Improved Sample Indexing for High Fidelity Demultiplexing to Increase Confidence in Low Frequency Variant Calling

Swift BIOSCIENCES™

Jonathan Irish, Jordan RoseFigura, Sukhinder Sandhu, Bita Lahann, Laurie Kurihara, Vladimir Makarov

Swift Biosciences Inc, 58 Parkland Plaza, Suite 100, Ann Arbor, MI 48103

Introduction

Due to recent scientific advances, there is a growing demand for new indices with high fidelity to eliminate errors associated with certain index combinations and sequencing chemistries. These errors are due to insufficient edit distances (the minimum number of changes required to transform one index sequence to the other) between index sequences or index hopping. Recent publications have highlighted how sequencing reads are misassigned due to "index hopping" on Illumina® patterned flow cells, such as the Novaseq™, which can generate billions of reads (>100 exomes) in a single run. This misassignment can lead to false positives in ultra-sensitive assays where low frequency variants and/or nucleic acid species are monitored. We have observed misassignment due to insufficient edit distance among Illumina TruSeq® HT i7 indices at a frequency up to 1.5%. Therefore, we developed 96 8nt i7 indices that can be paired with the existing TruSeq HT i5 indices to achieve 768 high throughput dual combinations, which were validated using a novel method on both Illumina's 2- and 4-channel technologies as both single and dual indices. This method involved the preparation of 96 libraries with unique, non-overlapping inserts to facilitate tracking of index misassignment. This allowed us to assess not only which index has misassigned library molecules, but also pinpoint the origin and rate of misassignment within a single run. With our 96 i7 indices, misassignment was observed at rates <0.1%. For even higher fidelity de-multiplexing, we have paired our 96 indices in a non-tandem manner for single use in both the i5 and i7 positions, known as Unique Dual Indices (UDIs). The use of UDIs further eliminates index read errors that misassign reads, enabling increased confidence in calling low frequency variants. UDIs can also eliminate PCR-induced chimerism that occurs during multiplexed post-hybridization PCR amplification, and can cause misassignment. Targeted sequencing, such as hybridization capture and amplicon technologies, enables low variant calling. Therefore, maximizing confidence during targeted sequencing experiments is of utmost importance. We have validated 96 new indices as UDIs for avoidance of index hopping and for eliminating PCR-induced chimerism during multiplexed library amplification.

Need for More Accurate Demultiplexing

Multiplexing relies on labeling genomic sequences from distinct samples with specific barcodes, also known as indices. The indices are short sequences, 6 to 8 nucleotides, that are incorporated into each DNA fragment during library preparation. This allows large numbers of libraries to be pooled and sequenced simultaneously during a single sequencing run. Gains in throughput come with an added layer of complexity, as sequencing reads from pooled libraries need to be accurately sorted and assigned to each sample in the pool, in a process called de-multiplexing.

Highly accurate de-multiplexing is imperative when:

- Detecting a low frequency nucleic acid, including SNV, indel, transcripts
- Pooling diverse samples for detection of events such as:
  - Germline mutations or amplification events vs. somatic mutation
  - Highly expressed vs. non-expressed

Sources of Read Misassignment

- **Sequencing errors**
  - Misassign one index read to another due to insufficient hamming distance or systematic errors intrinsic to each sequencing instrument
- **Index hopping**
  - Re-assigns an index on a library molecule on patterned flow cells due to indexed adapter carryover into the clustering reaction
- **PCR-induced chimerism**
  - Occurs during multiplexed library amplification due to incomplete extension and re-annealing or template switching
- **Sample cross-contamination**
  - Originates from errors in liquid handling or carryover from the previous sequencing runs

Solutions Offered by Swift Biosciences

The Accel-NGS® 2S Indexed Adapters are a unique set of 96 single indices optimized to work in combination with the Illumina TruSeq HT P5 adapter to provide up to 768 dual index combinations. Additionally, the unique design reduces misassignment and index hopping, an issue specific to Illumina's patterned flow cells, by incorporating an additional clean-up step. To provide greater sensitivity to detect low frequency variants and greater fidelity to accelerate demultiplexing during data analysis, Swift Biosciences offers two indexing strategies in combination with Accel-NGS 2S workflow for optimal sequencing performance:

I. 96 new single i7 indices compatible Illumina TruSeq HT i5 index sequences to provide up to 768 dual index combinations

II. 96 new index combinations for unique dual indexing (UDI) when each index is used only once and errors in index reads won't match another sample. Instead, they will be assigned to the undetermined bin.
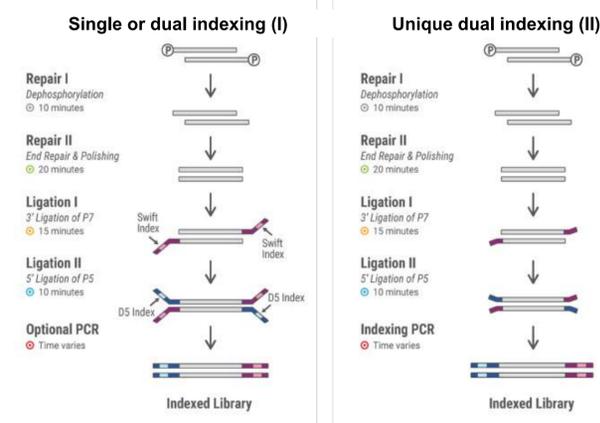

Table 1: Utility for addressing the sources of read misassignment

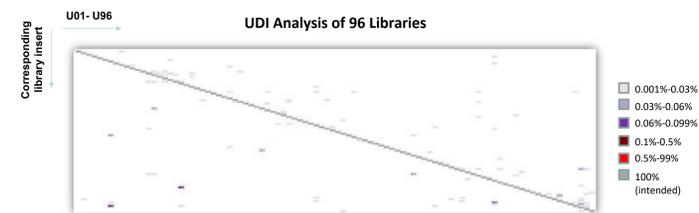| Workflow	Sequencing errors	Index hopping	PCR-induced chimerism	Sample Carryover
I	minimized	minimized	NA	Illumina sanctioned wash
II	eliminated	eliminated	eliminated	Illumina sanctioned wash

Examples of read misassignment in TruSeq HT indices

Reciprocal misassignment (>0.7%) occurred between D702 (TCCGGAGA) and D710 (TCCGCGAA) on both 2-channel and 4-channel instruments. Reads normalized to % of the Reads demultiplexed correctly.
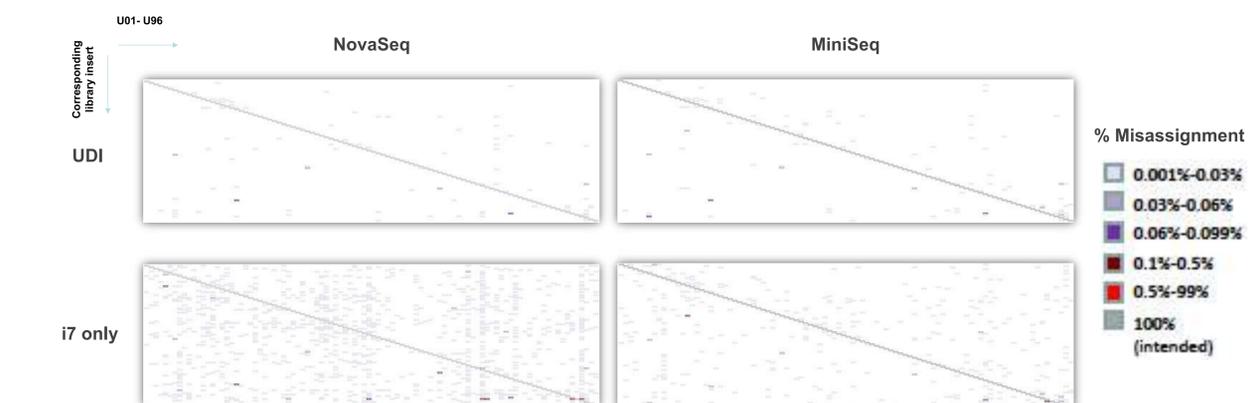

Results
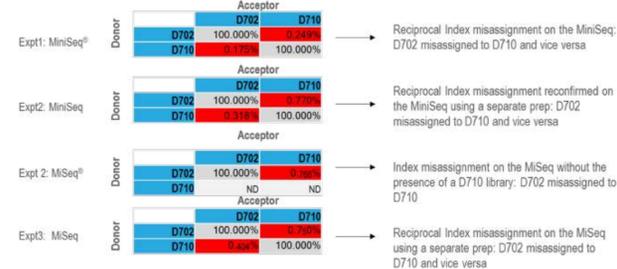
Validation Data

Unlike the Illumina HT indices, Swift i7 indices (S701-S796) do not exhibit index misassignment (one-way or reciprocal) greater than 0.1% with either the Illumina HT indices or with themselves on either a 2-channel instrument (shown right) or a 4-channel instrument (not shown). In addition, the Swift indices do not exhibit read misassignment in the i5 position (one-way or reciprocal) greater than 0.1% on either a 2-channel instrument or a 4-channel instrument (not shown). This has facilitated the development of unique dual indices (UDIs).

UDI Analysis of 96 Libraries

The use of the Swift indices, in a distinct manner in both the i7 and i5 positions to form Unique Dual Indices (UDIs), enables further elimination of index read errors that misassign reads. 96 UDIs were validated on both 2- and 4- channel chemistry and did not show read misassignment greater than 0.1%. (2-channel, MiniSeq shown Left)

To simulate PCR-induced chimerism, library pools of 12 and 24 libraries were subjected to 12 cycles of PCR. The resulting libraries were co-sequenced on a MiniSeq®. Demultiplexing with the UDIs maintained low levels of read misassignment, while single index (i7) analysis shows increased read misassignment (Right). The libraries only demonstrated read misassignment within the set that was co-amplified. The percent of reads that were undetermined are shown below.

% Undetermined		PCR-free	12 cycles PCR
	UDI	2.0%	3.1%
	i7 only	0.9%	0.7%

Libraries indexed with UDIs were constructed. The resulting libraries were co-sequenced on a MiniSeq and NovaSeq. De-multiplexing with the UDIs maintains low levels of read misassignment, while single index (P7) analysis shows increased read misassignment.

Conclusion

- 96 new single i7 indices were designed with an edit distance of 3 to demonstrate index read misassignment to < 0.1%
- These unique indices was validated for use as UDIs
- UDIs eliminate read misassignment due to PCR-induced chimerism and index hopping on patterned flow cells
- UDIs are commercially available and are compatible with the Accel-NGS 2S Plus, 1S Plus, and Methyl-Seq kits.

www.swiftbiosci.com