ARUP LABORATORIES

An enterprise of the University of Utah and its Department of Pathology

The University of Utah
Department of Pathology

Next Generation Sequencing of FFPE Tissue

Whole Genome and Kinome Capture

Christine L. Baker[1], Rebecca Margraf[1], Jacob Durtschi[1], Whitney L. Donahue[1], Cecily P. Vaughn[1], Karl Voelkerding, [1, 2] Wade S. Samowitz[1, 2]
[1]ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT
[2]Department of Pathology, University of Utah Health Sciences Center, 50 North Medical Drive, Salt Lake City, UT

Abstract (Updated)

INTRODUCTION
Formalin fixation and paraffin embedding (FFPE) is the standard for processing tumor tissue in pathology departments. Archived tissue blocks provide an important resource for molecular oncology studies, as tissue in this form and some degree of clinical follow-up is available from nearly all tumors. However, the fixation procedures that are necessary for tissue preservation compromise the quality of the DNA, making FFPE a challenging sample for many molecular applications. Clinical evaluation of FFPE tissue for oncogene mutations, such as those in KRAS, has been achieved, but this typically involves evaluation of a small number of nucleotides. Evaluation of an entire genome from FFPE tissue, as with next generation sequencing (NGS), is more challenging especially as compared to NGS on higher quality DNA, such as from peripheral blood leukocytes. We developed a protocol for NGS on routinely processed FFPE.

MATERIALS AND METHODS
We first developed a protocol to prepare an NGS sequencing library from FFPE tissue from a melanoma sample. The sample was microdissected to enrich for tumor cells from a 5 um slice of tissue. DNA was extracted using the Qiagen DNA extraction kit with extended (four days) Proteinase K incubation. DNA was then sheared to approximately 800 base pairs with the Covaris adapted, focused, acoustic technology. The SPRI-TE automated nucleic acid extractor was used for Illumina adaptor ligation. Six cycles of PCR was then performed across the adaptors. The resultant amplicon was gel purified to enrich for a size of 400 base pairs. Kapa Biosystems quantitative PCR was used to assess validated library quantity. The Illumina c-Bot generated clusters on 8 lanes of a flow cell and the Illumina Hi Seq generated 2x100 paired end reads.
Additionally, four tumor samples with mutations in hotspot oncogenes were prepared with the same extraction protocol, sheared to about 200 base pairs, and ligated manually with Illumina® adaptor sequence tags. PCR was performed with primers specific to the sequence tags for 6 cycles. The adapted sequences were then hybridized with biotinylated RNA baits from the Agilent SureSelect Human Kinome Kit® to enrich for 612 kinase and assorted cancer genes . The hybridized samples were then conjugated to strepdavidin T1 Dynabeads® The beads were captured with a magnetic field, the non-target DNA was washed away and the RNA baits removed by a neutralizing buffer. The samples were then PCR'd with four unique index tags for 20 cycles. The resulting prepared libraries were pyrosequenced to determine mutation allele percentage. Using concentration figures determined by Kapa Biosystems® quanitative PCR the samples were pooled into one equi-molar, multiplexed sample. The Illumina c-Bot generated clusters on 1 lanes of a flow cell and the Illumina Hi Seq generated 2x100 paired end reads

RESULTS
The entire genome of the melanoma was sequenced. Over 90% of the reads had Q scores greater than 30, corresponding to 99.9% base call accuracy. Average coverage of the genome was 30 reads per base. A *BRAF* c.1799A>T (V600E) mutation was detected in 17/29 reads (59%). BRAF mutation detection of the original DNA extraction before library preparation showed the same mutation at 57%. The four samples that were enriched for 612 genes, 3.2 Megabases and sequenced had read coverage ranging from several hundred to several thousand reads for the hotspot oncogene mutations. In all 5 samples the expected mutation was detected, and none of the samples had mutations in regions expected to be wild type.

CONCLUSIONS
The protocol described here is a method for next generation sequencing of an entire tumor genome from FFPE. The Q scores and coverage were comparable to sequence from genomic DNA from whole blood or snap frozen tissue. Targeted sequencing using the Agilent SureSelect Human Kinome Kit® provided adequate coverage for somatic mutation detection in FFPE samples.

Introduction

Formalin fixation and paraffin embedding (FFPE) is the standard for processing tumor tissue in pathology departments. Archived tissue blocks provide an important resource for molecular oncology studies, as tissue in this form and some degree of clinical follow-up is available from nearly all tumors. However, the fixation procedures that are necessary for tissue preservation compromise the quality of the DNA, making FFPE a challenging sample for many molecular applications. Clinical evaluation of FFPE tissue for oncogene mutations, such as those in KRAS, has been achieved, but this typically involves evaluation of a small number of nucleotides. Evaluation of an entire genome or targeted subset of the genome, from FFPE tissue, as with next generation sequencing (NGS), is more challenging. Especially as compared to NGS on higher quality DNA, such as from peripheral blood leukocytes or frozen tumor tissue. We developed a protocol for NGS on routinely processed FFPE.

Materials and Methods

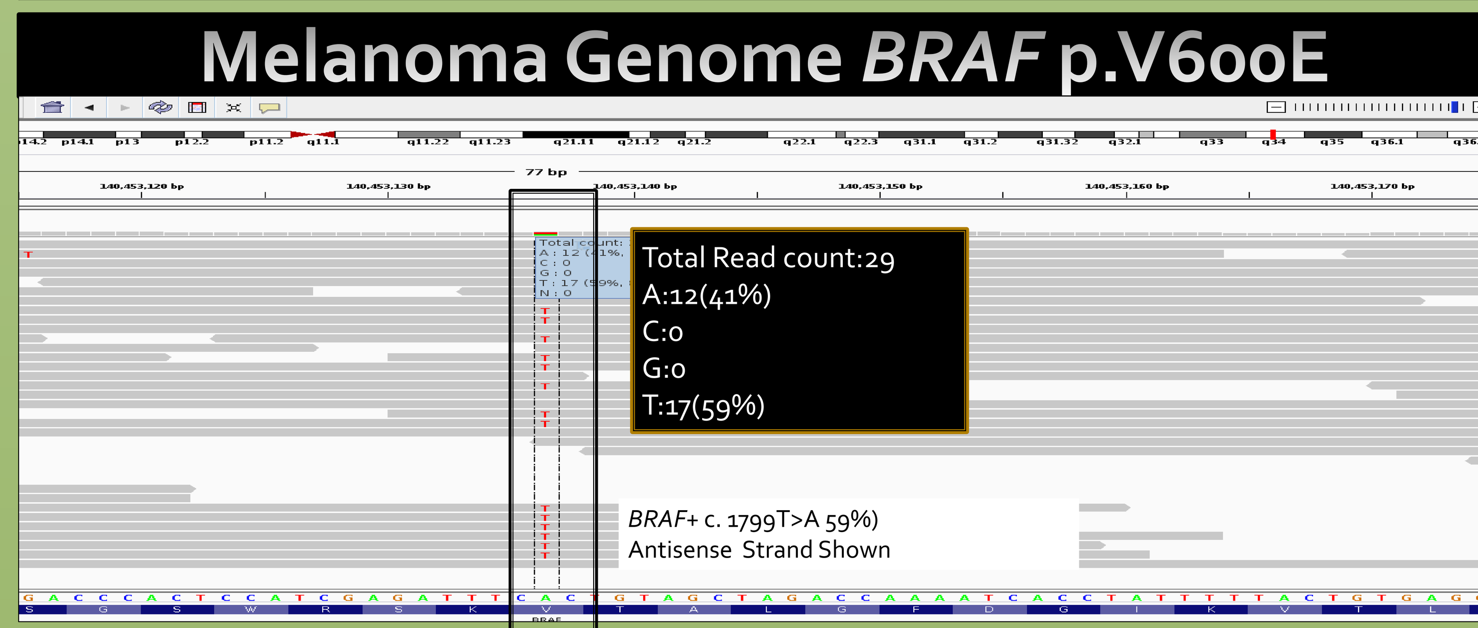
We developed a protocol to prepare an NGS sequencing library from FFPE tissue from a melanoma sample. The sample was microdissected to enrich for tumor cells from a 5 um slice of tissue. DNA was extracted using the Qiagen DNA extraction kit with an extended (four days) Proteinase K incubation. DNA was then sheared to approximately 800 base pairs with the Covaris adapted, focused, acoustic technology. The SPRI-TE automated nucleic acid extractor was used for Illumina adaptor ligation. PCR was done across the adaptors and gel extraction was done. Kapa Biosystems quantitative PCR was used to assess validated library quantity. The Illumina c-Bot generated clusters on 8 lanes of a flow cell and the Illumina Hi Seq generated 2x100 paired end reads.

Additionally, four tumor samples with mutations in hotspot oncogenes were prepared with the same extraction protocol, sheared to about 200 base pairs, ligated manually with Illumina® adaptor sequence tags. PCR was done with primers specific to the sequence tags for 6 cycles. The adapted sequences were then hybridized with biotinylated RNA baits from the Agilent SureSelect Human Kinome Kit® to enrich for 612 kinase and assorted cancer genes . The hybridized samples were then conjugated to strepdavidin T1 Dynabeads® The beads were captured with a magnetic field and the non-target DNA was washed away. The samples were then PCR'd with four unique index tags for 20 cycles. The resulting prepared libraries were pyrosequenced to determine mutation allele percentage. Using concentration figures determined by Kapa Biosystems® quantitative PCR the samples were pooled into one equi-molar, multiplexed sample. The Illumina c-Bot generated clusters on 1 lane of a flow cell and the Illumina Hi Seq generated 2x100 paired end reads.

Results

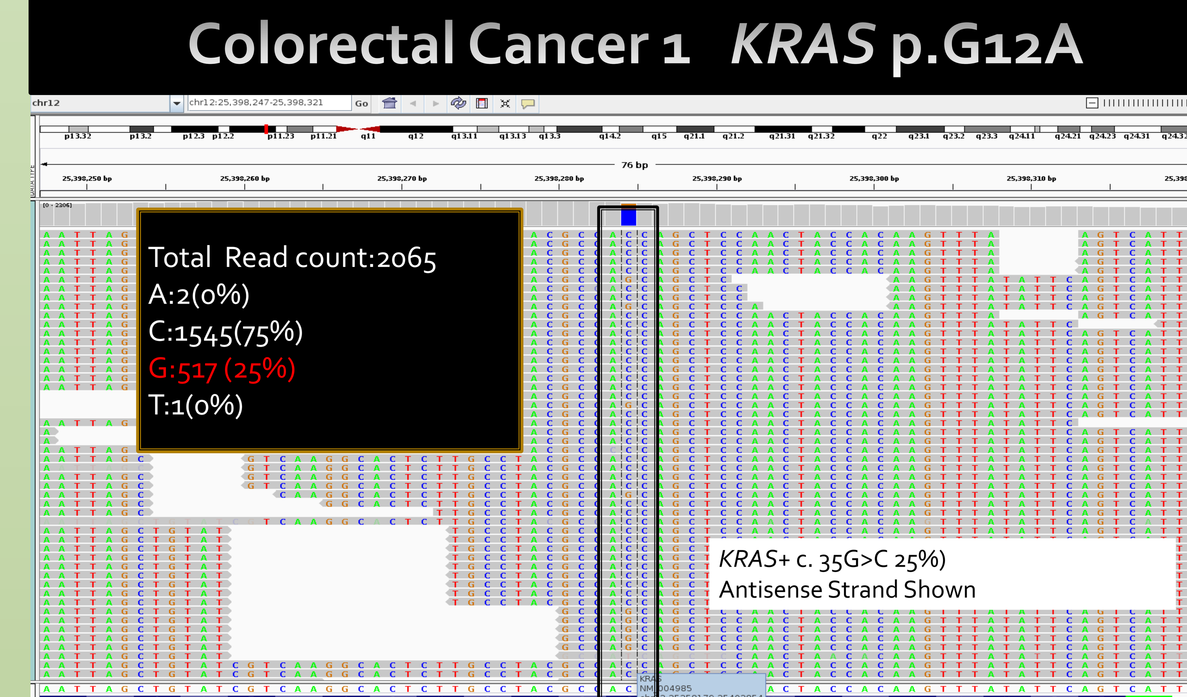
Sample Name (Library Pyrosequencing mutation %)	Read Depth Mutation % by NGS *KRAS*	Read Depth Mutation % by NGS *BRAF*	Read Depth Mutation % by NGS *NRAS*	Read Depth Mutation % by NGS *PIK3CA*	Read Depth Mutation % by NGS *c-KIT*
Melanoma Entire Genome (BRAF + c.1799T>A 68%)	19 WT	29 59%A	15 WT	9 WT	24 WT
Kinome Capture Samples					
Colorectal Cancer 1 (KRAS+ c. 35G>C 34%)	2065 25%C	352 WT	616 WT	1714 WT	1261 WT
Colorectal Cancer 2 (KRAS+ c.35G>A 63%)	3576 65%A	514 WT	531 WT	1799 WT	1279 WT
Melanoma 1 (BRAF + c.1799T>A 48-45%)	2315 WT	482 96%A	722 WT	1514 WT	565 WT
Melanoma 2 (NRAS+ c. 182A>G 38%)	780 WT	386 WT	381 34%	839 WT	508 WT

Melanoma Genome *BRAF* p.V600E

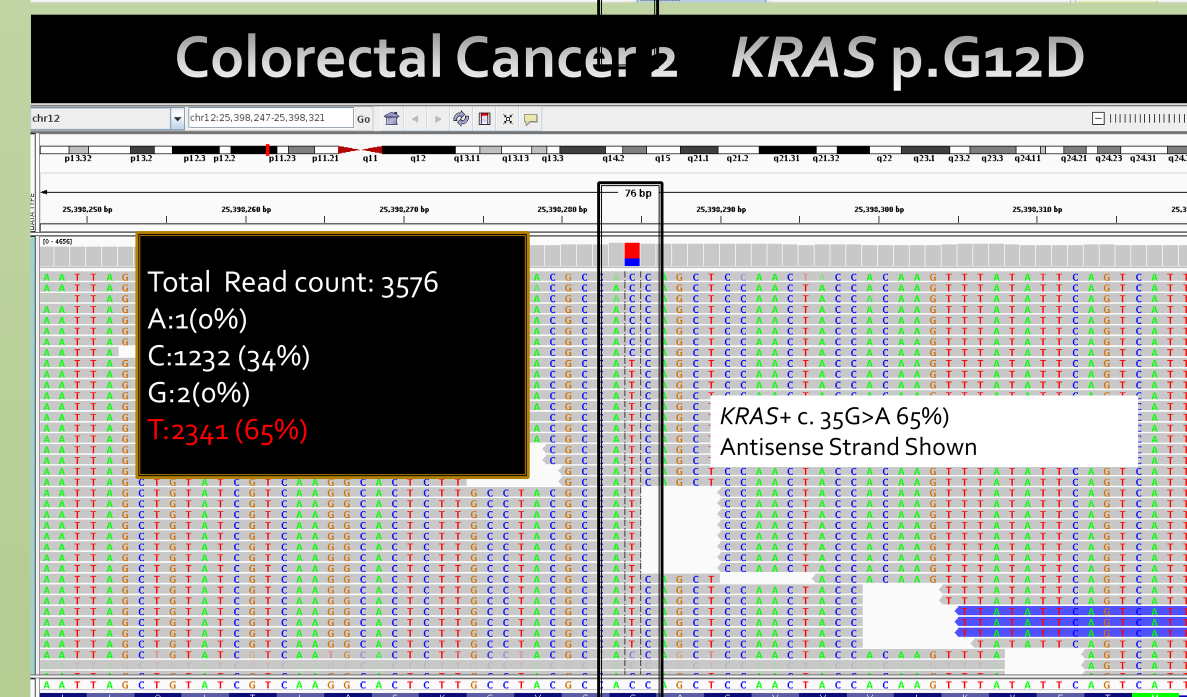


Results Continued; SureSelect Kinome Capture

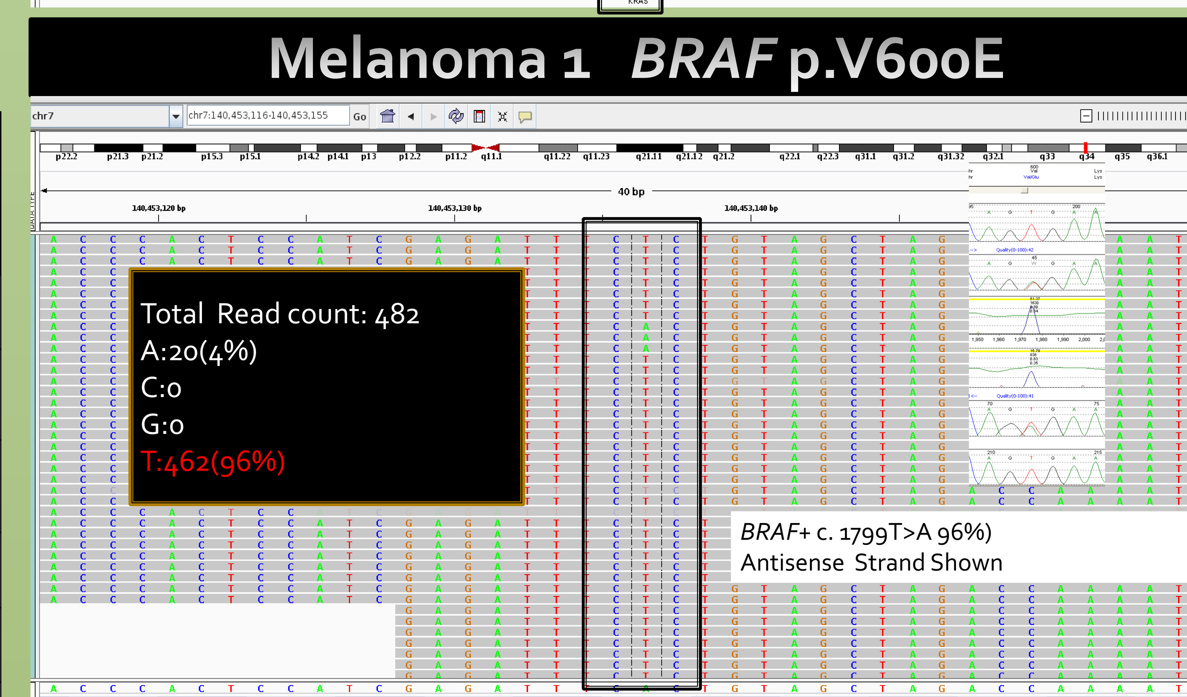
Colorectal Cancer 1 *KRAS* p.G12A



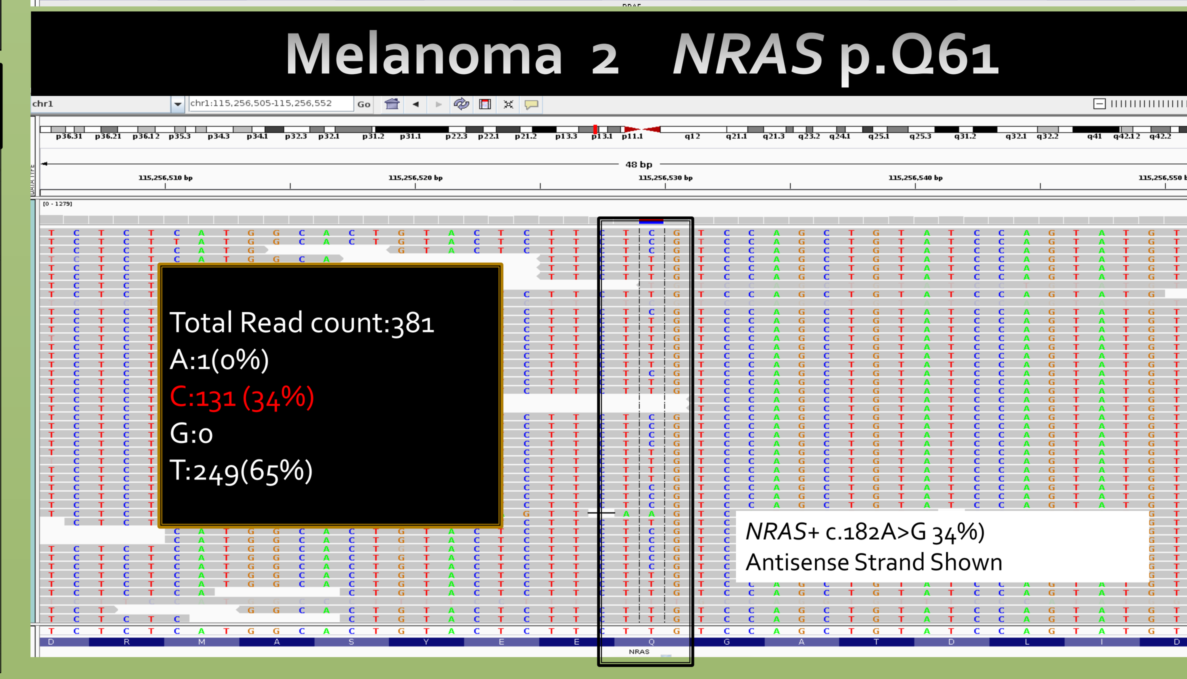
Colorectal Cancer 2 *KRAS* p.G12D



Melanoma 1 *BRAF* p.V600E



Melanoma 2 *NRAS* p.Q61



Conclusions

The protocol described here is a method for next generation sequencing of an entire tumor genome or targeted subset of genes from FFPE. The Q scores and coverage were comparable to sequence from genomic DNA from whole blood or snap frozen tissue for the whole genome and kinome capture samples. The current Illumina chemistry does not give adequate coverage of the whole genome for oncology purposes. Targeted sequencing of the 3.2 Mega bases in the Agilent SureSelect Human Kinome Kit® for 612 kinase and assorted cancer genes when paired with the extraction and Illumina sequencing protocol, provides ample coverage ranging from hundreds to thousands of reads for somatic mutation detection in FFPE samples. At this time the capture and enrichment for genes of interest is necessary to obtain the high read depth necessary for oncology mutation detection.

The BRAF mutated melanoma library was Sanger sequenced because the NGS mutation allele percentage is significantly discordant with the pyrosequencing mutation allele percentage. The Sanger data is in a box on the right side of the NGS data for the BRAF mutated Melanoma. The inflated mutation detection percentage in the NGS data for the BRAF mutated melanoma kinome sample does not appear to be a capture or library preparation artifact, as the pyrosequencing and Sanger sequence results for the library indicate a mutated allele percentage of 45-48%. It is possible that some reads are aligning improperly at the bioinformatics step or that the flow cell chemistry was anomalous for this sample.

References

Schweiger, M. R.; Kerick, M.; Timmermann, B.; Albrecht, M. W.; Borodina, T.; Parkhomchuk, D.; Zatloukal, K.; Lehrach, H., Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. *PLoS One* **2009,** *4* (5), e5548.

Wood, H. M.; Belvedere, O.; Conway, C.; Daly, C.; Chalkley, R.; Bickerdike, M.; McKinley, C.; Egan, P.; Ross, L.; Hayward, B.; Morgan, J.; Davidson, L.; MacLennan, K.; Ong, T. K.; Papagiannopoulos, K.; Cook, I.; Adams, D. J.; Taylor, G. R.; Rabbitts, P., Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res 38* (14), e151.

Feldman, M. Y., Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog Nucleic Acid Res Mol Biol* **1973,** *13*, 1-49.

http://www.chem.agilent.com/Library/flyers/Public/5990-6675en_lo.pdf

Acknowledgments

ARUP Institute for Clinical & Experimental Pathology®

Genomics-Rebecca Margraf, Jacob Durtschi, Karl Voelkerding

Genetics-Whitney Donahue

Anatomic Pathology Oncology-Cecily Vaughn, Noah Welker