Unveiling superior NGS library complexity and target coverage with the xGen™ cfDNA & FFPE DNA Library Prep Kit and hybridization capture solution

Abstract

Data from next generation sequencing (NGS) is highly dependent on the quality of the libraries. The xGen cfDNA & FFPE DNA Library Prep Kit paired with the IDT xGen NGS Hybridization Capture workflow provides researchers with an industry-leading solution to generate high-quality sequencing libraries. IDT partnered with a third-party research organization to compare the xGen cfDNA & FFPE DNA Library Prep Kit and custom hybridization capture (hyb cap) against two competitors. The IDT xGen NGS Workflow Solution outperforms both competitors' workflows due to the high sample conversion efficiency and a comprehensive xGen Custom Hybridization Capture panel design. xGen reagents are designed to be used together, and although they improve library complexity and sequencing metrics when used in competitor hyb cap workflows, the highest quality data comes from using the complete IDT xGen NGS Workflow Solution.

Introduction

Next generation sequencing (NGS) has been an epicenter of growth and innovation for the biotech and pharmaceutical research industries. It has revolutionized genomics and clinical research applications by enabling rapid and cost-effective sequencing of DNA and RNA, providing unprecedented insights into genetic variation, disease mechanisms, biological sciences, and personalized medicine. The speed and capacity of NGS have expanded exponentially in the past decade, enabling the sequencing of millions of DNA fragments on a single run, thus shifting the paradigm of genomics from addressing biological questions for a single gene or pathway to a genome-wide scale. Maximizing efficiency, cost-effectiveness, and data quality are crucial components of continued growth in NGS applications and are highly dependent on the construction of quality sequencing libraries.

Powered by novel chemistry (Figure 1), the xGen cfDNA & FFPE DNA Library Prep Kit is optimized for low input or challenging sample types like cell-free DNA (cfDNA) and formalin-fixed, paraffin-embedded (FFPE) samples. This industry-leading library prep solution provides ultra-high sample conversion efficiency and better sample representation.

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

IDT INTEGRATED DNA TECHNOLOGIES

The superior results of the IDT xGen cfDNA & FFPE DNA Library Prep Kit are driven by:

- Sequential, single-strand splint-ligation for maximized sample conversion efficiency
- An exclusive, novel ligase that prevents DNA chimera formation
- Thoughtful adapter modifications preventing dimer formation
- Unique molecular identifiers (UMIs) that enable enhanced resolution and the ability to remove library prep artifacts and sequencing errors

This combined higher library diversity and coverage coupled with stringent error correction boosts confidence in ultra-low frequency variant detection.

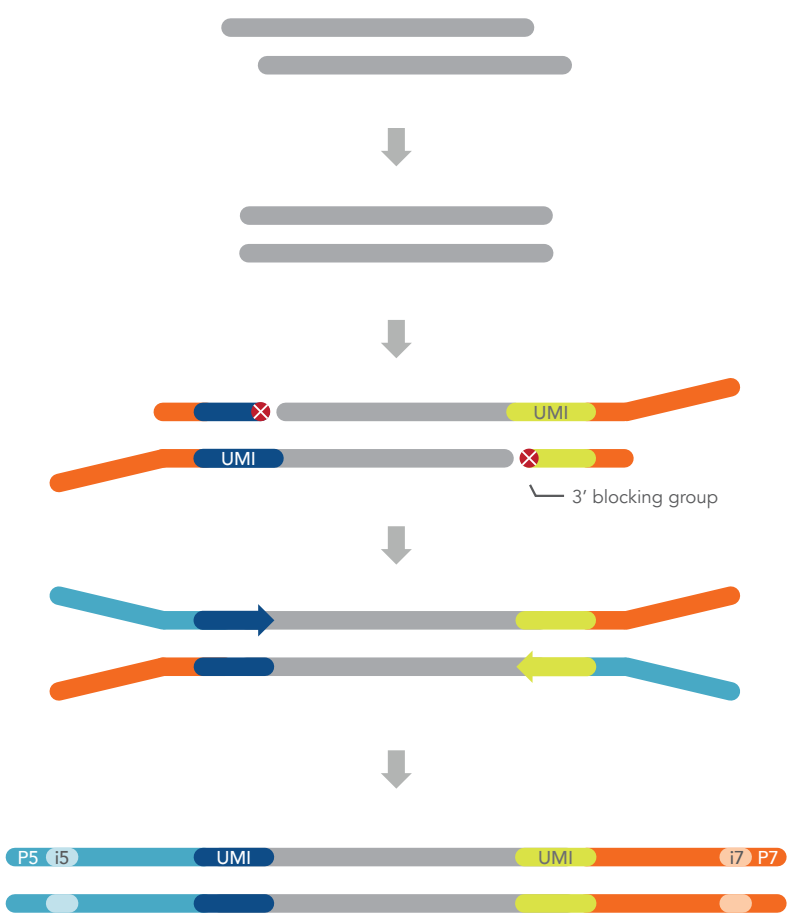


Figure 1. xGen cfDNA & FFPE DNA Library Prep Kit workflow. The workflow begins with end repair of cfDNA, FFPE DNA, or sheared gDNA to prep for blunt end ligation. Then the Ligation 1 Enzyme catalyzes the single-stranded addition of a Ligation 1 Adapter to the 3′ end of the DNA. Of note this novel enzyme is unable to ligate inserts together minimizing chimera formation. Additionally, the 3′ end of the Ligation 1 Adapter contains a blocking group to prevent adapter-dimer formation. The Ligation 2 Adapter acts as a primer to gap-fill the bases complementary to the Ligation 1 Adapter. This is followed by ligation of the 5′ end of the DNA insert creating a double-stranded product. In a final step, PCR incorporates sample index sequences for sequencing.

Combining the high conversion efficiency of xGen cfDNA & FFPE DNA library preparation with IDT xGen hybridization capture unlocks additional benefits driven by expertly designed target enrichment panels. Target enrichment via **hybridization capture** is a popular method that enables the sequencing of only the regions of interest in the genome, significantly reducing sequencing costs while simultaneously increasing throughput capability and target resolution.

xGen Custom Hybridization Capture Panels are made of high-fidelity, individually synthesized, and highly concentrated 5′-biotinylated oligos which do not require PCR amplification, thus reducing probe duplication errors. Capture panels are thoughtfully crafted by mixing equimolar concentrations of each probe resulting in uniform representation of all targets. IDT has carefully developed a proprietary panel design artificial intelligence algorithm optimized to capture more target regions while minimizing potential off-target effects (learn more about the design algorithm in this **white paper**). This strategy provides high target specificity and uniform capture, a hallmark of high-quality hybridization capture panel design.

Results

The complete xGen NGS Workflow Solution delivers more comprehensive coverage and sequencing efficiency compared to tested competitor workflows

For a non-biased comparison, a third-party research organization was used to perform experiments and sequencing using IDT's xGen NGS Workflow Solution and two NGS competitor's workflows. Each vendor generated a custom hybridization capture panel designed to the same list of 587 oncology-relevant genomic targets. IDT's hybridization capture panel design strategy employs proprietary algorithms and can be coupled with technical support and custom services from IDT scientists to help build an intuitive panel with optimized target coverage—even if that includes complex genomic elements such as repetitive regions, microsatellite instability (MSI) sites, insertions, and deletions. Leveraging these optimized algorithms and design expertise in this head-to-head panel comparison, IDT designed a panel that delivered 99.9% design coverage of desired target sequence, the highest of the three vendors (**Table 1**).

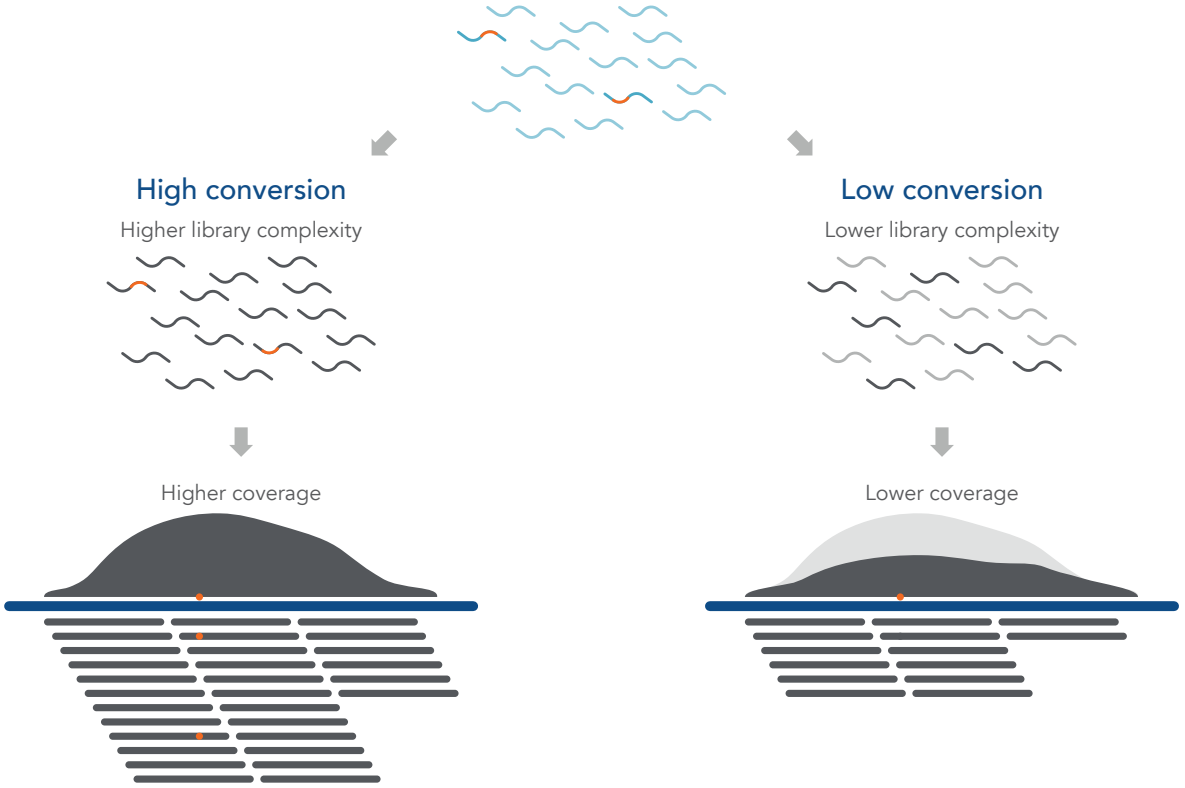
Table 1. Vendor custom hybridization capture panel design report metrics.

	IDT	Competitor A	Competitor T
Number of targets	587	587	587
Target bases covered	112,337	109,026	111,133
% Target bases covered	99.90%	96.96%	98.84%

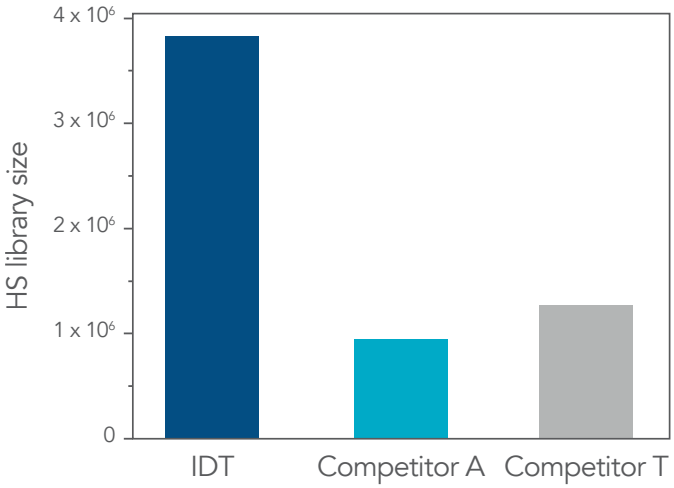
Complexity is an important characteristic of NGS libraries and can affect the quality of several key NGS metrics (**Figure 2A**). Complexity represents the number of unique molecules that were converted from DNA into usable library; the higher the library complexity, the better representation of the original sample input, which also leads to greater confidence in variant calling. As illustrated in **Figure 1**, the xGen cfDNA & FFPE DNA Library Prep Kit uses an efficient ligation strategy that leads to high conversion of sample DNA molecules into library molecules. This high conversion capability resulted in superior library complexity compared to other vendors, with IDT's solution generating libraries with almost four times higher complexity than competitors A and T's solutions (**Figure 2B**).

Complexity also directly affects the mean target coverage, the average number of reads that align to the target space. The higher the complexity of the library, the higher the coverage (Figure 2A). Using an equivalent number of subsampled reads, the xGen NGS Workflow Solution results in significantly higher mean target coverage than either competitor A or T's solutions (Figure 2C), translating to a higher efficiency and more cost-effective workflow.

A.



B.



C.

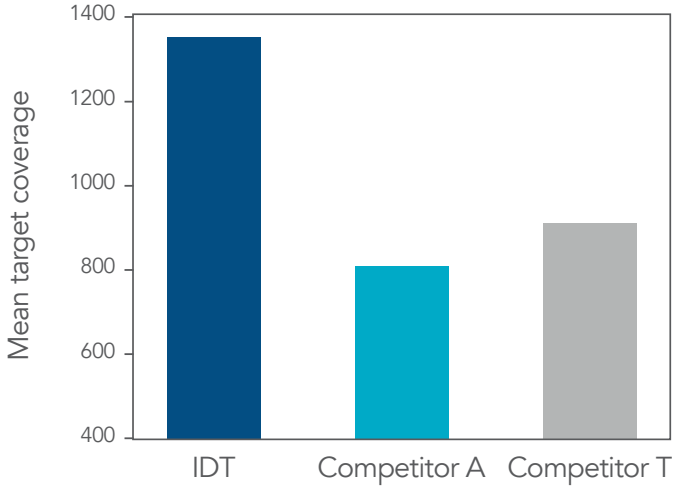


Figure 2. The xGen NGS Workflow Solution generates superior library complexity and mean target coverage. (**A**) Representation of how conversion efficiency and library complexity affect target coverage. The variant (orange) can be identified in the high conversion, high complexity library on the left, but was lost in the library on the right with low conversion and low complexity, which led to lower target coverage. (**B–C**) Hybridization captured libraries from each vendor were directly compared. The results show that the high HS library size (**B**) produced by the xGen NGS Workflow Solution results in higher mean target coverage (**C**) than both competitor A and T's sequencing solutions. The bar graphs display the mean of 24 replicates per vendor.

Library yield can also be used an indicator of conversion efficiency, with a higher library yield indicating higher library complexity. For a direct comparison of library yield, all libraries were made with the same amount of input material and underwent the same number of PCR cycles (see Methods). The IDT libraries had, on average, a final library yield over two times that of the competitors (IDT = 118 ng/µl, competitor A = 55 ng/µl, competitor T = 49 ng/µl).

During analysis, duplicate reads are removed to focus only on information from unique molecules. Duplicates can arise from multiple steps during the process including PCR and during sequencing chemistry. Higher duplication rates translate to more wasted sequencing reads, limiting the amount of data achievable and decreasing cost efficiency. Typically, higher library complexity trends with lower duplication rate. The duplication rates of all three vendors were compared to determine which NGS workflow provided the most efficient solution. On average, competitor T's sequencing solution resulted in almost two times the duplication rate as IDT's solution, while competitor A's solution had almost three times the duplication rate (data not shown). That translated to an average of IDT's libraries including 1.26×10^6 more reads for analysis than competitor T's, and 2.39×10^6 more reads included for analysis than competitor A's libraries.

Flanked-on target is a measure of sequencing reads that align to the region of the genome that is in the target space plus 150 bases on either side. Off-target refers to sequencing reads that align to regions of the genome that were not in the target space. A lower flanked on-target metric (and thus higher off-target) means that more reads are being aligned to regions outside of the target space, and consequently unused during analysis. The more reads that are removed during analysis, the more resources that are wasted during sequencing (i.e., paying for sequencing reads that can not be used). A direct comparison of NGS workflows demonstrated that using IDT's library prep and hybrid capture solution resulted in an average of 5.7% higher flanked on-target rate than competitor A's workflow, and 14.1% higher flanked on-target than competitor T's (data not shown). This reflects the increased target sequence-specificity of IDT hybridization capture panel design when compared to the competitors and will result in cost-savings during sequencing. The on-target rate also affects the mean target coverage (Figure 2C); when given the same number of reads, any read aligning to off-target regions will not contribute to the target coverage.

High target coverage is a key NGS quality metric that provides confidence in variant calling, especially in cases where identifying rare cell populations is important, such as cancer research or microbial studies. Uniformity and target coverage are often paired when evaluating the quality of hybridization capture. Ideally, all target spaces would have the same coverage. When hybridization capture probes are performing similarly, a high level of uniformity is seen, and genomic targets are not over- or under-sequenced. When a hybridization capture panel is imbalanced or probes enrich with different efficiencies, certain target regions may not reach the desired threshold for coverage and will require additional sequencing. Additional sequencing of regions with adequate coverage leads to over-sequencing of those regions and cost is driven up while efficiency goes down (Figure 3A).

Uniformity metrics like Fold-80 base penalty and percent target bases within a range of the mean coverage measure the evenness of sequencing and are a way to evaluate if under- or over-sequencing of any regions has occurred. Understanding the percentage of targets which receive coverage within +/–0.5X of the overall mean target coverage helps build a picture of sequencing uniformity. The higher this percentage, the more evenly regions within the target space are represented. Fold-80 represents the fold over-coverage necessary to raise 80% of bases in non-zero coverage targets to the mean coverage level. The closer the Fold-80 score is to 1, the more uniform the sequencing was, with 1 representing perfect uniformity across the target space. These metrics were used to compare the quality of custom hybridization capture workflows from IDT and two NGS competitors.

All three tested hybridization capture panels resulted in <1.2 % of the target space with zero coverage. When comparing the percent target bases uniformity metric, the xGen NGS Workflow Solution resulted in an average of 97% of target bases within +/–0.5X of the mean coverage, higher than either competitor T's and A's sequencing workflows which come in at 96% and 81%, respectively (**Figure 3B**). The average Fold-80 base penalty is nearest to 1 for IDT's libraries, followed by competitor T's and then competitor A's, further demonstrating that IDT hybridization capture is the most efficient and uniform solution for targeted sequencing (**Figure 3C**). These results highlight the evenness of the probe capture efficiency and design strategy of the IDT xGen Custom Hyb Capture Panel.

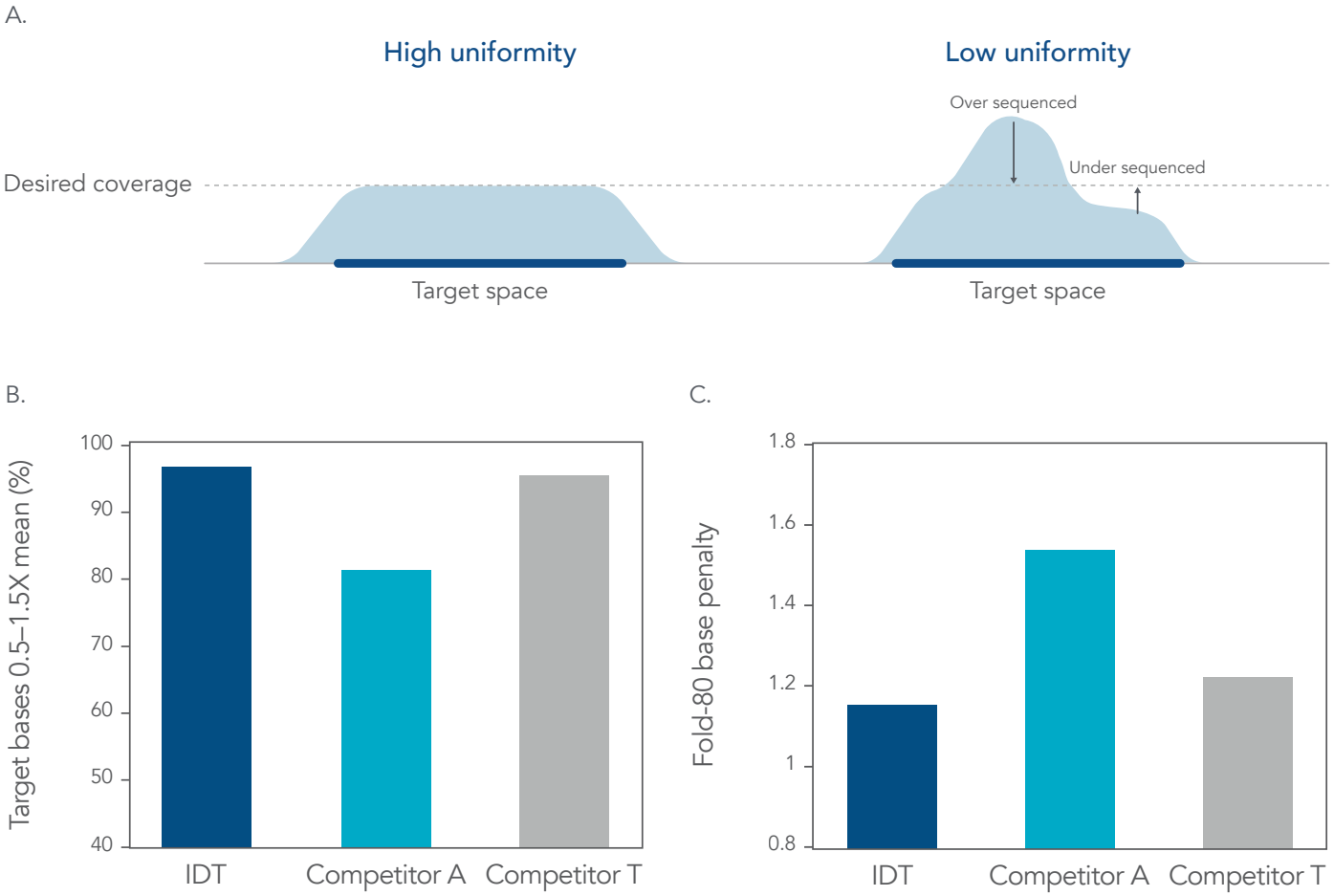


Figure 3. The xGen NGS Workflow Solution produces highly uniform sequencing. (**A**) Representation of how increased uniformity across target regions results in more efficient sequencing. Hybridization captured libraries from each vendor were directly compared. The xGen NGS Workflow Solution shows the highest level of uniformity of base coverage (**B**) and the lowest Fold-80 base penalty (**C**). The bar graphs represent the mean of the 24 replicates per vendor.

To demonstrate how the individual metrics highlighted above work together to generate quality sequencing results, a subset of targets from the sequencing data was visualized in the Integrative Genome Viewer (IGV, Broad Institute). Due to the high conversion rate of the xGen cfDNA & FFPE DNA Lib Prep Kit, a high library complexity and low duplication rate were observed, also resulting in the highest target coverage. The target sequence specificity and strategic design method of the xGen Custom Hyb Cap Panel generated data with great uniformity and minimal off-target reads across the target space. Competitors A and T's data both have lower coverage with the same amount of sequencing, and lower uniformity across targets when compared to IDT (Figure 4).

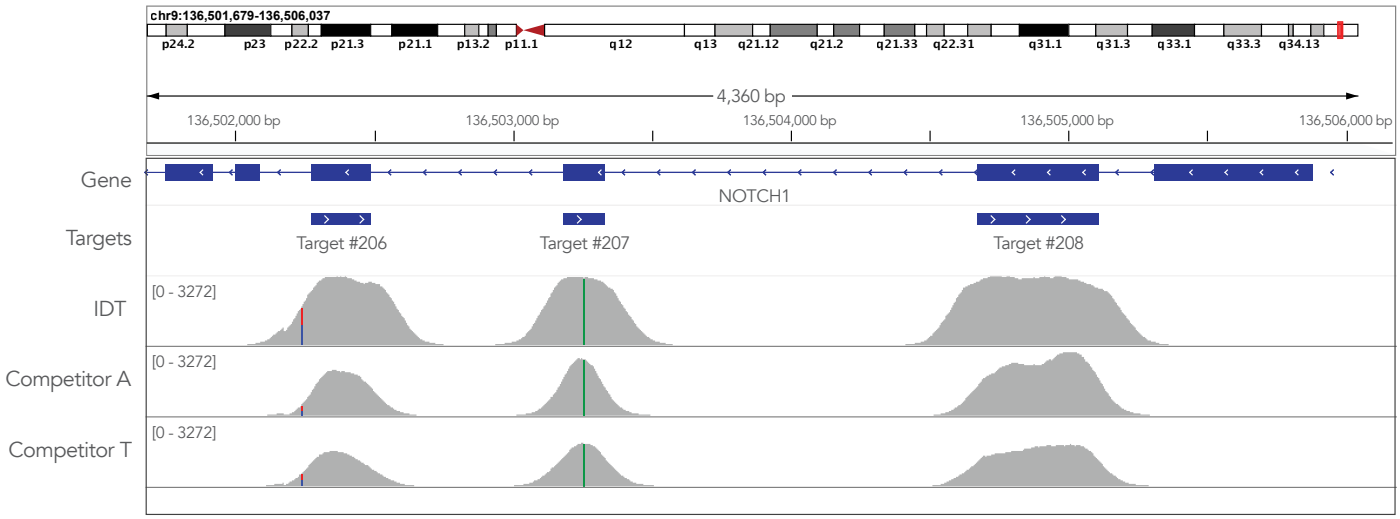


Figure 4. The xGen NGS Workflow Solution results in high target coverage and even uniformity. Sequencing data from individual target-enriched libraries (1 per vendor) was loaded into IGV for visualization. The representative hybridization capture library from IDT and each of the two competitors are set to the same coverage track on the y-scale (left side of image) for easy comparison. The grey color represents the coverage across the target space and multi-colored vertical bars represent bases that differ from the human reference genome (hg38).

Taken together, the single-strand splint-ligation strategy and unique enzymes in the xGen cfDNA & FFPE DNA Library Prep Kit paired with the xGen Custom Hyb Cap Panel result in high conversion of DNA into library, dramatically increasing library complexity resulting in higher target coverage and lower duplication rates when compared to competitors A and T, providing researchers a reliable workflow optimized to provide the highest efficiency for processing challenging samples.

Benefits of the xGen cfDNA & FFPE DNA Library Prep Kit are maximized when using the complete IDT targeted sequencing workflow

Due to the unique library technology and intelligent capture panel design, IDT's comprehensive xGen NGS Workflow Solution delivers excellent results when compared to the competitors' full solutions (Table 1, Figures 2–4). To better understand how the xGen library prep kit itself compares to the competitor's library prep kits, each competitor's hybridization capture technology was tested with either their own or with the xGen library prep kit to see how the performance was affected. The data shows the benefits of the IDT xGen cfDNA & FFPE DNA Library Preparation Kit and how the complete xGen NGS Workflow Solution is designed to work together for optimal performance. In Figures 5 and 6, the xGen NGS Workflow Solution data from Figures 2–4 is used and the data showing the xGen libraries with competitors' hybridization capture is added for comparison.

An increase in complexity translates to more unique molecules and thus a better representation of the original sample, which can be credited to better ligation efficiency and a higher quality library preparation. As shown in Figure 5, the xGen cfDNA & FFPE Library DNA Library Prep Kit produces NGS libraries with the highest complexity of all three tested kits and improves library complexity in both competitor workflows. The complete xGen NGS Workflow Solution results in a substantially more diverse library than the other evaluated workflows, reaffirming that components in the xGen NGS Workflow Solution are designed to work together to deliver high-quality sequencing data (Figure 5).

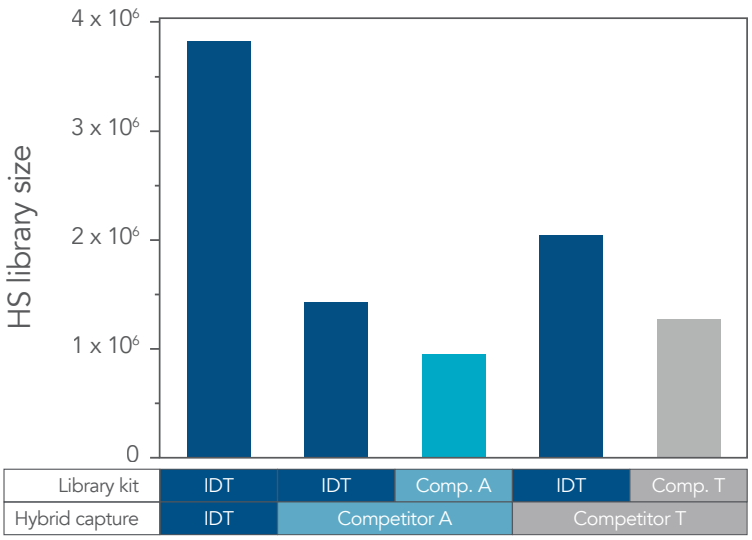


Figure 5. IDT's xGen NGS Workflow Solution's superior conversion efficiency results in the highest library complexity. An increase in HS library size is seen in competitor workflows when using the xGen cfDNA & FFPE DNA Library Prep Kit instead of the competitor's library kit. The xGen NGS Workflow Solution, which pairs xGen library prep with an xGen Custom Hyb Capture Panel, generates libraries with substantially higher complexity than all other tested workflows, highlighting the benefits of the complete IDT workflow. The bar graph represents the mean of the 24 replicates per workflow.

The benefits of using IDT's xGen cfDNA & FFPE DNA Library Prep Kit can also be demonstrated through the observed increase in target coverage (Figure 6A) and lower duplication rates when used as a component of each competitor's workflow (Figure 6B). Using the xGen library prep kit with competitor A's and competitor T's hyb capture workflows resulted in a ~14% decrease and 8% decrease in duplication rate compared to the full solutions from competitor A and competitor T, respectively. The full library prep plus hyb capture solution from IDT resulted in a ~10% lower duplication rate than the next best workflow combination (Figure 6B). This data highlights that the xGen unique library preparation chemistry results in high conversion of DNA to library molecules, which increases the complexity of the resulting hybridization captured library. Effectively, IDT library preparation promotes cost savings and rescues otherwise wasted reads, with optimal efficiency occurring when the full xGen NGS Workflow Solution is used.

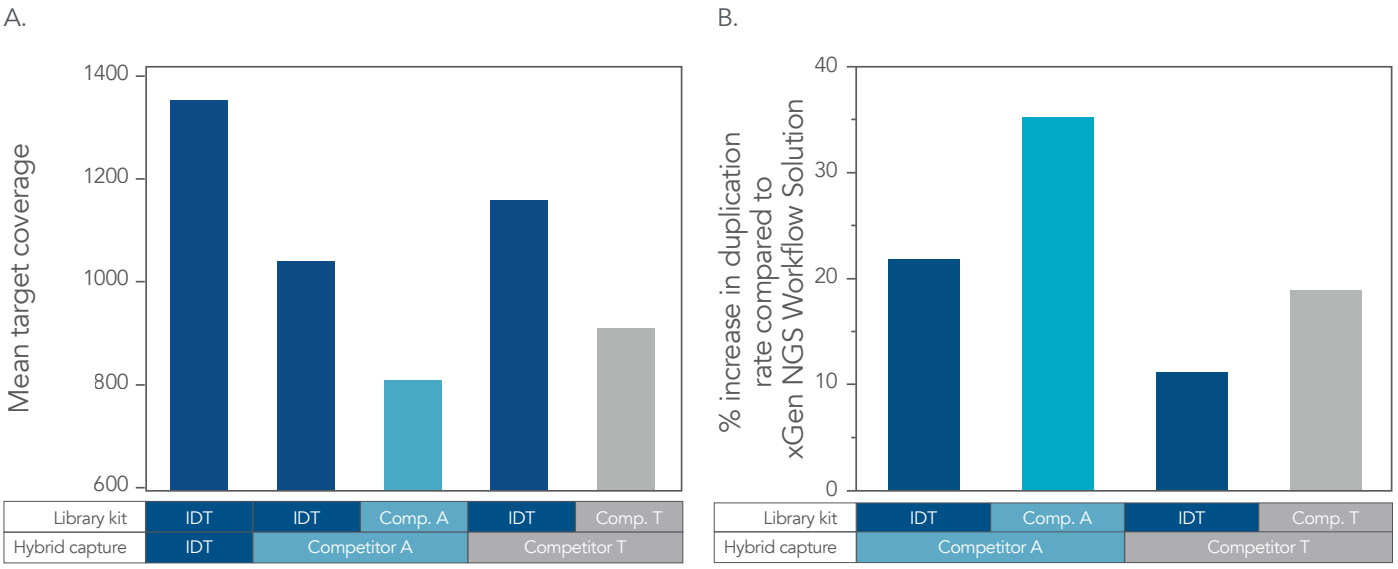


Figure 6. IDT's xGen NGS Workflow Solution's conversion efficiency results in higher target coverage and lower duplication rates. When the IDT xGen cfDNA & FFPE DNA Library Prep Kit was used in place of the competitors' library kits within each competitor's workflows, there was an increase in mean target coverage (**A**) and decrease in duplication rate (**B**) compared to the full workflows from competitor A and T. The bar graph represents the mean of the 24 replicates per workflow.

Conclusion

Due to the unique splint-ligation process, novel enzyme, and thoughtful adapter design, the xGen cfDNA & FFPE DNA Library Prep Kit results in a high conversion efficiency leading to higher complexity libraries than both tested competitors. The xGen NGS Hybridization Capture Panel's unique design strategy results in efficient hybridization capture which provides researchers with higher target-specificity and better sequencing uniformity. By testing the competitors' capture technology with both their own DNA library prep kit and with the IDT xGen cfDNA & FFPE DNA Library Prep Kit, the data clearly shows that the IDT xGen library prep kit results in superior library complexity, highlighting the beneficial chemistry of the kit. Notably, the xGen NGS Workflow Solution data presented here, exhibited the best quality across key NGS metrics, demonstrating that IDT xGen components are optimized to work together, and that researchers can expect the highest quality results by combining these technologies in a streamlined sequencing workflow.

For more information about NGS workflows, methods, and applications, download IDT's Targeted sequencing guide or visit www.idtdna.com/NGS.

Methods

General parameters and hybrid capture panel design

For an unbiased evaluation, a third-party research organization was used to perform library preparation, hybridization capture, and sequencing for all experiments. All vendors assessed (IDT, competitor A, and competitor T) were provided the same 587 oncology-relevant genomic targets and each independently designed a custom hybridization capture panel.

Library prep and hybrid capture

NGS libraries (*n* = 24 per vendor) were generated using 100 ng of Covaris sheared gDNA (Coriell NA12878) and 7 cycles of PCR. Libraries were captured as two, 12-plexes per vendor using their respective custom hybridization capture panel, associated reagents, and protocol from each vendor, with 14 cycles of post-capture PCR, allowing for a direct comparison of each vendor's complete workflow solution.

To evaluate the benefits of the xGen cfDNA & FFPE DNA Library Prep Kit as part of the xGen Workflow Solution as well as with competitor hybridization capture panels, an additional set of libraries (*n* = 48) were generated from Covaris sheared gDNA (Coriell NA12878) as described above. These libraries were coupled with competitor A and competitor T hybridization capture workflows as two, 12-plexes per competitor. All PCR cycling was the same as described above.

Sequencing and data analysis

All libraries were sequenced on an Illumina® NovaSeq® and subsampled to 7 million reads per sample. Data was analyzed using a common target BED file in Picard (Broad Institute) with hg38 as the reference genome.

Unveiling superior NGS library complexity and target coverage with the xGen™ cfDNA & FFPE DNA Library Prep Kit and hybridization capture solution

For more information, go to: **www.idtdna.com/ContactUs**

> 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 use in 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.

© 2023 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: RUO23-2238_001 08/23

IDT
INTEGRATED DNA TECHNOLOGIES