Streamlining NGS workflows by the application of the DNA Integrity Number (DIN) from the Genomic DNA ScreenTape Assay

Hyunju Jung[1], Sumin Ji[1], Sanghoon Song[1], Yeji Park[1], Ji-Won Yang[2] , Josh Wang[3] and Eva Schmidt[4]
1) Theragen Etex Co., Ltd. 2) Agilent Technologies, Inc., Korea 3) Agilent Technologies, USA 4) Agilent Technologies, Germany

Agilent Technologies

Introduction

Genomic DNA (gDNA) is used as starting material in the experimental workflow of many applications in molecular biology. The integrity of the DNA critically affects the success of many downstream experiments like array CGH or next generation sequencing. Purification techniques for gDNA are fundamental to ensure best results in downstream procedures, but optimized quality control of the purified gDNA is equally important. Therefore electrophoretic analysis of the gDNA sample is highly recommended as the respective downstream applications can be expensive and time consuming. The 2200 TapeStation system in conjunction with the Genomic DNA ScreenTape assay provides an excellent solution for assessing the quantity, integrity and overall quality of gDNA starting material. The ScreenTape device is a pre-packaged microfluidic consumable designed for performing electrophoretic applications in a microscale format. It is used in combination with the 2200 TapeStation instrument.

Degradation of gDNA is typically a gradual process in which high-molecular weight DNA is fragmented into smaller species. It can occur either enzymatically, chemically or mechanically. Judging the integrity of DNA by visual evaluation of the electropherogram trace is subjective and can be error-prone. In order to standardize this assessment, a novel algorithm was developed to score gDNA samples on the 2200 TapeStation system. The DNA integrity number (DIN) is calculated from several features obtained from the electrophoretic trace and provides a numerical value from 1 (degraded) to 10 (intact). The DIN is independent from instrument, reagent batch and sample concentration and can be used as objective measure for determining the integrity of gDNA.

Material and Methods

Material

The M220 Focused-ultrasonicator™ was obtained from Covaris Inc. (Woburn, MA, USA). The 2200 TapeStation system (PN G2965AA) with the TapeStation Analysis software (revision A01.05), Genomic DNA ScreenTape (PN 5067-5365) and Genomic DNA Reagents (PN 5067-5366), SureSelect^{XT} Reagent kit HSQ (PN G9611A), were obtained from Agilent Technologies Inc. (Santa Clara, CA, USA). The HiSeq 2500 system was purchased from Illumina, Inc. (San Diego, CA, USA).

Sample workflow

Genomic DNA was extracted from FFPE blocks from cancer tissue. and then subjected to the SureSelect workflow. QC was performed using the 2200 TapeStation system in combination with the Genomic DNA ScreenTape assay, including the automatic determination of DIN. Library preparation QC was performed using the High Sensitivity D1000 ScreenTape assays. Sequencing was performed with the HiSeq 2500 system according to the manufactures instructions.

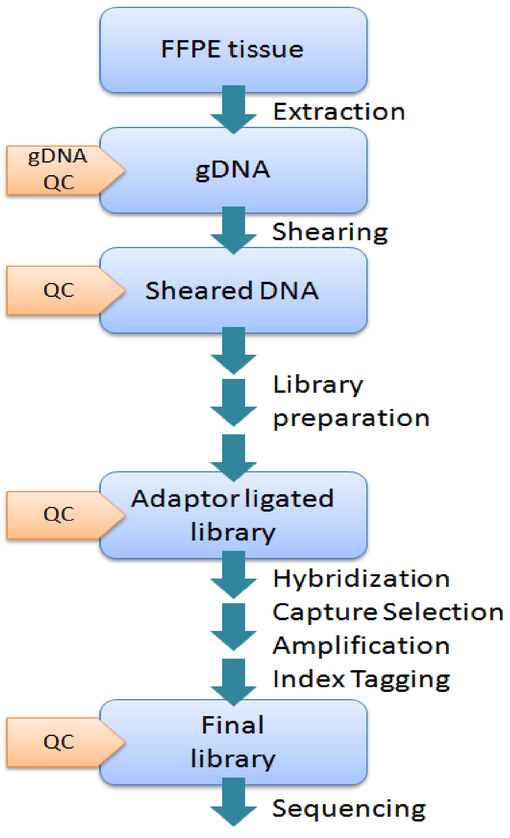

Figure 1. Overview of the overall SureSelect XT sequencing sample preparation workflow indicating the recommended sample QC steps.

Results

Influence of sample integrity on sequencing results

As part of the NGS library preparation workflow, input gDNA is fragmented (Figure 1). So it could be anticipated that that the original size distribution of a sample has no significant effect on the quality of the sequencing results. To test this 197 samples of the large sample pool were randomly selected and further analyzed to determine if there was a correlation between the initial gDNA integrity, calculated as DIN, and the sequencing quality criteria. This subset of samples had a range of DIN 1.1 to 6.6.

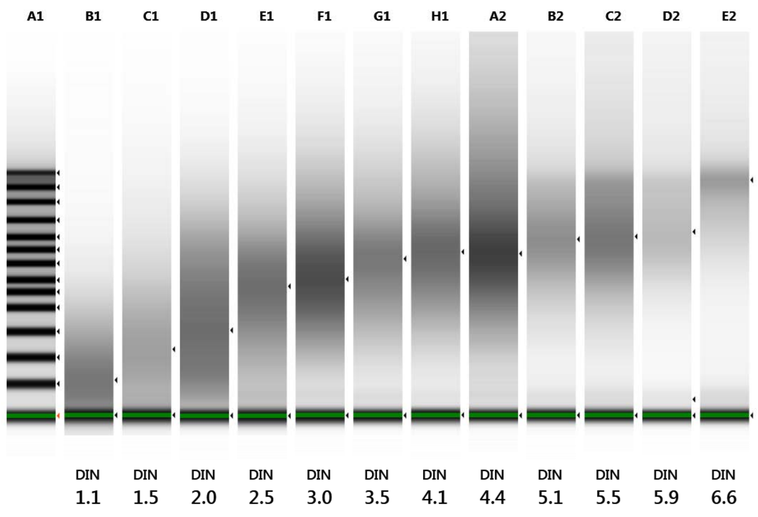

Figure 2: Gel image representative for the sample degradation range which was subjected to the sample integrity and sequencing correlation analysis

An on-target rate above 70% was defined as the sequencing success parameter. 63 samples (31%) of the analyzed sample subset did not meet the on-target rate criteria of at least 70%. Among these the integrity analysis utilizing DIN showed that 61 samples (97%) received a DNA Integrity Number below 3 (Figure 3).

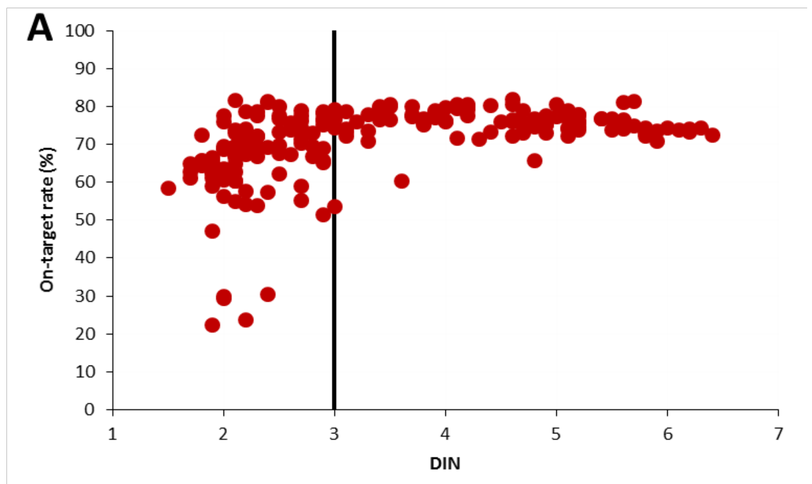

Figure 3: Gel image representative for the sample degradation range which was subjected to the sample integrity and sequencing correlation analysis

Furthermore the coverage at 10x (target above 90%) was analyzed as additional sequencing quality criteria. The majority of samples did pass the 90% target coverage rate at 10x, but greater variation was observed for samples with DIN below 3 (Figure 4).

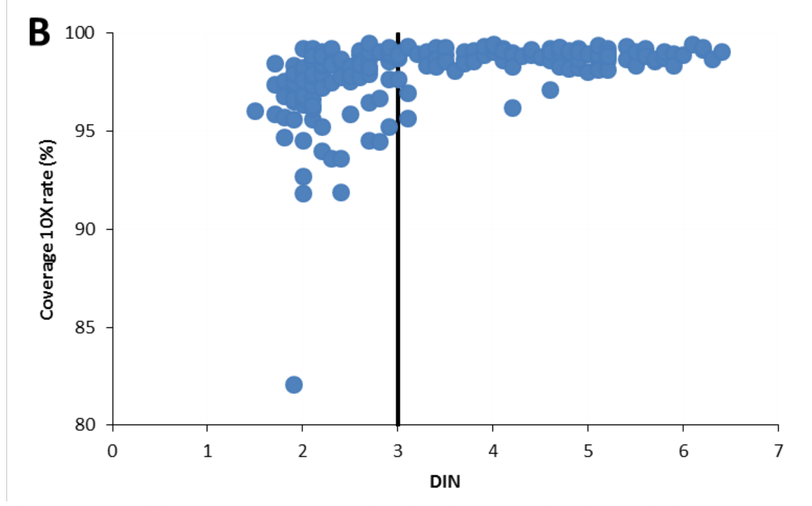

Figure 4: Gel image representative for the sample degradation range which was subjected to the sample integrity and sequencing correlation analysis

Objective analysis of genomic DNA integrity for FFPE samples

Based on subjective interpretation of the electropherograms or gel images it would have been difficult to reject samples, especially when handling results in a high throughput format. The determined DIN threshold allows objective exclusion of samples from further processing and sequencing and ensures correct repeatability of the DNA integrity assessment. In addition to DIN the amount of gDNA extracted from the FFPE tissue is critical for the success of the downstream sequencing. The 2200 TapeStation system determines DIN and the total sample concentration in a single step. In total 751 samples were analyzed and the DIN determined ranged from 1 to 7.8.

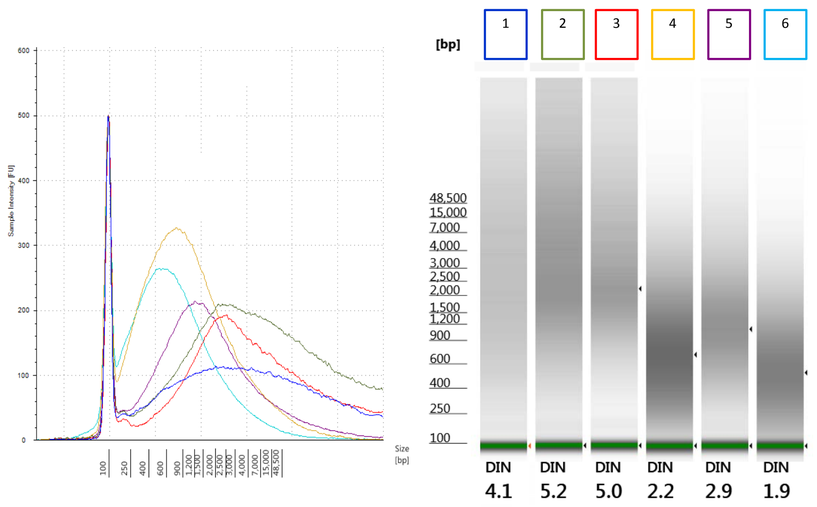

Figure 5: Gel image representative for the sample degradation range which was subjected to the sample integrity and sequencing correlation analysis

Samples	DIN	On-target rate % >70%	10x coverage rate % >90%
1	4.1	79.7	98.6
2	5.2	78.0	99.0
3	5.0	80.7	98.1
4	2.2	23.8	94.0
5	2.9	51.5	97.7
6	1.9	47.4	96.6

Table 1 summarizes detailed results for the gDNA samples shown in Figure 3. Three samples with DIN>3 showed >70% on-target rate and the three samples DIN<3 showed a diminished on-target rate.

Based on the analysis of the initial sample integrity and the correlation to the on-target and coverage 10x rate it was decided to set a DIN threshold of >3 for subsequent sequencing. Among the correlation data set (n=197) 49% did not pass the DIN>3 criteria. Among the total sample set (n=751) samples 65% did not pass the DIN>3 criteria. Samples with a DIN below 3 were not further processed and sequenced. The described NGS sample preparation workflow, including the several QC steps is relative work intensive and the cost per sample is relatively high, therefore it is of great advantage to be able to qualify samples for acceptance and rejection criteria utilizing DIN.

DIN > 3	Samples directly proceeded to sequencing workflow
DIN < 3	Samples only proceeded to sequencing workflow after consultation of the client who send in the sample

Conclusions

- **The automatically determined DIN provided by the 2200 Tape Station system for the Genomic DNA ScreenTape assay provides an optimal tool to screen gDNA samples extracted from FFPE tissue in a high throughput format.**
- **DIN correlates with key sequencing quality metrics and thus enables the definition of an integrity thresholds for the processing of FFPE samples.**
- **DIN can be integrated as a selection criteria on whether or not to proceed individual samples for the downstream workflow.**
- **Thus, utilizing DIN, specific sample integrity requirements can be established for respective NGS applications and easily communicated to customers or collaborators.**