Correlation between mutations found in FFPE tumor tissue and paired cfDNA samples

Brittany Niccum PhD, Cindy Heath, Lauren Saunders PhD, Antonia Hur, and Asmita Patel
Beckman Coulter Life Sciences, Indianapolis, IN

Introduction

Liquid biopsies represent a promising area of facilitating cancer research as blood collection is less invasive than tumor biopsies. Cell free DNA (cfDNA) consists of small (150 – 500 bp) DNA fragments that circulate in the blood. cfDNA levels tend to be low in healthy, non-pregnant patients, and increase in patients with cancer, pregnancy, or extensive damage to tissue. cfDNA is believed to be derive mostly from apoptotic cells for which biomarkers for a variety of diseases have been found in cfDNA.

Apoptotic or necrotic cell death results in near-complete digestion of native chromatin from normal cell, tumor or fetus.

Each 160-175 bp DNA is wrapped ~1.67 times around one nucleosome. These protein-bound DNA fragments preferentially survive digestion and are released into the circulation, and can be recovered from peripheral blood plasma as cfDNA.

Typical cfDNA peaks characterized by Agilent 2100 Bioanalyzer, with a main peak at 175 bp, second and third peaks at 350 and 525 bp.

FFPE tissue is often used to look for cancer-associated mutations despite invasiveness; however it does not always correlate with the mutations seen in cfDNA. In this poster we present a comparison of matched FFPE and plasma samples to determine how many mutation are seen in both tissues. We also look at where the mutational mismatches appear in the chromosome. Different chromosomal regions can have different mismatch rates, and we use this to draw conclusions about the best chromosomal locations for biomarkers. We automated from extraction through sequencing in collaboration with Swift biosciences.

As cfDNA is extracted from blood, it is a non-invasive way to detect disease; however, there is some concern that cfDNA does not contain the same biomarkers as tumor tissue. Tumor tissue is typically removed and stored as formalin-fixed, paraffin-embedded tissue, a process that preserves the morphological structures well but chemically modifies and degrades the nucleic acids.

Methods

Sample Preparation

DNA was extracted from FFPE tissues using *FormaPure XL* Total automated on a Biomek i5 multichannel workstation. DNA concentrations for 4 10 µM curls were estimated using the Quant-iT Picogreen; yields varied between blocks. Some blocks had very low yields, most likely due to tissue distribution in the block; for these samples the extraction was repeated with 7 10 uM curls. DNA from FFPE was sheared on a Covaris S220 following the 200bp shear protocol. cfDNA was extracted from 1 mL of plasma using the *Apostle MiniMax™ High Efficiency cfDNA* kit. cfDNA yield varied between samples as expected. Concentration was estimated using the Kapa hg-Quant kit

Library Construction

cfDNA and sheared FFPE DNA was processed with *Swift Accel-NGS 2S Hyb DNA Library Kit*. DNA input was 100ng. Following the library construction, the *Swift Pan-Cancer Hyb Panel* was used for gene enrichment. The hybridized libraries were sequenced on the Illumina NextSeq 550.

Analysis

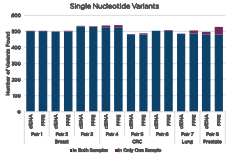
The sequencing data was analyzed using the BWA-enrichment tool on Illumina BaseSpace mapped to the genes in the panel.

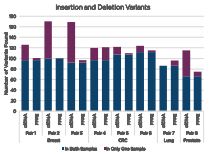
The two extractions, library construction and hybridization panel have all been automated on a Biomek Workstation. The approximate time for an automated workflows are compared to manual work flows in the table below.

	Throughput per run	Manual Timings		Automated Timings	
		Hands on Time	Total Time	Hands on Time	Total Time
Apostle MiniMax™ High Efficiency cfDNA kit	24 Manual/ 96 Automation	45 minutes	75 minutes	30 minutes	5.3 hours
FormaPure XL Total	24 Manual/ 96 Automation	3.5 hours	6.5 hours	30 minutes	8.5 hours
Swift Accel-NGS 2S Hyb DNA Library Kit	96	2.7 hours	4.6 hours	30 minutes	6 hours
Swift Pan-Cancer Hyb Panel	16	2.5 hours	8.5 hours	30 minutes	8.8 hours

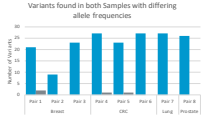
Majority of Variants are identified in both sample types

We identified both single nucleotide variants (SNV) and insertion and deletion (indel) events in the sequencing results from all the samples. First we compared the SNVs found in both sample types. The majority of the SNVs and indels were identified in both sampling methods and all tissue types (Figure 1 and Figure 2). In all of the samples unique SNVs and indels were identified in both cfDNA and FFPE DNA. More SNVs were identified when sequencing DNA from FFPE than cfDNA, however this difference wasn't statistically significant. Conversely, more indels were identified when sequencing cfDNA and this difference is statistically significant.



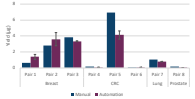


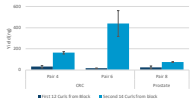
We also wanted to verify that the variants that were found in both sample types were found at equal frequency. This was determined by comparing the allele frequency identified by sequencing cfDNA and DNA from FFPE tissue. We compared the ratio of the two allele frequencies. For all 8 pairs the ratio of cfDNA allele frequency to FFPE allele frequency was 1:1; the allele frequency for variants identified by cfDNA and FFPE were not significantly different than each other. Interestingly though. When looking at variants with allele frequency 2 standard deviations from the mean, cfDNA showed allele frequencies higher than FFPE for more variants.



Methods Continued: Extraction Results and Sequencing Coverage

DNA extracted from FFPE was done both manually and on a Biomek i5 multichannel workstation. Both the manual user and the liquid handler extracted similar amounts of DNA (Figure 1); they are not significantly different as determined by a student t-test (P=0.79). Three of the blocks had very low yields and were extracted again. The initial low yield were most likely due to in block tissue distribution. In figure 2, the first 12 curls had much lower yield of DNA per curl than the second 14 curls did.





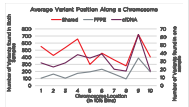
Below is a table of the sequencing coverage for all the samples. All samples had sequencing coverage of 50x at least 92% of the bases sequenced.

Tumor Tissue		Sample Material	Mean Region Coverage Depth	Uniformity of Coverage	Target Coverage at			
					1X	10X	20X	50X
Breast	Pair 1	cfDNA	1946.5	99.50%	100.00%	100.00%	100.00%	100.00%
		FFPE	1322.4	98.10%	100.00%	100.00%	100.00%	99.90%
	Pair 2	cfDNA	1677.2	89.50%	100.00%	100.00%	99.90%	99.50%
		FFPE	1306.5	99.10%	100.00%	100.00%	100.00%	100.00%
	Pair 3	cfDNA	1051.4	90.50%	100.00%	99.90%	99.80%	99.20%
		FFPE	1931	98.50%	100.00%	100.00%	100.00%	100.00%
CRC	Pair 4	cfDNA	1973.2	99.30%	100.00%	100.00%	100.00%	100.00%
		FFPE	1239.8	89.50%	100.00%	100.00%	99.90%	99.50%
	Pair 5	cfDNA	2108.1	99.40%	100.00%	100.00%	100.00%	100.00%
		FFPE	1480.1	98.70%	100.00%	100.00%	100.00%	100.00%
	Pair 6	cfDNA	1772.4	99.00%	100.00%	100.00%	100.00%	100.00%
		FFPE	2656	99.60%	100.00%	100.00%	100.00%	100.00%
Lung	Pair 7	cfDNA	1775.8	99.50%	100.00%	100.00%	100.00%	100.00%
		FFPE	336.9	87.90%	100.00%	99.60%	98.60%	92.30%
Prostate	Pair 8	cfDNA	1436.4	99.10%	100.00%	100.00%	100.00%	99.90%
		FFPE	598.6	91.70%	99.90%	99.80%	99.60%	98.60%

Variant location

To determine if mutations were evenly distributed throughout chromosomes, each chromosome was divided into 10 bins; the first 10% of each of the 23 chromosomes were treated as one bin. The mutations were pooled into the 10 bins for each chromosome. This was done separately for mutations found in only DNA from only FFPE or cfDNA and found in both FFPE DNA and cfDNA.

The graph to the right shows how the mutations mapped across the 10 bins. A Pearson's correlation was done to test for how differently the mutations mapped across the 10 bins. Mutations found only in FFPE tissue did not correlate (P=0.5) as well as mutations found only in cfDNA (P=0.8) to the position of mutations found in both FFPE DNA and cfDNA.



Variants Found at ClinVar Pathogenic Locations

We wanted to determine if we found any pathogenically relevant variants. To do this we compared against a list of variants from the NCBI ClinVar database. We only used variants that had the clinical significance of conflicting interpretations, uncertain significance, likely pathogenic, pathogenic and risk factor. We did this for all four of the tumor types. We were able to identify 2 variants in the ClinVar database only found using the DNA from FFPE tissue.

Tissue	Pair	cfDNA, FFPE, or Both	Gene	Chromosome	Position	Mutation	ClinVar mutation	Conditions associated
CRC	Pair 4	FFPE	APC	chr5	1.12E+08	C>T	C>T	Familial adenomatous polyposis
CRC	Pair 6	FFPE	TP53	chr17	7578406	C>T	C>T	Carcinoma of colon|Adrenocortical carcinoma, hereditary,Li-Fraumeni syndrome

In most of the samples however, we were unable to find any variants identified in the ClinVar database. We were however, able to find variants at the same position as variants in the ClinVar database. These could also alter the encoded protein and would be good candidates for further study into phenotype. Interestingly all of these variants were found in the FFPE DNA sequences; one of them was only found in FFPE DNA sequences; higher sequencing coverage of the cfDNA could help to identify these variants.

Tissue	Pair	cfDNA, FFPE, or Both	Gene	Chromosome	Position	Mutation	ClinVar mutation
Breast	Pair 1	Both	BRCA2	chr13	32911888	A>G	A Deletion
		Both	BRCA1	chr17	41223094	T>C	GCC insertion
		Both	BRCA1	chr17	41244936	G>A	G Duplication or TT insertion
		Both	CHEK2	chr22	29130458	T>C	T Deletion
		Both	TP53	chr17	7579472	G>C	G>T
		Both	BRCA1	chr17	41223094	T>C	T>A
		Both	BRCA1	chr17	41244000	T>C	T>A
		Both	BRCA1	chr17	41244435	T>C	T>G
		Both	BRCA1	chr17	41244936	G>A	G>C or G>T
	Pair 2	Both	BRCA1	chr17	41223094	T>C	T Deletion
		Both	TP53	chr17	7579472	G>C	G>T
		Both	BRCA1	chr17	41223094	T>C	T>A
		Both	BRCA1	chr17	41244000	T>C	T>A
		Both	BRCA1	chr17	41244435	T>C	T>G
	Pair 3	Both	STK11	chr19	1222012	G>C	G>A or G>T
		Both	BRCA2	chr13	32911888	A>G	A Deletion
		Both	STK11	chr19	1222012	G>C	G>A
Lung	Pair 7	FFPE	TP53	chr17	7579471	G>GC	G Duplication
Prostate	Pair 8	Both	BRCA1	chr17	41223094	T>C	GCC insertion
		Both	BRCA1	chr17	41244936	G>A	G Duplication or TT insertion
		Both	CHEK2	chr22	29130458	T>C	T Deletion

Conclusions

Here we show:

- Sequencing of cfDNA captures the majority of variants that found in sequenced FFPE DNA
- More indels are identified using cfDNA than with FFPE DNA, especially with breast tissue
- Distribution of variants across the genome differs when sequencing FFPE DNA
- More previously identified clinically relevant variants, as identified by the ClinVar database were found when sequencing FFPE DNA

This study is small and further work should be done using larger data sets to gain more conclusive information.