How to prevent index hopping

Summary

- Index hopping causes misassignment of sequencing data from one sample to another
- Use of unique dual indexes (UDIs) mitigates concerns about index hopping
- Invitrogen™ Collibri™ DNA library preparation kits offer UDIs for sample multiplexing

Introduction

Next-generation sequencing (NGS) has become a prominent tool in the investigation of disease at the molecular level. The technology involves millions of nucleic acid strands being read in parallel, one base at a time. Depending on the method used, the DNA strand is read from one or both ends of the DNA molecule. Over the past 10 years, DNA sequencing systems 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.

Historically, Illumina™ sequencers oriented single-stranded DNA in a random fashion in 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 microwells in patterned flow cells [2].

Soon after introducing 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 the DNA index of one sample is switched with the index of another sample 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 0.3 to 3% [7].

Considerations

Three strategies have been developed to minimize the effect of index hopping: (1) using UDIs, (2) minimizing the amount of free, unligated adapter in the samples, and (3) pooling libraries together right before sequencing. The easiest and most relable solution is to ensure that all libraries in the sequencing pool contain UDIs. UDIs require that index 1 and index 2—the specific 8-base sequences used to distinguish the DNA sequence of a unique sample—are used only once per sample (Figure 1).

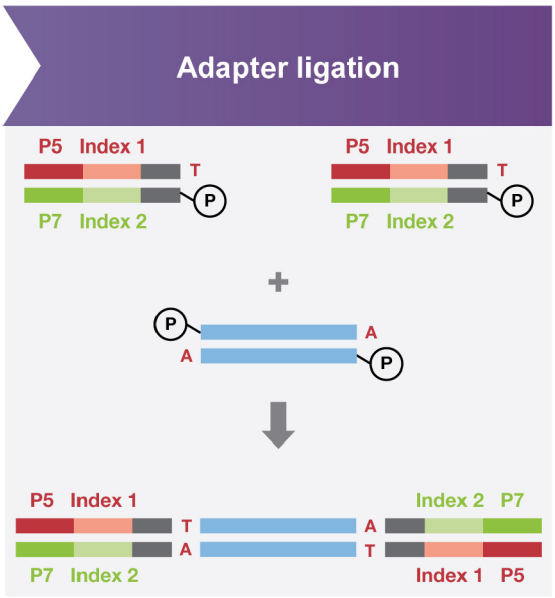

Figure 1. A sequencing library that includes the P5 and P7 sequences used to attach the DNA monotemplate in the flow cell and amplify it. The index 1 and index 2 sequences are 8-base sequences unique to each sample. To have a UDI, these two index sequences can only be used once during sample pooling, prior to loading the sequencer.

The second important strategy in minimizing index hopping is to reduce any unligated adapters that may be contaminating the sample prior to loading it onto the sequencer. In library preparation, a stoichiometric excess of adapter molecules relative to sample molecules is used to drive the ligation reaction as close to completion as possible. The issue with using a stoichiometric excess of adapter is that it leaves unligated adapters free in solution, and their complete removal can be difficult. The unligated adapters can be measured via agarose gel electrophoresis or with an Agilent™ Bioanalyzer™ system prior to sample pooling and the sequencing run.

One other consideration is that PCR-free libraries are known to be more susceptible to index hopping than PCR-amplified libraries [6]. The simplest hypothesis is that the inclusion of PCR results in an increased number of cleanup steps to remove the amplification primers, which, as an additional benefit, also remove any unligated adapter contaminants that may remain after the adapter ligation cleanup steps. These steps all occur prior to the pooling and loading of the sample into the flow cell, where the ExAmp reaction occurs.

Conclusions

- UDIs are recommended, especially when using PCR-free library preparation kits and patterned flow cells
- Collibri DNA library preparation kits support the inclusion of UDIs into libraries

References

1. An introduction to next-generation sequencing technology. **illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf** (accessed July 2019).
2. ExAMP Cluster Amplification Workflow course. **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. **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. **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. **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. **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 = human, M = mouse, R = rat

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

ThermoFisher SCIENTIFIC