Transcriptome Analysis in Fixed Tumor Biopsy Samples

March 2022

Frontage

Table of Contents

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

Introduction

Transcriptomic profiling, also known as RNA sequencing (RNA-seq), is a powerful technology to assess gene(Melé et al., 2015; Sultan et al., 2008), microRNA (Riccardi et al., 2016), and non-coding RNA (Clark et al., 2015; Esposti et al., 2016) expression profiles in disease (Costa et al., 2013; Lin et al., 2016) and holds promise for clinical diagnostic use (Byron et al., 2016) and biomarker discovery (Liang et al., 2015). Formalin-fixed paraffin-embedded (FFPE) tissues are a widely available source of clinical tissue samples because FFPE is the preferred preservation method for pathological diagnostic and archival purposes. Profiling gene expression patterns in FFPE tissues using RNA-seq is challenging because the fixation process used in preparing FFPE tissues crosslinks and chemically modifies RNA (Masuda et al., 1999) and tissue processing can fragment the RNA (von Ahlfen et al., 2007). Furthermore, the yield of RNA isolated from FFPE tissue is often low. New technologies to prepare libraries from RNA isolated from FFPE have recently emerged and enable facile processing of FFPE tissue for mRNA sequencing.

In this whitepaper, we present data comparing three commercially-available library preparation kits—Illumina's TruSeq RNAExome kit, Takara's SMARTer Stranded Total RNA-Seq Kit v2-Pico Input Mammalian, and NuGen's Ovation Human FFPE RNA-Seq Multiplex System 1-8—for suitability in gene expression profiling using FFPE RNA and discuss the advantages and limitations of each kit.

Preserving tissue samples using formalin fixation and paraffin embedding causes chemical modifications to the RNA molecules that inhibit subsequent molecular biology assays, including RNA-seq. Formalin fixation covalently attaches mono-methylol groups to the amino groups on RNA bases, primarily adenine (Masuda et al., 1999), and also modifies or attenuates poly (A) tails (Klopfleisch et al., 2011). The mono-methylol group additions could be partially reversed by heating (Masuda et al., 1999), a fact leveraged in current FFPE RNA isolation methods. Formalin also causes methylene bridges to crosslink nucleic acids to proteins and other biomolecules (von Ahlfen et al., 2007; Masuda et al., 1999). Moreover, the time between surgical removal and fixation, the duration of fixation, the temperature of FFPE storage, and the length of FFPE storage are all critical factors that contribute to the integrity (or degradation) of RNA in FFPE samples (von Ahlfen et al., 2007).

RNA isolated from FFPE sections has long been used for gene expression measurements using PCR (Stanta and Schneider, 1991). However, the quality of FFPE RNA fragments was generally considered too poor, and the size distribution (100-200 nt) too short, to perform RNA-seq for mRNA expression profiling. This was due in part to the requirement to deplete ribosomal RNA (rRNA), which accounts for a large fraction of total RNA. The most widespread method, initially, to deplete rRNA relied on enriching poly(A) RNA using oligo (dT) primers, which were inefficient at capturing partially degraded RNA. Alternative methods to deplete rRNA that did not rely on oligo (dT) selection were subsequently devised, including SDRNA; (Morlan et al., 2012), Ribo-Zero (Huang et al., 2011), duplex-specific nuclease degradation (Yi et al., 2011; Zhulidov et al., 2004), Smart-Seq (Ramsköld et al., 2012), and NuGen Ovation. These methods stimulated a wave of publications describing RNA-sequencing from FFPE tissue (Adiconis et al., 2013; Hedegaard et al., 2014; Norton et al., 2013; Sinicropi et al., 2012; Zhao et al., 2014). Importantly, these findings revealed a strong correlation between expression levels measured in the FFPE tissue with expression in match fresh-frozen tissue (Hedegaard et al., 2014; Hester et al., 2016; Norton et al., 2013; Webster et al., 2015; Zhao et al., 2014). Subsequent studies largely corroborated these results (Bossel Ben-Moshe et al., 2018; Li et al., 2018), although generating RNA-seq libraries from FFPE samples remains challenging (Esteve-Codina et al., 2017; Kresse et al., 2018) and library quality seems to diminish with the length of time samples have been archived (Jovanović et al., 2017).

The TruSeq RNA Exome kit works by first generating stranded RNA-seq libraries and then hybridizing the libraries to biotinylated probes targeting exonic regions. The probes and the hybridized libraries are then captured with streptavidin beads to enriching for libraries covering coding RNA regions. Two enrichment steps are used followed by amplification of the captured libraries. The SMARTer Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian first generates stranded RNA-seq libraries using Takara's SMART (Switching Mechanism At 5' end of RNA Template) cDNA synthesis technology and then depletes ribosomal cDNA libraries using R-Probes, which target mammalian ribosomal RNA and human mitochondrial rRNA sequences, coupled with its proprietary ZapR technology, which cleaves the rRNA libraries. Similar to the TruSeq RNA Exome and SMARTer Stranded Total RNA-Seq Kit v2, the Ovation Human FFPE RNA-Seq System also first creates a stranded RNA-seq library. Then, NuGEN's Insert-Dependent Adaptor Cleavage (InDA-C) technology is utilized to enzymatically deplete ribosomal rRNA transcripts.

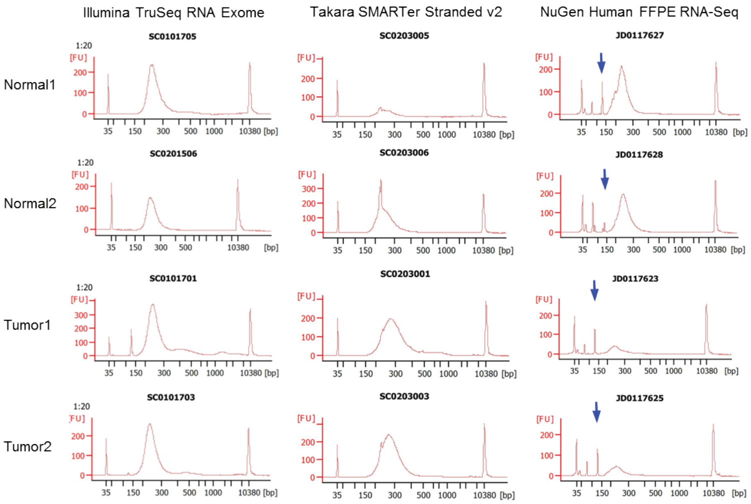
Results

Total RNA was isolated from four tumor FFPE samples and four normal tissue FFPE samples. The total RNA showed varying levels of nucleic acid composition and size distributions (data not shown). To evaluate commercially available methods for generating RNA-seq libraries from FFPE tissues, sequencing libraries were prepared from the total RNA using three product suites: Illumina's TruSeq RNA Exome, NuGen's Ovation Human FFPE RNA-Seq, and Takara's SMARTer Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian.

Library Quality and Size Distribution

Sequencable libraries were generated from total RNA isolated from FFPE tissue using both the TruSeq RNA Exome and SMARTer Stranded v2 kits (Figure 1). By contrast, libraries prepared using the NuGen Ovation FFPE RNA-Seq system had substantial adapter dimers present (Figure 1). While additional cleanup using the Agencourt beads reduced the amount of adapter dimers, the library yields following bead cleanup was insufficient to continue with sequencing. This protocol was therefore excluded from further analysis.

FIGURE 1. LIBRARY QUALITY AND SIZE DISTRIBUTIONS



Libraries were run on an Agilent Bioanalyzer 2100 using a high sensitivity DNA chip and the electropherograms are shown. The NuGenHuman FFPE RNA-Seq libraries showed substantial primer dimer peaks. Arrows indicate the peaks containing adapter dimers in the libraries prepared using the NuGen Human FFPE RNA-Seq kit.

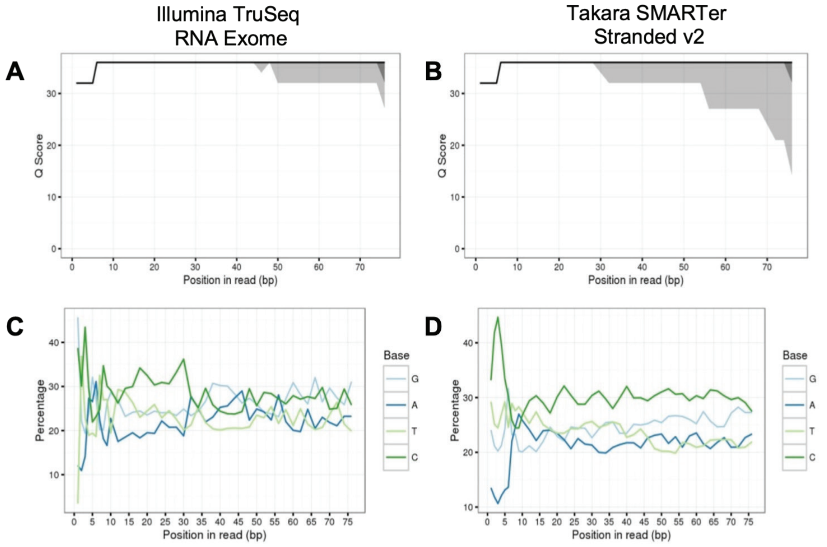
TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

Both the TruSeq RNA Exome and SMARTer Stranded v2 library preparation methods produced high-quality sequencing reads with some degree of nucleotide composition bias, especially at the start of the reads (Figure 2).

FIGURE 2. LIBRARY QUALITY

The distributions of quality scores are plotted at each position in the read (A-B). The line indicates the median quality score. The dark gray region extends from the lower to the upper quartile, and the light gray region extends from the 10th to the 90th percentile. The nucleotide percentages are plotted at each position in the read (C-D). An example R1 Illumina TruSeq RNA Exome library is shown on the left and an example Takara SMARTer R1 library is shown on the right.



Ribosomal RNA Content

To assess the percentage of ribosomal RNA present in libraries from each method, one million untrimmed reads from each sample were aligned to human 45S and 5S ribosomal sequences and the overall alignment percentage calculated. Takara SMARTer libraries had substantially higher ribosomal RNA content (27%) compared to Illumina TruSeq RNA Exome (1.3%; p = 0.0003, paired two-tailed T-test).

Transcriptome Mapping

Sequence reads were aligned to the hg38 human genome using HISAT2 and the number of reads mapping to annotated Ensembl version 83 human genes was counted using featureCounts. Illumina TruSeq RNA Exome libraries had more reads mapped to the hg38 reference genome (90%) vs. SMARTer Stranded v2 libraries (71%; p = 7e-4; paired two-tailed T-test; Figure 3). Moreover, Illumina TruSeq RNA Exome had more mapped reads assigned to annotated Ensembl v.83 genes (59% vs. 16%; p=1.1e-7 paired two-tailed T-test; Figure 3). In normal FFPE samples, 63% of the bases in mapped reads from Illumina TruSeq RNA Exome libraries overlapped coding regions whereas only 8% of bases in mapped reads from Takara SMARTer Stranded v2 libraries overlapped coding regions (Table 1). A similar disparity was observed in tumor samples, in which 69% of bases from Illumina TruSeq RNA Exome libraries overlapped coding regions compared to just 9% of bases from Takara SMARTer Stranded v2 libraries (Table 1). These data demonstrate that the Illumina coding RNA enrichment protocol works well.

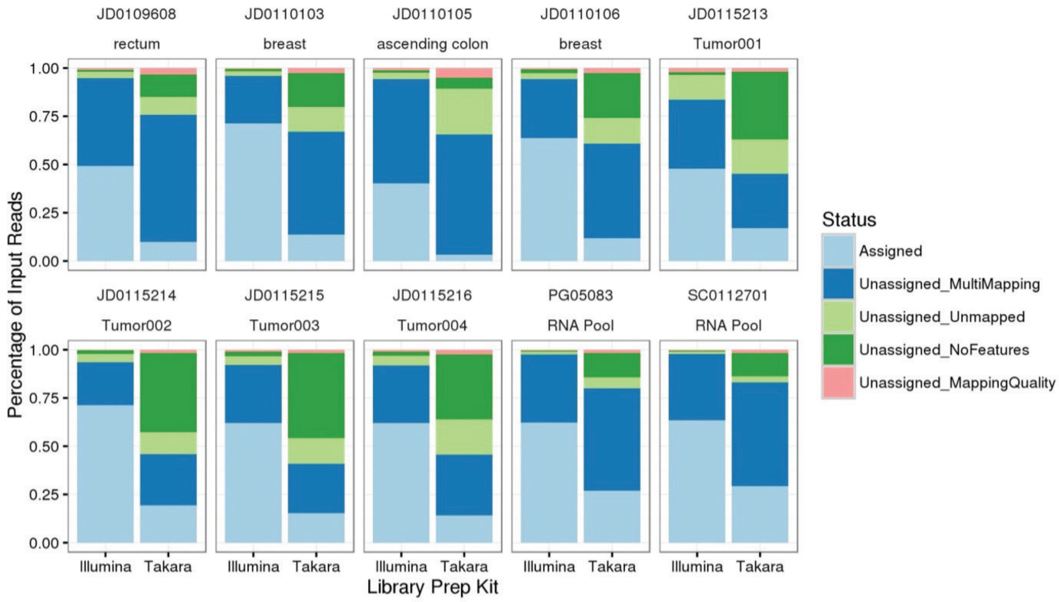
TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

TABLE 1. AVERAGE MAPPED BASE COMPOSITION FOR DIFFERENT LIBRARY METHODS AND SAMPLE TYPES

Library Prep Method	Sample Type	Ribosomal	Coding	UTR	Intronic	Intergenic
Illumina	NAT	0%	63%	25%	4%	7%
Takara	NAT	7%	8%	43%	31%	11%
Illumina	Tumor	0%	69%	19%	5%	7%
Takara	Tumor	2%	9%	19%	58%	12%

FIGURE 3. MAPPED READ SUMMARIZATION



The proportion of reads assigned to Ensembl v83 genes or unassigned for various reasons are shown for each library. Libraries from the same input total RNA are grouped together and the Total RNA ID and tissue type are listed above the group.

Assigned – the read mapped to an annotated gene;
Unassigned_MultiMapping – the read mapped to multiple locations in the genome and was therefore not counted;
Unassigned_Unmapped – the read did not align to the hg38 genome;
Unassigned_NoFeatures – the read aligned to the hg38 genome, but did not overlap an annotated gene region;
Unassigned_MappingQuality – the read aligned to the hg38 genome with a mapping quality less than 20 and was therefore not counted

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

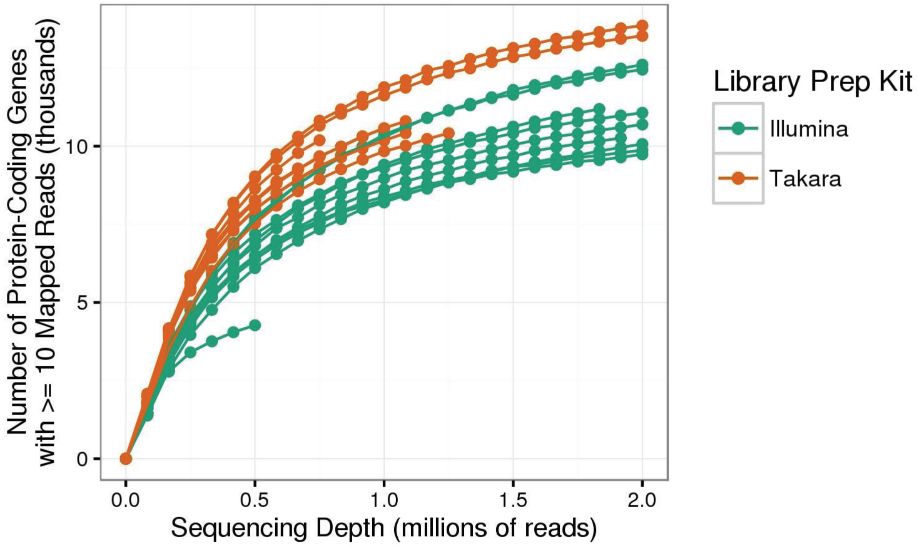
To assess the types of genes represented in each of the libraries, the number of reads mapping to Ensembl version 83 genes was totaled by the annotated biotype of each gene. The top 2 biotypes for Illumina libraries were protein_coding (76%) and snoRNA (9%), while the top 2 biotypes in Takara libraries were protein_coding (67%) and rRNA (7%) (not shown).

To estimate the coverage of protein-coding genes in each library, saturation curves for each library were generated. Random subsamples of the counted sequencing reads, consisting of between 1 read and 2 million reads, were generated and the number of genes with at least 10 mapped reads in each subsample was counted. The input data were gene counts for only the 20,253 nuclear protein-coding genes. At a given sequencing depth, libraries prepared using the Takara SMARTer Stranded v2 kit had more protein-coding genes with at least 10 mapped reads (Figure 4).

Differential Gene Expression

DESeq2 was used to evaluate the impact of library preparation method on the ability to detect genes differentially expressed between tumor and normal FFPE tissue. In Illumina libraries, 1,515 genes were found to be differentially expressed between tumor and normal samples with a false discovery rate (FDR) of 1% (Figure 5A). In Takara libraries, 434 genes were differentially expressed using the same FDR (Figure 5B). Of the genes whose expression differed between tumor and normal samples in either library type, 279 were found to be differentially expressed in both library types (Figure 6).

FIGURE 4. SATURATION CURVES FOR DIFFERENT LIBRARY PREPARATION METHODS



Random subsamples of the counted sequencing reads, consisting of between 1 read and 2 million reads, were generated for each library and the number of genes with at least 10 mapped reads in each subsample were counted.

FIGURE 5. AN OVERALL SUMMARY OF THE FOLD CHANGES AND FDRS IN BOTH LIBRARY TYPES

Volcano plots show the –log10(FDR) versus log2 fold change in expression between tumor and normal samples for Illumina TruSeq RNA Exome libraries (A) and Takara SMARTer Stranded v2 libraries (B). The gene symbols for genes with an FDR < 0.00001 are shown.

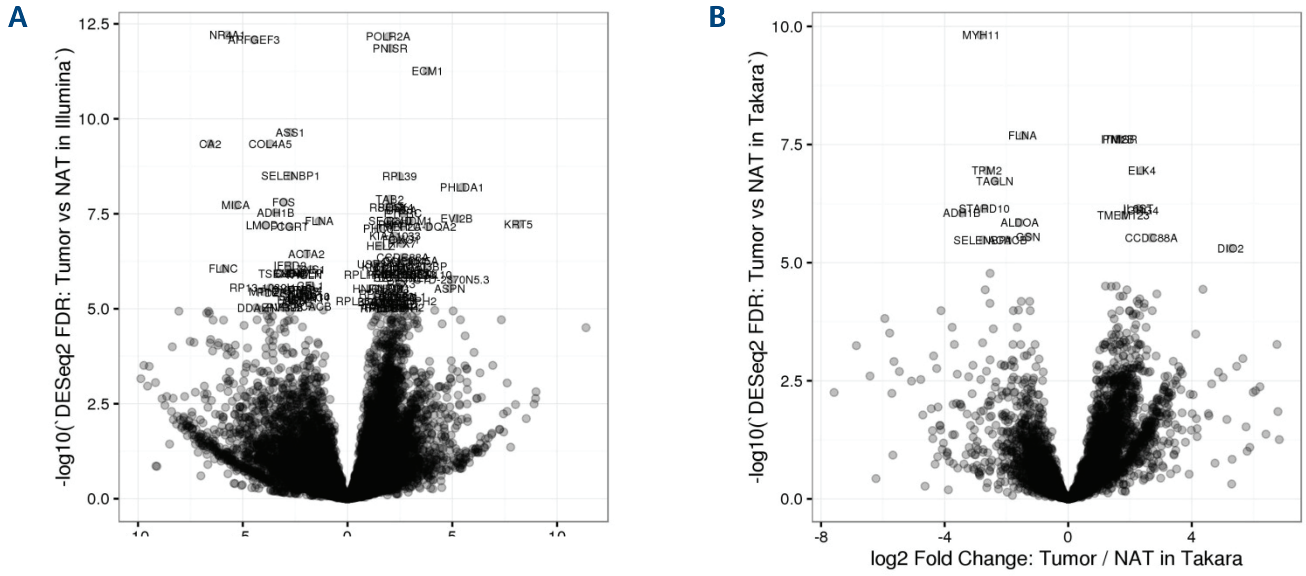
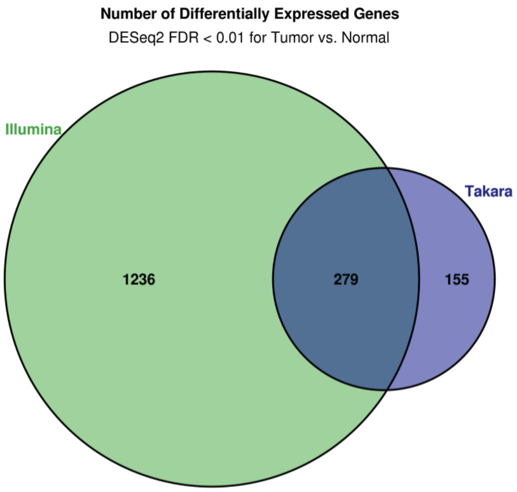


FIGURE 6. OVERLAP OF GENES DIFFERENTIALLY EXPRESSED BETWEEN TUMOR AND NORMAL SAMPLES

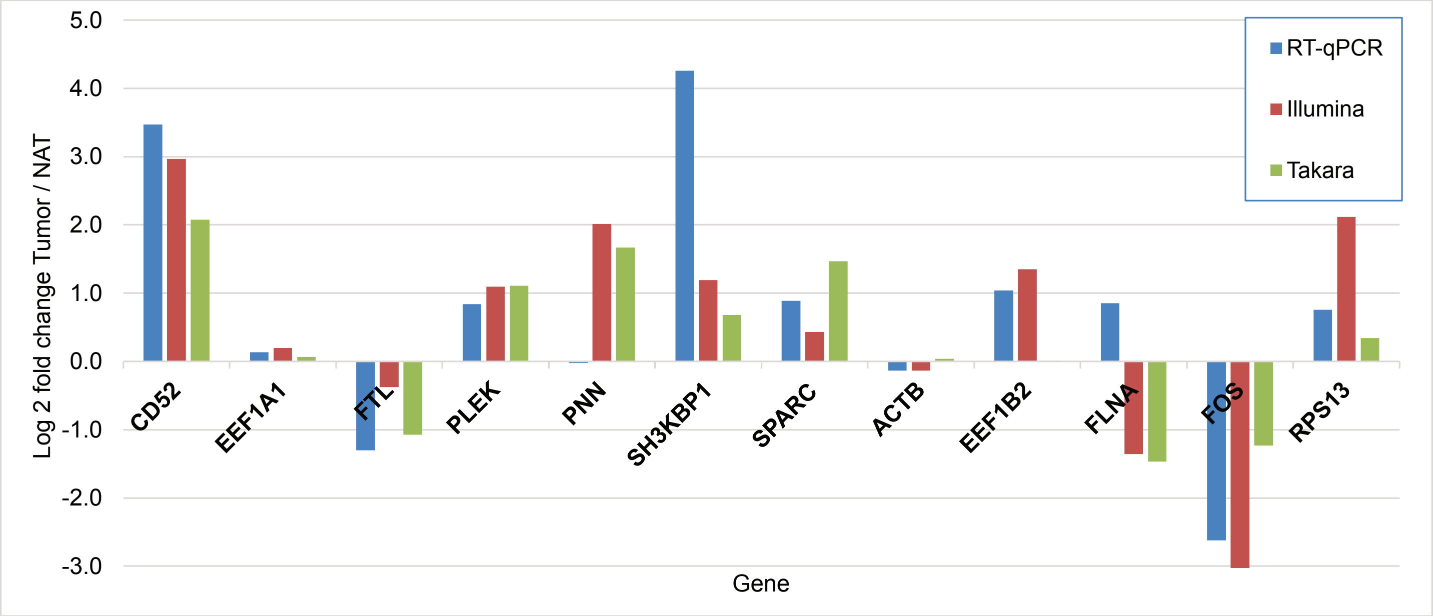
DESeq2 was used to evaluate how the gene expression levels depended on the tumor status and library preparation method of each sample. The Venn diagram shows how many of the 1,670 genes whose expression differed between tumor and normal samples were identified in Illumina TruSeq RNA Exome libraries, Takara SMARTer Stranded v2 libraries, or both.



Validation of RNA-seq Gene Expression by RT-qPCR

In order to confirm the gene expression levels measured by the TruSeq RNA Exome and SMARTer Stranded v2 libraries, RT-qPCR was run on the same RNA samples to measure the levels of 4 genes with detectable expression that were differentially expressed between NAT and Tumor (*CD52, FTL, PLEK,* and *SH3KBP1*), 6 genes with high expression that were differentially expressed between NAT and Tumor (*FLNA, PNN*, *FOS*, *RPS13*, *SPARC*, and *EEF1B2*), and 2 reference genes whose expression was similar between NAT and Tumor (*EEF1A1* and *ACTB*). The RT-qPCR data for 10 of the genes were consistent with the expression patterns observed with both the Illumina and Takara kits (Figure 7). Only two genes (*PNN* and *FLNA*) did not show the same expression patterns observed with Takara and Illumina kits.

FIGURE 7. VALIDATION OF FFPE RNA SEQUENCING DATA USING RT-QPCR.



The Log (2) of the fold change for Tumor / NAT, calculated from the average of all samples in each group, is plotted for 12 genes. The data used to calculate the fold changes were RPKM values from mRNA sequencing using the Illumina TruSeq RNA Exome and Takara SMARTer Stranded v2 library preparation kits or from RQ values generated from an RT-qPCR assay of the same RNA samples used for sequencing. ACTB and EEF1A1 genes were used as the reference genes while calculating the RQ values.

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

Conclusions

The most important finding of this study was that sequencable mRNA libraries could be generated from total RNA isolated from FFPE tissue using both the TruSeq RNA Exome and Takara SMARTer Stranded v2 library preparation methods. By contrast, the NuGen Ovation Universal Human FFPE RNA-Seq System generated libraries with substantial primer dimers that could not be removed with additional bead-based purification and therefore were not of sufficient quality to sequence.

The Illumina and Takara library preparation methods produced high-quality sequencing reads with some degree of nucleotide composition bias, especially at the start of the reads. Takara SMARTer libraries had substantially higher ribosomal RNA content (27%) compared to Illumina TruSeq RNA Exome (1.3%). Illumina TruSeq RNA Exome libraries had more reads mapped to the hg38 reference genome (90%) compared to Takara SMARTer (71%). Illumina TruSeq RNA Exome libraries had more mapped bases overlapping coding regions in annotated UCSC genes (68%) compared to SMARTer libraries (13%). Illumina TruSeq RNA Exome libraries had more mapped reads assigned to annotated Ensembl v.83 genes (59%) vs. Takara SMARTer Stranded v2 libraries (16%). The top 2 biotypes for Illumina libraries were protein_coding and snoRNA while the top 2 biotypes in Takara libraries were protein_coding and rRNA. However, Takara libraries had more diversity (covered more genes for a given sequencing depth). RT-qPCR validation of 10 genes that were differentially expressed between NAT and Tumor in the RNA-seq data confirmed the expression pattern of 8 out of 10 genes, indicating a high degree of fidelity in the RNA-seq libraries prepared using with Illumina's TruSeq RNA Exome or Takara's SMARTer Stranded RNA-Seq Kit v2.

Among the differentially expressed transcripts, the eight genes (*CD52*, *FTL*, *PLEK*, *SH3KBP1*, *SPARC*, *FOS* and *RPS13*) validated by both RNA-seq and RT-qPCR have been shown to be directly or indirectly involved in cancer. The *SPARC* gene encodes a cysteine-rich protein mediating interaction between cells and their extracellular matrix (*ECM*). Studies have shown that *SPARC* plays a role in the pathological responses in lung cancer (Grant et al., 2014; Wong and Sukkar, 2017). Our data shows that the *RPS13* gene is upregulated in tumor tissues and higher level of expression of *RPS13* has been reported in gastric, lung as well as colon cancer (Denis et al., 1993; Guo et al., 2011; Zhang et al., 2000). The *FOS* gene (a.k.a. *c-fos*) is a well-known proto-oncogene (Saez et al., 1995) and, interestingly, many of the signaling pathways involved in tumorigenesis leads to the activation of *FOS* (Bakin and Curran, 1999; Ordway et al., 2005). The *PLEK* gene is one of the differentially expressed we identified, and it has been shown that the expression pattern of *PLEK* is associated with low survival rates in patients displaying lung cancer (Vuong et al., 2014). Overall, the involvement of the validated genes in tumorigenesis shows that the data generated is robust, and the Illumina TruSeq RNA Exome as well as Takara SMARTer kits successfully identified cancer or cancer-related genes in the tumor tissues assayed.

Methods

FFPE Tissue Samples

FFPE tumor samples were provided by the Veteran's Administration Hospital at the University of Miami (n=4). Normal FFPE tissue samples were purchased from Geneticist, Inc. (Glendale, CA) and consisted of ascending colon (n=1), breast (n=2), and rectum (n=1).

RNA Isolation

FFPE samples were sectioned to 20 µm and at least four sections (≤ 35 mg) from each sample were deparafinized at 50 °C for 3 minutes in xylene. Total RNA was extracted from the samples using the RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Part No. AM1975, ThermoFisher Scientific, Waltham, MA) according to the manufacturer's instructions. The 8 total RNA samples were further processed by digestion with RNase-free DNase I (Part No. D9905K,

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

Lucigen Corporation, Middleton, WI) and re-purified using RNeasy MinElute purification columns according to the manufacturer's recommendations(Part No. 74204, QIAGEN, Valencia, CA), except that the final ethanol concentration of the sample was adjusted to 77% before loading on the column in order to retain RNAs less than 200 nucleotides in size . Each newly digested RNA sample was run on the Agilent 2100 Bioanalyzer to evaluate the electropherograms and obtain the RNA integrity number (RIN) as well as the percentage of RNA fragments longer than 200 nucleotides (DV_{200}) values. Following Bioanalyzer analysis, the total RNA samples were used as input for library preparation. Two positive controls consisting of pooled total RNA samples (Ambion) were included in library preparation along with a negative control.

mRNA Sequencing Library Preparation

Amplified cDNA libraries suitable for sequencing were prepared from DNA-free total RNA using three library preparation kits:

1. ***TruSeq RNA Exome (Illumina catalog no. 20020189)*** – The input for this protocol was 50 ng total RNA.
2. ***SMARTer Stranded Total RNA-Seq Kit v2*** - Pico Input Mammalian (Takara catalog no. 634412) – The input for this protocol was 50 ng total RNA.
3. ***Ovation Human FFPE RNA***-Seq Multiplex System 1-8 (NuGen catalog no. 0340-32) – The input for this protocol was 300 ng total RNA.

The quality and size distribution of the amplified libraries were determined by chip-based capillary electrophoresis (Bioanalyzer 2100, Agilent Technologies), and libraries were quantified using the KAPA Library Quantification Kit (Kapa Biosystems, Boston, MA). The 11 total RNA samples (4 FFPE tumor samples, 4 FFPE normal samples, 2 total RNA positive controls, and 1 negative control) were each split into three aliquots, one for each of the three library preparation methods. Hence, 3 libraries were generated from the same total RNA.

Sequencing

The libraries were pooled at equimolar concentrations and diluted prior to loading onto an Illumina NextSeq 500 v2.5 flow cell cartridge. The libraries were extended and bridge amplified to create sequence clusters and sequenced with 76 nt paired-end reads plus 8nt single-index reads using the Illumina NextSeq 500 High Output sequencing reagent kit v2 (Part # 15057931) controlled by the NextSeq Control Software version 2.2.0.4. Real time image analysis and base calling were performed on the instrument using the Real-Time Analysis (RTA) software version 2.4.11.

Read Filtering and Trimming

The FASTQ files generated from the sequencing base call files contained only reads that passed Illumina's chastity filter. Any Illumina adapters were trimmed from the 3'-end of both the R1 and R2 reads. Bases with a quality score less than Q20 were trimmed off the right end of each of R1 and R2. Read pairs in which either mate in the pair was less than 30 nt after trimming were discarded.

For TruSeq RNA Exome libraries, 7 bases were then trimmed from the left end of R1 and 1 base was trimmed from the left end of R2 due to a skewed nucleotide distribution in the first bases. For Takara SMARTer Stranded libraries, 6 bases were trimmed from the left end of R1 and 3 bases were trimmed from the left end of R2 due to a skewed nucleotide distribution in these bases. These quality-filtered and base-trimmed reads were then used for alignment.

Assessment of Ribosomal RNA Content

To assess the ribosomal RNA content, one million untrimmed reads from each sample were aligned to human ribosomal sequences, including 45S pre-ribosomal N4 (RNA45SN4) - NR_146117.1 - and 5S ribosomal 1 (RNA5S1) - NR_023363.1. Bowtie2 was used for alignment with the --**`very-sensitive`** settings and the overall alignment rate was calculated.

Alignment to Human Genome

Sequence alignment was performed using HISAT2 version 2.0.5 with the following settings:

```
--end-to-end -N 1 -L 20 -i S,1,0.5 -D 25 -R 5
--pen-noncansplice 12 --mp 6,3 --sp 3,0 --time
--reorder --known-splicesite-infile [SPLICESITES]
--novel-splicesite-outfile  splicesites.novel.txt
--novel-splicesite-infile splicesites.novel.txt -q
-x [hg38 HISAT2 INDEX] -1 [FASTQ1] -2 [FASTQ2] -S
[SAMOUT]
```

Where **SPLICESITES** is a BED file of known splice sites extracted from the Ensembl version 83 *Homo sapiens* GTF (gene transfer format) annotation file, **FASTQ1** and **FASTQ2** are files containing the 1st and 2nd sequencing reads, respectively, and **SAMOUT** is the alignment output file. The **hg38 HISAT2 INDEX** was generated from the hg38 reference genome using the command **hisat2-build hg38.fa hg38**.

Read Counting

The read summarization program featureCounts version 1.5.1 was used for exon- and gene-level counting. An Ensembl human version 83 GTF file was used for determination of exon boundaries and the exon-gene relationship during counting. The summarization level used for exon and gene counting was the feature and the meta-feature, respectively. To be counted for a exon or gene, a fragment must have aligned with a mapping quality of at least Q20, have aligned uniquely to the genome (i.e. multimapping fragments were not counted), and overlap the region by at least 1 nucleotide. Fragments overlapping multiple annotated genes or exons were counted once for each overlapping feature.

Mapped Nucleotide Composition

The Picard Tools version 1.141 CollectRnaSeqMetrics utility was used to calculate summary metrics describing the distribution of bases within transcripts annotated by UCSC. The input data were BAM alignment files resulting from mapping to hg38 using HISAT2. Ensembl version 83 genes with an rRNA biotype were used to define ribosomal coordinates – the 45S pre-ribosomal and 5S ribosomal genes are not annotated in Ensembl and therefore the percent rRNA reported by Picard Tools and the alignment method differ substantially for some samples.

Gene Biotype Composition of Mapped Reads

The number of reads mapping to Ensembl version 83 genes was totaled by the annotated biotype of each gene. The input data were raw gene counts for all 60,504 annotated human genes generated by featureCounts. Only the top 11 gene biotypes were considered separately and the remaining biotypes were grouped together into a single class, labelled in the figure as “other”.

Normalized RPKM Values Based on Counts from Nuclear Protein-Coding Genes

Normalized RPKM values were calculated from the raw featureCounts read counts using the formula $\frac{Gr/_{Gl}}{mRNArt/_{1,000,000}}$ for genes and $\frac{Er/_{El}}{mRNArt/_{1,000,000}}$ for exons, where Gr is raw read count for a gene, Gl is length of the exon model for the gene (i.e. the sum of all exon lengths in kilobases), Er is raw read count for an exon, El is length of the exon in kilobases, and mRNArt is the total read count for exons from nuclear-derived protein-coding mRNAs.

Number of Detectable Protein-Coding Genes

The raw gene counts were used to assess the number of protein-coding genes with at least 10 mapped reads at different sequencing depths. Random subsamples of the counted sequencing reads, consisting of between 1 and 2 million subsampled reads, were generated and the number of genes with at least 10 mapped reads in each subsample was counted. The input data were gene counts for only the 20,253 nuclear protein-coding genes.

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

Differential Expression Analysis

DESeq2 was used to examine differential expression between tumor and normal tissue samples, while controlling for the total RNA input into the library preparation. The hypothesis tested was that gene expression levels were proportional to the total RNA used for the library, the library preparation method, and the tumor status of the sample (i.e. **`expression ~ Total.RNA.ID + Library.Prep.Kit + Tumor.Status`**).

Real-time reverse-transcription-PCR analysis

A total of 450 nanograms of DNA-free total RNA from each sample was reverse transcribed with random primers using the High-Capacity Reverse Transcription kit (Thermo Fisher; Part # 4368813). The same input mass of a pool of human tissue RNAs and nuclease-free water were reverse transcribed alongside the experimental samples as positive and negative controls, respectively. The reverse transcription products were diluted in nuclease-free water and cDNA equivalent to 9.5 ng of template RNA per well were used to set up triplicate 10 ul PCR reactions containing gene-specific Taqman mRNA probe sets and 1X Universal Master Mix (Thermo Fisher; part # 43-643-43). Thermocycling and imaging were performed using the BioRad CFX384 quantitative PCR instrument; incubation conditions for quantitative PCR included denaturation at 95C for 10 minutes followed by 40 cycles of 95C 15 seconds – 60C 60 seconds. Cycle threshold (Ct) values were determined from the fluorescent signal intensities measured from each well after each PCR cycle using BioRad CFX Manager Software v3.1.1517.0823, using a baseline-subtracted curve fit and automatically determined thresholds for each set of genes run on the same plate. The mean values of the three replicate wells were averaged, using a Ct of 40 in the calculation for undetectable wells. For each sample, the mean of the normalization control CTs (ACTB, EEF1A1) was subtracted from each mRNA CT value to obtain ΔCT values. For each mRNA, the ΔCT of the mean of NAT samples was subtracted from each sample ΔCT to obtain the ΔΔCT values. The relative quantities (RQ) were calculated as: RQ=2^(-ΔΔCT), and these values were used to calculate the fold-changes for tumor / NAT.

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

REFERENCES:

1. Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysoker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat. Methods 10, 623–629.
2. von Ahlfen, S., Missel, A., Bendrat, K., and Schlumpberger, M. (2007). Determinants of RNA Quality from FFPE Samples. PLOS ONE 2, e1261.
3. Bakin, A.V., and Curran, T. (1999). Role of DNA 5-methylcytosine transferase in cell transformation by fos. Science 283, 387–390.
4. Bossel Ben-Moshe, N., Gilad, S., Perry, G., Benjamin, S., Balint-Lahat, N., Pavlovsky, A., Halperin, S., Markus, B., Yosepovich, A., Barshack, I., et al. (2018). mRNA-seq whole transcriptome profiling of fresh frozen versus archived fixed tissues. BMC Genomics 19, 419.
5. Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D., and Craig, D.W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat. Rev. Genet. 17, 257–271.
6. Clark, M.B., Mercer, T.R., Bussotti, G., Leonardi, T., Haynes, K.R., Crawford, J., Brunck, M.E., Cao, K.-A.L., Thomas, G.P., Chen, W.Y., et al. (2015). Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. Nat. Methods 12, 339–342.
7. Costa, V., Aprile, M., Esposito, R., and Ciccodicola, A. (2013). RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur. J. Hum. Genet. 21, 134–142.
8. Denis, M.G., Chadeneau, C., Lecabellec, M.T., LeMoullac, B., LeMevel, B., Meflah, K., and Lustenberger, P. (1993). Over-expression of the S13 ribosomal protein in actively growing cells. Int. J. Cancer 55, 275–280.
9. Esposti, D.D., Hernandez-Vargas, H., Voegele, C., Fernandez-Jimenez, N., Forey, N., Bancel, B., Calvez-Kelm, F.L., McKay, J., Merle, P., and Herceg, Z. (2016). Identification of novel long non-coding RNAs deregulated in hepatocellular carcinoma using RNA-sequencing. Oncotarget 7, 31862–31877.
10. Esteve-Codina, A., Arpi, O., Martinez-García, M., Pineda, E., Mallo, M., Gut, M., Carrato, C., Rovira, A., Lopez, R., Tortosa, A., et al. (2017). A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. PLOS ONE 12, e0170632.
11. Grant, J.L., Fishbein, M.C., Hong, L.-S., Krysan, K., Minna, J.D., Shay, J.W., Walser, T.C., and Dubinett, S.M. (2014). A Novel Molecular Pathway for Snail-Dependent, SPARC-Mediated Invasion in Non–Small Cell Lung Cancer Pathogenesis. Cancer Prev. Res. (Phila. Pa.) 7, 150–160.
12. Guo, X., Shi, Y., Gou, Y., Li, J., Han, S., Zhang, Y., Huo, J., Ning, X., Sun, L., Chen, Y., et al. (2011). Human ribosomal protein S13 promotes gastric cancer growth through down-regulating p27(Kip1). J. Cell. Mol. Med. 15, 296–306.
13. Hedegaard, J., Thorsen, K., Lund, M.K., Hein, A.-M.K., Hamilton-Dutoit, S.J., Vang, S., Nordentoft, I., Birkenkamp-Demtröder, K., Kruhøffer, M., Hager, H., et al. (2014). Next-Generation Sequencing of RNA and DNA Isolated from Paired Fresh-Frozen and Formalin-Fixed Paraffin-Embedded Samples of Human Cancer and Normal Tissue. PLOS ONE 9, e98187.
14. Hester, S.D., Bhat, V., Chorley, B.N., Carswell, G., Jones, W., Wehmas, L.C., and Wood, C.E. (2016). Editor's Highlight: Dose–Response Analysis of RNA-Seq Profiles in Archival Formalin-Fixed Paraffin-Embedded Samples. Toxicol. Sci. 154, 202–213

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

REFERENCES: *(cont.)*

15. Huang, R., Jaritz, M., Guenzl, P., Vlatkovic, I., Sommer, A., Tamir, I.M., Marks, H., Klampfl, T., Kralovics, R., Stunnenberg, H.G., et al. (2011). An RNA-Seq Strategy to Detect the Complete Coding and Non-Coding Transcriptome Including Full-Length Imprinted Macro ncRNAs. PLOS ONE 6, e27288.
16. Jovanović, B., Sheng, Q., Seitz, R.S., Lawrence, K.D., Morris, S.W., Thomas, L.R., Hout, D.R., Schweitzer, B.L., Guo, Y., Pietenpol, J.A., et al. (2017). Comparison of triple-negative breast cancer molecular subtyping using RNA from matched fresh-frozen versus formalin-fixed paraffin-embedded tissue. BMC Cancer 17, 241.
17. Klopfleisch, R., Weiss, A.T.A., and Gruber, A.D. (2011). Excavation of a buried treasure–DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. Histol. Histopathol. 26, 797–810.
18. Kresse, S.H., Namløs, H.M., Lorenz, S., Berner, J.-M., Myklebost, O., Bjerkehagen, B., and Meza-Zepeda, L.A. (2018). Evaluation of commercial DNA and RNA extraction methods for high-throughput sequencing of FFPE samples. PLOS ONE 13, e0197456.
19. Li, J., Fu, C., Speed, T.P., Wang, W., and Symmans, W.F. (2018). Accurate RNA Sequencing From Formalin-Fixed Cancer Tissue to Represent High-Quality Transcriptome From Frozen Tissue. JCO Precis. Oncol. 1–9.
20. Liang, J., Lv, J., and Liu, Z. (2015). Identification of stage-specific biomarkers in lung adenocarcinoma based on RNA-seq data. Tumor Biol. 36, 6391–6399.
21. Lin, L., Park, J.W., Ramachandran, S., Zhang, Y., Tseng, Y.-T., Shen, S., Waldvogel, H.J., Curtis, M.A., Faull, R.L.M., Troncoso, J.C., et al. (2016). Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease. Hum. Mol. Genet. 25, 3454–3466.
22. Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M., and Okubo, K. (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res. 27, 4436–4443.
23. Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). The human transcriptome across tissues and individuals. Science 348, 660–665.
24. Morlan, J.D., Qu, K., and Sinicropi, D.V. (2012). Selective Depletion of rRNA Enables Whole Transcriptome Profiling of Archival Fixed Tissue. PLOS ONE 7, e42882.
25. Norton, N., Sun, Z., Asmann, Y.W., Serie, D.J., Necela, B.M., Bhagwate, A., Jen, J., Eckloff, B.W., Kalari, K.R., Thompson, K.J., et al. (2013). Gene Expression, Single Nucleotide Variant and Fusion Transcript Discovery in Archival Material from Breast Tumors. PLOS ONE 8, e81925.
26. Ordway, J.M., Fenster, S.D., Ruan, H., and Curran, T. (2005). A transcriptome map of cellular transformation by the fos oncogene. Mol. Cancer 4, 19.
27. Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. 30, 777–782.
28. Riccardi, S., Bergling, S., Sigoillot, F., Beibel, M., Werner, A., Leighton-Davies, J., Knehr, J., Bouwmeester, T., Parker, C.N., Roma, G., et al. (2016). MiR-210 promotes sensory hair cell formation in the organ of corti. BMC Genomics 17, 309.

TRANSCRIPTOME ANALYSIS IN FIXED TUMOR BIOPSY SAMPLES

Frontage

REFERENCES: *(cont.)*

29. Saez, E., Rutberg, S.E., Mueller, E., Oppenheim, H., Smoluk, J., Yuspa, S.H., and Spiegelman, B.M. (1995). c-fos is required for malignant progression of skin tumors. Cell 82, 721–732.
30. Sinicropi, D., Qu, K., Collin, F., Crager, M., Liu, M.-L., Pelham, R.J., Pho, M., Rossi, A.D., Jeong, J., Scott, A., et al. (2012). Whole Transcriptome RNA-Seq Analysis of Breast Cancer Recurrence Risk Using Formalin-Fixed Paraffin-Embedded Tumor Tissue. PLOS ONE 7, e40092.
31. Stanta, G., and Schneider, C. (1991). RNA extracted from paraffin-embedded human tissues is amenable to analysis by PCR amplification. BioTechniques 11, 304, 306, 308–304, 306, 308.
32. Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome. Science 321, 956–960.
33. Vuong, H., Cheng, F., Lin, C.-C., and Zhao, Z. (2014). Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. Genome Med. 6, 81.
34. Webster, A.F., Zumbo, P., Fostel, J., Gandara, J., Hester, S.D., Recio, L., Williams, A., Wood, C.E., Yauk, C.L., and Mason, C.E. (2015). Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archival Formalin-Fixed Paraffin-Embedded Tissues. Toxicol. Sci. 148, 460–472.
35. Wong, S.L.I., and Sukkar, M.B. (2017). The SPARC protein: an overview of its role in lung cancer and pulmonary fibrosis and its potential role in chronic airways disease. Br. J. Pharmacol. 174, 3–14.
36. Yi, H., Cho, Y.-J., Won, S., Lee, J.-E., Jin Yu, H., Kim, S., Schroth, G.P., Luo, S., and Chun, J. (2011). Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. Nucleic Acids Res. 39, e140–e140.
37. Zhang, L., Cilley, R.E., and Chinoy, M.R. (2000). Suppression subtractive hybridization to identify gene expressions in variant and classic small cell lung cancer cell lines. J. Surg. Res. 93, 108–119.
38. Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N., and Perou, C.M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics 15, 419.
39. Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A., et al. (2004). Simple cDNA normalization using kamchatka crab duplex□specific nuclease. Nucleic Acids Res. 32, e37–e37.