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

Eva S. Schönegger ‡ †, Antony Crisp ‡, Markus Müller †, Jessica Fertl, Sascha Serdjukow, Stefano Croce, Michael Kollaschinski, Thomas Carell †, Thomas Frischmuth*

baseclick GmbH, Floriansbogen 2-4, 82061 Neuried (Munich), Germany

† Ludwig-Maximilians-Universität München, Institute for Chemical Epigenetics Munich, Butenandtstr. 5-13, 81377 Munich, Germany

‡ These authors contributed equally to this work.

Supporting Information Placeholder

ABSTRACT: Here we describe 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) reactions. Compared with prior efforts, our method involves neither random priming nor stochastic cDNA termination, thus enabling amplification of transcripts that were previously inaccessible via related click chemistry-based RNA sequencing techniques. A key modification involving the use of PCR primers containing two overhanging 3'-nucleotides substantially improved the read-through compatibility of the 1,4-disubstituted 1,2,3-triazole-containing cDNA, where such modifications typically hinder amplification. This allowed us to more than double the possible insert-size compared with the state-of-the art click chemistry-based technique, PAC-seq. Furthermore, our method performed on-par with a commercially available PCR-cDNA RNA sequencing kit, as determined by Oxford Nanopore sequencing. Given the known advantages of PAC-seq, namely suppression of PCR artefacts, we anticipate that our contribution could enable diverse applications including improved analyses of mRNA splicing variants and fusion transcripts.

Introduction

Beyond its initial synthesis, eukaryotic pre-mRNA undergoes a multitude of co- and post-transcriptional processing events that dictate how and when it will be translated by the cellular apparatus. Our understanding of these processes has matured largely thanks to advancements in RNA sequencing (RNA-seq) technologies.[1] In order to ascertain the function of a particular gene transcript, capture and sequencing of all possible variants is required, since the biological function of any particular spliced and polyadenylated form may be distinct.[2] Due to technical limitations associated with library generation, however, post-transcriptional regulatory events can often go undetected.[3] Given that the majority of mammalian transcripts are between 1-2 kb in length,[4,5] long-read technologies are crucial when it comes to disentangling the entire spectrum of mRNA isoforms.[5] The most widely-exploited strategy for cDNA library preparation from total RNA involves the use methods derived from the Smart-seq2 protocol.[6,7] As with many other approaches, Smart-seq2 involves a template switching (TS) reaction whereby primer-specific 3'-adapter sequences can be appended to nascent cDNA ends.[8] This process begins with an initial 3'-extension with three protruding 2'-deoxycytidine nucleotides, which may then subsequently hybridize to a reverse-complementary 'rGrGrG'-containing template-switching oligo (TSO). This allows the reverse transcriptase to shift between the DNA and RNA templates and to copy the entire TSO sequence to the 3'-end of the resulting cDNA.[5] While groundbreaking in its original implementation,[9] template switching introduces a host of concerns, including inefficient ligation,[10,11] overrepresentation of 5'-guanosine-containing sequences,[12] as well as major artefacts and false alternative transcripts.[13]

To address these drawbacks, we became interested in developing a full-length cDNA library preparation method for long read sequencing using click chemistry. Compared with enzymatic processes, Cu(I)-catalyzed azide–alkyne cycloaddition (CuAAC) reactions have several distinct advantages. Namely, CuAAC reactions are straightforward to execute, are stereospecific, are often quantitative in yield, and result in few to no biproducts.[14] We were particularly inspired by the work of Routh and coworkers, who recently reported a novel click chemistry-based approach for the capture and elucidation of 3´UTR/poly(A) tail junctions via short-read sequencing.[15,16] In their protocol titled 'PAC-seq', RT-PCR was performed in the presence of 3'-azido-ddNTPs, such that

stochastic chain termination always occurs upstream of the poly(dT)-primed 3´UTR. We sought to utilize a similar approach that would target not only the 3´-ends of polyadenylated RNAs, but also the entirety of a given transcript. We envisaged that this could be achieved via the introduction of a subsequent and separate terminal deoxynucleotidyl transferase (TdT)-catalyzed ligation reaction after reverse transcription to incorporate the relevant 3'-azido-ddNTP, as has previously been exploited within our laboratories.[17,18] The advantage in doing so would allow the entire transcript to be sequenced as a single read, harnessing the full capacity of Third Generation Sequencing technology, whereas stochastic termination cDNA synthesis can only provide reads of a certain predetermined, average length. Compared with other long-read library preparation methods including but not limited to Smart-Seq2 [10] or those described in commercially available kits from OXFORD NANOPORE TECHNOLOGIES, the potentially problematic one-step RT-PCR is avoided. Our method is also thus expected to exhibit the other ostensible benefits of ClickSeq methods such as circumvention of artefactual chimera formation.[15]

Results and Discussion

Development of cDNA library preparation using CuAAC

With the previous literature precedent [15,16,19,20] as well as our planned improvements in mind, we designed our cDNA library preparation method as shown in Figure 1 a. Reverse transcription, using eGFP mRNA as template (Figure S6), is initially performed under standard conditions, employing an anchored poly(dT)-containing primer to ensure full coverage of the 3'UTR. After cDNA synthesis and digestion of the RNA template, nascent cDNAs are treated with a buffered solution containing 3'-azido-2',3'-dideoxyguanosine-5'-triphosphate in the presence of TdT, which results in ligation of a single azido-modified nucleotide at the 3'-cDNA end. The mixture is then purified via a silica-membrane-based purification system to provide the single-stranded 3'-azido-terminated cDNAs as a concentrated solution in unbuffered-H_2O. The click ligation is accomplished using Reactor 25, a solid, heterogeneous Cu(I) catalyst system developed at BASECLICK GMBH in the presence of THPTA, $MgCl_2$, and an adapter oligodeoxynucleotide (ODN) containing a 5'-(5-hexynyl) phosphate moiety (Table S1), the sequence of which was closely derived from that which was designed by Routh and co-workers.[15] This reaction provides the 3'-adapter-ligated cDNA that will serve as the template for further PCR amplification (Figure 1 b).

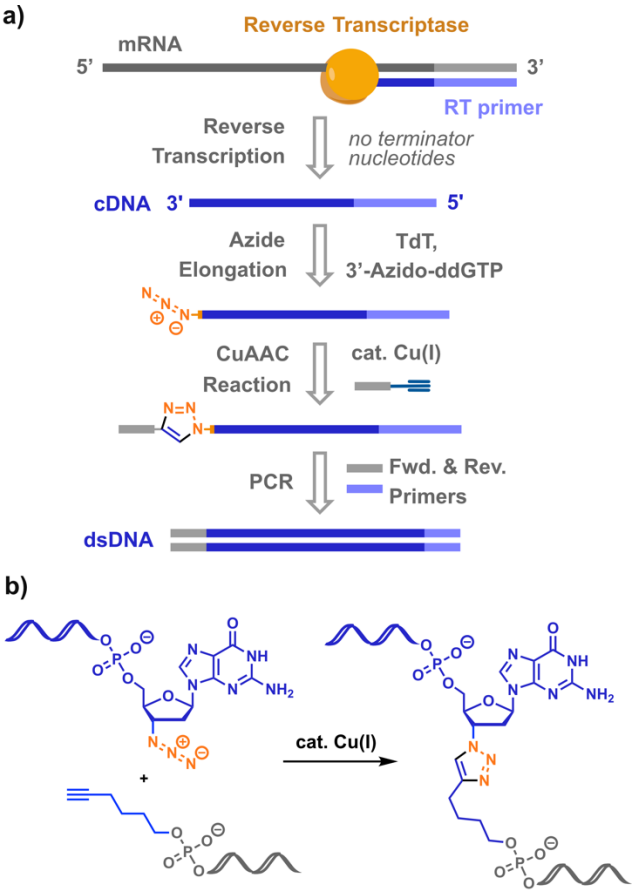

Figure 1. a) Schematic overview describing how dsDNA libraries are prepared from extracted total-RNA or isolated mRNA sequences. b) The copper-catalysed azide–alkyne cycloaddition (CuAAC) reaction of 3'-azido labelled cDNAs to provide the sequence-specific template required for PCR amplification of the ssDNA library.

Given that biocompatible triazole linkers are known to have a diminishing effect on the yields of enzymatic reactions that copy through the unnatural backbone,[21] certain considerations to primer design were made to ensure the highest possible yield during the PCR amplification. Firstly, a PCR forward primer was designed such that a single, overhanging deoxycytidine nucleotide will hybridize to the 3'-azido-2',3'-dideoxyguanosine nucleotide of the reverse-complimentary cDNA strand (Figure 2 a). This design ensures bridging of the triazole backbone with the primer and thus avoiding primer extension (or replication) across the potentially unfavorable linkage. The design also enables primer extension of the sense strand irrespective of the sequence context. To properly assess the capacity of this primer to facilitate PCR, short amplifications were set up involving a sequence specific reverse primer (Figure 2 b; Table S1). RTP3 was identified as the best-performing RT primer and thus 3'-clicked cDNA, obtained using RTP3 was amplified using each of three candidate forward primers (FP1-3) in combination with reverse primer RP3 (Figure 2 c) to further improve

the yield and reproducibility of the PCR. Agarose-gel visualization and quantification using a Qubit™ Fluorometer indicated that yields are substantially enhanced by the incorporation of an extra, fully degenerate (N) nucleotide at the 3'-end of the forward primer (Figure 2 d). Primer and adapter sequences are indicated in Table S1.

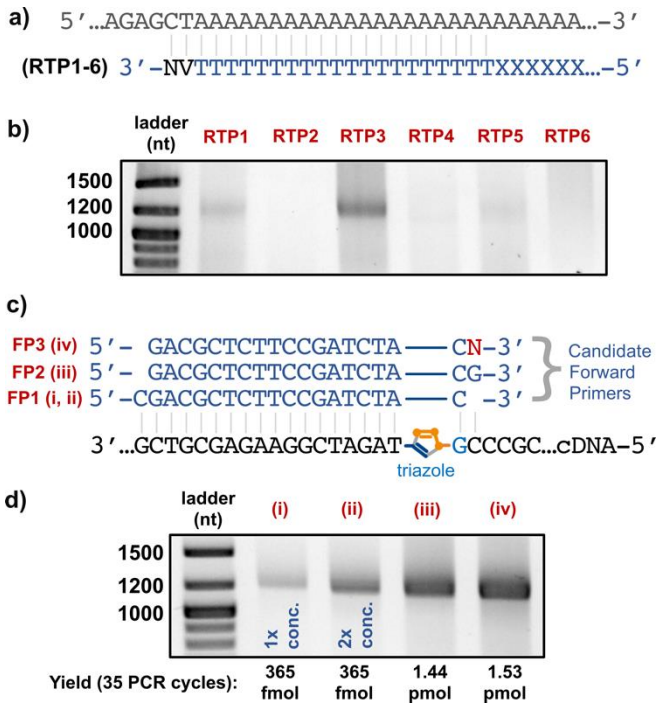

Figure 2. Optimization of RT- and PCR-primers allows the amplification of an in-vitro transcribed model 1.22 kB mRNA coding for enhanced green fluorescent protein (eGFP), which was first reverse transcribed using six candidate RT-primers a) (RTP1 – RTP6) and then labelled using 3'-azido-2',3'-ddGTP in combination with the enzyme TdT. After the Cu(I)-catalyzed click-ligation of the resulting 3'-azido-cDNAs with the alkyne-adapter AA1, the material was amplified by PCR. b) The crude PCR products originating from each RT-primer candidate were then visualized by agarose gel-electrophoresis. c) 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. d) Agarose-gel visualization and quantification using a Qubit™ Fluorometer indicated that yields are substantially enhanced by the incorporation of an extra, fully degenerate (N) nucleotide at the 3'- end of the forward primer. Primer and adapter sequences are indicated in Table S1.

Evaluation of click chemistry-based dsDNA libraries via ONT third generation sequencing

Having now demonstrated that triazole-conjugated ssDNA products of greater than 1 kB can successfully serve as templates for PCR, we next went about generating dsDNA libraries for sequencing with the OXFORD NANOPORE TECHNOLOGIES (ONT) platform. To this end, eGFP-encoding dsDNA products derived from clicked-cDNA were enzymatically repaired, and subsequently ligated to sequencing adaptors using a commercially available ligation-based sequencing kit designed for multiplexing samples (ONT Ligation sequencing gDNA kit: SQK-LSK109). The library was then sequenced using nanopore MinION device for 24 hours. An analogous experiment was subsequently conducted, in which the initial dsDNA library was instead prepared using a commercially available kit that ostensibly uses template switching (TS) with supplied adapter ODNs. The sequencing run was also performed for 24 hours. Coverage depth diagrams associated with the two experiments are shown in Figure 3a. Here, the relative read depths indicated that our click chemistry-based library preparation leads to comparable results to those obtained using the standard library preparation, except with a slight decline in coverage towards the 5'-end. Since the CuAAC reaction occurs irrespective of sequence context, this might suggest that the click protocol leads to successful adapter incorporation even in truncated cDNAs that would not ordinarily be ligated in the TS protocol. We attribute the slightly lower total read number in the click-library to normal variations in flow cell performance. Moreover, while 1363 pores yielded sequence data in the TS run, only 942 pores were productive for the Click library (active pore count was measured during the flow cell check via MinKNOW software).

To demonstrate the applicability of our method towards whole transcriptome RNA-seq, dsDNA libraries derived from clicked-cDNA, were prepared using T cell leukemia (Jurkat) total RNA. Following enzymatic end-repair and sequencing adapter ligation, as with our eGFP model system, nanopore sequencing was performed for 48 hours run time. Again, an analogous experiment was conducted, in which the sequencing library was generated by a commercially available kit using template switching (TS), and the library obtained was sequenced under the same conditions as the CuAAC-based sample for a total of 48 hours. The mapping quality of the individual sequencing experiments is shown in Figure 3b. Notably, we observed a higher percent reference identity for the click-generated libraries in comparison to standard library preparation. We speculate that this effect may at least partially be attributed to 5'-guanosine-containing sequences present only in the TS protocol (Figure S9). Read lengths (Figure 3c) of the CuAAC based library preparation were comparable to the standard library preparation, whereas PHRED base call quality scores were slightly higher in the case of the click- chemistry method (Figure 3d). The slight disparity in quality scores can perhaps be attributed to variation in the fidelity of the polymerases employed in each experiment. In the case of the proprietary TS kit, the identities of the polymerases were undisclosed, precluding

further investigation into this effect. Finally, to determine and compare isoform abundance from the nanopore data, we employed the Kallisto algorithm developed by Bray and colleagues (Figure 3e).[22] Gratifyingly, the results show that the vast majority of sequenced gene transcripts are shared between both the click-chemistry and TS method, further demonstrating the viability of click chemistry to enable whole transcriptome RNA sequencing.

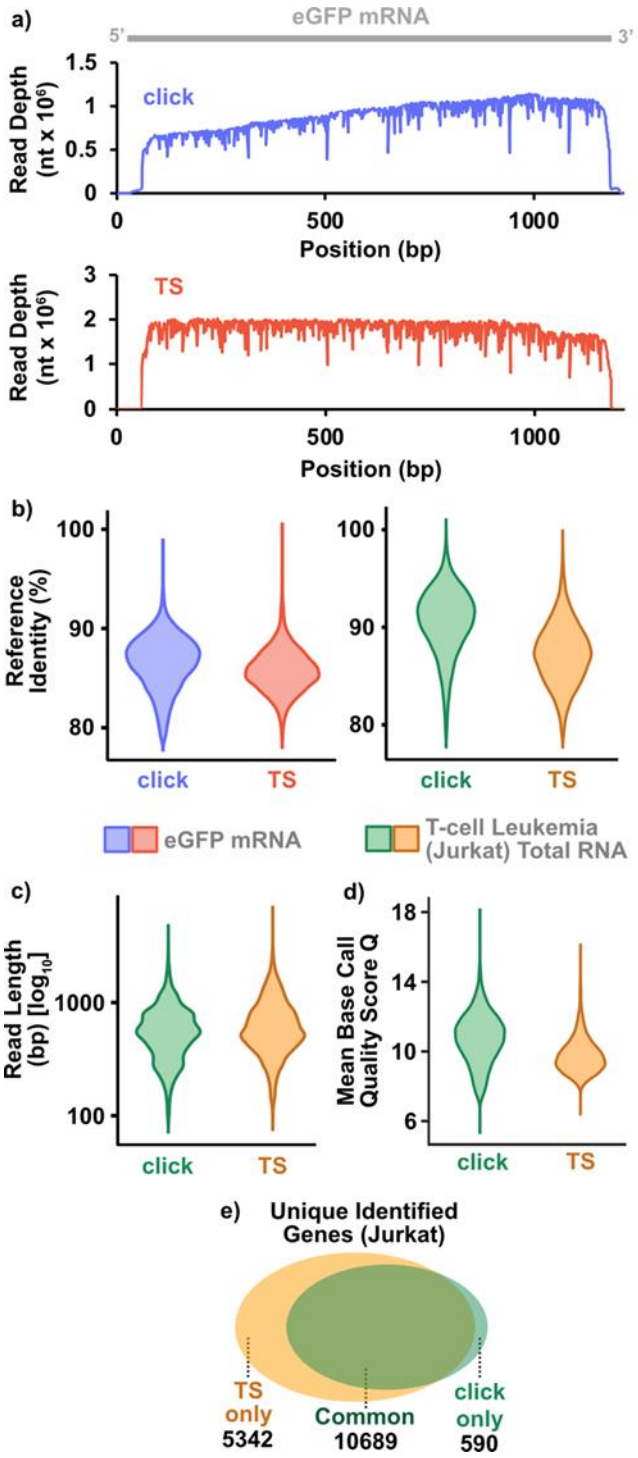

Figure 3. Performance of the Click-libraries in nanopore sequencing. a) Coverage depth (read numbers) on the eGFP model transcript of two independent runs on a MinION device (blue: click library, crimson: TS library). b) Comparison of mapping quality (percent reference identity) of the eGFP model libraries 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 transcripts identified with the Kallisto algorithm.[22]

Finally, an illustrative example of full-length transcript mapping is shown in Figure 4. Specifically, ONT reads derived from Jurkat total RNA were mapped to the GAPDH gene on chromosome 12 using data derived from both a) the click chemistry method, and b) the TS method. As anticipated, the data show that GAPDH exons are well represented using our click-chemistry based library preparation, with a slight bias towards the 3'-end, as compared with the TS-based kit.

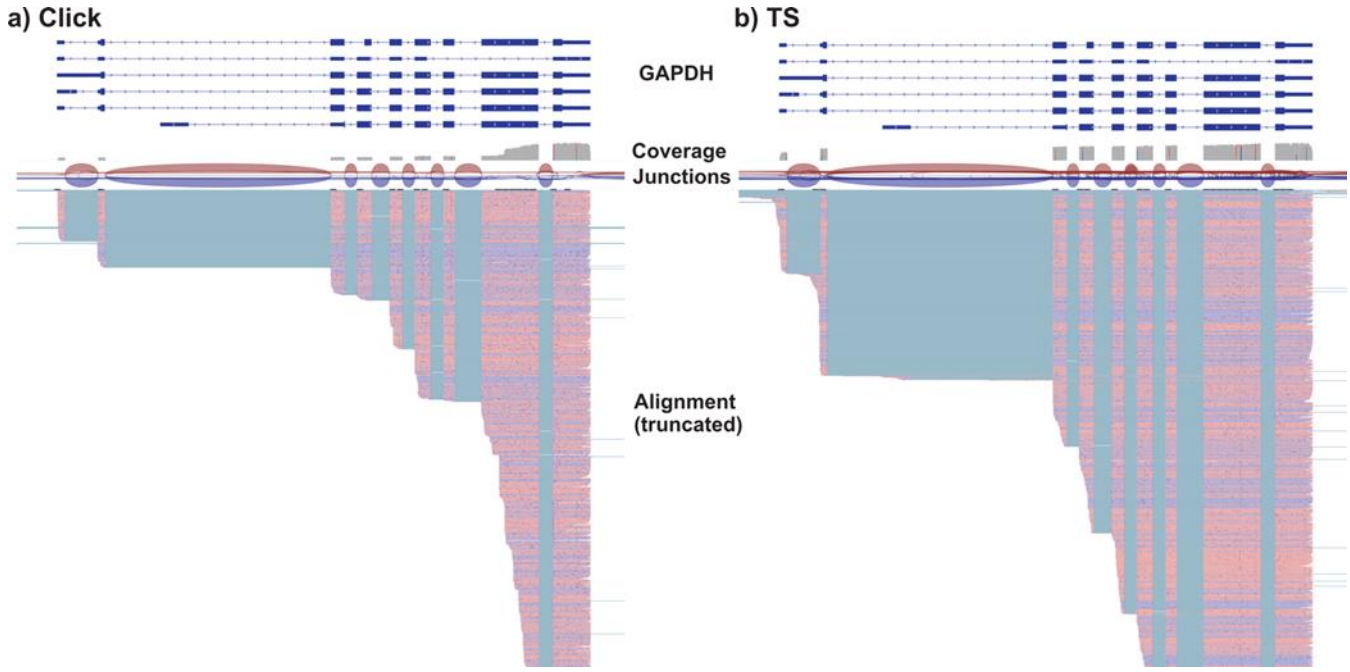

Figure 4. Mapping example, showing the GAPDH gene on chromosome 12 for two Jurkat total RNA libraries sequenced on a MinION nanopore sequencer. Mapping was performed with minimap2 against the GRCh38 reference genome. a) The library that was prepared using the click chemistry protocol. b) The library that was prepared with the standard template switching procedure (TS). The images were prepared with IGV 2.12.2. Reads in the forward direction (or 5' to 3', with respect to the mRNA) are colored red, and those in the reverse direction are colored blue. Splice junctions connecting exons are represented by arcs.

Conclusion

In summary, we here disclose a novel click chemistry-based method for the preparation of Third Generation Sequencing libraries from polyadenylated RNAs. Compared with the state-of-the-art, PAC-seq, full length transcripts are obtained, and thus, our methodology is for the very first time able to contend with existing methods based upon the Smart-Seq2 protocol. As a proof of principle, we successfully explore whole transcriptome sequencing with T-cell leukemia (Jurkat) total RNA. Due to the rapid and chemoselective nature of CuAAC reactions, we foresee possible applications in single cell RNA-seq with unparalleled sensitivity upon further optimization. Unbiased addition of sequencing adapters may also facilitate protocols that rely on blockade of reverse transcription like in SHAPE-seq or sequencing of RNA modifications (e.g. m1A or pseudouridine), which prohibit the application of template switching.[23–26]

Click chemistry-based dsDNA libraries for Third Generation Sequencing were prepared by 3' azido-labelling of reverse transcribed cDNA via anchored poly(dT) primer, followed by a CuAAC and subsequent PCR amplification. All individual steps of the introduced library preparation are reproducible, high yielding and easy to perform. The Cu(I) catalyzed azide-alkyne cycloaddition using BASECLICK's solid heterogeneous Cu(I) catalyst system enables straightforward click conditions without degassing the reaction mixture and simple workup conditions. By the incorporation of an extra, quadruple degenerated nucleotide at the 3'- end of the forward primer the quality of the PCR was enhanced significantly.

We directly compared Nanopore sequencing experiments of the two libraries (eGFP and Jurkat), generated via click chemistry -based and standard protocol (based on template switching reaction) library preparation, respectively. The generated sequencing data proves that the performance of the click-chemistry based library in both

amplicon and transcriptome sequencing can compete with standard protocols.

In conclusion, our novel library preparation method enables the reliable generation and amplification of full-length reverse transcripts for long read sequencing while avoiding templated ligation and overrepresentation of '5-guanosine-containing sequences. Our method offers an attractive alternative to standard methods, potentially alleviating G-bias in template switching protocols, as well as forming the basis for future development of more sophisticated workflows such as single cell- and long-read amplicon sequencing.[27] Given the modular-nature of click chemistry, an extremely diverse range of applications is in principle possible. Our library preparation could further be applied for analyzing differential gene expression and quantifying splicing variants as well as fusion transcripts.

Experimental Section

Custom Oligonucleotides

Short oligonucleotides and primers used in this study were provided by BASECLICK GMBH (Neuried, Germany) and were synthesised using standard phosphoramidite chemistry. HPLC-purified products were stored at -20 °C as lyophilised solids.

mRNA Synthesis

In order to generate cDNA libraries for Third Generation Sequencing, mRNAs encoding for the enhanced green fluorescent protein (eGFP) and T cell leukemia (Jurkat) total RNA were first prepared or purchased as model systems. Jurkat Total RNA was purchased from THERMOFISHER SCIENTIFIC, while the desired mRNA constructs of eGFP (Figure S6) were prepared by in vitro transcription from a linearised DNA template containing a T7 promoter and a 120mer poly(A)tail according to a modified version of a previously established protocol from Croce and co-workers (Figure S5).[17] Reaction mixtures were prepared with the TranscriptAid T7 High Yield Transcription Kit (THERMOFISHER SCIENTIFIC) by mixing the following components in a 200 µL PCR tube: linear plasmid DNA (1 µg); 8 µL NTP mix (2.5X; 100mM ATP/CTP/GTP/UTP solutions); 4 µLTranscriptAid Reaction Buffer (5X); 2 µL TranscriptAid Enzyme Mix and nuclease-free H2O to achieve a final volume of 20 µL. The samples were incubated for 2 h at 37 °C, cooled to 4 °C for 3 min and purified with the QIAquick PCR Purification Kit from QIAGEN. The concentrations were measured with the Qubit™ Fluorometer and the samples quality was analysed via agarose gel electrophoresis.

First Strand cDNA Synthesis

Prior to reverse transcription (RT), the poly(dT) primers (RTP1-RTP6) were hybridised to the template by mixing the following components in a 200 µL PCR microtube: 1 µL of eGFP mRNA (600- 650 ng) or Jurkat total RNA (1-2 µg); 1 µL dNTP Mix (10 mM); 1 µL (RTP1-RTP6; 100 µM), and 10 µL Nuclease-free H_2O to achieve a final volume of 13 µL. After mixing with a pipette, the mixture was briefly centrifuged and transferred to a thermal cycling block with a lid temperature of 105 °C. The samples were then heated to 65 °C for 5 min and cooled to 4°C for 3 min. Afterwards, 4 µL of SuperScript™ IV 5X Buffer, 1 µL of DTT (100 mM), 1 µL of SuperScript™ IV Reverse Transcriptase (200 U/µL) and 1 µL of RNaseOUT™ Recombinant Ribonuclease Inhibitor (40 U/µL) were added to reach a final volume of 20 µL. The resulting samples were mixed and briefly centrifuged, and the following temperature program for the RT was performed: 20 min at 50 °C, 10 min at 80 °C, cooled down to 4 °C and hold for 3 min.

For the following mRNA digestion, to each reaction mixture from the previous step were added to the following components: 3 µL RNase H Reaction Buffer (10X); 1.4 µL RNase H (5 U/µL); 1 µL RNase A (10 mg/µL) and 4.6 µL Nuclease-free H_2O to achieve a final volume of 30 µL. The digestion was performed as follows: 30 min at 37 °C, 15 min at 65 °C, cooled down at 4 °C and hold for 3 min. The digested cDNAs were then purified using a QIAquick PCR Purification Kit from QIAGEN. Elution was performed using nuclease-free H_2O.

Azide Elongation and CuAAC

The samples for the Azide Elongation were prepared by mixing the following components in a 200 µL PCR tube: 15 µL to 17 µL purified cDNA (eGFP: 1000-1300 ng; Jurkat: 900-950 ng); 1 µL 3'-Azido-2',3'-ddGTP (10 mM); 2 µL TdT (20 U/µL); 5 µL TdT Reaction Buffer (5X) and if necessary, sufficient nuclease-free H2O to achieve a final volume of 25 µL. Azide elongation was performed by incubating the sample for 1.5 h at 37 °C, then cooled down to 4 °C and hold for 3 min. The samples were purified as described above, again using nuclease-free H_2O for elution.

CuAAC reactions to ligate the 3'-azido-labelled cDNA with the alkyne adapter (AA1) were prepared by mixing the following components in 1.5 mL tube: a solution containing the single stranded N_3-labelled cDNA (eGFP: 440-460 ng; Jurkat: 450-650 ng) obtained from the previous step in nuclease free H_2O (9 to 28 µL) was combined by mixing with an appropriate volume of BASECLICK's proprietary '5X activator' solution to obtain a 1X final concentrated solution. Afterwards 2 pellets of BASECLICK's proprietary heterogeneous 'reactor' catalyst system; 0.5 µL AA1 (100 µM) and if necessary, nuclease-free H_2O, were added and mixed (final volume 12.5 - 40 µL). The click reactions were performed in the

thermomixer at 45 °C for 1.5 h at 600 rpm. Subsequent purification was performed using a QIAquick PCR Purification Kit from QIAGEN.

PCR Amplification

PCR reactions to generate dsDNA libraries were prepared using a modified Taq DNA Polymerase (LongAmp® Taq 2X Master Mix from NEW ENGLAND BIOLABS) according to the manufacturer's instructions, and with the following profile: 30 s at 94 °C; 35 x [20 s at 94 °C, 30 s at 57 °C, 1-2 min at 65 °C]; final extension 10 min at 65 °C. Individual primer concentrations (FP1-3 and RP1-6 respectively; see table S1) for each experiment were 0.4 µM, and 2.6 to 8.2 ng clicked-cDNA was used as the template for each 25 µL reaction. Annealing temperatures were calculated with respect to the specific primer pair using the OligoAnalyzer™ Tool from INTEGRATED DNA TECHNOLOGIES. Elongation times were calculated according to the anticipated PCR product length using the formula t(s) = fragment size(kB) x 60. PCR products were imaged via agarose gel electrophoresis in Tris-acetate-EDTA buffer and stained using an aqueous solution of approximately 0.2 µg/mL ethidium bromide. PCR products were further purified using the QIAquick PCR Purification Kit from QIAGEN.

ONT Sequencing

Amplified PCR products derived from clicked-cDNAs were converted into sequencing libraries using the Ligation Sequencing Kit (SQK-LSK109), according to the Genomic DNA by Ligation protocol and subsequently sequenced on a MinION flow cell (R10.3) with a MinION sequencing device (OXFORD NANOPORE TECHNOLOGIES). For individual samples, 1 µg of PCR-amplified cDNA was carried through to the first step of the kit protocol.

For standard mRNA sequencing without click chemistry (eGFP mRNA: see mRNA Synthesis; Jurkat Total RNA: purchased from THERMO FISCHER SCIENTIFIC) library preparation was performed according to the cDNA-PCR-Sequencing (SQK-PCS109) kit instructions from OXFORD NANOPORE TECHNOLOGIES. The supplied protocol employs strand switching (TS) with kit-supplied oligonucleotides, as with the Smart-Seq2 protocol. The library was subsequently sequenced on a MinION flow cell (R10.3) with a MinION sequencing device (OXFORD NANOPORE TECHNOLOGIES).

The individual steps of the commercially available kits were performed as directed, with the exception of sample purification, where we instead employed CleanNGS beads from CLEANNA instead of the recommended product Agencourt AMPure XP beads from BECKMAN COULTER.

Bioinformatic analysis

Jurkat cDNA reads derived from the TS library on the MinION device were cleaned for adapter contamination with Porechop (https://github.com/rrwick/Porechop) with default parameters. The click-library was trimmed with Porechop using the custom adapter sequence ACGCTCTTCCGATCTAC/GTAGATCGGAAGAGCG T which can be found on both cDNA ends after PCR amplification with primer FP1. Mapping of reads against the Homo sapiens DNA primary assembly release 105[28] was performed with Minimap2 (https://github.com/lh3/minimap2)[29] using the parameter set "-ax splice". The eGFP amplicon reads were trimmed with Porechop for the custom sequence "GACGCTCTTCCGATCT/AGATCGGAAGAGCGTC" , which is present on both ends of each fragment. For mapping of the eGFP amplicon (see Figure S8) with minimap2, the parameter set "-ax map-ont" was used.

Dataset statistics were calculated and visualised with NanoComp.[30] Mapping graphics were generated with IGV 2.12.2 [31] and assembled with Adobe Illustrator CS6 and Affinity Designer.

Transcript abundance was estimated by generating pseudoalignments against the Ensemble cDNA release 105[28] with kallisto[22] using read-length statistics from the NanoComp output. The Venn diagram was generated with R version 4.1.3 using the packages limma[32]and VennDiagram[33] and adapted with Adobe Illustrator CS6 and Affinity Designer.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Oligonucleotide sequences, agarose gel images, electropherogram data, map of the plasmid, sequences of the mRNA model system (eGFP) and additional sequencing data analysis (PDF).

AUTHOR INFORMATION

Corresponding Author

Thomas Frischmuth - baseclick GmbH
Floriansbogen 2-4, 82061 Neuried (Munich), Germany
* E-mail: t.frischmuth@baseclick.eu

Author Contributions

E. S. S. and A. C. contributed equally to this work.

Funding Sources

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 861381 (NATURE-ETN) and 642023 (ClickGene) and by baseclick GmbH.

Notes

We declare that Thomas Frischmuth and Thomas Carell are shareholders of baseclick GmbH.

ACKNOWLEDGMENT

A. C. and E. S. gratefully acknowledge Dr. Pascal Giehr for his thoughtful advice in the preparation of ONT sequencing libraries. Álvaro Cuadros-Inostrozais thanked for conducting preliminary BLAST searches.

REFERENCES

(1) Green, M. R. Pre-MRNA Splicing. *Annu. Rev. Genet.* **1986**, *20*, 671–708. https://doi.org/10.1146/annurev.ge.20.120186.003323.

(2) Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. Function of Alternative Splicing. *Gene* **2013**, *514* (1), 1–30. https://doi.org/10.1016/j.gene.2012.07.083.

(3) Tian, Y.; Wen, H.; Qi, X.; Zhang, X.; Liu, S.; Li, B.; Sun, Y.; Li, J.; He, F.; Yang, W.; Li, Y. Characterization of Full-Length Transcriptome Sequences and Splice Variants of Lateolabrax Maculatus by Single-Molecule Long-Read Sequencing and Their Involvement in Salinity Regulation. *Front. Genet.* **2019**, *10*. https://doi.org/10.3389/fgene.2019.01126.

(4) Harrow, J.; Frankish, A.; Gonzalez, J. M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B. L.; Barrell, D.; Zadissa, A.; Searle, S.; Barnes, I.; Bignell, A.; Boychenko, V.; Hunt, T.; Kay, M.; Mukherjee, G.; Rajan, J.; Despacio-Reyes, G.; Saunders, G.; Steward, C.; Harte, R.; Lin, M.; Howald, C.; Tanzer, A.; Derrien, T.; Chrast, J.; Walters, N.; Balasubramanian, S.; Pei, B.; Tress, M.; Rodriguez, J. M.; Ezkurdia, I.; van Baren, J.; Brent, M.; Haussler, D.; Kellis, M.; Valencia, A.; Reymond, A.; Gerstein, M.; Guigó, R.; Hubbard, T. J. GENCODE: The Reference Human Genome Annotation for the ENCODE Project. *Genome Res.* **2012**, *22* (9), 1760–1774. https://doi.org/10.1101/gr.135350.111.

(5) Oikonomopoulos, S.; Bayega, A.; Fahiminiya, S.; Djambazian, H.; Berube, P.; Ragoussis, J. Methodologies for Transcript Profiling Using Long-Read Technologies. *Front. Genet.* **2020**, *11*. https://doi.org/10.3389/fgene.2020.00606.

(6) Picelli, S.; Björklund, Å. K.; Faridani, O. R.; Sagasser, S.; Winberg, G.; Sandberg, R. Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat. Methods* **2013**, *10* (11), 1096–1100. https://doi.org/10.1038/nmeth.2639.

(7) Ramsköld, D.; Luo, S.; Wang, Y. C.; Li, R.; Deng, Q.; Faridani, O. R.; Daniels, G. A.; Khrebtukova, I.; Loring, J. F.; Laurent, L. C.; Schroth, G. P.; Sandberg, R. Full-Length MRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells. *Nat. Biotechnol.* **2012**, *30* (8), 777–782. https://doi.org/10.1038/nbt.2282.

(8) Picelli, S.; Faridani, O. R.; Björklund, Å. K.; Winberg, G.; Sagasser, S.; Sandberg, R. Full-Length RNA-Seq from Single Cells Using Smart-Seq2. *Nat. Protoc.* **2014**, *9* (1), 171–181. https://doi.org/10.1038/nprot.2014.006.

(9) Zhu, Y. Y.; Machleder, E. M.; Chenchik, A.; Li, R.; Siebert, P. D. Reverse Transcriptase Template Switching: A SMART™ Approach for Full-Length CDNA Library Construction. *Biotechniques* **2001**, *30* (4), 892–897. https://doi.org/10.2144/01304pf02.

(10) Saliba, A. E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-Cell RNA-Seq: Advances and Future Challenges. *Nucleic Acids Res.* 2014, pp 8845–8860. https://doi.org/10.1093/nar/gku555.

(11) Wu, A. R.; Wang, J.; Streets, A. M.; Huang, Y. Single-Cell Transcriptional Analysis. *Annu. Rev. Anal. Chem.* **2017**, *10*, 439–462. https://doi.org/10.1146/annurev-anchem-061516-045228.

(12) Wulf, M. G.; Maguire, S.; Humbert, P.; Dai, N.; Bei, Y.; Nichols, N. M.; Corrêa, I. R.; Guan, S. Non-Templated Addition and Template Switching by Moloney Murine Leukemia Virus (MMLV)-Based Reverse Transcriptases Co-Occur and Compete with Each Other. *J. Biol. Chem.* **2019**, *294* (48), 18220–18231. https://doi.org/10.1074/jbc.RA119.010676.

(13) Cocquet, J.; Chong, A.; Zhang, G.; Veitia, R. A. Reverse Transcriptase Template Switching and False Alternative Transcripts. *Genomics* **2006**, *88* (1), 127–131. https://doi.org/10.1016/j.ygeno.2005.12.013.

(14) Fantoni, N. Z.; El-Sagheer, A. H.; Brown, T. A Hitchhiker's Guide to Click-Chemistry with Nucleic Acids. *Chem. Rev.* **2021**, *121* (12), 7122–7154. https://doi.org/10.1021/acs.chemrev.0c00928.

(15) Jaworski, E.; Routh, A. ClickSeq: Replacing Fragmentation and Enzymatic Ligation with Click-Chemistry to Prevent Sequence Chimeras. *Methods Mol. Biol.* **2018**, *1712*, 71–85. https://doi.org/10.1007/978-1-4939-7514-3_6.

(16) Routh, A.; Head, S. R.; Ordoukhanian, P.; Johnson, J. E. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3′-Azido CDNAs. *J. Mol. Biol.* **2015**, *427* (16), 2610–2616. https://doi.org/10.1016/j.jmb.2015.06.011.

(17) Croce, S.; Serdjukow, S.; Carell, T.; Frischmuth, T. Chemoenzymatic Preparation of Functional Click-Labeled Messenger RNA. *ChemBioChem* **2020**, *21* (11), 1641–1646. https://doi.org/10.1002/cbic.201900718.

(18) Kollaschinski, M.; Sobotta, J.; Schalk, A.; Frischmuth, T.; Graf, B.; Serdjukow, S. Efficient DNA Click Reaction Replaces Enzymatic Ligation. *Bioconjugate Chem.* **2020**, *31* (3), 507–512. https://doi.org/10.1021/acs.bioconjchem.9b00805.

(19) Elrod, N. D.; Jaworski, E. A.; Ji, P.; Wagner, E. J.; Routh, A. Development of Poly(A)-ClickSeq as a Tool Enabling Simultaneous Genome-Wide Poly(A)-Site Identification and Differential Expression Analysis. *Methods* **2019**, *155*, 20–29. https://doi.org/10.1016/j.ymeth.2019.01.002.

(20) Routh, A.; Ji, P.; Jaworski, E.; Xia, Z.; Li, W.; Wagner, E. J. Poly(A)-ClickSeq: Click-Chemistry for next-Generation 3'-End Sequencing without RNA Enrichment or Fragmentation. *Nucleic Acids Res.* **2017**, *45* (12). https://doi.org/10.1093/nar/gkx286.

(21) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. Biocompatible Artificial DNA Linker That Is Read through by DNA Polymerases and Is Functional in Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (28), 11338–11343. https://doi.org/10.1073/pnas.1101519108.

(22) Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* **2016**, *34* (5). https://doi.org/10.1038/nbt.3519.

(23) Li, X.; Xiong, X.; Wang, K.; Wang, L.; Shu, X.; Ma, S.; Yi, C. Transcriptome-Wide Mapping Reveals Reversible and Dynamic N1-Methyladenosine Methylome. *Nat. Chem. Biol.* **2016**, *12* (5). https://doi.org/10.1038/nchembio.2040.

(24) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; León-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A. Transcriptome-Wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of NcRNA and MRNA. *Cell* **2014**, *159* (1). https://doi.org/10.1016/j.cell.2014.08.028.

(25) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. v. Pseudouridine Profiling Reveals Regulated MRNA Pseudouridylation in Yeast and Human Cells. *Nature* **2014**, *515* (7525). https://doi.org/10.1038/nature13802.

(26) Lucks, J. B.; Mortimer, S. A.; Trapnell, C.; Luo, S.; Aviran, S.; Schroth, G. P.; Pachter, L.; Doudna, J. A.; Arkin, A. P. Multiplexed RNA Structure Characterization with Selective 2′-Hydroxyl Acylation Analyzed by Primer Extension Sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (27). https://doi.org/10.1073/pnas.1106501108.

(27) Jensen, M. K.; Elrod, N. D.; Yalamanchili, H. K.; Ji, P.; Lin, A.; Liu, Z.; Wagner, E. J. Application and Design Considerations for 3′-End Sequencing Using Click-Chemistry. In *Methods Enzymol.*; Academic Press Inc., 2021; Vol. 655, pp 1–23. https://doi.org/10.1016/bs.mie.2021.03.012.

(28) Cunningham, F.; Allen, J. E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M. R.; Armean, I. M.; Austine-Orimoloye, O.; Azov, A. G.; Barnes, I.; Bennett, R.; Berry, A.; Bhai, J.; Bignell, A.; Billis, K.; Boddu, S.; Brooks, L.; Charkhchi, M.; Cummins, C.; da Rin Fioretto, L.; Davidson, C.; Dodiya, K.; Donaldson, S.; el Houdaigui, B.; el Naboulsi, T.; Fatima, R.; Giron, C. G.; Genez, T.; Martinez, J. G.; Guijarro-Clarke, C.; Gymer, A.; Hardy, M.; Hollis, Z.; Hourlier, T.; Hunt, T.; Juettemann, T.; Kaikala, V.; Kay, M.; Lavidas, I.; Le, T.; Lemos, D.; Marugán, J. C.; Mohanan, S.; Mushtaq, A.; Naven, M.; Ogeh, D. N.; Parker, A.; Parton, A.; Perry, M.; Pilizota, I.; Prosovetskaia, I.; Sakthivel, M. P.; Salam, A. I. A.; Schmitt, B. M.; Schuilenburg, H.; Sheppard, D.; Perez-Silva, J. G.; Stark, W.; Steed, E.; Sutinen, K.; Sukumaran, R.; Sumathipala, D.; Suner, M. M.; Szpak, M.; Thormann, A.; Tricomi, F. F.; Urbina-Gómez, D.; Veidenberg, A.; Walsh, T. A.; Walts, B.; Willhoft, N.; Winterbottom, A.; Wass, E.; Chakiachvili, M.; Flint, B.; Frankish, A.; Giorgetti, S.; Haggerty, L.; Hunt, S. E.; Iisley, G. R.; Loveland, J. E.; Martin, F. J.; Moore, B.; Mudge, J. M.; Muffato, M.; Perry, E.; Ruffier, M.; Tate, J.; Thybert, D.; Trevanion, S. J.; Dyer, S.; Harrison, P. W.; Howe, K. L.; Yates, A. D.; Zerbino, D. R.; Flicek, P. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50* (D1). https://doi.org/10.1093/nar/gkab1049.

(29) Li, H. New Strategies to Improve Minimap2 Alignment Accuracy. *Bioinformatics* **2021**, *37* (23). https://doi.org/10.1093/bioinformatics/btab705.

(30) de Coster, W.; D'Hert, S.; Schultz, D. T.; Cruts, M.; van Broeckhoven, C. NanoPack: Visualizing and Processing Long-Read Sequencing Data. *Bioinformatics* **2018**, *34* (15). https://doi.org/10.1093/bioinformatics/bty149.

(31) Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P.

Integrative Genomics Viewer. *Nat. Biotechnol.* 2011. https://doi.org/10.1038/nbt.1754.

(32) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* **2015**, *43* (7). https://doi.org/10.1093/nar/gkv007.

(33) Hanbo Chen. *VennDiagram: Generate High-Resolution Venn and Euler Plots*. https://CRAN.R-project.org/package=VennDiagram . https://CRAN.R-project.org/package=VennDiagram (accessed 2022-06-15).

Table of contents graphic