Article

Next-Generation Sequencing Workflow for NSCLC Critical Samples Using a Targeted Sequencing Approach by Ion Torrent PGM™ Platform

Irene Vanni [1],*,[†], Simona Coco [1],[†], Anna Truini [1,2], Marta Rusmini [3], Maria Giovanna Dal Bello [1], Angela Alama [1], Barbara Banelli [4], Marco Mora [5], Erika Rijavec [1], Giulia Barletta [1], Carlo Genova [1], Federica Biello [1], Claudia Maggioni [1] and Francesco Grossi [1]

[1] Lung Cancer Unit, IRCCS AOU San Martino-IST National Cancer Research Institute, L. go R. Benzi 10, 16132 Genoa, Italy; simona.coco@hsanmartino.it (S.C.); anna.truini@hsanmartino.it (A.T.); mariagiovanna.dalbello@hsanmartino.it (M.G.D.B.); angela.alama@hsanmartino.it (A.A.); ery80x@yahoo.it (E.R.); giulia.barletta@yahoo.it (G.B.); carlo.genova1985@gmail.com (C.G.); febiello@gmail.com (F.B.); claudia_87m@yahoo.it (C.M.); francesco.grossi@hsanmartino.it (F.G.)
[2] Department of Internal Medicine and Medical Specialties (DIMI), University of Genoa, IRCCS AOU San Martino-IST National Cancer Research Institute, L. go R. Benzi 10, 16132 Genoa, Italy
[3] Laboratory of Molecular Genetics, IRCCS, Giannina Gaslini Institute, L. go G. Gaslini 5, 16148 Genoa, Italy; marta.rusmini@gmail.com
[4] Laboratory of Tumor Epigenetics, IRCCS AOU San Martino-IST National Cancer Research Institute, L. go R. Benzi 10, 16132 Genoa, Italy; barbara.banelli@hsanmartino.it
[5] Department of Pathology, IRCCS AOU San Martino-IST National Cancer Research Institute, L. go R. Benzi 10, 16132 Genoa, Italy; marco.mora@hsanmartino.it
* Correspondence: irene.vanni@hsanmartino.it; Tel.: +39-010-555-8463; Fax: +39-010-555-8335
† These authors contributed equally to this work.

Abstract: Next-generation sequencing (NGS) is a cost-effective technology capable of screening several genes simultaneously; however, its application in a clinical context requires an established workflow to acquire reliable sequencing results. Here, we report an optimized NGS workflow analyzing 22 lung cancer-related genes to sequence critical samples such as DNA from formalin-fixed paraffin-embedded (FFPE) blocks and circulating free DNA (cfDNA). Snap frozen and matched FFPE gDNA from 12 non-small cell lung cancer (NSCLC) patients, whose gDNA fragmentation status was previously evaluated using a multiplex PCR-based quality control, were successfully sequenced with Ion Torrent PGM™. The robust bioinformatic pipeline allowed us to correctly call both Single Nucleotide Variants (SNVs) and indels with a detection limit of 5%, achieving 100% specificity and 96% sensitivity. This workflow was also validated in 13 FFPE NSCLC biopsies. Furthermore, a specific protocol for low input gDNA capable of producing good sequencing data with high coverage, high uniformity, and a low error rate was also optimized. In conclusion, we demonstrate the feasibility of obtaining gDNA from FFPE samples suitable for NGS by performing appropriate quality controls. The optimized workflow, capable of screening low input gDNA, highlights NGS as a potential tool in the detection, disease monitoring, and treatment of NSCLC.

Keywords: next-generation sequencing; NGS workflow; NSCLC; Ion Torrent PGM; FFPE; cfDNA

1. Introduction

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related death worldwide, and the majority of patients are diagnosed at advanced stages of NSCLC, with a dramatically poor five-year overall survival rate of 15% [1].

Recently, the identification of the molecular pathology of individual patient tumors has emerged as a promising opportunity for treatment of patients affected by advanced disease. Indeed, specific genetic alterations that drive cancer development can be targeted by selected inhibitors. In particular, the signaling pathway downstream of the Epidermal Growth factor Receptor (EGFR) has been investigated in advanced NSCLC patients harboring activating EGFR mutations (approximately 15% of Caucasian and 50% of Asian patients) treated with first-line tyrosine kinase inhibitors (TKIs), such as gefitinib, erlotinib, and afatinib. These patients report significant improvement in response and progression-free survival compared with patients treated with traditional chemotherapy [2].

Additionally, the application of molecular-based targeted approaches directed to other driver mutations, such as *KRAS*, *BRAF*, and the PI3K/AKT/mTOR pathway, is currently under evaluation, and several trials have been designed to assess the pharmacological efficacies of these compounds, reporting encouraging results [3].

Currently, traditional tools such as Sanger Sequencing (SS), real-time PCR, and pyosequencing (PS) can screen mutations on few genes per run with a sensitivity of approximately 20%, 10%, and 1%–2%, respectively [4–7]. However, the increasing number of novel potential targets results in more time-consuming and expensive diagnostic tests [8]. The advent of NGS has revolutionized the approach to detect gene mutations but requires high-quality DNA. The majority of advanced NSCLC tumor samples are usually represented by FFPE biopsies that result in limited quantity and low-quality DNA, affecting NGS feasibility. Recently, the evaluation of mutations in circulating free DNA (cfDNA) in the plasma of patients affected by advanced NSCLC has emerged as a hypothetical alternative approach to detect gene mutations from a liquid biopsy [9]. However, the feasibility of performing NGS on critical samples remains challenging, and increased efforts should be made in this direction to offer the best individual therapeutic option.

Here we present an optimized workflow to perform deep sequencing using the Ion AmpliSeq Colon and Lung Cancer Panel *v.*1 (Life Technologies, Carlsbad, CA, USA), which analyzes hotspot and targeted regions of 22 genes frequently mutated in colon and lung cancer. Analyses were performed on DNA resulting from critical samples, such as FFPE biopsy blocks and cfDNA, including a quality control step to assess the DNA quality from these samples.

2. Results

2.1. Assessment of NGS (Next-Generation Sequencing) Detection Limit, Specificity, Sensitivity, and Positive Predictive Value

gDNAs obtained from FFPE sections of engineered cell lines with clinical relevant mutations in tyrosine kinase domain of *EGFR*, such as p.T790M and the p.Glu746_Ala750del, were used to assess the limit of detection (LOD) of our NGS workflow. Firstly, dilution series of p.Thr790Met *EGFR* mutated gDNA with wild-type EGFR gDNA were used to obtain mutant allele frequencies of 50%, 10%, 5%, and 1%, respectively. All dilutions underwent to library preparation and sequenced by Ion PGM, achieving an average coverage per amplicon >1200× and uniformity >96% across the samples. Notably, the p.Thr790Met *EGFR* mutation was detected up to the third dilution (5%) with allele frequencies of 49.6%, 10.7%, and 5.6%, respectively, whereas the call of p.Thr790Met mutation was missed in the lowest dilution (1%). Similarly, we also assessed the LOD of our assay to correctly call the indel, applying a serial dilution of the p.Glu746_Ala750del gDNA with EGFR wild type, confirming the previous LOD of 5% also for indel detection. To evaluate the specificity, sensitivity, and Positive Predictive Value (PPV), we also sequenced two NSCLC cell lines (NCI-H1650 and NCI-H1975) and the two reference standards with EGFR mutation at 50% of frequency. NGS

data analysis revealed a specificity of 100% for the detection of SNVs and INDELs compared to the genotypes of NCI-H1650 and NCI-H1975 cell lines (genotype data assessed by Coco *et al.*, using NGS platform) and the two FFPE reference standards (genotype data assessed by Horizon using Droplet Digital PCR), whereas the sensitivity for detecting variants at the LOD of 5% was 96% (Confidence Interval (CI) 95% = 80.5, 99.3) with PPV of 100% (95% CI = 86.7, 100.0).

2.2. FFPE (Formalin-Fixed Paraffin-Embedded) Quality Control

gDNA from FFPE tissues reported very high purity (both 260/280 and 260/230 ratios > 1.7) and underwent multiplex PCR-based quality control assay, with degradation status ranging from 0.3 to 0.8 Average Yield Ratio (AYR). To identify an AYR cut-off suitable for library preparation, we chose three FFPE DNA samples with different AYR values (0.3, 0.6, and 0.8) and prepared the libraries starting from 10 and 20 ng of gDNA. As expected, all samples were successfully amplified when starting from 20 ng, whereas only two out of three were amplified when starting from 10 ng; indeed, the sample that did not show any amplification products was the most fragmented (0.3 AYR, corresponding to 70% degradation). Additionally, to set the gDNA degradation status cutoff, as a prerequisite for proceeding to a downstream workflow, we increased the fragmentation status of gDNA with the lowest AYR (0.3) by heat. In more detail, by adjusting the fragmentation time we obtained a gDNA test with AYR of 0.2 that was amplified starting from 20 ng as input DNA for the library construction. However, this sample did not display any amplification products, suggesting that 0.3 AYR was a reliable decision criterion to predict the NGS suitability of FFPE samples.

Then, to evaluate the degradation effect on sequencing output, all positive libraries were sequenced. Notably, the sample with 0.6 AYR, although successfully amplified starting from 10 and 20 ng of gDNA, displayed much higher coverage (2/90 amplicons < 500×) using 20 ng compared with 10 ng (6/90 amplicons < 500×). Therefore, we decided to use a DNA input of 20 ng for the library construction of all FFPE DNA samples.

2.3. FFPE and Snap Frozen (SF) NGS

Twelve FFPE (AYR range 0.3–0.8) and matched SF gDNAs were sequenced. The coverage and uniformity of each sample run are reported in Supplementary Table S1. The total number of reads was 6,993,993 and 6,538,054 for SF and FFPE samples, respectively. NGS analysis disclosed an average number of reads per amplicon of 6475.9 in SF samples compared with 6050.5 in FFPE samples. Despite this high coverage, seven of 12 FFPE samples revealed at least one amplicon (range: 1–5) with coverage less than 500×, whereas only one SF sample showed one amplicon with unsatisfactory coverage. More specifically, the FFPE sample ID12 with an AYR = 0.3, showed 19 amplicons with a coverage <500×, suggesting that low quality of gDNA could affect sequencing results. Notably, the aforementioned seven FFPE DNAs reported recurrent amplicons not fully covered: AKT1_1 (chr14: 105246446-105246583), FGFR3_2 (chr4: 1806082-1806187), NOTCH1_1 (content; chr9: 139399338-139399447), STK11_3 (chr19: 1220481-1220603), and TP53_2 (chr17: 7579351-7579485). Interestingly, these amplicons displayed high GC content (greater than 61%). Notably, the coverage uniformity also correlated with gDNA fragmentation status; more specifically, all SF samples displayed 100% coverage uniformity, except for one with 98.9%, whereas FFPE samples ranged from 73.3% to 100%, referring to 0.4 and 0.8 AYR samples, respectively.

The Variant Caller (VC) plugin reported a total of 102 genetic variants (100 SNVs and two indels) (Supplementary Table S2). Notably, all the *EGFR* and *KRAS* mutations previously detected by SS were confirmed in all NSCLC samples reporting identical results starting from SF and matched FFPE tissue, although with different allele frequencies due to intra-tumor heterogeneity (Table 1), whereas the other seven genes present in the colon-lung panel showed discordant results between FFPE and matched SF samples. In particular, 11 discordant nucleotide substitutions were identified: six genetic variants at a frequency less than 5% (1 in SF; 5 in FFPE), and five ranging from 5% to 20% (three in SF; two in FFPE) (Supplementary Table S3).

Table 1. *EGFR* and *KRAS* mutational status comparison between Sanger Sequencing and NGS (Next-Generation Sequencing) Variant Caller software. This table reports the *EGFR* or *KRAS* mutational status, previously defined by Sanger Sequencing from the 12 patients compared to the Variant Caller NGS results. The *EGFR* or *KRAS* mutation nomenclature is based on the convention recommended by the Human Genome Variation Society [10,11].

Patient	Tumor Type	*EGFR* and *KRAS* Status by Sanger Sequencing	*EGFR* and *KRAS* Status Confirmed by VC Software	Allele Name
1	SF	NM_033360.3 (KRAS): c.34G>C (p.Gly12Arg)	Confirmed *KRAS* mutation AF: 29.8%	COSM518
1	FFPE	NM_033360.3 (KRAS): c.34G>C (p.Gly12Arg)	Confirmed *KRAS* mutation AF: 58.6%	COSM518
2	SF	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 24.1%	COSM516
2	FFPE	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 6.4%	COSM516
3	SF	NM_005228.3 (EGFR): c.2239_2240TT>CC (p.Lys747Pro)	Confirmed *EGFR* mutation AF: 37.8%	COSM24267
3	FFPE	NM_005228.3 (EGFR): c.2239_2240TT>CC (p.Lys747Pro)	Confirmed *EGFR* mutation AF: 45.6%	COSM24267
4	SF	NM_033360.3 (KRAS): c.35G>T (p.Gly12Val)	Confirmed *KRAS* mutation AF: 24.0%	COSM520
4	FFPE	NM_033360.3 (KRAS): c.35G>T (p.Gly12Val)	Confirmed *KRAS* mutation AF: 23.2%	COSM520
5	SF	NM_005228.3 (EGFR): c.2237_2255delinsT (p.Glu746_Ser752delinsVal)	Confirmed *EGFR* mutation AF: 5.6%	COSM12384
5	FFPE	NM_005228.3 (EGFR): c.2237_2255delinsT (p.Glu746_Ser752delinsVal)	Confirmed *EGFR* mutation AF: 10.5%	COSM12384
6	SF	NM_005228.3 (EGFR): c.2236_2250del (p.Glu746_Ala750del)	Confirmed *EGFR* mutation AF: 7.8%	COSM6225

Table 1. *Cont.*

Patient	Tumor Type	*EGFR* and *KRAS* Status by Sanger Sequencing	*EGFR* and *KRAS* Status Confirmed by VC Software	Allele Name
6	FFPE	NM_005228.3 (EGFR): c.2236_2250del (p.Glu746_Ala750del)	Confirmed *EGFR* mutation AF: 45.8%	COSM6225
7	SF	NM_033360.3 (KRAS): c.35G>A (p.Gly12Asp)	Confirmed *KRAS* mutation AF: 38.6%	COSM521
7	FFPE	NM_033360.3 (KRAS): c.35G>A (p.Gly12Asp)	Confirmed *KRAS* mutation AF: 26.3%	COSM521
8	SF	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 50.5%	COSM516
8	FFPE	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 24.5%	COSM516
9	SF	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 42.0%	COSM516
9	FFPE	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 41.7%	COSM516
10	SF	NM_005228.3 (EGFR): c.2573T>G (p.Leu858Arg)	Confirmed *EGFR* mutation AF: 45.3%	COSM6224
10	FFPE	NM_005228.3 (EGFR): c.2573T>G (p.Leu858Arg)	Confirmed *EGFR* mutation AF: 40.2%	COSM6224
11	SF	NM_033360.3 (KRAS): c.37G>T (p.Gly13Cys)	Confirmed *KRAS* mutation AF: 32.0%	COSM527
11	FFPE	NM_033360.3 (KRAS): c.37G>T (p.Gly13Cys)	Confirmed *KRAS* mutation AF: 24.2%	COSM527
12	SF	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 31.0%	COSM516
12	FFPE	NM_033360.3 (KRAS): c.34G>T (p.Gly12Cys)	Confirmed *KRAS* mutation AF: 42.6%	COSM516

Abbreviations: SF: Snap Frozen; FFPE: Formalin-Fixed Paraffin-Embedded; VC: Variant Caller; AF: Allele Frequency.

To confirm the Single Nucleotide Variants (SNVs) (excluding indels) identified by the VC software, we performed a second round of analysis using the GATK software, which identified 103 nucleotide substitutions. Notably, GATK validated 93 SNVs identified by VC: 89 FFPE-SF matching SNVs and four of five VC-SNVs (5%–20% frequency) FFPE-SF non-matching, reported as pathogenic mutations (COSM516; COSM1579024; COSM710; COSM11496). Conversely, this pipeline did not confirm the six VC-SNVs (<5% frequency) and one VC-SNV (5%–20% frequency). Moreover, the GATK software identified an additional 10 non-matching novel SNVs (six with a frequency <20% and four with a frequency higher than 20%; range: 20%–35%). However, since none of those four (>20%) was confirmed by SS, we consider them as false positives (data not shown).

We also applied the PS to further validate the five genetic variants (5%–20% frequency) discordant between the FF and matching FFPE samples, confirming all VC-SNVs as true positives (Supplementary Table S3). Overall these data suggest the superiority of the VC pipeline to the call genetic variants generated by Ion PGM. Therefore, we considered 94 SNVs and two indels called by VC to define the tumor samples' mutational profile, excluding six genetic variants at a frequency lower than 5%, according to our assay LOD.

Data reported an average of 7.4 genetic variants (range 5–12) per patient, including Single Nucleotide Polymorphisms (SNPs) and pathogenic mutations. More specifically, only two SNPs involving the *DDR2* intron region (rs200983597 and rs143729297) reported a Minor Allele Frequency (MAF) smaller than 0.01 defined as "rare" variants [12]. Among the pathogenic variants (Figure 1A), tumor protein p53 (*TP53*) was the most altered gene: five out of 12 patients harbored pathogenic mutations (COSM44571; COSM43814; COSM45413; COSM40942; COSM10769) and all reported a SNP (rs1042522). Additionally, five patients harbored *MET* proto-oncogene receptor tyrosine kinase (*MET*) gene mutations (COSM706; COSM710; COSM1579024). Notably, COSM710 and COSM1579024 coexisted with *KRAS* mutations (COSM518; COSM520; COSM516). Further somatic mutations involved the serine/threonine kinase 11 (COSM49004), catenin (cadherin-associated protein) beta 1 (COSM5670), and SMAD family member 4 (COSM14151).
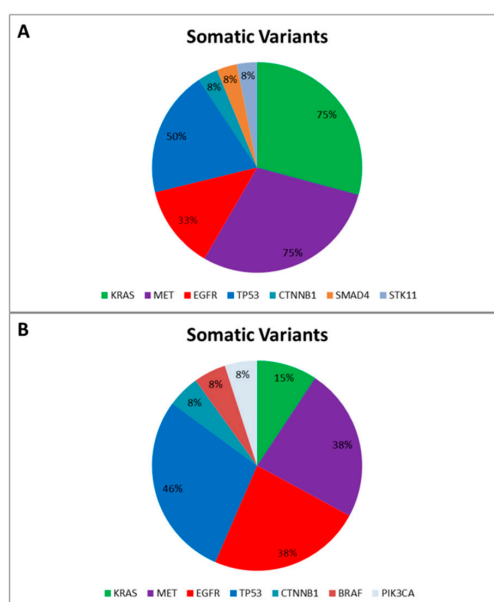

Figure 1. Somatic Variants identified in the "Training Set" (**A**) and in the "Validation Set" (**B**), respectively. The pie charts report the frequency of mutated genes in the "Training Set" (**A**) and in the "Validation Set" (**B**) based on the Catalogue of Somatic Mutation in Cancer (COSMIC) [13].

Finally, to correctly call genetic variants also including indel, we sequenced a further cohort of 13 NSCLC patients, previously tested for *EGFR* mutational status by SS in order to validate our optimized

workflow and test the ability of the bioinformatic pipeline. All gDNA underwent to a multiplex-PCR quality control step, reporting AYR values greater than 0.3, predictive of NGS suitability. Then all FFPE gDNAs were subjected to the Ion PGM workflow and successfully sequenced. In particular, the data analysis properly called four *EGFR* variants detected by SS. Similarly to the training set, the data analysis identified a mean of 7.0 genetic variants (ranging from five to 10) per patient, and, besides the *EGFR* activating mutations, we identified further pathogenic variants involving *BRAF*, *CTNNB1*, *KRAS*, *MET*, *PI3KCA*, and *TP53* genes (Figure 1B).

2.4. cfDNA NGS

Plasma samples, collected at diagnosis, were available for nine out of 12 patients for cfDNA extraction. Quantification of cfDNA was evaluated by qPCR using the *hTERT* assay, showing total concentrations ranging from 0.08 to 10 ng. Because the gDNA input recommended for library construction is 10 ng, we used a Whole Genome Amplification (WGA) technique to increase the starting cfDNA amount. First, we evaluated the WGA technique performance by a scalar dilution of NCI-H1650 human cell line gDNA (10, 1, and 0.1 ng), reporting final gDNA yields of approximately 20 μg regardless of the starting gDNA input. Then, 10 ng of each WGA NCI-H1650 was sequenced to evaluate the specificity and sensitivity of the WGA technique. The sequencing outputs reported a coverage of WGA gDNA greater than 500× and uniformity of 95.2%, except for the WGA 0.1 ng input gDNA sample that displayed four unsatisfactory covered amplicons and a uniformity of 81.1%. The NGS data analysis showed an optimum variant call concordance and identical results in terms of frequencies of variant calls starting from 10, 1, and 0.1 ng of NCI-H1650 gDNA (data not shown). However, the comparison between WGA and the original unamplified NCI-H1650 gDNA displayed five novel insertions in common among all WGA samples. Because these insertions occurred in homopolymeric regions (five or more nucleotide repeats), we hypothesized them to be amplification artefacts due to the WGA technique. Additionally, four out of five novel insertions displayed an AF lower than our LOD (5%); therefore, these insertions were excluded from the performance analysis of NGS workflow on WGA samples resulting in specificity, sensitivity, and PPV of 87.5% (CI 95% = 52.9, 97.7), 100.0% (CI 95% = 67.5, 100), and 100.0% (CI 95% = 67.5, 100), respectively. Therefore, due to the introduction of amplification artefacts in WGA cell lines, we decided to limit the use of the WGA approach to the three samples characterized by low input cfDNA (<1 ng). The remaining six samples were then enriched starting from a lower amount of recommended input cfDNA (2–10 ng) by increasing the cycle number of library PCRs. This modified protocol allowed us to obtain suitable no-WGA libraries; the coverage and uniformity of each run are displayed in Supplementary Table S1.

Although the average reads per amplicon for WGA cfDNA was higher compared with no-WGA cfDNA samples (6719.7 *vs.* 4858.5), the coverage uniformity was better in the no-WGA cfDNA; more specifically, WGA cfDNA showed an average of 83.7% (81.1%–85.6%) *vs.* 99.4% of the no-WGA (98.9%–100%). Indeed, no-WGA cfDNA sample exhibited coverage greater than 500× for all amplicons, except for one sample with 64 amplicons<500×, likely because of a low input of cfDNA (2.1 ng). Conversely, WGA cfDNAs always displayed five amplicons with coverage less than 500×. Notably, as previously observed in the seven FFPE samples, the recurrent uncovered amplicons (AKT1_1, FGFR3_2, NOTCH1_1, STK11_3, and STK11_5) contained high GC-content. Data obtained from VC were compared to the matched tumor mutational profile. As expected, the WGA cfDNA displayed additional insertions; notably, these insertions, likely due to non-specific amplification artefacts of WGA, were the same observed in the NGS analysis of WGA NCI-H1650 and therefore filtered out. Moreover, the comparison analysis identified two additional SNVs with frequencies of 4.4% and 20.6%. However, the second variant call by GATK (Supplementary Table S4) validated all SNVs identified by VC, excluding the nucleotide variant below 5%. Although the remaining discordant variant (20.6%) was confirmed by GATK, it occurred in a WGA-cfDNA sample, suggesting a possible PCR artefact. Unexpectedly, eight out of nine cfDNAs did not disclose any somatic mutations either in the gatekeeper gene *TP53* or in the other genes mutated in their

matched tumors; by contrast, one cfDNA from a patient with metastatic disease presented the identical mutational profile reported in the corresponding tumor.

3. Discussion

Recently, targeted therapy has emerged as a promising therapeutic option for the subgroup of advanced NSCLC patients carrying activating mutations in druggable genes [2]. Currently, the screening of clinically-actionable mutations is performed on FFPE tumor biopsies, but the amount of tumor tissue is often limited, and the gDNA quality may not be always optimal. Moreover, the increasing number of candidate targets requires the validation of highly sensitive throughput technologies, such as NGS, into clinical settings. To date, several studies have investigated targeted deep sequencing on FFPE samples from different cancer types [14–16], including NSCLC [17]; here, we propose an optimized workflow using a specific 22-gene enrichment lung cancer panel to successfully sequence critical samples, such as FFPE biopsies and cfDNA, assuring good quality of NGS data. The workflow includes correct quantification and quality control of input gDNA, a WGA approach to linearly increase the gDNA for low input of starting gDNA (<1 ng), library protocols tailored for specific gDNA type (FFPE and cfDNA), and a robust bioinformatic pipeline to correctly call both SNVs and indels (Figure 2). Similarly to other NGS workflows [18], our assay was able to detect up to 5% of AF either SNV and indel, achieving an optimal assay performance for FFPE samples (100% of specificity and PPV; 96% of sensitivity). The NGS workflow was retrospectively evaluated on DNA extracted from SF and matching FFPE and cfDNA belonging to 12 resected NSCLC patients previously tested for *EGFR* and *KRAS* mutations by SS.
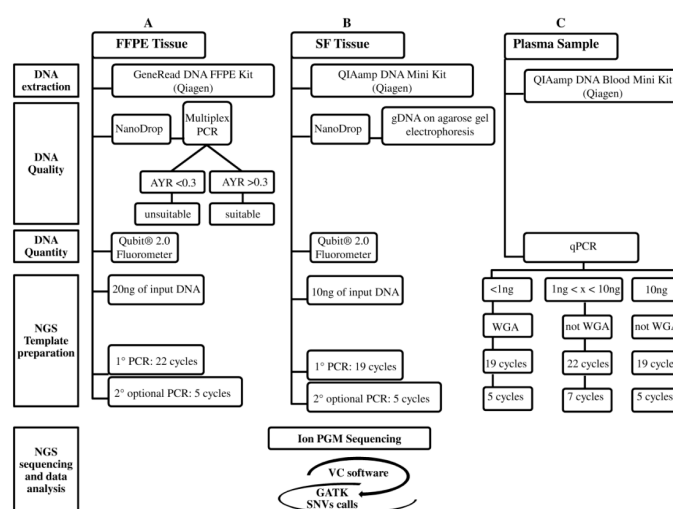

Figure 2. The three analytical NGS (Next-Generation Sequencing) workflows are: (**A**) FFPE (Formalin-Fixed Paraffin-Embedded); (**B**) SF (Snap Frozen); and (**C**) Plasma. These workflows, currently designed for the Ion PGM platform, are divided into five steps: DNA extraction, gDNA Quality, gDNA Quantity, NGS Template preparation, and NGS sequencing and data analysis. The latter, in common among three workflows, includes: Ion PGM sequencing and variant identification using VC software and then GATK for SNVs call. The NGS cost (from DNA extraction to NGS data analysis), excluding personnel, is approximately 250€ for each FFPE, SF, and plasma sample (**A**–**C**). The NGS turnaround time for processing SF (**B**) and no-WGA cfDNA (**C**) samples is approximately five days, whereas for FFPE (**A**) and WGA cfDNA (**C**) samples one more day is required for the multiplex PCR DNA quality control (**A**) and WGA technique (**C**).

Pre-analytic quality control to define suitable DNA for NGS application is crucial, particularly when it is applied to highly degraded DNA, such as FFPE DNA. The majority of studies, aimed to define FFPE deep sequencing guidelines, focused on the correct DNA quantification method [19,20], whereas

our workflow, which involves a multiplex-PCR approach, defined an AYR parameter, indicative of gDNA degradation status, that highly correlates with sequencing quality metrics, such as uniformity and coverage. Notably, we identified an integrity threshold for the processing of FFPE gDNAs estimated at 0.3 AYR, which corresponds to 70% of gDNA fragmentation. Moreover, although the library protocol requires 10 ng of input FFPE gDNA, we demonstrated that when doubling the input FFPE DNA amount (20 ng), highly fragmented DNAs up to 0.3 AYR were successfully sequenced. Notably, this improvement increased the number of amplicons covered to at least 500×, a parameter required for any clinically-oriented cancer panel [18]. However, we identified five recurrent amplicons that were insufficiently covered in highly-degraded FFPE gDNAs because of high GC content (>60%), as previously reported [21].

One critical point of the NGS analysis is generating a list of trustworthy variants; therefore, we tested whether the integration of two different bioinformatic pipelines could improve the SNVs call with higher sensitivity and accuracy compared with single software performance. Specifically, we compared the VC, the Ion PGM plugin with optimized pre-set parameters, with GATK, one of the most used free NGS tools. Each pipeline showed different results in the detection of SNVs (100 SNVs by VC *vs.* 103 by GATK (95 in common between the two pipelines). Differences between the pipelines could be due to the different alignment and calling algorithms [22]. Furthermore, the GATK software exhibited a very high error rate in indel calling (approximately 20 indels per sample; data not shown), leading us to exclude it in the indel calling. A possible explanation is that the GATK pipeline was developed for Illumina data analysis, in contrast to the VC pipeline that has been set up for the analysis of FastQ data generated from Ion-platforms [23]. In this context, some evidence has been previously reported regarding the high rate of false positive variants detected by GATK with non Illumina-FastQ data. These false positives particularly refer to small insertions and deletions (indels) in homopolymeric regions where the difficulty in detecting and determining the genotype of short indels remains a challenge because of the propensity for polymerase slipping during PCR, resulting in sequence artifacts that GATK cannot distinguish [24]. The validation with the PS confirmed all VC discordant genetic variants between SF and matching FFPE samples (>5% frequency), demonstrating that the VC pipeline is optimized for Ion PGM data. Indeed, these discordant variants among SF and matching FFPE samples may be explained by intra-tumor heterogeneity, as previously observed in NSCLC [25–29]. Although the integration of GATK did not exclude any false positive SNVsby VC, it removed all the SF/FFPE discordant SNVs detected by the VC pipeline with frequencies less than 5%, confirming our LOD assay.

EGFR and *KRAS* somatic mutations, previously detected by SS, were confirmed by both pipelines either in SF and matched FFPE, validating the high specificity of the Ion AmpliSeq Colon and Lung Cancer panel for the identification of these mutations. Our data revealed that *TP53* was the most frequent pathogenically-mutated gene [30,31], and all patients harbored the SNP rs1042522 in *TP53*. Interestingly, this SNP, responsible for a substitution of a proline residue with an arginine residue in codon 72 of *TP53*, has been associated with lung cancer susceptibility [32] and a risk of toxicity during platinum-based chemotherapy treatment in advanced NSCLC patients [33]. Notably, we also observed a recurrent association between *MET* genetic variants and *KRAS* activating mutations, as already described in NSCLC by Govindan *et al.* [34]. Due to its high sensitivity, NGS represents an optimal approach to investigate the mutational profile of cfDNA as a non-invasive diagnostic approach to monitor disease progression during treatment. Indeed, a fraction of cfDNA derives from tumor cells and is referred to as ctDNA in cancer patients [35]. To date, several studies have demonstrated the feasibility of using ctDNA as a potential tumor marker in a limited number of patients with various solid cancers; however, few studies focused on NSCLC have been performed [9]. Therefore, we developed an optimized workflow for cfDNA mutational screening, combining a WGA approach for low input samples (<1 ng) and a specific protocol for sequencing library construction, resulting in high numbers of good-quality reads. Currently, several WGA approaches, differing in assay chemistry and complexity, have been developed [36–40]. Here, we used a WGA based on the MDA technology that employs the phi29 DNA polymerase and

random primers to amplify gDNA without thermocycling [40,41]. Several studies revealed that MDA can amplify the genome with higher uniformity, longer length, and less bias compared with PCR-based WGA methods [41–45]. Therefore, MDA-based WGA has been shown to be the best method to obtain high quality cfDNA for different applications, such as the NGS [44,46,47]. Another important concern of the WGA approach is the correct quantification of the amplified DNA. Indeed, WGA techniques report non-specific side products and residual WGA reagents that may influence DNA quantification by fluorometric methods [43,48]. Therefore, we suggested performing WGA-DNA quantification using real-time PCR, which has been demonstrated to be the most reliable method [49]. The sequencing output obtained from the WGA NCI-H1650 displayed a similar performance to the standard no-WGA cell line, irrespective of input gDNA. Similar SNV frequencies were reported across the WGA NCI-H1650 samples, suggesting that this technology may be applied to the colon and lung assay when the starting gDNA amount is limited. However, it should be taken into account that the MDA-based WGA introduces some amplification biases that mainly occur in DNA homopolymeric regions, as previously observed, resulting in a lower specificity of the assay [41,50,51]. Because the WGA application might be necessary to yield sufficient DNA for the subsequent sequencing workflow [52], we tried to limit all critical points. Our optimized workflow and NGS analysis on cfDNAs allowed for detection of the same mutational profile of the corresponding tumor in one cfDNA sample. However, among nine NSCLC patients with available plasma, only two were at stage IV of disease. Because several studies revealed that the presence of ctDNA was mainly detected in aggressive disease [9,53], this observation could explain why we detected ctDNA in only one of two patients with advanced NSCLC. The unidentified somatic mutations in localized (stage I–III) cfDNA samples could also be explained by the detection limit of mutation allele frequencies of our NGS assay, estimated at 5% [54].

However, since the Ion PGM technology utilizes scalable semiconductor chips, a lower number of samples in the same run could increase the amplicon coverage; this might allow for the detection of variants with lower frequency (<5%) in the hypothesis to also apply the NGS workflow to the ctDNA screening. Such higher throughput, might also permit us to detect tumor heterogeneity e.g., in the profiling of ctDNA and to call variants at very low frequencies. Furthermore, since the correct amplicon size is critical to the success of sequencing experiments, designing custom panels with a shorter length of amplicons could improve the sequencing performances, especially for highly fragmented DNA such as FFPE and cfDNA. Finally, it should be kept in mind that customizing panels for a subset of druggable genes of a specific cancer type may offer testing at a lower cost, thereby yielding a higher coverage.

4. Experimental Section

4.1. Sample Collection

Twelve patients who underwent surgery for NSCLC at different stage of the disease were enrolled in our study as the "Training Set", based on their previously known *EGFR* and *KRAS* mutational statuses. The clinical characteristics of the patients are reported in Table 2. SF tumors from all patients were available in the tissue biobank of our Institute. Additionally, matched FFPE tissues were obtained from the pathology archive. All samples were evaluated by a pathologist for malignant cell content to macrodissect and enrich the tumor component up to 50%, when necessary. Nine of the 12 plasma samples from patients were also collected at time of diagnosis.

We also enrolled a further 13 NSCLC patients as the "Validation Set" (Table 3), and for each case the FFPE tissue block was revised by the pathologist to verify the tumor cell content. Moreover, three Horizon FFPE standards (*EGFR* p.Thr790Met, *EGFR* p.Glu746_Ala750del, and *EGFR* wild-type reference standards; Horizon, Cambridge, UK) and two human NSCLC cell lines (NCI-H1650 p.Glu746_Ala750del and NCI-H1975 p.Thr790Met/p.Leu858Arg; American Type Culture Collection, Manassas, VA, USA) were used to assess the parameter performances of our assay. The present study was approved by the local ethics committee (TP-01-2014; 255REG2014) and written informed consent was obtained from every patient.

Table 2. Clinical characteristics of NSCLC (Non-small cell lung cancer) in the "Training Set".

Patient	Block Age	% TC	SF	FFPE	cfDNA	Histology	TNM	Stage	Age at Diagnosis	G	Smoking Habits	PS	OS	Status
1	2012	>50	x	x	x	SCC	T1b, N0, M0	IA	77	M	S	1	30	A
2	2010	>50	x	x	x	ADK	T1a,N2,M0	IIIA	67	M	NS	1	18	DOD
3	2013	>50	x	x	x	ADK	T2, N2, M1a	IV	76	F	NS	1	14	DOD
4	2010	>50	x	x	x	ADK	T2b,N0,MX	IIA	76	M	FS	1	11	DOD
5	2013	>50	x	x	NA	ADK	T2a,N2,M1b	IV	75	F	NS	1	13	A
6	2011	>50	x	x	NA	ADK	T2a, N0, MX	IB	78	F	NS	1	29	DOD
7	2013	>50	x	x	x	ADK	T2a, N2, M1b	IV	65	M	S	1	3	DOD
8	2013	>50	x	x	x	ADK	T2b, N1, M0	IIB	59	F	S	1	23	A
9	2012	>50	x	x	x	ADK	T2a, N0, MX	IB	72	M	FS	0	28	A
10	2011	>50	x	x	NA	ADK	T2b, N0, M0	IIB	66	F	FS	1	42	A
11	2010	>50	x	x	x	ADK	T1a, N0, M0	IA	77	M	FS	0	49	DOD
12	2010	>50	x	x	x	ADK	T1a, N0, M0	IA	60	F	S	1	32	DOD

Table 3. Clinical characteristics of NSCLC patients in the "Validation Set".

Patient	Block Age	% TC	SF	FFPE	cfDNA	Histology	TNM	Stage	Age at Diagnosis	G	Smoking Habits	PS	OS	Status
13	2015	>50	NA	x	NA	ADK	T2, N2, M0	IIIB	59	M	S	0	5	A
14	2014	>50	NA	x	NA	ADK	Tx, Nx, M1a	IV	42	M	NS	0	14	A
15	2015	>50	NA	x	NA	ADK	T3, N3, M1a	IV	64	M	FS	1	4	A
16	2014	>50	NA	x	NA	ADK	T4, N2, M1b	IV	77	F	NS	0	12	A
17	2014	>50	NA	x	NA	ADK	T3, N2, M0	IIIA	70	F	NS	1	8	A
18	2012	>50	NA	x	NA	ADK	T3, N2, M1b	IV	63	M	S	1	37	A
19	2014	>50	NA	x	NA	ADK	T4, N2, M1b	IV	70	M	FS	1	2	DOD
20	2014	>50	NA	x	NA	ADK	T4, N0, M0	IIIA	68	M	FS	0	16	A
21	2011	>50	NA	x	NA	ADK	T1a, N0, M0	IA	71	F	NS	1	42	DOD
22	2014	>50	NA	x	NA	ADK	T2, N2, M1b	IV	81	F	NS	1	10	DOD
23	2013	>50	NA	x	NA	ADK	T1b, N2, M0	IIIA	72	M	FS	1	22	A
24	2014	>50	NA	x	NA	ADK	T3, N3, M1b	IV	61	M	NS	1	11	DOD
25	2015	>50	NA	x	NA	ADK	Tx, Nx, M1b	IV	55	F	FS	0	7	A

Abbreviations: TC: Tumor Content; ADK: Adenocarcinoma; SCC: squamous cell carcinoma; TNM: Tumor-Nodes-Metastasis; G: Gender; F: Female; M: Male; S: Smoker; NS: never smoker; FS: Former smoker; PS: Performance Status; OS: Overall Survival (Month); A: Alive; DOD: Dead of disease; x: present; NA: not available.

4.2. DNA Extraction and Quality Control

gDNA from SF samples and the NCI-H1650 cell line was isolated using a QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany), whereas gDNA from FFPE tissues was extracted using the GeneRead DNA FFPE Kit (Qiagen), according to the manufacturer's recommendations. All gDNA samples were then quantified by a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using a Qubit® dsDNA HS Assay Kit. cfDNA was extracted from 400 µL of plasma with a QIAamp DNA Blood Mini Kit (Qiagen) and quantified by quantitative PCR (qPCR) using the human telomerase reverse transcriptase (*hTERT*) TaqMan® Copy Number Reference Assay (Catalog: 4403316; Life Technologies). The calibration curve was defined based on a dilution series (10–1×10^6 pg) of a control human genomic DNA standard (Promega, Madison, WI, USA). The qPCR reaction was performed in a final reaction volume of 10 µL using 5 µL of TaqMan Universal Mastermix (Life Technologies, Carlsbad, CA, USA), 1 µL of the TaqMan® Copy Number Reference Assay *hTERT* (Life Technologies), and 4 µL of cfDNA on a RealPlex (Eppendorf, Hamburg, Germany) device. Each sample was run in duplicate, and the final concentration calculated by interpolation of the CT value with the aforementioned calibration curve. Both SF and FFPE gDNA underwent a quality control assay. DNA quality was first assessed using a NanoDrop ND-1000 (Thermo Scientific, Wilmington, DE, USA) to measure the whole absorption spectrum (220–750 nm) and calculating absorbance ratios at both 260/280 and 260/230. SF gDNA was analyzed on 0.8% agarose gel electrophoresis to visualize the DNA size distribution, whereas we performed a multiplex PCR-based quality control assay using a slight modification of the van Beers *et al.* [55] protocol to assess FFPE gDNA fragmentation status. Briefly, 30 ng of FFPE gDNA, as measured by Qubit, was analyzed by a multiplex PCR reaction performed using three independent sets of *GAPDH* primers that amplify 200, 300, and 400 bp fragments. The amplification was performed in a reaction volume of 30 µL with final concentrations of 0.133 µM of each primer, 1.5 U AmpliTaq Gold® DNA Polymerase (Life Technologies, Carlsbad, CA, USA), GeneAmp® $10\times$ PCR Gold Buffer, 1.5 mmol/L $MgCl_2$, and 0.2 mmol/L dNTPs. PCR was conducted in a Mastercycler® nexus Thermal Cycler (Eppendorf) as follows: 95 °C for 7 min, 35 cycles of 1 min at 94 °C, 1 min at 56 °C, and 3 min at 72 °C, followed by a hold at 72 °C for 7 min. PCR products were then analyzed with an Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA). The yield of the resulting three amplicons was evaluated by comparison with the yield of amplicons from a reference DNA template (Promega). The Average Yield Ratio (AYR), calculated for each amplicon, was used as a quantitative indicator of FFPE DNA integrity.

4.3. Whole Genome Amplification (WGA)

cfDNA less than 1 ng, estimated by qPCR, underwent WGA using the REPLI-g Single Cell Kit (Qiagen). The REPLI-g Single Cell Kit is based on a Multiple Displacement Amplification (MDA) technology that achieves highly uniform amplification across the entire genome with minimal locus bias and reduced nonspecific amplification artefacts [41]. To evaluate the performance of this assay for subsequent NGS application, we previously amplified a scalar dilution of NCI-H1650 cell line gDNA (10, 1, and 0.1 ng). The amplified gDNAs were diluted 1:100 and evaluated by qPCR using the *hTERT* assay, as previously described.

4.4. Next Generation Sequencing (NGS)

Libraries were amplified using the Ion AmpliSeq Colon and Lung Cancer Panel *v*.1 (Life Technologies), which analyzes 90 amplicons in hotspots and target regions of 22 genes (*AKT1*, *ALK*, *BRAF*, *CTNNB1*, *DDR2*, *EGFR*, *ERBB2*, *ERBB4*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *KRAS*, *MAP2K1*, *MET*, *NOTCH1*, *NRAS*, *PIK3CA*, *PTEN*, *SMAD4*, *STK11*, and *TP53*) with a coverage of over 500 mutations involved in colon and lung cancers. Overall, 20 ng for FFPE, 10 ng for SF gDNA and WGA cfDNA, and a range from 1 to 10 ng for no-WGA cfDNA were amplified using the Ion AmpliSeq™ Library Kit 2.0

(Life Technologies), barcoding each sample. Different cycling conditions were performed according to the DNA type: 22 cycles for FFPE gDNA and no-WGA cfDNA and 19 cycles for SF gDNA and WGA cfDNA in the first multiplex PCR, whereas in the second, optional PCR, the SF, FFPE and WGA cfDNA were subjected to five cycles, and the no-WGA cfDNA underwent seven cycles.

The library size was checked using the Agilent High Sensitivity DNA Kit by the Bioanalyzer 2100 instrument (Agilent Technologies), and library concentration was evaluated with a Qubit® 2.0 Fluorometer using the Agilent High Sensitivity DNA Kit (Life Technologies). Each diluted library (100 pM) was amplified through emulsion PCR using the OneTouch™ Instrument (Life Technologies) and enriched by the OneTouch™ ES Instrument (Life Technologies) using the Ion PGM Template OT2 200 KIT, following the manufacturer's instructions. Finally, sequencing was performed on the Ion PGM (Life Technologies) with the Ion PGM 200 Sequencing Kit (Life Technologies), loading barcoded samples into a 314*v*.2 or 316*v*.2 chip.

4.5. NGS Data Analysis

The PGM sequencing data were analyzed by the Ion Torrent Software Suite *v*.4.2 (Life Technologies) using the plugin Variant Caller (VC) *v*.4.2-r88446. Hotspot and targeted regions, together with the json parameters files associated with the Ion AmpliSeq Colon and Lung Cancer Panel *v*.1 (Life Technologies), were then loaded into the VC plugin. All identified variants were visually confirmed by the Integrative Genomics viewer (IGV) [56]. To validate data, a second round of analysis was performed as follows: FastQ data, generated by the Ion PGM™ semiconductor, were aligned to the Hg19 genome reference using BWA (Burrows-Wheeler Alignment)-MEM tool optimized for 70 bp–1 Mb long reads. Then, the resulting SAM (Sequence Alignment Map) files were converted and sorted to the BAM (binary version of SAM) format by SAMtools. Post-alignment steps included local realignment in the proximity of known indels and recalibration of the quality score, both performed by the Genome Analysis Toolkit (GATK) 2.5 Unified Genotyper (Realigner Target Creator, Indel Realigner and Base Recalibrator options) starting from sorted BAM files. Finally, GATK called variants in the target regions included in the BED file with Phred scores greater than 5 (those with a Phred score between 5 and 30 were marked as low-quality variants). Only Single Nucleotide Variants (SNVs) were considered, annotated by SNPEff, and compared to those called with the Ion Reporter™ tool [57]. All the samples belonging to the same patients were analyzed in multi-sample mode. Variants were annotated using the PolyPhen-2) [58], SIFT [59], and Genomic Evolutionary Rate Profiling (GERP) tools to predict the effect of missense mutations on the protein and calculate their conservation scores. Additionally, variants were named according to the Catalogue of Somatic Mutation in Cancer (COSMIC) [13] and Single Nucleotide Polymorphism Database (dbSNP). To distinguish common polymorphisms from rare variants, Global Minor Allele Frequency (GMAF) was calculated using the allele frequencies across all 1000 Genomes Phase I populations [60].

4.6. NGS Parameter Performance

To assess the performance parameters of our assay, two NSCLC cell lines and two DNA FFPE reference standards were sequenced according to our workflow. NGS data from the cell lines (NCI-H1650 and NCI-1975) were compared to genotypes obtained from published NGS data using Ion AmpliSeq™ Cancer Panel on Ion PGM sequencer (Life Technologies) [61]. Similarly, the data from the two FFPE reference standards were compared to genotype assessed by Horizon using droplet digital PCR. We also evaluated the performance parameters of our assay on WGA samples by comparing the mutation profiles of WGA amplified of NCI-H1650 scalar dilutions (10, 1, and 0.1 ng) with the corresponding unamplified cell line. The sensitivity, specificity, and PPV of the tests were calculated as described [62] and confidence intervals were calculated using the Confidence Interval Calculator (accessed on 1 February 2015).

4.7. Sanger Sequencing (SS) and Pyrosequencing (PS)

NGS variants with frequencies higher than 20% were validated by SS, using the specific gene primer of the panel as described by Coco (2012) [63]. Genetic variants with a frequency range from 5% to 20% were validated by PS. The primer sets for the PS assays (which include a PCR primer pair and a sequencing primer) were designed with the Pyrosequencing Assay Design software (Biotage, Uppsala, Sweden). All the primer sequences, the PCR amplification conditions, and the sequence to analyze are reported in Supplementary Table S5. The PCR reactions were performed by amplifying 100 ng of gDNA in a final volume of 50 μL containing 200 mol/L dNTPs, 1× Taq buffer, 1.5 mM MgCl$_2$, 0.2 μM of each PCR primer, 1.5 U of AmpliTaq Gold® DNA Polymerase (Life Technologies). The PCR program consists of 10 min at 95 °C and 45 cycles with 30 s at 95 °C, 30 s at specific annealing temperature of primer, and 30 s at 72 °C, followed by 5 min at 72 °C. The PS assays were performed with a PSQ 96MA instrument (Qiagen); the sequencing reactions were performed with the Pyro Gold reagent kit PSQ 96MA, according to the manufacturer's instructions, and the sequencing analysis conducted with the PSQTM 96MA software *v.*2.02 (Biotage AB, Uppsala, Sweden).

5. Conclusions

In conclusion, here we demonstrate the feasibility of obtaining DNA suitable for NGS from FFPE tissues by performing an appropriate DNA quality control assay based on a multiplex PCR. Indeed, adequate quality control of extracted DNA represents the most critical step of the entire procedure and in particular gDNA from an FFPE biopsy, which is highly fragmented and limited. The assessment of the suitability of the starting DNA for subsequent sequencing is mandatory for reducing the time and costs of the downstream processes due to a possible sequencing failure.

In particular, our NGS-based workflow was successfully able to screen critical samples, such as FFPE tissues, demonstrating the possibility of transferring this novel technology into routine clinical context. Our criterion for considering a variant as true positive was a variant allele frequency greater than 5% with coverage higher than 500×.

We also report that VC is a reliable software program for identifying SNVs and indels using Ion PGM NGS data, that streamline and simplify the data analysis, annotation, and reporting of data, reducing the time to results. However, a second round of variant calling by GATK software is suggested, particularly in a starting setup. The inclusion of a reference standard, harboring mutation with AF at 5%, to sequence in each run as well as a validation step with an independent test to confirm clinically actionable mutations, are recommended to confidently call mutations in a clinical contest.

We also report a feasible sequencing workflow by starting from low-input gDNA to be applicable in the detection of ctDNA mutational profiles. However, to translate this technology into the disease monitoring and treatment of NSCLC, further validation studies on a greater number of samples, specifically on cfDNA from metastatic NSCLC patients, are required.

Author Contributions: Irene Vanni, Simona Coco, and Francesco Grossi conceived and planned the study. Marco Mora revised the cell content of all tumors; Maria Giovanna Dal Bello performed the gDNA quality control. Irene Vanni, Simona Coco, and Anna Truini performed sequencing analysis by Next generation Sequencing and Sanger, and data interpretation. Barbara Banelli performed the validation analysis and data interpretation by PS. Marta Rusmini set up the bioinformatics pipeline for NGS data analysis. Irene Vanni, Simona Coco, Anna Truini, and Angela Alama contributed to the interpretation of the data and wrote the manuscript. Francesco Grossi, Erika Rijavec, Carlo Genova, Federica Biello, and Claudia Maggioni contributed to the interpretation of the data and revised the manuscript. All authors discussed the results and reviewed and approved the manuscript.

References

1. Siegel, R.; Ma, J.; Zou, Z.; Jemal, A. Cancer statistics, 2014. *CA Cancer J. Clin.* **2014**, *64*, 9–29. [CrossRef] [PubMed]

2. Pao, W.; Miller, V.; Zakowski, M.; Doherty, J.; Politi, K.; Sarkaria, I.; Singh, B.; Heelan, R.; Rusch, V.; Fulton, L.; *et al*. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13306–13311. [CrossRef] [PubMed]

3. Coco, S.; Truini, A.; Vanni, I.; dal Bello, M.G.; Alama, A.; Rijavec, E.; Genova, C.; Barletta, G.; Sini, C.; Burrafato, G.; *et al*. Next generation sequencing in non-small cell lung cancer: New avenues toward the personalized medicine. *Curr. Drug Targets* **2015**, *16*, 47–59. [CrossRef] [PubMed]

4. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef] [PubMed]

5. Higuchi, R.; Fockler, C.; Dollinger, G.; Watson, R. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Biotechnology* **1993**, *11*, 1026–1030. [CrossRef] [PubMed]

6. Ronaghi, M.; Uhlén, M.; Nyrén, P. A sequencing method based on real-time pyrophosphate. *Science* **1998**, *281*, 363–365. [CrossRef] [PubMed]

7. Hantson, I.; Dooms, C.; Verbeken, E.; Vandenberghe, P.; Vliegen, L.; Roskams, T.; Borght, S.V.; Nackaerts, K.; Wauters, I.; Vansteenkiste, J. Performance of standard procedures in detection of *EGFR* mutations in daily practice in advanced NSCLC patients selected according to the ESMO guideline: A large Caucasian cohort study. *Transl. Respir. Med.* **2014**, *2*. [CrossRef] [PubMed]

8. Cronin, M.; Ross, J.S. Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology. *Biomark. Med.* **2011**, *5*, 293–305. [CrossRef] [PubMed]

9. Couraud, S.; Vaca-Paniagua, F.; Villar, S.; Oliver, J.; Schuster, T.; Blanché, H.; Girard, N.; Trédaniel, J.; Guilleminault, L.; Gervais, R.; *et al*. BioCAST/IFCT-1002 investigators: Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: A proof-of-concept study from BioCAST/IFCT-1002. *Clin. Cancer Res.* **2014**, *20*, 4613–4624. [CrossRef] [PubMed]

10. Nomenclature for the description of sequence variants. Available online: http://www.hgvs.org/mutnomen/ (accessed on 10 April 2015).

11. Den Dunnen, J.T.; Antonarakis, S.E. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum. Mutat.* **2000**, *15*, 7–12. [CrossRef]

12. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; *et al*. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [CrossRef] [PubMed]

13. COSMIC (Catalogue of Somatic Mutations in Cancer). Available online: http://cancer.sanger.ac.uk/cosmic (accessed on 28 February 2015).

14. Choudhary, A.; Mambo, E.; Sanford, T.; Boedigheimer, M.; Twomey, B.; Califano, J.; Hadd, A.; Oliner, K.S.; Beaudenon, S.; Latham, G.J.; *et al*. Evaluation of an integrated clinical workflow for targeted next-generation sequencing of low-quality tumor DNA using a 51-gene enrichment panel. *BMC Med. Genom.* **2014**, *7*. [CrossRef] [PubMed]

15. Zhang, L.; Chen, L.; Sah, S.; Latham, G.J.; Patel, R.; Song, Q.; Koeppen, H.; Tam, R.; Schleifman, E.; Mashhedi, H.; *et al*. Profiling cancer gene mutations in clinical formalin-fixed, paraffin-embedded colorectal tumor specimens using targeted next-generation sequencing. *Oncologist* **2014**, *19*, 336–343. [CrossRef] [PubMed]

16. Hadd, A.G.; Houghton, J.; Choudhary, A.; Sah, S.; Chen, L.; Marko, A.C.; Sanford, T.; Buddavarapu, K.; Krosting, J.; Garmirek, L.; *et al*. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J. Mol. Diagn.* **2013**, *15*, 234–247. [CrossRef] [PubMed]

17. Hagemann, I.S.; Devarakonda, S.; Lockwood, C.M.; Spencer, D.H.; Guebert, K.; Bredemeyer, A.J.; al-Kateb, H.; Nguyen, T.T.; Duncavage, E.J.; Cottrell, C.E.; *et al*. Clinical next-generation sequencing in patients with non-small cell lung cancer. *Cancer* **2015**, *121*, 631–639. [CrossRef] [PubMed]

18. D'Haene, N.; le Mercier, M.; de Nève, N.; Blanchard, O.; Delaunoy, M.; el Housni, H.; Dessars, B.; Heimann, P.; Remmelink, M.; Demetter, P.; *et al.* Clinical Validation of Targeted Next Generation Sequencing for Colon and Lung Cancers. *PLoS ONE* **2015**, *10*, e0138245. [CrossRef] [PubMed]

19. Sah, S.; Chen, L.; Houghton, J.; Kemppainen, J.; Marko, A.C.; Zeigler, R.; Latham, G.J. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Med.* **2013**, *5*, 77. [CrossRef] [PubMed]

20. Simbolo, M.; Gottardi, M.; Corbo, V.; Fassan, M.; Mafficini, A.; Malpeli, G.; Lawlor, R.T.; Scarpa, A. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE* **2013**, *8*, e62692. [CrossRef] [PubMed]

21. Ross, M.G.; Russ, C.; Costello, M.; Hollinger, A.; Lennon, N.J.; Hegarty, R.; Nusbaum, C.; Jaffe, D.B. Characterizing and measuring bias in sequence data. *Genome Biol.* **2013**, *14*. [CrossRef] [PubMed]

22. Wilm, A.; Aw, P.P.; Bertrand, D.; Yeo, G.H.; Ong, S.H.; Wong, C.H.; Khor, C.C.; Petric, R.; Hibberd, M.L.; Nagarajan, N. LoFreq: A sequence quality aware, ultra-sensitive variant caller for uncovering cell population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **2012**, *40*, 11189–11201. [CrossRef] [PubMed]

23. Genome Analysis Toolkit (GATK). Available online: https://www.broadinstitute.org/gatk/index.php (accessed on 3 February 2015).

24. Boland, J.F.; Chung, C.C.; Roberson, D.; Mitchell, J.; Zhang, X.; Im, K.M.; He, J.; Chanock, S.J.; Yeager, M.; Dean, M. The new sequencer on the block: Comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum. Genet.* **2013**, *132*, 1153–1163. [CrossRef] [PubMed]

25. Gerlinger, M.; Rowan, A.J.; Horswell, S.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; Tarpey, P.; *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **2012**, *366*, 883–892. [CrossRef] [PubMed]

26. Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A., Jr.; Kinzler, K.W. Cancer genome landscapes. *Science* **2013**, *339*, 1546–1558. [CrossRef] [PubMed]

27. Sakurada, A.; Lara-Guerra, H.; Liu, N.; Shepherd, F.A.; Tsao, M.S. Tissue heterogeneity of EGFR mutation in lung adenocarcinoma. *J. Thorac. Oncol.* **2008**, *3*, 527–529. [CrossRef] [PubMed]

28. Zhang, J.; Fujimoto, J.; Zhang, J.; Wedge, D.C.; Song, X.; Zhang, J.; Seth, S.; Chow, C.W.; Cao, Y.; Gumbs, C.; *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **2014**, *346*, 256–259. [CrossRef] [PubMed]

29. De Bruin, E.C.; McGranahan, N.; Mitter, R.; Salm, M.; Wedge, D.C.; Yates, L.; Jamal-Hanjani, M.; Shafi, S.; Murugaesu, N.; Rowan, A.J.; *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **2014**, *346*, 251–256. [CrossRef] [PubMed]

30. Imielinski, M.; Berger, A.H.; Hammerman, P.S.; Hernandez, B.; Pugh, T.J.; Hodis, E.; Cho, J.; Suh, J.; Capelletti, M.; Sivachenko, A.; *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **2012**, *150*, 1107–1120. [CrossRef] [PubMed]

31. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**, *489*, 519–525.

32. Zhang, X.; Miao, X.; Guo, Y.; Tan, W.; Zhou, Y.; Sun, T.; Wang, Y.; Lin, D. Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Hum. Mutat.* **2006**, *27*, 110–117. [CrossRef] [PubMed]

33. Zheng, D.; Chen, Y.; Gao, C.; Wei, Y.; Cao, G.; Lu, N.; Hou, Y.; Jiang, X.; Wang, J. Polymorphisms of *p53* and *MDM2* genes are associated with severe toxicities in patients with non-small cell lung cancer. *Cancer Biol. Ther.* **2014**, *15*, 1542–1551. [CrossRef] [PubMed]

34. Govindan, R.; Ding, L.; Griffith, M.; Subramanian, J.; Dees, N.D.; Kanchi, K.L.; Maher, C.A.; Fulton, R.; Fulton, L.; Wallis, J.; *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **2012**, *50*, 1121–1134. [CrossRef] [PubMed]

35. Diaz, L.A., Jr.; Bardelli, A. Liquid biopsies: Genotyping circulating tumor DNA. *J. Clin. Oncol.* **2014**, *32*, 579–586. [CrossRef] [PubMed]

36. Zhang, L.; Cui, X.; Schmitt, K.; Hubert, R.; Navidi, W.; Arnheim, N. Whole genome amplification from a single cell: Implications for genetic analysis. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5847–5851. [CrossRef] [PubMed]

37. Telenius, H.; Carter, N.P.; Bebb, C.E.; Nordenskjold, M.; Ponder, B.A.; Tunnacliffe, A. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* **1992**, *13*, 718–725. [CrossRef]

38. Kamberov, E.; Sleptsova, I.; Suchyta, S.; Bruening, E.; Zeihler, W.; Nagel, J.S.; Langmore, J.; Makarov, V. Use of *in vitro* OmniPlex Libraries for high-throughput comparative genomics and molecular haplotyping. *Proc. SPIE* **2002**, *4626*, 340–351.

39. Langmore, J.P. Rubicon Genomics, Inc. *Pharmacogenomics* **2002**, *3*, 557–560. [CrossRef] [PubMed]

40. Lizardi, P.M.; Huang, X.; Zhu, Z.; Bray-Ward, P.; Thomas, D.C.; Ward, D.C. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **1998**, *19*, 225–232. [CrossRef] [PubMed]

41. Dean, F.B.; Hosono, S.; Fang, L.; Wu, X.; Faruqi, A.F.; Bray-Ward, P.; Sun, Z.; Zong, Q.; Du, Y.; Du, J.; *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5261–5266. [CrossRef] [PubMed]

42. Barker, D.L.; Hansen, M.S.; Faruqi, A.F.; Giannola, D.; Irsula, O.R.; Lasken, R.S.; Latterich, M.; Makarov, V.; Oliphant, A.; Pinter, J.H.; *et al.* Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.* **2004**, *14*, 901–907. [CrossRef] [PubMed]

43. Bergen, A.W.; Haque, K.A.; Qi, Y.; Beerman, M.B.; Garcia-Closas, M.; Rothman, N.; Chanock, S.J. Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources. *Hum. Mutat.* **2005**, *26*, 262–270. [CrossRef] [PubMed]

44. Park, J.W.; Beaty, T.H.; Boyce, P.; Scott, A.F.; McIntosh, I. Comparing whole-genome amplification methods and sources of biological samples for single-nucleotide polymorphism genotyping. *Clin. Chem.* **2005**, *51*, 1520–1523. [CrossRef] [PubMed]

45. Pask, R.; Rance, H.E.; Barratt, B.J.; Nutland, S.; Smyth, D.J.; Sebastian, M.; Twells, R.C.; Smith, A.; Lam, A.C.; Smink, L.J.; *et al.* Investigating the utility of combining Phi29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray genotyping. *BMC Biotechnol.* **2004**, *4*. [CrossRef] [PubMed]

46. Pinard, R.; de Winter, A.; Sarkis, G.J.; Gerstein, M.B.; Tartaro, K.R.; Plant, R.N.; Egholm, M.; Rothberg, J.M.; Leamon, J.H. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genom.* **2006**, *7*. [CrossRef] [PubMed]

47. Yilmaz, S.; Singh, A.K. Single cell genome sequencing. *Curr. Opin. Biotechnol.* **2011**, *23*, 1–7. [CrossRef] [PubMed]

48. Bergen, A.W.; Qi, Y.; Haque, K.A.; Welch, R.A.; Chanock, S.J. Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. *BMC Biotechnol.* **2005**, *5*. [CrossRef] [PubMed]

49. Hansen, H.M.; Wiemels, J.L.; Wrensch, M.; Wiencke, J.K. DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 1686–1690. [CrossRef] [PubMed]

50. Paez, J.G.; Lin, M.; Beroukhim, R.; Lee, J.C.; Zhao, X.; Richter, D.J.; Gabriel, S.; Herman, P.; Sasaki, H.; Altshuler, D.; *et al.* Genome coverage and sequence fidelity of Phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res.* **2004**, *32*. [CrossRef] [PubMed]

51. Hosono, S.; Faruqi, A.F.; Dean, F.B.; Du, Y.; Sun, Z.; Wu, X.; Du, J.; Kingsmore, S.F.; Egholm, M.; Lasken, R.S. Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* **2003**, *13*, 954–964. [CrossRef] [PubMed]

52. Lim, E.H.; Zhang, S.L.; Li, J.L.; Yap, W.S.; Howe, T.C.; Tan, B.P.; Lee, Y.S.; Wong, D.; Khoo, K.L.; Seto, K.Y.; *et al.* Using whole genome amplification (WGA) of low-volume biopsies to assess the prognostic role of EGFR, KRAS, p53, and CMET mutations in advanced-stage non-small cell lung cancer (NSCLC). *J. Thorac. Oncol.* **2009**, *4*, 12–21. [CrossRef] [PubMed]

53. Bettegowda, C.; Sausen, M.; Leary, R.J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B.R.; Wang, H.; Luber, B.; Alani, R.M.; *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **2014**, *6*. [CrossRef] [PubMed]

54. Diehl, F.; Schmidt, K.; Choti, M.A.; Romans, K.; Goodman, S.; Li, M.; Thornton, K.; Agrawal, N.; Sokoll, L.; Szabo, S.A.; *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **2008**, *14*, 985–990. [CrossRef] [PubMed]

55. Van Beers, E.H.; Joosse, S.A.; Ligtenberg, M.J.; Fles, R.; Hogervorst, F.B.; Verhoef, S.; Nederlof, P.M. A multiplex PCR predictor for aCGH success of FFPE samples. *Br. J. Cancer* **2006**, *94*, 333–337. [CrossRef] [PubMed]

56. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [CrossRef] [PubMed]

57. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]

58. PolyPhen-2 prediction of functional effects of human nsSNPs. Available online: http://genetics.bwh. harvard.edu/pph2/ (accessed on 28 February 2015).

59. SIFT Blink. Available online: http://sift.jcvi.org/ (accessed on 28 February 2015).

60. 1000Genomes. Available online: http://www.1000genomes.org/ (accessed on 28 February 2015).

61. Coco, S.; Truini, A.; Alama, A.; dal-Bello, M.G.; Venè, R.; Garuti, A.; Carminati, E.; Rijavec, E.; Genova, C.; Barletta, G.; *et al*. Afatinib resistance in non-small cell lung cancer involves the PI3K/AKT and MAPK/ERK signalling pathways and epithelial-to-mesenchymal transition. *Target Oncol.* **2015**, *10*, 393–404. [CrossRef] [PubMed]

62. Rehm, H.L.; Bale, S.J.; Bayrak-Toydemir, P.; Berg, J.S.; Brown, K.K.; Deignan, J.L.; Friez, J.; Funke, B.H.; Hegde, M.R.; Lyon, E.; *et al*. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **2013**, *15*, 733–747. [CrossRef] [PubMed]

63. Coco, S.; de Mariano, M.; Valdora, F.; Servidei, T.; Ridola, V.; Andolfo, I.; Oberthuer, A.; Tonini, G.P.; Longo, L. Identification of ALK germline mutation (3605delG) in pediatric anaplastic medulloblastoma. *J. Hum. Genet.* **2012**, *57*, 682–684. [CrossRef] [PubMed]