Validation of the Spotlight 59™ Cancer Panel Applied to IsoFlux Liquid Biopsies

An ultra-sensitivity somatic mutation detection workflow applicable to liquid biopsy samples.

Introduction

Whether derived from cell free DNA present in the extracellular fraction (cfDNA) or isolated circulating tumor cells (CTCs), "liquid biopsy" cancer samples present the challenge of detecting somatic variants present at very low concentrations by next-generation sequencing (NGS). While liquid biopsy variants are sometimes detected in the range of 1% - 5% allele frequency (AF, mutant as a percent of wild type), they are often present at much lower concentrations (0.1% -1% and below). Standard sequencing and data analysis approaches cannot address this range, as false positive calls are present between 1-5% in standard analysis pipelines (Frampton et al), and they overwhelm the true positive signal below 1%.

A number of different approaches have been put forward to improve accuracy, including deep sequencing combined with reduced mutation search space and various implementations of molecular barcoding. Molecular barcoding identifies which reads originate from the same starting sample molecule and use this information to reduce errors present in that set of reads. While molecular barcoding results in significantly lower error rates for consensus sequences, published analytical validation studies show reduced effectiveness below 1% allele frequency (Hiatt et al., Peng et al.). Recently introduced barcoding kits have been commercialized that cite sensitivities to 0.5%, without providing specificity in the space below 1% (Haloplex Data Sheet, Agilent). Barcoding strategies are limited by low efficiency of the molecular barcoding reaction. This means that only a relatively small percentage of sample molecules are incorporated into sequence-able fragments, reducing the biological diversity of the starting sample and limiting sensitivity. High depth and dedicated bioinformatics tools are also required (Hiatt et al., Peng et al.).

Here we introduce Spotlight 59, an oncology targeted amplification panel designed to leverage the statistical power of replicates and improved bioinformatics to

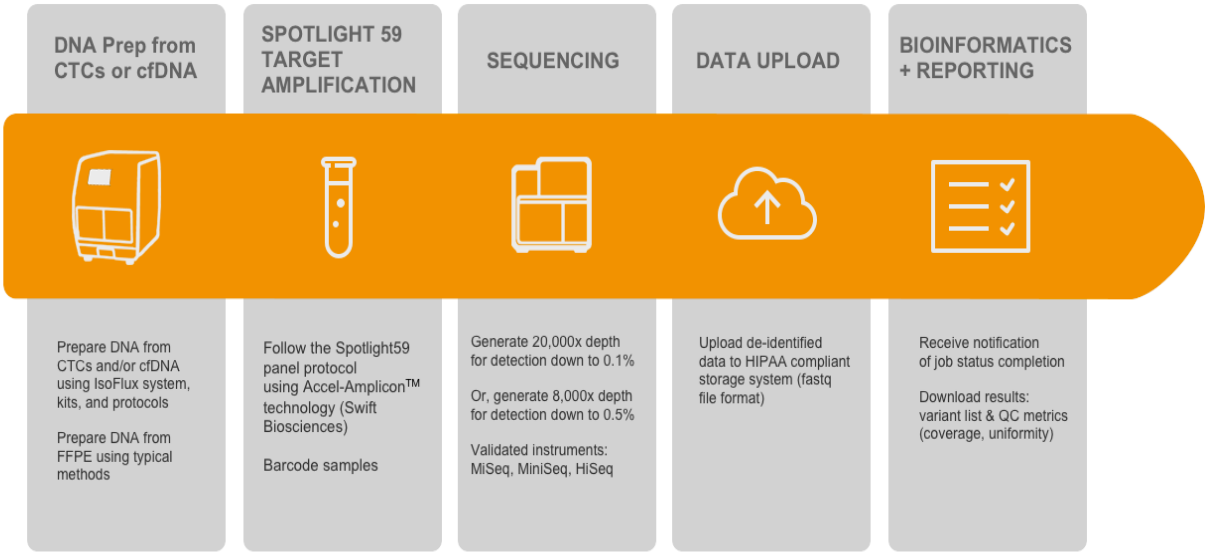


Figure 1: The Spotlight 59 liquid biopsy workflow for ultra-accurate detection of rare variants starts with isolation of CTCs and/or cfDNA using the IsoFlux system and kits. This is followed by targeted amplification and sequencing at 20,000X per sample. Fluxion's ERASE-Seq bioinformatics solution is accessed free of charge via secure upload to a HIPAA-compliant storage solution. Statistical comparison to a well-characterized normal control specific to the Spotlight 59 panel is performed, allowing accurate variant calling to 0.1% allele frequency. Variant calls are available for download within 24 hours. The analysis service is provided free of charge for Spotlight 59 customers.

Figure 2: The key technical specifications of the Spotlight 59 panel, and list of covered genes.

Feature	Specification
Input DNA Required	20-100 ng
Minimum starting cell number	200 total (CTC plus background)
Starting sample	CTC, cfDNA, FFPE
Time Required	2 hours
Number of Amplicons	277 (Spotlight 59) + 104 (sample ID)
Amplicon Size	Average 140 bp
Number of Genes Covered	59
Total Target Size	25.75 kb
On Target Percentage	> 90%
Coverage Uniformity at > 20% of Mean	> 90%
Limit of Detection (at 40 ng input, for base substitutions)	0.1%, 0.3% using WGA
Multiplexing on MiSeq v3 at 20,000X Average Depth	5 samples
ERASE-Seq bioinformatics	HIPAA-compliant cloud-based analysis included

ABL1	5	DDR2	1	FLT3	4	IDH2	2	MSH6	4	RET	6
AKT1	2	DNMT3A	1	FOXL2	1	JAK2	2	NOTCH1	3	STK11	5
ALK	2	EGFR	9	GNA11	2	JAK3	3	NPM1	1	SMAD4	10
APC	9	ERBB2	4	GNAQ	2	KDR	9	NRAS	3	SMARCB1	4
ATM	19	ERBB4	8	GNAS	2	KIT	14	PDGFRA	4	SMO	5
BRAF	2	EZH2	1	H3F3A	2	KRAS	3	PIK3CA	11	SRC	1
CDH1	3	FBXW7	6	HIST1H3B	4	MAP2K1	5	PIK3R1	8	TP53	21
CDKN2A	2	FGFR1	2	HNF1A	4	MET	6	PTEN	14	TSC1	1
CSF1R	2	FGFR2	4	HRAS	2	MLH1	1	PTPN11	2	VHL	3
CTNNB1	1	FGFR3	6	IDH1	1	MPL	1	RB1	12		

The Spotlight 59 Oncology Panel includes both clinically relevant hotspot loci and regions of contiguous coverage, depending on the allele distribution across each target gene. The table depicts the genes represented, followed by the number of amplicons for each gene.

Contiguous, overlapping coverage is included for APC, ATM, EGFR, FBXW7, FGFR3, H3F3A, HNF1A, KIT, MSH6, PIK3CA, PTEN, SMAD4, and TP53.

Comprehensive coding exon coverage is included for TP53.

detect mutations present at allele frequencies (AF) down to 0.1%. The computational method, termed ERASE-Seq (Elimination of Recurrent Artifacts and Stochastic Errors) relies on the utilization of a high-fidelity amplicon panel for which the background error rate has been well characterized. Spotlight 59 is a 277-amplicon cancer mutation panel developed in collaboration with Swift Biosciences and tailored to work with the ERASE-Seq approach. It includes 59 genes commonly associated with actionable mutations present in common cancer type.

The key specifications for the panel are shown in Fig. 2. Important advantages include a 2 hour preparation time and the short mean amplicon size (140 bp); the panel is applicable to both cell-derived DNA and shorter cell free DNA fragments that are commonly found in cell-free DNA. The key specifications and list of covered genes are shown in Fig 2

Workflow and Panel Specifications

The Spotlight 59 panel and ERASE-Seq workflow provide extraordinary sensitivity and specificity in the detection of rare mutant alleles in the range of 0.1-1%.

Variants					Cell lines (spiked at 1%)				Allelic Frequency			
chr	position	ref	alt	GENE	A549	H1975	12878	MDA	Expected	Spotlight 59 (20,000X) AF/ mut reads		Standard (25,000X)
9	21971153	C	A	CDKN2A		100%			1.0%	0.27%	76	nc
19	1207021	C	T	STK11	98%				1.0%	0.66%	238	0.6%
11	534242	A	G	HRAS	32%	54%			0.9%	0.78%	134	0.7%
4	55602765	G	C	KIT			46%	35%	0.8%	0.76%	80	0.7%
12	25398281	C	T	KRAS				58%	0.6%	0.47%	108	nc
7	140481417	C	A	BRAF				56%	0.6%	0.82%	416	0.8%
18	48586344	C	T	SMAD4				51%	0.5%	0.63%	60	0.6%
2	48030692	T	A	MSH6			50%		0.5%	0.55%	80	0.6%
2	209113192	G	A	IDH1	23%				0.2%	0.27%	38	nc
1	43815051	G	C	MPL					nc	nc	-	0.7%
3	178916894	T	C	PIK3CA					nc	nc	-	0.6%
19	1220379	T	A	STK11					nc	nc	-	0.7%
19	1220380	G	A	STK11					nc	nc	-	0.7%
19	1220398	T	A	STK11					nc	nc	-	0.8%
19	1220400	G	A	STK11					nc	nc	-	0.7%

Table 1. Analytical experiments are based on a spike-in of four different cell lines. Standard sequencing was performed on each of the pure cell line DNA samples and a mix containing each of the lines at 1% concentration each into a wild type background. The results of standard sequencing and variant calling, limited to 0.5% to reduce false positive rates, is compared to Spotlight 59 results. The standard caller had 6 false positive calls and missed 3 true positives of 9 true positive calls in the range of 0.2 to 1%.

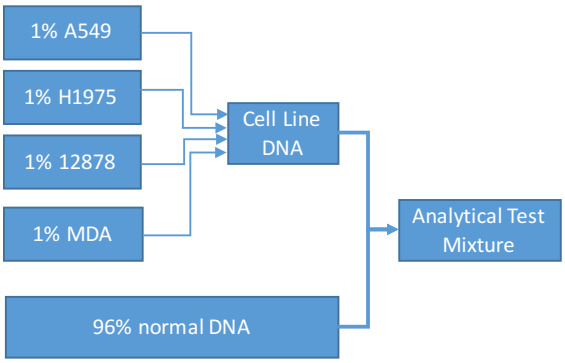


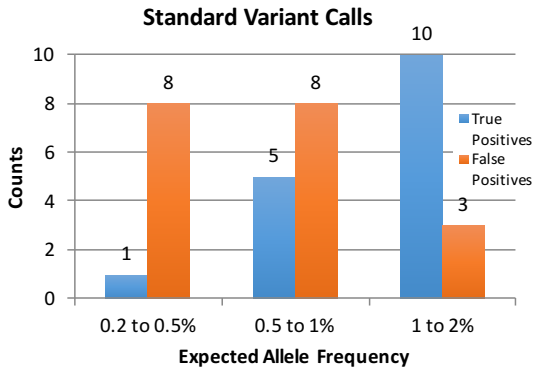
Figure 3: Analytical validation experiments. Tumor cell line DNA spiked into a normal DNA standard at a final concentration of 1% of the total DNA content.

The process consists of 5 different steps (Fig. 1). First, DNA is obtained from CTC, cfDNA or solid tumor samples. The second step is targeted amplification utilizing the Spotlight 59 kit, a single tube amplification reaction repeated over 4 technical replicates, each starting with with a DNA input of 10ng. The targeted amplification was developed in collaboration with Swift Biosciences, using the low noise Accel-Amplicon™ technology. The kit includes sample barcoding and provides the reagents required for full preparation of multiplex libraries. Each sample now consists of 4x different reactions.

Samples are then sequenced on Illumina instruments to a mean depth of 5,000x per replicate, 20,000x per sample; validated instruments include MiSeq, MiniSeq, and HiSeq.

A MiSeq v3 run can analyze 5 samples at a mean depth of 20,000x per sample. A HiSeq run can analyze 60 samples per lane at the required depth, resulting in a much lower cost per sample.

Sequencing data is then uploaded to Fluxion's HIPAA compliant storage solution. Each sample consists of several fastq files, representing each of the amplification reactions. A fully encrypted upload to



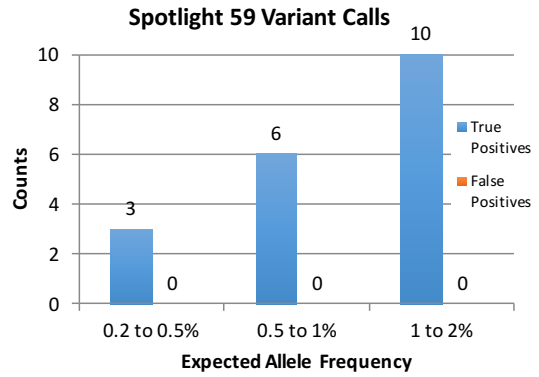


Figure 4: A histogram of true positive calls and false positive calls comparing Spotlight 59 panel results (and ERASE-Seq informatics) to a standard analysis at a similar At a high sensitivity setting, the standard caller missed only 3 true positive calls, but allowed 19 false positives.

HIPAA-compliant storage is provided, enabling subsequent data analysis on a secure server. A QC pipeline is used to analyze each replicate to assure adequate coverage and uniformity (must be >90%) followed by application of the proprietary ERASE-Seq algorithm for low-frequency variant calling.

The resulting variant calls are made available for download within 24 hours. Access to the ERASE-Seq bioinformatics solution is provided free of charge for

Approach	Panel	Allelic Frequency	Sensitivity	False Positives / 10kB
Traditional Caller[4]	Foundation Medicine	2-5%	98%	0.25
Molecular Barcoding[3]	Qiagen Custom Panel	1-2%	85%	0.76
Molecular Barcoding[2]	Custom Panel	0.75-1.5%	72%	0.56
Traditional Caller	56G Oncology Panel	0.3-0.5%	33%	3.37
Traditional Caller	56G Oncology Panel	0.5-1%	83%	3.37
ERASE-Seq	**Spotlight 59**	**0.2-0.5%**	**100%**	**0**
ERASE-Seq	**Spotlight 59**	**0.5-1%**	**100%**	**0**

Table 2. A comparison of the sensitivity and false positive rate of barcoding approaches, traditional callers and the Spotlight 59 data analyzed using the ERASE-Seq approach. Swift 56G results using a traditional caller are also reported because the panel chemistry for the 56G Oncology Panel is the same high fidelity chemistry used in the development of the Spotlight 59 panel.

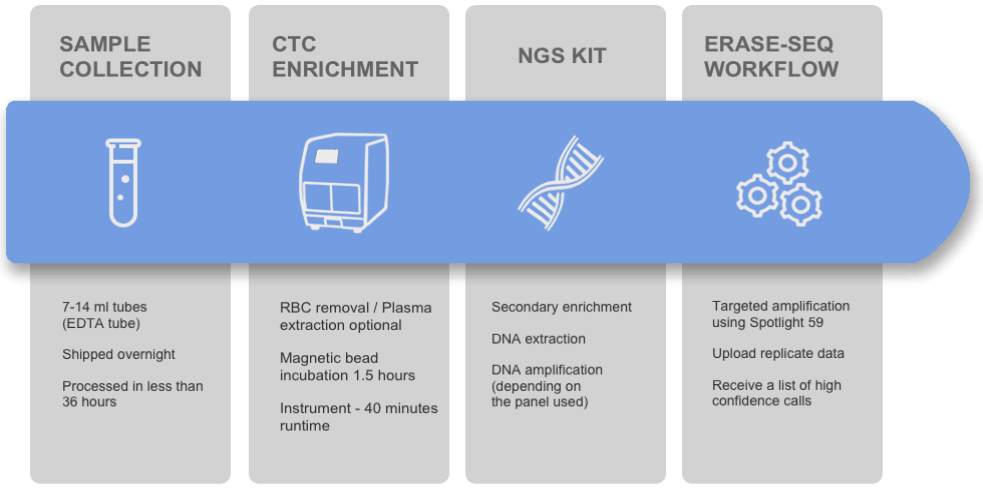


Figure 5. A workflow for utilizing the Spotlight 59 panel with IsoFlux liquid biopsy samples is detailed here. Blood collection, CTC enrichment and DNA preparation via Fluxion's NGS Kit precede the Spotlight amplification step.

Spotlight 59 customers. Please contact Fluxion for applications of ERASE-Seq on a custom panel.

Analytical Validation

The limit of detection and false positive rates of the Spotlight 59 workflow were determined using control samples created by spiking DNA from 4 cell lines (3 tumor derived and one B-lymphocyte line) from the International HapMap Project into healthy control DNA (NA19129 DNA - Coriell Institute for Medical Research). While each cell line's DNA was spiked in at a 1% concentration with respect to the total number of copies (Fig 3), allele distribution within the cell line itself ranged from 20 to 100%. The final expected allelic fractions for the mutations analyzed spanned a range from 0.2% up to 1%. The distribution of variants is shown in greater detail in Table 1.

A comparison between Spotlight 59 results and typical deep sequencing results using the same set of amplicons is also shown in Table 1. A standard caller (LoFreq) was used on data at a similar read depth as the ERASE-Seq data; the quality and AF cutoffs were optimized to maximize both sensitivity and specificity. When using Spotlight 59 and ERASE-Seq, the caller was allowed to operate without a frequency cutoff (down to below 0.1%). The output had 100% sensitivity and no false positive calls (100% specificity) for all variants studied. Variants were detected down to the lowest expected allele frequency (AF = 0.2%), so the sensitivity was verified down to this level. Note that the standard run was at a similar depth as the

Sample	Indication	Variants Detected
#01	PC	No Mutations Detected
#02	PC	17:7577139 (G>A) TP53 Missense; 10:89692949 (T>A) PTEN Missense; 1:43815041 (C>T) MPL Intronic; 9:133748451 (C>T) ABL1 Intronic; 11:108170563 (T>C) ATM Missense; 15:90631917 (T>C) IDH2 Missense
#03	PC	2:25457243 (G>A) DNMT3A Missense; 3:178952042 (C>A) PIK3CA Missense; 8:38282179 (G>T) FGFR1 Missense; 13:28610195 (G>A) FLT3 Intronic; 13:48936995 (C>T) RB1 Nonsense
#04	PC	4:153249456 (C>T) FBXW7 Missense; 7: chr7 (G>A) EGFR Missense
#05	PC	12:25398284 (C>T) KRAS Missense; 4:55953771 (T>C) KDR Intronic
#06	PC	9:21971099 (G>A) CDKN2A Missense; 2:212530147 (T>C) ERBB4 Missense
#07	PC	9:139399399 (G>A) NOTCH1 Missense
#08	RCC	10:123258056 (C>A) FGFR2 Missense
#09	RCC	No Mutations Detected
#10	RCC	No Mutations Detected
#11	RCC	3:10183812 (A>G) VHL Missense; 17:7577094 (G>A) TP53 Missense; 17:7578399 (G>A) TP53 Synonymous
#12	BC	5:149433690 (C>T) CSF1R Missense; 11:108123566 (G>T) ATM Nonsense; 17:7578503 (C>T) TP53 Missense
#13	BC	19:1207094 (G>C) STK11 Missense
#14	BC	2:212576808 (C>T) ERBB4 Missense; 17:7577568 (C>T) TP53 Missense
#15	BC	7:55221802 (G>A) EGFR Synonymous
#16	CRC	11:108236087 (G>A) ATM Missense; 7:55259502 (A>G) EGFR Missense
#17	CRC	No Mutations Detected

Table 3. Results for 17 clinical samples processed according to the IsoFlux liquid biopsy workflow and the high sensitivity Spotlight 59 panel. Variants were detected in a majority of samples, including common mutations for the indications listed (i.e TP53, KRAS, EGFR).

Spotlight panel, yet had a much higher false positive rate even at a cutoff of AF = 0.5%; for the ideal parameters shown in Table 1 Standard Run column, 6 false positives are present across this panel, while two of the true positives are missed. Another representation showing the comparison to standard sequencing and variant calling is shown in Fig. 4. Here we have a histogram of the true and false positives detected by either method for key expected allele frequency ranges: 0.1-0.5%, 0.5%-1% and 1% to 2%. While the Spotlight 59 workflow detects all expected variants, standard sequencing workflows have significant false positive rates and loss of sensitivity in this range (Fig. 4).

Clinical Validation & the IsoFlux Liquid Biopsy Solution

One of the key applications of Spotlight 59 and the ERASE-Seq methodology is mutation detection in DNA derived from liquid biopsy samples, where the allele frequency is often found to be between 0.1% and 1%. For this data set, samples were obtained using Fluxion's IsoFlux CTC workflow, which has been previously presented (Harb et al.). A schematic of the workflow is shown in Fig. 7. Blood samples are collected in EDTA tubes and must be processed within 36 hours of draw. This allows for overnight shipping from collection to processing centers, and preserves viable CTCs. The enrichment step is completed using the IsoFlux instrument, a process that includes a microfluidic separation step performed on up to 4 samples in parallel. For each set of samples, the overall processing time is about 4 hours. Successful sequencing requires the use of the IsoFlux NGS kit post enrichment. The NGS protocol includes a secondary enrichment step that reduces the white blood cell background (WBC) from a few thousand cells to 300-700 cells. The kit also provides the necessary reagents for cell lysis, DNA extraction and high-fidelity whole genome amplification (WGA). The starting sample for the WGA reaction is a set of approximately 500 intact genomes, of which up to 25% are of tumor origin.

The resulting samples after the IsoFlux NGS workflow contain 50-150µg of WGA-amplified DNA. A small portion of this DNA becomes the input of the Spotlight 59 workflow, which was described in detail above and in Fig. 1. For this validation study, 17 samples from patients already diagnosed with a solid tumor cancer (bladder, prostate and colorectal) at a range of stages were analyzed using the Spotlight 59 panel. Abnormalities were detected with an allele frequency range of 0.2 – 5%, and present in 13/17 samples. A summary of the variants called are presented in Table 3. Calls were made using the ERASE-Seq pipeline; when the same analysis parameters were applied to negative and spiked control samples, no false positive calls were made. For five of the samples shown, calls were verified by re-amplification and re-sequencing of the same sample using a different amplicon panel and sequencing instrument, at a different site (data not shown).

Conclusions

- Spotlight59 is a robust and low noise tumor profiling amplicon panel consisting of 277 amplicons, tailor-built to work w/ ERASE-Seq and liquid biopsy samples
- ERASE-Seq is a method for accurate and sensitive detection of low frequency DNA variants that is immediately implementable to existing amplicon panel workflows. It represents a performance improvement with respect to molecular barcoding methods
- Analytical validation using DNA mixtures demonstrates variant calling down to 0.2% AF with perfect sensitivity and specificity (no false positives across the panel), at a depth of 20,000x per sample
- When applied to clinical CTC samples resulting from the IsoFlux CTC workflow, this approach detects a number of variants in the range of 0.1-5 %. This approach detects alterations in a large fraction of the samples studied.
- ERASE-Seq and the Spotlight 59 panel are suitable for low allele frequency mutation detection from CTC- and cfDNA-derived liquid biopsy samples.

References

1. **Harb**, Wael, et al. "Mutational Analysis of Circulating Tumor Cells Using a Novel Microfluidic Collection Device and qPCR Assay" Translational Oncology 6.5 (2013): 1-11
2. **Hiatt**, Joseph B., et al. "Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation." Genome research 23.5 (2013): 843-854.
3. **Peng**, Quan, et al. "Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes." BMC genomics 16.1 (2015): 1.
4. **Frampton** et al. "Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing" Nature Biotechnology 31.11 (2013): 1023
5. **Haloplex Data Sheet** "High sensitivity rare variant detection using molecular barcodes" - data sheet (http://www.agilent.com/cs/library)