diagenode
Innovating Epigenetics Solutions

TECHNICAL NOTE

Low input RNA-seq library preparation provides higher small non-coding RNA diversity and greatly reduced hands-on time

INTRODUCTION

RNA-seq is a next-generation sequencing technique that aims for the identification and quantification of RNA in biological samples. RNA-seq is advantageous over many alternative methods as it allows the detection of undetermined genomic sequences, transcription boundaries with single-base resolution, and sequence variations. Furthermore, RNA-seq has a large dynamic range for quantifying gene expression and has been shown to be highly sensitive and reproducible.

However, prior to sequencing, libraries that are suitable for NGS platforms need to be prepared. The library preparation typically requires the conversion of the template RNA into cDNA, the fusion of adapters to both extremities, and PCR amplification to generate enough material.

One of the main concerns for small RNA-seq is that the addition of the adapters by RNA ligase is heavily biased because of its uneven affinity towards the different template sequences.

In addition, ligation-based techniques necessitate large amounts of RNA input and a number of time-consuming purification steps, posing a challenge for low input samples, degraded samples, and circulating RNAs.

To remedy this problem, we have developed a solution utilizing the straightforward, cost-effective, highly efficient protocol "Capture and Amplification by Tailing and Switching" (CATS) to easily generate ready-to-sequence NGS libraries from picogram quantities of RNA. This ligation-free library preparation protocol only requires 5 hours and is fully compatible with automation platforms thanks to its simple bead-based purification of the final product. The combination of all these features makes CATS an ideal method for more accurate representation of a sample transcriptome. The CATS method is particularly relevant for challenging samples such as FFPE tissue and liquid biopsy, making it an excellent tool for biomarker discovery.

METHODS

CATS was compared to two commercially and well-established library preparation methods: NEBNext® Small RNA Library Prep Set for Illumina® (New England Biolabs) and SMARTer® smRNA-seq Kit for Illumina® (Takara). The template used in this comparison was human universal total RNA (Agilent, 740000). Both total RNA and the isolated small RNA fraction (< 200nt) were used. However, different inputs were chosen in line with the protocol recommendations (see Table 1).

Table 1. Total RNA, small RNA inputs used for comparison.

	NEBNext®	SMARTer®
Total RNA	100 ng	1 ng
Isolated small RNA (< 200nt)	1 ng	100 pg

The small RNA fraction was extracted from the total RNA with the miRNeasy kit (Qiagen) following the manufacturer's instructions. Afterwards, library preparation was carried out following protocol instructions for every kit. The libraries were quality checked using a Qubit 2.0 with HS dsDNA kit (ThermoFisher) and a BioAnalyzer 2100 with HS DNA kit (Agilent). The single end-sequencing was carried out on a HiSeq 2500 (Illumina®) sequencer with 50 bp reads.

CATS WORKFLOW

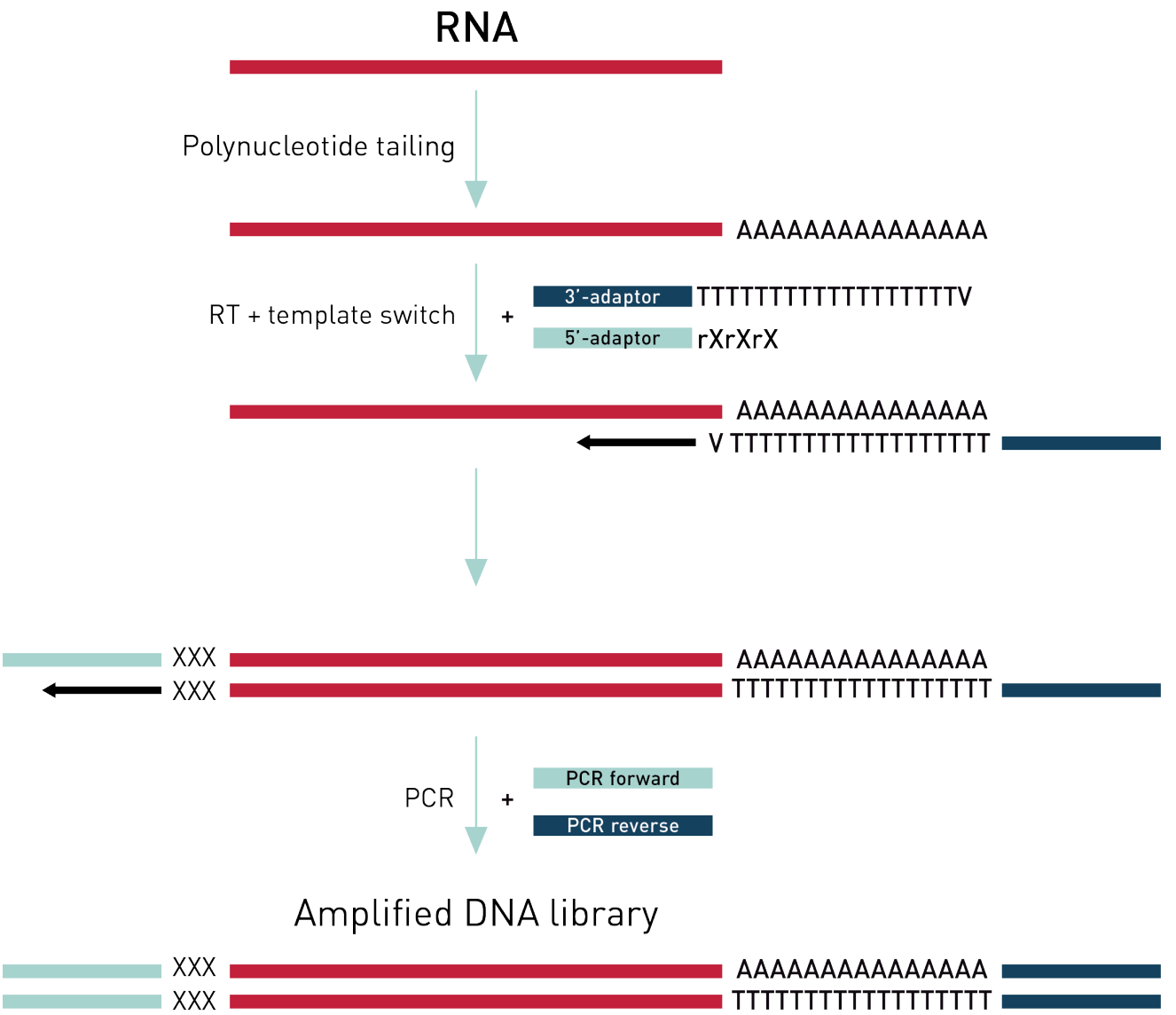

Figure 1: Schematic representation of the workflow used by the CATS Small RNA-seq library preparation. Single stranded RNAs are first dephosphorylated (end-repaired) and polyadenylated at the 3'-end. Subsequently, a cDNA strand synthesis is performed in the presence of the anchored poly(dT) oligonucleotide containing terminal P7 Illumina® adapter sequence. When the reverse transcriptase reaches the 5'-end of the RNA it switches the template and continues DNA synthesis over the template-switching oligonucleotide (TSO). The TSO contains three 3'-terminal ribonucleotides X (rX) which facilitate the template switching and carry the terminal P5 Illumina® adapter sequence. During PCR pre-amplification of the first cDNA strand, Illumina® adapters carrying P5 and P7 terminal sequences (required for clustering on an Illumina® flow cell) as well as index sequences are incorporated into the library. The sum size of the adapters (the size of "empty" library) is 143 bp.

RESULTS

CATS - a highly robust technology

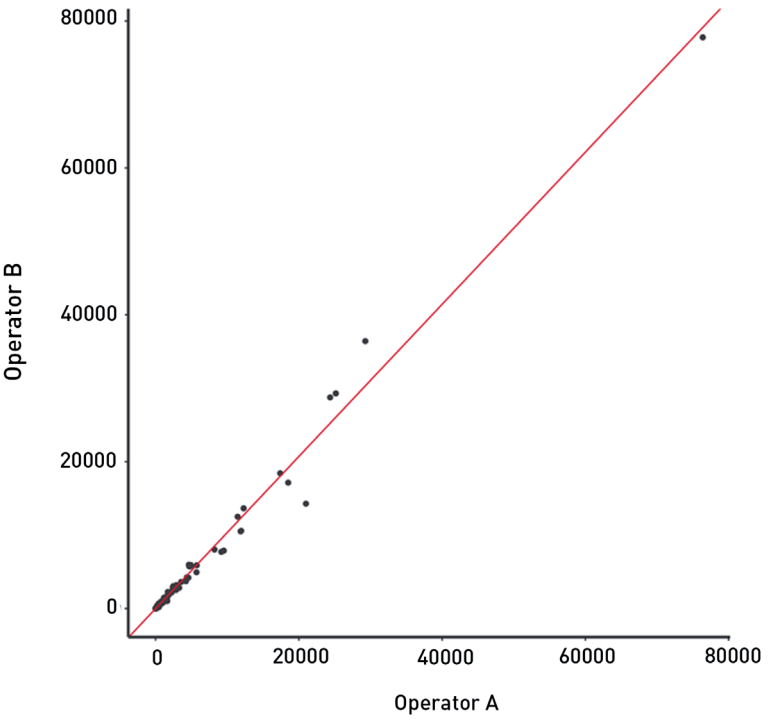

Figure 2: Correlation between 2 different operators using CATS technology on 1 ng of isolated small RNA for ncRNA detected at TPM ≥2. R^2 = 0.99.

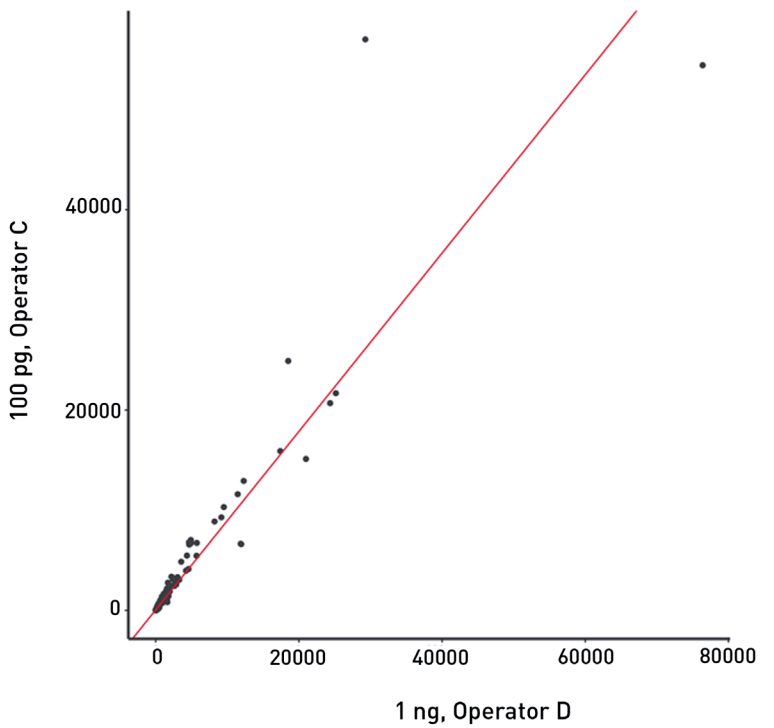

Figure 3: Correlation between 2 different operators using CATS on two different inputs for ncRNA detected at TPM ≥2. R^2 = 0.86.

COMPARISON WITH NEBNext® SMALL RNA LIBRARY PREP SET FOR ILLUMINA®

Library yield assessment

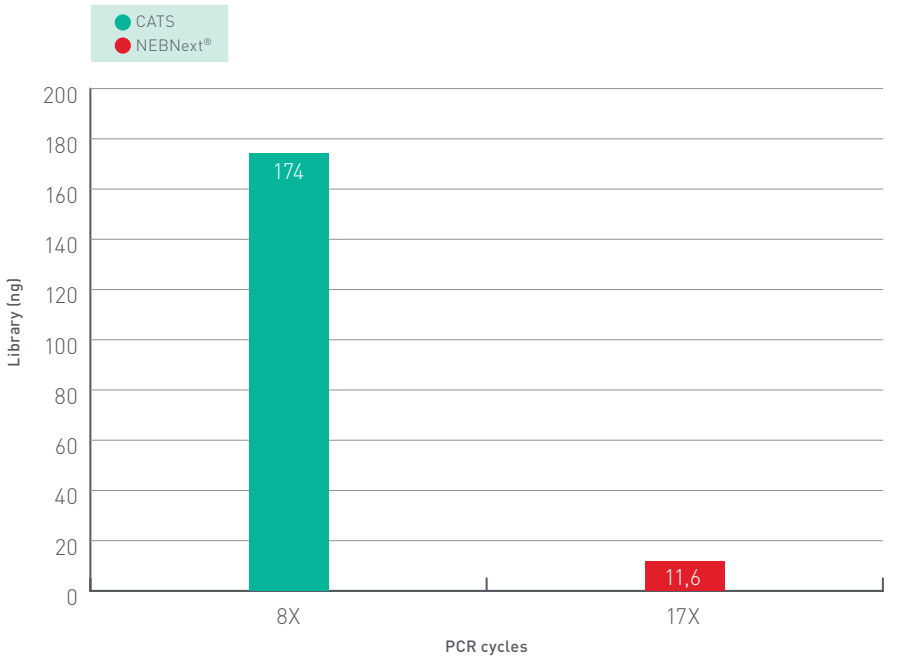

Figure 4: Representation of the library preparation yield according to different methods for 100 ng total RNA. Number of PCR cycles have been applied following manufacturer's instructions.

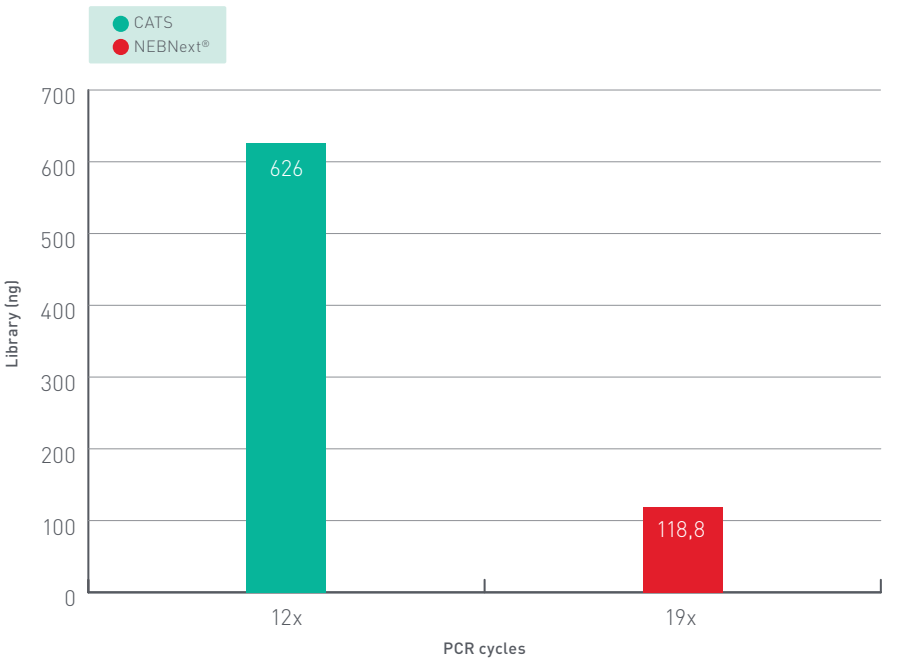

Figure 5: Representation of the library preparation yield according to different methods for 1 ng isolated small RNA. Number of PCR cycles have been applied following manufacturer's instructions.

Detection of small non-coding transcripts and their biotypes

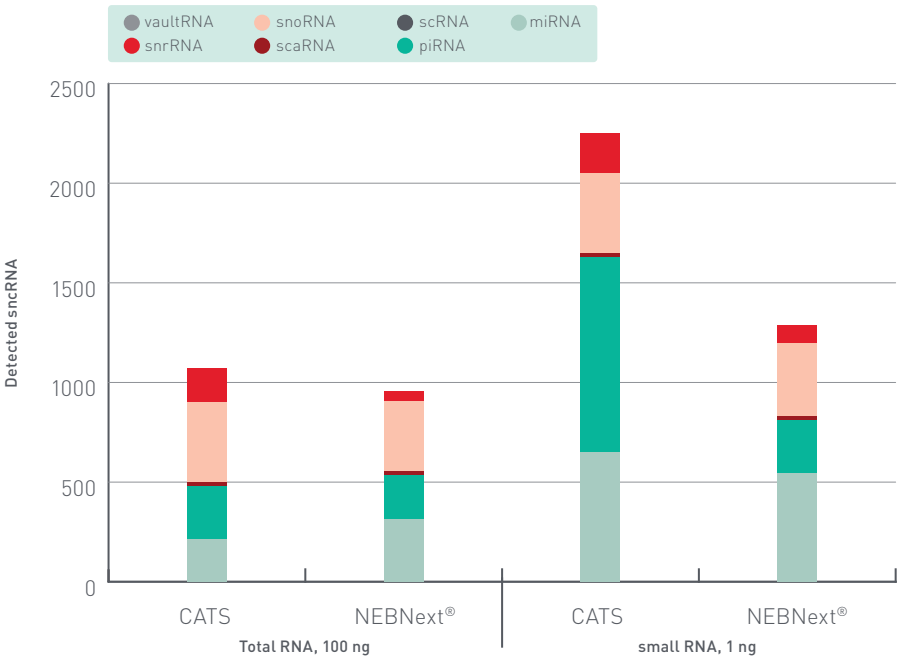

Figure 6: Representation of the number of detected transcripts according to different biotypes at TPM≥2. Different RNA input and template are represented for CATS and NEBNext®.

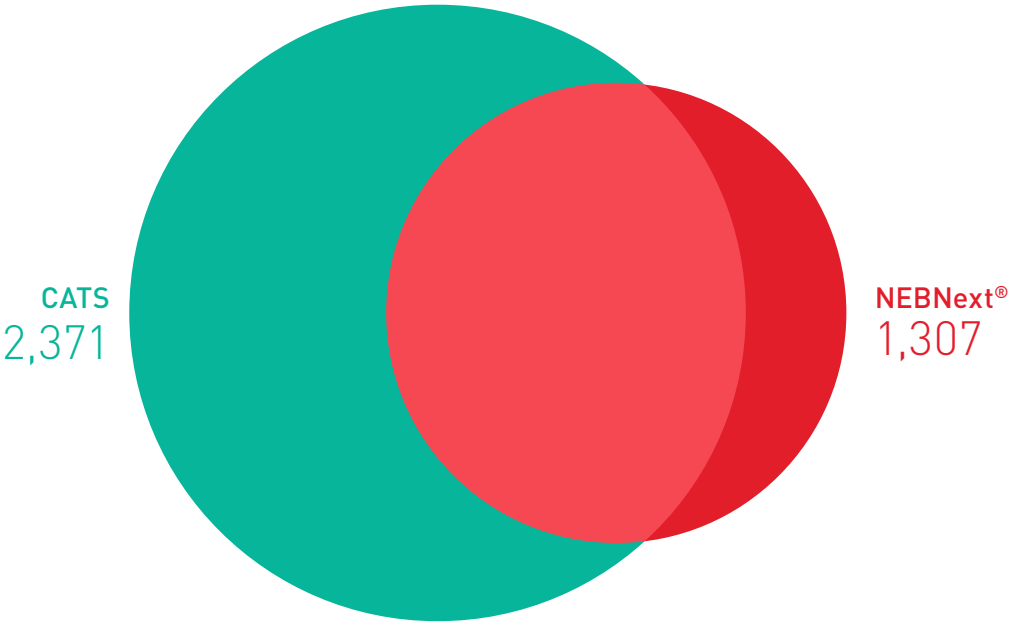

Figure 7: Comparison of the number of detected small non-coding RNAs at TPM ≥2 with CATS and NEBNext® on 1 ng isolated small RNA. Transcripts taken into consideration are miRNA, piRNA, scRNA, scaRNA, snoRNA, snRNA and vaultRNA. 1,010 transcripts were found in both libraries at this expression level.

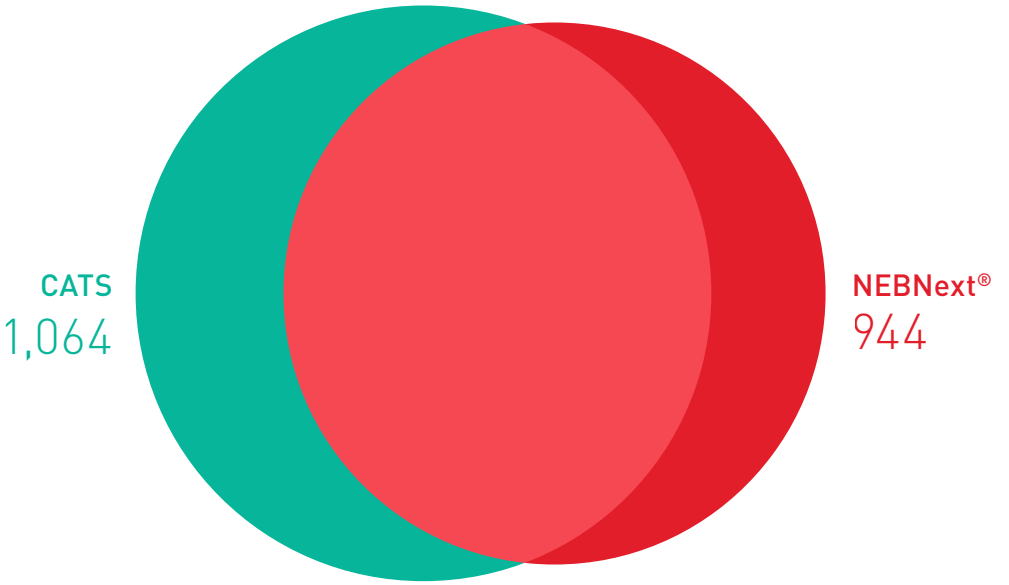

Figure 8: Comparison of the number of detected small non-coding RNAs at TPM ≥2 with CATS and NEBNext® on 100 ng total RNA. Transcripts taken into consideration are miRNA, piRNA, scRNA, scaRNA, snoRNA, snRNA and vaultRNA. 704 transcripts were found in both libraries at this expression level.

COMPARISON WITH SMARTer® smRNA-seq KIT FOR ILLUMINA®

Library yield assessment

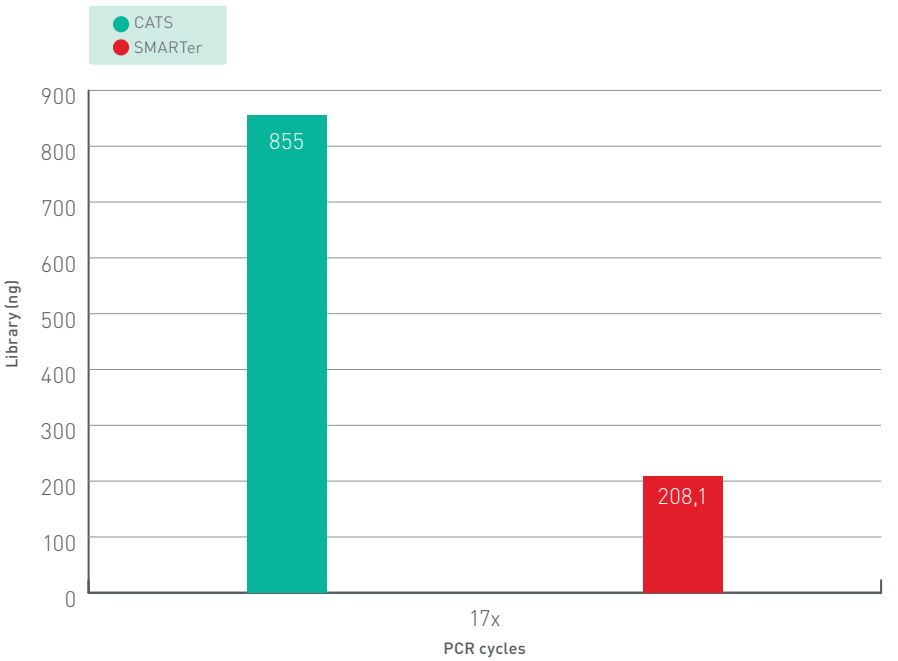

Figure 9: Representation of the library preparation yield according to different methods for 100 pg isolated small RNA. Number of PCR cycles have been applied following manufacturer's instructions.

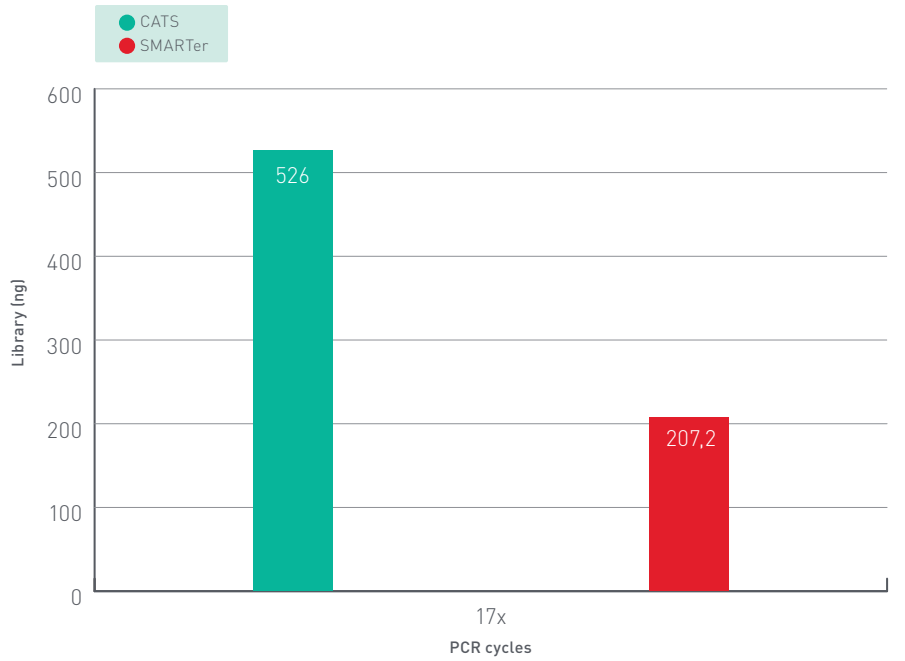

Figure 10: Representation of the library preparation yield according to different methods for 1 ng total RNA. Number of PCR cycles have been applied following manufacturer's instructions.

Detected small non-coding transcripts and their biotypes

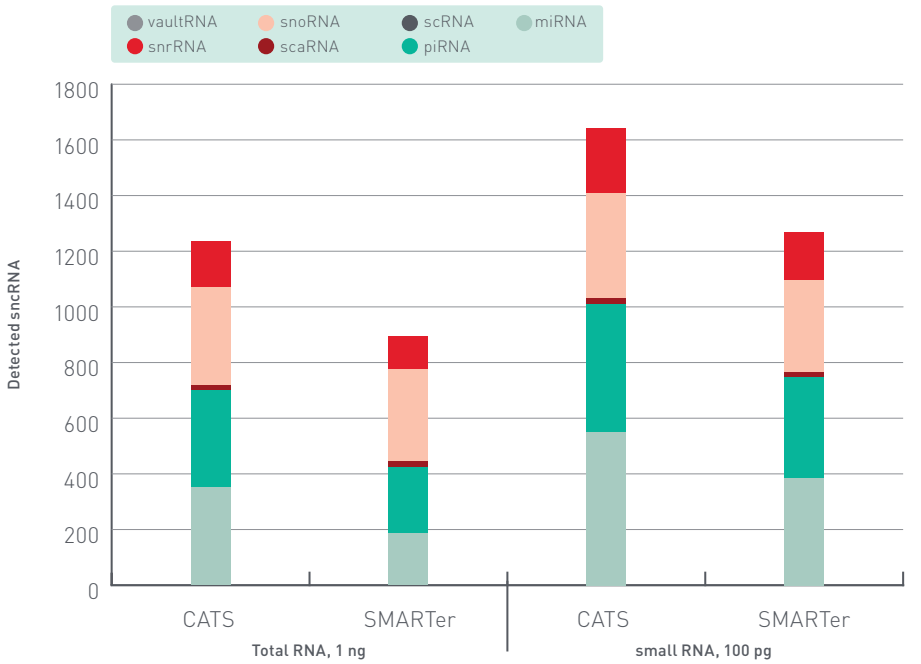

Figure 11: Representation of the number of detected transcripts according to different biotypes at TPM≥2. Different RNA input and template are represented for CATS and SMARTer.

Overlap detection of small non-coding transcripts

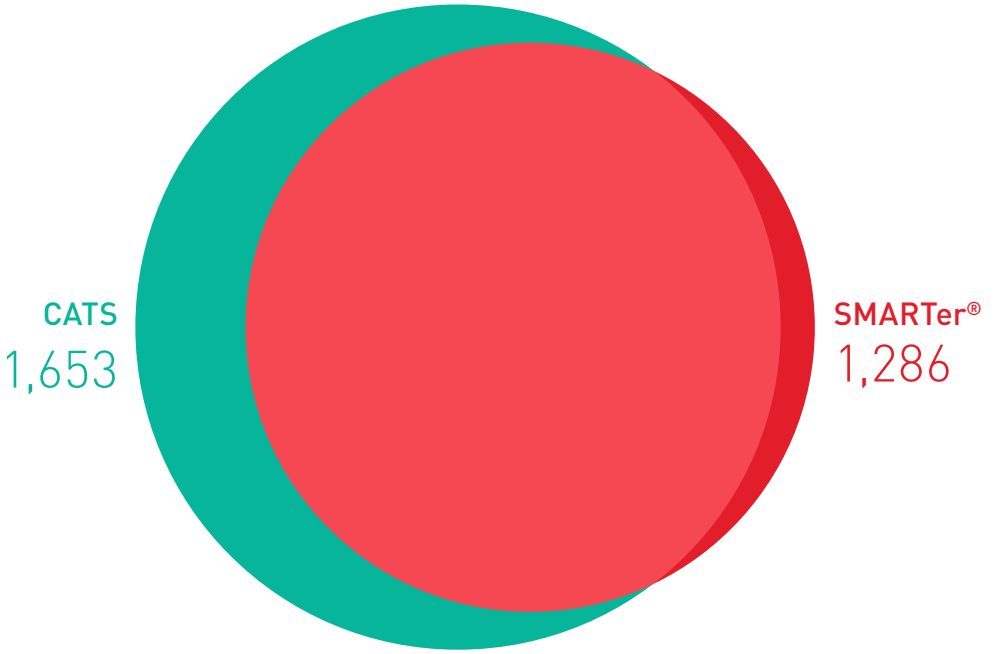

Figure 12: Comparison of the number of detected small non-coding RNAs at TPM ≥2 with CATS and SMARTer on 100 pg isolated small RNA. Transcripts taken into consideration are miRNA, piRNA, scRNA, scaRNA, snoRNA, snRNA and vaultRNA. 1,217 transcripts were found in both libraries at this expression level.

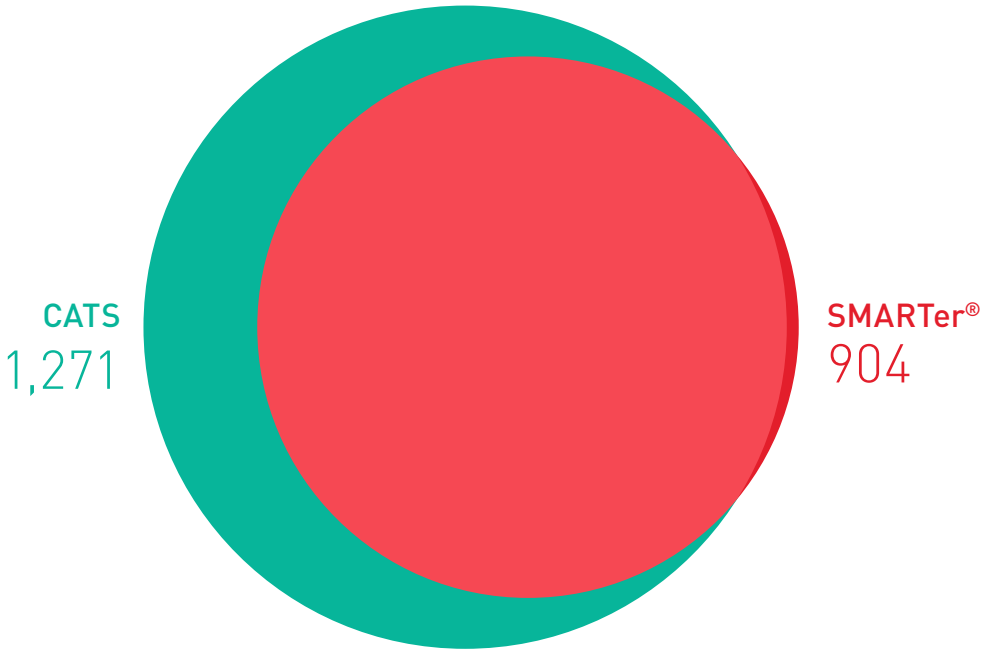

Figure 13: Comparison of the number of detected small non-coding RNAs at TPM ≥2 with CATS and SMARTer on 1 ng total RNA. Transcripts taken into consideration are miRNA, piRNA, scRNA, scaRNA, snoRNA, snRNA and vaultRNA. 892 transcripts were found in both libraries at this expression level.

DISCUSSION

CATS - a highly robust technology

The correlation for different users in Figure 2 shows that the CATS is a robust library preparation method and that reproducible results can be achieved independent of operator or location.

Furthermore, the correlation in Figure 3 indicates that the vast majority (R^2 =0.86) of the transcripts detected in 1 ng are also detected with 100 pg input of small RNA. This shows that CATS retains complexity when using less starting material for library preparation.

Comparison with NEBNext® Small RNA Library Prep Set for Illumina®

The CATS Small RNA-seq library preparation performs better overall than the NEBNext® Small RNA library preparation kit.

First of all, starting with the same RNA input amount, the CATS method requires far fewer amplification cycles, and is therefore potentially less biased. Furthermore it also indicates that a larger proportion of the sample is incorporated during library preparation (Figures 4 and 5).

This trend is confirmed in Figure 6 which clearly shows that CATS provides superior transcript diversity. This figure shows that for equivalent starting material, the CATS has more reads mapped to sncRNA. Moreover, the highest detection level of small non-coding RNAs is reached when using the CATS on 1 ng isolated small RNA. Although different templates can be processed for library preparation, we advise to start with the small RNA isolated fraction (< 200 nt) but one can still use total RNA and get satisfactory results.

Figures 7 and 8 show that the vast majority of transcripts detected with the NEBNext® kit are also detected with the CATS method. Figure 7 (1 ng isolated small RNA) illustrates that significantly more transcripts are uniquely detected with CATS rather than with the NEBNext® technology, which positions CATS as a valuable tool for transcript discovery.

Comparison with SMARTer® smRNA-seq kit for Illumina®

The comparison with the SMARTer® smRNA-seq kit for Illumina® shows similar results to the first comparison with NEBNext® Small RNA Library Prep Set for Illumina®.

The library output is higher for CATS than for SMARTer although the same number of PCR cycles has been applied (Figures 9 and 10). Again, this indicates that a higher proportion of the sample is incorporated during library preparation and demonstrates overall a more efficient library preparation process.

In addition, Figure 11 illustrates that with equal starting amounts, a significantly higher number of reads are mapped to small non-coding RNAs with CATS rather than with SMARTer.

Figures 12 and 13 show that the overwhelming majority of all detected transcripts by SMARTer are equally detected by CATS. However, CATS also uniquely detects a significant number of other transcripts which the SMARTer technology is unable to identify.

CONCLUSION

It has been clearly shown that CATS identifies the most transcripts when compared to competing methods and also allows for detection of a significantly higher number of small RNA which are not captured at all when using competing solutions. These results confirm CATS as a method of choice for small non-coding RNA discovery.

Furthermore, this superior detection is achieved by performing fewer PCR cycles prior to sequencing which reduces the risk of erroneous identification of amplification duplicates as true biological duplicates.

In conclusion, CATS Small RNA-seq library preparation is a highly efficient and reproducible method for producing libraries for small RNA sequencing. This technology not only offers a streamlined protocol with minimal hands-on time but also results that outperform competing methods.