xGen® Dual Index UMI Adapters resolve index hopping and enable low-frequency variant detection

M Light[1]*, M McNeill[2], K Lai[1], K Giorda[1], K Shah[1], S Groenewold[2], and M Jarosz[1]

1 Research & Development, Integrated DNA Technologies, Redwood City, CA
2 Research & Development, Integrated DNA Technologies, Coralville, IA

* Corresponding author: mlight@idtdna.com

Introduction

NGS adapters designed with unique molecular identifiers (UMIs)—in combination with unique, dual-index sequences—enable superior low-frequency variant detection in difficult-to-use and low-input samples. UMIs are random sequences within the adapter that uniquely tag individual molecules in a sample. Unique, dual-indexes are a combination of P5 and P7 index sequences that are both unique to a single sample. These adapters are compatible with commercially available library construction kits and can be used for PCR-free applications. They can be analyzed as single-index (1), dual-index (2), or dual-index with UMI (3) (Figure 1).

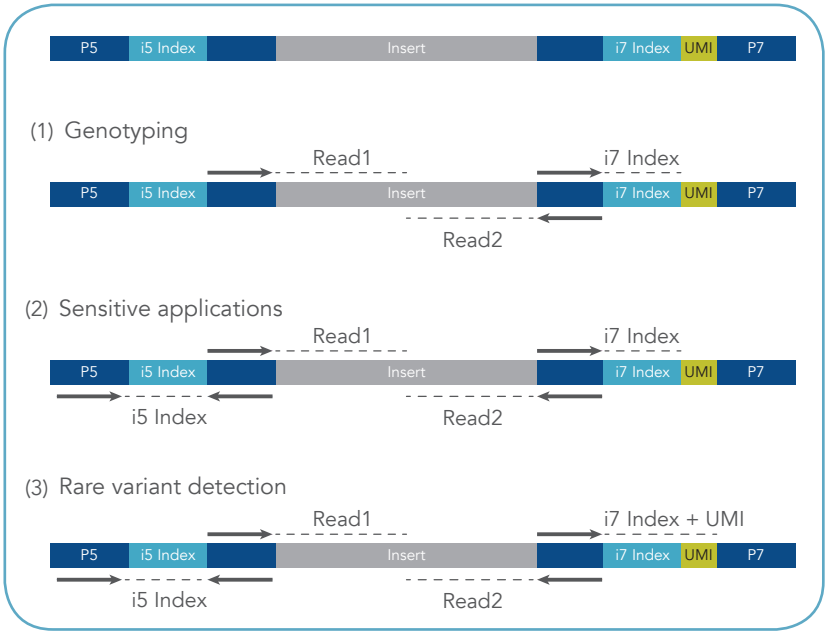

Figure 1. Unique, dual-index adapters with UMIs can be read in one of three modes, depending on the sensitivity requirements of the assay.

Technical review

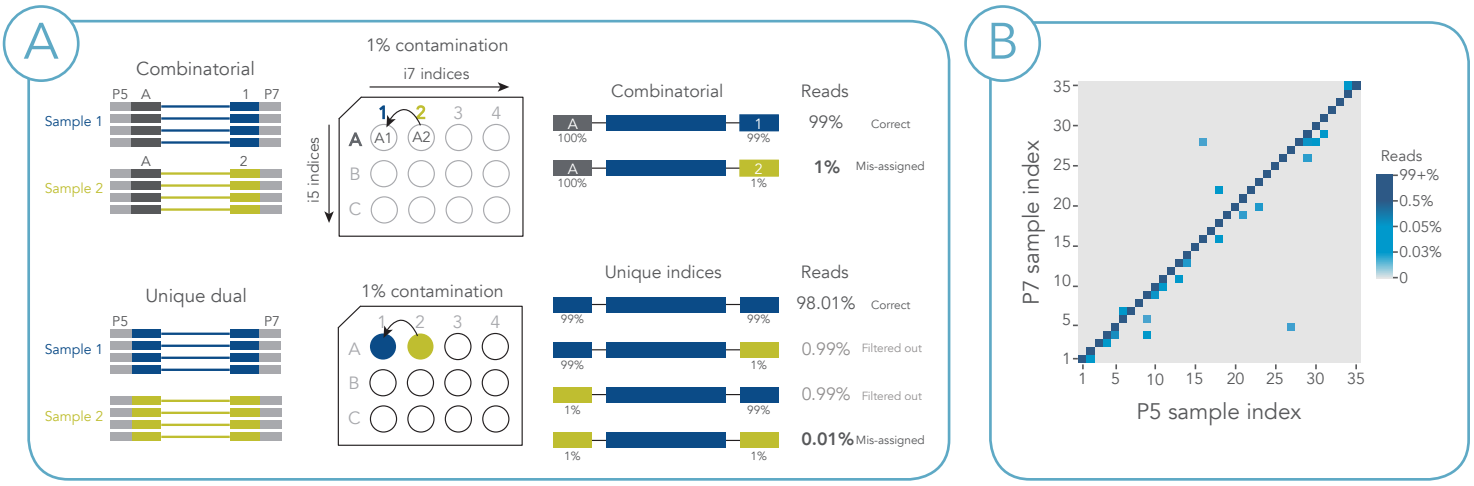

Figure 2. Unique dual indexes resolve barcode contamination. Sample misassignment can be caused by barcode contamination or index hopping during multiplexed target capture or sequencing. (**A**) Schematics show that when barcodes are used again within a set of samples (top), barcode contamination reads cannot be resolved. However, use of unique barcodes (bottom) prevents this problem. (**B**) Sequence data from PCR-free, whole genome libraries show that adapter contamination or index hopping are correctly filtered with dual-index adapters.

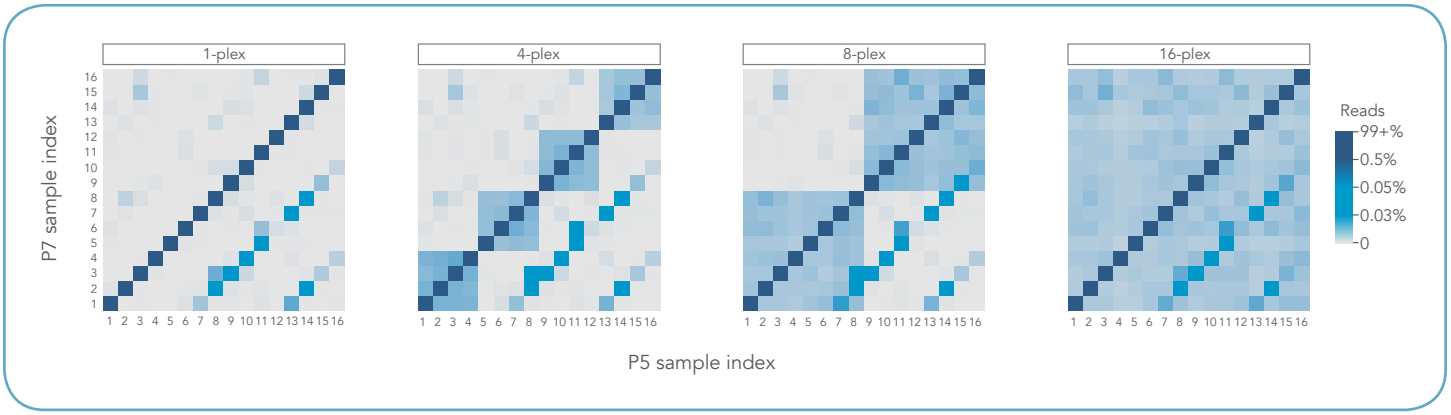

Figure 3. Unique dual indexes resolve index hopping. Index hopping may occur during multiplexed target capture or sequencing. Libraries shown were made from 250 ng of sheared gDNA and captured as 1-plex, 4-plex, 8-plex, or 16-plex captures using the xGen® AML Cancer Panel. Higher levels of multiplexing leads to increased rates of index hopping, which can be resolved with dual-index filtering during analysis.

Methods for variant detection study

Sample mixtures were created to mimic 1% minor allele frequency using well-characterized Genome in a Bottle (GIAB) cell lines or individually genotyped tumor-derived FFPE samples. Following library construction, singleplex hybrid capture was performed on all samples using a custom xGen Lockdown Panel that spans 75 KB and targets 288 common SNPs. Variants were called in high GIAB confidence regions using VarDict with start/stop deduplication or UMI consensus analysis.

Table1. Details on sample mixtures used for assessing adapter performance with cell-line and FFPE DNA.

Sample information	Cell-line	FFPE
Sample source	99% NA12878 / 1% NA24835	99% Breast / 1% Stomach
Number of alternate SNPs	10 (1%) / 44 (0.5%)	20 (1%) / 56 (0.5%)
Library input	25 ng	25 / 50 / 100 ng
Total number of reads	~18 million	12 / 20 / 38 million

Variant detection results with UMIs

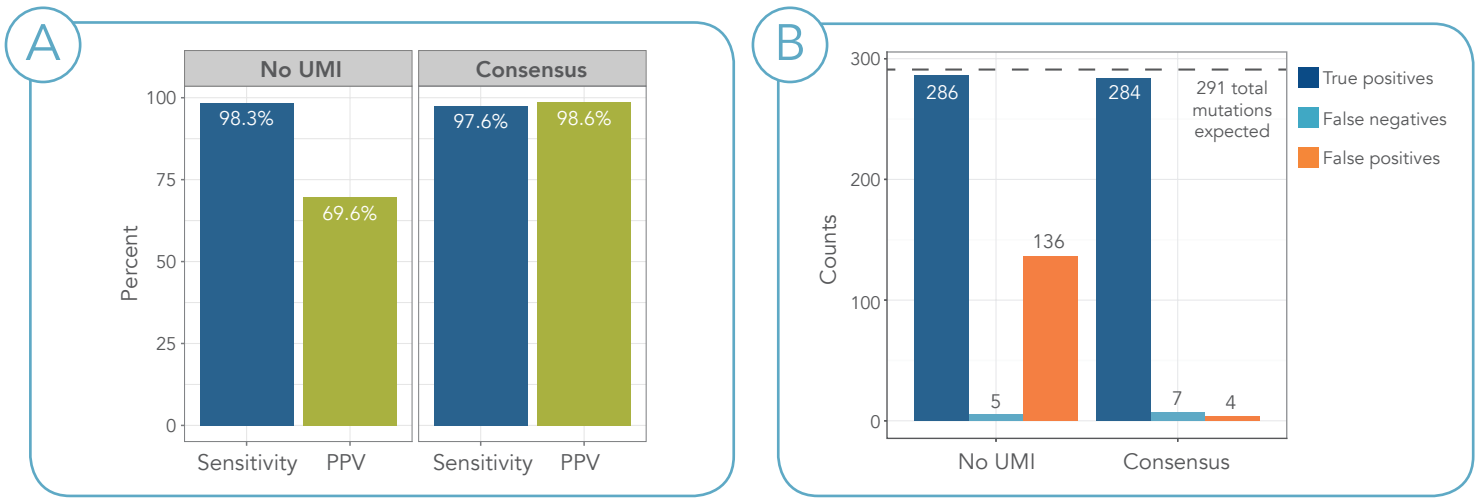

Figure 4. UMI error correction in cell-line samples. (**A**) UMI consensus calling improved PPV from 69.6% to 98.6% with minimal effect on sensitivity. (**B**) Total false-positive calls dropped from 136 to 4 with UMI consensus calling. (PPV = positive predictive value)

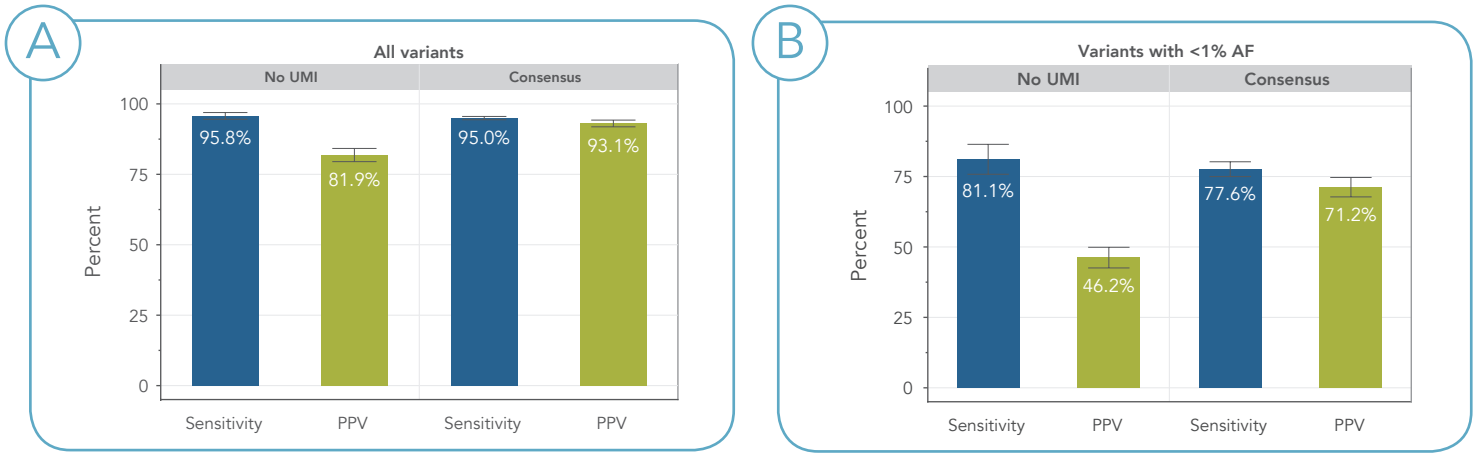

Figure 5. UMI error correction in FFPE samples. (**A**) Graph shows sensitivity and PPV using 0.6% min allele frequency for all mutations (n = 340). (**B**) Graph shows sensitivity and PPV using 0.6% min allele frequency for mutations present <1% (n = 76). (AF = allele frequency)

Coverage results with FFPE DNA

Libraries were generated using FFPE DNA isolated from breast cancer tumor tissue with inputs of 25 ng, 50 ng, and 100 ng. Following library construction, hybrid capture was performed on all samples using a custom xGen Lockdown Panel spanning 75 kb.

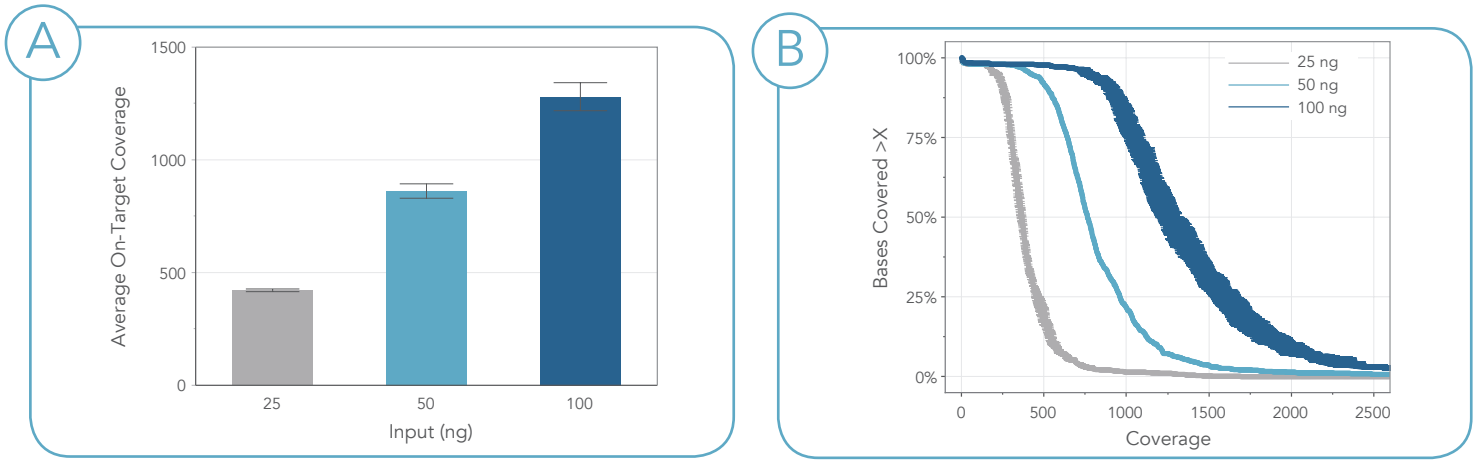

Figure 6. Deduplicated coverage of FFPE samples. (**A**) Mean deduplicated coverage is shown for 25 ng, 50 ng, and 100 ng inputs. Error bars reflect standard deviation of replicates (n = 3). (**B**) Mean cumulative coverage by sample input is shown. Error bars on x-axis reflect standard deviation of coverage.

Coverage results with cell-free DNA

Libraries were generated using cell-free DNA isolated from healthy donors with inputs of 5 ng, 10 ng, and 25 ng. Following library construction, all samples underwent hybrid capture using a custom xGen Lockdown Panel spanning 75 kb.

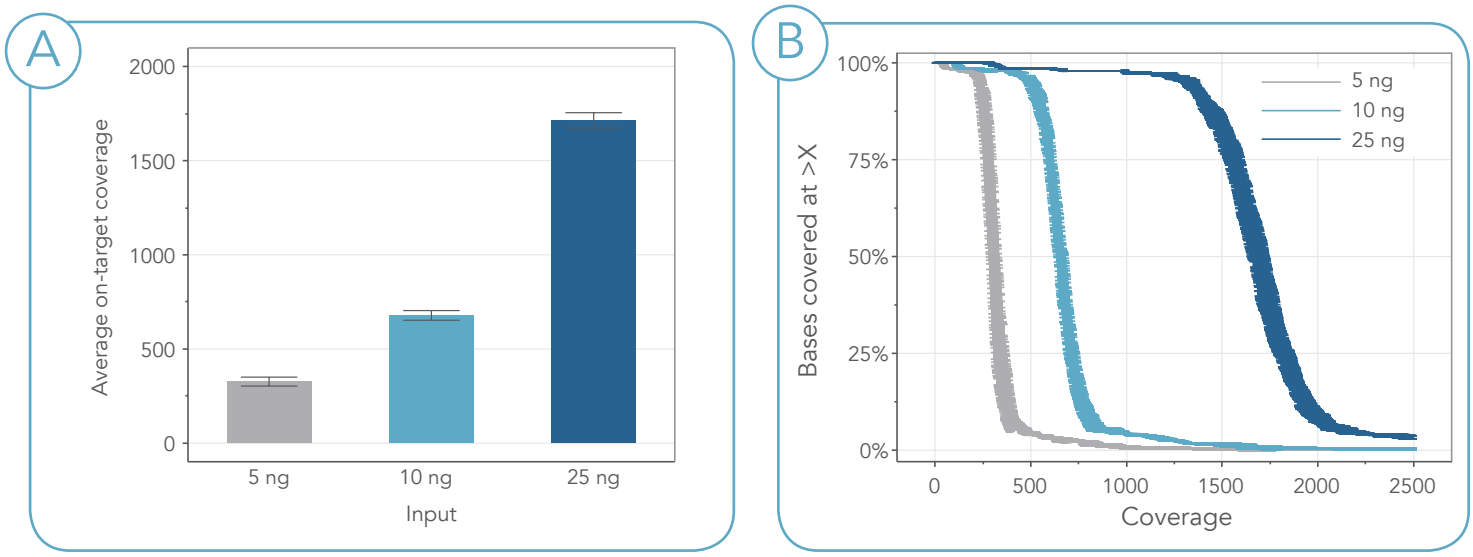

Figure 7. Deduplicated coverage of cell-free samples. (**A**) Mean deduplicated coverage is shown for 5 ng,10 ng, and 25 ng inputs using 12 M, 20 M, and 38 M total reads, respectively. Error bars reflect standard deviation of replicates (n = 3). (**B**) Graph shows the mean cumulative coverage by sample input. Error bars on x-axis reflect standard deviation of coverage. The percent of targets covered remained high at 100X, 300X and 500X, reflecting good uniformity of coverage.

Conclusions

- Unique dual-index adapters accurately flag barcode contamination and index hopping, and are required to accurately assign rare variants in multiplexed studies.
- Building consensus sequences enables *in silico* error correction, dramatically increasing variant calling specificity.
- Dual-index with UMI error correction be can successfully applied to a variety of sample inputs and difficult-to-use samples, including FFPE and cell-free DNA.

Find more information on xGen Dual Index UMI Adapters at **www.idtdna.com/UMI-techaccess**.