Rare variant detection in low input samples using a high conversion library preparation and quality sequencing from the PacBio® Onso™ system

Abstract

Combining IDT's xGen™ high-conversion NGS library prep technology and the short-read PacBio Onso platform has demonstrated an ability to detect low allelic frequency variants in low input samples.

Key points:

- Liquid biopsy applications offer oncology researchers important insights, but NGS-based analyses of these samples can be challenging due to the low concentration of cfDNA in samples.
- An effective NGS workflow that uses liquid biopsy samples requires high conversion library preparation approaches and highly accurate sequencing technology.
- The high-conversion capabilities of the xGen cfDNA & FFPE Library Prep Kit coupled with the PacBio Onso sequencer can enable sensitive and accurate variant detection for cfDNA applications.

Using only 1 ng for library prep resulted in high sensitivity and specificity across a range of sequencing depths; and 100% sensitivity rates were achieved for ultra-low allele-frequency of 0.5% using a cfDNA reference standard.

Introduction

As next generation sequencing (NGS) technology continues to advance and sequencing costs continue to decrease, applications of this technology have become increasingly sophisticated. One of the benefits of the advancement of NGS technology in oncology research is the ability to identify rare variants in samples with a low level of genetic input material, e.g., cell-free DNA (cfDNA). Although detecting rare variants in cfDNA can be difficult due to the low abundance of material, it is a highly desired sample type due to the ease of sample procurement (liquid biopsy) relative to the more invasive approach of a tumor biopsy.

Robust NGS workflows in the oncology research space that utilize liquid biopsy samples require workflows that can:

- Efficiently convert low-input samples into sequencing libraries
- Consistently generate high-quality sequencing data
- Reliably detect low-frequency allele variants

To achieve all of these, a purposefully designed NGS workflow from library preparation to sequencing is needed. The work presented here illustrates the utility of the xGen cfDNA & FFPE Library Prep Kit and xGen Custom Hybrid capture panel, paired with the PacBio Onso sequencer, to support oncology research for low-input (1 ng), low copy variants (15 copies), and rare-variant detection (0.5% AF) using cfDNA reference standards.

IDT xGen NGS Solutions coupled with PacBio Onso Sequencing

IDT xGen library preparation and hybrid capture technologies are seamlessly compatible with the PacBio Onso sequencing platform, resulting in a full-solution workflow from template to sequencing data. After library preparation or hybridization capture, the Onso library conversion kit is used to add Onso-compatible sequencing adapters to the prepared libraries. Once sequencing is complete, the Onso system completes on-board demultiplexing of samples and generates FASTQ files that are compatible with most publicly available short-read analysis pipelines (Figure 1).

Because accuracy is crucial for high sensitivity, having purpose-fit technology to generate data with minimal sequencing errors is necessary. The PacBio short-read sequencing platform was designed to focus on high accuracy applications such as with liquid biopsies. The technology is enabled by a novel Sequencing By Binding (SBB) chemistry which achieves greater accuracy through multiple innovations. First, SBB uses rolling-circle amplification for the clustering step, which reduces the propagation of errors introduced during clustering by continuously priming from the original template molecule, rather than error-containing copies. Additionally, SBB separates the interrogation and incorporation steps of the sequencing cycle, such that fluorescently-labeled nucleotides used for base interrogation are not incorporated into the growing DNA strand. This results in a more native DNA molecule, improving performance of the sequencing polymerase. Because of its high accuracy, the SBB chemistry reduces false positive variant calls and increases sensitivity for rare variant detection applications.

The xGen cfDNA & FFPE Library Preparation Kit which utilizes a sequential, single-stranded splint-ligation for maximized sample conversion efficiency, a novel ligase that prevents chimera formations, and adapter modifications that prevent dimer formation is ideally suited to pair with the Onso system. Combining the high conversion library preparation kit with the accuracy of the Onso platform allows for identification of rare events in liquid biopsy research.

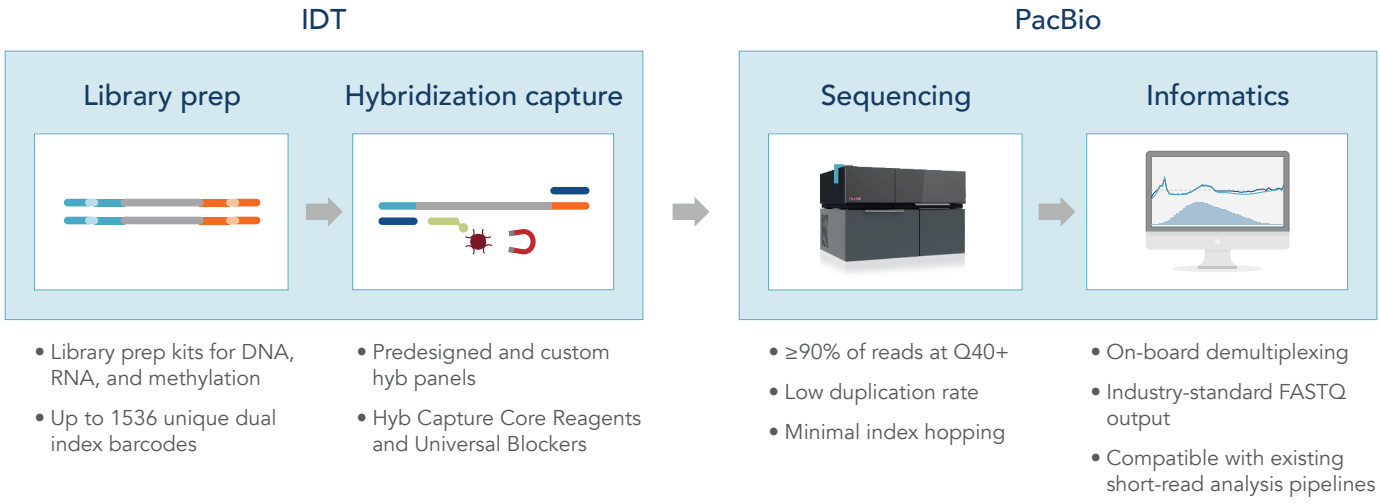

Figure 1. IDT library and hybrid capture technologies pair seamlessly with the PacBio Onso sequencing instrument.

Methods

To test the sensitivity of the workflow based on mass input, a Myeloid DNA Reference Standard (HD829) from Horizon was used to generate libraries using the xGen cfDNA & FFPE Library Prep Kit with xGen UDI Primer Pairs. The reference standard was sheared targeting a 150 bp insert size using a Covaris® to mimic the size of typical cfDNA samples. Libraries were generated in triplicate using low inputs of 1 ng and 10 ng into library preparation.

Similarly, to determine the sensitivity of this workflow based on low allele frequencies, a Myeloid cfDNA Reference Standard (HD838) from Horizon was used to generate libraries with the xGen cfDNA & FFPE Library Prep Kit with xGen UDI Primer Pairs. Libraries were generated in triplicate using 25 ng and 50 ng inputs.

An xGen Custom Hyb Panel was designed to target all variants in the reference standards (HD838 and HD829) as well as other known oncological variants. The hot-spot panel targets space covers 8,890 bases using 111 probes. Triplicate libraries were multiplexed, and hybrid captures (n = 4) using xGen Hybridization Capture Core Reagents were performed using 16-hour hybridizations following the xGen Hybridization Capture Protocol. Captured libraries were then converted using the Onso library conversion kit, which uses 5 cycles of PCR to add the Onso adapter sequences using primers complementary to the existing P5/P7 adapter sequences. All samples were subsampled (1 ng and 10 ng samples: 250K reads/sample, and 25 ng and 50 ng samples: 50M reads/sample) hybrid capture performance metrics were calculated using Picard [1], and variant detection was completed using VarDict (v.1.8.3) without structural variant calling, a minimum allele frequency of 0.001 and minimum allele depth of 2, and with the use of the -A parameter to output multi-allelic variants [2]

Results

Quality performance of the xGen and Onso sequencing workflow

Using cfDNA for sequencing can be challenging as it is usually present at low concentrations in a liquid biopsy sample, providing minimal starting material for library preparation. This was mimicked through the use of low reference standard inputs (1 ng and 10 ng) into library preparation and using Covaris®-sheared samples with a mean insert size of 165 bp, mimicking the typical insert size of cfDNA. All libraries, even the low input of 1 ng, had >99.9% reads aligned to the genome, with <0.3% chimera rate. This demonstrates the ability of the xGen cfDNA & FFPE Library Prep Kit to minimize dimers and chimera formation even with low inputs in library preparation, thus resulting in fewer wasted sequencing reads.

Coverage and library complexity are directly related to library preparation input. The fewer molecules in, the lower the complexity, and the lower the mean target coverage. As expected, the coverage and complexity increased when input increases (Table 1 and Table 2). The libraries which used 1 ng of sample input were still able to achieve a high mean target coverage and a complexity of > 9.52E+03, which was sufficient for rare variant detection. The ability to retain high complexity and coverage with the 1 ng input samples highlights the conversion efficiency of the xGen cfDNA & FFPE Library Prep Kit. This level of conversion efficiency resulted in a high number of the original library molecules being sequenced, thus indicating the sequencing data represents the original sample.

Table 1. Hybrid capture metrics for low input library preparation.

Input (ng)	Mean target coverage	HS library size	Selected bases	Fold-80 base penalty
1	143	9.52E+03	76%	1.10
10	917	6.26E+04	78%	1.00

Table 2. Consistent hybrid capture metrics in cfDNA reference standards.

Input (ng)	Mean target coverage	HS library size	Selected bases	Fold-80 base penalty
25	2260	1.76E+05	77%	1.02
50	4271	3.18E+05	78%	1.01

For this work, an xGen Custom Hyb panel was designed to target AML related variants resulting in a small panel containing 111 probes covering a ~9 kb target space. One difficulty with small target capture panels is the potential for high amounts of off-target reads. Reads outside of the target space are removed during analysis, and therefore reduce sequencing efficiency. Here, the xGen Custom Hyb panel had an on-target rate (percent selected bases, a measure of design sequence specificity [1]) of ≥75% across all replicates with a range of 2.7% (Table 1 and Table 2); highlighting the consistency of the panel performance across multiple captures and thus minimizing the number of reads lost during analysis.

Uniformity of sequencing reads across the target space is important for downstream analyses including variant detection. Ideally sequencing results in a high level of uniform coverage across all the targets. If within a panel, some targets are consistently under-sequenced, more reads will be needed to reach the desired coverage level, meaning that other high-performing targets will be over-sequenced resulting in waste of sequencing reads. Fold-80 base penalty is a metric that measures uniformity by calculating the fold over-coverage necessary to raise 80% of bases in non-zero coverage targets to the mean coverage with the optimal fold-80 score being 1 [1]. The xGen Custom Hyb panel tested here had a fold-80 between 1.0 and 1.1 (Table 1 and Table 2). The panel had one target with zero-coverage that was excluded from analysis. These uniformity metrics are exceptional, representing a highly uniform coverage across the target space and thus a more even distribution of sequencing resources. Primary sequencing metrics from the Onso system were similarly impressive, with 96% of reads at ≥Q40 (predicted error rate of 1 in 10,000 bases) and 83% of reads at ≥Q50 (predicted error rate of 1 in 100,000 bases).

Taken together, the consistent and high on-target rate with the exceptional uniformity (Table 1 and Table 2) highlights the ability of IDT's proprietary panel design to maximize on-target rates and uniform capture. The quality hybrid capture metrics generated using the xGen cfDNA & FFPE Library Prep Kit with xGen Custom Hyb panel combined with the remarkable accuracy of the Onso sequencing platform, demonstrate the high-quality sequencing results generated using the IDT and PacBio workflow.

Detection of low frequency variants using a high conversion library preparation with an accurate sequencing technology

To test the sensitivity of variant calling, libraries were prepared with low inputs using either 1 ng or 10 ng of reference standard (HD829) that contained known variants at a frequency down to 5%. Table 3 shows the conversion of a 5% variant allele frequency into an estimated genomic copy number. The lowest input used (1 ng) resulted in 15 copies of the variants present going into library preparation.

Table 3. Conversion of library input in nanograms into estimated genomic copy number of the lowest frequency variants.

Input (ng)	Genomic equivalents	Variant copy number (5%)
1	300	15
10	3000	150

The reference material contained 20 expected variants including single nucleotide variants (SNV), and insertions and deletions (indels), ranging in allele frequencies from 5–70%. All variants in all replicates were identified in the 10 ng samples, and all variants were identified in two of the three 1 ng replicates, with only one variant missing in the third replicate (Table 4), indicating the remarkable conversion efficiency of the original starting material.

Table 4. Mean observed allele frequencies for expected variants with low input samples.

					1 ng		10 ng	
Gene	ID	Ref	Alt	Expected AF (%)	Observed AF % (replicates with variant call)			
ABL1	COSV59323790	C	T	5	5.7	(3/3)	4.5	(3/3)
ASXL1	COSV60102155	A	AG	40	19.0	(3/3)	8.0	(3/3)
ASXL1	COSV60115219	G	T	5	4.8	(3/3)	4.7	(3/3)
BCOR	COSV60698892	G	GT	70	64.0	(3/3)	60.4	(3/3)
CBL	COSV50649507	C	T	5	6.3	(3/3)	4.9	(3/3)
CBL	COSV50649507	C	A	5	4.1	(3/3)	5.1	(3/3)
DNMT3A	COSV53036332	G	A	5	4.4	(3/3)	4.0	(3/3)
EZH2	COSV57458095	C	T	5	4.5	(3/3)	4.8	(3/3)
FLT3	COSV54042116	C	A	5	4.8	(3/3)	4.6	(3/3)
GATA1	COSV64962576	C	T	10	7.9	(3/3)	10.5	(3/3)
GATA2	COSV62003348	AC	A	35	33.6	(3/3)	32.9	(3/3)
IDH1	COSV61615256	G	A	5	4.7	(3/3)	3.1	(3/3)
IDH2	COSV57468734	C	T	5	4.1	(3/3)	4.5	(3/3)
JAK2	COSV67579858	TTCACAA	T	5	5.2	(3/3)	3.5	(3/3)
JAK2	COSV67569051	G	T	5	6.2	(2/3)	4.9	(3/3)
KRAS	COSV55497388	C	T	40	41.3	(3/3)	38.4	(3/3)
NPM1	COSV51542664	C	CTCTG	5	4.9	(3/3)	4.1	(3/3)
NRAS	COSV54736624	T	A	10	11.8	(3/3)	9.6	(3/3)
RUNX1	COSV55866591	C	T	35	35.3	(3/3)	32.1	(3/3)
SF3B1	COSV59205460	C	T	5	4.5	(3/3)	5.8	(3/3)
TET2	COSV54396706	G	A	5	4.9	(3/3)	4.4	(3/3)
TP53	COSV52661688	G	A	5	5.1	(3/3)	4.7	(3/3)

Simulating a range of sequencing depths, the data was subsampled across a variety of levels from 250K/sample to 10K/sample to understand what the lowest number of reads that could be used before observing a decrease in sensitivity. At 250K reads, the specificity rate for all samples was >99.7%; more than 99.9% sensitivity for 10 ng samples, and >99.5% for the 1 ng samples. High sensitivity was maintained until 24K reads/sample at which point there was a slight decrease in sensitivity rates, however high specificity was maintained (Figure 2), illustrating the ability of the Onso system to reduce false positive variant calls through its improved SBB chemistry. With the sequencing chemistry of the Onso system and the high conversion efficiency of the xGen library preparation kit, fewer reads could be used while maintaining high sensitivity even with low-input samples with rare alleles. This benefit offers the potential to reduce sequencing costs for labs by reducing the reads needed per sequencing run.

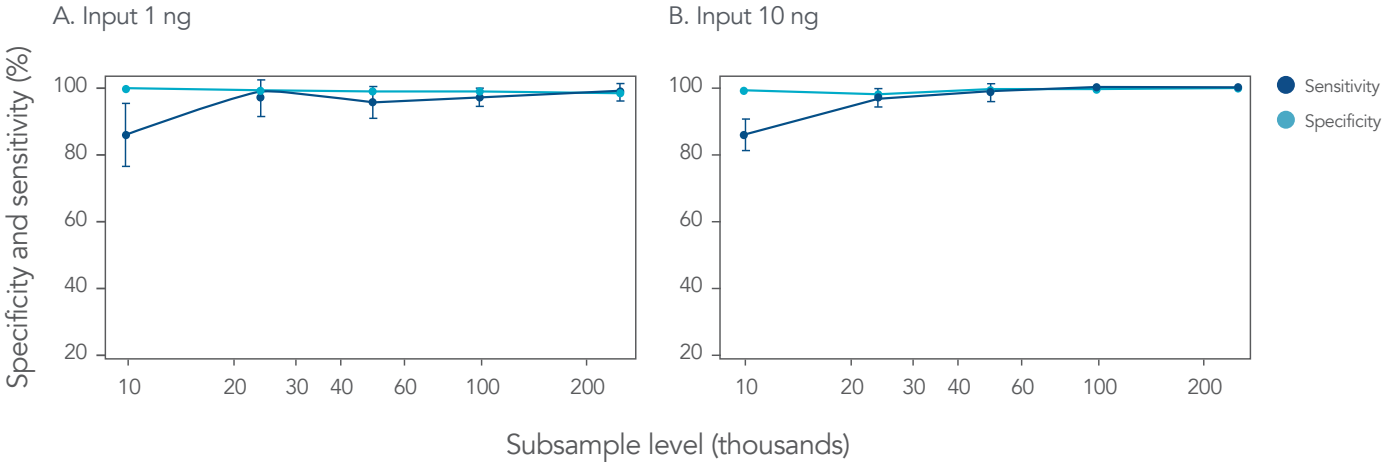

Figure 2. High sensitivity and specificity across a range of sequencing depths. Three libraries were generated with the xGen cfDNA & FFPE Library Prep Kit using 1 ng and 10 ng inputs of a reference standard with 20 known variants in our target space. Libraries were multiplexed (*n* = 3), captured with a xGen Custom Hyb panel (*n* = 2), and sequenced on the PacBio Onso system using 2 x 150 PE reads. Each library was subsampled to various depths and variants were called using VarDict [2]. Graph represents the mean of the three replicates and the error bars represent standard deviations.

To test the sensitivity of the workflow at ultra-low allele frequencies a reference standard (HD838) that contained known variants at a frequency down to 0.5% was used. The reference material contained 15 expected variants including SNV, indels, ranging in allele frequencies from 0.5–1%. All variants in all replicates were identified across inputs (Table 5). The high conversion library preparation kit resulted in a sensitivity of 100% and a specificity of >99.7% for all replicates (Figure 3), again highlighting the ability for the xGen with PacBio workflow to support liquid biopsy oncology research.

Table 5. Mean observed allele frequencies with ultra-low variant allele frequencies.

					25 ng		50 ng	
Gene	ID	Ref	Alt	Expected AF (%)	Observed AF % (replicates with variant call)			
ABL1	COSV59323790	C	T	0.5	0.6	(3/3)	0.4	(3/3)
ASXL1	COSV60115219	G	T	0.5	0.5	(3/3)	0.4	(3/3)
CBL	COSV50649507	C	T	0.5	0.3	(3/3)	0.4	(3/3)
DNMT3A	COSV53036332	G	A	0.5	0.4	(3/3)	0.4	(3/3)
EZH2	COSV57458095	C	T	0.5	0.5	(3/3)	0.4	(3/3)
FLT3	COSV54042116	C	A	0.5	0.4	(3/3)	0.4	(3/3)
IDH1	COSV61615256	G	A	0.5	0.3	(3/3)	0.3	(3/3)
IDH2	COSV57468734	C	T	0.5	0.5	(3/3)	0.5	(3/3)
JAK2	COSV67579858	TTCACAA	T	0.5	0.2	(3/3)	0.3	(3/3)
JAK2	COSV67569051	G	T	0.5	0.5	(3/3)	0.3	(3/3)
NPM1	COSV51542664	C	CTCTG	0.5	0.3	(3/3)	0.3	(3/3)
NRAS	COSV54736624	T	A	1	1.0	(3/3)	1.0	(3/3)
SF3B1	COSV59205460	C	T	0.5	0.4	(3/3)	0.5	(3/3)
TET2	COSV54396706	G	A	0.5	0.4	(3/3)	0.4	(3/3)
TP53	COSV52661688	G	A	0.5	0.4	(3/3)	0.6	(3/3)

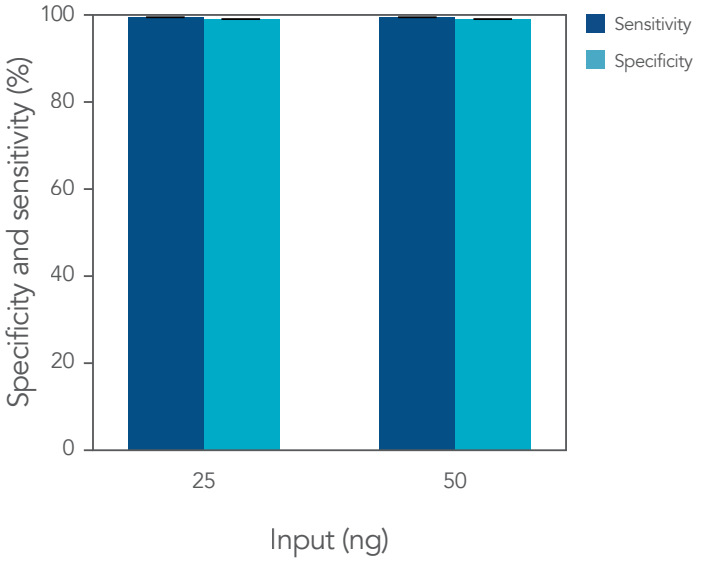

Figure 3. High sensitivity and specificity with ultra-low variants. Three libraries were generated with the xGen cfDNA & FFPE Library Prep Kit using 25 ng and 50 ng inputs of a reference standard with 15 known variants in our target space. Libraries were multiplexed (*n* = 3), captured with a xGen Custom Hyb panel (*n* = 2), and sequenced on the PacBio Onso system using 2 x 150 PE reads. Each library was subsampled to 50 M reads and variants were called using VarDict [2]. Error bars represent standard deviations.

Conclusions

The rise of new, short-read sequencing technologies has allowed oncology scientists to quickly identify relevant genetic variants using sampling approaches that are easier and more accessible like liquid biopsy. While these approaches have proven to be advantageous, they present nontrivial challenges such as low input concentrations, limited fragment size, and low quality–i.e., cfDNA in liquid biopsy samples. Therefore, it is critical researchers have an NGS workflow that generates high-quality sequencing libraries with accurate sequencing technology for reliable insights.

The data presented above illustrates the efficacy of an NGS oncology workflow that utilizes the xGen cfDNA & FFPE Library Prep Kit, xGen Custom Hyb panel, and PacBio Onso sequencer. With inputs as low as 1 ng of reference standard, this workflow resulted in high-quality sequencing libraries, high uniformity and on-target rates, consistent identification of all expected low frequency variants, and maintained detection sensitivity using as little as 24K reads/sample. The workflow was able to correctly identify all ultra-low frequency variants (0.5%) in as little as 25 ng of a cfDNA reference standard. The high conversion xGen NGS technology coupled with the remarkable accuracy of the PacBio Onso platform provides researchers with an efficient solution for low frequency variant detection in low input samples.

References

1. "Picard Toolkit.". Broad Institute, GitHub repository: Broad Institute; 2019.
2. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44(11):e108.

Rare variant detection in low input samples using a high conversion library preparation and quality sequencing from the PacBio® Onso™ system

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**.

PacBio Compatible

IDT INTEGRATED DNA TECHNOLOGIES