Improved whole exome sequencing with IDT's xGen™ Exome Hybridization Panel v2

Human exome sequencing enables deep sequencing of the protein coding regions of the human genome. Although the exome makes up approximately 1% of the genome, a recent study examining >450,000 samples from the UK Biobank reported over 12 million variants (median per sample of about 20,000 variants) [1]. This very rich dataset indicates that exome sequencing can provide a wealth of knowledge around understanding human diseases [1]. Sequencing the protein-coding exons streamlines research while saving time and sequencing-related costs.

Here we compare two versions of IDT's xGen Exome Hybridization Capture Panel used as part of a human whole exome sequencing workflow. There are several content and functional differences between the two panels. The xGen Exome Hyb Panel v2 offers numerous improvements that consist of a highly efficient capture probe design strategy based on an updated human reference genome build, Refseq 109 hg38. The xGen Exome Hyb Panel v2 offers more cost-effective sequencing than the first version for multiplexing or performing individual captures. Moreover, the xGen Exome Hyb Panel v2 panel shows equivalent functionality between single and 12-plex captures and gives better functionality over the xGen Exome Hyb Panel v1 in all key NGS metrics examined.

xGen Exome Hyb Panel design and content

Both xGen Exome Hyb Panels use the RefSeq database at the National Center for Biotechnology Information (NCBI) for the human reference genome for probe design. However, each panel uses different annotation for their respective designs. The first version of the panel uses the genome annotation Refseq 68 and assembly hg19, while the xGen Exome Hyb Panel v2 was designed using RefSeq 109 assembly hg38, an updated annotation offering significant improvements over hg19 [2].

Capture probe design strategy is another feature that differs between the two exome panel versions. The xGen Exome Hyb Panel v2 was created using a novel capture-aware probe design algorithm that reduces off-target binding while maximizing coverage and uniformity. The xGen Exome Hyb Panel v1 was designed using a 1X end-to-end probe tiling strategy and proprietary repeat masking algorithm.

The target space size differs between xGen Exome Hyb Panel v2 and v1 at 34 Mb and 39 Mb, respectively. The difference in target space is largely due to differences in the human reference genome builds, which the panel designs are based on. Differences in panel size can also be attributed to the alternate approach in capture probe design. Furthermore, the xGen Exome Hyb Panel v2 contains the addition of several non-coding targets, the *RMRP*, and the *TERC* exons as well as the *TERT* promoter in the target space. Both capture panels are comprised of fully quality controlled (by ESI-MS and OD) 5′ biotinylated 120-mer probes. These improvements have led to increased efficiency and cost-effectiveness of the xGen Exome Hyb Panel v2 compared to the xGen Exome Hyb Panel v1.

A unique feature available to customers using either version of the xGen Exome Hyb Panel is the ability to create a customized Exome+ panel. This allows users of the xGen Exome Hyb Panel to use the stocked panel and seamlessly add custom content, enabling aliquoting and storage from a single bulk formulation made at IDT. A comparison of the xGen Exome Hyb Panels features is summarized in Table 1.

Table 1. Capture panel content differences for xGen Exome Hyb Panel v2 and xGen Exome Hyb Panel v1.

	xGen Exome Hyb Panel v2	xGen Exome Hyb Panel v1
Genome annotation for design	hg38	hg19
RefSeq annotation for design	RefSeq 109	RefSeq 68
Probe design	Capture-aware algorithm	End-to-end, 1X tiling
Genes covered	19,433	19,936
Target space size (Mb)	34	39
Probe footprint (Mb)	39	51
Number of probes	415,115	429,826
Exome+ formulation available	Yes	Yes

Exome capture panel functionality

The DNA libraries were generated using 100 ng of Coriell NA12878 gDNA. DNA was fragmented enzymatically to 200 bp and ligated using xGen Stubby Adapters with unique dual index (UDI) primer pairs to create full length libraries for Illumina® sequencing. Captures were performed following the xGen Exome Hyb Panel protocol using 500 ng of each library in either a single or 12-plex hybridization capture for 4 hours. Sequencing was performed using a NextSeq® 550 (Illumina). Each library was subsampled to 50 M total reads, which were bioinformatically trimmed to 2 x 100 bp and analyzed using Picard (v2.18.9) [3].

Exome capture panel function was evaluated using key NGS metrics. The xGen Exome Hyb Panel v2 shows improvements in all metrics evaluated including capture specificity, uniformity, coverage depth and percentage of exons fully covered. The standard metric (PCT_SELECTED_BASES) from Picard was used to investigate the on-target rate. The first version of the panel shows ~83% selected bases, and the xGen Exome Hyb Panel v2 panel shows improved on target rates of just over 90% (Figure 1), with a minimal change in performance seen between single to multiplex captures.

This improvement indicates a more efficient capture and leads to fewer wasted sequencing reads, increasing the cost-effectiveness of the updated panel. In addition to an increase in the number of useable reads for xGen Exome Hyb Panel v2, we also see an increase in coverage depth which helps provide confidence in variant calling. There is an increased coverage of target regions at 20X for both single and 12-plex captured libraries when using xGen Exome Hyb Panel v2 (Figure 2A), as well as an increase in the number of exons covered end-to-end (Figure 2B). Fold-80 base penalty measures the extra sequencing needed to bring the bottom 20% of non-zero coverage probes up to the mean coverage of the panel and is a measurement of enrichment uniformity, where a smaller Fold-80 score indicates better uniformity of capture. We see an improvement in uniformity for the xGen Exome Hyb Panel v2 with a Fold-80 of 1.45, a decrease from the first version of the panel which shows a Fold-80 of 1.67 (Figure 3).

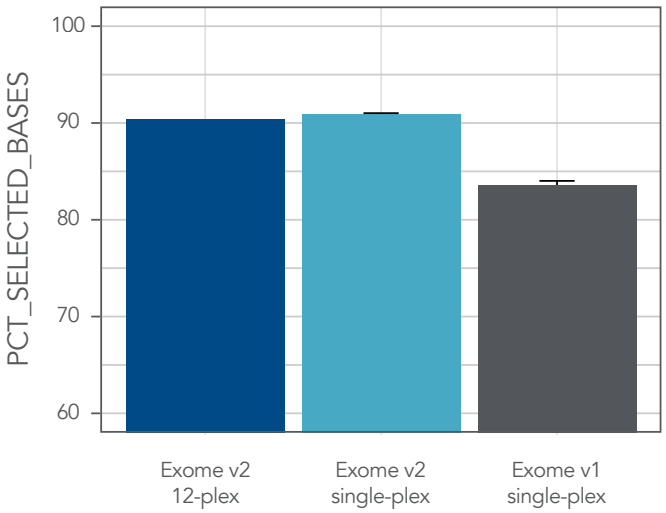

Figure 1. xGen Exome Hyb Panel v2 shows superior on-target rate for single and 12-plex captures when compared to xGen Exome Hyb Panel v1. Single and 12-plex whole exome sequencing captures were performed using xGen Exome Hyb Panels v1 and v2. All captures used 500 ng of each library and a 4-hr hybridization time. Sequencing was performed using a NextSeq® 550 (Illumina). Each library was subsampled to 50 M total reads, which were bioinformatically trimmed to 2 x 100 bp and analyzed using Picard (v2.18.9). Both single and multiplex captures for xGen Exome Hyb Panel v2 have an on-target rate of >90%, xGen Exome Hyb Panel v1 has an on-target rate between 80–85%. Eighteen libraries were generated from 7 total captures (3 single-plex captures using the xGen Exome Hyb Panel v1, 3 single-plex captures using the xGen Exome Hyb Panel v2, and one 12-plex capture xGen Exome Hyb Panel v2). Error bars represent standard deviation.

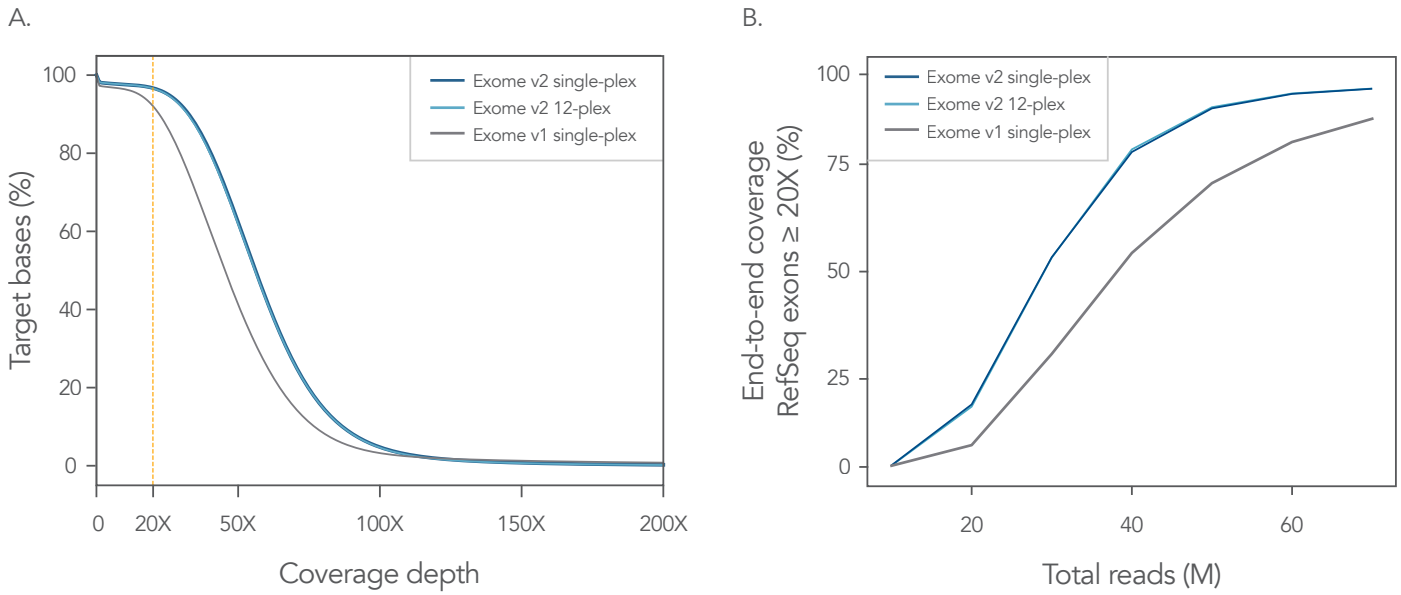

Figure 2. xGen Exome Hyb Panel v2 shows an improvement in target coverage. Single and 12-plex whole exome sequencing captures were performed using xGen Exome Hyb Panels v1 and v2. All captures used 500 ng of each library and a 4-hr hybridization time. Sequencing was performed using a NextSeq® 550 (Illumina). Each library was subsampled to 50 M total reads, which were bioinformatically trimmed to 2 x 100 bp and analyzed using Picard (v2.18.9). (**A**) A higher percentage of target bases are covered at 20X for both the single and 12-plex captures using the xGen Exome Hyb Panel v2 when subsampling to 50 M reads. (**B**) The xGen Exome Hyb Panel v2 shows more complete end-to-end coverage of targeted RefSeq exons than the first version regardless of sub-sampling to different read counts. Eighteen libraries were generated from 7 total captures (3 single-plex captures using the xGen Exome Hyb Panel v1, 3 single-plex captures using the xGen Exome Hyb Panel v2, and one 12-plex capture xGen Exome Hyb Panel v2).

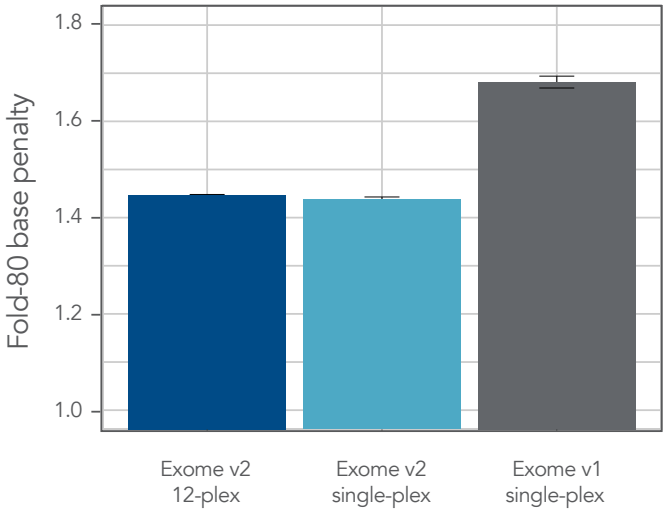

Figure 3. xGen Exome Hyb Panel v2 shows improved uniformity. Single and 12-plex whole exome sequencing captures were performed using xGen Exome Hyb Panels v1 and v2. All captures used 500 ng of each library and a 4-hr hybridization time. Sequencing was performed using a NextSeq® 550 (Illumina). Each library was subsampled to 50 M total reads, which were bioinformatically trimmed to 2 x 100 bp and analyzed using Picard (v2.18.9). The xGen Exome Hyb Panel v2 has a Fold-80 of just over 1.4 for both the 12-plex and singleplex captures, lower than the Fold-80 for the first version of just under 1.7 for all single-plex captures. Eighteen libraries were generated from 7 total captures (3 single-plex captures using the xGen Exome Hyb Panel v1, 3 single-plex captures using the xGen Exome Hyb Panel v2, and one 12-plex capture xGen Exome Hyb Panel v2). Error bars represent standard deviation.

A summary of key NGS metrics for the libraries resulting from the xGen Exome Hyb Panel v2 and v1 is included in Table 2.

Table 2. Key NGS metrics summary for xGen Exome Hyb Panel v2 and xGen Exome Hyb Panel v1.

	xGen Exome Hyb Panel v2	xGen Exome Hyb Panel v1
On target	91%	83%
Fold-80 base penalty	1.45	1.67
Bases covered ≥ 20X	97%	92%
End-to-end exon coverage	91%	74%

Each value in Table 2 represents the mean of 3 single-plex captures. Overall, the xGen Exome Hyb Panel v2 contains several benefits over the xGen Exome Hyb Panel v1 including an optimized probe design based on an updated human reference genome build, Refseq 109 hg38. The xGen Exome Hyb Panel v2 offers more cost-effective sequencing than the first version of the panel for multiplexing or performing individual captures, which shows equivalent performance between single and 12-plex captures, and also gives superior performance over the first version in all key NGS metrics examined.

References

1. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599(7886):628-634. doi:10.1038/s41586-021-04103-z
2. Schneider VA, Graves-Lindsay T, Howe K, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res*. 2017;27(5):849-864. doi:10.1101/gr.213611.116
3. "Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute

Improved whole exome sequencing with IDT's xGen™ Exome Hybridization Panel v2

Technical support: www.idtdna.com/ContactUs

Doc ID: RUO22-1079_001 08/22