Low-level somatic variant detection in tumor FFPE samples by Sanger sequencing

This application note demonstrates:

- The power of high-sensitivity Sanger sequencing in one of its ideal applications—oncology research
- The detection of low-level (down to 5%) somatic variants using Sanger sequencing with Applied Biosystems™ Minor Variant Finder Software
- The utility of Sanger sequencing for first-line screening in addition to its role as the gold-standard confirmatory method of next-generation sequencing
- The successful interrogation of minimal amounts (as low as 0.1 ng) of DNA from formalin-fixed, paraffin-embedded (FFPE) samples
- The advantages of Sanger sequencing, including simple workflow and data analysis, low cost per sample (at a limited number of targets), and fast turnaround time

Introduction

DNA sequence variants play an important role in the initiation and progression of many different cancer types. These alterations could also predict prognosis, response to treatment, and specific therapy-associated toxicities. Molecular profiling of cancers is becoming more and more important not only as a diagnostic tool but also in research areas of personalized cancer therapy, or precision oncology. The detection of germline variants at a fixed ratio by gold-standard Sanger sequencing has been well established; however, the detection of somatic mutations, especially in heterogeneous tumor samples where variants may be present at a lower level, has been more challenging. Minor Variant Finder Software (MVF) enables calling of low-frequency variants at a detection level as low as 5% using Sanger sequencing. The improved sensitivity achieved through Minor Variant Finder Software makes Sanger sequencing an ideal choice for oncology research applications.

Next-generation sequencing (NGS) provides a valuable method for high-throughput applications when many targets and samples need to be multiplexed and screened. However, for very focused applications such as single-analyte or single-gene targets, there is a clear need for fast, simple, and affordable methods for detecting low-level somatic variants. The improved sensitivity achieved through Minor Variant Finder Software makes Sanger sequencing an ideal technology to fill this gap.

Somatic mutations typically drive carcinogenesis by deactivation of proteins that normally suppress tumorigenesis or by constitutive activation of proteins that drive carcinogenesis. Molecular profiling of cancer cells and the detection of variants in specific genes (e.g., *TP53*, *KRAS*, *NRAS*, *BRAF*, and *EGFR*) could be as important as identifying the tumor histological type. Depending on the tumor type, the entire coding sequences of some genes might need to be screened (e.g., mutations in the tumor suppressor gene *TP53* can be distributed throughout the entire coding sequence, often with functional impact), while only specific nucleotide positions need to be analyzed in other genes.

Thermo Fisher Scientific provides a complete workflow for tumor genetic analysis

We have developed a complete workflow for tumor genetic analysis using Sanger sequencing (Figure 1). For all the experiments described below, DNA was extracted from commercially obtained FFPE slides using Invitrogen™ RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE, and the DNA concentration was measured by Invitrogen™ Qubit™ fluorometric quantitation. PCR and sequencing reactions were performed using the Applied Biosystems™ BigDye™ Direct Cycle Sequencing Kit and a Veriti™ Thermal Cycler. Sequencing reactions were cleaned up using the Applied Biosystems™ BigDye XTerminator™ Purification Kit. Sequencing reactions were electrophoresed on the Applied Biosystems™ 3500xL Genetic Analyzer. FFPE test samples were referenced to DNA control CEPH 1347-02 sequenced in both forward and reverse directions, and processed under similar conditions throughout the entire workflow on the same Applied Biosystems™ MicroAmp™ Fast Optical 96-Well Reaction Plate sealed with MicroAmp™ Clear Adhesive Film.

Minor Variant Finder Software

Analysis of allele frequencies was performed using Minor Variant Finder Software. Minor Variant Finder Software is a user-friendly desktop software specifically designed for the detection and reporting of single nucleotide variants (SNV) in Sanger sequencing traces with a detection level as low as 5%. On a test set of 632,452 base positions, it exhibits a 5% limit of detection with 95.3% sensitivity and 99.83% specificity. The sophisticated algorithm filters out systematic noise components in bidirectional traces and highlights and presents genuine somatic variant candidates for review and reporting. Minor Variant Finder Software can also readily align sequences with the human reference genome and VCF files from NGS experiments, providing a smooth workflow for NGS confirmation with annotations in the dbSNP database.

Gene-specific panels

We have developed gene-specific Sanger sequencing panels covering the entire coding region (all exons) of specific genes (e.g., *TP53*, *KRAS*, and *NRAS*) implicated in tumorigenesis. The *TP53* coding sequence was covered by 24 short amplicons (Figure 2). Similarly, all *KRAS* exons were covered by 12 amplicons, and all *NRAS* exons were covered by 9 amplicons. Similar gene-specific panels can be built for any other genes of interest. We designed M13-tailed primers generating short amplicons in the size range of 50–151 bp in the *KRAS* and *NRAS* panels and 115–200 bp in the *TP53* panel to facilitate the analysis of potentially degraded DNA obtained from FFPE samples. To demonstrate the workflow of these panels with Minor Variant Finder Software, we analyzed DNA extracted from lung tumor FFPE samples. We initially determined variants of *TP53* and *KRAS* in these samples by NGS using the Ion PGM™ System.

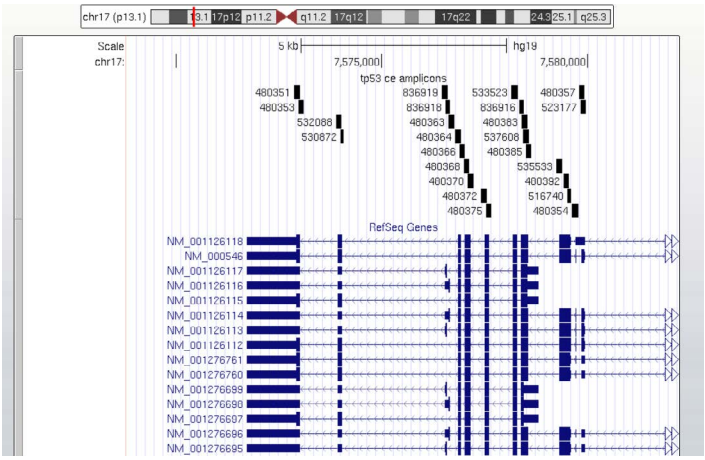

Figure 2. Single-gene Sanger sequencing panel of *TP53*. The entire coding sequence of *TP53* was covered by 24 amplicons. Similar gene-specific panels were built for *KRAS* and *NRAS* genes.

Isolate DNA using RecoverAll Total Nucleic Acid Isolation Kit

Perform PCR with primers designed for Sanger sequencing panels

Prepare for sequencing using BigDye Direct Cycle Sequencing Kit

Run sequencing using 3500xL Genetic Analyzer

Analyze data using Minor Variant Finder Software

Figure 1. A simple and fast workflow for tumor genetic analysis. First, DNA is extracted from FFPE slides. Next, PCR is used to amplify the DNA region of interest, which is followed by Sanger sequencing. The purified sequencing products are then separated by capillary electrophoresis. Finally, the sequencing traces are analyzed using Minor Variant Finder Software.

We confirmed the identity and minor allele frequency of these variants by the two gene-specific Sanger sequencing panels analyzed by Minor Variant Finder Software (Table 1 and Figure 3). Allele frequencies determined by Minor Variant Finder Software are comparable with variant allele frequencies (VAF) found by NGS. For example, variant c.1024C>T in FFPE sample 2162 was detected at 13.4% in the colon and lung cancer NGS panel and 12.6% in the Oncomine NGS panel, and was measured at 11% in the forward trace and 14% in the reverse trace by Sanger sequencing.

Table 1. Comparison of minor variants detected using NGS vs. Sanger sequencing.

TP53					
FFPE sample	**Variant**	**Primer ID**	**VAF, colon and lung cancer NGS panel**	**VAF, Oncomine NGS panel**	**VAF, Sanger sequencing with MVF**
2162	c.1024C>T	530872	13.4%	12.6%	Forward: 11% Reverse: 14%
2182	c.517G>T	836916	10%	Not detected	Forward: 8.2% Reverse: 8.4%
KRAS					
FFPE sample	**Variant**	**Primer ID**	**VAF, Colon and lung cancer NGS panel**	**VAF, Oncomine NGS panel**	**VAF, Sanger sequencing with MVF**
2182	c.35G>C	Hs00532827_CE	34.2%*	33.1%*, 41.1%**	Forward: 41% Reverse: 55%**
PB	c.182A>G	Hs00477023_CE	31.7%†	Not tested	Forward: 33% Reverse: 33%**

* Whole-tissue scrape (2014).
** Whole-tissue scrape (2015).
† Microdissected (2014), average of 17.2%, 36.5%, and 41.5%.

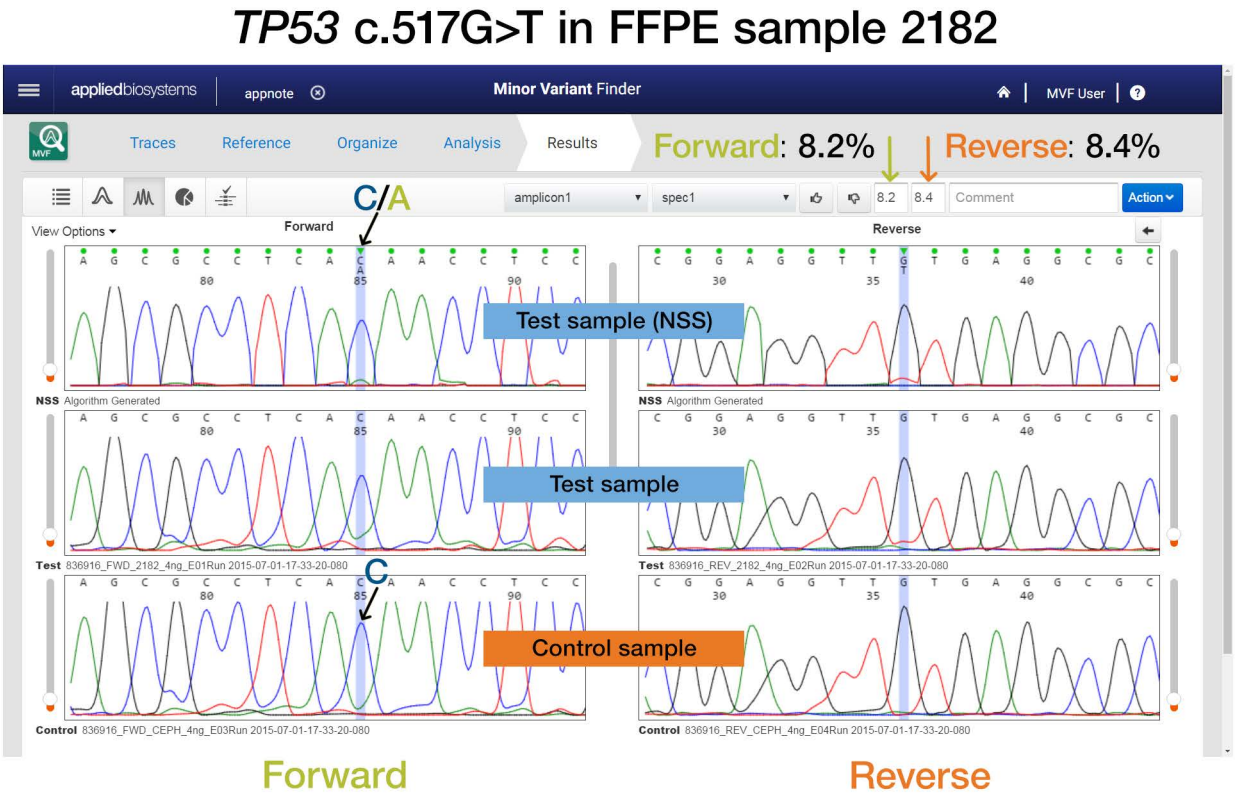

Figure 3. Electropherograms generated by Minor Variant Finder Software. Minor variant c.517G>T was detected in FFPE sample 2182 using amplicon 836916 from the *TP53* panel. The variant was detected at 8.2% in the forward and 8.4% in the reverse direction by Minor Variant Finder Software compared to the primary base C (or G in the corresponding reverse reaction). C (or G) was detected in the control sample at the allelic ratio of 100% (bottom electropherograms). Minor variant A in the forward reaction (similarly, the corresponding T in the reverse reaction) of the test specimen would have been easily missed by visual inspection of the electropherograms of the test sample (middle electropherograms). However, the Minor Variant Finder algorithm is able to identify the A (or T) allele as a minor variant candidate as shown in the algorithm-generated electropherograms (top) after noise subtraction and submission (NSS).

We also tested this workflow with lower amounts of DNA input from one of these samples. We confirmed the allele frequency of *KRAS* variant c.182A>G in FFPE sample PB with 1 ng, 0.5 ng, and 0.1 ng of DNA input amount (Table 2). The overall sequencing quality and the allele ratios appear to be more variable for reactions with starting materials lower than 1 ng DNA; however, the minor variant of interest was still detectable even with DNA input as low as 0.1 ng DNA.

Table 2. Effect of low DNA input amount on minor variant detection.

DNA input	VAF, forward reaction	VAF, reverse reaction
1 ng	33%	33%
0.5 ng	31%	33%
0.1 ng	23%	39%

Limit of detection

To establish a limit of detection (LOD), we made serial dilutions using DNA from one of these cancer FFPE samples mixed with a control DNA. Specifically, a G/T variant (forward: 48%, reverse: 51%) in FFPE sample 2162 at position Chr17: 7,579,619 in amplicon 480354 of the *TP53* panel was diluted with a 100% G base found in the CEPH DNA control sample. We mixed these two samples to generate a minor variant T at position Chr17: 7,579,619 with minor variant ratio at 50% (not shown), 25%, 12.5%, 6.25%, 3.125%, and 1.56% (not shown). Allelic ratios measured by Minor Variant Finder Software correlated well with the expected ratios in both forward and reverse reactions. These data show that this workflow can detect as little as 3% of a minor variant in an FFPE sample (Figure 4).

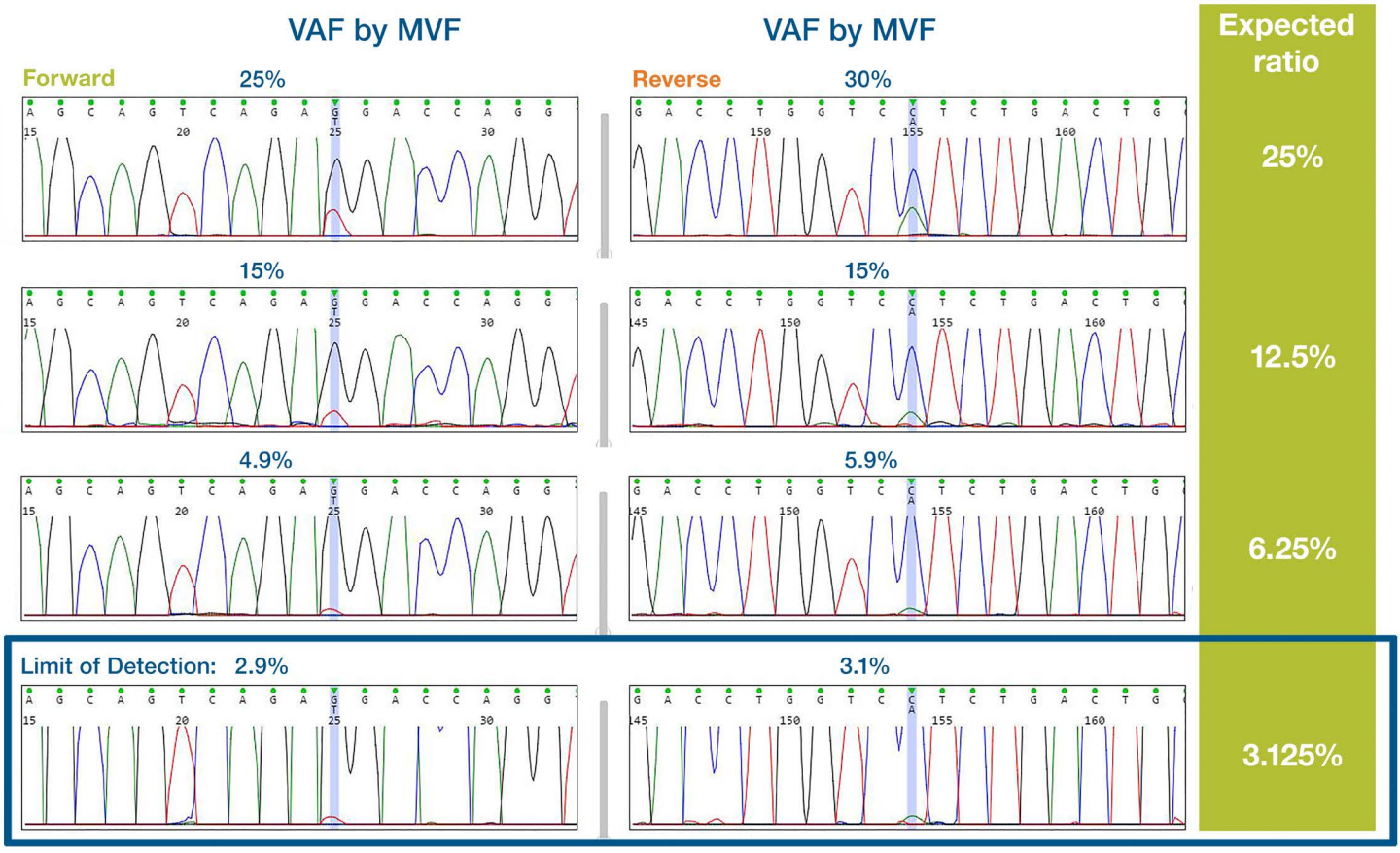

Figure 4. Limit of detection (LOD) study. The variant allele frequencies (VAF) detected by Minor Variant Finder Software were in line with the expected ratios and the LOD was found to be at ~3%.

Pan-cancer panel

Although a single gene could be the major pathogenic allele (e.g., *TP53* tumor suppressor), in many cases a combination of pathogenic SNVs of several different genes could contribute to the molecular profile of a tumor. To demonstrate the robustness and flexibility of using Sanger sequencing for oncology research, we included variants across many different solid tumor types in a pan-cancer panel. We narrowed our focus to alleles present at frequencies greater than 1% in common cancers. Primers flanking these variants were identified using the online Applied Biosystems™ Primer Designer™ tool, and the primers generating the shortest amplicon size were selected to facilitate analysis of potentially degraded FFPE DNA. Selected amplicons were in the size range of 126–179 bp. This resulted in a panel consisting of 26 amplicons that queried 66 COSMIC variants from 18 different cancer genes (Table 3). Note that these amplicons also encompass an additional 1,906 less-frequently found (<1%) COSMIC variants (including indels).

Table 3. Pan-cancer panel for high-sensitivity Sanger sequencing.

Gene	Assay ID*	Amplicon name	Amplicon length	COSMIC ID
AKT1	Hs00532882_CE	H	179	33765
BRAF	Hs00518328_CE	Y	172	476, 473
CTNNB1	Hs00532885_CE	L	171	5668, 5672, 5661, 5681, 5682, 5677, 5686, 5671, 5670, 5662, 5679, 5676, 5663, 5667
EGFR	Hs00532896_CE	V	170	6224
FGFR2	Hs00532795_CE	D	154	36903
FGFR3	Hs00532939_CE	P	137	714, 715
FGFR3	Hs00643836_CE	Q	232	716, 718
FGFR3	Hs00532901_CE	R	136	719
FGFR3	Hs00532870_CE	S	127	24802
GNA11	Hs00532801_CE	J	137	52969
GNAQ	Hs00527928_CE	Z	157	28758, 28757
HRAS	Hs00532842_CE	E	161	499, 498, 496
HRAS	Hs00532817-CE	F	133	483, 480
IDH1	Hs00532826_CE	K	150	28746, 28747, 28749
IDH2	Hs00532791_CE	I	173	34090, 33733
KIT	Hs00532832_CE	T	171	1217, 1290
KIT	Hs00532972_CE	U	128	1311, 1314
KRAS	Hs00532827_CE	G	172	532, 521, 520, 522, 518, 516, 517
MET	Hs00532963_CE	W	134	700
NRAS	Hs00532802_CE	A	126	584, 580
NRAS	Hs00532811_CE	B	134	564, 563
PIK3CA	Hs00532809_CE	M	160	27502, 746
PIK3CA	Hs00532895_CE	N	136	760, 763, 12458, 29315
PIK3CA	Hs00532909_CE	O	157	775, 776, 777
RET	Hs00532977_CE	C	168	965
SMO	Hs00532881_CE	X	138	13146

* The assay ID refers to the primer pairs provided by the Primer Designer Tool.

To test the panel, we identified 34 commercially available FFPE samples representing 12 different tissue types where allele frequency information was also available from our previous NGS studies. These FFPE samples were used to screen for 14 variants represented by 10 amplicons. Allelic ratios calculated by Minor Variant Finder Software were in line with variant allele frequencies found by NGS (Table 4).

Table 4. Sanger sequencing confirmation of NGS allele frequency data.

FFPE name	Tissue type	Amplicon name	Gene	Amino acid change	Reference/ variant	VAF, Sanger forward reaction (%)*	VAF, Sanger reverse reaction (%)*	VAF, NGS (%)
1191289	Uterus	H	*AKT1*	p.Glu17Lys	C/T	64.3	67.4	52.40
364	Melanoma	Y	*BRAF*	p.Val600Glu	A/T	2.7	16.9	4.71
367	Melanoma	Y	*BRAF*	p.Val600Glu	A/T	23.9	40.6	21.95
369	Melanoma	Y	*BRAF*	p.Val600Glu	A/T	39.3	50	23.63
372	Melanoma	Y	*BRAF*	p.Val600Glu	A/T	63.5	78.8	66.99
1181219	Brain	Y	*BRAF*	p.Val600Glu	A/T	49.3	43.8	37.19
1186528	Colon	Y	*BRAF*	p.Val600Glu	A/T	53.5	61.8	40.77
1188466	Uterus	L	*CTNNB1*	p.Ser33Cys	C/G	14.5	14.3	14.27
1200313	Lung	L	*CTNNB1*	p.Asp32Asn	G/A	34	28	25.06
1181943	Brain	K	*IDH1*	p.Arg132His	C/T	40.5	35.5	40.90
1188945	Brain	K	*IDH1*	p.Arg132His	C/T	27.3	22.5	25.11
1193124	Brain	K	*IDH1*	p.Arg132His	C/T	47.1	40.5	46.40
1160628	Lung	G	*KRAS*	p.Gly12Val	C/A	23.1	26.3	18.97
1182647	Colon	G	*KRAS*	p.Gly12Asp	C/T	24.2	35.6	22.04
1185114	Colon	G	*KRAS*	p.Gly12Asp	C/T	49.2	56.2	41.14
1187394	Lung	G	*KRAS*	p.Gly12Cys	C/A	17.6	17.9	20.39
1189051	Lung	G	*KRAS*	p.Gly12Asp	C/T	56.6	51.7	70.59
1191289	Uterus	G	*KRAS*	p.Gly12Val	C/A	35.3	26	24.76
1193721	Liver	G	*KRAS*	p.Gly12Val	C/A	8.4	8.7	7.27
1198166	Stomach	G	*KRAS*	p.Gly13Asp	C/T	6.1	14	13.33
1199157	Lung	G	*KRAS*	p.Gly12Val	C/A	22	25	20.39
1199257	Lung	G	*KRAS*	p.Gly12Asp	C/T	10.2	12.7	11.18
1200313	Lung	G	*KRAS*	p.Gly12Asp	C/T	37	46.8	36.04
330686m1	Pancreas	G	*KRAS*	p.Gly12Asp	C/T	17.6	29	11.17
365	Melanoma	A	*NRAS*	p.Gln61Lys	G/T	35.1	32.3	41.42
1181647	Brain	B	*NRAS*	p.Gly12Asp	C/T	9.6	6	6.76
1171721	Cervix	N	*PIK3CA*	p.Glu545Lys	G/A	30	27.2	29.70
1179932	Cervix	N	*PIK3CA*	p.Glu545Lys	G/A	33.4	34.6	34.01
1183966	Uterus	O	*PIK3CA*	p.His1047Arg	A/G	12.3	15.6	13.62
1183966	Uterus	O	*PIK3CA*	p.His1047Arg	A/G	15.1	15	13.62
1184127	Uterus	N	*PIK3CA*	p.Glu545Lys	G/A	35.3	26.8	32.41
1185114	Colon	N	*PIK3CA*	p.Glu545Lys	G/A	15.9	13.4	13.10
1186427	Ovary	M	*PIK3CA*	p.Arg88Gln	G/A	16.7	24.4	30.18
1188466	Uterus	N	*PIK3CA*	p.Glu545Lys	G/A	3.9	4	6.80
1191227	Esophagus	O	*PIK3CA*	p.His1047Arg	A/G	38.3	32.8	30.33
1194253	Lung	N	*PIK3CA*	p.Glu545Lys	G/A	8.4	10	14.25
1197327	Stomach	O	*PIC3CA*	p.His1047Arg	A/G	2.5	3.7	5.50
1199851	Stomach	N	*PIK3CA*	p.Glu542Lys	G/A	12.9	10.8	11.21
308452a4	Breast	N	*PIK3CA*	p.Glu545Lys	G/A	32.1	26.1	25.41

* Highlighted cells represent variants where the variant allele frequency (VAF) was significantly different between forward and reverse Sanger sequencing reactions. This is likely due to local sequence context–specific nucleotide incorporation differences.

Importantly, the correlation between the variant allele frequencies generated by Sanger sequencing coupled with Minor Variant Finder Software versus NGS was similar to the correlation between the variant allele frequencies of forward versus reverse Sanger sequencing reactions (Figure 5). This illustrates that there is as much variance in allele frequency between NGS and Minor Variant Finder Software results as there is in sequencing opposite DNA strands.

In many cases, the amount of DNA obtained from a tumor FFPE sample is very limited. We therefore analyzed the accuracy of minor variant detection using very low amounts of DNA input across FFPE samples from different tissue types and different genes. We selected three amplicons and four samples that covered common COSMIC mutations at differing frequencies. We then tested the ability to detect allelic variants using 10 ng, 3 ng, 1 ng, 0.3 ng, and 0.1 ng of input DNA (Table 5). The lower the amount of DNA sequenced, the higher the variation of allele frequency calls were observed. This could be explained simply by the stochastic variation in template sampling. For example, 0.1 ng of human genomic DNA contains only approximately 15 diploid genome copies.

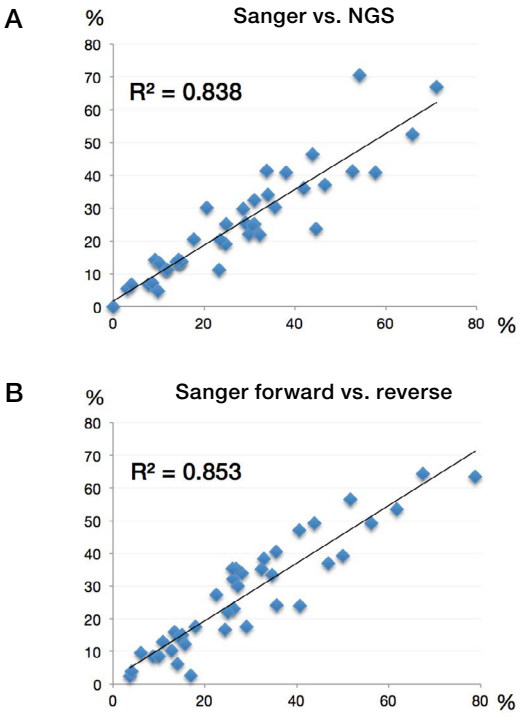

Figure 5. Correlation between allele frequencies. The correlation between **(A)** the average of allele frequencies generated from forward and reverse Sanger sequencing reactions vs. NGS allele frequencies is similar to the correlation between that of **(B)** allele frequencies generated by Sanger forward vs. reverse sequencing reactions.

Table 5. Effect of low DNA input amount on allele frequency across different FFPE samples.

FFPE sample	Variant allele	Tissue type	VAF, NGS (%)	DNA amount	Diploid copy number	VAF, Sanger forward reaction (%)	VAF, Sanger reverse reaction (%)
1187394	KRAS p.G12C	Lung tumor	20.39	10 ng	1,500	17.6	17.9
				3 ng	450	16.7	21.1
				1 ng	150	16.6	15.6
				0.3 ng	45	24.2	27.5
				0.1 ng	15	30.9	11.7
308452a4	PIK3CA p.E545K	Breast tumor	25.41	10 ng	1,500	32.1	36.1
				3 ng	450	27.8	29.3
				1 ng	150	30.4	23.9
				0.3 ng	45	29	23.1
				0.1 ng	15	11.9	27.7
372	BRAF p.V600E	Melanoma	66.99	10 ng	1,500	63.5	78.8
				3 ng	450	63.5	80.8
				1 ng	150	63.8	82.9
				0.3 ng	45	53.8	80.4
				0.1 ng	15	84.9	77.5
364	BRAF p.V600E	Melanoma	4.71	10 ng	1,500	2	8.2
				3 ng	450	5.5	10.8
				1 ng	150	4.8	No data
				0.3 ng	45	21.5	5.5
				0.1 ng	15	59	13.6

applied biosystems

Based on our findings, we recommend a minimum of 1 ng of input DNA for high-quality data. However, in situations where only a limited amount of biopsy or FFPE material is available, this approach could still be helpful in getting an actionable result. The overall sequencing quality and the allele ratios appear to be more variable for reactions with starting materials lower than 1 ng DNA, but the minor variants of interest were still detectable even with DNA input as low as 0.1 ng.

Conclusions

Sanger sequencing with Minor Variant Finder Software is not only an ideal tool for confirmation of minor variants detected by NGS, but it is also an attractive first-line screening choice when working with a limited number of targets. The workflow presented here can be used to detect as little as 5% of a minor variant in an FFPE sample using 1 ng (or less) DNA per reaction. This robust and simple Sanger sequencing approach also offers fast turnaround time (~4 hours including data analysis) at a low cost per sample, providing a valuable tool for oncology researchers.

Find out more at **thermofisher.com/sangerapplications**

ThermoFisher SCIENTIFIC