Automated Extraction of DNA from FFPE for Whole Exome Sequencing

Shawn K. Westaway¹, Sumit Punj², Charisma Jenkins¹, Arpita Kulkarni², John H. Letaw², Amiee B. Potter², C. Sue Richards², Sheila Weinmann¹

¹Center for Health Research, Kaiser Permanente Northwest, Portland, OR ²Oregon Health & Science University Knight Diagnostic Laboratories, Portland, OR

INTRODUCTION

- Extracting DNA suitable for next generation sequencing from formalin-fixed, paraffin-embedded (FFPE) tissue remains a challenge due to reduced yield, fragmentation, and protein-crosslinking. The degree of DNA decay can be affected by:¹
  - the fixation time length
  - pH of the preservative
  - formalin strength
  - fixation temperature
  - age of the sample at fixation
  - age of the FFPE sample on extraction
  - storage condition of the FFPE sample
- Breast tissue is particularly refractory due to lower cell numbers and higher fat content², creating complications for studying breast cancer specimens.
- Manual methods for extraction from frozen tissue have been examined and compared in the literature but are difficult for high throughput biorepositories³.
- This pilot study sought to optimize a semi-automated FFPE DNA extraction protocol from breast and lymph node tissue for downstream genomic testing.
- Genes in the P450 pathway were chosen due to their role in metabolizing tamoxifen, a medication prescribed to prevent breast cancer recurrence.

METHODS

Extractions

- DNA was extracted from 5- or 10-micron thin sections of formalin-fixed, paraffin-embedded (FFPE) normal tissue from breast cancer subjects (see **Table 1**).
  - Ten sample blocks were extracted in duplicate for comparison of extraction kit efficiency.
  - Section surgery collection dates ranged from 1985 to 2016, and fixation times ranged from 6 to 72 hours.
  - One to five sections were incubated overnight at 70°C in 200 µl incubation buffer with proteinase K (Promega kit).
  - The following day, the 20 samples were lysed and extracted with two different Promega FFPE DNA Purification kits tested in parallel on a Maxwell® 16 automated DNA extractor (Maxwell® 16 FFPE Plus LEV and Maxwell® 16 FFPE Tissue LEV, Promega Corp., Madison, WI).
  - Low elution volume (LEV) resulted in samples with approximately 45 µl of DNA.
- ### Quantifications
- Two µl DNA was quantified via NanoDrop and concentrations (ng/µl) are reported in **Table 1** with DNA/protein ratios (A 260/280).
  - As measured by NanoDrop, 500 pg were further tested in triplicate by qPCR for 16 samples

- (Agilent NGS FFPE QC kit; Agilent, Santa Clara, CA). Samples with five sections were excluded from further analysis (see Discussion).
- Samples were run on an Applied Biosystems StepOnePlus™, which allows quantification of the DNA by standard curve analysis based on reference DNA PCR amplification of a 42-bp amplicon.
  - Calculation of a DNA integrity score (ΔΔCq in **Table 1**) is based on concurrent amplification of a 123-bp amplicon in the same genomic region.
- ### Whole exome methods
- Three representative samples purified with the FFPE **Plus** LEV kit were used to obtain whole exome sequence (WES) at the Knight Diagnostics Laboratories at Oregon Health & Science University.
  - Samples were chosen to represent all available tissue types, a range of DNA/protein ratios (A 260/280), a range of DNA integrity scores, and average DNA concentrations. Libraries were prepared from 200 ng DNA by a protocol modified for Low Input FFPE DNA.
  - Samples were run on the Illumina NextSeq 500, and NGS metrics were collected and assessed. Sequencing data were processed through Galaxy-based, research grade analysis pipelines and filtered for the specified gene list for this project.

RESULTS

- ### Results from extraction, quantification, and quality assessment by qPCR are shown in Table 1.
- Samples are arranged in pairs with the Plus kit in color. The three samples chosen for further testing are listed at the top (#'s **1, 3, and 5**).
  - Averages and ranges for concentrations and A260/280 ratios are shown separately for the Plus and the Tissue kits.
  - DNA integrity scores (ΔΔCq) are calculated from the differential amplification of two different-sized overlapping regions of the genome, then compared to the reference sample run in each assay. The resulting scores should range from 0 to >5, with lower scores indicating better quality DNA.
- ### A standard curve example from Agilent NGS FFPE qPCR assay is shown in Figure 1.
- CT (cycle threshold) is plotted against quantity of DNA in reaction in picograms. The cycle threshold is defined as the number of cycles required for the fluorescent signal to exceed background level.
  - Standards are shown in green and experimental samples in blue.

Results from whole exome sequencing.

- Three representative DNA samples were selected for further testing by whole exome sequencing from the Maxwell® 16 FFPE **Plus** LEV kit extractions. One sample was used from normal lymph, normal breast, and normal lymph/other.
  - Sample concentration range: 43.9 to 113.7 ng/µl.
  - A 260/280 ratio range: 1.73 - 1.82.
- DNA integrity score range: 1.69 (best) to 2.38 (borderline).
- Age of FFPE sample at extraction range: 10 to 30+ years old.

Descriptive statistics on depth of sequencing coverage were computed for each sample and across samples by gene (Table 2).

- Average depth for all genes for the set was 409.641. Average depth for the three samples was not significantly different (F = 0.27186, p=0.05). However, average depth for all genes was lower for sample **5** (Average depth = 386.45).
- Distribution fitting of depth across samples by gene was performed and found to be normally distributed.
- To detect outliers, Grubbs Outlier Test was conducted across samples by gene. The observed G value was significant (p<0.05) for one gene (UGT1A4) for sample **5** and was thus identified as an outlier. However, clinical reporting requires depth >10X, therefore, this sample is within accepted clinical QC metrics.
- Descriptive statistics on gene coverage at >10X depth for each sample and across samples by gene are shown in **Table 2**. Gene coverage of greater than 98% is considered fully covered. Average gene coverage at 10X depth or greater for the set for all genes was 99.935%. Average gene coverage at 10X depth by sample was as follows: 99.864% (**1**), 100% (**3**), and 99.94% (**5**). Average gene coverage for the three samples was not significantly different (F = 0.62989, p>0.05).
- Variant calls from the gene panel list in **Table 2** were filtered for population minor allele frequency (MAF) and are shown in **Table 3**. The sample numbers (**1, 3**, and/or **5**) indicate which has a particular variant. Chromosome, reference base, alternate base, called genotype, gene, NCBI SNP database “rs” ID, substituted function and protein change in the gene are also shown, followed by two deleterious measures (PolyPhen & SIFT). Variant severity measures for the PolyPhen and SIFT are: PolyPhen (>.95 deleterious), SIFT (<0.05 deleterious). Population minor allele frequencies for both European populations and local (Pacific Northwest) populations are given.

DISCUSSION

- On average, twice as much DNA was obtained using the Promega FFPE Plus kit when compared to the FFPE Tissue kit, with only a slight drop in DNA quality as measured by A260/280 ratio. A260/280 ratios under 1.80 did not seem to impact the coverage of whole exome sequence results.
- Not surprisingly, lymph tissue yielded as much as 10 times that for breast tissue. The best yield by qPCR QC assay was from one 10-µm section of lymph tissue using the Plus kit for extraction.
- Using more than three sections in the incubation sometimes resulted in a paraffin plug developing inside the tube. This may have caused reduced protease or lysis efficiency during the overnight incubation and created difficulty when transferring the solutions to the DNA purification cartridges.
- The age of the tissue block did not seem to impact results. Impact of fixation time on results is unknown, as this information was not available at individual specimen level.
- Average gene coverage at 10X depth or greater for a subset of genes in the whole exome sequence results was 99.94% and did not significantly differ over the three samples submitted. While one of two normal lymph DNA samples had an outlier on depth of coverage for one gene from the subset, it was still within clinical requirements.

CONCLUSIONS

- Extraction from one 10-µm section for normal lymph tissue and three 10-µm sections for normal breast tissue using the Promega Maxwell® 16 FFPE Plus LEV kit gave the best yields, including analysis by qPCR.
- Similar gene sequence coverage and depth statistics were obtained for both tissue types and for a range of tissue block ages from 1985 to 2008 and were well above clinical QC metrics.

ACKNOWLEDGMENTS

We would like to thank Lina Issak, Tami Roland, and Deepannita Roy at Kaiser Permanente Northwest Airport Way Regional Laboratory for providing the thin sections, the Kaiser Permanente Northwest pathologists for the use of the tissue blocks, and the sales and technical team at Promega Corporation for help in selecting and applying the correct kits for our analysis.

REFERENCES

- Campos PF, Gilbert TM (2012) *DNA extraction from formalin-fixed material*. **Methods Mol Biol.** **840**:81-5.
- Mathot L, Wallin M, Sjöblom T (2013) *Automated serial extraction of DNA and RNA from biobanked tissue specimens*. **BMC Biotechnology** **13**: 66.

KAISER PERMANENTE
Center for Health Research

Table 1. FFPE Sample Extraction and Quantification Results DNA was extracted from 10 FFPE tissue blocks in duplicate using two types of extraction kits (Plus kit in purple/green). Three representative samples were chosen for further analysis by whole exome sequencing (purple).										
Sample #	Tissue Type	Year of Surgery	Section Size (microns)	# of Sections	Extraction Kit Type	NanoDrop Conc. (ng/uL)	A 260/280	Agilent Conc. (ng/uL)	ΔΔCq	Used for WES
1	normal lymph/other	1985	10	1	Plus	210.5	1.81	51.63	1.87	yes
2	normal lymph/other	1985	10	1	Tissue	151.7	1.83	47.44	2.89	
3	normal breast	1996	10	3	Plus	146.8	1.73	43.90	2.38	yes
4	normal breast	1996	10	3	Tissue	87.4	1.81	29.79	1.75	
5	normal lymph	2008	10	1	Plus	188.6	1.82	113.66	1.69	yes
6	normal lymph	2008	10	1	Tissue	117.8	1.85	64.51	1.29	
7	normal breast	2000	5	5	Plus	22.4	1.83	nt	nt	
8	normal breast	2000	5	5	Tissue	12.9	1.76	nt	nt	
9	normal breast	2000	10	1	Plus	14.3	1.87	1.62	3.14	
10	normal breast	2000	10	1	Tissue	7.4	1.86	0.92	5.38	
11	normal breast	2006	10	3	Plus	42.7	1.66	10.06	1.65	
12	normal breast	2006	10	3	Tissue	23.2	1.80	11.91	1.99	
13	normal breast	2016	10	3	Plus	91.3	1.74	30.98	0.84	
14	normal breast	2016	10	3	Tissue	36.3	1.85	18.38	0.78	
15	normal lymph	2002	5	5	Plus	340.8	1.79	nt	nt	
16	normal lymph	2002	5	5	Tissue	127.6	1.81	nt	nt	
17	normal lymph	2002	10	1	Plus	247.9	1.78	29.07	3.99	
18	normal lymph	2002	10	1	Tissue	119.1	1.80	14.64	3.18	
19	normal lymph	2016	10	1	Plus	22.0	1.68	7.81	1.13	
20	normal lymph	2016	10	1	Tissue	15.3	1.73	10.46	0.84	

nt = not tested

Average for Plus Kit:	132.7	1.77
Range for Plus Kit:	14.3 - 340.8	1.66 - 1.87
Average for Tissue Kit:	69.9	1.81
Range for Tissue Kit:	7.4 - 151.7	1.73 - 1.86

Figure 1. Sample Standard Curve from Agilent NGS FFPE qPCR analysis

Shawn K. Westaway, PhD
Laboratory Manager
Northwest Research Laboratory
Kaiser Permanente
Center for Health Research
7201 N Interstate Aveue
Portland, OR 97217-5599
503-286-6891 | Phone
503-286-6842 | Fax
503-286-6813 | Lab
shawn.k.westaway@kpchr.org
www.kpchr.org

Table 2. Whole Exome Gene Panel Coverage
For three DNA samples analyzed by whole exome sequencing (sample # 1, 3, 5; see Table 1), we report the depth of coverage for each gene in the experimental panel, as well as the percentage of the gene that had greater than 10X coverage.

Gene	Overall Depth of Coverage:			>10X Depth of Coverage (%):			Mean	SD
	1	3	5	1	3	5		
CYP1A2	231.70	240.20	277.70	100	100	100	100	0
CYP2B6	238.90	248.00	271.10	100	100	100	100	0
CYP2C19	571.90	570.70	534.40	100	100	100	100	0
CYP2C9	581.90	616.20	586.20	100	100	100	100	0
CYP3A5	359.30	377.40	345.30	100	100	100	100	0
UGT2B7	671.30	628.10	571.00	100	100	100	100	0
UGT2B15	477.90	447.30	336.80	100	100	100	100	0
FM01	381.80	393.60	362.70	100	100	100	100	0
SULT2A1	322.70	314.50	315.50	100	100	100	100	0
SULT1E1	312.30	293.90	253.40	100	100	100	100	0
UGT1A4	521.10	523.20	371.70	100	100	100	100	0
UGT1A10	636.20	634.30	640.00	100	100	100	100	0
UGT1A8	642.90	638.00	637.50	100	100	100	100	0
UGT1A3	533.80	560.30	386.30	100	100	100	100	0
SULT1A1	202.10	175.50	204.10	100	100	100	100	0
FM03	442.90	428.00	401.10	100	100	100	100	0
CYP2D6	76.60	120.20	98.60	97.56	100	98.92	98.83	1.22
CYP3A4	367.50	382.30	362.70	100	100	100	100	0
Average Depth All Genes	420.71	421.76	386.45					
Standard Deviation	171.25	166.55	152.02					

Table 3. Whole Exome Gene Panel Variant Calls

A sampling of variants called in genes in the P450 pathway is shown. Only variants with a minor allele frequency (MAF) less than 5% are shown, and headings are described in the text.

Sample #	Chr	Ref. Base	Alt. Base	Genotype	Gene	SNP ID	Function	Protein Change	PolyPhen	SIFT	Minor Allele Freq. Eur (%)	Minor Allele Freq. Local (%)
3; 5	1	G	T	G/T	FM03	75904274	missense	GLY/VAL	0.716	0.18	1.73	1.68
5	1	A	G	A/G	FM01	16864314	missense	ILE/VAL	0.005	1	0.13	0.10
1	2	G	A	G/A	UGT1A8	17863762	missense	CYS/TYR	1	0	2.22	0.74
1	2	A	G	A/G	UGT1A isoforms	17868336	synonymous	THR	unknown	0.07	3.88	2.17
3	4	C	T	C/T	UGT2B15	149314105	synonymous	ALA	unknown	1	1.60	NA
5	7	G	A	G/A	CYP3A5	28383468	missense	HIS/TYR	0.009	0.46	0.95	0.79
5	10	G	C	G/C	CYP2C19	17878459	missense	GLU/ASP	0.001	0.38	3.05	2.76
1; 3; 5	16	C	T	T/T	SULT1A1	1801030	missense	VAL/MET	0.073	0.09	0.21	100.00
1; 3; 5	19	A	G	G/G	SULT2A1	11083907	synonymous	PHE	unknown	1	0.07	100.00