# Development of a modified SMART system for robust transcriptome library preparation from limited quantities of compromised samples

Magnolia Bostick, Cynthia Chang and Andrew Farmer

Clontech® Laboratories, Inc. 1290 Terra Bella Avenue, Mountain View CA 94043

## Abstract

Next Generation Sequencing has revolutionized biomedical research by providing sequence data on millions of short DNA fragments, in parallel. In particular, NGS has enabled RNA expression analysis over the entire transcriptome with high sensitivity and dynamic range. Currently, the field is seeking methods to utilize challenging samples that are either compromised or are only available in limited amounts. Overcoming these constraints will require highly sensitive and robust sample preparation methods.

One powerful method for cDNA preparation is SMART™ technology (**S**witching **M**echanism **A**t the 5' end of the **R**NA **T**emplate), which utilizes the template switching activity of reverse transcriptase to enable the direct addition of a PCR adaptor to the 3' end of the first-strand cDNA, thus avoiding inefficient ligation steps. One drawback of SMART technology is its current inability to work with compromised samples, owing to its dependence on an oligo dT primer for first strand synthesis.

A modified SMART system has been developed including the use of random primers to work with samples containing compromised or degraded RNA. Data (including gene body coverage, reproducibility, and mappability metrics) will be presented for both chemically degraded RNA samples and FFPE RNA prepared using the modified SMART system.

This modified SMART protocol will be especially useful for small samples of degraded RNA. It is capable of generating cDNA libraries for transcriptome profiling from as little as 10 ng of total RNA.

## Introduction

Sequencing RNA is an important method for learning how any particular cell functions. A variety of techniques have been developed over the last few decades to identify RNA, ranging from single-sequence identification via reverse transcriptase-dependent cloning or quantitative PCR to microarray analysis of populations of genes and, most recently, next generation sequencing of all of the RNA in a cell. For many samples containing well-preserved mRNA, a dT oligo-primed reverse transcriptase reaction is an excellent method for identifying full length cDNA. The SMARTer® Ultra Low RNA Kit for Illumina® Sequencing allows researchers to push the limits of this method down to the single-cell level (2; see also Poster #42). Unfortunately, this powerful technology requires RNA with little to no degradation—but many common sample preparation techniques, including FFPE (formaldehyde fixed paraffin embedded tissue) and laser-capture dissection, lead to degraded samples. We are developing a SMARTer Universal Low Input RNA Kit to address the unmet need for a cDNA production kit that accommodates low concentrations of degraded samples (Figures 1 and 2). We have tested the kit on sheared control RNA and on FFPE samples, and have confirmed library production results with sequencing (Figures 3 and 5). The results of these experiments are presented.

## Results

We tested several different random primers for first strand cDNA production, and the N6 primer gave the greatest amplification with the least primer dimers (data not shown). Therefore, we adapted the protocol for the SMARTer Ultra Low RNA Kit for Illumina Sequencing for use with the universal N6 primer (Figures 1 and 2). We created control sheared Human Universal RNA and compared it to RNA extracted from Breast Carcinoma FFPE slides. Both had a distribution of lengths, with a low peak (less than 200 bp) and an RNA Integrity Number (RIN) of ~2 (Figure 3, Panel A). We then created cDNA libraries with varying inputs of the same Breast Carcinoma FFPE RNA, ranging from 50 ng down to 2 ng. The DNA concentrations of these libraries indicated that we were within the linear range of amplification, perfect for sequencing (Figure 3, Panel B). We also sequenced two different cDNA libraries made from 20 ng each of the same Human Universal RNA. The high correlation (Pearson, 0.959) between the expression levels of the two libraries indicates good reproducibility from our kit (Figure 4). We compared these results to sequencing of RNA extracted from a flash frozen sample from the same tumor. This RNA was sequenced with either the SMARTer Ultra Low RNA Kit for Illumina Sequencing (dT) or the SMARTer Universal Low RNA Kit for Sequencing (N6). The gene body coverage was more complete for the N6-primed library compared to the dT-primed library (Figure 5, Panel B). A summary of the sequencing runs is presented in Table I. All three libraries had a high percentage of reads mapping to the human genome, a high percentage of unique reads, and an appropriate number of genes identified in the library.

## References

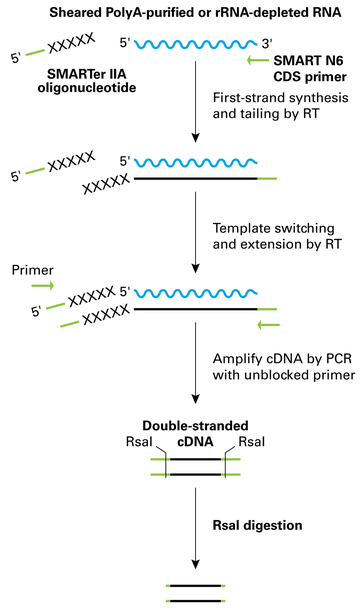
1. Chenchik, A. *et al.,* (1998) In RT-PCR Methods for Gene Cloning and Analysis. Eds. Siebert, P. & Larrick, J. (BioTechniques Books, MA), pp. 305–319.

2. Ramsköld, D. *et al.* (2012) *Nat. Biotechnol.* **30**(8):777–782.



633576

## 1 SMART Technology for cDNA Synthesis from Small Amounts of Total RNA




1. **First strand synthesis and tailing by SMARTScribe™ Reverse Transcriptase.** The SMARTer Universal Low Input RNA Kit for Sequencing starts with sheared polyA-purified or rRNA depleted RNA and a modified N6 primer called the SMART N6 CDS Primer.

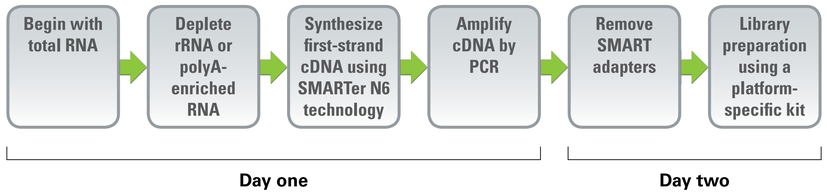
2. **Template switching and extension by SMARTScribe Reverse Transcriptase.** When SMARTScribe reaches the 5' end of the RNA, its terminal transferase activity adds a few additional nucleotides to the 3' end of the cDNA. The SMARTer Oligonucleotide base-pairs with this non-template nucleotide stretch, creating an extended template to enable SMARTScribe to continue replicating to the end of the oligonucleotide (1).

3. **cDNA amplification.** The SMARTer anchor sequence and the modified N6 sequence serve as universal priming sites for cDNA amplification.

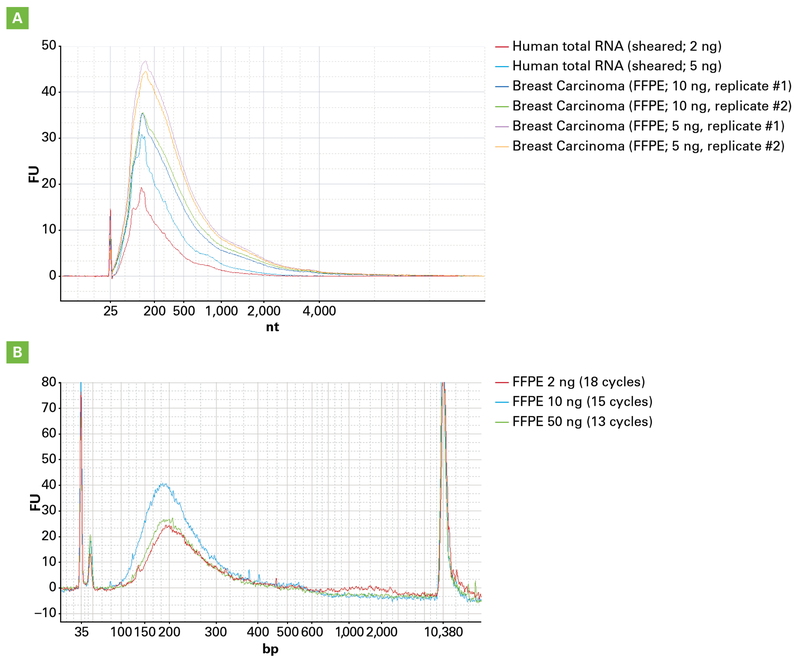
4. **Adapter removal.** Finally, amplified cDNA is digested with RsaI to remove the adapter prior to sequencing.

**SMARTer technology for cDNA synthesis from extremely small samples.** The SMARTer Universal Low Input RNA Kit for Sequencing starts with 200 pg–10 ng of input RNA and a modified N6 primer (where N = A, G, T, or C) called the SMART N6 CDS Primer, and produces cDNA libraries suitable for transcriptome profiling.

## 2 NGS Workflow: SMARTer Universal Low Input RNA Kit for Sequencing




## 3 FFPE & Degraded Samples



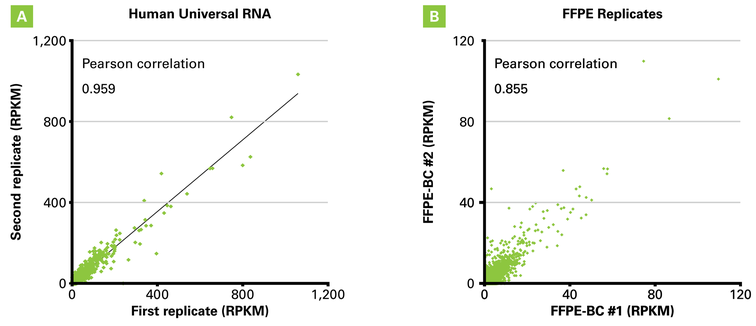

**The degraded nature of FFPE RNA, illustrated using Agilent 2100 Bioanalyzer electropherograms (RNA 6000 Pico chip). Panel A.** 10 ng (green and aqua lines) or 5 ng of RNA (orange and pink lines) of Breast Carcinoma FFPE RNA and 2 ng (red line) or 5 ng (blue line) of sheared Human Universal Reference RNA are shown to illustrate the degraded nature of the samples. RINs (RNA Integrity Numbers) were between 2.0–2.2. **Panel B.** 500 ng of total FFPE RNA was treated with the Epicentre RiboZero kit, concentrated to 10 µl, and used to make three SMARTer N6-primed cDNA libraries. Green line: cDNA library created from 1 µl RiboZero-treated, concentrated FFPE RNA (equivalent to 50 ng of total RNA) using 13 cycles of PCR. Blue line: cDNA library created from 200 nl RiboZero-treated, concentrated FFPE RNA (equivalent to 10 ng of total RNA) using 15 cycles of PCR. Red line: cDNA library created from 40 nl RiboZero-treated, concentrated FFPE RNA (equivalent to 2 ng of total RNA) using 18 cycles of PCR.

**Table I: Sequencing Results for N6 Libraries**

| | Human Universal (20 ng) | FFPE Breast Cancer (20 ng) | Human Brain (10 ng) |
|---|---|---|---|
| **Number of Reads** | 7,580,325 | 5,485,865 | 3,006,898 |
| **Mapped to Genome (%)** | 99 | 98.7 | 94.3 |
| **Mapped to ORFs (%)** | 85 | 54 | 81 |
| **Mapped Uniquely (%)** | 77 | 48 | 72 |
| **RNA (%)** | 1.9 | 0.5 | 3.4 |
| **Number of Genes** | 15,574 | 17,963 | 16,050 |

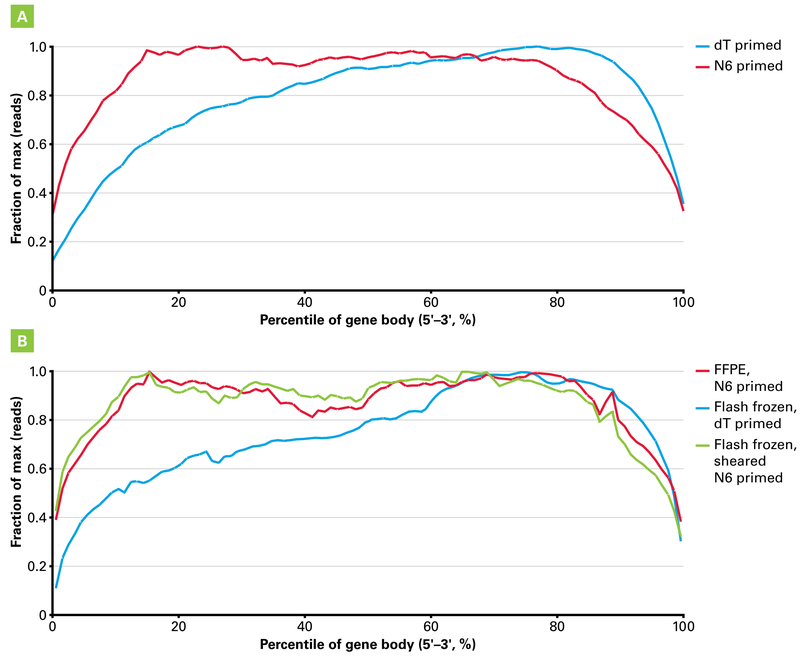
**Sequencing alignment metrics of three N6 libraries.** Human Universal Reference RNA, FFPE Breast Carcinoma RNA purified from Cureline slides using the MACHEREY-NAGEL NucleoSpin FFPE RNA kit, and Human Brain Reference RNA were treated with the RiboZero kit to remove rRNA and converted to cDNA as described in Figures 1 and 2. Illumina adapters and indices were added using ThruPlex technology from Rubicon. cDNA libraries were sequenced on an Illumina MiSeq® Platform with 1 x 50 bp reads. Unmapped reads were trimmed by CLC Genomics Workbench. Reads were mapped to rRNA and the mitochondrial genome with CLC (% reads indicated). Unmapped reads were then mapped to the human genome (UCSC hg19), and % reads that mapped are indicated. In parallel, the trimmed reads were mapped to the human genome with the RefSeq masking with CLC, producing uniquely mapped reads and % reads that mapped to RefSeq annotations. Number of genes identified in each library was determined by the number of genes with an RPKM (Read per Kilobase of exon per Million of reads) ≥ 0.1.

## 4 Correlation between N6 Libraries




**Pearson correlation statistics of expression (RPKM) between replicate libraries.** Pearson correlation between libraries made with the N6 primer from Human Universal RNA. **Panel A.** Two independent libraries were made with 20 ng of total Human Universal Reference RNA. **Panel B.** Two libraries were made with 20 ng of total RNA extracted from two independent slides taken from the same Breast Carcinoma (BC) tumor. All four samples were treated as described in Figure 3 and Table I.

## 5 Gene Body Coverage




**Gene body coverage.** Genes were scaled to 100% and the number of reads within each percentile bin was counted and normalized to the highest percentile within each library. **Panel A.** Gene body coverage generated with control Human Universal Reference RNA using the dT CDS SMART primer (blue line) or the N6 CDS SMART primer (red line). Both libraries had good coverage over the entire gene body without strong biases. **Panel B.** Gene body coverage generated with three samples of breast carcinoma RNA taken from a single donor and either FFPE-treated (red line) or flash frozen. The RNA from the flash frozen sample was either left full length (RIN of 7.9) and converted to cDNA with the dT primer (blue line) or chemically sheared and converted to cDNA with the N6 primer (green line). dT coverage was not as good as N6 coverage, due to the lower quality of the flash frozen library.