LEXOGEN

Enabling complete transcriptome sequencing

SENSE mRNA-Seq V2: A fast and automatable RNA-Seq library preparation protocol with superior strand-specificity

The SENSE mRNA-Seq Library Preparation Kit generates ready-to-sequence libraries from low amounts of total RNA in less than 5 hours. SENSE libraries are significantly more strand-specific (>99.9 %) than leading alternatives and are virtually free of rRNA contamination. The protocol does not rely on RNA or cDNA fragmentation and doesn't require additional kits for rRNA depletion, library amplification, or size selection.

Introduction

With the rapid development of Next Generation Sequencing (NGS) technologies, RNA-Seq has become the method of choice for transcriptome analysis. Most NGS platforms require library construction prior to sequencing, and several current mRNA-Seq library preparation protocols conserve the strand-orientation of transcripts. This allows the specific assignment of reads to one strand of the genome and enables the discovery and quantification of antisense transcripts as well as overlapping genes transcribed from opposite strands.

However, current RNA-Seq protocols do not provide sufficient strand-specificity; reads mapping to both strands are generated during library preparation through a number of mechanisms, and this background noise obscures the detection of antisense transcripts.

SENSE provides unprecedented strand-specificity, allowing the detection and accurate quantification of antisense transcripts and overlapping genes by minimizing false-positive reads.

The SENSE Workflow

SENSE is a complete total RNA-to-sequencer solution that requires no additional kits (Fig. 1). It contains an integrated poly(A) selection on magnetic beads which outperforms existing methods. Libraries are then generated with a random priming approach using Lexogen's strand displacement stop/ligation technology.

In a single-tube reaction, starter/stopper heterodimers containing platform-specific linkers are hybridized to the mRNA, where the starters serve as primers for reverse transcription. Reverse transcription terminates upon reaching the stopper from the next heterodimer, at which point the newly-synthesized cDNA and the stopper are ligated while still bound to the RNA template. No time-consuming fragmentation step is required, eliminating the need for mechanical shearing or additional enzymes. Library size is determined by the protocol itself, not with *post hoc* size selection. Libraries can be sequenced with standard single-end or paired-end reagents. External barcodes are included in all SENSE kits, allowing multiplexing of up to 96 libraries. SENSE uses magnetic bead-based purifications rendering it compatible with automation.

One person can prepare 16 sequence-ready libraries starting from total RNA within 5 hours with minimal equipment and no additional kits. In addition to SENSE mRNA-Seq Lexogen also offers SENSE Total RNA-Seq (Cat. No. 009) for rRNA-depleted RNA input if non-coding RNA is of particular interest.

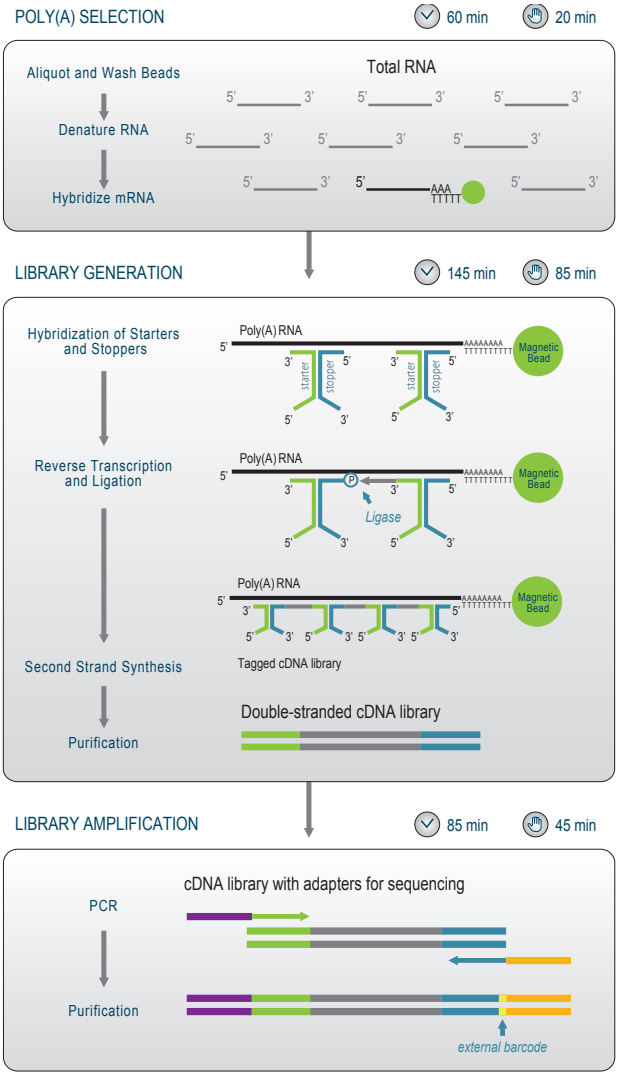

Figure 1 I Overview of the SENSE protocol

SENSE mRNA-Seq V2 vs standard mRNA-Seq

We generated multiplexed SENSE libraries from U.S. Food and Drug Administration (FDA) Sequencing Quality Control (SEQC) standard samples A and B in three technical replicates. Sample A is a mixture of Universal Human Reference RNA (UHRR) and External RNA Controls Consortium (ERCC) spike-in control mix 1. Sample B is a mixture of Human Brain Reference RNA (HBRR) and ERCC spike-in control mix 2 (SEQC samples A and B were received from the FDA prepared according to the FDA/National Center for Toxicological Research SEQC RNA Sample Preparation and Testing SOP_20110804). Libraries were sequenced on one lane of an upgraded Illumina® HiSeq 2000 using 125 bp paired-end (PE125) reagents (V4 chemistry). Pass filter reads (278 million) were de-multiplexed and mapped to the human and ERCC reference genomes. Cytoplasmic rRNA content was very low, indicating efficient poly(A) mRNA selection **(Tab. 1)**.

The mRNA-Seq data sets were made available by a laboratory that participated in the recently published Association of Biomolecular Resource Facilities (ABRF) NGS study[2]. In that study, the researchers performed a stranded RNA-Seq library preparation with poly(A) enrichment in 2 technical triplicates, obtaining 50 bp paired-end (PE50) reads on an Illumina HiSeq 2000 (ref. 2; from the GSE48035 data set samples SRR903178–80 from GSM1166109 and SRR903210–12 from GSM1166113 were used in this comparison).

The ERCC spike-in transcripts allow the accurate calculation of strandedness as all antisense ERCC reads can be considered false positives introduced during library preparation. In contrast, genome-wide calculations of strandedness are conflated by true antisense transcription[2]. Strand-specificity was therefore calculated based on ERCC data only, and was exceptionally high for SENSE (>99.9 %) whereas the standard stranded mRNA-Seq protocol only reached 93.3 % - 97.8 % strandedness **(Tab. 1)**. A high and consistent strandedness is extremely crucial for the detection of true antisense transcripts. One of the primary causes of antisense background and low strand-specificity is spurious second strand synthesis during reverse transcription. Lexogen's strand displacement stop/ligation technology effectively suppresses this background reaction, providing SENSE libraries with the exceptionally high strand-specificity required for the detection and accurate quantification of antisense transcripts. SENSE includes a highly specific poly(A) selection as demonstrated by the low number of rRNA reads (maximal 0.0004 %).

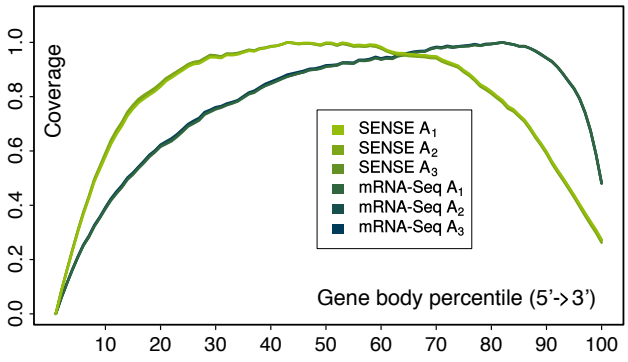

Figure 2 | Gene Body Coverage: Standard mRNA-Seq shows a 3' bias and a lower coverage at the 5' end. SENSE mRNA-Seq V2 shows a much more homogenous coverage across the entire transcript.

Another crucial quality parameter for RNA-Seq analysis is gene-body coverage. Ideally, reads are evenly distributed across the entire gene-body. Most RNA-Seq methods show decreased coverage at the 3' and 5' ends of transcripts. We could show that SENSE mRNA-Seq V2 provides a better gene body coverage than standard mRNA-Seq, with a better coverage of the 5' ends and less bias at the 3' ends. SENSE mRNA-Seq V2 achieves this better coverage as it is a fragmentation-free protocol and Starter/Stopper Heterodimers are hybridized across the entire transcript length **(Fig. 2)**.

Conclusions

The SENSE mRNA-Seq V2 library preparation protocol is a fast all-in-one solution for the production of strand-specific mRNA-Seq libraries starting from total RNA. The integrated poly(A) selection virtually eliminates rRNA contamination without relying on additional selection or depletion protocols. SENSE libraries exhibit exceptional strand-specificity, reduce experimental noise and empower the detection of antisense transcription. Additionally, SENSE provides a better gene-body coverage than standard mRNA-Seq.

1. Baker, S. *et al.* (2005). External RNA Controls Consortium: a progress report. *Nat. Methods* 2, 731-734
2. Jiang, L. *et al.* (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* 21, 1543-1551
3. Li, S. et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. Nat. Biotechnol. 32, 915–925 (2014).

Lexogen GmbH · Campus Vienna Biocenter 5 · 1030 Vienna · Austria

Find more about SENSE at www.lexogen.com.
Contact us at info@lexogen.com or +43 1 345 1212-41.

Table 1 | Summary of sequencing results

	SENSE mRNA-Seq V2					Standard mRNA-Seq				
500 ng Total RNA	Number of Reads	Strand-specificity[a]	% uniquely mapping Reads	% rRNA Reads	% Protein Coding	Number of Reads	Strand-specificity[a]	% uniquely mapping Reads	% rRNA Reads	% Protein Coding
FDA-A_1	48,353,384	99.981 %	89.11 %	0.0004 %	94.12 %	5,303,785	93.567 %	86.96 %	0.0003 %	94.93 %
FDA-A_2	38,011,481	99.989 %	89.03 %	0.0003 %	94.24 %	5,406,559	93.348 %	86.89 %	0.0003 %	94.94 %
FDA-A_3	45,786,988	99.995 %	88.85 %	0.0002 %	94.26 %	5,067,031	93.505 %	86.83 %	0.0003 %	94.89 %
FDA-B_1	58,183,338	99.990 %	89.00 %	0.0002 %	93.28 %	6,288,544	97.818 %	91.53 %	0.0002 %	92.58 %
FDA-B_2	49,813,976	99.994 %	89.99 %	0.0001 %	93.37 %	6,266,933	97.840 %	91.48 %	0.0002 %	92.61 %
FDA-B_3	38,189,107	99.984 %	89.68 %	0.0002 %	93.23 %	5,897,307	97.839 %	91.45 %	0.0003 %	92.57 %

[a] number of reads mapping to ERCC genes in the sense direction divided by total number of reads mapped to the ERCC genome.