SureSelect strand specific RNA library prep kit provides a fast and streamlined workflow for preparing directional libraries from total RNA

Bahram Arezi[1], Bilan Hsue[1], Francisco Useche[2], Anya Tsalenko[2], Barbara Novak[2], Anne Lucas[2], Keith Chen[1], Huong Tang[1], Alex Siebold[2], Alex Wong[2], D. Neil Hayes[3], David A. Eberhard[3], Melissa Flowers[3], Holly Hogrefe[1]
[1]Agilent Technologies, la Jolla, CA, 92037, USA
[2]Agilent Technologies, Santa Clara, CA, 95051, USA
[3]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Agilent Technologies

Abstract

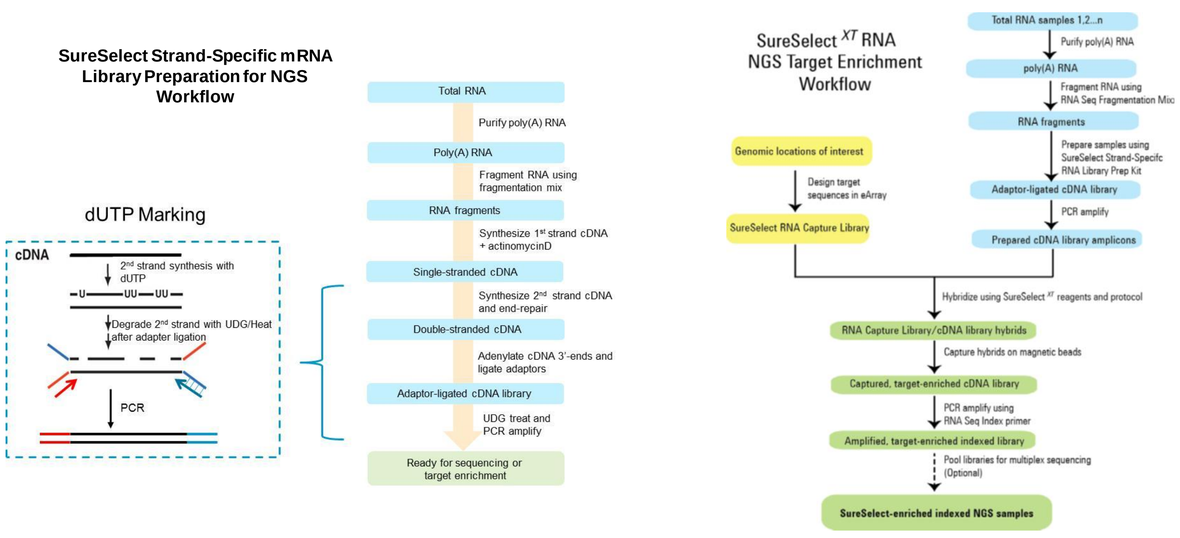
RNA-Seq is a revolutionary technology for whole transcriptome analysis. Deep sequencing of cDNA has been used to quantify transcript levels, confirm gene annotation, and identify novel transcripts, splice variants, and SNPs. RNA-Seq protocols that preserve strand information are critical for identifying antisense transcripts that play a role in gene regulation, determining the exact boundaries of genes transcribed on opposite strands, and accurately measuring expression levels of overlapping genes. Here, we present the SureSelect Strand Specific RNA Library Prep Kit, which includes all reagents required for isolating polyA RNA and preparing directional RNA libraries for Illumina sequencing from as low as 50ng total RNA, using the dUTP marking method (Parkomchuk et al (09) NAR 37: e123). In a comprehensive comparison of directional RNA-Seq methods, the dUTP marking method was found to provide the most compelling overall balance across all sequencing metrics examined, including strand specificity, library complexity, evenness and continuity of coverage, and accuracy of gene expression profiling (Levin et al (10) Nature Methods 7: 709-715). With our streamlined automation-friendly protocol, which includes master mixes and combined enzymatic steps, researchers can prepare directional cDNA libraries with >99% strand specificity from total RNA in about 5 hours, without the need for gel size selection. A comparison of RNA-Seq to gene expression array shows good correlation (R>0.8) between differential gene expression ratios (MAQCA and B) obtained using Agilent's SureSelect Strand Specific RNA Library Prep kit and SurePrint G3 Human Gene Expression Microarray 8x60K. Finally, we show that SureSelect Strand Specific RNA Library Prep is compatible with target enrichment by performing capture using the SureSelect Human Kinome bait library. Enriched libraries show high target and strand specificity (80% on-target reads; >99% strand specificity), and correlation of mean gene RPKM values between target-enriched and un-enriched libraries (R>0.8).

Methods/Results

Part 1: Workflows for SureSelect Strand-Specific mRNA NGS and SureSelect^{XT} RNA NGS Target Enrichment

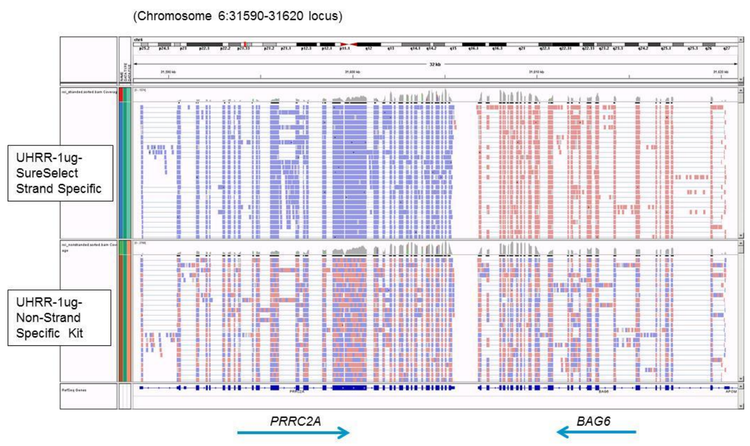
Key Features:

1. Paired-end multiplexed sequencing on Illumina platforms.
2. Whole mRNA analysis with total RNA input from 0.05 to 4 µg. The kit includes oligo(dT) beads for polyA-RNA isolation.
3. Strand-specific method employing dUTP marking of the second strand cDNA.
4. Streamlined protocol featuring master mixes and combined enzymatic steps.
5. Optimal target enrichment module including up to 48 indexes. It is the only kit on the market that can be used for both whole mRNA and target enrichment sequencing. For target enrichment, the index is added post capture (similar to SureSelect^{XT}).
6. Automated protocols are available for Agilent NGS Workstation.



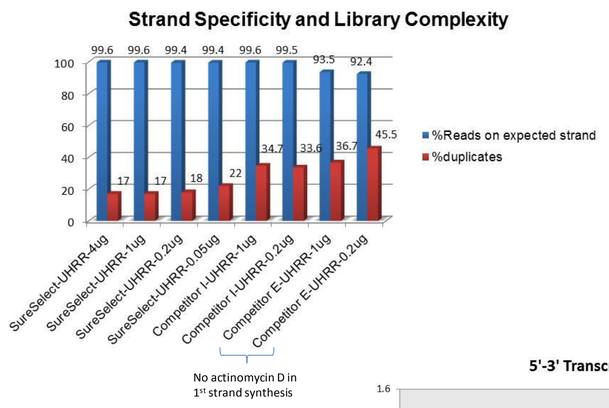
Part 2: Visualizing Strand-Specific sequencing using IGV:

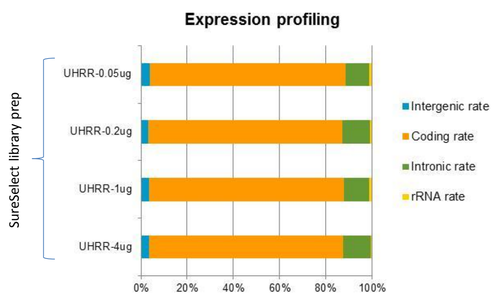
In order to visualize strand specificity, we used Integrative Genomic Viewer (Broad) to examine an RNA library that we generated using SureSelect Strand-Specific mRNA library prep and a non-strand specific library prep kit, using Universal Human Reference total RNA (UHRR) as input. Reads mapping to the positive strand are shown in orange and reads mapping to the negative strand are shown in blue. Arrows indicate the direction of transcription for each RefSeq gene.

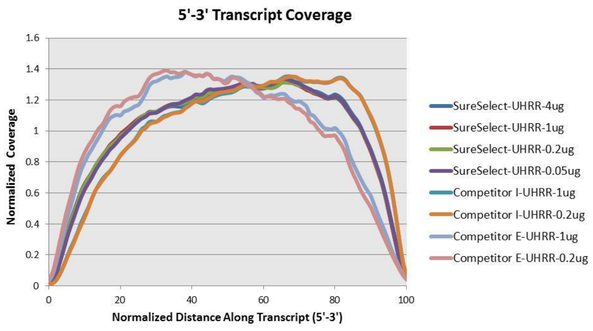


Part 3: Whole mRNA Sequencing using SureSelect Strand Specific and competitor kits:

SureSelect strand specific RNA library prep kit is optimized for total RNA input ranging from 4 ug to 50 ng. In order to determine kit performance, we used UHRR (MAQC-A, RIN=8.7) as input. We also used two different commercially available directional RNA preparation kits for these comparisons (Competitor I and Competitor E). Poly(A) purification was performed for all samples before library preparation. All reads were normalized to 20 million/library (2X100bp sequencing) for comparison purposes. Intragenic rate refers to the fraction of reads that map within genes (within introns or exons). Exonic rate is the fraction mapping within exons. Intronic rate is the fraction mapping within introns.



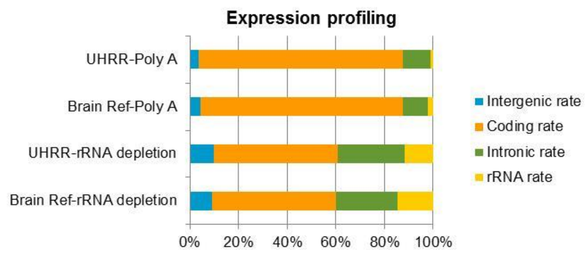


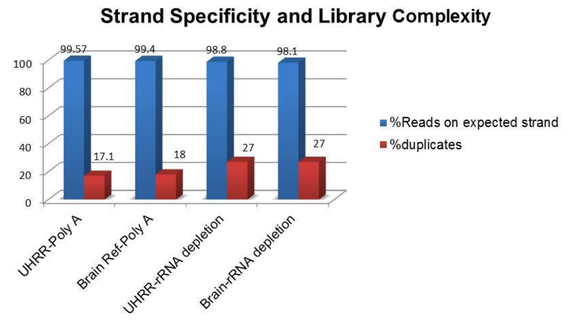


Methods/Results

Part 4: Poly(A) selection versus rRNA depletion before library prep:

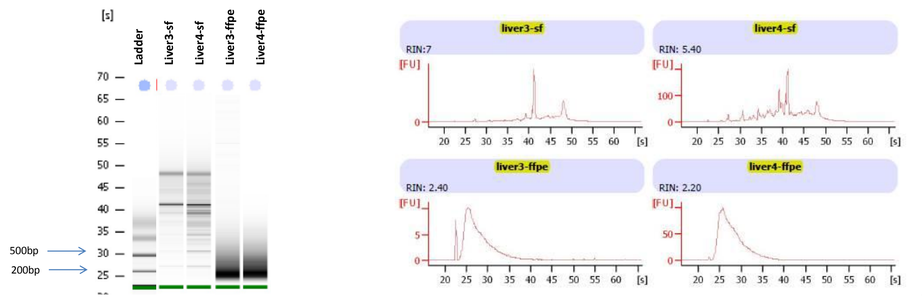
We looked at the effect of using rRNA depletion instead of poly(A) selection prior to library prep. 1ug of UHRR (MAQC-A; RIN=8.7) and Ambion's brain reference RNA (MAQC-B; RIN=7.4) were used as input. All libraries were prepared using SureSelect Strand Specific Library prep kit. All reads were normalized to 20 million/library (2X100bp sequencing) for comparison purposes.

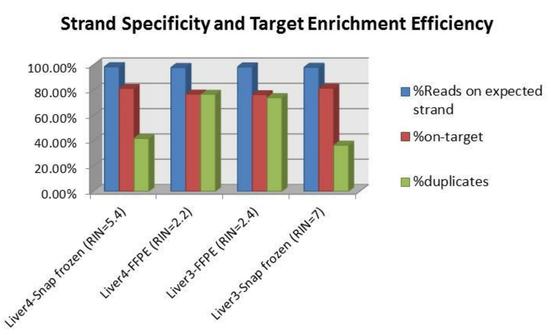


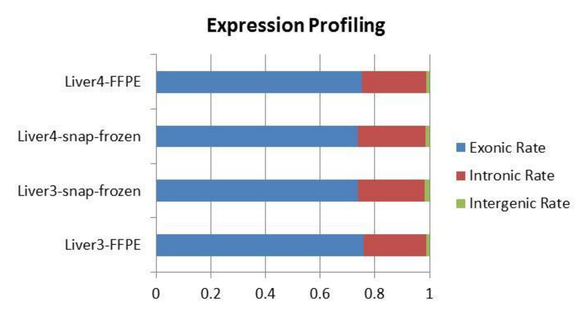


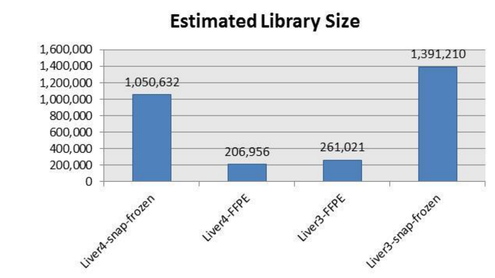
Part 5: Library preparation from FFPE samples and target-enriched sequencing:

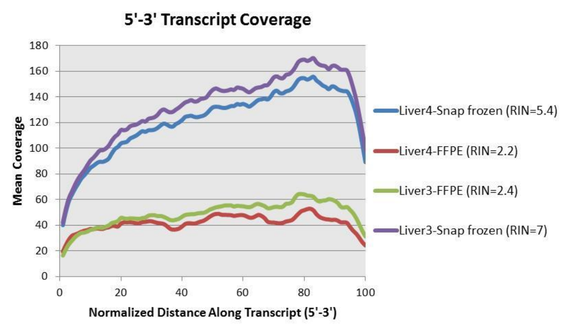
SureSelect strand specific RNA library prep kit was used in order to prepare libraries from highly degraded FFPE samples and their snap-frozen counterparts. Slight adjustments were made to our standard protocol in order to prep from FFPE samples. 4ug of FFPE and 2ug of snap-frozen samples were used as input. A SureSelect bait library of 51229 probes at 2.88Mb was used for all captures. Number of reads were normalized to 2.88 million/library (2 x 100bp sequencing) for comparison purposes. Estimated library size is the number of expected unique fragments based upon the total number of reads and duplication rate assuming a Poisson distribution. Mean per base coverage is averaged across each transcript and averaged again across all top 1000 expressed transcripts





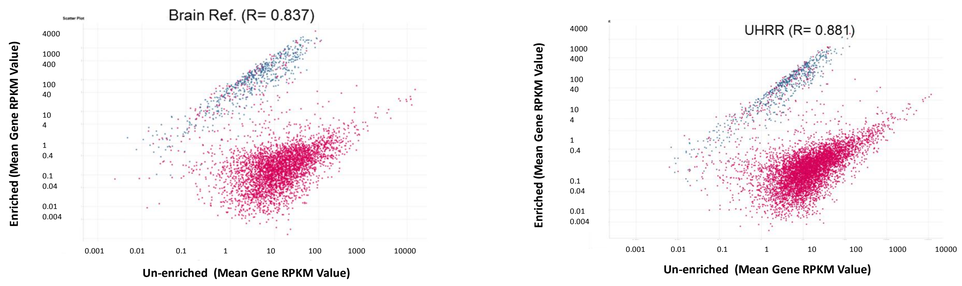






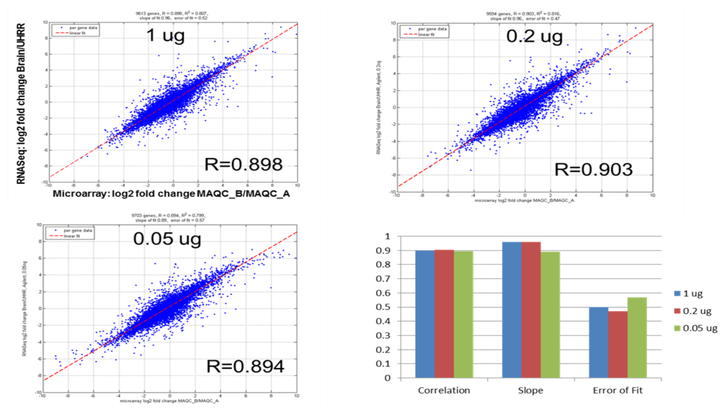
Part 6: Correlation between target-enriched and un-enriched libraries:

We looked at the correlation between enriched and un-enriched libraries using UHRR and brain reference RNA. All captures were performed using the 2.9 Mb Human RNA kinome (612 genes). Genome Studio (Illumina) was used to create files with normalized reads using gene names. Normalized count (RPKM) equals (raw count x read length) / (feature length x number of mapped reads in millions). The scatter plots compare normalized gene counts from a Human RNA Kinome enriched sample vs. un-enriched sample. Blue dots represent targeted genes and red dots represent all other genes. Both the enriched UHRR and brain reference RNA maintain a good correlation with their un-enriched samples (R>0.8).



Part 7: Correlation of RNA sequencing and microarray data:

Different total RNA (UHRR: MAQC-A and brain reference: MAQC-B) input amounts ranging from 50ng to 1ug were used to prep libraries for RNA sequencing. SurePrint G3 Human Gene Expression 8x60K microarray was used to generate microarray data. In Figure below, the x-axis represents the log2 fold change of MAQC_B over MAQC_A for the microarray data, while the y-axis represents the log2 fold change of MAQC-B over MAQC-A for the RNA sequencing data. The R values of ~0.9 indicate that the RNA sequencing data maintains a good correlation with the microarray data.



Conclusions

SureSelect Strand Specific RNA Library Prep Kit provides easy-to-use master mixes and a streamlined protocol for customers. It is the only kit on the market that can be used for strand specific whole mRNA sequencing as well target enrichment. Based on the sequencing results, we show that SureSelect Strand Specific RNA Library Pre Kit outperforms the competition by by generating significantly higher library complexity. Furthermore, Competitor I shows slightly higher 3'-bias and Competitor E exhibits slightly higher 5'-bias compared to SureSelect strand specific RNA kit. We also show excellent correlation between enriched and un-enriched libraries using the 2.9 Mb RNA kinome library, and a linear correlation between RNA sequencing and microarray data using Agilent SurePrint G3 Human Gene Expression 8x60K microarray. With a few minor adjustments to the protocol, SureSelect Strand Specific kit can be successfully used to generate libraries from highly degraded FFPE samples.