

Lexogen 12 nt Unique Dual Index System

- 384 Unique Dual Indices cover all barcoding needs in RNA-Seq
- Rescue of reads maximizes data output and saves sequencing costs
- Adjustable index read-out length accommodates each experiment scale
- Optimized distances translate into superior index error correction

Introduction

A critical consideration for any multiplexed RNA-Seq workflow is to avoid errors in the index read-out, which can result in the mis-assignment of sequencing reads to the wrong samples. While the majority of the raw reads will have the expected index combination (Fig. 1A), read mis-assignment can occur on all Illumina platforms. This happens due to two main events: Index Hopping and random Index Sequence Errors.

During Index Hopping an index sequence of one library is incorrectly added to another library, which may affect 0.1 – 2 % of all reads [1]. Only the use of Unique Dual Indexing (UDI), where each library in a given pool is barcoded with unique i7 and unique i5 index sequences, unambiguously identifies reads with hopped indices. Such reads are removed from downstream analysis and discarded (Fig. 1C).

Read mis-assignment due to random Index Sequence Errors occurs

when an error in one index sequence transforms the index into another one that is present within the same multiplexed sample pool. UDIs resolve such mis-assignment and the read is discarded.

More frequently, an Index Sequence Error results in an index sequence that does not match any other index in the pool, and the read is initially classified as undetermined. If the index sequence in question is different enough from the other index sequences in this pool, then error correction can be applied to recover a significant share of these reads (4 - 7 % of the initial reads, Fig. 1B). The performance of this error correction depends predominantly on the quality of the index design, as deficient index design can result in a higher rate of faulty error correction. Due to their unique design Lexogen's 12 nt UDIs minimize the impact of Index Sequence Errors and enable maximal data output gain by error correction.

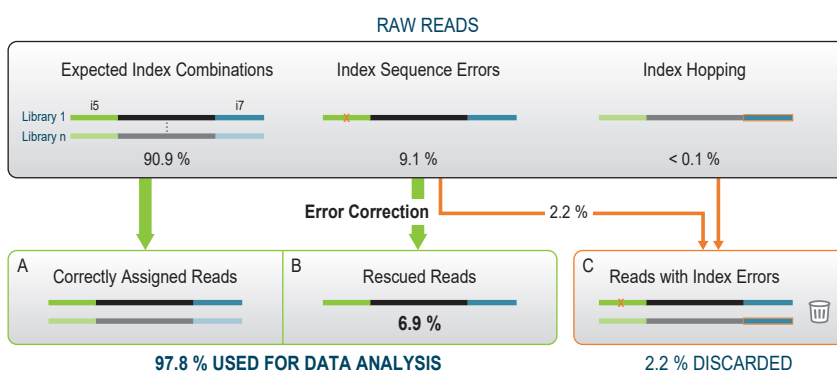


Figure 1 | The effects of Index Hopping and Index Sequence Errors in a pool of libraries with Unique Dual Indexing. Read mis-assignment caused by Index Hopping can be avoided by using Unique Dual Indexing (UDI). Reads with hopped indices are irreversibly discarded (C). Reads with random Index Sequence Errors resulting in an index not present in the pool are classified undetermined. Accurate error correction can rescue most of these reads making them available for downstream data analysis (B). The percentage values were derived from an RNA-Seq experiment pooling 96 libraries with Lexogen's 12 nt UDIs and full 12 nucleotide index read-out on an Illumina NextSeq500.

Superior Error Correction Maximizes Sequencing Yield

Lexogen's 12 nt UDIs are 12 nucleotides (nt) long and designed to maximize inter-index distance for different sample numbers and index read-out lengths. In a typical experiment using the full 12 nt index read-out around 9.1 % of the initial raw reads contain a random Index Sequence Error (Fig. 2A). This renders them undetermined, hence removing these reads from downstream analysis.

Lexogen's advanced index design enables the rescue of 76 % of these undetermined reads (6.9% of the initial reads), even if multiple nucleotides of the index contain errors. The useful output thereby increases to 97.8 % of the initial reads, an unprecedented performance due to the cutting-edge index design (Fig. 2B).



Figure 2 | Maximizing read output with Lexogen's 12 nt UDIs and error correction. 96 multiplexed libraries were sequenced on an Illumina NextSeq500 with 12 nt UDI read-out. A) In a standard RNA-Seq experiment a significant number of reads is undetermined (orange) due to random Index Sequence Errors. B) Lexogen's 12 nt Unique Dual Indices are optimized for maximal error correction with highest accuracy. Lexogen's Error Correction Tool allows almost 7% of originally undetermined reads to be confidently rescued and correctly assigned to the respective library.

Scalable Index Read-out Length

The design of Lexogen's 12 nt UDIs enables scalable read-out lengths of 12, 10, and 8 nucleotides. The UDIs therefore support all kinds of requirements for multiplexing, which depend on experiment type, sequencing equipment, desired read depth, and/or the number of pooled libraries. For small sample sizes (e.g., 24 samples) short indices (e.g., 8 or 10 nt) are sufficient to ensure high accuracy and reliable error correction. For more than 96 samples however, 8 nt index read-out does not allow reliable error correction, and 10 or 12 nt read-outs are required.

While needing slightly more sequencing cycles, 12 nt long index sequences also provide the ability to correct up to two (or in very small sets, three) Index Sequence Errors. Adjustable index read-

Summary

Lexogen's 12 nt UDI system adapts to the user's needs while always providing highest inter-index distance and maximal error correction capacity. Read mis-assignment due to Index Hopping is avoided, and Index Sequence Errors can be corrected with highest accuracy. Thereby, the system provides the optimized indexing solution for current and future barcoding requirements.

Subset	12 nt		10 nt		8 nt	
	D	ec	D	ec	D	ec
384	4	1	3	1	2	1*
↳ 96	5	2	4	1	3	1
↳ 24	6	2	5	2	4	1
↳ 4	7	3	6	2	5	2

Order Information

Lexogen's 12 nt UDIs are available as Add-on kits for QuantSeq, CORALL, SENSE and other library prep kits. Dedicated QuantSeq-FWD and CORALL kits including the 12 nt UDIs are also available:

Cat. No.	Lexogen UDI 12 nt Unique Dual Indexing Solutions	Illumina Instruments
107.96 – 110.96 113.96, 115.384 117.96, 119.384	Add-on Kits, Set A1, A2, A3, or A4 (UDI12A_0001-0384) QuantSeq 3' mRNA-Seq Library Prep Kit FWD with 12 nt UDI Sets A1 or A1-A4, (UDI12A_0001-0384) CORALL Total RNA-Seq Library Prep Kit with 12 nt UDI Sets A1 or A1-A4, (UDI12A_0001-0384)	NovaSeq™ 6000 HiSeq® 3000/4000 (SR) HiSeq® 2000/2500 MiSeq®
111.96 114.96 118.96	Add-on Kit, Set B1 (UDI12B_0001-0096) QuantSeq 3' mRNA-Seq Library Prep Kit FWD with 12 nt UDI Set B1, (UDI12B_0001-0096) CORALL Total RNA-Seq Library Prep Kit with 12 nt UDI Set B1, (UDI12B_0001-0096)	NextSeq™ 500/550 HiSeq® 3000/4000 (PE) MiniSeq™ 6000

References

[1] Illumina, Effects of Index Misassignment on Multiplexing and Downstream Analysis (2017) 770-2017-004-D.

out-length allows tuning your indexing needs to the experiment design, without the need to purchase separate indexing sets.

Nested Index Set Design for Highest Accuracy

To provide optimal index subsets for these varying multiplexing needs, Lexogen has designed the 12 nt UDIs in a nested approach. Small subsets benefit from Lexogen's nested index system by having the largest inter-index distance and highest error correction capacity while larger subsets provide for higher multiplexing needs.

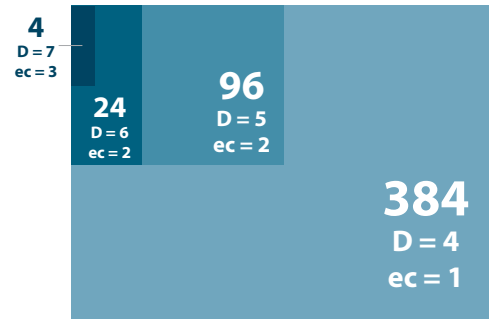


Figure 3 | Distance and error correction in Lexogen's nested 12 nt UDI sets. Illustration of inter-index distance (D) and number of possible error corrections (ec) in a nested index set with 12 nt read-out. An optimized set of 384 indices contains a subset of 96 indices with larger distances and enhanced error correction. Within these 96 indices, a subset of 24 indices is optimized even further, while a 4 index subset features the highest possible inter-index distance and error correction capacity.

Moreover, all subsets are nucleotide-balanced at each index position for optimized cluster identification in the NGS run. Using a proprietary algorithm, Lexogen has designed more than 9,216 UDIs (24x 384 subsets) with the capacity of correcting at least one error. Such sets with more than 384 UDIs are available upon request and enable extreme levels of multiplexing while still providing excellent error correction.

Table 1 | Comparison of distance and error correction capacity in Lexogen's nested 12 nt UDI sets with 8, 10, and 12 nt read-out. Inter-index distance (D) and number of errors that can be corrected (ec) are compared for subsets of 384, 96, 24, and 4 libraries and the three possible read-out lengths. The ec values represent the number of all errors (including substitutions, insertions, and deletions) that can be confidently corrected, except for *. In this case error correction can only address substitutions.

Find more about 12 nt UDIs at www.lexogen.com.

Contact us at info@lexogen.com, +43 1 345 1212-41, or find your local contact at www.lexogen.com/distributors.