Unlocking the transcriptome in formalin-fixed paraffin embedded tissue using mRNA capture sequencing

Formalin fixation and paraffin embedding (FFPE) is the clinical standard for preparing tissue samples for histopathological assessment. Such samples represent a vast repository of tissue material, often with long-term clinical follow-up. With the advent of high-throughput molecular profiling technologies, there is a unique opportunity to screen and comprehensively evaluate biomarkers. Such studies typically require a large sample size and long-term outcome data, both key features of FFPE tissue archives. Unfortunately, the process of tissue fixation induces chemical changes and fragmentation in both DNA and RNA, making subsequent analysis unreliable. Because of this highly fragmented state of RNA in FFPE tissue, most gene expression studies have instead focused on intact RNA from fresh frozen (FF) material.

To make use of the rich resource of FFPE specimens, Biogazelle previously developed a sensitive and accurate method for targeted gene expression analysis on FFPE tissue using a dedicated RT-qPCR workflow, compatible with fragmented and low input RNA samples. To accommodate unbiased mRNA gene expression profiling, we have now successfully implemented a workflow for mRNA capture sequencing on FFPE tissue using the TruSeq RNA Access Library Preparation Kit (Illumina). Using proven TruSeq stranded RNA library preparation chemistry combined with efficient sequence-specific exon capture, the TruSeq RNA Access Library Preparation Kit generates RNA sequencing libraries from degraded samples that focus on the protein coding regions of the transcriptome. Isolating these high-value content regions maximizes discovery power, while requiring only a fraction of the read depth of total RNA sequencing.

This tech note describes the technical assessment of the workflow implementation and zooms in on differential gene expression in colon cancer compared to normal colon FFPE tissue.

Performing mRNA capture sequencing

To assess the performance of the workflow, we performed mRNA capture sequencing on RNA isolated from 4 colon cancer FFPE samples and 4 matching normal colon FFPE samples.

The isolated RNA was of particularly low quality with DV200 values between 8 and 26 (Table 1 and Fig. 1). Libraries for mRNA capture sequencing were prepared starting from 100 ng of total RNA using the TruSeq RNA Access Library Preparation Kit (Illumina) according to the manufacturer instructions. Paired-end sequencing was performed on a NextSeq 500 instrument (Illumina) with a read length of 75 base pairs to a sequencing depth of 50 million read pairs. Read mapping to the reference genome (Ensembl 78) was performed by TopHat, genes were quantified with HTSeq and raw read counts were normalized using the DESeq size factor.

Messenger RNA capture sequencing using the TruSeq RNA Access Library Preparation Kit (Illumina) results in stranded sequencing data with high exonic coverage (Fig. 2). Stranded sequencing data allows for precise measurement of strand orientation, enhancing transcript annotation, and increasing alignment efficiency.

sample	RIN	DV200 *
colon cancer 1	2.2	26
normal colon 1	2.3	18
colon cancer 2	2.5	15
normal colon 2	2.6	11
colon cancer 3	2.3	15
normal colon 3	2.5	20
colon cancer 4	2.4	23
normal colon 4	2.6	8

Table 1: RIN and DV200 values from FFPE samples. *"DV200" is the percentage of RNA fragments > 200 nucleotides.

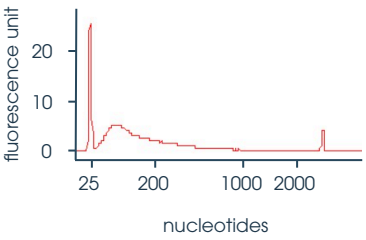


Figure 1: RNA quality from FFPE sample. RNA isolated from sample "colon cancer 1" examined using the Agilent Bioanalyzer.

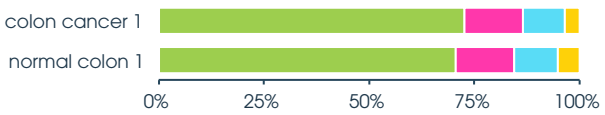


Figure 2: Percentage of bases aligned to genomic regions. More than 85% of the mRNA capture sequencing data aligns to mRNA transcripts (coding and UTR regions). Green: exon; pink: UTR; blue: intron; yellow: intergenic.

An interesting feature of RNA sequencing is the ability to identify alternative splicing events. Although the exonic capture probes in the TruSeq RNA Access Library Preparation Kit are not designed to cover splice junctions, a high fraction of reads map to splice junction sequences (Fig. 3). As a consequence, mRNA capture sequencing using the TruSeq RNA Access Library Preparation Kit is highly efficient for detecting alternative splicing events.

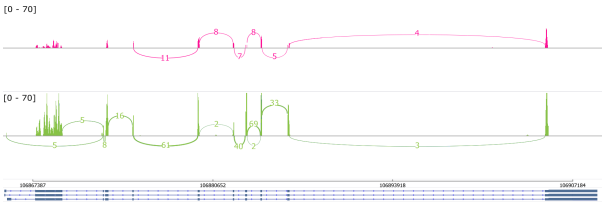


Figure 3: Example of reduced expression and splice junction reads in the PIK3CG gene in colon cancer (pink) compared to matched normal colon FFPE tissue (green). Sashimi plots quantitatively visualize splice junctions from alignment data, along side genomic coordinates and annotation.

Highly sensitive and reproducible workflow

To assess the reproducibility of our workflow, a technical replicate of one FFPE sample was included in the entire workflow. We observed excellent correlation between the normalized gene expression counts from technical replicates of a low quality normal colon FFPE sample (Fig. 4A). At sequencing depth of 20 million subsampled paired reads, we detect on average 14,241 mRNA genes (with a minimum read count of 10) per sample. Next, we analyzed the gene expression data of the 4 matched colon tumor-normal pairs. We found 2738 genes to be differentially expressed between colon cancer and colon control samples (FDR < 5%; Fig. 4B). Functional annotation of the differentially expressed genes using Gene Set Enrichment Analysis (GSEA) shows that pathways involved in colon cancer and cancer in general (i.e. cell cycle, proliferation, metastasis, resistance to chemotherapy, and lymphocyte infiltration) are associated with upregulated genes (FDR < 5%). No pathways were found to be associated with downregulated genes (FDR < 5%).

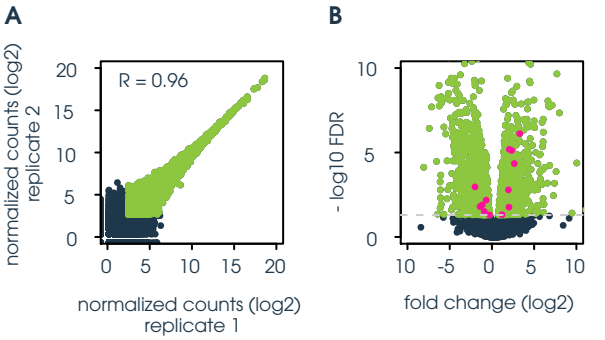


Figure 4: Highly sensitive and reproducible workflow. (A) Normalized gene expression counts from technical replicates of a low quality colon control FFPE sample. Gray dots: genes below the detection threshold (10 reads) in either sample replicate. (B) Volcano plot; green dots: FDR < 5%; pink dots: differential genes belonging to KEGG colon cancer pathway.

mRNA capture seq on FFPE tissue versus polyA+ seq on fresh frozen tissue

Given that mRNA capture sequencing of the matched colon tumor-normal FFPE pairs resulted in a high number of differentially expressed genes, we next evaluated how these results compare to RNA sequencing data from fresh frozen (FF) colon cancer samples. To this end, we made use of publically available RNA sequencing data of four FF colon tumor-normal sample pairs from The Cancer Genome Atlas (TCGA). Normalized polyA+ RNA sequencing data were downloaded from TCGA data portal for selected tumor-normal pairs (Table 2).

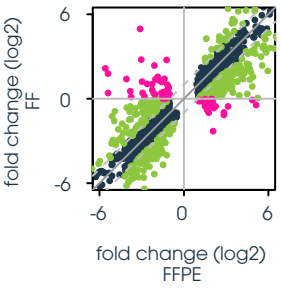
Table 2: Overview of Biogazelle mRNA capture sequencing data and TCGA polyA+ sequencing data of matched colon tumor-normal pairs. The four TCGA sample pairs were selected from a larger cohort of matching tumor-normal pairs (n = 26) to match our Biogazelle sample pairs according to TNM tumor staging: size and local invasiveness of the tumor (T3 or T4), spreading to the lymph nodes (N1 or N2) and extent of metastasis (M1). FF: fresh frozen; FFPE: formalin-fixed paraffin embedded; PE: paired-end

	Biogazelle data	TCGA data
sample type	FFPE	FF
samples	4 pairs	4 pairs
sequencing instrument	NextSeq	HiSeq
library prep	mRNA capture	polyA+ selection
sequencing depth	~50 million	~50 million
reads	PE75	PE50
pipeline	TopHat + HTSeq	MapSplice + RSEM
differential expression	DESeq2	DESeq2

Differential gene expression between FF colon cancer tumor and normal samples identifies 5200 differentially expressed genes (FDR < 5%), of which 1753 genes (33.7%) were also detected as differentially expressed in the FFPE data. Subsequently we evaluated how concordant are the log fold changes between FFPE and FF data for differentially expressed genes in the FFPE data. Two aspects are evaluated when comparing fold changes obtained in both sample types: log fold change concordance and direction concordance. The former is measured as the absolute difference between log2 fold change obtained for each gene from both datasets; the absolute difference should be lower than 1 (log2 scale) for concordance. Directional concordance is measured as the sign of the fold change obtained from both datasets. A high fraction (72.5 %) of genes differentially expressed in the FFPE data (FDR < 5 %) shows concordance in log fold change and direction compared to genes differentially expressed in the FF data (Fig. 5). In addition, 23.7 % of genes differentially expressed in the FFPE data shows concordance in direction. Only a small minority (3.8 %) of genes differentially expressed in the FFPE data shows no concordance in fold change or direction compared to genes differentially expressed in the FF data. This high concordance in differentially expressed genes is reflected in the pathways that are deregulated in both datasets (data not shown).

Despite the many differences in the FFPE and FF datasets (Table 2), we observe a very good correspondence between both datasets. Thus, gene expression changes measured using mRNA capture sequencing on FFPE samples nicely represent those measured using polyA+ RNA sequencing on FF samples.

Figure 6: High concordance of fold changes upon differential gene expression analysis in fresh frozen and FFPE tissue. Dark blue: FC and directional concordance; green: directional concordance; pink: non-concordant.



Conclusion

To enable mRNA biomarker discovery in FFPE tissue, Biogazelle has implemented a workflow for mRNA capture sequencing. Using a highly reproducible and sensitive method, we detect known and potentially new mRNA biomarkers for colon cancer based on the study of FFPE tissue. This optimized workflow enables researchers to apply the power of next-generation sequencing technology to mRNA expression studies on RNA isolated from FFPE samples. Beyond gene expression analysis, RNA capture sequencing can also be used for discovery applications such as identifying alternative splicing events, fusion genes and expressed mutations. Taken together, mRNA capture sequencing opens new and powerful ways of analyzing mRNA from FFPE samples. Reach for your FFPE archives as patients from the past can provide solutions for the future.

Learn more

To learn more about the value of analyzing mRNA in archived formalin-fixed paraffin embedded tissue, visit Biogazelle's Knowledge Center:

https://www.biogazelle.com/knowledge-center

Your mRNA capture sequencing project @ Biogazelle – 8 steps to success

1. upfront discussion with our PhD-level project managers
2. RNA extraction (optional)
3. library preparation
4. library quality control
5. sequencing
6. data processing and quality control
7. reporting
8. discussion of results with our PhD-level project managers

biogazelle

Biogazelle, NV
Technologiepark 3
B-9052 Zwijnaarde, Belgium

www.biogazelle.com
info@biogazelle.com
+32-9-245 96 44