LEXOGEN
Enabling complete transcriptome sequencing

SENSE: a fast 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 within 4 h. 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 or additional kits for rRNA depletion, library amplification or size selection.

Introduction

With the rapid development of next-generation sequencing technologies, RNA-seq has become the method of choice for transcriptome analysis. Most next-generation sequencing 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 by minimizing false-positive reads.

The SENSE workflow

SENSE is a complete RNA-to-sequencer solution that requires no additional kits (**Fig. 1**). It contains an integrated poly(A) selection on magnetic beads that 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. SENSE supports multiplexing with in-line barcodes that can be introduced with no additional effort and do not require a third barcode-specific sequencing read. One person can prepare 8 sequence-ready libraries starting from total RNA within 4 h with minimal equipment and no additional kits.
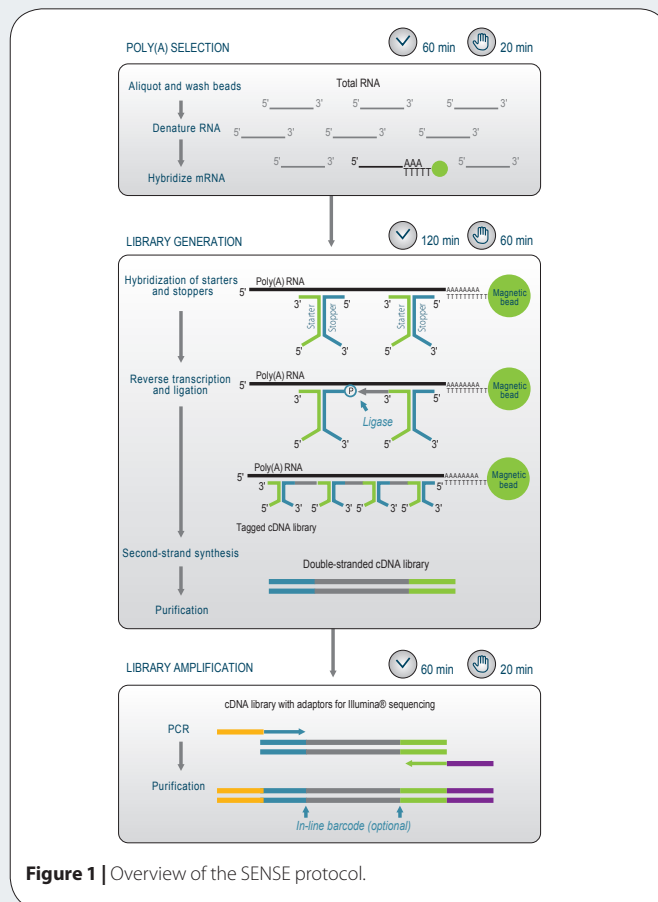

Figure 1 | Overview of the SENSE protocol.

Courtney Nadeau & Alexander Seitz

Lexogen GmbH, Vienna, Austria. Correspondence should be addressed to A.S. (alexander.seitz@lexogen.com).

APPLICATION NOTES

Table 1 | Summary of sequencing results.

Total RNA	Strand specificity[a]	False antisense reads[b]	Sense reads (genome-wide)[c]	rRNA content[d]
2 µg	99.989±0.008%	0.011±0.008%	99.888±0.002%	0.00014±0.00003%
1 µg	99.990±0.004%	0.010±0.004%	99.832±0.017%	0.00036±0.000005%
500 ng	99.996±0.001%	0.004±0.001%	99.857±0.036%	0.00027±0.00002%
50 ng	99.965%	0.035%	99.779%	0.00044%

[a]Number of reads mapping to ERCC genes in the sense direction divided by total number of reads mapped to the ERCC genome. [b]Number of antisense reads mapping to ERCC transcripts divided by the total number of reads mapped to the ERCC genome. [c]Number of reads mapping to annotated genes in the sense orientation divided by reads mapping in both directions. [d]Total number of reads mapping to rRNA genes divided by the number of reads mapped to any annotation.

Experiment

We generated multiplexed SENSE libraries from various amounts of Universal Human Reference RNA (Agilent Technologies) with ERCC spike-in RNA controls[1] (Ambion Inc.) and sequenced these on an Illumina® HiSeq using 100 bp single-end reagents. Pass filter reads (178 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 (**Table 1**).

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 with all amounts of input RNA (**Table 1**).

With the 99.99% strand specificity determined from ERCC data, we detected 2,904 antisense transcripts in the pooled dataset. As the abundance of a sense transcript is often several orders of magnitude higher than its corresponding antisense, signal from actual antisense transcripts can easily drown in false-positive noise generated due to low strand specificity. To examine the relationship between strandedness and antisense detection, we implied lower levels of strand specificity to the data and recalculated the number of antisense transcripts detected (**Fig. 2**).

Decreasing strand specificity from 99.99% to 99% greatly reduced sensitivity and only half as many (1,506) antisense transcripts could be detected. The large loss of sensitivity with even 99% directionality highlights the extreme importance of maintaining high strand specificity when analyzing antisense gene expression.

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.

Conclusions

The SENSE mRNA-seq 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
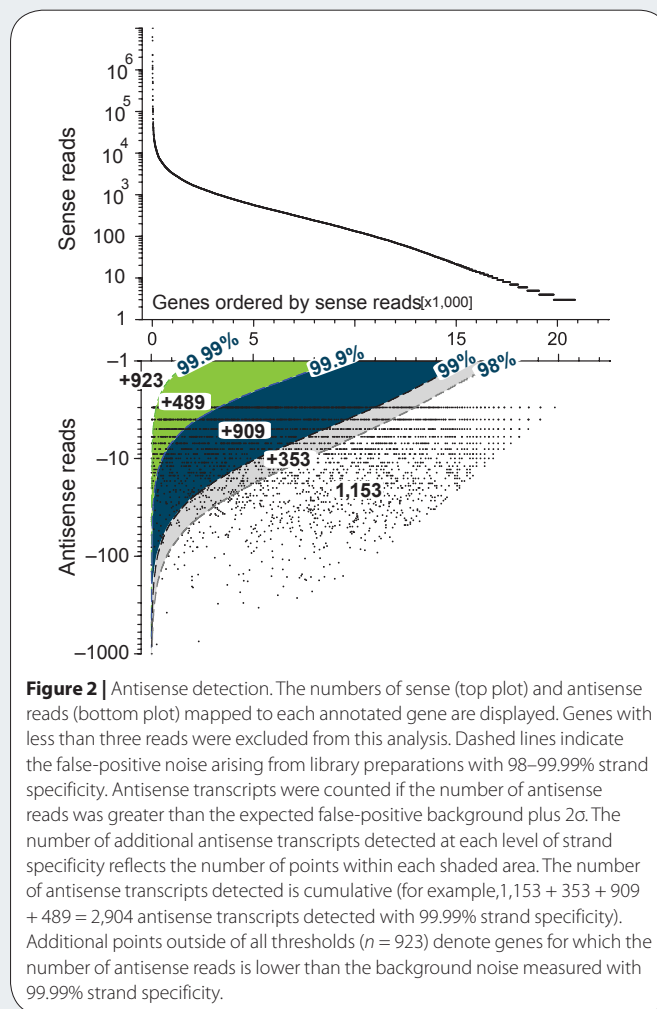

Figure 2 | Antisense detection. The numbers of sense (top plot) and antisense reads (bottom plot) mapped to each annotated gene are displayed. Genes with less than three reads were excluded from this analysis. Dashed lines indicate the false-positive noise arising from library preparations with 98–99.99% strand specificity. Antisense transcripts were counted if the number of antisense reads was greater than the expected false-positive background plus 2σ. The number of additional antisense transcripts detected at each level of strand specificity reflects the number of points within each shaded area. The number of antisense transcripts detected is cumulative (for example,1,153 + 353 + 909 + 489 = 2,904 antisense transcripts detected with 99.99% strand specificity). Additional points outside of all thresholds (*n* = 923) denote genes for which the number of antisense reads is lower than the background noise measured with 99.99% strand specificity.

specificity, reduce experimental noise and empower the detection of antisense transcription.

1.   Baker, S. *et al*. External RNA Controls Consortium: a progress report. *Nat. Methods* **2**, 731–734 (2005).
2.   Jiang, L. *et al*. Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).