Application Note

Genomics

Agilent

Trusted Answers

Optimization of an Agilent NGS Automated Workflow for the Characterization of NSCLC DNA Samples in a Molecular Diagnostic and Anatomic Pathology Laboratory

Authors

Antonio Marchetti, MD, PhD,
Professor,
Anatomic Pathology Laboratory,
Ospedale SS Annunziata Chieti,
Italy

Fiamma Buttitta, MD, PhD,
Professor,
Anatomic Pathology Laboratory,
Ospedale SS Annunziata Chieti,
Italy

Lara Felicioni, PhD,
Biologist,
Anatomic Pathology Laboratory,
Ospedale SS Annunziata Chieti,
Italy

Alessandro Agostino, PhD,
NGS Clinical WF Specialist,
Agilent Technologies, Inc.,
Italy

Introduction

The Anatomical Pathology laboratory of the SS Annunziata Hospital in Chieti is a reference center for lung cancer. For some time now, this lab has used next-generation sequencing (NGS) methods to characterize formalin-fixed paraffin-embedded (FFPE) and plasma samples with the aim of searching for alterations in genes that are the target of new drugs. A series of 39 previously characterized samples from 37 patients with advanced non-small cell lung cancer (NSCLC) were employed in this study. These samples consisted of 2 plasma samples and 37 FFPE samples, 3 of which belonged to the same patient. The samples were characterized using a complete Agilent workflow, which consisted of the 4200 TapeStation system for sample quality control of genomic DNA (gDNA), cell-free DNA (cfDNA), and NGS libraries, and the MagnisDx NGS Prep system for automated library preparation and target enrichment with the SureSelect Cancer All-In-One assay. The sequencing data were analyzed and variants were annotated by the secondary and tertiary analysis modules, respectively, of the Alissa Informatics platform. The data obtained demonstrates the feasibility and advantage of introducing this Agilent workflow into our routine practice.

Material and Methods

Sample selection, gDNA/cfDNA extraction and qualification

Table 1 includes the list of samples characterized with the TapeStation system using either tissue gDNA or cfDNA ScreenTape devices. gDNA samples were extracted from either biopsy-derived FFPE samples or plasma tissue when biopsies were not available. All samples came from NSCLC patients with advanced metastatic lung tumors. A representative collection of "average" quality samples that display different levels of degradation are shown in Figure 1. Sample concentration and integrity were very diverse (Figures 1 and 2; Table 1). cfDNA samples isolated from plasma demonstrated a very high proportion of cfDNA.

DNA extraction and purification was performed with either the QIAamp DNA FFPE Tissue Kit (for sections taken from FFPE tissues) or the Cobas cfDNA Sample Preparation Kit (CE-IVD) (for cfDNA extraction from 2 ml of plasma samples). gDNA and cfDNA samples were qualified with gDNA and cfDNA ScreenTape assays, respectively, on a 4200 TapeStation system.

DNA fragmentation

Where possible, gDNA samples were normalized to a final concentration of 9 to 10 ng/µL, allowing 60 to 70 ng of gDNA to be fragmented following the recommended protocol for the Agilent SureSelectXT HS Enzymatic Fragmentation kit. For samples with concentration lower than 7 ng/µL, the entire sample volume (7 µL) was loaded. cfDNA samples extracted from plasma were not fragmented and 50 µL of sample was used in the library preparation step.

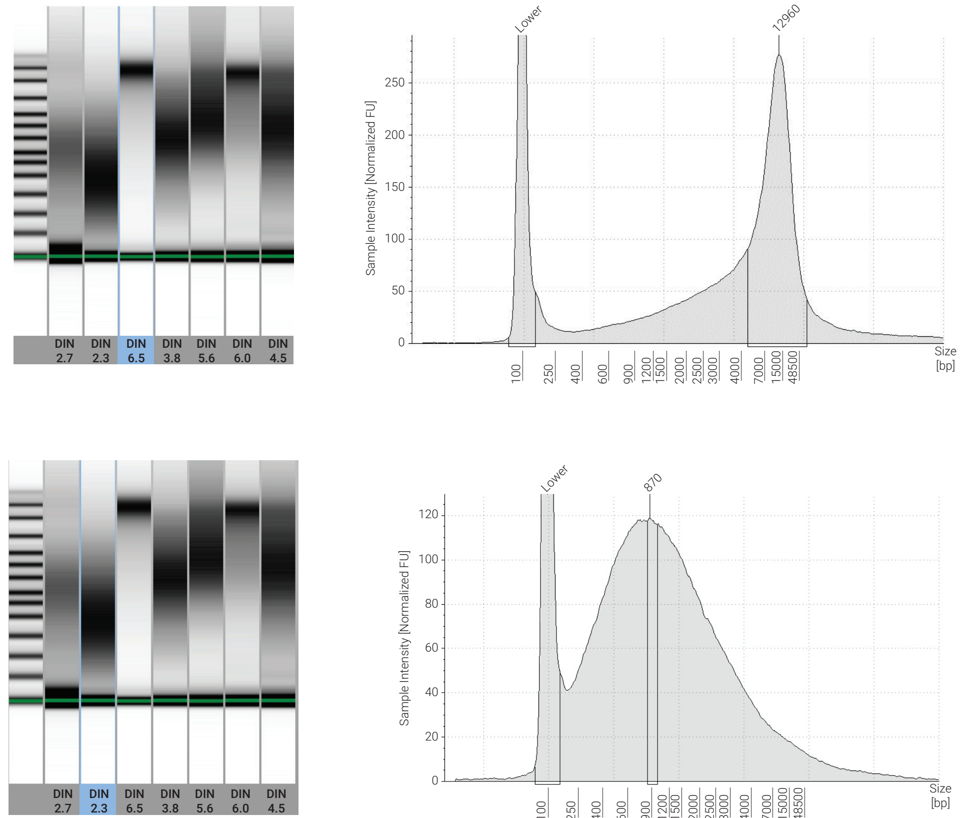


Figure 1. Examples of good and bad quality DNA samples. gDNA was extracted from FFPE of lung tumor biopsies. Samples were characterized using Genomic DNA ScreenTape assay which resulted in a DNA Integrity Number (DIN) of 6.5 (top panel) and 2.3 (bottom panel). Lower DIN number represents bad quality DNA within a scale of 1-10.

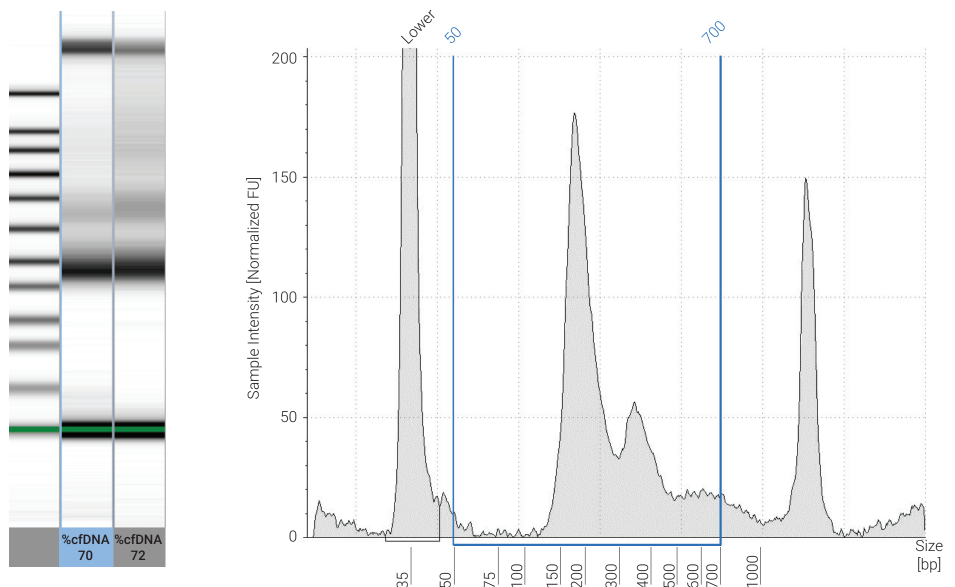


Figure 2. An example of cfDNA extracted from plasma and analyzed with the Agilent Cell-free DNA ScreenTape assay. The highlighted sample contained cfDNA equal to 70% of the total extracted DNA, indicated by the %cfDNA quality metric.

HER2 Immunohistochemistry and FISH

Immunohistochemistry (IHC) on ERBB2 amplified samples was performed with HER2 antibodies on the DAKO Omnis with FFPE sections taken from the same tissue block that was previously used for DNA extraction. For FISH analysis, HER2 IQFISH pharmDx ready-to-use probes were used on the DAKO Omnis starting from a section taken for the IHC from the same tissue block.

Library preparation and qualification

Library samples were prepared on the MagnisDx NGS Prep system following the manufacturer's protocol for the SureSelect Cancer All-In-One Lung HS assay. 12 PCR cycles were used for pre-capture amplification and 14 cycles were used for post-capture amplification. 40 samples were processed in five independent runs.

Sequencing

Libraries obtained from the MagnisDx NGS Prep system were quantified with the High Sensitivity D1000 ScreenTape assay on a 4200 TapeStation system and normalized to 4 nM pools of eight samples each. These samples were sequenced on an Illumina MiSeq system in five different runs using V3 600 flow cells, consistent with the SureSelect HS user manual.

Analysis

FASTQ data were analyzed with the Alissa Align & Call software using the All-In-One module and VCF files were generated. VCF files were imported into the Alissa Interpret software and a specific filtering tree was applied to report the pathogenic variants encountered with an allelic frequency higher than 5%.

Table 1. Sample qualification utilizing the Agilent TapeStation system.

Sample ID	N° exp.	DIN	ng/µL
AR	1	8	200
S01	1	2.5	49.7
S02	1	3.3	35.1
S03	1	3.4	32.5
S04	1	3.1	9,17
S05	1	2.6	73.9
S06	1	2.3	24.5
S07	1	5.4	106
S08	2	5.9	123
S09	2	6.4	363
S10	2	6.0	23.9
S11	2	4.5	155
S12	2	5.0	9.94
S13	2	5.5	56.3
S14 (cfDNA)	2	70%*	0.33
S15 (cfDNA)	2	72%*	0.38
S16	3	2.9	77.3
S17	3	5.2	204
S18	3	5.6	170
S19	3	4.8	147
S20	3	5.6	78.8
S21	3	3.1	66
S22	3	1.7	9.19
S23	3	5.1	10.7
S24	4	6.4	91.5
S25	4	3.4	6.66
S26	4	4.1	118
S27	4	5.4	25.1
S28	4	3.9	5.36
S29	4	3.2	14.3
S30	4	5.7	8.66
S31	4	5.7	10.6
S32	5	5.6	72
S33	5	5.9	167
S34	5	4.9	95.1
S35	5	6	137
S36	5	5.1	19.5
S37	5	2.7	7.39
S38	5	6.5	29.4
S39	5	2.3	29.4

**DIN is not available for cfDNA samples*
% of cfDNA over total gDNA was reported instead

Results

Library preparation and yields

The SureSelect Cancer All-In-One Lung HS assay allows the simultaneous detection of single nucleotide variants (SNVs), copy number variants (CNVs), insertions and deletions, and DNA translocations in 20 genes relevant to the characterization of lung tumors. This panel was used to analyze 40 samples, including one reference DNA sample (Agilent Female Reference, AR) that allowed the calculation of CNVs and served as a validity check for the experiment.

In order to reduce the variability linked to sample diversity, we attempted to standardize the starting mass of material where possible. However, the amount of DNA used for library preparation ranged from 16.15 to 100 ng total and DNA samples exhibited heterogeneous degradation profiles (measured by the DNA integrity number (DIN)).

Five different runs were performed with eight samples at a time. Library quantification highlighted a great deal of heterogeneity in terms of recovery that was not always consistent with the starting DNA input mass (Figure 3). This is likely due to the diversity of DINs observed in the samples. While the integrity of some samples was severely compromised (e.g., DINs of 1.7 and 2.3), in all cases the workflow produced libraries of sufficient quality and quantity to proceed with sequencing.

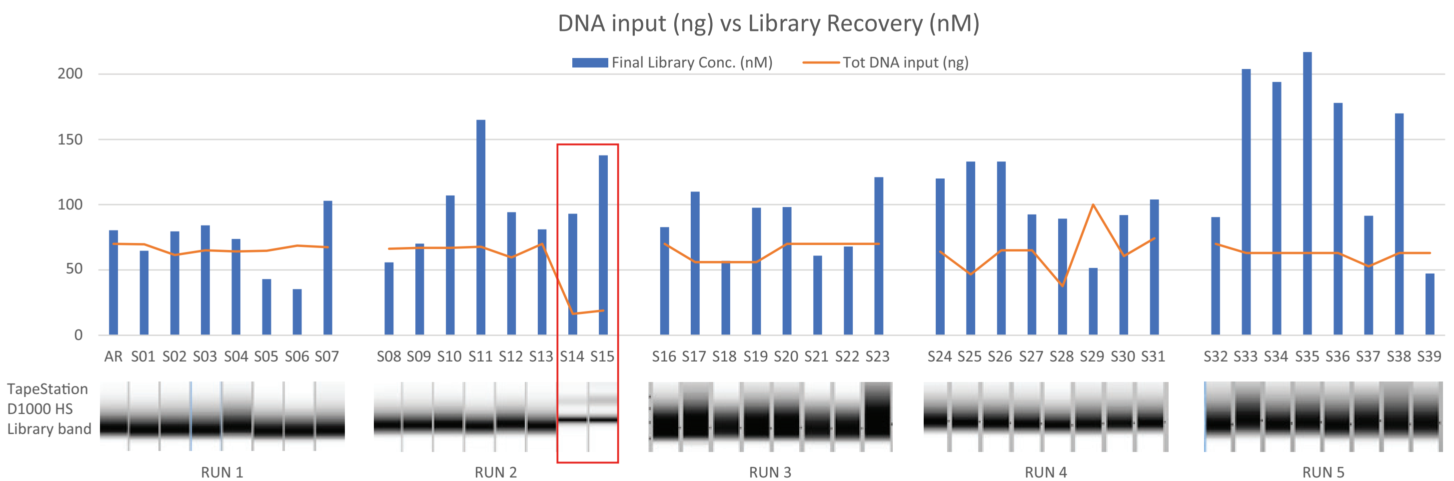


Figure 3. DNA input versus library recovery from five different runs on the MagnisDx NGS Prep system. The red rectangle highlights the two cfDNA samples.

Sequencing metrics

To make the panel usable in a routine analyses and to reduce sequencing costs, eight samples were processed at a time. Average on-target read rate was 65% with a duplicate read rate of 40%. The average total number of usable reads was 2,549,654 which enabled an average on-target coverage of 1124x. With this data a multitude of variants were detected with a sensitivity limit of 5%. As shown in Figure 4, the percentage of duplicates and the number of on-target reads show high correlation with the DIN. The performance analysis in terms of on-target read percentage in the different runs highlights the consistency and uniformity of the results despite the variability linked to the nature of the samples (Figure 5). This is due to the robust sample preparation by the MagnisDx NGS Prep system.

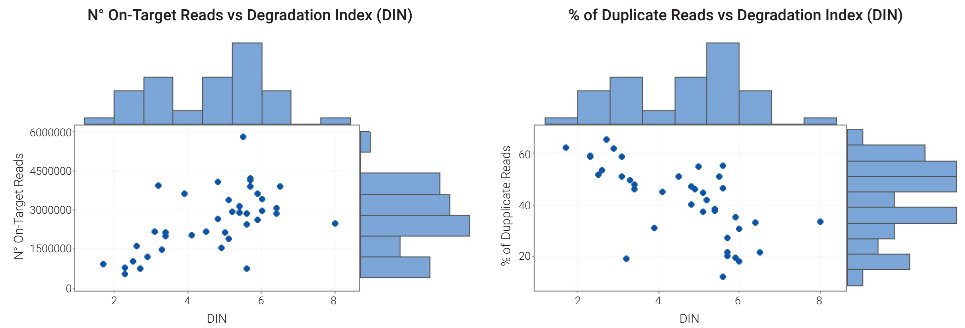


Figure 4. Analysis of on-target and duplicate reads. Graphs depict the number of on-target reads (left) and % of duplicate reads (right), with each demonstrating correlation with the DNA integrity number (DIN).

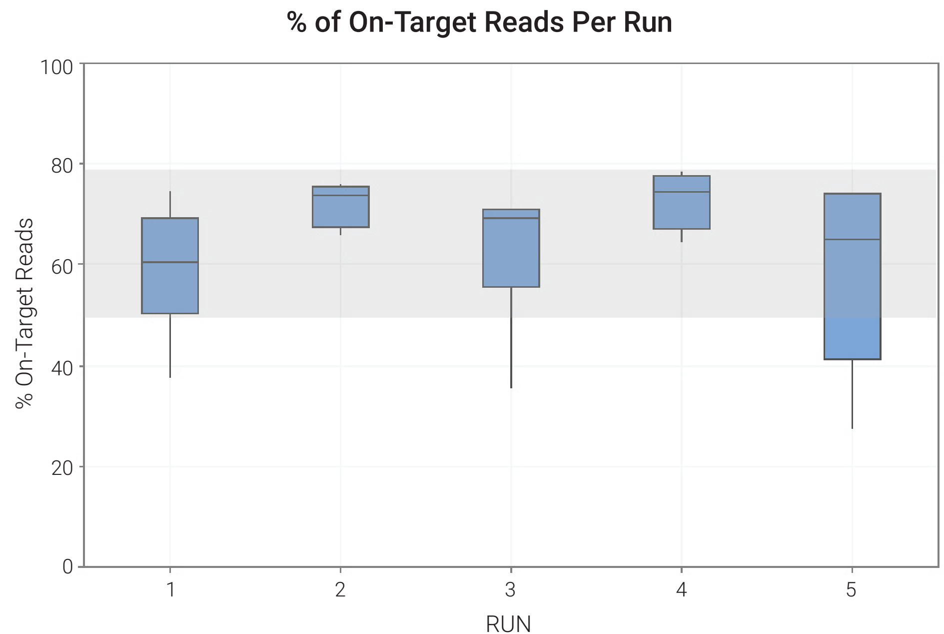


Figure 5. Consistency of results from different Magnis runs. In spite of samples being separated across five different runs, samples exhibited consistency in percentage of on-target reads.

Variant analysis and interpretation

The samples used in this study had been previously characterized in the diagnosis phase with both immunohistochemistry (IHC) and molecular methods. The progression of the disease made it necessary to investigate multiple markers in these cases using NGS. After identifying the variants in the Alissa Align & Call software, variants were filtered by variant allele fraction (VAF) greater than 5% using a decision tree in the Alissa Interpret software (built specifically for these type of samples). 40 pathogenetic variants were distributed across 10 genes in 19 out of the 37 patients analyzed. The distribution of the variants found and the relative allele frequencies are detailed in Table 2.

Table 2.

Gene	VAF range	N°Found	Type of variant	Description
EGFR	24.50%	1	SNV	missense
ALK	8%	1	SNV	Found in cfDNA
BRAF	19.3% to 49.7%	2	SNV	missense
PIK3CA	6.2% to 59.2%	5	SNV	missense
KRAS	8.8% to 43.8%	8	SNV	missense
TP53	12% to 79.3%	11	SNV	4 found as unique mutation 7 found in conjunction with other driver mutations
EGFR	21.4% to 65.3%	4	DEL	3 x 15 nt del ex19 (1 in cfDNA) 1 x 18 nt del ex19
ALK	36/711 reads*	2	FUSION	1 *ALK-EML4* was found in cfDNA
NTRK1	30/3119 reads*	1	FUSION	<10% Tumor fraction
ERBB2	NA	3	CNV (amp)	1 amp 20x - FISH not possible 1 amp 7x confirmed by FISH and IHC 1 amp 6x - FISH not possible
PTEN	NA	2	CNV (amp)	1 amp 10x - to be confirmed

** For fusions only the fraction % of split reads can be counted*

Special cases

Several of the analyzed cases exhibited genomic aberrations that were both clinically relevant and difficult to detect using traditional means. These cases included:

Plasma Samples

It was possible to identify the *ALK-EML4* fusion, which is associated with a resistance mutation in the *ALK* gene, in one of the two plasma samples. In the second plasma sample a 15 nt deletion in the *EGFR* gene was identified.

Case 1 (*HER2* amplification)

In a sample with a tumor fraction of less than 20%, an amplification of *HER2*—subsequently confirmed by fluorescent in situ hybridization (FISH) and IHC—was highlighted (Figure 6). In two other cases, the amplifications observed in *HER2* could not be confirmed due to the lack of material required to perform a FISH experiment.

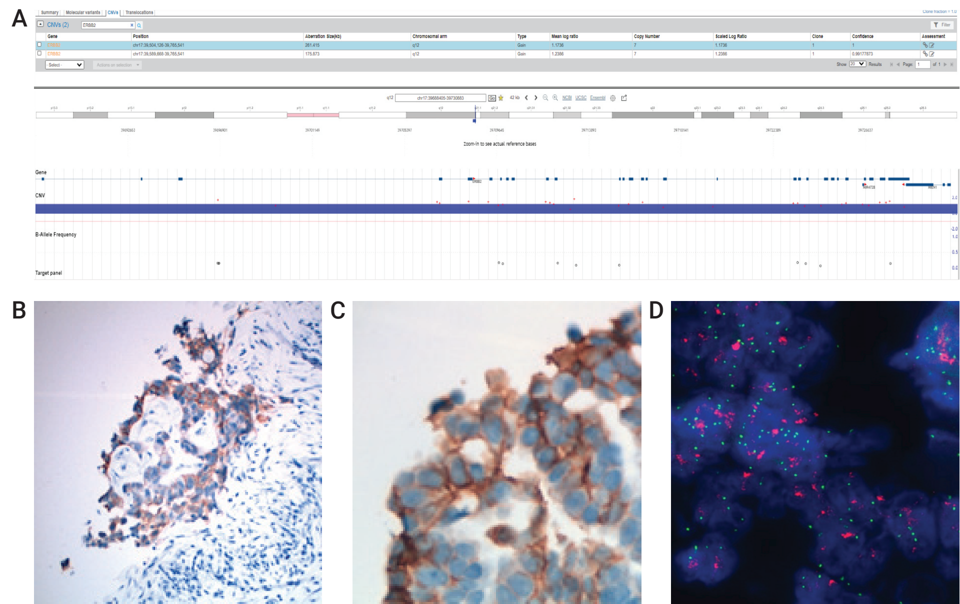
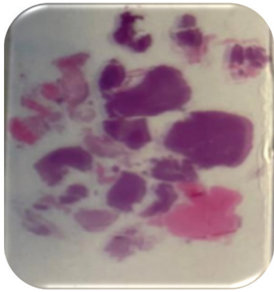


Figure 6. *HER2* gene amplification was confirmed by immunohistochemistry and FISH. Alissa Align & Call identified an *ERBB2* amplification (panel A) that was validated by immunohistochemistry (panels B and C) and FISH (panel D) from the same tissue used for NGS analysis.



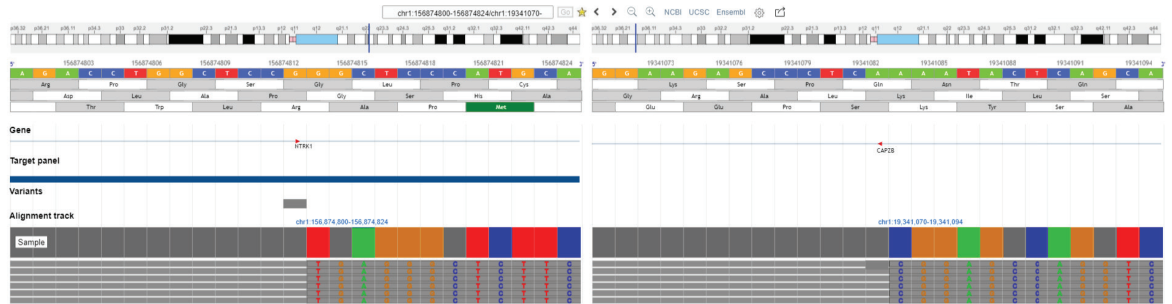


Figure 7. Examination of an *NTRK1* gene fusion. Metastatic undifferentiated sarcoma with pulmonary manifestation fragments of neoplastic tissue were observed with hematoxylin and eosin staining, but exhibited a very low tumor fraction (<10%) (left panel). DNA extracted from this tissue and analyzed with the SureSelect Cancer All-In-One assay displayed an *NTRK1* gene fusion (right panel). Rearrangement of *NTRK1* was confirmed in two independent experiments performed with two separate DNA extractions from the same tissue sample.

Case 2 (*NTRK1* fusion)

In a case that exhibited lung metastasis of sarcoma, a translocation of the *NTRK1* gene was found in a very small tumor fraction (less than 10%) (Figure 7).

Case 3 (*KRAS* + *TP53* in different histotypes)

In one case that was analyzed, the tumor presented with two different histotypes. In one area of the tumor, a pathogenic mutation in the *KRAS* gene was noted while in the other it was observed that two mutations in *TP53* exhibited the same allelic frequency in two different portions of tissue (Figure 8).

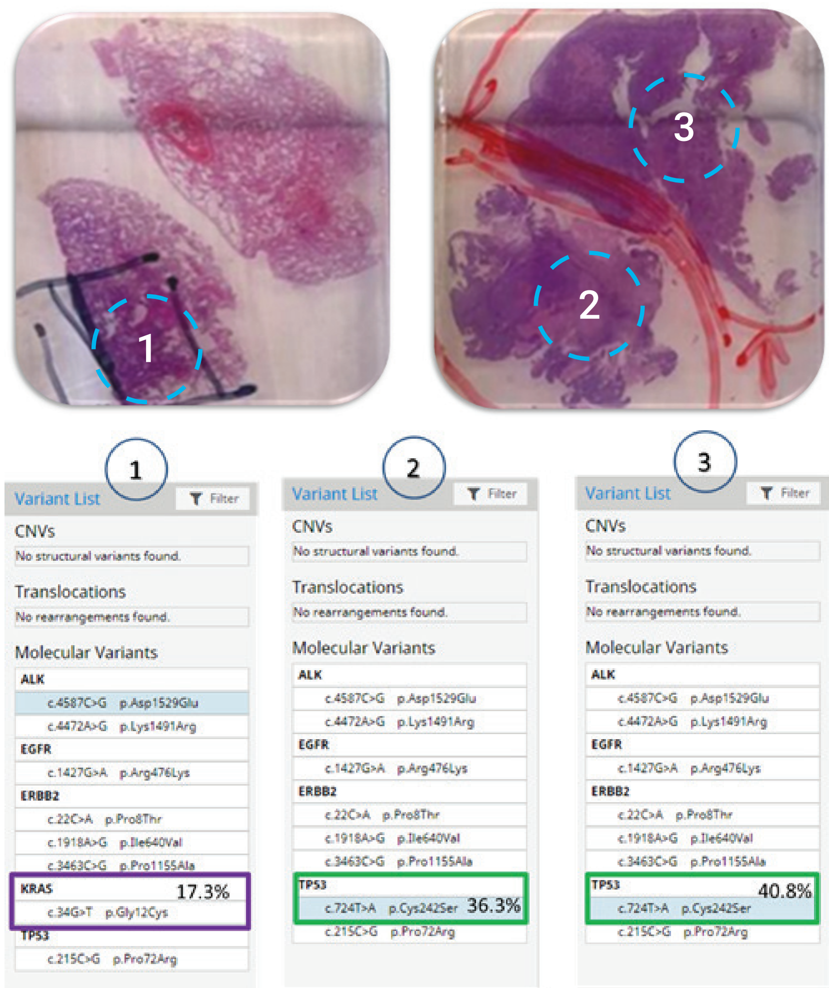


Figure 8. Analysis of different histo-types from *NSCLC* tumor. Two different histo-types of *NSCLC* tumor were analyzed by collecting DNA from three different areas (top panel). DNA samples were analyzed with the SureSelect Cancer All-In-One assay. Two different pathogenic mutations were found: one in the *KRAS* gene with VAF of 17.3% and another in the *TP53* gene. The *TP53* mutation was found in two distinct areas of the tissue (bottom panel) with VAFs of 36.3% and 40.8%.

Conclusions

This study of the cases described allowed optimization of the workflow for the detection of somatic variants in lung tumor samples. The SureSelect Cancer All-In-One assay was able to reliably detect variants that are not easily detectable with other systems in a single DNA-based assay. This was particularly true in the detection of gene fusions (such as the *ALK-EML4* and *NTRK1* fusions described).

The ease-of-use and reliability of the MagnisDx NGS Prep system and the 4200 TapeStation system maximized laboratory productivity and increased reproducibility of sample qualification and library preparations. Most importantly, this workflow allowed a complete characterization of these samples in only three working days, enabling full integration of hybrid capture-based targeted NGS analyses into clinical practice. Compared to traditional single-marker methods, this approach saves not only time but also money, producing a large amount of information from a single experiment even when starting from very small samples.

Table 3. The following list of products were utilized in this publication:

Product name	Part number
Sample QC	
Genomic DNA ScreenTape assay	5067-5365
Genomic DNA Reagents	5067-5366
Cell-free DNA ScreenTape assay	5067-5630
Cell-free DNA Reagents	5067-5631
4200 TapeStation system	G2991BA
D1000 ScreenTape	5067-5582
D1000 Reagents	5067-5583
Library Preparation and Target Enrichment	
Magnis Dx NGS Prep system	K1007AA
SureSelect XT HS Enzymatic Fragmentation kit	5191-4080
Magnis SureSelect XT HS, 1 - 500 kb, ILMN, 96	G9731B (design ID A3097591 All in One Lung Assay)
Data Analysis and Reporting	
Aliss A&C Tier 1	G5357AA-103
Alissa Interpret Tier 1	K5852AA-103
IHC/FISH Analysis	
HER2 Antibodies	PN Omnis GE001
HER2 IQFISH PharmDx Ready-to-use probes	PN GM33311-2
DAKO Omnis	

www.agilent.com

For In Vitro Diagnostic Use.
PR7000-2758

This information is subject to change without notice.

© Agilent Technologies, Inc. 2021
Printed in the USA, July 15, 2021
5994-3181EN

Agilent
Trusted Answers