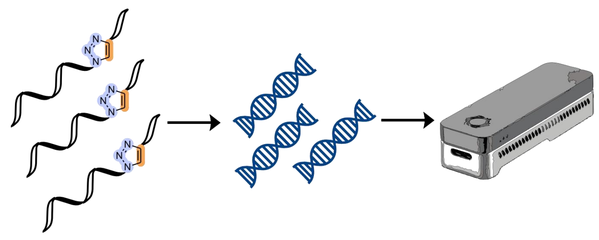
Click Chemistry enables rapid amplification of full-length reverse transcripts for long read Next Generation Sequencing

E. Schönegger [1,2] , A. Crisp [1], M. Müller [2], T. Carell [2],T. Frischmuth [1]

[1] baseclick GmbH, Neuried 82061 Germany, [2] Ludwig-Maximilians-Universität München (LMU), Institute for Chemical Epigenetics (ICEM), 81375 Munich, Germany

INTRODUCTION

Eukaryotic pre-mRNA undergoes numerous co- and post-transcriptional processing events that dictate when and how it will be translated by the cellular apparatus. Our understanding of these processes has matured largely based on advancements in RNA sequencing (RNA-seq) technologies.[1] To grasp the full biological role of a particular gene, capture and sequencing of all possible transcripts is thus required, since the biological function of any particular spliced and polyadenylated form may be distinct.[2] Post-transcriptionally regulatory events, however, often remain undetected due to technical limitations associated with library generation.[3]



We describe here the development of a novel click chemistry-based method for the generation and amplification of full-length cDNA libraries from total RNA, while avoiding the need for problematic template-switching (TS, standard library preparation based on the Smart-Seq2 protocol) reactions. Our method avoids random priming as well as stochastic cDNA termination, thus enabling the amplification of transcripts that were previously inaccessible by related click chemistry-based RNA sequencing techniques. [4]

TECHNOLOGY

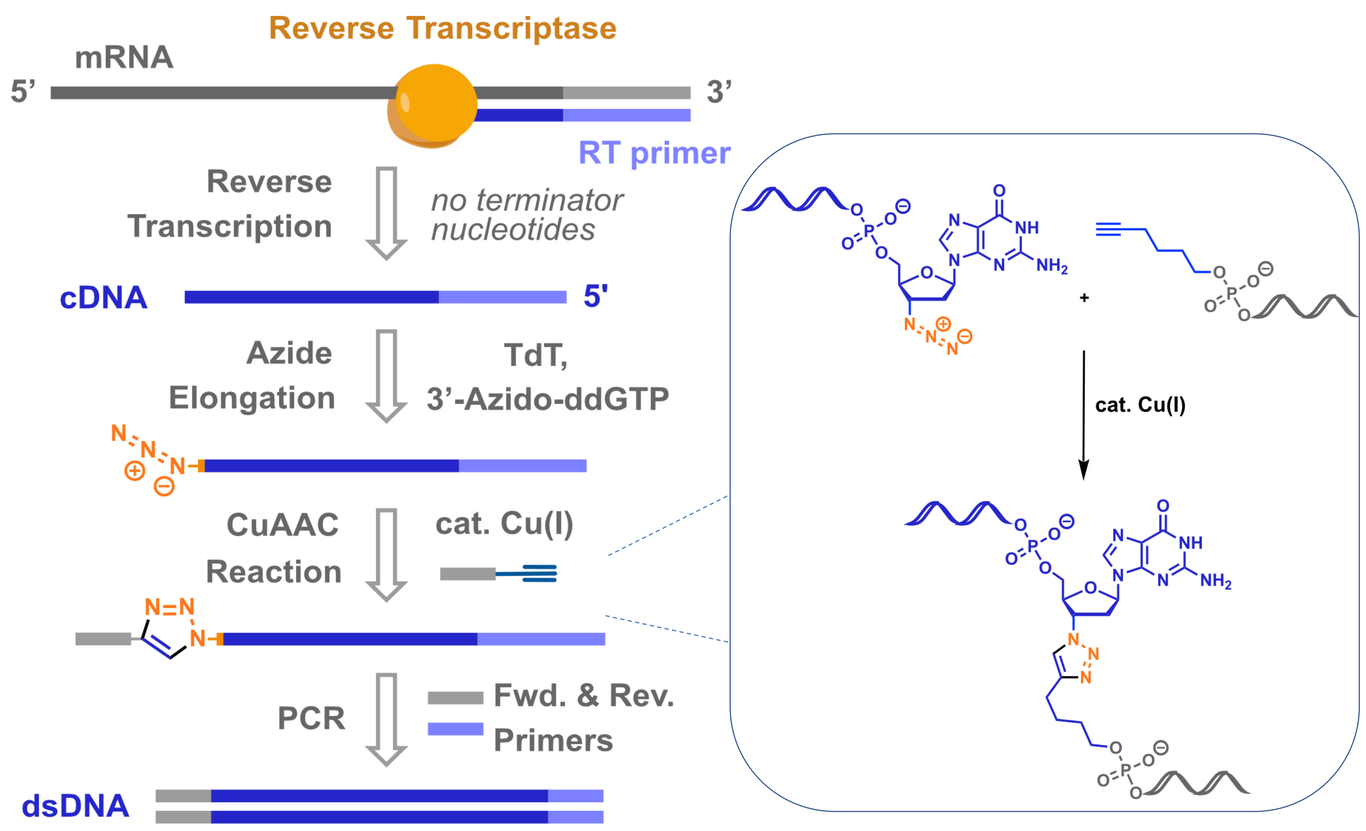


Figure 1. Schematic overview describing the preparation of dsDNA libraries from extracted total-RNA or isolated mRNA sequences. The sequence specific template, required for PCR amplification of the ssDNA library, is provided by copper-catalyzed azide–alkyne cycloaddition (CuAAC) of 3'-azido labelled cDNAs with an adapter oligo bearing a 5'-hexynyl moiety.

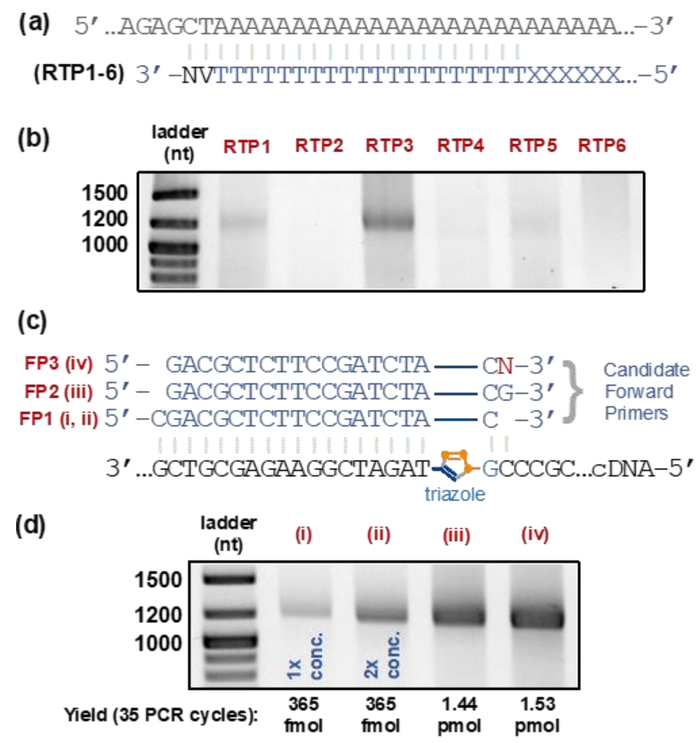


Figure 2. Optimization of RT- and PCR- primers. Reverse transcribed eGFP mRNA (1.22 kB) using six RT-primers (RTP1–RTP6) **(a).** Crude PCR products originating from each RT-primer visualized by agarose gel-electrophoresis **(b)**. 3'-Clicked cDNA obtained using the best-performing RT primer (RTP3) was amplified using each of three candidate forward primers (FP1-3) in combination with reverse primer RP3 **(c)**. Agarose-gel visualization and quantification indicating significantly enhanced yields by the incorporation of an extra, quadruple degenerated nucleotide at the 3'- end of the forward primer **(d)**.

RESULTS

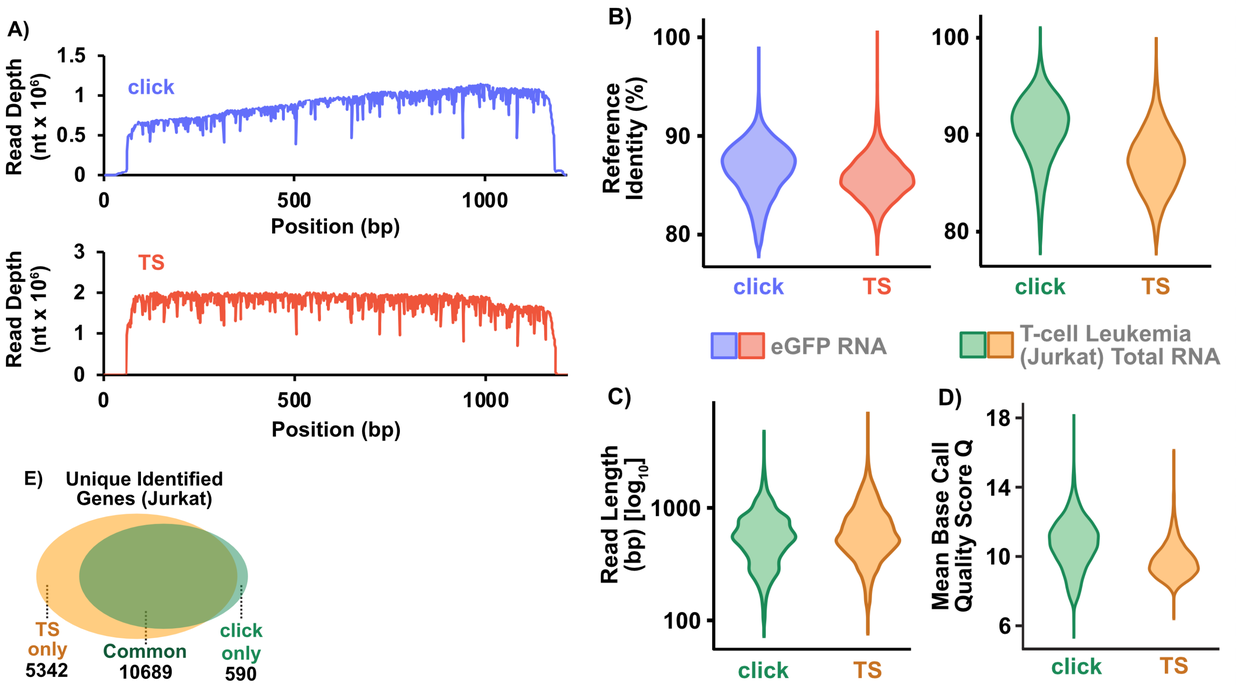


Figure 3. Performance of the Click-libraries in nanopore sequencing. **A)** Coverage depth (read numbers) on the eGFP model transcript of two independent runs (blue: click library, orange: TS). **B)** Comparison of mapping quality (percent reference identity) of the eGFP model libraries and Jurkat total RNA libraries, prepared by click chemistry (click) or standard protocol (TS). **C)** Read-length distribution of the two Jurkat total RNA libraries. **D)** PHRED base call quality score comparison of click vs. standard library. **E)** Number and distribution of identified transcripts.

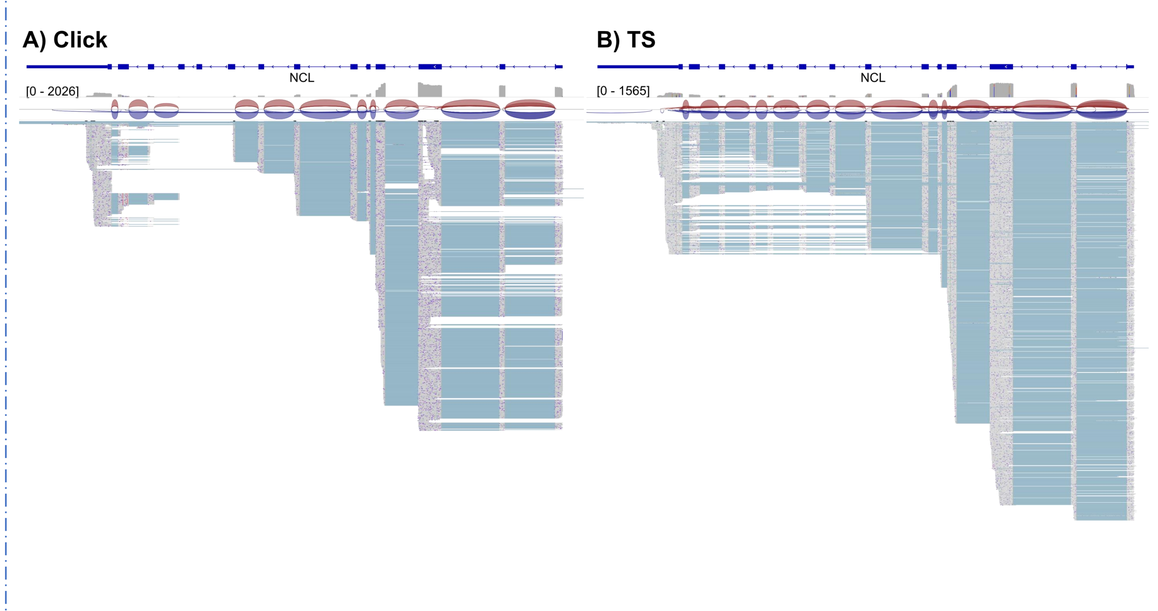


Figure 4. Mapping example, showing the nucleolin (NCL) gene on chromosome 2 for two Jurkat total RNA libraries sequenced on a MinION device. Mapping was performed with minimap2 against the GRCh38 reference genome. **A)** Library prepared with the Click protocol. **B)** Library prepared with the standard procedure.

CONCLUSION

The developed click chemistry-based method for generating full-length cDNA libraries allows the entire transcript to be sequenced as a single read, while overcoming the limitations of template switching oligos. Our data shows that the performance of the click library in both amplicon and transcriptome sequencing can compete with the standard protocols. Instead of losing accuracy due to Cu-mediated damage, our technology performs even better than the standard protocol.

REFERNCES

[1] M. R. Green, *Annu. Rev. Genet*. **1986**, *20*, 671–708.
[2] O. Kelemen, P. Convertini, Z. Zhang, Y. Wen, M. Shen, M. Falaleeva, S. Stamm, *Gene* **2013**, *514*, 1–30.
[3] Y. Tian, H. Wen, X. Qi, X. Zhang, S. Liu, B. Li, Y. Sun, J. Li, F. He, W. Yang, et al**., Front. Genet. 2019**, 10.
[4] A. Routh, P. Ji, E. Jaworski, Z. Xia, W. Li, E. J. Wagner, *Nucleic Acids Res*. **2017**, 45, 12.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861381.