Abstract

Whole exome sequencing (WES) of the human genome with consistently deep coverage is essential for research, especially for researchers studying cancer or rare genetic variations [1–3]. Therefore, the panels used to target the human exome must be carefully designed to efficiently capture these protein-coding regions of the genome. IDT has updated their human exome capture panel using the most up-to-date exome reference combined with IDT’s proprietary bioinformatics design pipeline. Here, we compare the xGen Exome Hyb Panel v2 to other commercially available human exome panels. The results showed that the xGen Exome Hyb Panel v2 provided higher on-target rates and the most complete coverage compared to the other panels tested, while also offering greater flexibility in multiplexing levels and hybridization times.

Introduction

The ability to generate large amounts of sequencing data in a relatively short amount of time enables a wide range of genetic analysis applications and accelerates advances in clinical and applied markets research. However, despite the continual decrease in sequencing prices, the cost associated to sequencing the human genome with enough depth to precisely identify variants associated to different disease states, such as cancer, remains high and is thus impractical for large-scale analyses [4]. An alternative to whole-genome sequencing (WGS) is to sequence only certain regions of interest in a genome, for example the protein-coding regions or the exome also called whole exome sequencing (WES) [4,5]. This technique helps enables deep sequencing and the subsequent identification of important variations in those regions [4,5], and has been invaluable for advancing research in cancer and understanding other rare genetic variations in the human genome [3,6]. WES uses modified oligonucleotide probes to “capture” and enrich the region of interest, in this case exons in the human genome. These probes are then separated from the rest of the genetic material in the sample, permitting researchers to focus on the exons for downstream next generation sequencing (NGS) and analyses (Figure 1).
Add blocking oligos

Hybridize targets to capture probes

Incubate with magnetic streptavidin beads

Isolate targets with magnet

Figure 1. Schematic of a target capture workflow.

Focusing on protein-coding exons can lower the cost and decrease the 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 the variants that are associated with disease [7]. However, when performing WES, it is important to attain high rates of on-target read mapping as well as, deep and uniform coverage of the target space to be able to reliably identify mutations [2,8]. The metrics that are commonly used to assess WES are further explained in Box 1 Performance metrics for targeted whole-exome sequencing.

IDT has updated their panel of hybridization capture probes, the xGen Exome Hyb Panel v2, for WES of the human genome. This panel contains 415,115 biotin-modified oligonucleotide probes that are individually synthesized and assessed for quality, assuring that each probe within the panel meets the necessary requirements for high quality sequencing results. More specifically, each probe within the xGen Exome Hyb Panel v2 is analyzed using both electrospray ionization-mass spectrometry (ESI-MS) and optical density (OD) measurements. ESI-MS provides confidence that the probe meets IDT’s high-quality standards and has minimal deletions, additions, or base substitutions. Pooling probes after synthesis supports appropriate representation within the final panel. Furthermore, IDT’s approach to probe synthesis limits the number of truncated probes in each panel and helps to ensure reproducibility between panels.
Here, the IDT xGen Exome Hyb Panel v2 is compared to other vendors’ exome hybridization panels. Each panel’s protocol was followed using the same genomic DNA (gDNA) input, sequenced, and subsequently compared using the RefSeq human exome target space (v109) [9]. The capability of the panels were assessed using the metrics in Box 1. Additionally, the xGen Exome Hyb Panel v2 was tested using a range of multiplexed samples (1 to 16), and hybridization times to ascertain its flexibility. The results show that the xGen Exome Hyb Panel v2 provides the highest percentage of on-target reads and the best coverage of the human exome in comparison to other tested panels. Further, the xGen Exome Hyb Panel v2 permits researchers to multiplex up to 16 samples and can be used in combination with a variety of hybridization times, meaning researchers can go from library preparation to sequencing using flexible workflow parameters.

**Box 1. Performance metrics for targeted whole-exome sequencing**

**Reproducibility between lots.** It is not uncommon for researchers to repeat experiments, which could result in the need to purchase panels from different production lots. However, variation between lots in target coverage requires you to revalidate your new panel and can be expensive. Thus, it is imperative that panels contain probes which provide similar results.

**Percentage of on-target reads mapped.** The measurement of on-target bases or reads is typically represented as the ratio of number of reads within a target region to total number of reads aligned to the reference genome, expressed as a percentage [8]. A higher percentage of on-target reads indicates more efficient sequencing data.

**Coverage depth.** Coverage represents the number of times a sequenced DNA fragment (a read) maps to a genomic target. The deeper the coverage of a target region (i.e., the more times the region is sequenced), the greater the reliability of the sequencing data [8].

**Coverage uniformity.** Coverage uniformity is a metric that should be considered in combination with other measures. One measure of uniformity is the fold-80—the amount of extra coverage required for 80% of the target sequences to reach the mean coverage depth. It is calculated by dividing the mean coverage by the 20th percentile coverage. Fold-80 can be a misleading measure of sequencing efficiency since regions with zero coverage are not included in the calculation. Sequencing data with a significant lack of coverage could still obtain a fold-80 score close to 1.0, which indicates perfect uniformity. This caveat illustrates the importance of evaluating targeted NGS panel data outputs considering multiple measurements.

**Effectively targets challenging regions.** Exons associated to the same protein coding region can have differing amount of GC or AT pairs (i.e., the first exon of a region tends to have higher GC content than following exons [10]), which can result in poor coverage if probes are not designed properly.

**Flexibility of workflow.** Ideally, experiments produce similar data when variables like hybridization time or multiplex level are changed for the hybridization-capture workflow. Flexibility of timing permits users to plan experiments around shifts. The ability to multiplex provides similar flexibility, helping to enable users to adjust the workflow to the number of samples, as well as providing cost-effectiveness.
Methods

Comparison between panels

One hundred nanograms of human gDNA (NA12878, Coriell) was used to prepare libraries with the corresponding library prep kit from each manufacturer for their specific exome panel—e.g., the xGen™ DNA Library Prep Kit EZ was used with the xGen Exome Hyb Panel v2. For the comparison study, libraries were multiplexed and captured using overnight hybridization according to each vendor’s exome panel capture protocol. An 8-plex capture was used since this is the maximum complexity supported across all panels.

Enriched libraries were sequenced on a NextSeq® 500 instrument (Illumina®) in high output mode using 2 x 150 paired-end (PE) reads. 50 M reads per sample were analyzed (5 Gb) using open-access tools (Picard, v2.18.9 [11]) against a universal human exome target space (RefSeq v109 [9]).

Workflow flexibility

In addition to the panel comparison study, input and conditions were tested for the xGen Exome Hyb Panel v2 to learn more about the capabilities of the updated panel. As with the competitor study, 100 ng of human gDNA (NA12878, Coriell) was used to prepare libraries with the xGen DNA Library Prep Kit EZ which were then enriched with the xGen Exome Hyb Panel v2. Five hundred nanograms of prepared libraries were used for 1-, 8-, 12-, and 16-plex captures to understand the full multiplexing range supported by the panel and hybridized for 16 hours. Data was then subsampled to 5 Gb for analysis.

To investigate the flexibility of hybridization time in the xGen Exome Hyb Panel v2, 1-plex xGen DNA Library Prep Kit EZ libraries were prepared and enriched using the xGen Exome Hyb Panel v2 with hybridization times of 0.5, 1, 1.5, 2, and 4 hours.

Results and discussion

These experiments were preformed to assess the functionality and flexibility of the updated xGen Exome Hyb Panel v2, relative to other panels available on the market. More specifically, the goal was to determine how the xGen Exome Hyb Panel v2 compared to other vendors in terms of lot-to-lot variation, on-target reads, coverage (uniformity and depth), targeting GC rich regions, and flexibility in complexity and hybridization time.

Minimal variation between panel lots

Lot-to-lot consistency is important to understand when selecting a hybridization panel. The xGen Exome Hyb Panel v2 is generated in a large-scale batch synthesis event and separated into aliquots which IDT numbers as different “lots”. These lot numbers are only used for tracking since the panels are generated during a single synthesis event. This is different from other vendors who generate different lots in multiple synthesis events. To illustrate the difference between these production processes, two independent researchers performed hybridization captures on different days, at different locations, using different lots of the xGen Exome Hyb v2 panel as well as different lots of vendor X's panel (Figure 2). The xGen Exome Hyb v2 panel was more consistent between lots than vendor X (Figure 2). This illustrates that IDT’s approach of using a single-batch synthesis event permits researchers to reproduce their results in a consistent manner by reducing lot-to-lot variation.
The xGen Exome Hyb Panel v2—human exome sequencing with consistent coverage

Figure 2. IDT’s xGen Exome Hyb Panel v2 resulted in more consistent coverage between lots compared to a competitor. 100 ng gDNA (Coriell) was used to make libraries, which were captured in 8-plex. Two independent users performed the captures on different days in different locations (n = 3). Bioinformatics analysis was done for 3-way coverage correlation comparison, a representative pair from each vendor is shown here. (A) The xGen Exome Research Panel v2 shows a linear regression line that mimics the predicted correlation line with an $R^2$ value of 0.76. (B) In comparison, 2 lots of supplier X showed less consistency and indicated more lot-to-lot variation.

Highest on-target rates and coverage of the human exome

DNA libraries were enriched using a multiplex of eight samples to compare between different exome panels (Figure 3). Relative to vendors X, R, and A the xGen Exome Hyb v2 panel resulted in the largest percentage of on-target reads (Figure 3A) as well as the highest mean target coverage (Figure 3B). These are two important metrics to consider when investigators are selecting panels for WES, as an increase in on-target reads and coverage helps enable researchers reliably identify mutations.

Figure 3. The IDT whole exome sequencing workflow resulted in high on-target percentage and consistent sequence coverage. 8-plexed enriched libraries were sequenced on a NextSeq® 500 instrument (Illumina) in high output mode using 2 x 150 bp PE reads for analysis. On-target bases and mean target coverage were determined using Picard tools [11] using 5 Gb per library (25 M read pairs). (A) On-target rate of the different panels. (B) The mean target coverage of the different panels.
Comprehensive and uniform coverage of the RefSeq human exome

Lack of coverage and uneven coverage can prevent reliable variant calling and may require samples to be re-sequenced [12]. Therefore, it is imperative that WES provides uniform and deep coverage to enable variant identification. The xGen Exome Hyb Panel v2 provided uniform sequence coverage of both 1-plex and 12-plex samples (Figure 4A). When comparing the percentage of complete exons covered in the RefSeq human exome (v109), the xGen Exome Hyb Panel v2 captured the highest percentage of exons at each analyzed depth (Figure 4B).

Figure 4. Libraries enriched with the xGen Exome Hyb Panel v2 resulted in the most complete coverage. (A) DNA libraries were created from 100 ng of human genomic DNA (Coriell) and enriched either as 1-plex captures (n = 3) or as a single 12-plex (n = 2) capture using the xGen Exome Hyb Panel v2. The enriched libraries were sequenced (2 x 150 bp reads) on a NextSeq® 500 instrument (Illumina) and subsampled to 5 Gb. Uniform coverage on the same run obtained a flanked on-target rate of 94.7%, mean target coverage of 64.5X, and a duplication rate of 3.3% (calculated with Picard). (B) 8-plexed enriched libraries were sequenced with 5 Gb per sample, and the percent of exons covered end-to-end at each read depth were calculated. The xGen Exome Research Panel v2 showed the highest percentage of exons covered at each indicated depth, compared to panels from suppliers X, R, and A.
Captured challenging regions of the exome as well as WGS

An advantage of WES is that it permits for deep sequencing of targeted regions of a genome, thus saving money on sequencing compared to WGS. However, as discussed in Box 1 it has been shown that exons within the same protein-coding region of a genome contain varying levels of GC [10], which can result in poor sequence coverage. The xGen Exome Hyb Panel v2 more closely resembles WGS, relative to vendor A (Figure 5A), and results in higher coverage of exons with varying levels of GC content (Figure 5B).

![Graph showing mean capture coverage (log10 scale) between exome panels and WGS. The xGen Exome Hyb Panel V2 matches the WGS more closely than supplier A (8-plex, n = 1). The coloration of the dots indicates the percentage of GC content. The xGen Exome Hyb Panel captured exons with higher GC content more effectively.](image)

![Graph showing RB1 exons 1 and 2 show extremes of GC content with ~76% in exon 1 and ~38% in exon 2. A comparison of capture between IDT and supplier A shows a higher read depth across exon 1 for the xGen Exome Hyb Panel v2 on the Integrative Genomics Viewer (Broad Institute), whereas supplier A coverage has a lower number of reads of this exon with high GC content (single-plex, n = 1).](image)

Figure 5. Coverage profile of the xGen Exome Hyb Panel v2 closely resembles whole genome data (WGS), and effectively captured exons with a high percentage GC content. (A) A plot of the mean capture coverage (log10 scale) between exome panels and WGS shows that the xGen Exome Hyb Panel V2 matches the WGS more closely than supplier A (8-plex, n = 1). The coloration of the dots indicates the percentage of GC content. The xGen Exome Hyb Panel captured exons with higher GC content more effectively. (B) RB1 exons 1 and 2 show extremes of GC content with ~76% in exon 1 and ~38% in exon 2. A comparison of capture between IDT and supplier A shows a higher read depth across exon 1 for the xGen Exome Hyb Panel v2 on the Integrative Genomics Viewer (Broad Institute), whereas supplier A coverage has a lower number of reads of this exon with high GC content (single-plex, n = 1).
Flexible range of multiplexing and hybridization times using the same workflow

The xGen Exome Hyb Panel v2, not only resulted in higher on-target read coverage, the most complete exon coverage of the RefSeq human exome, and effectively captured GC rich exons relative to other vendor’s panels (Figures 2–5), but it also offers researchers a flexible workflow in terms of hybridization time and multiplexing (Figure 6). To test the range of multiplexing that can be used by this workflow, libraries were prepared with the xGen DNA Library Prep Kit EZ and pooled at either 1-, 8-, 12-, or 16-plex using 500 ng from each library before hybridization with the xGen Exome Hyb Panel v2. Hybridization was performed for 16 hours, and data was subsampled to 5 Gb for analysis. The xGen Exome Hyb Panel v2 helps enable researchers to increase multiplexing in their library preparation while maintaining a high on-target rate (Figure 6A) and without noticeably increasing the rate of duplication (3–5.2%) (Figure 6B). The target coverage at ≥20X is also not reduced with an increase of multiplexing (Figure 6C). This consistency across a range of multiplexing levels permits researchers to maximize experimental efficiency.

Time is another factor that needs to be considered by researchers when performing hybridization experiments. The panels from other vendors that were tested against the xGen Exome Hyb Panel v2 call for a hybridization time of 16–24 hr. Even though the recommended hybridization time of the xGen Exome Hyb Panel v2 is already greatly reduced compared to the other vendors (4 hr), we investigated how reducing hybridization time even further impacted the key NGS metrics discussed in Box 1. Single-plex xGen DNA Library Prep Kit EZ libraries were tested using the xGen Exome Hyb Panel v2 with the 4 hr hybridization recommended in the protocol and with shorter hybridization times of 2, 1.5, 1, and 0.5 hours. On-target rate and mean target coverage are marginally impacted by reduced hybridization times (Figure 6D, E); and target coverage at ≥20X remains stable after 1-hour hybridization (Figure 6F). These results illustrate that the updated IDT exome solution helps to enable users to more easily go from capture to sequencing in a single day using the same xGen reagents and workflow protocol.

Figure 6. xGen Exome Hyb Panel v2 offers a range of flexible workflows, such as increased complexity and shorter hybridization times. (A–C) DNA libraries were pooled at 1-plex, 8-plex, 12-plex, and 16-plex using 500 ng of each indexed library before hybridization. (A) On-target rate remains consistently high across all multiplex levels tested. (B) The duplication rate is approximately 3% for up to 12-plex and rises to 5.2% for the 16-plex capture, as expected from higher multiplexing. (C) However, the slight increase in duplication rate does not impact target coverage at ≥20X which remains high at all multiplex levels. (D–F) Single-plex IDT xGen DNA Library Prep Kit EZ libraries were captured with the xGen Exome Hyb Panel v2 at the recommended 4 hour (hr) hybridization in the protocol and at shorter hybridization times of 2, 1.5, 1, and 0.5 hrs (n = 12). (D) On-target rate and (E) mean target coverage are only slightly decreased when hybridization time is below 1.5 hrs. (F) Target coverage at ≥20X remains high down to 1 hr hybridization.
Conclusions

WES is an application that permits researchers to identify variations in protein-coding regions of the human genome and has resulted in breakthroughs in cancer research as well as advanced the study of rare genetic variants [3,6]. However, this approach can be limited by the efficacy of the hybridization panel utilized to enrich the exome, as poor probe design can lead to lack of coverage depth and uniformity, low on-target rates, among other issues [4,5]. A comparison of the updated IDT exome panel, the xGen Exome Hyb Panel v2, shows that this panel outperformed similar panels from other vendors in every metric listed in Box 1. More specifically, use of the xGen Exome Hyb Panel v2 resulted:

- Highest on-target rate and mean coverage of the panels tested
- Increased depth coverage across a higher percentage of exons
- Expanded flexibility of workflow in terms of multiplexing range and hybridization times

More samples per capture saves time and cost without negatively impacting data. Because the sample number is not fixed in the xGen Exome Research Panel v2 protocol, hybridization capture can be performed without waiting for additional samples to reach the ideal multiplexing level, allowing researchers to maintain an efficient turnaround time. Smaller pools can attain the same great results as larger pools without additional kits or major changes in the original workflow.

Finally, since each probe in the panel is manufactured and assessed individually, and produce in a single, large-scale synthesis event the xGen Exome Hyb Panel v2 offers greater consistency for researchers using a WES approach.

Find more information on the xGen Exome Hyb Panel v2 at www.idtdna.com.

References


The xGen™ Exome Hyb Panel v2—human exome sequencing with consistent coverage

Technical support: applicationsupport@idtdna.com

For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service.

> SEE WHAT MORE WE CAN DO FOR YOU AT WWW.IDTDNA.COM.

For Research Use Only. Not for diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

© 2022 Integrated DNA Technologies, Inc. All rights reserved. Trademarks contained herein are the property of Integrated DNA Technologies, Inc. or their respective owners. For specific trademark and licensing information, see www.idtdna.com/trademarks.
Doc ID: RUO22-0790_001 05/22