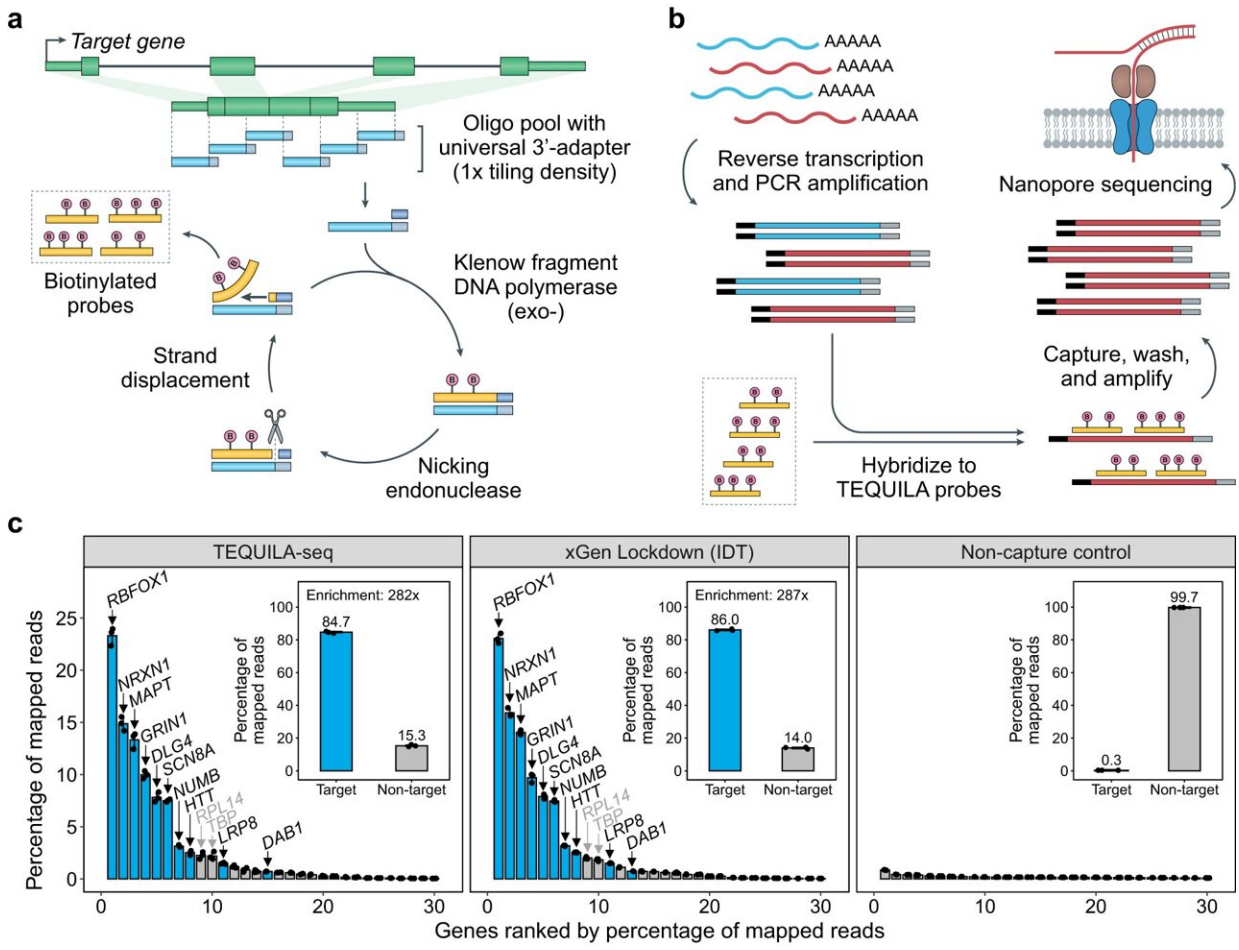
Researchers develop versatile and low-cost technology for targeted long-read RNA sequencing

August 15 2023



Overview of TEQUILA-seq. a, b Schematic of TEQUILA-seq. a Single-stranded DNA (ssDNA) oligonucleotides are designed to tile across all annotated exons of target genes and synthesized using an array-based DNA synthesis technology. TEQUILA probes are amplified from ssDNA oligo templates in a single pool

using nickase-triggered strand displacement amplification with universal primers and biotin-dUTPs. b Full-length cDNAs are synthesized from poly(A)+ RNAs by reverse transcription and PCR amplification. TEQUILA probes are then hybridized to cDNAs. Upon capture and washing, cDNA-to-probe hybrids are immobilized to streptavidin magnetic beads, whereas unbound cDNAs are washed away. Captured cDNAs are further amplified by PCR and subjected to nanopore 1D library preparation and sequencing. c Comparison of TEQUILA-seq vs xGen Lockdown (IDT) probe-based target enrichment and sequencing. Main graphs: percentage of reads mapped to a given gene (mean ± s.d. of n = 3 replicates), for the top 30 genes with the highest number of mapped reads. Insets: percentage of reads mapped to target genes and non-target genes (mean ± s.d. of n = 3 replicates). Blue: target genes. Gray: non-target genes. Target gene panel: ten human genes with long transcripts in the brain. All sequencing methods were applied to the same human brain RNA mix from multiple donors. Credit: *Nature Communications* (2023). DOI: 10.1038/s41467-023-40083-6

In a development that could accelerate the discovery of new diagnostics and treatments, researchers at Children's Hospital of Philadelphia (CHOP) have developed a versatile and low-cost technology for targeted sequencing of full-length RNA molecules. The technology, called TEQUILA-seq, is highly cost-effective compared to commercially available solutions for targeted RNA sequencing and can be adapted for different research and clinical purposes. The details are described in a paper in *Nature Communications*.

On the journey from gene to protein, an RNA molecule can be cut and joined in different ways before being translated into a protein. This process is known as alternative splicing, and it allows a single gene to encode several different proteins. Although alternative splicing occurs in many biological processes, it can be dysregulated in diseases like cancer, leading to pathogenic RNA molecules. To understand how alternative splicing might lead to disease, researchers need to have accurate

accounting of all the RNA molecules (known as "transcript isoforms") that emanate from a single gene.

One way of doing so is using "long-read" RNA sequencing platforms, which sequence RNA molecules over 10,000 bases in length end-to-end, capturing the entirety of the transcript isoforms. However, these long-read platforms have modest sequencing yield, which has hampered their widespread adoption, especially in the clinical setting, as generating long-read RNA sequencing data at clinically informative depth could be prohibitively expensive. Targeted sequencing, which involves enriching specific nucleic acid sequences of interest prior to sequencing, is a useful strategy that can substantially enhance coverage of predefined targets, but the cost and complexity of target capture have been barriers to wider use.

"Targeted long-read RNA sequencing is a powerful strategy for elucidating the RNA repertoire for any predefined set of genes. However, existing technologies for targeted sequencing of full-length RNA molecules are either expensive or difficult to set up, putting them out of reach for many labs," said co-senior author Lan Lin, Ph.D., Assistant Professor of Pathology and Laboratory Medicine and a member of the Raymond G. Perelman Center for Cellular and Molecular Therapeutics at CHOP. "TEQUILA-seq solves that problem by being both inexpensive and easy to use. The technology can be adapted by users for different purposes, and researchers can choose which genes they want to sequence and make the reagents for target capture in their own labs. This has the potential to accelerate discovery of new diagnostic and therapeutic solutions for a wide range of diseases."

One method that allows for targeted sequencing is called hybridization capture-based enrichment, which uses short pieces of nucleic acids called oligonucleotides as capture probes. These oligonucleotides (often simply referred to as "oligos") are tagged with biotin molecules and

designed to hybridize to their targets based on nucleic acid sequence complementarity, which allows for easy capture and isolation of their target sequences from a biological sample. However, although hybridization capture-based enrichment is an efficient method for targeted sequencing, commercially synthesized biotinylated capture probes are expensive and can only be used for a limited number of reactions, making the per-sample cost high for each capture reaction.

To address this limitation, the CHOP researchers developed TEQUILA-seq (Transcript Enrichment and Quantification Utilizing Isothermally Linear-Amplified probes in conjunction with long-read sequencing). A key innovation in TEQUILA-seq is a nicking-endonuclease triggered isothermal strand displacement amplification reaction, which can synthesize large quantities of biotinylated capture probes from a cheap pool of non-biotinylated oligos as the templates. Using an input of only 2 ng of template oligos, the researchers can generate 25 ug of TEQUILA probes, which can be used for at least 250 capture reactions. This innovative strategy for synthesizing capture probes makes TEQUILA-seq highly cost-effective and scalable for large target panels and many biological samples.

To benchmark its performance, the researchers performed TEQUILA-seq for multiple gene panels on synthetic RNAs or human RNAs. TEQUILA probes performed as well as commercial capture probes in target capture and enrichment, while being hundreds of times cheaper for each capture reaction. Moreover, the researchers demonstrated that TEQUILA-seq can substantially enhance detection while preserving quantification of target RNA molecules.

"Using cheap reagents and an easy experimental workflow, TEQUILA-seq allows us to deeply sequence the full-length RNA molecules for any gene set across many biological samples," said co-senior author Yi Xing, Ph.D., director of theCenter for Computational and Genomic Medicine

at CHOP. "This is very exciting and enables a wide range of medical applications, from RNA-guided genetic diagnosis to therapy development."

To illustrate its biomedical utility, the researchers applied TEQUILA-seq to profile full-length RNA molecules of 468 actionable cancer genes across 40 breast cancer cell lines. They discovered previously unknown transcript isoforms in extensively studied cancer genes that may shed light on how genes that protect the body from cancer are inactivated in individual tumors.

"Our work shows that TEQUILA-seq can be broadly used for targeted sequencing of full-length RNA molecules," Dr. Lin said. "Moreover, TEQUILA probes are general-purpose capture probes. They are compatible for both targeted RNA and DNA sequencing, on both long-read and short-read sequencing platforms. The ability to easily generate large quantities of biotinylated capture probes for any target panel at a low cost and with ease can facilitate large-scale and population-level studies for many basic, translational, and clinical applications."

More information: Feng Wang et al, TEQUILA-seq: a versatile and low-cost method for targeted long-read RNA sequencing, *Nature Communications* (2023). DOI: 10.1038/s41467-023-40083-6

Provided by Children's Hospital of Philadelphia