Importance of removing free adapters when sequencing on instruments that use patterned flow cells

Summary

- Libraries with higher levels of free adapters are prone to increased index hopping.
- Higher levels of free adapters can increase the level of adapter-dimer and chimeric reads observed in a sequencing run.
- Free adapters are a bigger problem with patterned flow cells.
- Invitrogen™ Collibri™ DNA library preparation kits include a library cleanup step to create libraries with minimum amounts of unligated adapters.

Introduction

Next-generation sequencing (NGS) has become a prominent tool in the investigation of diseases at the molecular level. The technology is based on millions of nucleic acid strands being read simultaneously and in parallel, one base at a time. Over the past 10 years, DNA sequencing instruments have evolved from a throughput of several megabases per day to a throughput of terabases per day. To achieve this massive increase in data output, one key change was made—instead of DNA being randomly oriented in a flow cell, it is added to predefined arrays, enabling known locations and dimensions. This change led to a major increase in output and a decrease in sequencing time, equivalent to approximately 6x the data output per unit time. The inclusion of unique index sequences during library construction also allowed multiple samples to be pooled and sequenced in a single run, further increasing throughput (Figure 1).

Historically, Illumina™ sequencers oriented single-stranded DNA in a random fashion on a flow cell and amplified the DNA using a technology known as bridge amplification [1]. This technology and its iterative improvements are used on many of the legacy Illumina sequencers, including the MiSeq™, NextSeq™, and HiSeq™ 2500 instruments. The adoption of patterned flow cells on the HiSeq™ 4000, HiSeq X, and NovaSeq™ 6000 systems necessitated a change in the isothermal amplification chemistry to a technology known as exclusion amplification (ExAmp) chemistry. The ExAmp chemistry enables isothermal amplification of unique DNA molecules in individual patterned flow cell wells [2].

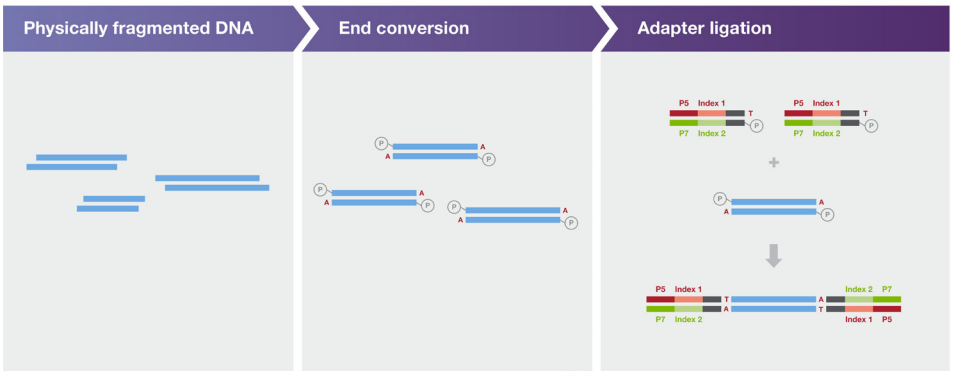

Figure 1. Overview of the Collibri PS DNA library preparation process.

ThermoFisher SCIENTIFIC

Soon after the introduction of this new amplification chemistry, the scientific community began to report the misassignment of reads in multiplexed libraries, resulting in a *Wired* article [3] and several extensive blog posts [4,5]. A phenomenon known as index hopping causes this misassignment and occurs when indexes are switched between DNA fragments of two different samples in the same sequencing run [6]. This switching can lead to data integrity issues, since sequences are assigned to the wrong samples. Depending on the library preparation approach, the rate of index hopping has been reported to range from about 0.3 to 3% [7].

The presence of unligated adapters—or, worse yet, adapter-dimers (two adapters ligated to each other)—in solution can also reduce the ability of the libraries of interest to be amplified on the patterned flow cell. The adapter-dimer contaminants contain all of the elements needed to bind and amplify on a patterned flow cell. These contaminants can reduce the output of the sequencing run and reduce the apparent quality of the sequencing data; these reads must be filtered out during the data analysis. Because of the numerous complicating factors, confirming the purity of the library prior to pooling and loading onto the flow cell is recommended.

Considerations

To reduce the possibility of misaligned reads through index hopping, the removal of unligated adapters using a kit such as one of the Collibri DNA library preparation kits (Figure 2), or another method, is important. Several complications can arise from unligated adapters or adapter-dimers.

The presence of these contaminating species in any samples being pooled together in a sequencing run can have multiple consequences. First, it can reduce the ability of the whole pooled library to attach and amplify on a patterned array. This effect is due to quantitation issues because these species can be present at a much higher concentration than the library elements; the higher concentration is a consequence of the mass difference between an adapter-dimer with a length of approximately 140 bp and a library element with a length of 450 bp. Another effect of contaminating species is that they can reduce the apparent quality metrics of the sequencing run. This effect is mainly caused by the sequencer lacking sufficient template to sequence and read a monotemplate, which the sequencer analyzes as exhibiting poor quality. Thus, if contamination is excessive, the whole sequencing run could fail or appear to have very poor quality.

Avoiding these situations altogether—that is, avoiding any contaminant species, either unligated adapters from the library preparation process or adapter-dimers that may have been created—is recommended.

If unligated adapters or adapter-dimers are apparent during the quantitation and QC of prepared libraries, similar to Figure 2A, an additional purification step should be added. Although this additional step will reduce the overall yield, it can save the library and mitigate the downstream issues highlighted here.

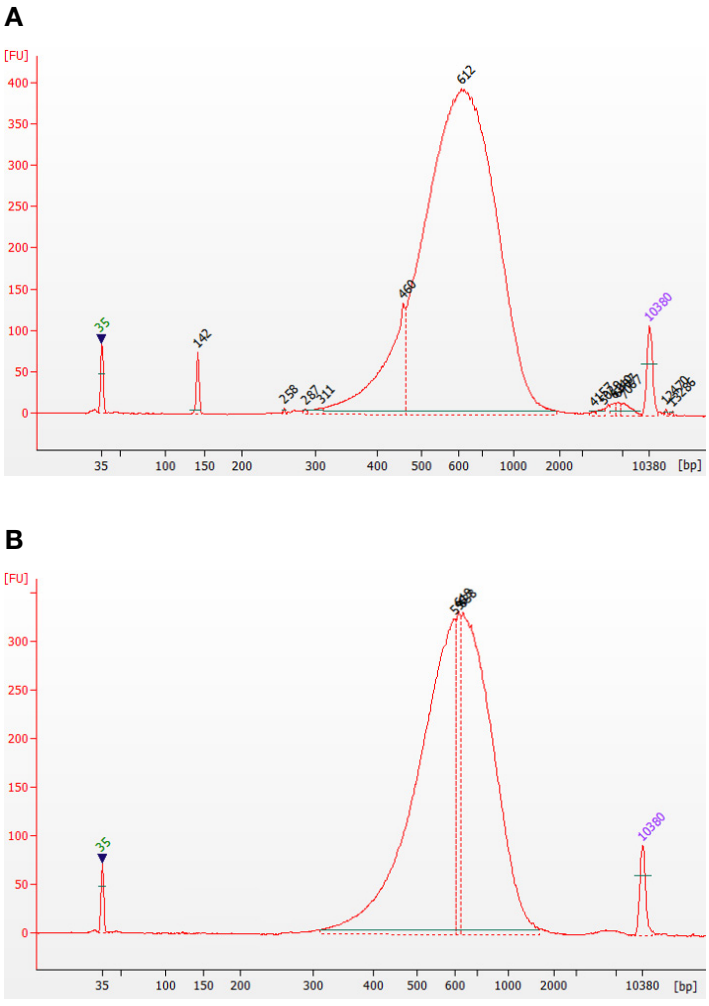

Figure 2. Traces of libraries prepared with a Collibri DNA library preparation kit, as recorded by a high-sensitivity Bioanalyzer™ system. Traces showing **(A)** libraries with spiked-in unligated adapters, which are present between the 35 bp marker (indicated in green) and the library, and **(B)** libraries prepared with the Collibri DNA library preparation kit, showing no detectable free adapters in the prepared libraries.

Conclusions

- The presence of unligated adapters can lead to increased index hopping on patterned flow cells.
- Collibri library preparation kits feature very efficient removal of unligated adapter and adapter-dimer contaminants.
- To minimize the possibility of misaligned reads due to index hopping, it is important to remove all detectable unligated adapters and to utilize unique dual indexes (UDIs) prior to loading samples onto patterned flow cells.

References

1. An introduction to next-generation sequencing technology. **https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf** (accessed July 2019).
2. ExAmp Cluster Amplification Workflow course. **https://support.illumina.com/content/dam/illumina-support/courses/examp-cluster-workflow/story_html5.html** (accessed July 2019).
3. The go-to gene sequencing machine with very strange results. **https://www.wired.com/2017/04/geneticists-fear-illuminas-sequencers-may-distort-results/** (accessed July 2019).
4. Index mis-assignment between samples on HiSeq 4000 and X-Ten. **http://enseqlopedia.com/2016/12/index-mis-assignment-between-samples-on-hiseq-4000-and-x-ten/** (accessed July 2019).
5. Mixing sample types in a flowcell lane generates cross contamination artefacts. **https://sequencing.qcfail.com/articles/mixing-sample-types-in-a-flowcell-lane-generates-cross-contamination-artefacts/** (accessed July 2019).
6. Minimize index hopping in multiplexed runs. **https://www.illumina.com/science/education/minimizing-index-hopping.html** (accessed July 2019).
7. Costello M, Fleharty M, Abreu J et al. (2018) Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 19:332.

Ordering information

Product		Quantity	Cat. No.
DNA-Seq kits for Illumina systems			
Collibri ES DNA Library Prep Kits	with CD Indexes	24 preps	A38605024
	with CD Indexes	96 preps	A38607096
	with UD Indexes, Set A (1-24)	24 preps	A38606024
	with UD Indexes, Set B (25-48)	24 preps	A43605024
	with UD Indexes, Set C (49-72)	24 preps	A43606024
	with UD Indexes, Set D (73-96)	24 preps	A43607024
Collibri PCR-Free ES DNA Library Prep Kits	with CD Indexes	24 preps	A38545024
	with CD Indexes	96 preps	A38603096
	with UD Indexes, Set A (1-24)	24 preps	A38602024
	with UD Indexes, Set B (25-48)	24 preps	A43602024
	with UD Indexes, Set C (49-72)	24 preps	A43603024
	with UD Indexes, Set D (73-96)	24 preps	A43604024
Collibri PS DNA Library Prep Kits	with CD Indexes	24 preps	A38612024
	with CD Indexes	96 preps	A38614096
	with UD Indexes, Set A (1-24)	24 preps	A38613024
	with UD Indexes, Set B (25-48)	24 preps	A43611024
	with UD Indexes, Set C (49-72)	24 preps	A43612024
	with UD Indexes, Set D (73-96)	24 preps	A43613024
	with UD Indexes, Set A-D (1-96)	96 preps	A38614196
Collibri PCR-Free PS DNA Library Prep Kits	with CD Indexes	24 preps	A38608024
	with CD Indexes	96 preps	A38610096
	with UD Indexes, Set A (1-24)	24 preps	A38609024
	with UD Indexes, Set B (25-48)	24 preps	A43608024
	with UD Indexes, Set C (49-72)	24 preps	A43609024
	with UD Indexes, Set D (73-96)	24 preps	A43610024
	with UD Indexes, Set A-D (1-96)	96 preps	A38615196

CD = combinatorial dual, UD = unique dual

invitrogen

Ordering information (continued)

Product	Quantity	Cat. No.
RNA-Seq kits for Illumina systems		
Collibri Stranded RNA Library Prep Kit for Illumina Systems	24 preps	A38994024
	96 preps	A38994096
Collibri Stranded RNA Library Prep Kit for Illumina Systems with H/M/R rRNA Depletion Kit	24 preps	A39003024
	96 preps	A39003096
ERCC RNA Spike-In Mix	1 kit	4456740
ERCC ExFold RNA Spike-In Mixes	1 kit	4456739
Library quantification		
Collibri Library Quantification Kit	100 rxns	A38524100
	500 rxns	A38524500
Qubit 4 Fluorometer, with WiFi	1 fluorometer	Q33238
Qubit 4 NGS Starter Kit, with WiFi	1 kit	Q33240
Library amplification		
Collibri Library Amplification Master Mix	50 rxns	A38539050
	250 rxns	A38539250
Collibri Library Amplification Master Mix with Primer Mix	50 rxns	A38540050
	250 rxns	A38540250

H/M/R = human/mouse/rat

Find out more at **thermofisher.com/collibri**

ThermoFisher SCIENTIFIC