Application Note

Whole Genome Sequencing and High-Efficiency Library Preparation for Generating and Detecting Minimal Residual Disease (MRD) Signatures

Abstract

Genomic analysis of cancer FFPE tissue and cell-free DNA (cfDNA) is redefining early detection, personalized treatment, and disease monitoring of cancer. The potential utility of these assays in clinical settings relies on simple and efficient workflows as well as low operation costs. Here, we use the IDT xGen™ cfDNA & FFPE Library Prep Kit to prepare FFPE and cfDNA libraries for Whole Genome Sequencing (WGS) using the UG 100™ sequencer and demonstrate the utility of these combined technologies for identifying mutation signatures in low-quality FFPE and sensitive detection of circulating tumor DNA (ctDNA) in clinical research plasma samples with low tumor fractions.

Highlights

- IDT's xGen cfDNA & FFPE Library Prep Kit with xGen Indexing Primers for Ultima produce high complexity libraries compatible with the UG 100™ sequencer
- Whole Genome Sequencing (WGS) from FFPE tumor samples improves MRD detection via large mutational signatures
- Combination of IDT's xGen cfDNA & FFPE Library Prep Kit and UG 100 sequencing technology detects circulating tumor DNA (ctDNA) in plasma samples at tumor fraction as low as $3x10^{-5}$

Introduction

Minimal Residual Disease (MRD) analysis detects the presence of cancer by interrogating cfDNA isolated from plasma. Plasma contains variable amounts of cfDNA that are shed from cells in healthy tissues, inflamed or injured tissues, tumors, or other diseased tissues when they are present. Identification of ctDNA in plasma can be used for both therapeutic decision-making as well as for early detection of cancer to improve the prognostic outcome. To detect ctDNA within cfDNA, a tumor mutational signature is generated from solid-tumor and nondiseased tissue samples based on analysis of somatic and germline mutations. This mutational signature is then used to differentiate ctDNA from the cfDNA. A major challenge in using liquid biopsy for tumor DNA, especially after treatment, is that the ctDNA is a small fraction of the circulating cfDNA and the amount of cfDNA is often very low. Therefore, it is crucial to select a library prep method that can convert as many genomic fragments as possible into sequenceable library. In this way, maximal information can be extracted from these limited samples.

The molecular characterization of genomic DNA from cancer samples has largely focused on the presence of driver mutations (SNVs, indels and CNVs) within coding regions of genes, uncovering the association of specific mutations with mechanisms of oncogenic processes. As a result, there is a natural tendency to focus on these coding regions for signatures of cancer mutations using target enrichment. However, the use of WGS allows for the detection of many more passenger mutations across the genome that may be useful in defining larger oncological signatures and has been shown to increase the sensitivity of circulating tumor DNA

detection. While historically using WGS for tumor characterization and ctDNA detection has been cost prohibitive, the UG 100 sequencing platform changes that paradigm. By delivering affordable deep whole genome sequencing and extremely high data quality, the UG 100 enables WGS at scale for tumor characterization and MRD detection. Furthermore, the UG 100 flow chemistry results in very low error rates for single nucleotide substitutions (SNVs) facilitating extremely low background noise and high sensitivity for this tumor informed detection method.

The IDT xGen cfDNA & FFPE Library Prep Kit uses patented chemistry to drive library conversion to high levels. The kit achieves this with a 2-step ligation using a novel ligase and highly modified adapters to reduce adapter dimers and ensure efficient addition of both adapter strands. The adapter includes inline Unique Molecular Identifiers (UMIs) to enable error correction.

In this retrospective MRD study we utilize the throughput of the UG 100 sequencer to get WGS data for matched samples from donors. First, we generate tumor-specific mutation signatures using low-quality FFPE and, second, we use these signatures to estimate probable tumor fraction in cfDNA.

Results

Solid tumor mutational signatures in low-quality FFPE

Matched samples were obtained from 9 research donors, with a variety of confirmed cancer types (Table 1), from the MIDGAM biobank (Israel National Biobank for Research; MID-116-2019). An FFPE fixed tumor section and a sample of whole blood were collected from each donor, and for 4 donors a fresh frozen (FF) tumor section was also obtained. Fresh frozen samples are a simpler source of tumor DNA as they are more amenable for DNA extraction and do not incur the same damage as an FFPE sample that undergoes the fixation process. However, FF tissue is not always available, and it is much more common for a tumor biopsy to be fixed and embedded for traditional pathology assays and for easier archival storage. This means that an approach which can utilize FFPE samples is more likely to have broader practical application.

Donor	Tumor Type
Pa_46	Lung
Pa_47	Lung
Pa_67	Bladder
Pa_68	Colon
Pa_69	Colon
Pa_70	Colon
Pa_71	Lung
Pa_73	Colon
Pa_75	Colon

Table 1. List of donors and confirmed cancer type.

Because extracted DNA from FFPE samples tends to be damaged and of low-quality, it can be difficult to obtain quality libraries for sequencing, and quality libraries are essential for identifying mutation signatures. To demonstrate the ability of the xGen cfDNA & FFPE Library Prep Kit to generate quality libraries from FFPE samples, DNA was extracted from ~50 µm of FFPE blocks with RecoverAll Total Nucleic Acid Isolation kit. DIN scores (DNA Integrity Number with 1 being degraded and 10 being intact) of the FFPE extracted DNA were determined using Agilent Genomic DNA ScreenTape for TapeStation Systems. Whole-genome libraries were generated from extracted FFPE DNA (library input range from 95–250 ng) using Covaris shearing, NEBNext FFPE DNA Repair kit (part # M6630) followed by the xGen cfDNA & FFPE DNA Library Prep Kit (part # 10010207) with xGen Indexing Primers for Ultima (part # 10016992) and the UG Library Amplification Kit following the published protocol (link to protocol). Sequencing was performed using the UG 100 sequencer and analyzed using start/stop location with UMIs for duplicate removal. Here, even severely degraded (DIN scores range: 1.4–3.1) FFPE samples produced libraries with high complexity as shown by deep coverage with low duplication rate (Figure 1), highlighting a high sample conversion rate in library prep.

Key next generation sequencing (NGS) metrics for coverage and duplication rate were used to conclude the generation of high complexity libraries (Figure 1) Tumor mutational signatures were determined by comparing variant detection results of the FFPE sample against those of a normal

(non-tumor) sample from the same donor. The normal DNA was extracted from > 10^6 peripheral blood mononuclear cells (PBMCs), using Qiagen DNeasy Blood & Tissue Kits (part # 69504), and was sequenced to at least 50X mean coverage. Variants were called using Deep Variant and a signature of tumor-specific mutations was created for each sample. To confirm the utility of tumor mutation signatures from the FFPE samples, four donor-matched pairs of fresh frozen DNA was extracted from 18–40 mg of matched FF tissue using AllPrep DNA/RNA Mini Kit and libraries were generated (inputs range from 300–500 ng). Deep Variant was used to call variants from the FF WGS data and a tumor signature was generated by comparing variants called in these samples to those of the matched PBMC data.

To compare the FFPE mutation signature to the FF mutation signature, allele frequencies of the tumor-specific mutations were compared for variants found in both tumor samples. The FF samples incur less

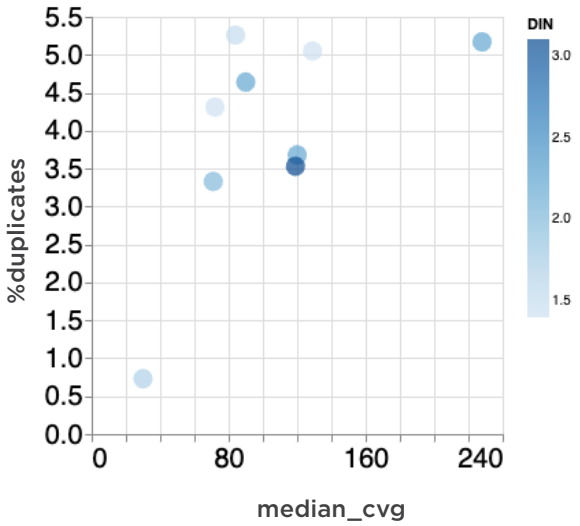
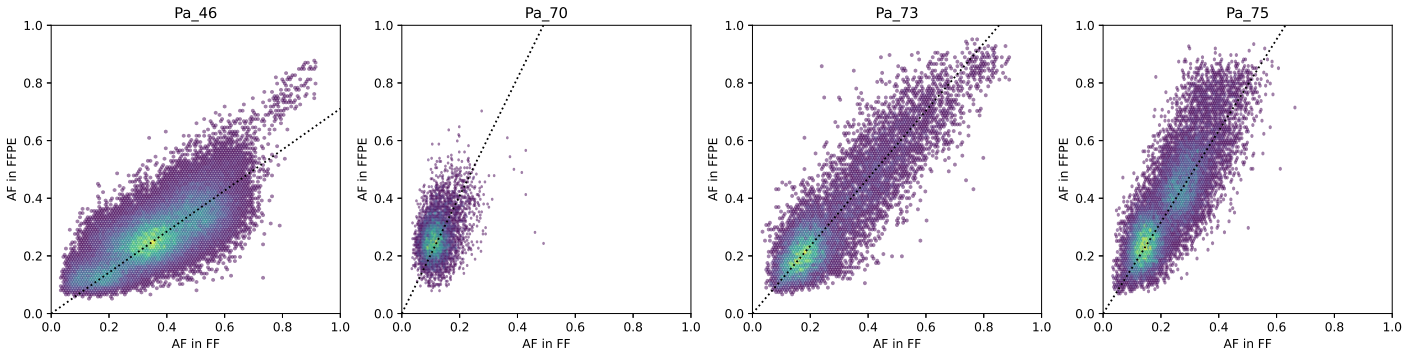


Figure 1: Low duplication rates and high coverage with degraded FFPE samples. The percent duplication, as a function of median coverage per FFPE sample. Color scale represents DIN value for each FFPE sample.



Donor	DIN score	Expected	Evaluated	False-negative	False-positive	True-positive	Recall %	Precision %
		(Variants detected in FF)	(Variants detected in FFPE)	(Variants detected in both FF and FFPE)	(Variants detected in both FF and FFPE)	(Variants detected in both FF and FFPE)	(FF variants detected in FFPE)	(FFPE variants detected in FF)
Pa_46	2.2	68,110	65,942	2,931	763	65,179	95.7	98.86
Pa_70	1.4	4,723	5,383	175	835	4,548	96.29	84.51
Pa_73	1.4	15,155	14,800	1,704	1,349	13,451	88.76	90.93
Pa_75	1.5	20,815	23,129	1,557	3,871	19,258	92.52	83.34

Figure 2: Confirmation of mutational signatures. Shown is the allele fraction (AF) as measured in the FFPE sample (y-axis) and the matching AF as measured in the fresh frozen sample. Bottom table shows quantitative evaluation of the FFPE variants when using the FF variants as ground truth.

damage than the FFPE samples and are used as a ground truth in this experiment. The degraded FFPE samples maintain similar allele frequencies as the matched FF samples, showing that using xGen cfDNA & FFPE Library Prep Kit retains mutation signatures from damaged samples with low DIN scores (Figure 2). The allele frequencies are correlated but are not identical, likely due to different tumor purity levels in the FF and FFPE sample material. More importantly, mutations in both tissue sources are detected with high levels of consistency (recall and precision) (Figure 2).

Low tumor fraction detection of ctDNA for Minimal Residual Disease (MRD)

Plasma samples collected from cancer donors from Table 1, as well as from an additional 9 healthy donors, were sourced from the MIDGAM biobank. cfDNA was extracted from these plasma samples with cfPure® V2 Cell Free DNA Extraction Kit. Whole-genome libraries were prepared with the xGen cfDNA & FFPE DNA Library Prep Kit (part # 10010207) with xGen Indexing Primers for Ultima (part # 10016992) and the UG Library Amplification Kit following the published protocol (link to protocol), using a range of sample input amounts (4.2–10 ng). Libraries were sequenced on the UG 100 and duplicates were removed using start/stop locations and UMI analysis. The resulting WGS data shows high complexity libraries despite the low cfDNA sample input into library preparation, highlighting the efficiency of this library prep method.

To support the clinical research utility case of the xGen cfDNA & FFPE Library Prep Kit paired with whole-genome sequencing on the UG 100 for MRD, it is essential that the tumor fraction for cfDNA be distinguishable from background noise. Here, the list of FFPE mutations detected in each cancer donor was compared to the list of variants detected in matched cfDNA. As ctDNA makes up a small fraction of overall cfDNA, a highly sensitive technology is needed for accurate detection. The FFPE sample variant list was also compared to variants found in cfDNA extracted from other cancer and healthy donors to test the background level of this assay, ensuring that the ctDNA fraction from the matched donor is above background noise. Because these controls contain mutational signatures of cancer,

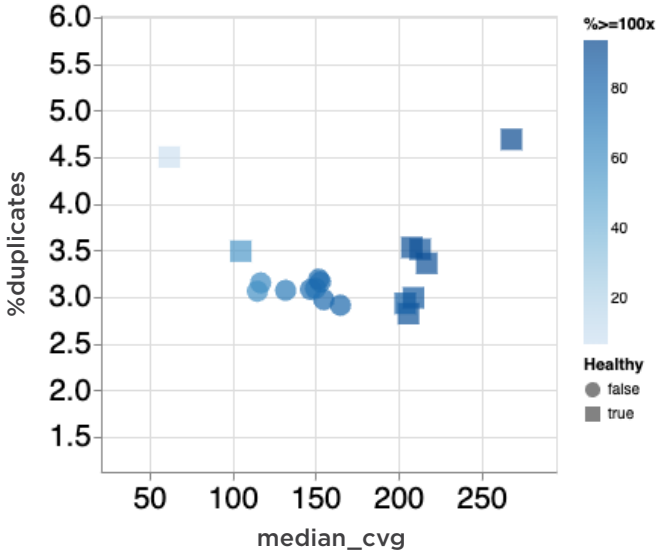
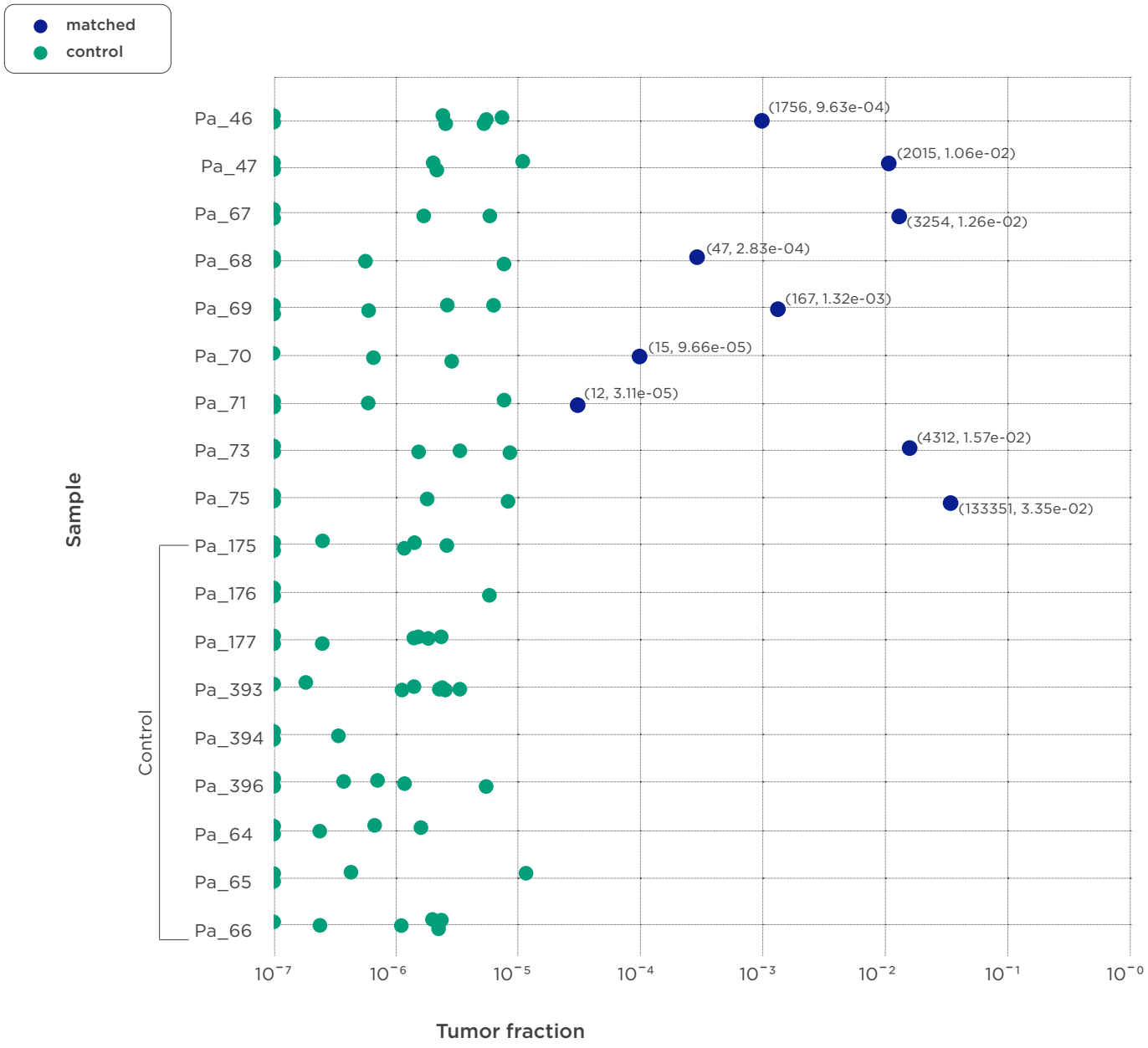


Figure 3: Low duplication rates and high WGS coverage in low input cfDNA samples. Duplication rate as a function of median coverage for the cfDNA extracted from plasma isolated from 9 healthy donors (squares) and 9 cancer sample donors (circles). Color intensity of the marks show the percent of the genome covered at >100X.

quantifying the difference between the matched cfDNA variants and different cancer signatures will demonstrate the specificity of this method. It is expected that the mutational profile for matched tumor FFPE and cfDNA samples will have a higher overlap than when comparing with non-matched samples or samples from healthy donors. This is because the cfDNA should contain the ctDNA that has the same mutational signature as the FFPE tumor within the individual, allowing for longitudinal monitoring using blood draws as opposed to biopsy.

Tumor fractions were calculated as the fraction of observed tumor-supporting reads out of the total relevant reads at these positions given the tumor mutational signature size and the sequencing depth. As Figure 4 shows, for all matched samples from the same donor (FFPE vs cfDNA), the estimated cfDNA tumor fraction (blue) was above that of the background (green), showing the ability of this method to identify ctDNA in cfDNA across multiple cancer types even at very low tumor fraction levels. Furthermore, the control samples did not show tumor fractions higher than the matched tumor samples highlighting the tumor fractions for ctDNA being higher than background noise. Impressively, tumor fractions as low as $3x10^{-5}$ were detected above the noise level highlighting the utility of this technology for MRD research.

Figure 4: Tumor fraction levels in matched cfDNA compared to background. Each row shows an evaluation of the variants detected in FFPE tumor compared against the mutational signature of cfDNA extracted from either the matching donor (blue circles) or from mutational signatures taken from other cancer donors (green). The x-axis shows the estimated tumor fraction. For cancer donors the numbers over each blue circle specify the number of overlapping variants detected in both cfDNA and FFPE (left) as well as the resulting estimated tumor fraction (right).



Conclusions

Here, we present confident identification of low-frequency variants in matched FFPE tissue and cfDNA samples by combining the IDT xGen cfDNA & FFPE DNA Library Prep Kit with Ultima Genomics UG100 sequencing. Matched biobank cancer samples were prepared for sequencing using the xGen cfDNA & FFPE DNA Library Prep Kit, which is optimized for low-input and degraded samples such as FFPE and cfDNA. The library kit generated high-complexity libraries, even with the most degraded FFPE samples, allowing for tumor-specific variant identification with deep whole genome sequencing. Using variants identified in the FFPE tumor sample and comparing to the matching subjects' plasma cfDNA and control cfDNA, the tumor fraction of the matched ctDNA was found to be above the background level of the assay, suggesting a workflow suitable for sensitive liquid biopsy applications, such as MRD research. UG 100 sequencing of whole genome xGen cfDNA & FFPE DNA libraries demonstrated:

- High conversion rates resulting from novel ligase and proprietary adapters
- High-complexity libraries enabling generation of whole genome mutational signatures
- High library complexity from severely degraded FFPE samples and low input samples
- A single, streamlined workflow for analysis of tumor-associated variants in matched cfDNA and FFPE samples
- A sensitive and straightforward assay for MRD research workflows
- Detection of tumor presence in clinical research samples with tumor fraction as low as $3x10^{-5}$ above the background level.