technical note

Multiplexed hybrid capture for whole exome sequencing

Optimal performance and cost efficiency using IDT's exome panel and Illumina's library prep kits

Allen Nguyen, Steve Groenewold, Layla Katiraee, Hana Johannesen
Integrated DNA Technologies, 1710 Commercial Park, Coralville, IA, 52241, USA

Abstract

Whole exome sequencing (WES) using next generation sequencing (NGS) empowers researchers to interrogate the entire exome to potentially reveal important variations of a genome. A higher level of sample multiplexing is enabled with WES compared to whole genome sequencing (WGS) to help maximize throughput while reducing cost per sample. Here, we tested the robustness of a combined Illumina library prep kit and IDT exome panel hybrid capture system, comparing various levels of library precapture multiplexing from one to 12 libraries. We found no significant differences in the NGS data metrics between the various levels of multiplexing during the capture. These results indicate that the combined IDT–Illumina exome solution accommodates a wide range of exome projects for optimum throughput and cost efficiency without sacrificing data quality.

Introduction

Exome sequencing, also known as whole exome sequencing, has become an established, cost-effective alternative to whole genome sequencing for identifying variants in a wide range of applications from population genetics to cancer studies. Exome sequencing has the advantage of focusing distinctly on coding regions in the genome (representing ~2% of total genome content, but containing ~85% of known disease-causing mutations) to detect highly relevant, disease-associated variants at a fraction of the sequencing output needed for WGS.

Because of the reduced sequencing output required for WES compared to WGS, researchers are able to increase efficiency and throughput by processing multiple samples in a single run (multiplexing). This practice, now routine for many labs, requires the use of indexing (barcoding) samples before sequencing, so that each read is bioinformatically assigned to the appropriate sample before downstream analysis.

Multiplexing can be performed before (precapture), or after (postcapture), hybrid capture for targeted, exome sequencing. Postcapture multiplexing has become standard with exome sequencing; however, performing precapture multiplexing (Figure 1) can be advantageous because it increases efficiency of the capture while conserving enrichment reagents.

See what more we can do for **you** at **www.idtdna.com.**

Precapture multiplexing is only effective when the combination of library preparation and target enrichment—specifically the DNA probes used—is robust enough to support pooling multiple samples in a single capture reaction without adversely affecting the results.

IDT and Illumina co-market a comprehensive workflow for exome sequencing. This workflow couples the Illumina TruSeq® and Nextera® library preparation kits with the IDT xGen® Exome Research Panel. Here, we present performance data for the combined workflow, comparing precapture multiplexing results with up to 12 samples per pool (12-plex).

Figure 1. Illustration of the pre capture multiplexing method before NGS. Libraries are pooled at equal mass before target enrichment to conserve reagents while increasing workflow efficiency.

Materials and methods

Libraries were prepared using the TruSeq and TruSeq Rapid Exome Kits (now TruSeq and Nextera DNA Library Prep for Enrichment Kits, respectively) with NA12878 genomic DNA (Coriell Institute). The protocol from these kits was followed, including the suggested size for DNA fragmentation in the TruSeq workflow. However, the final elution volume after PCR was increased to 20 µL to allow for enough material for quantification and size verification.

Replicate multiplexed pools of 4, 8, and 12 samples were created using 500 ng of each indexed library, for both TruSeq and Nextera samples. Individual libraries were included in the study as non-multiplexed controls. These pools were mixed with 2 µL of xGen Universal Blockers—TS Mix or xGen Universal Blockers—NXT Mix, for the TruSeq and Nextera pools, respectively. Multiplexed pools and blockers were dried using a vacuum concentrator.

Samples were hybridized with the xGen Exome Research Panel at 65°C for 4 hours. The IDT protocol for capture and washing of libraries was performed using the xGen-specific modules developed on the SciClone® G3 NGS Workstation (Perkin Elmer). These automation scripts were developed in partnership with Perkin Elmer and have been qualified on the SciClone workstation.

Streptavidin bead-captured libraries were PCR-amplified with the number of PCR cycles titrated according to the number of samples in the multiplexed pool. Pools with 12, 8, 4, and 1 sample were amplified by 5, 6, 7, and 9 PCR cycles, respectively. The amplified material was purified using AMPure® XP beads (Beckman Coulter).

Analysis

BWA v0.7.15 (Heng Li) was used to align a subsample of 50-million paired-end reads from each sample to an hg19 reference genome. After removing PCR duplicates, secondary alignments, and supplemental alignments

with Picard v2.9.0 (Broad Institute) processing software, SAMTools v1.3.1 (Heng Li) was used to count the number of remaining reads in the primary alignment BAM file as well as one intersected with the IDT xGen Exome Targets BED file (BEDTools v2.25, Aaron Quinlan), to calculate the on-target read percentage. A 150 bp pad on both sides of each target was used for the flanked on-target percentage. BEDTools v2.25 was also used to calculate per-base coverage depths and panel-wide mean coverage depth.

Results

Ninety-two exomes, multiplexed in pools of 1, 4, 8, or 12 samples, and captured with the xGen Exome Research Panel, were sequenced on the NextSeq 550 sequencer. A combination of up to 14 exomes were sequenced per run and normalized to 50-million paired-end reads per sample. The primary metrics examined were coverage and percent of reads on-target (Figure 2). Dunnett's analysis of means was performed using non-multiplexed samples as the control ($p<0.05$). The data suggests that there is no significant difference between 1, 4, 8, and 12-plexed samples in the flanked on-target percentage or mean target coverage depth.
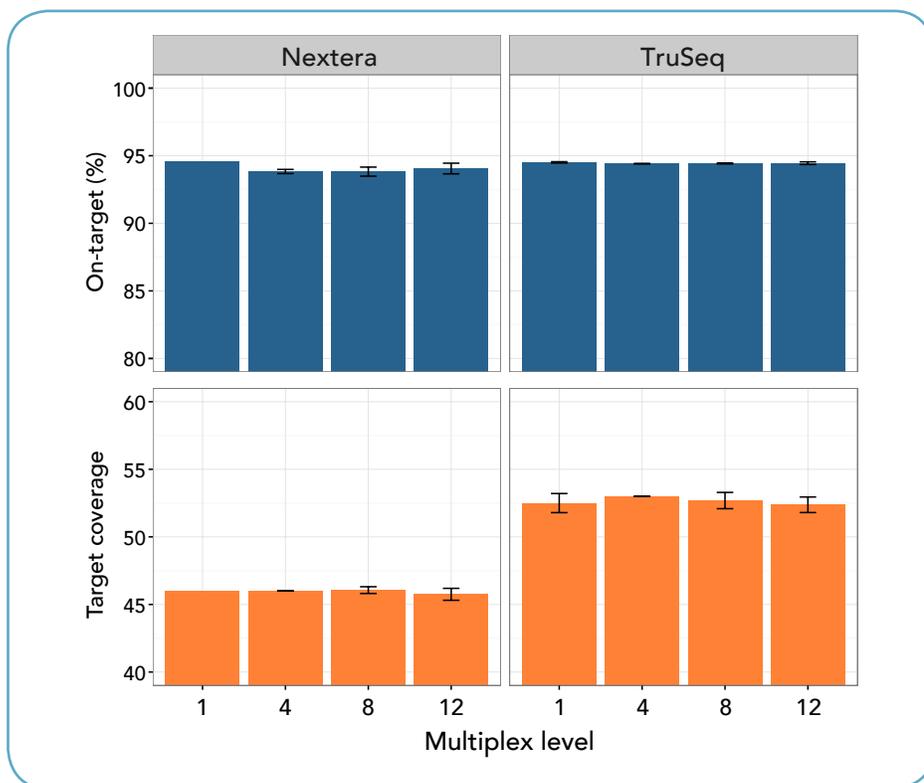

Figure 2. Consistent on-target rates and coverage with increased levels of pre capture multiplexing. The Nextera and TruSeq libraries have been enriched with the xGen Exome Research Panel, from a singleplex to a 12-plex capture.

Uniformity was not adversely affected by the higher levels of plexity, as observed in this study (Figure 3). We also found no significant increases in duplicate rate, in either TruSeq or Nextera libraries, as a function of increasing multiplex level (data not shown). The data suggest that users can confidently multiplex up to 12 samples in a single capture, with no significant impact on these metrics. The results highlight the robustness of the combined Illumina-IDT exome workflow supported by the automation protocol performed on the Perkin Elmer SciClone system. Variability between replicates within the pools, and from pool to pool, was minimal.
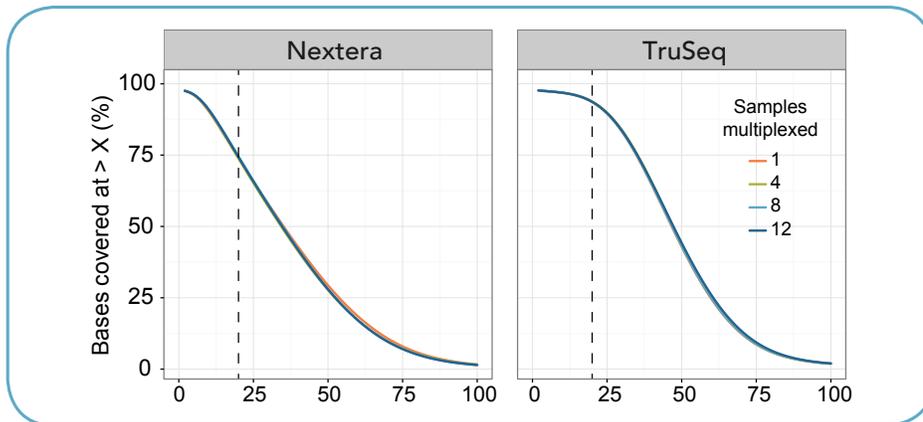
Figure 3. Coverage plots for Nextera and TruSeq libraries. Similar coverage curves were observed across the different levels of sample multiplexing within Nextera or TruSeq library samples before capture, suggesting that combining samples pre capture had minimum effect on overall uniformity.

Conclusion

Precapture multiplexing is a useful technique to increase the efficiency of NGS hybridization capture experiments by reducing reagent consumption and increasing throughput. Here, we demonstrated the ability to multiplex up to 12 libraries in a single exome capture reaction with no significant loss of data quality using the combined IDT–Illumina exome workflow.

Although previous contradicting studies have shown a reduction in capture efficiency with pre-enrichment multiplexing [1], we found no statistically significant difference in coverage or bases on-target with increased level of multiplexing. We attribute the different conclusions to the overall quality of our system, including the TruSeq and Nextera library prep kits (Illumina) and the xGen Exome Research Panel (IDT) enrichment kits and protocols.

Based on our testing, precapture multiplexing allows for significant cost savings in whole exome sequencing.

References

1. Shearer AE, Hildebrand MS, et al. (2012) **Pre-capture multiplexing improves efficiency and cost-effectiveness of targeted genomic enrichment**. BMC Genomics 2012, 13:618.

Products used in this technical note

xGen Exome Research Panel

xGen Universal Blockers—TS Mix

xGen Universal Blockers—NXT Mix

TruSeq DNA Library Prep for Enrichment

Nextera DNA Library Prep for Enrichment

IDT for Illumina—TruSeq DNA UD Indexes

See what more we can do for **you** at **www.idtdna.com.**