iontorrent

Detection of somatic mutations at 0.1% frequency from cfDNA in peripheral blood with a multiplex next-generation sequencing assay

Dumitru Brinza, Ann Mongan, Richard Chien, Dalia Dhingra, Jian Gu, Janice Au-Young, Fiona Hyland and Kelli Bramlett - Thermo Fisher Scientific · 180 Oyster Point Blvd South San Francisco CA 94080

ABSTRACT

Effective blood screening for the purpose of tracking tumor recurrence and resistance of tumors may improve outcomes in the future. Research studies suggest that virtually all tumors carry somatic DNA mutations, and these may serve as biomarkers that also can be tracked in blood. One of the sources containing tumor DNA in blood is circulating cell-free DNA (cfDNA). Tumor DNA comes from different tumor clones, and its abundance in plasma can be very low at critical stages such as early recurrence or development of resistance. Hence, there is great interest in being able to detect mutation biomarkers at very low frequency from cfDNA for detection and characterization of tumor clones.

We present a research use only analysis workflow for peripheral monitoring that enables detection of low frequency DNA variants. We developed an analysis algorithm that models errors accumulated during amplification and sequencing, and accurately reconstructs sequence of original DNA molecules based on multiple next generation sequencing reads. The reads contain genomic sequence and an adaptor that allows identification of reads originated from the same DNA molecule. We then developed a variant calling method that uses accurately reconstructed sequences to enable sensitive and specific detection of somatic mutations to 0.1% allele ratio. We demonstrate the analysis in control and archived cfDNA and FFPE research samples.

Sample-to-variant 2 day cfDNA Lung Assay Workflow

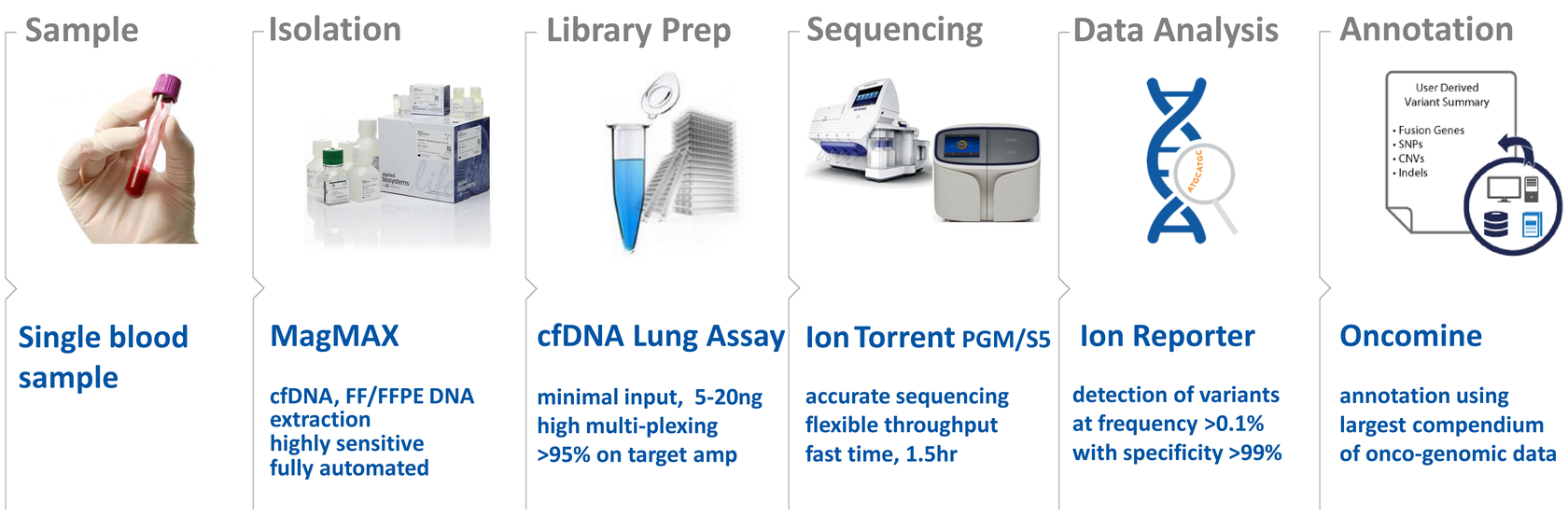

Figure 1. The cfDNA Lung Assay workflow enables detection of low frequency variants from DNA samples. We validated the workflow on 5-20ng cfDNA/FF/FFPE samples. We demonstrated high sensitivity/specificity for variants at frequency >0.1% with 20ng input DNA, and >0.5% with 5ng. The workflow is compatible with lower frequency variant detection, but requires higher amount of input DNA (Figure 3).

RESULTS

We achieved >95% sensitivity with >20ng input DNA and >85% sensitivity with 20ng input DNA and <1 false call per sample for variants in hotspot positions present at frequency 0.1% in the sample (Table 1). Due to sampling variability the detected frequency of the variants ranges between 0.05% and 0.15% (Figure 2). The workflow delivers 100% sensitivity and 100% specificity with 10ng input for variants at frequency above 0.5% Control samples were used for sensitivity calculations, cfDNA samples were used for specificity and FP rate calculations.

Sample Input	cfDNA 20ng	FFPE/cfDNA 10ng
LOD	0.10%	0.50%
Sensitivity (%)	89.6±5.8	100%
Specificity (%)	99.4±0.3	100%
FP/sample	0.25	0

Table 1. Performance of the cfDNA Lung Assay workflow. Reported are mean and one standard deviation.

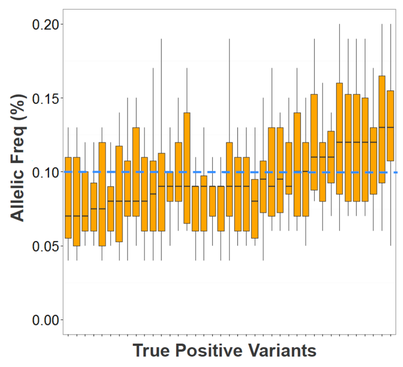

Figure 2. Detected variant frequencies of 40 control variants in fragmented 0.1% AcroMetrix™ Oncology Hotspot Control mix.

Library preparation and Sequencing

We used centrifugation at 1600 x g for 10 min at 4°C to extract plasma from blood. cfDNA was extracted from the plasma fraction using MagMAX™ cfDNA isolation protocol.

We used next generation sequencing cfDNA Lung Assay that allows interrogation of 171 biomarkers relevant in lung from COSMIC and Oncomine™ databases, and de-novo variant detection at ~1,700 genomic positions in 11 genes implicated in non-small cell lung cancer (NSCLC). The assay delivers >95% on target reads and highly uniform amplification across targeted cfDNA molecules (Figure 6).

We barcoded 8/32 samples and ran them on a single Ion S5™ 530/540 sequencing chip, that enables very deep (>50,000x coverage) and accurate sequencing. We also tested 4 samples on a single Ion PGM™ 318 chip that delivers (>25,000x coverage). The research assay requires a small amount of input DNA (~20ng for 0.1% LOD, Figure 3), and has a fast turnaround time from extracted DNA to variants of less than 24 hr.

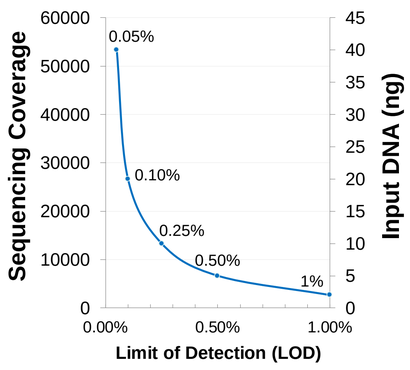

Figure 3. The cfDNA Lung Assay sequencing and input requirements for LOD levels ranging from 0.01% to 1%.

Verification Data

We tested the limits of variant detection in controlled dilution series, in cfDNA, and in FFPE cell lines.

1. First, we diluted engineered plasmid controls (AcroMetrix™ Oncology Hotspot Control) in background GM24385 genomic DNA down to 0.1% or 0.5% frequency, and then fragmented the DNA mix into fragments with average size of 170bp. AcroMetrix sample contains 40 common tumor mutations interrogated by our assay. The distribution of fragment size looks similar to Horizon's cfDNA reference sample (Figure 4).
2. Next, we used 0.1%, 1%, 5% Horizon's (HD780) cfDNA reference sample that contains 8 low frequency mutations at our hotspot positions including two large insertion and deletion variants of size >10bp.
3. Finally, we performed analytical verification of variant detection performance in normal cfDNA samples and FF/FFPE tumor samples.

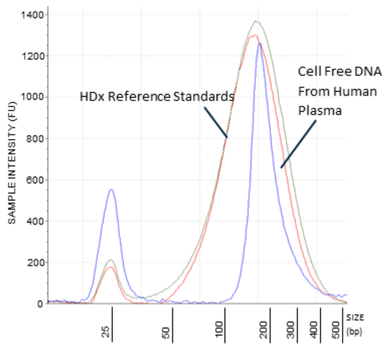

Figure 4. DNA fragment size distribution in Horizon control sample and human cfDNA from Plasma. For control samples we doubled input DNA amount in order to match the number of DNA fragments longer than 110bp in human cfDNA.

cfDNA Lung Assay

The cfDNA Lung Assay is a PCR based assay where forward and reverse gene specific primers are located 40-60bp apart to accommodate the fragmented nature of cfDNA/FFPE DNA. In addition to gene specific sequence primers contain a random sequence (forming 4K possibilities) followed by A/P1 adaptors for forward/reverse primers. We first perform two PCR cycles to copy each strand of original DNA fragment into a fragment with random sequences (Tags) and A/P1 adaptors attached to 5'/3' ends. Two Tags on each side of the fragment enable 16 million unique identifiers (UID) per amplicon. This number is significantly higher than the number of fragmented genomes in input DNA (we expect ~5K genomes in 20ng), thus most DNA fragments from the same genomic location will have an unique UID. We then perform 18 PCR cycles with universal A/P1 adapters followed by sequencing.

ANALYSIS - SOFTWARE

Reads with the same UID originate from the same DNA fragment (same family). We call a family functional if it contains at least 3 reads, which enables accurate reconstruction of the sequence of original DNA fragment (Figure 5). We require sufficient input DNA amount and sufficient sequencing to generate large number of functional families in order to accurately and robustly call variants at desired frequency (Figure 3). For 0.1% LOD we require 20ng input DNA and >25,000x read coverage. That generates more than 2500 functional families (molecular coverage) on each target (Figure 6).

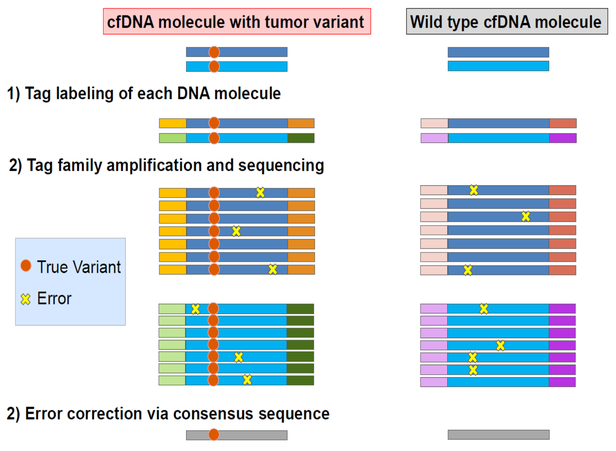

Figure 5. cfDNA Lung Assay Workflow utilizes molecular tagging to label individual DNA molecules. Tag families are generated after amplification, and consensus sequences are built to eliminate errors and represent the original molecules.

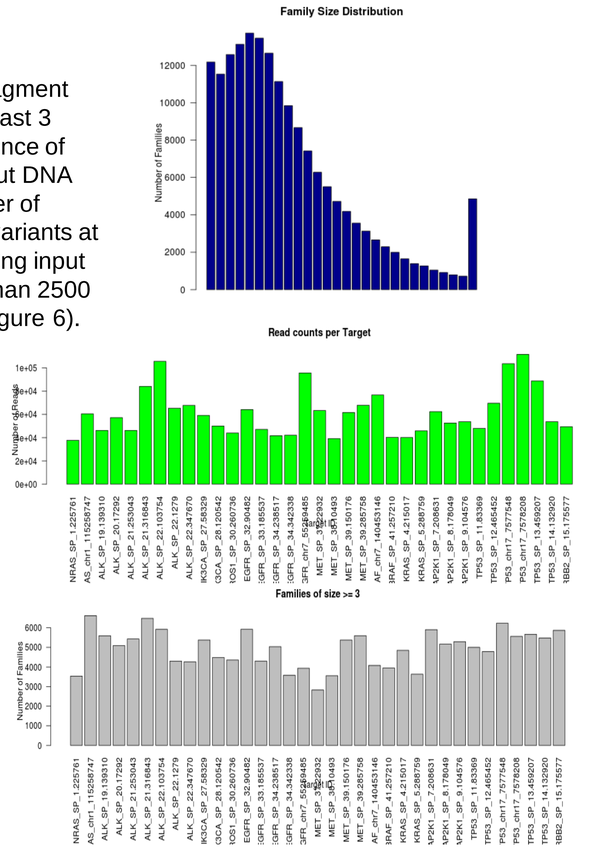

Figure 6. Family size distribution, amplicon read coverage, and amplicon molecular coverage generated from 20ng of human cfDNA using cfDNA Lung Assay containing 35 target regions, sequenced at 55,000x average coverage.

Position	Ref	Variant	Allele	Frequency	Quality	LOD	Coverage	Allele Read Coverage	Allele Read Frequency	Molecular Coverage	Allele Mol Cov
chr1:115256530	G	T	NRAS p.Q61K	0.07 %	28.0	0.05 %	70,497	45	0.06 %	6,884	5
chr1:115256536	C	T	NRAS p.A59T	0.23 %	39.0	0.05 %	70,428	155	0.22 %	6,888	16
chr3:178936091	G	A	PIK3CA p.E545K	0.13 %	32.0	0.05 %	65,741	92	0.14 %	6,208	8
chr3:178952085	A	G	PIK3CA p.H1047R	24.32 %	85.0	0.05 %	64,537	15,310	23.72 %	6,159	1,498
chr7:55241707	G	A	EGFR p.G719S	19.78 %	85.0	0.05 %	78,625	15,536	19.76 %	7,983	1,579
chr7:55242465	GGAATTAAGAGAAGC	-	EGFR p.E746_A750delELREA	0.08 %	29.0	0.05 %	79,867	84	0.11 %	7,553	6
chr7:55249010	-	GCCAGCGTG	EGFR p.V769_D770insASV	0.09 %	29.0	0.05 %	54,287	123	0.23 %	6,935	6
chr7:55249071	C	T	EGFR p.T790M	0.08 %	29.0	0.05 %	82,428	64	0.08 %	7,788	6
chr7:55259515	T	G	EGFR p.L858R	0.06 %	18.0	0.05 %	68,584	32	0.05 %	3,398	2
chr7:140453136	A	T	BRAF p.V600E	31.91 %	84.0	0.05 %	97,184	31,817	32.74 %	4,487	1,432
chr12:25398284	C	T	KRAS p.G12D	0.08 %	29.0	0.05 %	82,277	85	0.10 %	7,477	6
chr15:66727451	A	C	MAP2K1 p.Q56P	24.31 %	85.0	0.05 %	59,465	14,421	24.25 %	6,155	1,496

Figure 6. The cfDNA Lung Assay Workflow is a fully automated 2 days workflow available in Ion Reporter and Torrent Suite Software. Example of variant caller view on Horizon 0.1% control sample showing 8 true calls expected at 0.1% frequency and 4 true calls expected at 20-30% frequency.

CONCLUSIONS

The cfDNA Lung Assay Workflow with the Ion Torrent™ platform is a comprehensive 2 days sample-to-variant solution that facilitates researchers to study relevant biomarkers at 0.1% frequency in cfDNA/FF/FFPE DNA. Analysis is compatible with lower frequency variant detection, but will require higher input DNA amount and higher sequencing coverage (Figure 3).

ThermoFisher SCIENTIFIC