovaSeq 6000 System.

Summary

The iSeq 100 System enables library pool rebalancing before NovaSeq 6000 runs. The high correlation of index representation between the two platforms enables prediction of NovaSeq 6000 index representation for a given set of index pairs. Systematic index by index differences between the two platforms are informative, enabling calculation of scaling factors that can be used in an iterative fashion to improve the predictive power of index representation. These library QC functions empower users to increase the sample yield consistency of NovaSeq 6000 sequencing runs.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.