Direct RNA sequencing of dsRNA using Oxford Nanopore Technologies

Ilia Gelfat, Luis Ramirez-Tapia, Emiko Kurosawa, Mark Sheff, Shivani Khatu, Diana Posadas, Nicholas Skizim

GreenLight Biosciences, Medford, MA, USA, 02155

ABSTRACT

Double-stranded RNA (dsRNA) occurs in several biological phenomena, such as dsRNA viruses, intermediates of replicating single-stranded RNA viruses, and long cellular dsRNA[1,2]. In recent decades, there has been increased interest in utilizing dsRNA for RNA interference, particularly for pest control in agricultural settings[3], necessitating the large-scale production of dsRNA. Novel methods for dsRNA characterization are therefore of interest for RNA manufacturing as well as for the study of RNA biology.

In this work, we describe the development of a protocol for direct RNA sequencing (DRS) of dsRNA using Oxford Nanopore Technologies. A simple computational pipeline is then used to separate reads corresponding to the Sense and Antisense strands. Finally, we show that through the combination of strand identity, read length distribution, and sequence information, we can gain insight into and distinguish between various molecular mechanisms that influence RNA structure, such as early termination and multiple transcription start sites, while also providing sequence identity and quantification of Sense vs Antisense. The work described here therefore provides a powerful and versatile tool for the characterization and study of dsRNA.

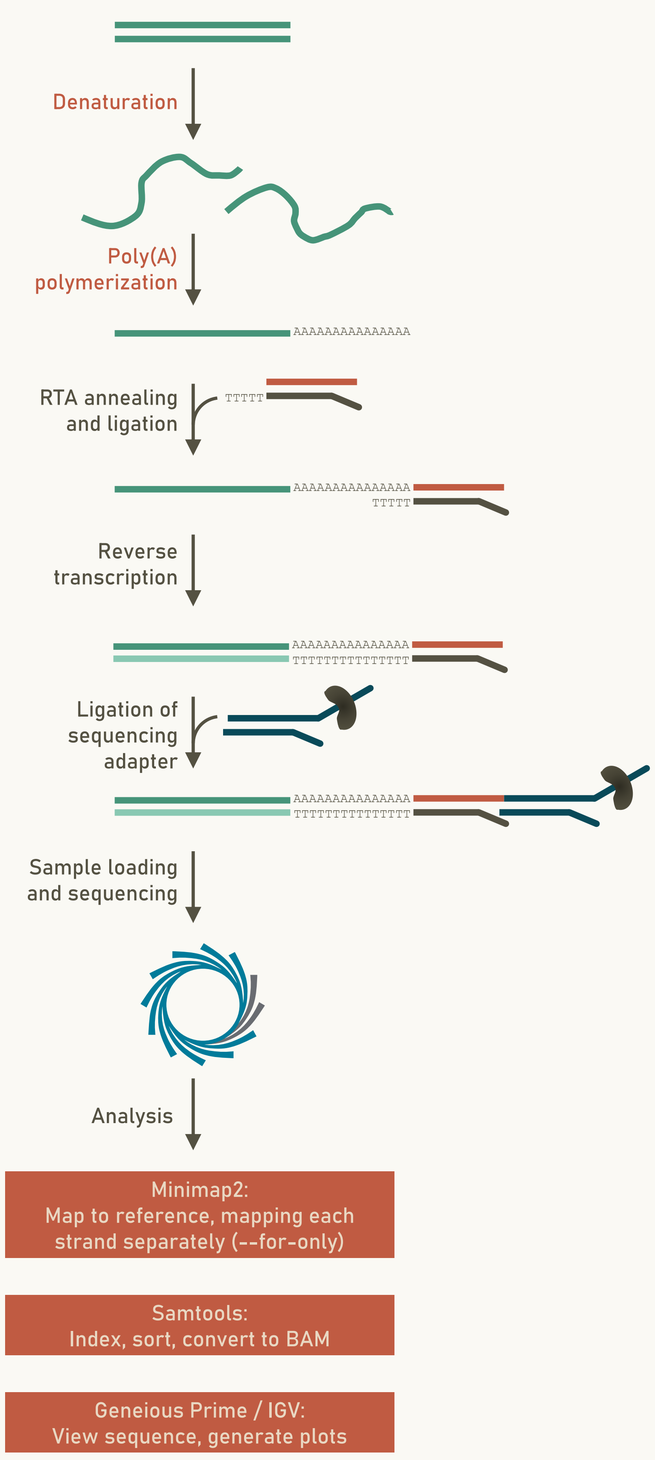
LIBRARY PREPARATION

Our dsRNA sequencing method is based on the standard Oxford Nanopore protocol for DRS. However, since the ONT protocol requires the input material to be single-stranded and polyadenylated, the following modifications were made:

1. **Denaturation:** 10µL of 50% DMSO are added to 10µL of dsRNA sample. The sample is then denatured at 90°C for 5 minutes, followed by rapid cooling on an ice block.
2. **Polyadenylation:** *E. coli* poly(A) polymerase (New England Biolabs, USA) is used according to vendor protocol to add a poly(A) tail to the denatured RNA sample.

A schematic of the complete library preparation protocol can be found in Figure 1 below. Following library preparation and sequencing, Minimap2 is used to separately map reads to either Sense or Antisense. This can be achieved by using either a Sense or Antisense reference sequence and the Minimap2 "--for-only" option (or, alternatively, using a single reference sequence, but calling Minimap2 with either "--for-only" or "--rev-only").

Figure 1: library preparation, sequencing, and analysis workflow



MATERIALS

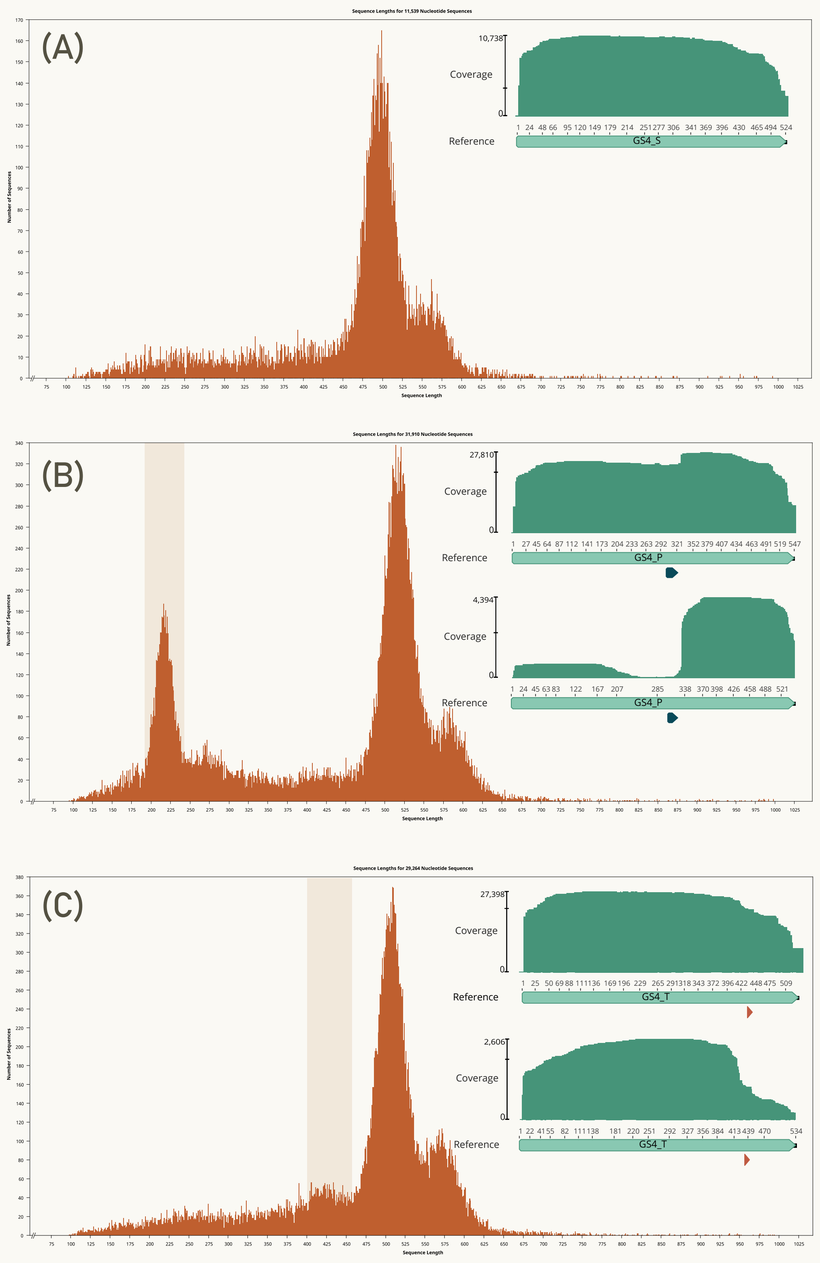
For this work, several sequences were designed based on 524bp segment taken from a GFP-encoding sequence. A T7 promoter sequence was added upstream of the desired transcripts, and DNA templates were purchased from IDT (Integrated DNA Technologies, USA) for the Sense and Antisense transcripts. These sequences were labeled GS4_S and GS4_AS, respectively.

Two additional sequences were designed based on GS4_S:

- GS4_P: an additional T7 promoter sequence (TAATTGAACTCACTAAAGGGAGA) was inserted 245bp upstream of the 3'-end (Fig. 2B, inset, dark blue annotation).
- GS4_T: a 10-bp terminator sequence (CATCTGTTTT) was inserted 102bp upstream of the 3'-end (Fig. 2C, inset, orange annotation).

All RNA materials were synthesized using in vitro transcription (IVT) according to a previously published protocol[4]. Each strand was produced in a separate reaction. GS4_AS was mixed with either GS4_S, GS4_P, or GS4_T at a desired molar ratio, and annealed by holding the materials at 80°C for 20 minutes followed by cooling to 20°C at a rate of 1°C/min.

Figure 2: strand-specific sequence analysis. Read-length distributions and coverage maps of the Sense strand for samples containing (A) GS4_S, (B) GS4_P, and (C) GS4_T. In panel (B), the top coverage map represents all Sense reads, while the bottom map represents reads of lengths 190-240nt (also highlighted in distribution). In panel (C), the top coverage map represents all Sense reads, while the bottom map represents reads of lengths 400-460nt (also highlighted in distribution).



STRAND-SPECIFIC CHARACTERIZATION

Three dsRNA samples were prepared as described above by combining GS4_AS with GS4_S, GS4_P, or GS4_T in a 1:1 ratio. Each sample was prepared according to the dsRNA library preparation protocol (Fig. 1), sequenced using Flongle flow cells (R9.4.1), and Sense reads were separated from Antisense reads. Geneious Prime was used to generate read-length distribution plots and coverage maps for the Sense reads (Fig. 2).

For all three samples, most reads were close in size to full-length transcripts, exhibiting a peak in the distribution around ~500nt. The GS4_P-containing sample contained an additional population around ~220nt in length, closely matching the size of a partial transcript initiated at the inserted T7 promoter. The coverage map demonstrates increased coverage downstream of the added promoter sequence, allowing identification of the inserted sequence as a promoter. This is further supported by extracting reads of lengths 190-240nt and separately mapping them to the reference, which shows these reads overwhelmingly matching the partial transcript.

Similar analysis was performed for the GS4_T-containing sample. In addition to the main population of reads, a small subpopulation was present at ~425nt. A modest drop in coverage was also observed at the inserted terminator region, consistent with the presence of a functional weak terminator. Mapping the reads of length 400-460nt to the reference shows alignment with the suspected partial transcript upstream of the added terminator.

STRENGTHS

- Allows correlation of structural and sequence information
- Strand-specific
- Polyadenylation – agnostic to sequence and integrity of 3'-end
- DRS allows for detection of modified bases
- No PCR or RT bias

LIMITATIONS

- Lower basecall accuracy compared to cDNA-based methods
- Obtaining full-length reads may be challenging for longer RNA species

SENSE VS ANTISENSE QUANTIFICATION

Five dsRNA samples were prepared as described by combining GS4_S and GS4_AS in different ratios: 9:1, 3:1, 1:1, 1:3, and 1:9. These correspond to a GS4_S percentage of 10%, 25%, 50%, 75%, and 90%, respectively. Following library construction (Fig. 1), samples were sequenced using Flongle flow cells (R9.4.1), and Sense reads were separated from Antisense reads.

The proportion of GS4_S was estimated by dividing the number of Sense reads by total reads. As shown in Figure 3, these estimates closely matched the known GS4_S proportions of each sample. Linear regression between the expected proportion and observed proportion was carried out in Python 3 (scipy.stats package). The regression yielded the following results:

- Slope = 0.96 (95% confidence interval [0.696, 1.224])
- Intercept = -0.009 (95% confidence interval [-0.163, 0.145])
- R^2 = 0.978

Overall, the observed proportions provided a good estimate for all samples within the tested range (10%-90%), with a mean error of 3%.

Figure 3: Sense vs Antisense quantification. Read numbers from five GS4 samples with varying Sense proportions (between 10-90%) were used to estimate the proportion of Sense strands (dark green). Linear regression yielded the following parameters:
$y = 0.96 \cdot x - 0.009$, $R^2 = 0.978$ (light green).

CONCLUSIONS

- In this work, we developed a protocol for direct RNA sequencing of dsRNA.
- This method is agnostic to 3'-end sequence, can distinguish between Sense and Antisense strands, and comes with all the strength and limitations of standard DRS.
- The combination of sequence, strand identity, and read length distribution can provide insights into dsRNA structure. We hope this method can serve as a tool for the study of dsRNA in its many biological roles and potential applications.
- Future work will aim to utilize this method to characterize various mechanisms of RNA degradation.

REFERENCES

1. Yu, N., Christiaens, O., Liu, J., Niu, J., Cappelle, K., Caccia, S., ... & Smagghe, G. (2013). Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect science, 20(1)*, 4-14.
2. Reich, D. P., & Bass, B. L. (2019). Mapping the dsRNA world. *Cold Spring Harbor perspectives in biology, 11(3)*, a035352.
3. Huvenne, H., & Smagghe, G. (2010). Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. *Journal of insect physiology, 56(3)*, 227-235.
4. Beckert, B., & Masquida, B. (2010). Synthesis of RNA by In Vitro Transcription in RNA in Methods in Molecular Biology.
5. Geneious Prime version 2023.0 created by Biomatters. Available from: https://www.geneious.com

© 2023 | GreenLight Biosciences Inc. | All rights reserved

GreenLight BIOSCIENCES™