illumina®

APPLICATION NOTE

Unlock whole-genome sequencing from FFPE samples

A modified Illumina DNA Prep workflow enables accurate sequencing from a range of DNA input quality and amounts

Optimized library preparation protocol enables sensitive variant calling from FFPE DNA

Powerful sequencing on Illumina high-throughput systems provides comprehensive coverage

End-to-end workflow includes highly efficient DRAGEN™ analysis pipelines for accurate WGS results

Introduction

Whole-genome sequencing (WGS) gives researchers the ability to investigate all variant types in a single assay, including copy number variation, structural variants, potential fusion genes, and homologous recombination deficiencies, which may be missed by targeted gene panels.

High-quality DNA input from tissue samples is recommended to obtain accurate and reliable WGS data. Formalin-fixed paraffin-embedded (FFPE) tumor samples are one of the most accessible sources of tissue for molecular profiling. However, the fixation, embedding, and extraction process can lead to DNA degradation, fragmentation, crosslinking, and chemical modifications that can compromise DNA quality,[1,2] making WGS challenging. Optimized workflows are necessary to enable WGS from DNA extracted from FFPE samples.

This application note demonstrates a streamlined, end-to-end workflow using Illumina DNA Prep to generate high-quality WGS data from FFPE-derived DNA. Minor modifications to the tagmentation-based Illumina DNA Prep protocol, sequencing on powerful Illumina high-throughput systems, and DRAGEN secondary analysis deliver a high performance WGS tumor–normal workflow when starting with FFPE-derived DNA (Figure 1).

Methods

Samples

Two fresh-frozen (FF) cell lines, HCC1187 (ATCC, Catalog no. CRL-2322) and COLO 829 (ATCC, Catalog no. CRL-1974), were used to assess the impact of the modified library preparation protocol on variant calling performance. DNA extracted from a cohort of 241 FFPE tumor samples with matched blood samples and three tumor FFPE DNA with matched FF tumor and blood DNA samples was provided by external clinical research collaborators. DNA extracted from FFPE tissue samples and matching blood samples was quantified using the Spectramax Gemini XPS microplate reader (Molecular Devices, Catalog no. XPS). DNA quality was measured using the 4200 TapeStation Instrument (Agilent, Catalog no. G2991BA) which provides a DNA integrity number (DIN) ranging from 1 for degraded DNA to 10 for intact DNA.

Library preparation

Libraries were prepared from DNA extracted from FFPE, FF, or blood samples using Illumina DNA Prep as detailed in the Illumina DNA Prep reference guide with some modifications to enable sensitive variant detection from FFPE samples with variable quality DNA (Table 1).

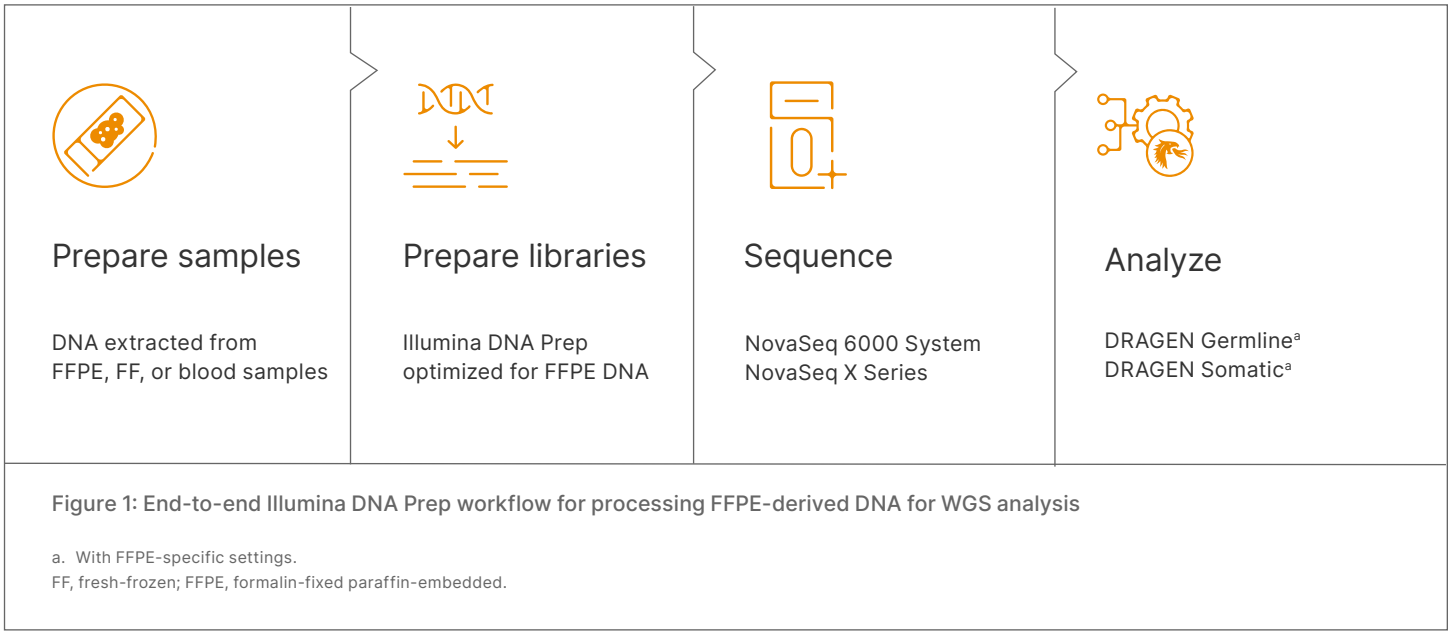

Figure 1: End-to-end Illumina DNA Prep workflow for processing FFPE-derived DNA for WGS analysis

a. With FFPE-specific settings.
FF, fresh-frozen; FFPE, formalin-fixed paraffin-embedded.

Table 1: Modifications made to the Illumina DNA Prep protocol to enable WGS from FFPE samples

	Illumina DNA Prep standard protocol	Illumina DNA Prep FFPE protocol
DNA input	1–500 ng	FFPE: 10–200 ng[a] FF/blood: 10–50 ng
PCR cycles	5–12[b]	6
Size selection	Double-sided bead cleanup	Single-sided bead cleanup with 0.7× size selection
Final elution volume	30 µl	20 µl

a. Recommended DNA input for FFPE samples is 100 ng.
b. Depending on sample input.

Sequencing

Prepared libraries were loaded on to S4 and 10B flow cells and sequenced on the NovaSeq™ 6000 or NovaSeq X Plus Systems, respectively, with a run configuration of 2 × 151 bp to enable high coverage and accuracy for WGS.

Data analysis

After sequencing was complete, data were analyzed using DRAGEN Germline and DRAGEN Somatic pipelines with FFPE-specific settings. For the DRAGEN Germline pipeline, the option “"--qc-coverage-ignore-overlaps true” was used to make sure effective coverage was measured correctly for samples with short fragments. Because FFPE samples show a high number of false positive calls for small somatic variants in specific repeat regions, including L1P and Alu repeats, the DRAGEN Somatic pipeline was run with the additional option “--vc-excluded-regions-bed" and a BED file with the location of L1P and Alu repeats. Somatic analyses were performed with ‘Nucleotide (NTD) Error Bias Estimation’ kept on to correct for systematic error patterns that can occur during the processing of FFPE samples. The systematic noise filter was used to reduce false-positive calls by accounting for site-specific noise. The Genomics England London (GEL) solid adult cancer susceptibility gene list was used for somatic variant detection.

Results

Somatic variant recall in FF cell lines using the modified library preparation protocol

The modified Illumina DNA Prep FFPE protocol was used to assess single nucleotide variant (SNV) and insertion-deletion (indel) recall from two FF cell lines, HCC1187 and COLO 829. The findings demonstrate high recall of somatic SNVs and indels for both samples, confirming that the modified FFPE protocol was suitable for use with FF cell lines (Figure 2).

WGS coverage across a range of FFPE-derived DNA quality and input amounts

The modified Illumina DNA Prep FFPE protocol was used to prepare libraries in a cohort of 241 breast cancer samples with varying levels of DNA degradation. DINs were used to assess DNA quality for all samples. The percentage of the genome with more than 50× effective coverage was calculated for all samples with variable DIN values across a range on input amounts. The results demonstrate that the FFPE-specific Illumina DNA Prep protocol enabled WGS with an effective depth exceeding 50× in samples with DIN greater than 2.5 (Figure 3). FFPE DNA samples with DIN below 2.5 showed increased variability in sequencing coverage. Increasing DNA input amounts resulted in improved effective coverage for samples with DIN greater than 2.5 (Figure 3). These findings highlight DIN and input DNA concentration as key predictors of sequencing data quality in FFPE-derived DNA samples.

High concordance between FF and FFPE samples

DNA extracted from FF samples is of high quality with reduced degradation and minimal artifacts, making it an ideal input for WGS studies. However, FF samples are not routinely used in clinical research due to cost and practical limitations. FFPE samples are a more accessible and commonly used source of archival tissue. Using the modified Illumina DNA Prep FFPE protocol, a high level of concordance was observed between somatic variants identified in FFPE and matched FF real world tumor samples (Figure 4). These findings demonstrate the feasibility of using this optimized protocol for somatic variant detection in FFPE samples for cancer research.

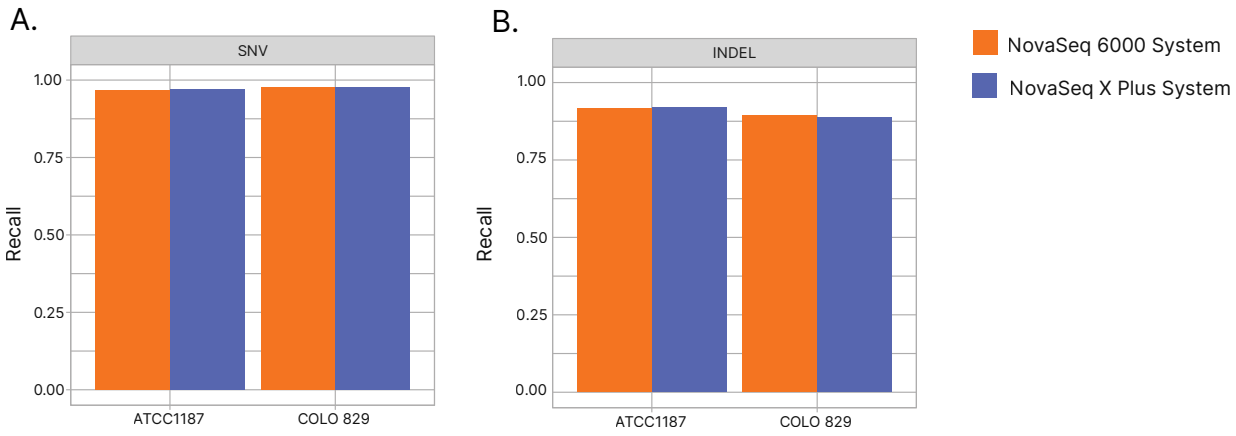

Figure 2: High somatic variant recall in FF cell lines using the modified Illumina DNA Prep FFPE workflow

FF-derived DNA from two cell lines (HCC1187 and COLO 829) prepared with the modified Illumina DNA Prep FFPE protocol and sequenced on NovaSeq 6000 and NovaSeq X Plus Systems shows high recall of single nucleotide variants (SNV) and insertion-deletions (indels) compared to the reference truth set.

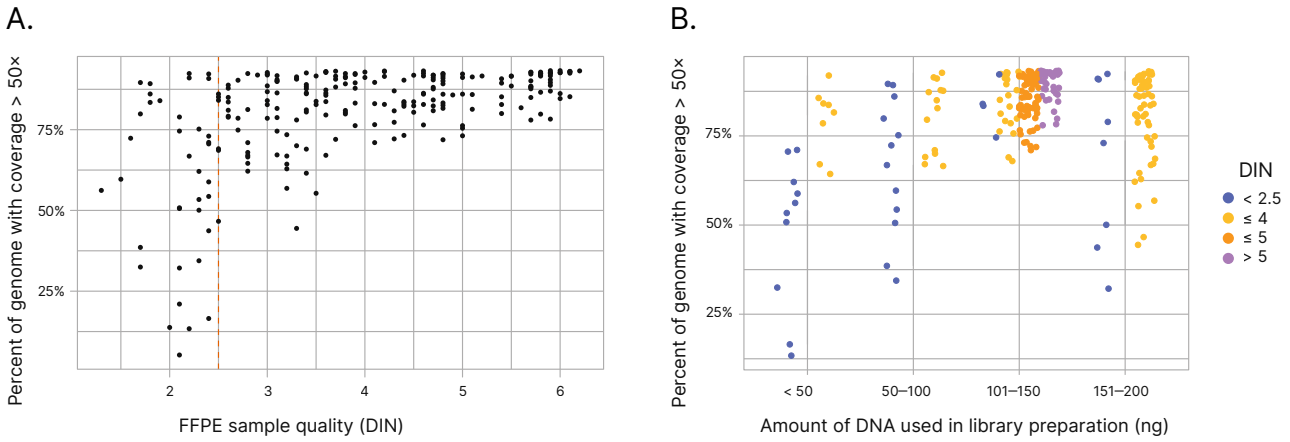

Figure 3: The Illumina DNA Prep solution enables WGS of FFPE-derived DNA samples across a range of quality and input amounts

Percent of the genome with more than 50× effective depth increases as (A) FFPE-derived DNA quality exceeds DIN of 2.5 and (B) DNA input amounts used for library preparation increase, particularly for samples with DIN greater than 2.5. Degraded samples with DIN below 2.5 show greater variability in sequencing performance.

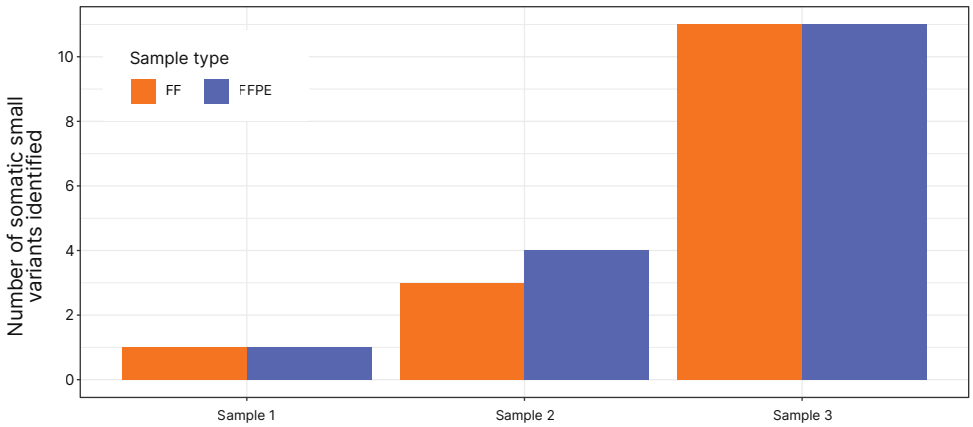

Figure 4: High level of concordance between somatic small variants detected using FFPE and FF samples

DNA from three FFPE and matched FF tumor samples was sequenced using the modified Illumina DNA Prep FFPE protocol for both sample types. The variant detection for variants with VAF ≥ 5% within the GEL adult solid tumors panel gene list for FFPE and FF DNA were highly concordant across all samples tested.

Summary

Generating high-quality WGS data from FFPE tissues requires optimized protocols for library preparation to account for degraded or low-quality input DNA. This application note details a modified FFPE-specific Illumina DNA Prep protocol for generating WGS data from FFPE-derived DNA of variable quality and input. The results demonstrate excellent variant recall and high concordance between somatic variants identified in FFPE and FF samples. With this end-to-end workflow, researchers are empowered to analyze FFPE tissue samples to generate high-quality WGS data.

Learn more →

Illumina DNA Prep

NovaSeq X Series

DRAGEN secondary analysis

DRAGEN recipe for analyzing WGS data from libraries prepared using a modified Illumina DNA Prep FFPE workflow

Prebuilt systematic noise file

References

1. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem*. 2015;61(1):64-71. doi:10.1373/clinchem.2014.223040
2. Bass BP, Engel KB, Greytak SR, Moore HM. A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen?. *Arch Pathol Lab Med*. 2014;138(11):1520-1530. doi:10.5858/arpa.2013-0691-RA

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-03852 v1.0