Whole-exome sequencing using the xGen™ DNA Library Preparation Kit family enables customizable workflows returning quality in consistent results

Abstract

The xGen DNA Library Preparation Kit family provides flexibility by utilizing enzymatic fragmentation (EZ) or mechanical shearing (MC) methods that support multiple indexing strategies and generates libraries suitable for PCR-free or PCR-amplified workflows. Here, we show that use of both xGen DNA Library Preparation kit workflows maintains sequencing performance when used with xGen Exome Hybridization Panel v2 while offering additional flexibility in multiplexing. These results demonstrate that versatile whole exome capture workflows provide high on-target and low duplication rates, and deep, uniform coverage across all regions, regardless of GC-content.

Introduction

Exome sequencing is invaluable for sequencing only the protein-coding regions of the human genome. This method is performed using hybridization capture—a technique that uses biotin-modified oligonucleotide probes to "capture" the region of interest for sequencing. Exome capture provides significant sequencing advantages over whole-genome sequencing. Focusing only on the protein-coding exons lowers the cost and time of sequencing, as exons make up approximately 1% of the genome. In contrast to their small size contribution to the genome, exons contain 85% of variants associated with disease [1].

Whole exome sequencing (WES) is a practical method for mapping variants that are rare in the population to better understand complex disorders [2]. It is a feasible option for population genetics and discovery science, or data mining, when searching for associations, or linking genes to phenotypes [3]. WES is also particularly useful in oncology studies, enabling researchers to identify germline and somatic variants within coding regions of genes.

For applications requiring high sensitivity, only captured libraries suitable for sequencing at a high coverage depth and uniformity should be used. At the same time, it is understood that researchers must balance cost and logistical considerations. In particular, multiplexing many libraries within a single hybridization capture will keep reagent costs down. If high sequencing coverage depth and uniformity can be maintained, a flexible multiplexing strategy will empower researchers to process more samples and sequence samples deeper—critical to variant calling and scaling large studies.

The xGen Exome Hybridization Panel v2 consists of 415,115 xGen Hyb Capture Probes that are 5′ biotin-modified and target only the coding sequences (CDS) of human coding genes in the RefSeq 109 annotation. Our protocol supports multiplexing up 16 samples, enabling researchers to process more samples without sacrificing data quality.

Flexible, comprehensive DNA library preparation workflows

A customizable library preparation system unlocks many benefits for exome researchers. For example, mechanical shearing can give users more control over fragmentation size, while enzymatic fragmentation may be easier to perform using standard laboratory equipment. Choosing PCR-free libraries can eliminate bias introduced through PCR amplification and provide more accurate sequencing data for highly sensitive applications.

The family of xGen DNA Library Prep kits for library generation give researchers complete control over library workflow design. Features that enable flexible workflows include:

- Choice of DNA fragmentation method: enzymatic (EZ), or mechanical (MC)
- Wide range of DNA input amounts into library prep:
  - xGen DNA Library Prep EZ supports 100 pg to 1 µg input
  - xGen DNA Library Prep MC supports 1 ng to 1 µg input
- Choice of xGen index adapter type: xGen Stubby Adapter, or xGen Full-Length (UNI) Adapter
- Up to 1536 Illumina® bar code options of 10 nt Unique Dual Index (UDI) that are stocked in single-use plates*
- Option to perform PCR amplification or create PCR-free libraries
- Compatible with our Normalase™ Module
- Ability to automate with automation scripts available on multiple platforms, including Sciclone® G3 NGS Workstation (Perkin Elmer) and other automation protocols**

* Custom barcodes, bulk ordering, and adapters for other sequencing platforms are available by custom order.

** IDT's Automation Team is available to assist on-site with automation script development, specifically tailored to your needs.

In this application note, we compare libraries prepared from both xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI Kits. After adding indexed libraries at differing plexity capture levels using xGen Exome Hybridization Panel v2 and then performing whole-exome sequencing, results were assessed with key NGS performance metrics.

Methods

Library preparation

Using 100 ng of human genomic DNA (Coriell NA 12878), twelve libraries each were prepared with either the xGen DNA Library Prep EZ Kit (Figure 1) or the xGen DNA Library Prep MC UNI Kit (Figure 2) protocol. High-quality gDNA was used as input for xGen DNA Library Prep EZ libraries, while gDNA mechanically sheared to 200 bp was used as input for xGen DNA Library Prep MC UNI libraries; both kits use whole or fragmented DNA as input. For xGen DNA Library Prep EZ libraries, stubby adapter was used followed by xGen Normalase™ Unique Dual Indexing primers (N-UDIs) with 6 cycles of PCR. For xGen DNA Library Prep MC UNI libraries, xGen UDI-UMI Full-Length Adapters and p5/p7 amplification primers were used for 6 cycles of PCR for library amplification. All libraries were quantified using Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific).

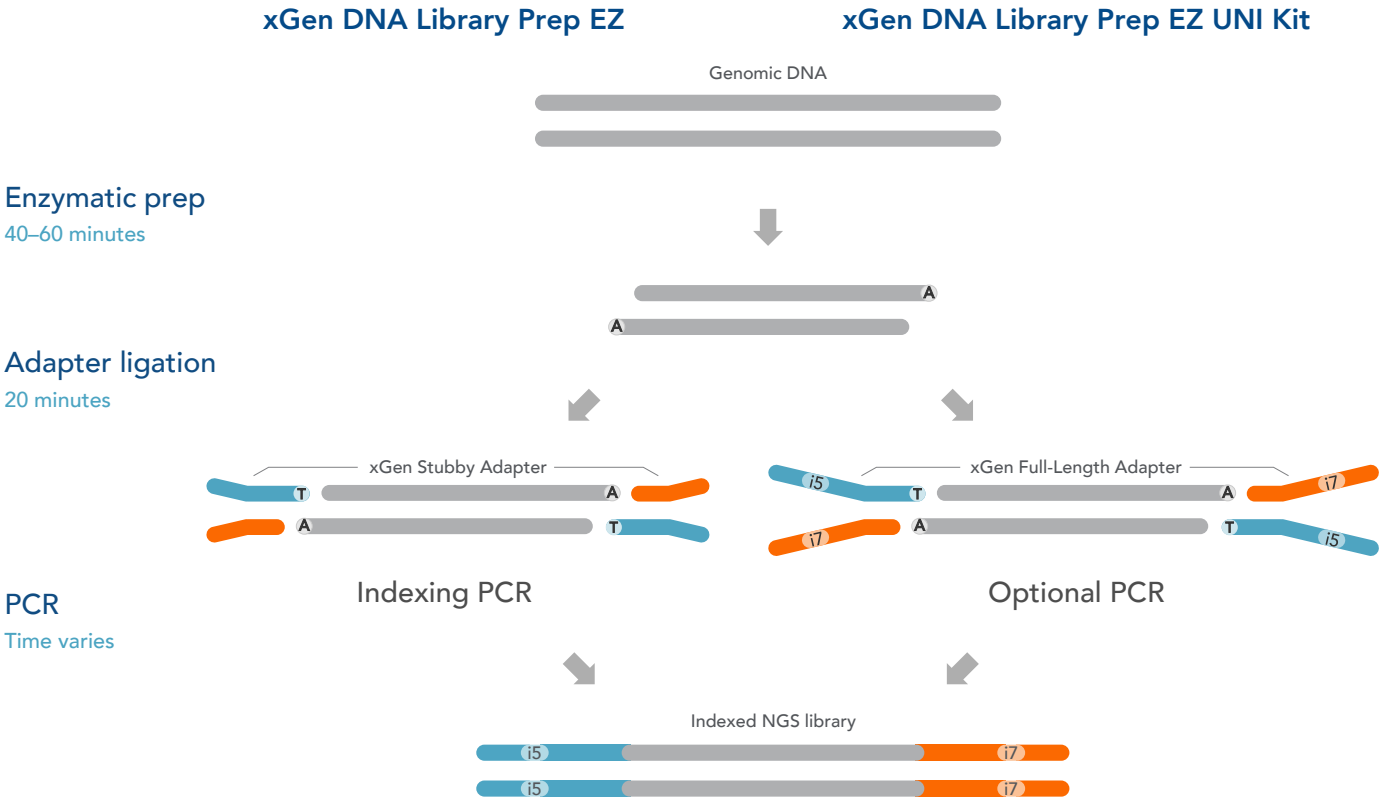

Figure 1. Workflow for the xGen DNA Library Prep EZ Kits. xGen DNA Library Prep EZ Kits are compatible with enzymatic fragmentation and use either an indexing by PCR workflow using xGen Stubby Adapters included in the kit and indexing primers supplied separately (left) or an indexing by ligation workflow using full length, indexed Y adapters (right). Both kits include amplification reagents for Indexing PCR and library amplification.

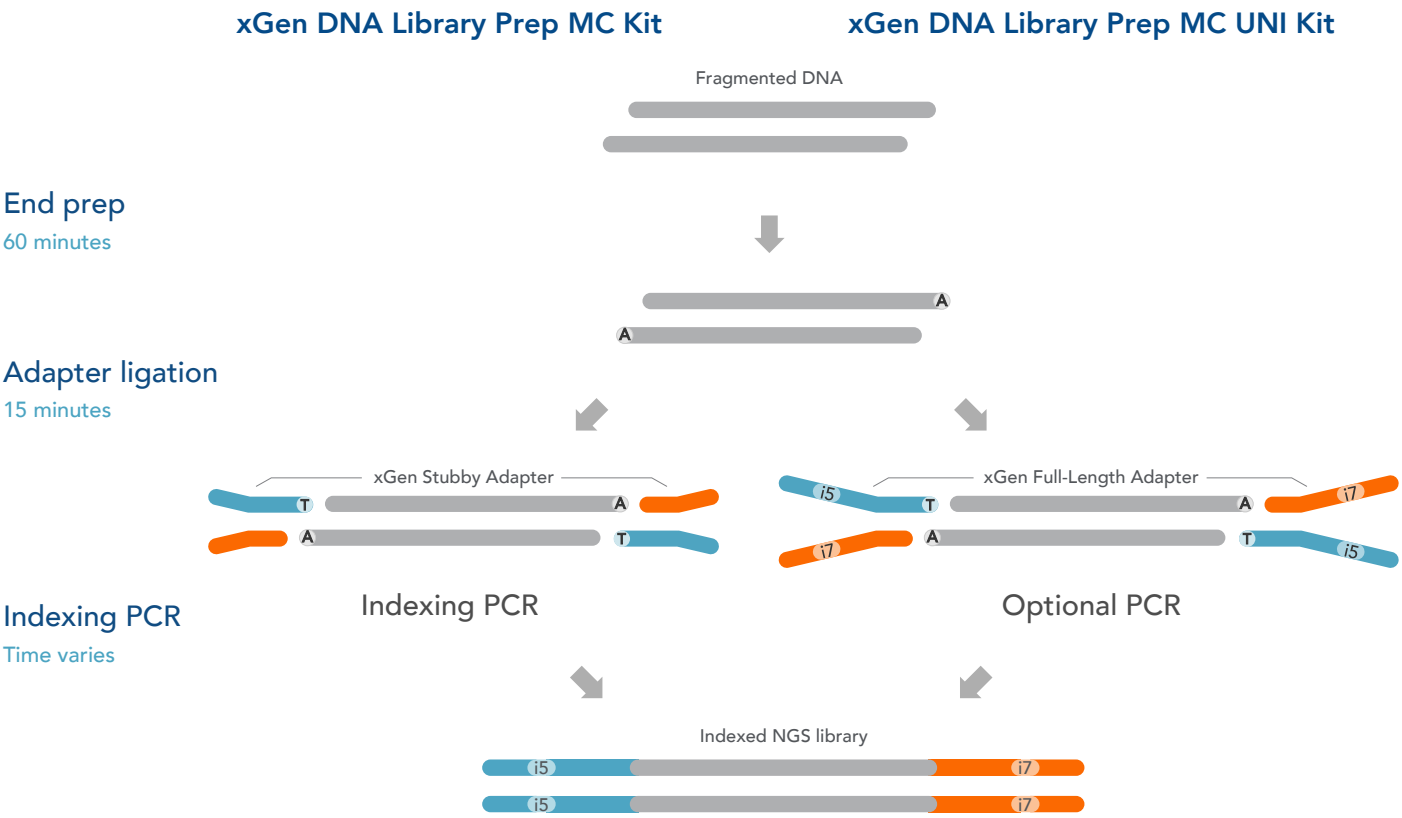

Figure 2. Workflow for the xGen DNA Library Prep MC Kits These kits are compatible with mechanically sheared DNA, and available in two configurations to support different indexing workflows. The incubation steps consist of end repair, polishing of dsDNA, and A-tailing, all performed in a single End Prep reaction followed by ligation of either a stubby Y adapter (xGen DNA Library Prep MC kit, left) or full-length indexed Y adapter (xGen DNA Library Prep MC UNI Kit, right). The xGen DNA Library Prep MC workflow incorporates an indexing PCR step after adapter ligation to complete the adapter sequences.

Hybridization capture

Indexed DNA libraries were captured and multiplexed according to the xGen Hybridization Capture of DNA Libraries protocol. For each library type, 500 ng of DNA library was used to generate four singleplex and one 8-plex capture (10 total captures). An 8-plex capture was included to demonstrate flexibility in library capture pooling strategies. Each pool underwent a 4 h hybridization capture reaction. PCR was performed on each capture using settings optimized for the panel size and number of pooled libraries in each capture (10 PCR cycles for singleplex captures; 7 PCR cycles for 8-plex captures). Final library size was determined using the TapeStation® 4200 system and High Sensitivity D1000 ScreenTape® assay (Agilent). Final library concentration was determined using AccuClear® Ultra High Sensitivity dsDNA Quantitation kit (Biotium).

Whole-exome sequencing

Enriched libraries were normalized to 4 nM and sequenced on a NextSeq® 550 instrument (Illumina) using 2 x 150 paired-end reads. Each library type required its own sequencing run, resulting in two total runs. Libraries were subsampled to 50 M total reads each before analysis against the product target space in human reference genome hg38. Sequencing metrics were calculated using Picard analysis tools [4].

Results

Flexible workflows maintain comprehensive and quality coverage

On-target rates represent bases or reads that align to the targeted region. A higher percentage of on-target reads gives increased confidence in the accuracy of sequencing data. To simplify analysis and compare across diverse samples, the target region was padded (150 bp in either direction) to calculate bases flanked on-target. Both xGen DNA Library Prep workflows demonstrated similar performance between singleplex and 8-plex captures, with ≥ 90% flanked on-target rate for each library type (Figure 3A).

In addition, comprehensive coverage at an adequate read depth is necessary for variant calling. Both xGen DNA Library Prep workflows demonstrated coverage of 97% of target bases at ≥ 20X coverage depth, with no noticeable drop-off in performance with 8-plex captures (Figure 3B). For germline exome sequencing, a minimum coverage depth of 20X is generally applied. Even at a higher coverage depth of 30X, libraries from both workflows maintained outstanding coverage of 94% of target bases.

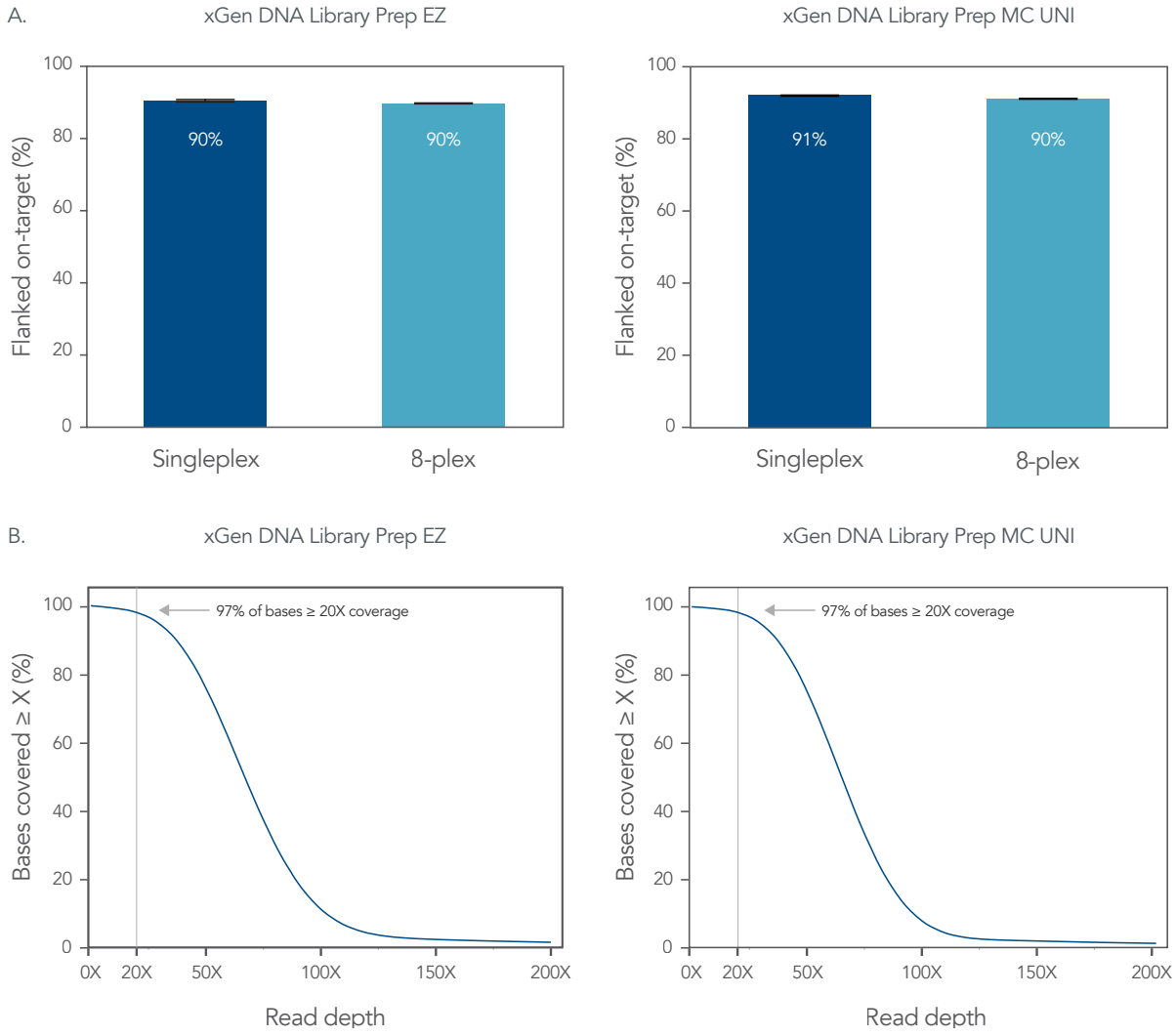

Figure 3. xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI libraries offers comprehensive coverage for whole exome sequencing. Four singleplex WES captures and one 8-plex WES capture were performed with the xGen Exome Hybridization Panel v2. All captures used 500 ng of library and a 4-hr hybridization time. Sequencing was performed on a NextSeq® 550 high-output flowcell (Illumina) to generate 2 x 150 paired-end reads. Reads for all samples were down sampled to 50 M total reads. (**A**) Comparable flanked on-target rates are seen with both singleplex and 8-plex capture conditions. (**B**) 97% of target bases are covered at ≥ 20X coverage. Data was analyzed using Picard tools. *n* = 24 libraries (12 libraries × 2 library prep workflows); 10 total captures (5 captures × 2 library prep workflows)

High library complexity was determined using hybrid-selection (HS) library size and shown as estimated unique molecules. xGen DNA Library Prep EZ libraries had over 200 M unique molecules; xGen DNA Library Prep MC UNI libraries yielded over 150 M unique molecules (Figure 4). Higher library complexity represents more unique molecules and an increased ability to identify rare genetic events or variants.

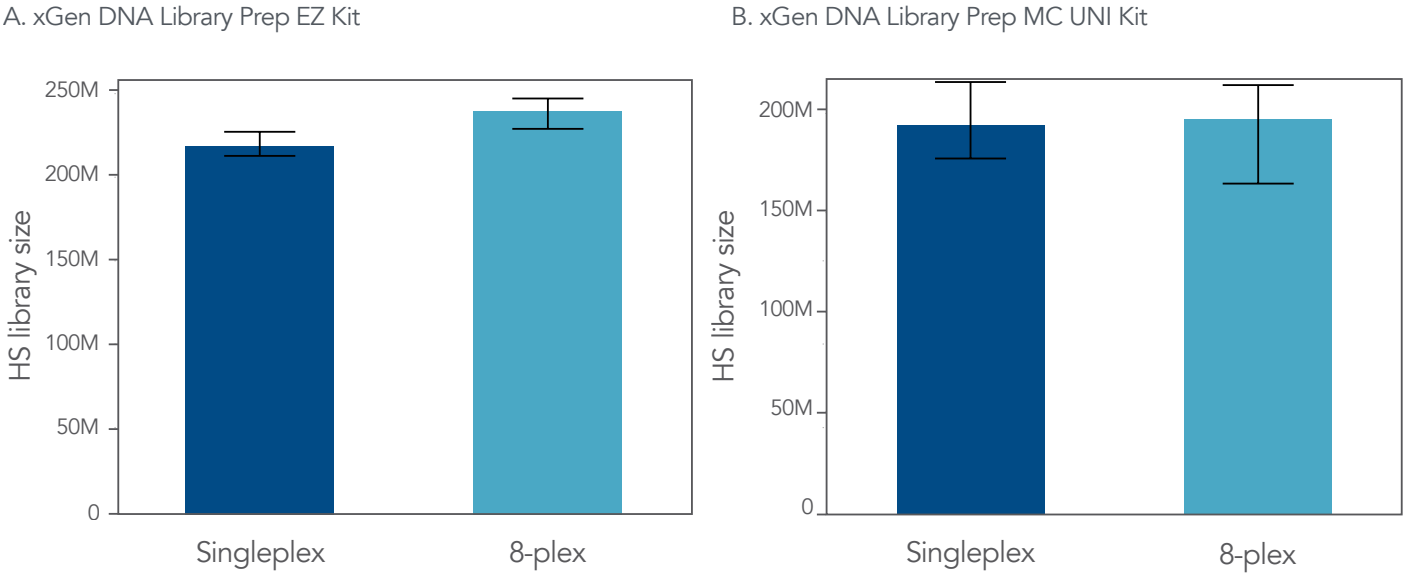

Figure 4. xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI workflows deliver high complexity libraries. Four singleplex WES captures and one 8-plex WES capture were performed with the xGen Exome Hybridization Panel v2. All captures used 500 ng of library and a 4 h hybridization time. Sequencing was completed on a NextSeq® 550 high-output flowcell (Illumina) to generate 2 x 150 paired-end reads. Reads for all samples were down sampled to 50 M total reads. Hybrid-selection (Hs) library size, reported as estimated unique molecules, was determined using Picard tools. *n* = 24 libraries (12 libraries × 2 library prep workflows); 10 total captures (5 captures × 2 library prep workflows)

Duplication rates for both library workflows were below 6% (Figure 5). No difference in duplication rate was observed between singleplex or 8-plex captures. Less duplicate reads produced mean less reads are wasted in a sequencing run, while the flexibility to increase multiplexing saves time and cost without negatively impacting data quality.

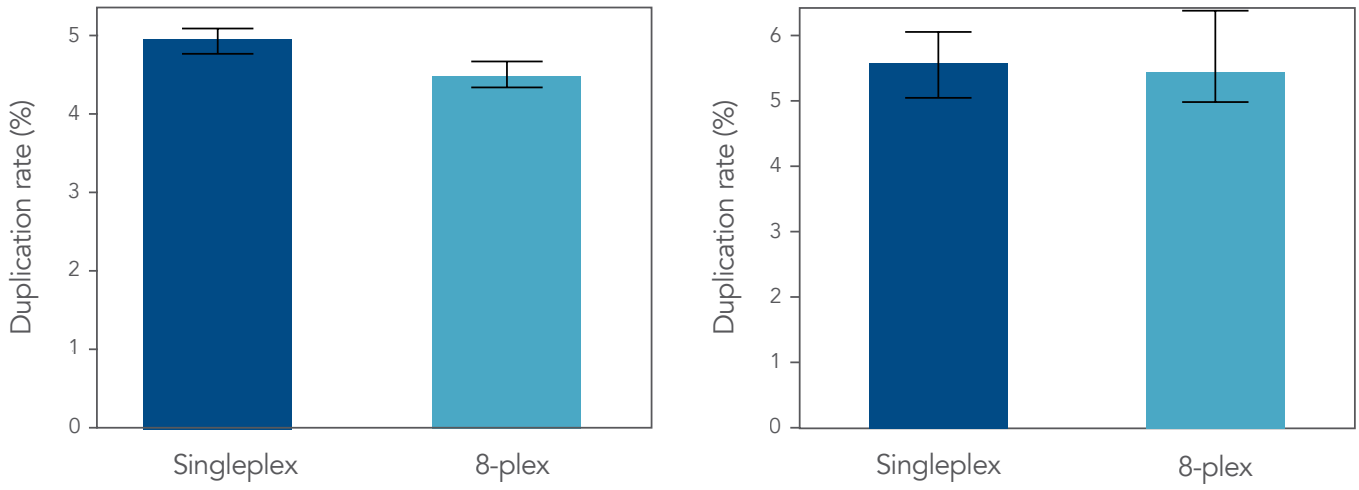

Figure 5. xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI libraries yields less wasted reads. Four singleplex WES captures and one 8-plex WES capture were performed with the xGen Exome Hybridization Panel v2. All captures used 500 ng of library and a 4-hr hybridization time. Sequencing was completed on a NextSeq® 550 high-output flowcell (Illumina) to generate 2 x 150 paired-end reads. Reads for all samples were down sampled to 50 M total reads. Duplication rates were determined using Picard tools. Duplication rates for both libraries were below 6%. *n* = 24 libraries (12 libraries × 2 library prep workflows); 10 total captures (5 captures × 2 library prep workflows).

Uniformity of coverage

To further evaluate sequencing performance, multiple uniformity metrics were assessed: the effect of GC-content, fold-80 base penalty, and percent of target bases between 0.5X and 1.5X of the mean.

Uniform coverage of both high and low GC regions indicates fewer dropouts and more complete coverage of targeted regions. To evaluate this, Exon 1 (GC-rich) and Exon 2 (AT-rich) of the RB1 gene target space were analyzed for both library workflows. Integrated Genome Viewer (IGV) [5] plots show ample coverage for both GC- and AT- rich regions, indicating low bias with either library prep workflow (Figure 6A). This coverage is maintained at the same level in the 8-plex capture (data not shown).

Fold-80 base penalty (fold-80) is the amount of extra coverage required for 80% of the target sequences to reach the mean coverage target coverage. Using Picard analysis tools, fold-80 was calculated for these two xGen DNA Library workflows. Both library prep workflows and pool levels yielded libraries with low fold-80 around 1.4 (Figure 6B). Zero coverage targets for both library types were approximately 1%. This is an important consideration since the fold-80 calculation does not include regions with zero coverage, and sequencing data with a significant lack of coverage could still achieve a fold-80 close to 1.0—perfect uniformity with no variance. This caveat illustrates the importance of assessing multiple metrics to adequately evaluate targeted sequencing performance.

An additional uniformity metric that can be considered is the percent target bases between 0.5X and 1.5X of the mean target coverage. A higher percentage of target bases closer to the mean indicates more uniform sequencing. Both library workflows demonstrated high uniformity; library samples from xGen DNA Library Prep EZ yielded 85% of target bases between 0.5X and 1.5X of the mean, while xGen DNA Library MC UNI libraries yielded 87% (Figure 6C).

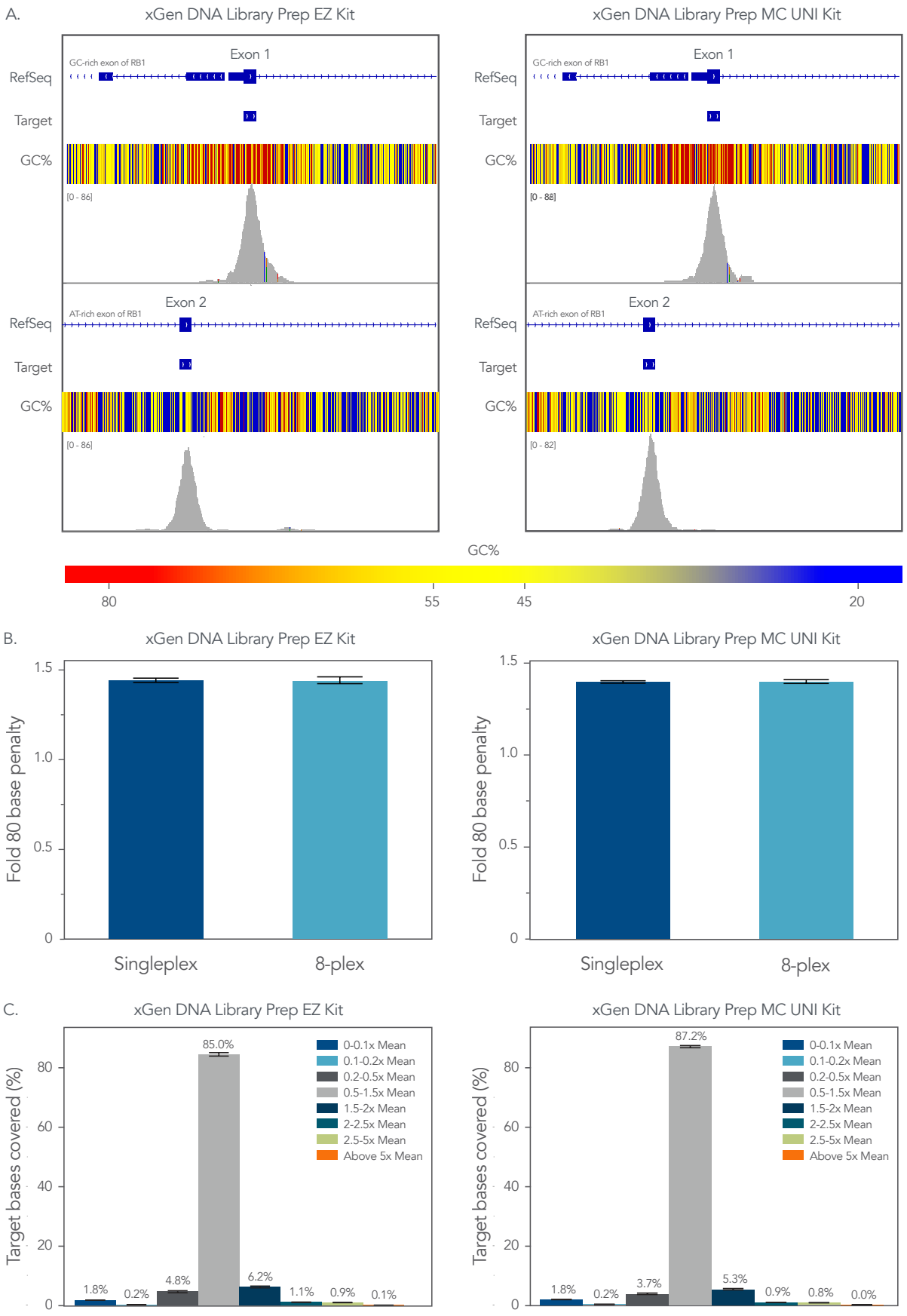

Figure 6. xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI libraries create highly uniform libraries for whole exome sequencing. (**A**) IGV coverage histograms showing ample coverage in both GC- and AT- rich regions of the RB1 gene. This coverage is maintained at the same level in the 8-plex captures. (**B**) Fold-80 base penalty of 1.4 for xGen DNA Library Prep MC UNI and 1.44 for xGen DNA Library Prep EZ libraries. (**C**) ≥ 85% of target bases for both libraries are between 0.5X and 1.5X of the mean. Data was analyzed using Picard tools. Note: One xGen DNA Library Prep MC UNI sample was determined to be an outlier (1.5X Interquartile Rule used to determine outlier) and was excluded from the data set. *n* = 24 libraries (12 libraries × 2 library prep workflows); 10 total captures (5 captures × 2 library prep workflows)

Conclusion

xGen DNA Library Prep EZ and xGen DNA Library Prep MC UNI Kits demonstrate similar high-quality performance for whole-exome sequencing, despite differences in workflow. This high level of performance is maintained between singleplex and multiplex captures, lowering cost and time without negative impact to data quality. Our workflows yield high on-target rates, high complexity libraries, and less wasted sequencing reads. Both workflows also demonstrate ample and uniform coverage for both high and low GC targets, while resulting in few dropouts due to zero coverage targets.

Moreover, whole-exome sequencing using xGen DNA Library Prep Kits for library generation and xGen Exome Hybridization Panel v2 for hybridization capture give users complete control over library workflow design with many customizable features, such as choices in DNA fragmentation, adapter indexing strategy, stocked or custom barcode options, and creation of PCR-free or PCR-amplified libraries.

Our **xGen NGS Solutions Builder Tool** can also simplify and enhance your ordering experience, ensuring you select the right reagents for your unique, custom project.

References

1. Choi M, Scholl UI, Ji W, et al. **Genetic diagnosis by whole exome capture and massively parallel DNA sequencing**. *Proc Natl Acad Sci*. 2009;106(45):19096-19101. doi:10.1073/PNAS.0910672106
2. Williams HJ, Hurst JR, Ocaka L, et al. **The use of whole-exome sequencing to disentangle complex phenotypes**. *Eur J Hum Genet 2016* 242. 2015;24(2):298-301. doi:10.1038/ejhg.2015.121
3. Bamshad MJ, Ng SB, Bigham AW, et al. **Exome sequencing as a tool for Mendelian disease gene discovery**. *Nat Rev Genet 2011* 1211. 2011;12(11):745-755. doi:10.1038/nrg3031
4. Picard Toolkit. Broad Institute, *GitHub Repository. 2019*. **https://github.com/broadinstitute/picard**. Broad Institute.
5. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. **Integrative Genomics Viewer**. *Nat Biotechnol*. 2011;29(1):24. doi:10.1038/NBT.1754

Whole-exome sequencing using the xGen™ DNA Library Preparation Kit family enables customizable workflows returning quality in consistent results

Technical support: applicationsupport@idtdna.com

Doc ID: RUO22-0856_001 06/22