SMARTer™ Ultra Low RNA Kit for Illumina® Sequencing

Two powerful technologies combine to enable sequencing with ultra-low levels of RNA

- The most sensitive cDNA synthesis technology, combined with next-generation sequencing
- Robust library generation starting from picogram quantities of total RNA
- Wide dynamic range, accurate gene quantification, and unparalleled sensitivity for ultra-low RNA input
- Powerful analysis of gene expression levels and alternatively spliced isoforms, characterization of coding SNPs, and gene discovery

Introduction

RNA Sequencing (RNA-Seq) has revolutionized gene expression analysis, providing an unprecedented view into the transcriptome complexity of diverse cell types. Although current RNA-Seq methods require far less RNA than other methods of gene expression analysis, they are still inadequate when analyzing samples of very limited size. The ability to study exceptionally rare or precious samples—including stem cells, circulating tumor cells, and brain tissue biopsies—requires extraordinarily sensitive and reproducible methods. To that end, Clontech has teamed its proven SMART technology with Illumina's industry-leading, next-generation sequencing platforms to provide a powerful new tool for the analysis of these exceedingly limited samples.

Clontech's **SMARTer Ultra Low RNA Kit for Illumina Sequencing** enables cDNA synthesis for library construction and sequencing with all Illumina HiSeq™, HiScanSQ™, and Genome Analyzer systems. The kit has been specially created to generate cDNA for sequencing from as little as 100 pg of purified total RNA. The resulting cDNA is fed directly into a modified Illumina library prep protocol for use with standard Illumina sequence-by-synthesis (SBS) sequencing chemistries.

SMART cDNA synthesis technology

Clontech's revolutionary SMART (Switching Mechanism At the 5' end of RNA Transcript) technology provides an exceptionally efficient method for accurately synthesizing full-length cDNA from RNA in a single reaction. The SMARTer Ultra Low RNA Kit combines well-established aspects of Clontech's SMART(er) kits with an optimized technique specifically designed for compatibility with Illumina sequencing.

SMART technology offers unparalleled sensitivity and unbiased amplification of mRNA transcripts, while enriching for full-length transcripts and maintaining the true representation of genes in the unamplified sample; these factors are critical for transcriptome sequencing and gene expression analysis. SMART technology is used to synthesize full-length cDNA by using the terminal transferase activity and template-switching ability of Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV RT): When the RT reaches the 5' end of the RNA template, it adds a few additional nucleotides to the 3' end of the cDNA. The carefully designed SMARTer II A oligo then base-pairs with these additional nucleotides, creating an extended template. The RT then switches templates and continues transcribing to the end of the oligonucleotide (**Figure 1**). The resulting full-length cDNA contains the complete 5' end of the mRNA, as well as an anchor sequence that serves as a universal priming site for second strand synthesis and amplification by LD-PCR (1).

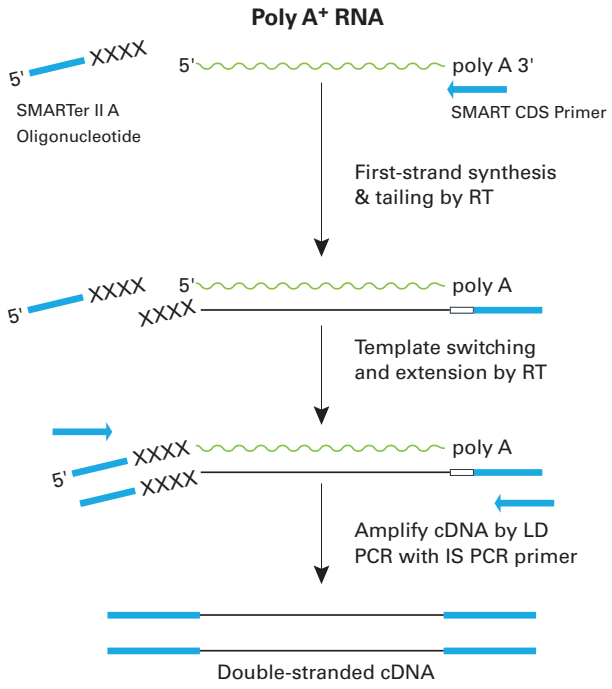


Figure 1. Flowchart of SMART cDNA Synthesis. When the reverse transcriptase (RT) reaches the 5' end of the mRNA template, it automatically adds a short tail of non-template nucleotides to the end of the newly synthesized cDNA strand. The carefully designed SMARTer II A Oligonucleotide anneals to the non-template nucleotides, creating an extended template for the RT. The RT then continues with first-strand synthesis until it reaches the end of the SMARTer oligo. Second-strand cDNA synthesis is accomplished by LD PCR using the IS PCR Primer, which contains the same 5' anchor sequence (blue) as the template-switching SMARTer oligo and the SMART CDS primer. The resulting full-length, double-stranded cDNA is then used for Illumina library preparation.

Clontech Laboratories, Inc.
A Takara Bio Company
www.clontech.com
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.543.6116

AN133852 IN

Clontech

SMARTer Ultra Low RNA Kit for Illumina Sequencing...continued

The cDNA produced by the kit is quickly and easily purified by immobilization onto polymer-coated, paramagnetic microparticles (Agencourt AMPure XP beads; Beckman Coulter) to remove reaction components and small DNA fragments (< 100 bp) without shearing the cDNA. The resulting purified cDNA has sufficient yield for Illumina library construction (**Figure 2**).

With the SMARTer Ultra Low RNA Kit for Illumina Sequencing, robust and reproducible cDNA synthesis is consistently achieved with as little as 100 pg of total RNA, representing approximately 5 cells' worth of RNA. Depending on many factors, including cell size, the amount of total RNA in a cell can be as little as 10 pg (see "How Much RNA is in a Single Cell?"), which is well below the levels needed for other RNA-Seq methods.

Sequence analysis of ultra-small amounts of RNA

The SMARTer Ultra Low RNA Kit provides highly reproducible results across a wide dynamic range, ensuring consistent quantification of both high- and low-abundance transcripts. When combined with Illumina sequencing, the system delivers dependable performance at extremely low RNA input levels, enabling the true diversity of gene expression within small cell populations to be explored. To demonstrate the exceptional performance of the system, libraries were made from mouse brain total RNA at input levels varying from 10 to 0.01 ng. After sequencing using one lane of Genome Analyzer *IIx* data per sample (~20 million reads/lane), the data were aligned to the mouse genome using the standard Illumina analysis pipeline. Even with just 10 pg of input RNA—which is an amount found in a typical single cell—over 90% of the data mapped to the genome, and the average transcript coverage was as uniform as that seen with much greater amounts of RNA (**Table I**, **Figure 3**). Also, under all conditions used, rRNA reads accounted for only 3–5% of the total reads, which is typical for standard poly(A)-selected library preparation methods. All of these results—high mapability, uniform read coverage, number of genes detected, and low amount of rRNA—are consistent with those typically achieved using much larger amounts of RNA.

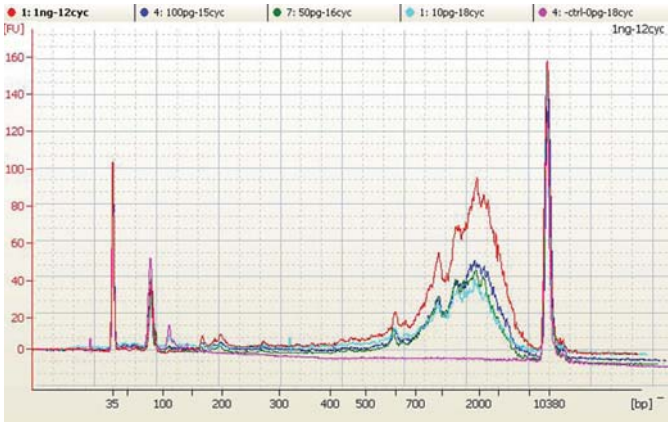
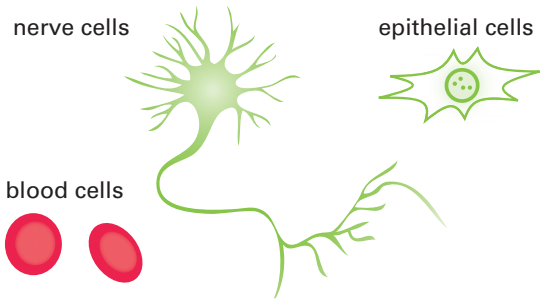


Figure 2. Electropherogram of amplified SMARTer cDNA. Various amounts of mouse brain total RNA were used as input for SMART cDNA synthesis. The cDNA samples were then analyzed for purity and yield on an Agilent 2100 Bioanalyzer. Shown are Bioanalyzer trace overlays of cDNA amplified from 1 ng (**red line**), 0.1 ng (**dark blue line**), 0.05 ng (**green line**), and 0.01 ng (**light blue line**) of total RNA and a no template control (NTC; **pink line**). The main peak indicates the purity and yield of cDNA between 0.4 and 9 kb—with the highest point at ~2 kb. There was no amplification in the negative control (pink line). Although the amount of input RNA can vary over quite a large range (e.g., 1 ng to 0.01 ng), comparable cDNA output can be obtained by adjusting the number of PCR cycles.

How Much RNA is in a Single Cell?



The amount of RNA in any given cell strongly depends upon the cell's size and metabolic state. In general, most cells in the body have 10 to 30 pg of RNA. The SMARTer Ultra Low RNA Kit for Illumina Sequencing allows researchers to generate robust, high-quality RNA sequencing results from just a few cells' worth of RNA.

SMARTer Ultra Low RNA Kit for Illumina Sequencing...continued

While this kit delivers highly accurate gene representation, ultra-low input samples (i.e., <100 pg RNA) yield results that are typical of ultra-low input RNA-Seq: a reduction in the number of genes detected (**Table I**) and increased variability (see scatter plots comparing assay reproducibility; **Figure 4**). These results are largely a consequence of the sampling noise associated with the capture and sequencing of so few RNA molecules. It has been suggested that an expression level of 1–3 RPKM (reads per kb of exon per million mapped reads) corresponds to a single transcript per cell (2). An RNA sample from thousands of cells contains thousands of copies of each mRNA molecule. In fact, standard RNA-Seq is able to easily detect mRNA molecules at less than one copy per cell in larger RNA samples (2), but at input levels of 2–5 cells, there is an expected decrease in the number of genes detected and the sampling noise associated with counting so few transcripts (see Table I).

Another contributing factor is likely to be the stochastic (i.e. random) cell-to-cell variability of gene expression, even among cells that are morphologically indistinguishable. This variability is not usually observed when looking at transcript levels in RNA obtained from a large population of cells (e.g., from 10 ng of mouse brain total RNA, which represents the RNA content of as many as 500 cells). Known as the "stochastic effect", this phenomenon has been noted in previous studies of gene expression in individual cells (3).

Gene expression differences in small populations of cells

Gene expression profiles are quite similar among replicates of large populations of cells (as represented in 10 ng of total RNA), but can vary greatly between small populations of cells. These differences are especially pronounced between single cells (2). Likewise, large samples of a particular tissue exhibit an average gene expression profile, even though individual cells or small populations of cells from

Table I. Primary sequencing metrics from very small amounts of total RNA

Sequencing Metric	10 ng	1 ng	0.1 ng	0.05 ng	0.01 ng
% Mapping to Genome	97%	96%	95%	94%	92%
% Ribosomal RNA	4.3%	5.0%	3.8%	3.9%	3.9%
Number of Genes Detected	16,610	15,425	12,847	11,516	8,618

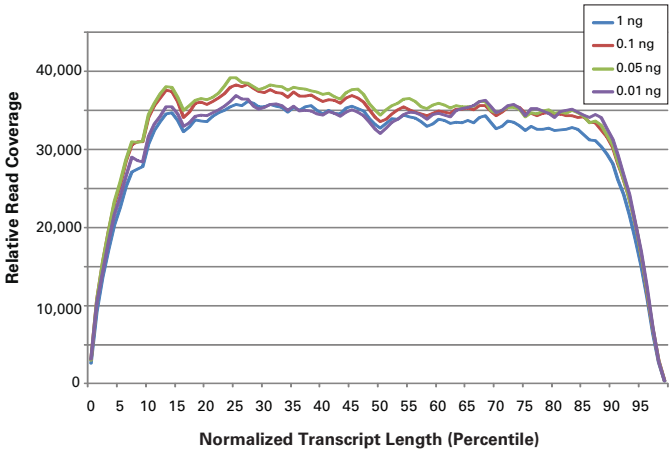


Figure 3. Comparison of transcript coverage with different amounts of input RNA. Shown are overlaid plots comparing the average read coverage from libraries made with 1 ng to 0.01 ng of mouse brain total RNA. The x-axis represents gene length normalized to 100%, where 0 is the 5'-end of each transcript and 100 is the 3'-end. The y-axis indicates the average coverage for a set of 724 genes that are moderately to highly expressed in brain tissue. The results are very consistent through the range of input RNA used, with full-length coverage of the transcripts reflecting no systematic 5'- or 3'-bias.

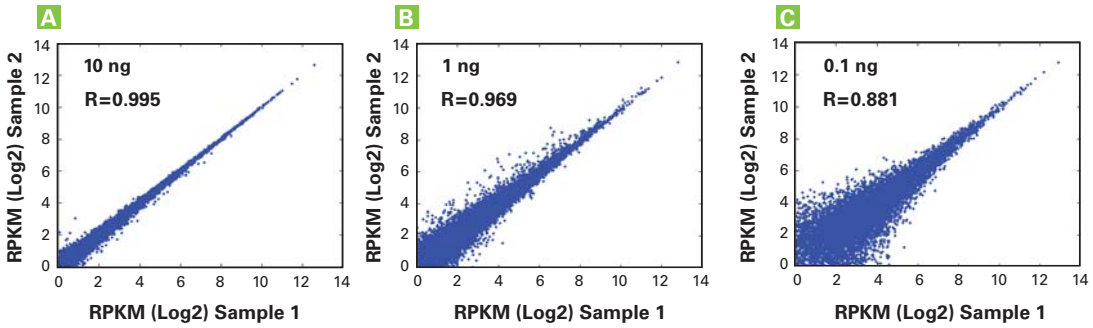


Figure 4. Reproducibility with various amounts of input RNA. Scatter plots were used to compare gene counts (i.e., log2 RPKM values) for replicate samples prepared using 10 ng, 1 ng, and 0.1 ng of mouse brain total RNA; these input levels represent the amount of RNA obtained from ~500, 50, and 5 cells, respectively (See "How Much RNA is in a Single Cell?"). When 10 ng of mouse brain total RNA was used (**Panel A**), the average gene counts were consistently reproduced, even at the lowest expression levels; however, when the amount of input RNA was decreased to 1 ng and 100 pg (**Panels B and C**, respectively), gene-count reproducibility decreased accordingly. RPKM = reads per kilobase of exon per million mapped reads (2).

SMARTer Ultra Low RNA Kit for Illumina Sequencing...continued

the same tissue have distinct profiles. The sensitivity exhibited by the SMARTer Ultra Low RNA Kit, combined with Illumina sequencing, allows researchers to accurately identify and measure transcript levels in very small amounts of RNA, providing higher confidence in gene expression and sequence data at lower input levels than ever before.

SMART template-switching technology enables the accurate synthesis of full-length cDNA in a single reverse transcription reaction. As a result, SMART technology provides faithfully reproduced, full-length cDNA for use as template in library sample preparation. The resulting exon coverage is equivalent to traditional RNA-seq methods requiring significantly more starting material (**Figures 5 and 6**).

The new SMARTer Ultra Low RNA Kit for Illumina sequencing allows the construction of quality libraries for sequencing, even with single-cell levels of total RNA. Even at ultra-low input levels, the kit produces sequencing data of extremely high-quality—good mapping, uniform transcript coverage, low rRNA, and accurate differential expression measurements—which allows researchers to investigate gene expression levels for biological processes that, until now, have been previously inaccessible to researchers who use RNA-Seq.

Summary

The SMARTer Ultra Low RNA protocol produces accurate, full-length cDNA that, when combined with Illumina sequencing, allows robust library generation, reproducible, high-quality sequence data, and accurate gene quantification from picograms of total RNA. The system provides researchers access to valuable samples with limited amounts of starting material, originating from sources such as laser capture microdissection, cell isolation technologies, and fine-needle biopsies. This method would be greatly beneficial for understanding the key biological processes involved in cancer, development, stem cells, and brain function.

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. Mortazavi, A., *et al.* (2010) *Nat Methods* **5**(7):621–628.
3. Janes, K. A., *et al.* (2010) *Nat Methods* **7**(4):311–317.
4. MAQC Consortium (2006) *Nat Biotechnol.* **24**(98):1151–1161.

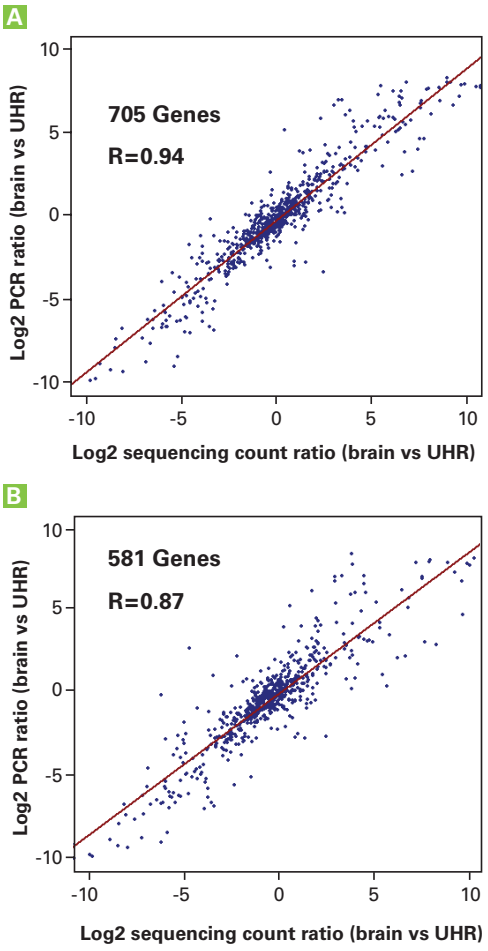


Figure 5. Gene expression data obtained from very low amounts of RNA correlate well with data obtained by qPCR. Scatter plots were used to compare differential expression data obtained by sequencing with the SMARTer Ultra Low RNA Kit (1ng and 0.1 ng total RNA input) and quantitative PCR (qPCR) data available for Universal Human Reference Total RNA (UHR) and Human Brain Reference RNA through the MicroArray Quality Control (MAQC) project (4). The differential expression of ~600-700 genes showed correlation values of 0.94 (1 ng; **Panel A**) and 0.87 (0.1 ng; **Panel B**), demonstrating that the sequencing results are consistent with orthogonal gene expression technologies.

SMARTer Ultra Low RNA Kit for Illumina Sequencing...continued

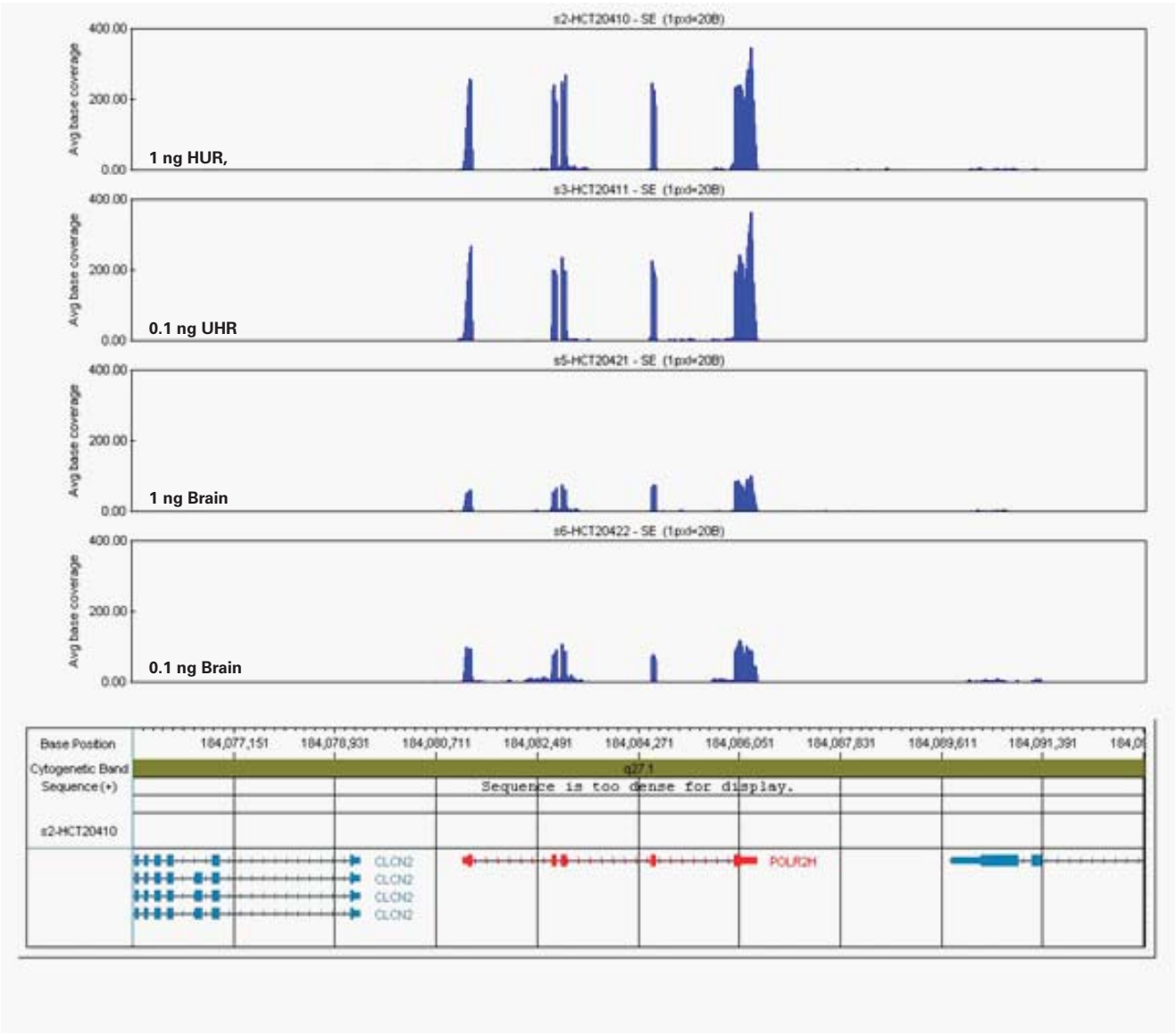


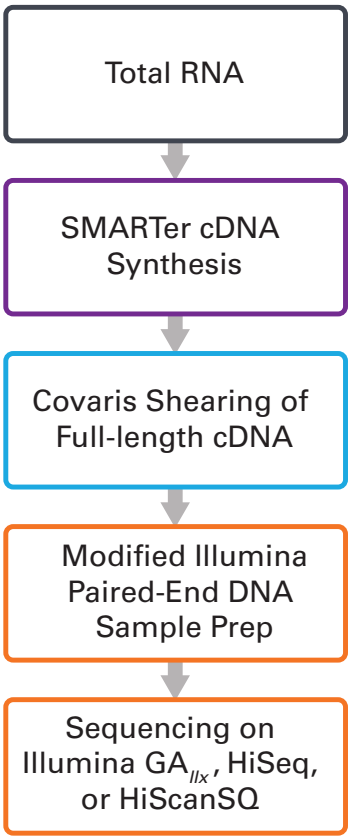
Figure 6. Quantitative, full-length coverage of expressed transcripts. Shown are sequencing data from UHR and human brain RNA samples (1 ng and 0.1 ng each) aligned at the POLR2H gene. As with standard mRNA-Seq, the data were aligned to the human genome and the reads mapped precisely to the exon regions of the gene. This data shows that POLR2H is more highly represented in UHR than in Brain RNA, and that the data are very similar at both input levels.

SMARTer Ultra Low RNA Kit for Illumina Sequencing

Two powerful technologies combine to enable sequencing with ultra-low levels of RNA

SMARTer Ultra Low RNA Kit and Illumina Sequencing Workflow.

- **Efficient: single-tube protocol integrated with powerful gene expression analysis by Illumina sequencing**
- **Sensitive: cDNA sample preparation from 100 pg of RNA**
- **Robust: high reproducibility between samples**
- **Accurate: complete coverage of the transcriptome**



Clontech's SMART technology and Illumina sequencing combine to provide the highest quality RNA-sequencing results from ultra-low levels of RNA. First, the SMARTer Ultra Low RNA protocol is used to generate high-quality cDNA from picogram quantities of total RNA. The resulting ds cDNA is then used in a modified Illumina paired-end library construction protocol to create a robust, high-quality library that is ready to be sequenced on any Illumina platform to provide the highest quality results.

The resulting library provides high mapability, uniform read coverage, a consistently high number of genes detected, and low amounts of rRNA, all of which are consistent with results typically achieved using much larger amounts of RNA.

Visit the Clontech product page at: **www.clontech.com/smartseq**

Clontech Products	Size	Cat. No.
SMARTer Ultra Low RNA Kit for Illumina Sequencing	10 rxns	634935

Illumina Products	Size	Cat. No.
1 Paired End DNA Sample Prep Kit	10 preps	PE-102-1001
4 Paired End DNA Sample Prep Kits	40 preps	PE-102-1002

Clontech