DANA-FARBER/BRIGHAM AND WOMEN'S CANCER CENTER

Automated TruSeq RNA Sample Preparation from FFPE tissue specimens utilizing the Biomek FX^P Liquid Handler

^^Michaela Bowden Ph.D., ^^Ashley Rossi, **Bee Na Lee Ph.D., **Sarah Simons ^^Svitlana Tyekucheva, ^^Ed Stack Ph.D., ^^Massimo Loda MD.,
^^Dana Farber Cancer Institute and Brigham, Women's Hospital and **Beckman Coulter Life Sciences

Overview

This poster describes a TruSeq RNA library preparation for FFPE samples using the Biomek automation platform, capable of constructing 96 libraries in 7 hr. The Biomek platform provides a solution for high throughput library construction, producing sensitive and reproducible sequencing data, which facilitates biomarker discovery in archival FFPE tissues.

Introduction

Next Generation Sequencing (NGS) technology facilitates high throughput, high speed and cost-effective sequencing of DNA and / or RNA. RNA library construction from archival FFPE tissue specimens can be automated utilizing the Beckman Coulter Biomek FX^P Liquid Handler. RNA is extracted using the Beckman Coulter Agencourt FormaPure kit and the library preparation method is based on the Illumina TruSeq RNA sample preparation protocol.

The Biomek TruSeq RNA method comprises three parts:

1. mRNA purification, fragmentation and cDNA synthesis
2. cDNA library construction (end repair, A-tailing and adaptor ligation)
3. PCR amplification and product purification

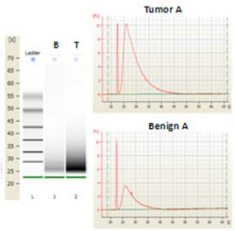
The data presented here benchmarks the novel automated approach using the Biomek FX^P Liquid Handler against a standard manual preparation. Archival FFPE tissue specimens have lower RNA yield and are highly degraded, which represents a major challenge in developing robust methodologies for sequencing. The findings of our study highlighted that libraries prepped on the Biomek automated platform were more reliable, providing end-users with a faster, more streamlined approach to generate libraries from precious clinical samples in a faster turnaround time.

Methods and Results

RNA preparation

Three Tumor/Normal pairs of archival FFPE prostate tissue specimens were selected and scored. Total RNA was extracted according to the protocol of the Agencourt FormaPure kits (Beckman Coulter Life Sciences, PN# A33343). Each sample was eluted in 55 µL dH_2O., RNA quality was determined on the Bioanalyzer 2100 system. Figure 1 shows the typical degraded profile of RNA isolated from archival FFPE tissue. The RN # was ≤ 2.4, which indicates the high level of degradation associated with these tissues.

Figure 1 Representative RNA QC plots with associated elution bands on the left for Tumor / Benign pair, A



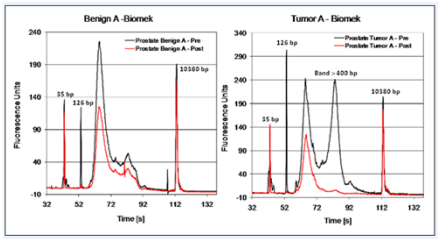
RNA concentration was determined by RiboGreen Assay. Concentrations ranged from 24.4 – 61.1 ng/µl, of which 500 ng of total RNA was utilized for each library as the input amount.

cDNA library quality clean up

Initial library quality showed an average peak size for the amplified cDNA library at approximately 260bp, however, there was a 120bp non-specific adaptor amplified peak also present in all 12 samples. These represent adaptor-dimer contamination, which can lead to junk reads when sequencing. In some cases, an additional fragment greater than 400bp was also detected, which could represent single-stranded library products that have self-annealed.

In order to improve the library quality, a two-step size selection cleanup protocol was implemented. In the first cleanup step, a 0.7X ratio of AMPure XP bead solution (Beckman Coulter Life Sciences, PN# A63881) was added to each sample. Under this condition, only larger fragments greater than 500bp DNA were bound to the beads. The bead bound DNA was discarded and the supernatant treated with a 1.1X volume of AMPure beads to rebind the DNA fragments between 150bp-400bp. Figure 2 shows the overlaid QC plots of the pre- and post-cleanup libraries for tumor / benign pair A with a side-by-side comparison of the Biomek and manual libraries.

Figure 2 Example of QC plots of Pre- and Post-Cleanup Library Preparations for Tumor / Benign pair, A



Single End Sequencing Performance

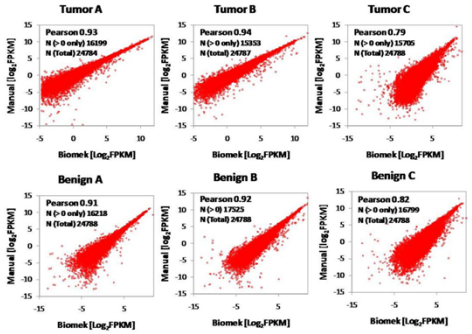
Sequencing was performed on the Illumina HiSeq platform. 50bp single reads were mapped using Tophat and transcript abundance in FPKM units (Fragments per Kilobase of mRNA per 10^6 reads) calculated using Cufflinks. Multiplexed sequencing was performed in such that each lane contained 3 samples. RNASeq library construction / sequencing were successful in 12 / 12 samples with aligned reads ranging from 21 % - 72 % relative to the total # reads, as shown in Table 1.

Table 1 Summary of basic sequencing performance metrics

Donor ID	A				B				C			
Tissue	Tumor	Benign	Tumor	Benign	Tumor	Benign	Tumor	Benign	Tumor	Benign	Tumor	Benign
Library Prep	Biomek		Manual		Biomek		Manual		Biomek		Manual	
Flowcell #	1	1	3	3	2	2	4	4	1	2	3	4
Adaptor ID	AR012	AR006	AR012	AR006	AR006	AR012	AR006	AR012	AR005	AR019	AR005	AR019
# Total Reads	17687022	29046556	48645830	41657865	59938621	49836308	50991836	46441369	19964567	41801701	57399460	38203892
% Aligned	58.55	59.61	50.94	65.81	47.60	57.18	62.94	70.35	21.97	71.88	42.79	66.03
% Intergenic	15.22	15.15	9.13	13.57	14.94	14.67	11.12	12.76	15.01	13.58	9.06	10.22
% Intronic	10.11	12.81	14.45	13.74	10.25	11.20	14.31	12.60	10.45	9.61	13.86	12.66
% Exon	68.46	66.09	24.31	51.02	68.98	71.35	37.92	58.72	62.69	73.32	23.07	44.70
% rRNA	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02

Reproducibility was measured by calculating the Pearson Correlation from transcript abundance values, presented as Log_2[FPKM] values. TruSeq libraries prepped by the Biomek automated platform and manually on the bench, exhibited good concordance, as evidenced in the scatterplots depicted in Figure 3.

Figure 3 Scatterplots comparing method preparation for each individual sample. Log_2 [FPKM] values were utilized to make the correlation.



The total # of transcripts is also displayed in Figure 3. In particular the poor library quality of tumor / normal pair, C, is verified in the corresponding Pearson correlation coefficients.

A wide ranging panel of endogenous controls [1] were found to have good concordance between the manual and Biomek sequenced libraries, with a Pearson coefficient of 0.99, calculated. The insets are plots of the average Log_2[FPKM] values of all samples for each library preparation method, which highlights the stability of these endogenous controls across all sample types, irrespective of preparatory method or tissue morphology.

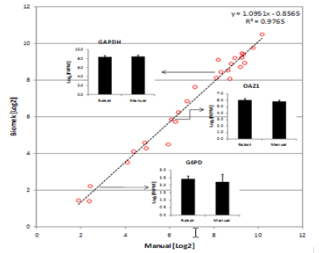


Figure 4 Scatterplots of panel of endogenous control as measured in Biomek and manual sequenced libraries (Insets: Log_2[FPKM] values plotted for low, medium and higher abundance endogenous controls, G6PD, OAZ1 and GAPDH, respectively.

Conclusion

This study shows that RNASeq library preparation carried out on the automated high throughput Biomek platform, results in sensitive and reproducible sequencing data, which facilitates biomarker discovery in archival FFPE tissue. Going forward it is envisaged that new coding and non-coding transcripts, as well as gene signaling networks that strongly associate with prostate cancer progression will be identified. The Biomek FX^P automated platform offers a viable high throughput alternative to traditional manual bench preparation of RNA libraries for sequencing.

Acknowledgement

We thank Beckman Coulter Life Sciences for technical and Biomek supports.

Reference

[1] De Jonge, HJM et al., Evidence Based Selection og Housekeeping Genes, *PLoS One*, **2007**, 2(8), e898.