



Revealing nascent RNA processing dynamics with nano-COP

Heather L. Drexler^{1,4}, Karine Choquet^{1,4}, Hope E. Merens¹, Paul S. Tang²,
Jared T. Simpson^{2,3} and L. Stirling Churchman¹✉

During maturation, eukaryotic precursor RNAs undergo processing events including intron splicing, 3'-end cleavage, and polyadenylation. Here we describe nanopore analysis of co-transcriptional processing (nano-COP), a method for probing the timing and patterns of RNA processing. An extension of native elongating transcript sequencing, which quantifies transcription genome-wide through short-read sequencing of nascent RNA 3' ends, nano-COP uses long-read nascent RNA sequencing to observe global patterns of RNA processing. First, nascent RNA is stringently purified through a combination of 4-thiouridine metabolic labeling and cellular fractionation. In contrast to cDNA or short-read-based approaches relying on reverse transcription or amplification, the sample is sequenced directly through nanopores to reveal the native context of nascent RNA. nano-COP identifies both active transcription sites and splice isoforms of single RNA molecules during synthesis, providing insight into patterns of intron removal and the physical coupling between transcription and splicing. The nano-COP protocol yields data within 3 d.

This protocol is an extension to: *Nat. Protoc.* **11**, 813–833 (2016): <https://doi.org/10.1038/nprot.2016.047>

Introduction

Splicing is largely co-transcriptional^{1–3}. Most studies that have characterized co-transcriptional splicing have measured the length of time between the synthesis of an intron and its removal by splicing in living cells^{4–8}. These studies, which have considerably advanced our understanding of splicing kinetics at individual introns, have shown that splicing in mammalian cells typically takes several minutes but that this interval varies substantially across different introns^{4–6,8–10}. However, current approaches are unable to report on the role of transcription in splicing and the coordination of splicing across introns. The ability to study the position of RNA polymerase (Pol) II relative to sites of splicing is critical for our understanding of the spatial coupling between transcription and splicing, which directly informs mechanistic models of co-transcriptional splicing. Furthermore, patterns of splicing across multi-intron genes have the potential to influence alternative splicing decisions¹¹ since differing kinetic rates and dependencies on neighboring introns contribute to mechanisms that control alternative splicing^{12–14}. However, it has been challenging to capture evidence of splicing coordination. Consequently, studies to date have been largely limited to analyzing the order of splicing within individual genes^{11,15,16} or between pairs of introns throughout the genome¹⁷. These efforts have revealed that human introns are frequently spliced out of order, consistent with the idea that splicing occurs on a timescale of minutes, enabling a downstream intron to be spliced before an upstream one. However, to expand these studies to the analysis of more complex splicing patterns, such as splicing coordination, as well as comprehensively understand the general rules (and exceptions) governing these patterns, we need to be able to analyze splicing over multiple introns as transcripts are synthesized.

Here we provide a step-by-step protocol for nanopore analysis of co-transcriptional processing (nano-COP)¹⁸, a method that we recently developed for monitoring early RNA processing by directly sequencing long nascent RNAs. nano-COP combines stringent purification of nascent RNA, long-read sequencing without reverse transcription (RT) or PCR amplification, and rigorous computational analyses. In developing this protocol, at every step we sought to minimize technical biases that

¹Department of Genetics, Harvard Medical School, Boston, MA, USA. ²Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ³Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. ⁴These authors contributed equally: Heather L. Drexler, Karine Choquet.
✉e-mail: churchman@genetics.med.harvard.edu

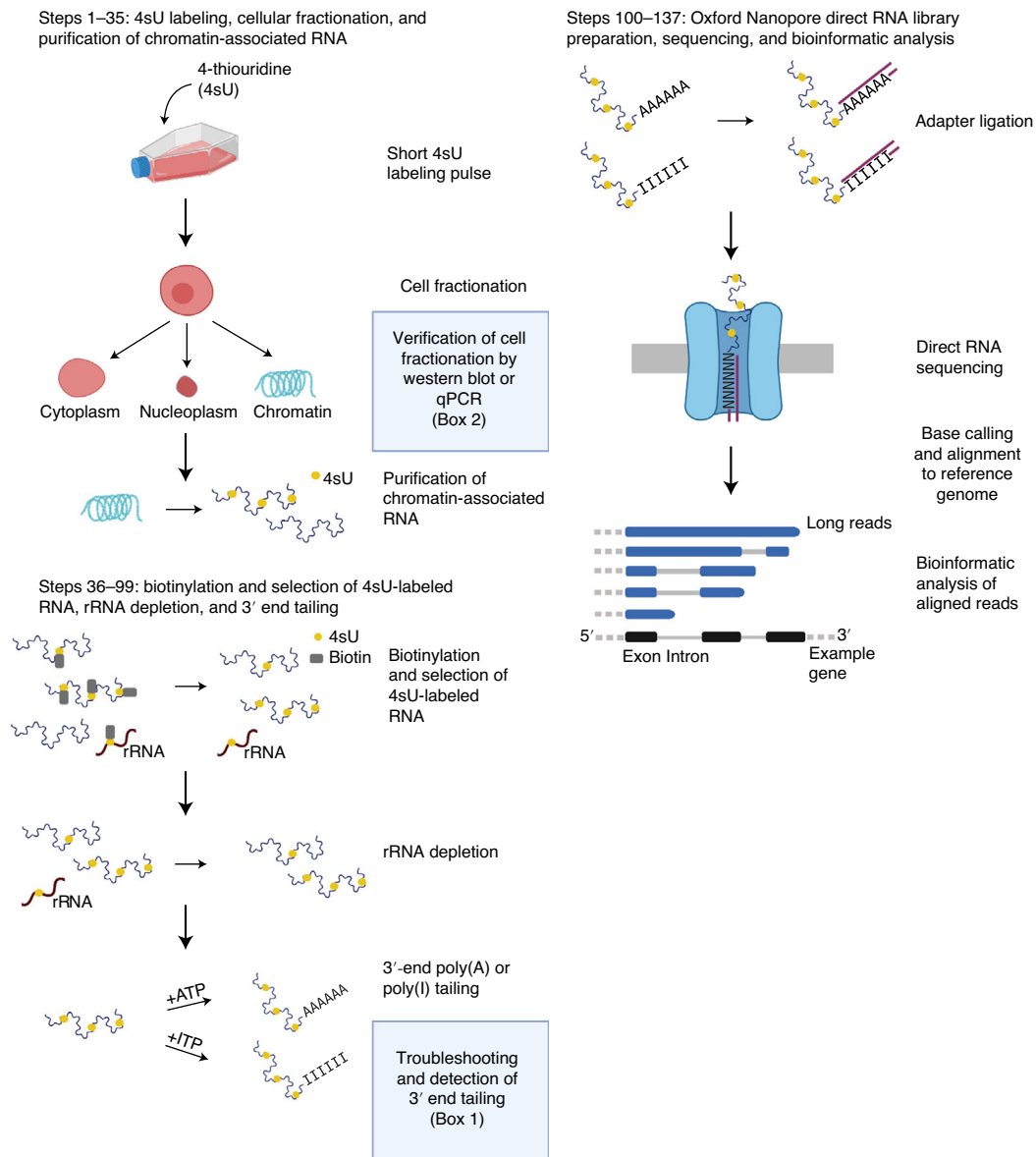


Fig. 1 | nano-COP schematic. Outline of the nano-COP procedure, providing an overview of the critical experimental and computational steps in the protocol. Created with BioRender.com.

have the potential to alter interpretation, thereby achieving a faithful representation of the splicing status of each RNA during its transcription.

Development of the protocol

We set out to build a sequencing approach that captures snapshots of nascent RNAs as they are transcribed and processed, in order to monitor the kinetics of splicing and reveal splicing dynamics across multi-intron genes (Fig. 1). To accomplish this goal, we needed a strategy to capture RNAs as they were synthesized. Initially, we turned to native elongating transcript sequencing (NET-seq), a technique previously developed by our group, which reveals the position of Pol II by sequencing the 3' ends of nascent RNA^{19,20}.

To capture nascent RNA in mammalian cells, NET-seq takes advantage of the strong interaction between the RNA–DNA–Pol II ternary complex on chromatin. The 3' end of nascent RNA represents the active site of Pol II transcription, as this is where new nucleotides are incorporated into the nascent transcript. When millions of nascent RNA 3' ends are aligned to the genome, NET-seq reads

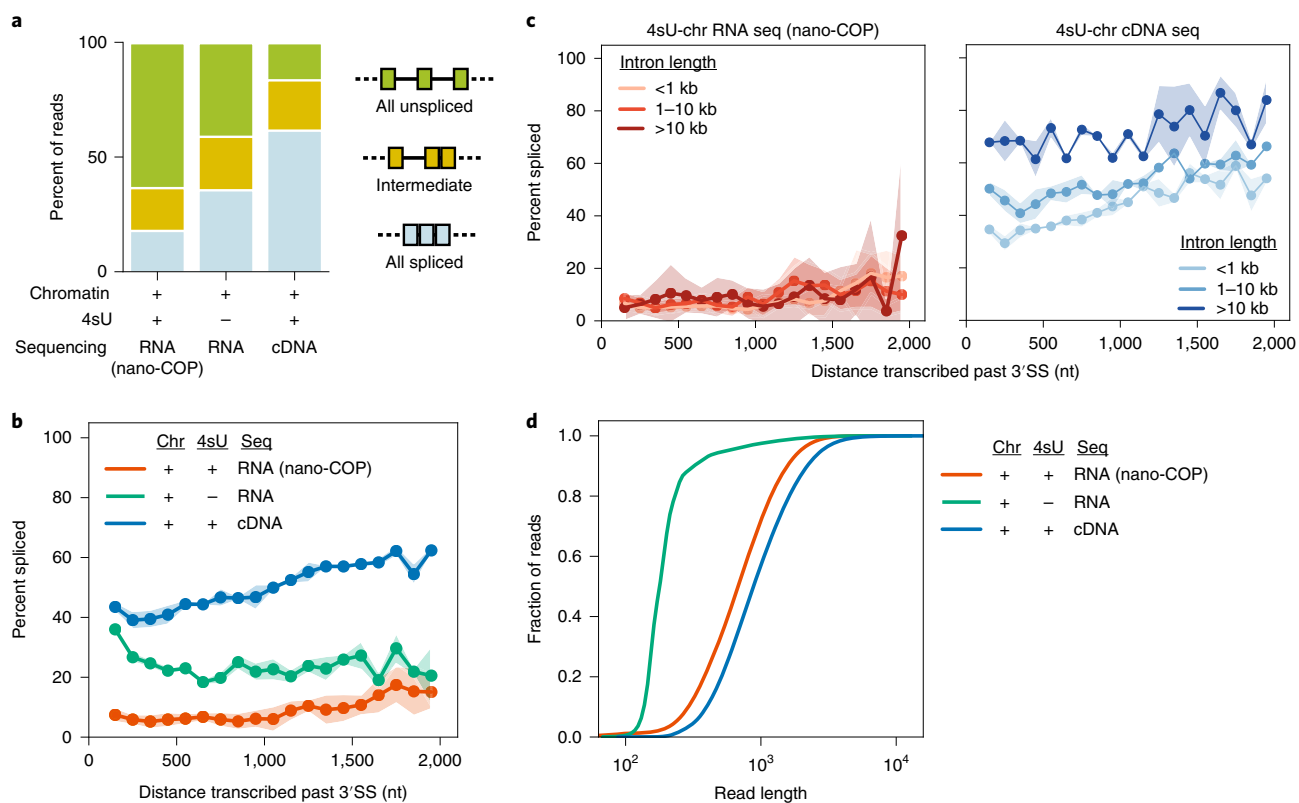


Fig. 2 | Direct RNA sequencing of 4sU-labeled chromatin-associated RNA provides the most accurate measurement of the nascent transcriptome. **a**, Distribution of splicing patterns obtained with different library preparation methods in human K562 cells. Among reads spanning at least two introns, ‘all spliced’ represents reads in which every intron within the read is spliced, ‘intermediate’ represents reads in which at least one intron is spliced and one intron is unspliced, and ‘all unspliced’ represents reads in which every intron is present and therefore unspliced. **b**, Global analysis of distance transcribed from the 3’ SS and the percentage of spliced molecules in human K562 cells using the indicated library preparation methods ($N = 72,937$ 4sU-chr RNA; $N = 14,227$ chr RNA; $N = 34,463$ 4sU-chr cDNA). **c**, Global analysis of distance transcribed from the 3’ SS and the percentage of spliced molecules separated by intron length for RNA or cDNA sequencing of 4sU-chr RNA ($N = 37,489$ < 1 kb 4sU-chr RNA; $N = 32,246$ 1–10 kb 4sU-chr RNA; $N = 3,202$ > 10 kb 4sU-chr RNA; $N = 15,952$ < 1 kb 4sU-chr cDNA; $N = 16,149$ 1–10 kb 4sU-chr cDNA; $N = 2,362$ > 10 kb 4sU-chr cDNA). Similar trends were observed across transcript lengths (data not shown). Shaded regions in **b** and **c** represent standard deviation across two to five biological replicates. Each category includes both poly(A)- and poly(I)-tailed replicates. **d**, Cumulative distribution plot of read lengths passing the default base calling threshold for one representative sample from each library preparation method using poly(A) tailing. The methods employed for the preparation of libraries other than nano-COP are detailed in Supplementary Methods. chr, chromatin purification; 4sU, 4sU labeling and purification; seq, sequencing. Nano-COP samples shown in this figure were previously published in ref. ¹⁸.

reveal the global population of active transcription sites. However, NET-seq is not suitable for analyzing splicing kinetics or patterns of intron removal since the reads are too short to observe long-range interactions.

To re-engineer the NET-seq methodology to observe nascent RNA processing patterns, we made the following modifications during the development of the nano-COP protocol:

- 1 *Direct RNA sequencing of long nascent RNAs.* Due to its reliance on short-read Illumina sequencing, NET-seq cannot capture long-distance splicing patterns or dynamics. Therefore, we modified the sequencing approach to obtain longer reads. Several methods for long-read RNA sequencing have been developed; but most approaches first convert the RNA to cDNA using enzymatic reactions, including RT, terminal transferase template switching, and PCR amplification. These library preparation steps are prone to adding size biases that favor the capture of shorter molecules, which tend to be spliced RNAs.

Currently, the two main long-read sequencing platforms are Pacific Biosciences (PacBio) SMRT sequencing for cDNA^{21,22} and Oxford Nanopore Technologies (ONT) sequencing for cDNA or RNA²³. Although the PacBio platform has higher accuracy derived from circular consensus sequencing, we decided to use the ONT platform due to its ability to sequence RNA directly.

When comparing the ONT direct cDNA sequencing approach, which relies on template-switching RT but not PCR amplification, with the direct RNA approach, which sequences long

nascent RNAs directly through nanopores without the need for RT or PCR amplification, we found that the cDNA approach overrepresents spliced RNAs (Fig. 2a and Supplementary Methods). Presumably, the requirement for the RT reaction to reach the 5' end and complete terminal transferase activity, an event that is more likely to occur with short spliced RNAs than with long unspliced RNAs, biases cDNA libraries towards spliced RNAs²⁴, especially in human genes with long introns. A long-read cDNA sequencing analysis of *Arabidopsis thaliana* chromatin-associated RNA did not appear to suffer from the same level of bias associated with cDNA sequencing that we observed in the development of nano-COP (Fig. 2), possibly because of the much shorter intron size in plants compared to human genes²⁵.

In cDNA libraries, even without PCR amplification, biases led to the detection of spliced introns before Pol II had transcribed 100 bp (Fig. 2b) with many ends even within the footprint of Pol II where splicing is physically inhibited. Based on average transcription rates of 1–4 kb/min^{26,27}, this would suggest that splicing was completed very rapidly i.e., within a few seconds. By contrast, direct RNA sequencing detected only a small fraction of spliced introns within 100 bp transcribed past the 3' splice site (3'SS), with most splicing occurring after Pol II had transcribed several kilobases; this is consistent with established splicing times of 5–10 min in mammalian cells^{4–6,8–10}. Presumably, partially processed pre-mRNAs or short fragmented RNAs are preferentially reverse-transcribed to completion during cDNA library preparation, confounding the interpretation of nascent RNA processing. Indeed, long introns and long transcripts tend to be spliced in cDNA libraries, whereas intron and transcript lengths do not have an influence on splicing levels with direct RNA sequencing (Fig. 2c). Although nanopore sequencing of cDNA produces longer reads than direct RNA sequencing (Fig. 2d), library preparation seems to bias the size distribution of the molecules that pass through the nanopore in the first place. Therefore, to capture the most accurate representation of the nascent transcriptome, nano-COP uses direct RNA nanopore sequencing which requires the least amount of library manipulation and in turn minimizes size biases prior to sequencing.

- 2 *Stringent purification of nascent RNA.* In designing nano-COP, we modified the purification of nascent RNA. In NET-seq, the purification procedure captures chromatin-associated RNA that is enriched for nascent RNA being actively transcribed. However, a large proportion of NET-seq reads are derived from short non-coding RNAs (ncRNAs), including snRNAs and tRNAs, which are sequenced alongside the nascent RNAs. NET-seq libraries are typically sequenced to >100 million reads, so a substantial number of reads representing the nascent RNAs of interest remain even after the computational removal of the reads corresponding to short ncRNAs. However, because direct RNA nanopore sequencing is lower-throughput (~1 million sequenced reads per run), chromatin RNA purification alone would not be sufficient to produce adequately informative nano-COP libraries.

Hence, nano-COP incorporates an additional nascent RNA purification step to decrease the proportion of small non-coding RNAs that are sequenced. In addition to cellular fractionation, metabolic labeling and selection is also performed in order to increase the stringency of nascent RNA purification^{28–30}. Briefly, we add to the medium a short pulse (8 min) of 4-thiouridine (4sU), which is incorporated into nascent RNA as it is transcribed. After chromatin fractionation and purification, the labeled molecules are isolated through biotinylation followed by streptavidin pulldown. Although 4sU is a non-canonical base and the ONT software is not currently designed to recognize it, we did not observe a detectable effect on base-calling uridines after direct RNA sequencing of nano-COP samples (Supplementary Fig. 1 in this paper and Supplemental Fig. S2d from ref. ¹⁸).

This combined approach to purifying nascent RNA enables a selection of RNAs that are both recently transcribed (by metabolic labeling) and localized to the site of transcription (by cellular fractionation). The RNA obtained by the combined purification of 4sU-labeled chromatin-associated (4sU-chr) RNA exhibits less splicing than RNA obtained by either purification strategy individually (Fig. 2a,b and ref. ¹⁸), indicating that the combination achieves more stringent purification of nascent RNA. Moreover, direct RNA sequencing of 4sU-chr RNA results in a substantially lower proportion of short (<200 nt) RNAs than in chromatin-associated RNA purified without metabolic labeling, yielding a higher proportion of reads usable in downstream analyses (Fig. 2d). Consistent with the idea that chromatin-associated RNA contains a population of older, more processed, RNA that has remained on chromatin, chromatin-associated RNA sequencing reveals more overall splicing than 4sU-chr RNA (Fig. 2a). In addition, as in the cDNA samples, direct sequencing of chromatin-associated RNA without additional purification yields reads with

3' ends close to 3' SSs, resulting in an apparent measurement of splicing within 100 bp transcribed past the 3' SS (Fig. 2b). Presumably, these reads are derived from older chromatin-associated RNAs that remain attached to chromatin after premature termination or are in the process of being degraded, or are undergoing post-transcriptional processing and are fragmented during purification. Taken together, these observations demonstrate that combining 4sU labeling with chromatin purification results in material that is much better suited for accurate measurement of co-transcriptional splicing dynamics.

- 3 *Enzymatic tailing.* The third major aspect of the NET-seq protocol that we modified during the development of nano-COP was the approach used to tag the 3' ends, which marks the location of Pol II. In NET-seq, a common nucleic acid primer is ligated to the nascent RNA 3' ends, facilitating library construction and 3'-end sequencing. The ligation method is not compatible with nanopore sequencing because the dideoxynucleotide at the 3' end of the primer, which prevents ligation concatemers, blocks the necessary subsequent ligation of the nanopore adapter. Instead, nano-COP uses enzymatic tailing to prepare 3' ends of nascent RNA for nanopore sequencing. The direct RNA nanopore adapter is designed to recognize and sequence poly(A) tails from mature mRNA. Therefore, our initial nano-COP protocol added artificial poly(A) tails to the purified 4sU-chr RNA using poly(A) polymerase (Box 1 and Fig. 3). In this approach naturally polyadenylated RNAs, such as mature mRNA or polyadenylated pre-mRNAs still associated with chromatin, are also sequenced. Hence, to avoid interpreting RNAs with natural poly(A) tails as RNAs that are still undergoing transcription, we computationally remove all reads that end near annotated poly(A) sites when using poly(A) tailing. However, a downside of this initial approach is the inability to discriminate natural poly(A) tails from those that are added *in vitro*. To circumvent this ambiguity, we developed an alternative tailing protocol that adds a stretch of inosines rather than adenosines. The ONT base calling software accurately distinguishes the entire tail region from the start of the transcript, regardless of the composition of the homopolymers. Importantly in this regard, we have developed an algorithm that distinguishes poly(I), poly(A), and poly(A)-poly(I) tails based on differences in the raw nanopore signal (Box 1 and Fig. 4). In this manner, reads derived from naturally polyadenylated mRNAs can be identified.

Below, we present a step-by-step protocol for performing nano-COP in cultured suspension cells (Fig. 1). The protocol was originally applied to human K562 and BL1184 and *Drosophila* S2 cells, and we expect that it will be applicable to other cell lines. We anticipate that our detailed descriptions of the technical development of nano-COP and the related computational analyses will help researchers apply this approach to their own studies.

Applications of the method

As nano-COP captures and sequences long RNAs that are being actively synthesized and processed, it has the potential to address many questions related to *in vivo* co-transcriptional RNA processing. Thus far, nano-COP analysis has focused on RNA splicing¹⁸, but it can also be used to investigate other RNA processing events, including microRNA, snoRNA, and rRNA processing.

nano-COP reveals multiple aspects of RNA splicing dynamics. First, it can provide insights into the physical coupling between transcription and splicing by capturing the distance between the position of Pol II (the 3' end of the nascent RNA) and the 3' SS of introns within a given RNA. After tallying the splicing status of each intron, nano-COP provides a global estimate of the distance Pol II transcribes before splicing. Currently, nanopore technology provides only a few reads per transcript. Accordingly, nano-COP has primarily been used to uncover global trends of splicing dynamics. However, as coverage increases, nano-COP will become capable of measuring splicing kinetics at individual introns. In addition, we anticipate that increases in nano-COP sequencing depth will allow analyses currently performed using NET-seq to be performed by nano-COP, e.g., measurement of the association of Pol II pausing with intron splicing or quantification of RNA processing in unstable antisense transcripts.

Second, nano-COP can be used to analyze the interplay between 3'-end processing and splicing. As a fraction of nano-COP reads are derived from RNAs that have completed transcription, the data can be used to assess whether splicing occurs before or after 3'-end processing. By analyzing RNA 3'-end mapping locations, nano-COP can determine whether introns are spliced co- or post-transcriptionally. Furthermore, the poly(I) tailing approach improves the identification of endogenously polyadenylated transcripts, and should therefore be able to probe the relationship between splicing, poly(A) site choice, cleavage, and polyadenylation. Nano-COP also captures intermediates of

Box 1 | Troubleshooting and detection of 3' end tailing

Nano-COP requires the addition of a 3' end tail to nascent RNA to allow for library preparation. Two alternative protocols are summarized below:

	poly(A) tailing	poly(I) tailing
Enzyme	<i>E. coli</i> poly(A) polymerase	Yeast poly(A) polymerase
Incubation time	7.5 min	30 min
Adapter	RTA from ONT kit (RTA-T ₁₀)	RTA-C ₁₀
Advantages	-Use of standard ONT kit and protocol -Shorter incubation time	-Higher proportion of nascent RNA -Discrimination between naturally polyadenylated and non-polyadenylated transcripts

If using another enzyme or tailing approach, we recommend verifying its efficiency by 3' end-tailing of a short RNA oligonucleotide, such as oGAB11, and visualizing the size shift by electrophoresis (Fig. 3). The goal is to identify an appropriate enzyme and incubation time for adding a 10–100 nt tail to all RNA molecules.

Procedure

Troubleshooting of 3' end-tailing with oGAB11

- Mix 0.5 µl of 20 µM oGAB11 with 9.5 µl of 10 mM Tris pH 7.0 for each condition.
- Prepare the necessary buffers with oGAB11 dilution from the previous step (step 1 in Box 1) as the RNA:
 - Poly(A)-tailing: follow Step 96A.
 - Poly(I)-tailing: follow Step 96B.
- Denature samples at 80 °C for 2 min in a thermal cycler.
- Add 10 µl of reaction buffer to the sample.
- Incubate at 37 °C for the desired time.
- Add 0.5 µl of 500 mM EDTA after the incubation is complete and leave on ice.
- Clean up by precipitation as in Steps 83–91 and resuspend in 10 µl of 10 mM Tris-HCl (pH 7).
- Perform electrophoresis on a TBE-urea polyacrylamide gel to check the size of the RNA (Fig. 3). Run the Low Range ssRNA ladder (NEB) and/or ss10 DNA ladder (Simplex Sciences) alongside RNA samples. See Steps 37–42 of the NET-seq protocol for more details³⁸.

Detection of poly(A) and poly(I) tails in nano-COP data

In previous work, we developed nanopolish-polya to computationally segment the raw nanopore signal using a Hidden Markov Model, which estimates the polyadenylated tail length⁵⁰. The raw signal from poly(A) and poly(I) tails can be distinguished visually (Fig. 4a). Thus, we extended nanopolish-polya to differentiate between poly(A) and poly(I) tails in nano-COP data (nanopolish-detect-polyI). Nanopolish-detect-polyI is a classifier that distinguishes between poly(A)-tailed, poly(I)-tailed, and poly(A)-poly(I)-tailed reads by combining two Hidden Markov models for segmentation and changepoint detection at the signal-trace level (for more details, see the Supplementary Note). To validate the method, we tailed the in vitro transcribed RNA ERCC-00048 in three different ways: (i) addition of a poly(A) tail; (ii) addition of a poly(I) tail; or (iii) addition of a poly(A) tail followed by addition of a poly(I) tail to mimic a naturally polyadenylated transcript that gets poly(I)-tailed in vitro. In the first two cases, nanopolish-detect-polyI correctly identified the tail identity in >93% of reads (Fig. 4b). In the last case, nanopolish-detect-polyI correctly identified the dual tail in 87% of reads (Fig. 4b). As expected, application of poly(I) tailing and the nanopolish-detect-polyI program to human nano-COP samples showed only a small proportion of poly(A)-poly(I) tails (1–3%), whereas poly(A)-selected mRNAs had poly(A)-poly(I) tails present on the majority of reads (Fig. 4c). One potential limitation is that poly(A)-poly(I) tails may be more difficult to identify in the current version if either tail is too short (< 30 nt).

To use nanopolish-detect-polyI:

- Download and install the version of nanopolish with the polyI detection module:

```
> git clone --recursive https://github.com/jts/nanopolish.git
> cd nanopolish
> make
```

- Index reads for use with nanopolish

```
> /path/to/nanopolish index --directory=/path/to/sample_name/fast5_pass/ --sequencing-summary=/path/to/sample_name.sequencing_summary.txt /path/to/sample_name.fastq
```

- Use nanopolish tail detector

```
> path/to/nanopolish detect-polyi --threads=8 --reads=/path/to/sample_name.fastq --bam=/path/to/sample_name.bam --genome=/path/to/reference_genome > /path/to/output_file.txt
```

In the output file, the column detected indicates the identity of the tail for each read. Importantly, in this first version of nanopolish-detect-polyI the column polya_length represents the estimated length of the entire tail, including poly(A) and poly(I) when they are both present in a read. Nanopolish-detect-polyI was trained with data obtained on R9.4 flow cells. When new flow cell models are released, we will update nanopolish-detect-polyI as necessary.

the splicing reaction through the tailing of free 3'-ends of upstream exons after the first catalytic step of splicing. Analysis of reads derived from this population provides insight into the efficiency and regulation of the catalytic steps of splicing under different conditions.

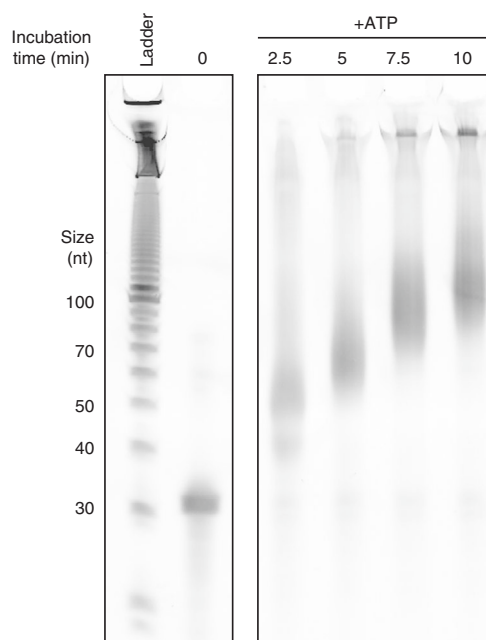


Fig. 3 | Troubleshooting the incubation time for 3' end poly(A) tailing with oGAB11. oGAB11 was polyadenylated using Clontech *E. coli* poly(A) polymerase in the presence of ATP as described in Box 2. Tailing resulted in a time-dependent size shift of oGAB11. After 7.5 min, the incubation time selected for nano-COP, >40 adenosines have been added to most RNAs, which is more than sufficient for ligation of the direct RNA sequencing RTA containing ten T's. The unprocessed image is shown in Source Data.

Third, nano-COP can detect the impact of perturbations on nascent RNA processing. For instance, nano-COP following treatment with the splicing inhibitor pladienolide B revealed a decrease in co-transcriptional splicing and the proportion of splicing intermediates¹⁸. Thus, through the introduction of genetic and environmental perturbations, nano-COP can reveal how co-transcriptional RNA processing is regulated by *cis*-elements, *trans*-acting factors, or environmental stress.

Finally, nano-COP can reveal the order of intron removal in single nascent RNA molecules by comparing the splicing status of neighboring introns in the same read. Furthermore, the long reads obtained by nanopore sequencing can elucidate higher-order splicing patterns over multiple introns within the same transcript. For example, analysis of reads spanning four or more introns can indicate whether introns are generally spliced in the order in which they are transcribed, or if instead other patterns are observed and at what frequency, providing insight into how isoforms are determined as they are produced.

Comparison with other approaches

Several alternative sequencing methods are available for analyzing splicing kinetics. Techniques relying on metabolic labeling, such as transient transcriptome sequencing (TT-seq), have paired 4sU labeling with RNA fragmentation and short-read sequencing over several time points in order to study gene expression dynamics^{30–32}, and these techniques have also been adapted to model splicing dynamics^{8,33}. Using computational models of gene transcription and RNA processing, these approaches are able to estimate splicing times faster than their sampling rates. Studies using these methods have confirmed that the timescale of splicing in mammalian cells is ~10 min^{8,9}, comparable to results obtained using more direct methods^{4–6}. Another technique, nano-ID, detects long RNAs that have been labeled with 5-ethynyl uridine (5EU) using direct RNA nanopore sequencing and computational modeling of the raw sequence³⁴. nano-ID is designed to detect recently synthesized RNA isoforms and uses 60-min 5EU pulse times and poly(A)-selected mRNA. Thus, nano-ID captures RNAs that have completed transcription and cannot analyze co-transcriptional RNA processing kinetics or patterns.

As an alternative to modeling kinetics based on timepoints, splicing kinetics can be estimated by determining the distance that Pol II transcribes past the 3'SS of an intron before it is spliced. Single molecule intron tracking (SMIT) achieved this goal by mapping the 3' ends of chromatin-associated RNA relative to the proportion of spliced molecules at individual loci³⁵. SMIT measures splicing

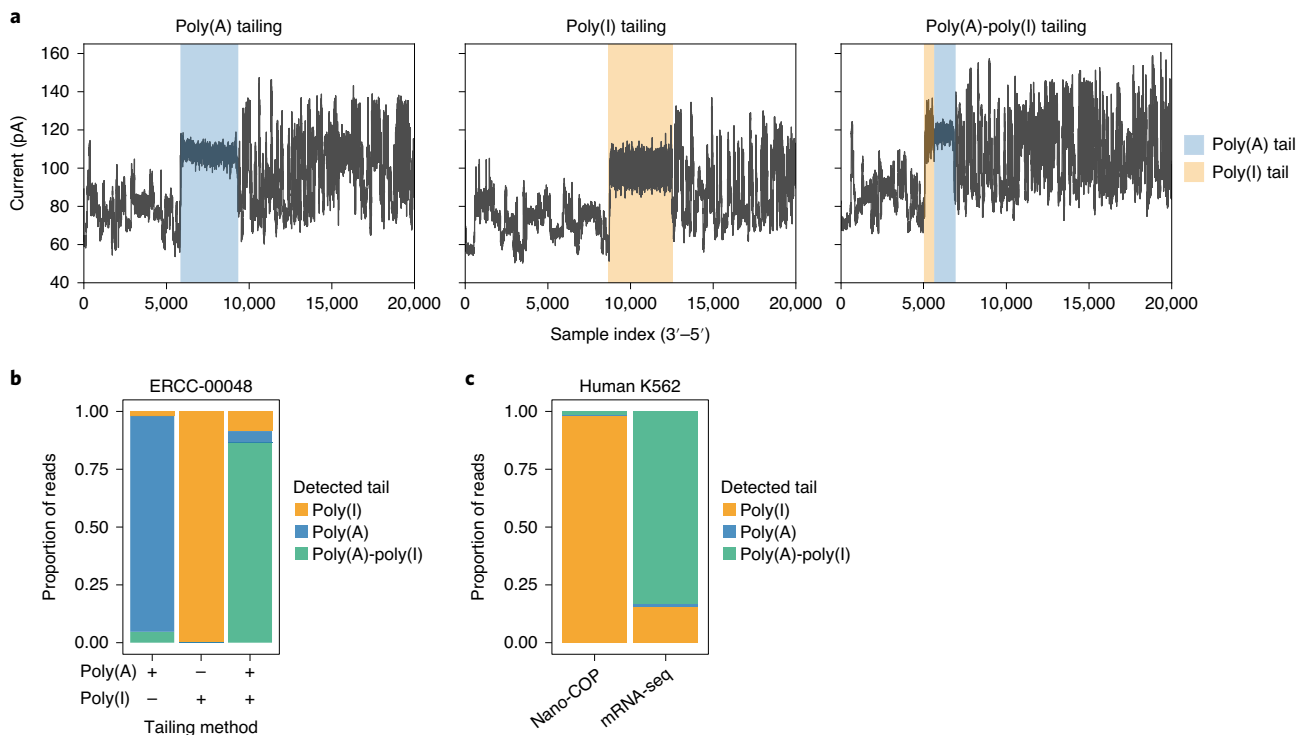


Fig. 4 | Detection of poly(A) and poly(I) tails in ONT direct RNA sequencing data. **a**, Ionic current traces (from 3' to 5') for reads obtained from sequencing of ERCC-00048 with a poly(A) tail (left), a poly(I) tail (middle), or poly(A) and poly(I) tails (right). Poly(A) tails result in a low variance region near the 3' end (highlighted in blue), whereas poly(I) tails result in a similar signal with higher variance (highlighted in orange). Only the first 20,000 samples of each read are shown. **b**, Detection of tail identity using nanopolish-detect-polyI on reads obtained from sequencing of ERCC-00048 with the three tailing conditions described in **a**. **c**, Detection of tail identity using nanopolish-detect-polyI on nano-COP with poly(I) tailing and direct mRNA-seq following poly(I) tailing of polyA⁺ RNA (Supplementary Methods). Both types of samples were from K562 cells. Tail identity is shown for reads ending in gene bodies, at poly(A) sites, or up to 500 nt downstream of poly(A) sites.

status at individual loci using RT-PCR with primers targeting both an upstream exon and the adapter ligated to the 3' ends of RNAs. However, to yield enough material for paired-end Illumina sequencing the sample is amplified using many rounds of PCR, restricting the method to short introns (~300 nt or less) in order to reduce size biases. Read-length distributions of sequenced intronless yeast genes have been used to estimate and subsequently correct for residual length bias³⁵. Unfortunately, SMIT remains unsuitable for most human introns, which typically span several kilobases.

As a parallel approach to sequencing-based methods, microscopy has been used to analyze splicing kinetics. Most commonly, studies using fluorescence microscopy use expression constructs harboring sequences containing RNA hairpins, such as MS2 loops, that are recognized and bound by fluorescently tagged proteins^{6,7,36}. Based on the positioning of the stem-loops, transcription and splicing kinetics for a particular transcript or intron can be estimated. For instance, by placing stem-loops within an intron, the appearance and subsequent drop-off of fluorescent signal reports on the transcription and splicing of the intron, respectively^{6,7,10}. Protein binding on stem-loops can decrease splicing efficiency⁷, and may therefore have an effect on splicing kinetics. Nevertheless, these approaches have also confirmed splicing in human cells to occur on a timescale of several minutes, comparable to results obtained by other methods. However, the insertion of stem-loops is time-intensive, and it is therefore impractical to use imaging methods to examine intron splicing throughout the genome.

Most of our understanding of the order of intron removal emerged from studies of single genes using RT-PCR^{11,15,16}. These studies determined the relative splicing order for up to four consecutive introns by comparing the size and abundance of different PCR products. More recently, short-read sequencing analysis of intron pairs has permitted global analysis of splicing order¹⁷. However, short-read sequencing does not allow analyses of higher-order splicing patterns with three introns or more. Indeed, only techniques based on long-read sequencing of nascent RNA, either by PacBio³⁷ or ONT direct RNA sequencing¹⁸, are able to globally analyze higher-order splicing patterns in multi-intron transcripts.

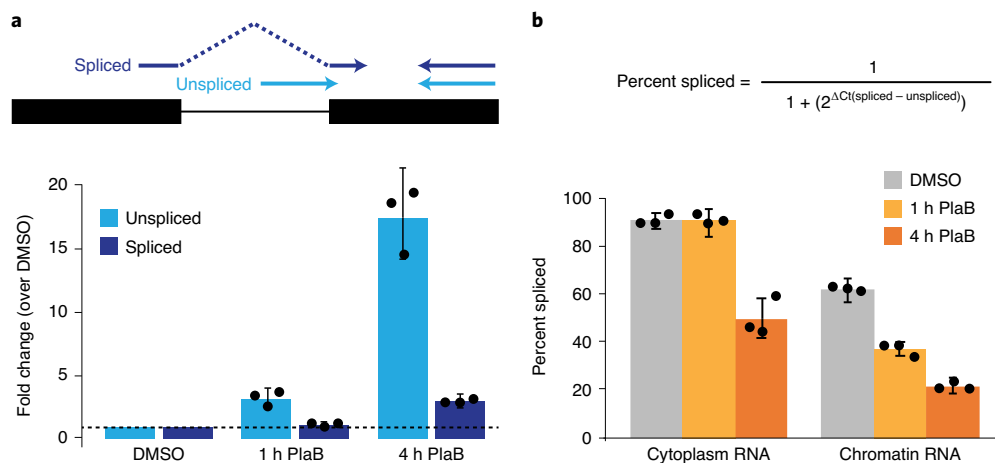


Fig. 5 | Representative RT-qPCR plots of RNA purified by cellular fractionation with varying incubation times in the presence of the splicing inhibitor pladienolide B (PlaB). **a**, The forward ‘spliced’ primer is designed over a splice junction, whereas the forward ‘unspliced’ primer is just upstream of the 3’ intron–exon junction. The reverse primer is the same for both PCR reactions in the downstream exon. The proportions of spliced and unspliced molecules for intron 5 of the *BRD2* gene are measured for the indicated durations of incubation with 100 mM PlaB before chromatin RNA purification, and are represented as fold change relative to the DMSO sample. Black dots represent individual qPCR reactions and error bars represent values calculated from standard deviation of the triplicate samples. **b**, Percentage spliced is determined by calculating the proportion of spliced molecules to total molecules. For intron 5 of the *BRD2* gene, percentage spliced was compared between cytoplasmic and chromatin-associated RNA for the indicated durations of incubation with 100 mM PlaB.

Experimental design

General considerations

The nano-COP protocol is modular. Hence, in order to suit their particular biological question and experimental needs, users may wish to follow only certain sections. One important consideration is that nano-COP requires a considerable amount of input material, as nascent RNA represents a very small proportion of total RNA in the cell and direct RNA sequencing currently requires a large amount of RNA as input. Thus, nano-COP may not be applicable to samples with low cell numbers or low growth rates. It may also be difficult to process multiple samples at once, especially for the cellular fractionation steps, and this should be taken into account when designing experiments. Furthermore, to ensure rapid collection and cell lysis following the short 4sU pulse, nano-COP was developed using suspension or semi-adherent cells. Although we expect the protocol to be easily adaptable to adherent cells, further optimization may be required.

Cell fractionation and purification of chromatin-associated RNA

Each cell fractionation reaction used in this protocol was initially optimized for an input of 10 million human K562 cells or 50 million S2 cells^{20,38}. However, because nano-COP combines cell fractionation with 4sU pulse labeling, a larger amount of input material is required. Therefore, we recommend performing 10–12 cell fractionation reactions per sample, for a total of 100–120 million K562 cells or 500–600 million S2 cells, in order to obtain sufficient material (‘reaction’ refers to each cell fractionation of 10 million human cells or 50 million *Drosophila* cells, and ‘sample’ refers to all the combined reactions for one condition). Increasing the number of cells per reaction is not recommended because it would decrease cell lysis efficiency and increase the risk of cross-contamination from other subcellular fractions²⁰. One reaction per sample can be used to verify cell fractionation by western blot^{20,38} or qRT-PCR (Box 2). Cell fractionation is performed in the presence of α -amanitin to avoid run-on transcription^{20,39} and RNase and protease inhibitors to avoid RNA and protein degradation. Furthermore, all steps are performed on ice or at 4 °C. Considering the number of reactions required per sample (10–12), we recommend performing the cellular fractionation for one sample at a time. Moreover, to avoid lengthy incubations between steps, we recommend performing no more than 12 reactions at a time. If it is preferable to process two samples simultaneously (i.e., control and treatment), we recommend performing a smaller number of reactions per sample (6) and repeating the experiment twice if necessary. After cellular fractionation, reactions are combined for

Box 2 | qRT-PCR verification of cell fractions

Procedure

Measuring splicing levels in RNA purified by cellular fractionation (Steps 12, 18 and 19), metabolic labeling (Steps 61 and 81), or a combined strategy can help troubleshoot reaction conditions across cell lines and experiments (Fig. 5).

- 1 For the cell fractionation samples kept from Steps 12, 18 and 19 as well as the metabolic labeling flow through from Step 61, purify RNA using the miRNeasy mini kit (QIAGEN) following the manufacturer's instructions, including the optional on-column DNase treatment using the RNase-free DNase set. For samples kept from metabolic labeling in Step 81, proceed directly to the next step (Box 2 step 2).
- 2 Perform RT of 0.1–1 µg of RNA using the Invitrogen Superscript III First Strand Synthesis System or an equivalent kit.
 - ▲ **CRITICAL STEP** Perform RT using random hexamers only and omit oligo(dT), as most nascent RNA do not have poly(A) tails.
- 3 Prepare quantitative PCR (qPCR) reactions(s) with the resultant cDNA using SsoFast EvaGreen Supermix (2×) or an equivalent. Typically, we dilute each cDNA sample 1:10 before performing qPCR. However, concentrations may vary depending on the starting amount of RNA and the expression level of the target gene.
 - ▲ **CRITICAL STEP** It is important to run a standard curve with each primer set to ensure that the results are quantitative. For the standard curve, prepare cDNA by making a dilution series of one sample (or a mix of multiple samples) with the following ratios 1:3, 1:9, 1:27, 1:81, 1:243, and 1:729. Run these dilutions with each primer set along with samples.
- 4 Run a qPCR thermal cycling program to capture quantitative measures of spliced and unspliced RNA species. We use the following program:
 - Initial denaturation: 95 °C for 30 s.
 - Amplification: 40 cycles of 95 °C for 5 s to 55 °C for 10 s (fluorescence captured after every extension step).
 - Melting curve: continuous step from 65 °C to 95 °C at 0.5 °C increments for 5 s each (fluorescence captured after every 5 s hold).

An example primer design is shown in Fig. 5a. For human K562 cells, we often use primers assessing the splicing levels of *BRD2* intron 5 (Fig. 5):

Name	Sequence (5'-3')	Description
BRD2_e5e6_junct	AAGTTGGCAGCGCTCCAGG	Exon-exon junction forward primer
BRD2_i5e6_junct	CTTTTTTCTAGCGCTCCAGG	Intron-exon junction forward primer
BRD2_e6_R	GGGAATGTTGAGGACAGTGG	Reverse primer in downstream exon

RNA purification using the QIAzol reagent and isopropanol precipitation based on a modification of a protocol described in ref. ³⁰.

4sU labeling, biotinylation, and selection of 4sU-labeled RNA

The metabolic labeling and purification section of the protocol is largely based on the TT-seq method³⁰, which uses a 5-min 4sU pulse to isolate newly transcribed RNAs. We recommend labeling cells for 8 min, which yields enough nascent RNA for nanopore sequencing (>500 ng) from 100 million human K562 cells or 500 million *Drosophila* S2 cells. The labeling time may need to be optimized for other cell lines to ensure that enough material is obtained. For example, approximately fourfold more 4sU-chr RNA is obtained from K562 cells than the same number of BL1184 cells (120 million). We also found that labeling B-lymphoblast cells for 10 min instead of 8 min resulted in a 1.5-fold increase in yield without changing the downstream results. The resultant 4sU labeled RNA is biotinylated and selected for.

3'-end tailing, library preparation, and direct RNA sequencing

Direct nanopore RNA sequencing requires a poly(A) tail for annealing and ligation of the ONT reverse transcription adapter (RTA). As most nascent RNA is not yet polyadenylated, we developed two alternative approaches for adding either a poly(A) or a poly(I) tail to the 3' ends of nascent RNAs in vitro before proceeding to library preparation. The characteristics and advantages of each approach are described in Box 1. We also recommend troubleshooting 3'-end tailing on a short RNA oligonucleotide before proceeding to library preparation, as we have observed that efficiencies differ among enzymes (Box 1).

After 3'end tailing, we follow the direct RNA sequencing library preparation protocol from ONT, using the instructions for a single sample with the MinION device or for several samples with the PromethION device. We opted to perform the optional RT step suggested in the ONT direct RNA sequencing protocol. The resultant cDNA strand is not sequenced, but is believed to minimize secondary structures in the sequenced RNA strand that can interfere with its threading through the pore, thereby improving sequencing throughput. The resulting library is then sequenced on a MinION or a PromethION for 48 h.

Data processing and analysis

Base calling, or conversion of the raw nanopore signal into nucleotide sequences, is performed either during sequencing ('live') using the ONT MinKNOW interface if computing power is sufficient

Box 3 | Identifying constitutively spliced introns from short-read RNA-seq data

When assessing splicing kinetics with nano-COP, such as the distance transcribed before splicing occurs (Fig. 2b), it may be useful to determine which introns are constitutively spliced in mature RNA in order to avoid the confounding effect of intron retention. Indeed, if certain introns are retained in mature transcripts this could give the illusion of slower splicing than what is observed for introns that are fully spliced.

To take this into account we developed a measure called 'intron stringency', which is calculated using short-read total RNA-seq from the cell line of interest. It is best to use total RNA-seq data rather than polyA+ mRNA-seq to obtain more accurate levels of intron retention from transcripts that may not be polyadenylated. Briefly, we analyze reads spanning 5' and 3' splice junctions for all introns to determine the proportion that exhibits evidence of splicing based on the CIGAR string. We defined four different stringency levels based on the number of aligned reads (minimum coverage), the number of nucleotides spanning 5' and 3' splice junctions (minimum overlap), and the percentage of spanning reads that are spliced:

Stringency level	Minimum coverage (reads)	Minimum overlap (nucleotides)	Minimum spliced (%)
High	50	5	90
Medium	20	4	80
Low	10	3	50
No	0	3	0

'High' and 'medium' stringency represent introns that are constitutively spliced in mature RNA in the cell line of interest. 'Low' stringency represents introns that show some intron retention level in mature RNA, which could confound analyses of splicing kinetics. 'No' stringency represents all introns that are covered in the RNA-seq dataset regardless of read depths or splicing rates, and so includes those with high levels of intron retention.

We recommend using the 'medium' stringency level for analyzing splicing kinetics of constitutively spliced introns. Alternatively, results from different stringency levels can be compared. We found that varying the stringency level led to modest differences in the global distance between transcription and splicing in *Drosophila* S2 cells, but only minor changes in human K562 cells¹⁸.

To measure intron stringency level, use the script `intronCoord_stringency_from_RNAseq.py`, available at <https://github.com/churchmanlab/nano-COP/>:

```
> python intronCoords_from_RNAseq.py transcriptomeCoordinates.txt input.bam output.txt --min_overlap
--min_cov --min_splice
```

The necessary files are:

- `transcriptomeCoordinates.txt`: input file with exon and intron coordinates in BED format (e.g., `<genome>_merge_parsed_sortByNameCoord.bed` produced in Step 114).
- `input.bam`: BAM file containing uniquely mapped reads from short-read RNA-seq aligned with the STAR aligner⁵¹ with default parameters, except: `readFilesCommand=cat, limitIObufferSize=200000000, limitBAMsortRAM=64000000000, outReadsUnmapped=Fastx, outSAMtype=BAM SortedByCoordinate, outSAMattributes=All, outFilterMultimapNmax=101, outSJfilterOverhangMin=3 1 1 1, outSJfilterDistToOtherSJmin=0 0 0, alignIntronMin=11, alignEndsType=EndToEnd`).

(see below), or after sequencing using Guppy (ONT). Base calling removes the poly(A) or poly(I) tails at the starts of the reads so that the reads can be aligned to the reference genome without an adapter removal step. We found that the alignment software `minimap2` (ref. ⁴⁰) is particularly well-suited to managing the higher error rate of nanopore sequencing and will sometimes remove portions of the start of reads until it is confident in the alignment. Due to the high error rate of nanopore sequencing, we developed a custom algorithm to determine the splicing status of introns in nano-COP reads¹⁸. This technique is based on existing intron annotations and does not support *de novo* splicing calls. We start by identifying reads that overlap annotated intron coordinates and determine splicing status based on the proportion of the reads that map to the exonic and intronic regions. Splicing status is then used to perform nano-COP analyses such as those in Fig. 2 or to answer new questions specific to the experiment. For instance, splicing status combined with analysis of RNA 3' ends determines the physical proximity between splicing and transcription. For this purpose, we recommend obtaining short-read total RNA-seq data from the cell line of interest to determine which introns are constitutively spliced in mature RNA, in order to avoid confusing delayed splicing kinetics with intron retention (Box 3). nano-COP also reveals higher-order splicing patterns across multiple introns in a transcript, such as splicing order or coordination.

Expertise needed to implement the protocol

This protocol can be implemented in any laboratory that has tissue culture facilities and access to a MinION or PromethION device with an appropriate computer (see 'Equipment'). In addition, a high-performance computing system is recommended for data processing (base calling/alignment). Basic command-line usage is necessary for initial data processing, and basic knowledge of Python is recommended to run available nano-COP scripts for downstream analyses.

Limitations

Nano-COP is a powerful method for analyzing co-transcriptional processing. However, as with all methods, it has limitations. Direct RNA sequencing requires a significant amount of input material (~500 ng). As nascent RNA represents a very small proportion of the total RNA in the cell (<0.5%)^{41,42}, a large number of cells are needed for each nano-COP experiment (~100 million human cells, ~500 million *Drosophila* S2 cells). This can be easily achieved with fast-growing cancer or immortalized cell lines, but may be more challenging for slower-growing cell lines or primary cells. In addition, because nano-COP requires rapid cell lysis following the short 4sU pulse to avoid continuous labeling, suspension cells are well-suited for analysis as they can be rapidly poured into conical tubes for further processing. Although we expect that the Protocol can be adapted for adherent cells, the pulse time may need to be adjusted to accommodate the time required to collect the cells. Long 4sU pulse times (24–48 h) result in cellular toxicities^{43,44}. However, after a 4 h 4sU pulse time, splicing levels are not appreciably changed (Supplementary Fig. 1b,c from ref. ¹⁸, with data from ref. ⁴⁵). Thus, longer pulse times can be considered for adherent or slow growing cells. Nevertheless, the 4sU pulse time is ideally limited to the shortest option possible, such as the 8-min pulse used for nano-COP analysis of K562 cells.

Nanopore sequencing currently yields up to 650 K (MinION) or up to 1.1 M (PromethION) uniquely mapped reads per nano-COP sample, resulting in a median of 1–4 (MinION) or 4–7 (PromethION) mapped reads per gene per sequencing run with samples from human cell lines. Although this allows an analysis of global splicing dynamics by aggregating all reads, it does not offer the read depth necessary to perform analyses at the single-gene or single-transcript level, with the exception of the most highly expressed genes. Nonetheless, finer-scale analyses may become feasible as the throughput of nanopore sequencing increases. We anticipate that single-intron splicing kinetic analyses would become possible with ~20× more coverage when using PromethION sequencing. Alternatively, targeted approaches such as UNCALLED (ref. ⁴⁶) or nanopore adaptive sequencing⁴⁷, which actively enrich for reads corresponding to specific sequences during nanopore sequencing, could be applied to nano-COP once they are adapted to direct RNA sequencing.

nano-COP resolution is limited by the current read lengths obtained by nanopore sequencing. Direct RNA sequencing by the ONT platform yields shorter median read lengths than cDNA sequencing for mRNA²⁴ and nascent RNA samples (Fig. 2d). It has yet to be determined whether incomplete read lengths are due to features of the input RNA or disruptions during sequencing (perhaps through RNA secondary structure). Nevertheless, we expect that read lengths will continue to increase as both the nano-COP protocol and direct RNA sequencing technologies continue to improve. With median nano-COP read lengths currently between 450 and 700 nt, analyses that require reads spanning the entire length of an intron tend to return a distribution of analyzed introns that is shorter than the average intron length in the genome, especially in human cells. Thus, some analyses are biased towards shorter introns and may not always reflect the behaviors of longer introns. In addition, analyses of higher-order splicing patterns have so far been limited to four consecutive introns in the same read, and studies of the distance between transcription and splicing cannot resolve distances longer than the reads themselves. Thus, in human cells, where splicing occurs distally from transcription (Fig. 2b), it is currently not possible to determine exactly when splicing happens for many introns because they are spliced after >2 kb have been transcribed.

Finally, nano-COP does not achieve the single-nucleotide resolution offered by NET-seq to map the Pol II position due to the poor alignment accuracy at the very ends of nanopore reads. Nonetheless, we previously found that the majority of RNA 3' ends correctly align within 25 nucleotides of the expected transcript end position¹⁸, indicating that the ends of nascent transcripts can be mapped accurately within 50 nucleotides. Lower base calling accuracy of nanopore sequencing data limits the proportion of short reads (e.g., <200 nt) that are accurately and uniquely aligned to the reference genome. However, nano-COP analyses generally require longer reads and so are not limited by the accuracy of nanopore sequencing. Nevertheless, as direct RNA-sequencing read lengths and ONT base calling accuracy continue to improve, we expect that the proportion of short reads aligning to the genome will also increase.

Materials

Biological materials

- The most ideal cell types for nano-COP are those that grow in suspension for easy handling and rapid collection. The biological material for our nano-COP study included human K562 cells (American

Type Culture Collection (ATCC), cat. no. CCL-243, RRID: [CVCL_0004](#)), human B lymphoblast BL1184 cells (ATCC, cat. no. CRL-5949, RRID: [CVCL_2635](#)), and *Drosophila* S2 cells (Expression Systems, cat. no. 94-005, RRID: [CVCL_Z232](#)) **! CAUTION** It is important to regularly check cell lines to ensure that they are authentic and are not infected with mycoplasma. Handle cell lines according to the supplier's instructions. Work in a biosafety cabinet, use sterile equipment, and wear gloves to minimize the risk of contamination.

Reagents

- RNase AWAY Surface Decontaminant (Thermo Fisher, cat. no. 7000TS1)
- RPMI 1640 medium (Thermo Fisher, cat. no. 11875119) for human cells or Schneider's *Drosophila* medium (Thermo Fisher, cat. no. 21720024) for *Drosophila* S2 cells
- Fetal bovine serum (FBS) (Thermo Fisher, cat. no. 10437036) for human cells or heat-inactivated FBS (Thermo Fisher, cat. no. 16140063) for *Drosophila* S2 cells
- Penicillin (10,000 units/ml) and streptomycin (10,000 µg/ml, Thermo Fisher, cat. no. 15140122) for human cells or 5,000 units/ml penicillin and 5,000 µg/ml streptomycin (Thermo Fisher, cat. no. 15070063) for *Drosophila* cells
- 4-Thiouridine (4sU, Sigma, cat. no. T4509)
- RNase/DNase-free H₂O (Life Technologies, cat. no. 10977-015)
- PBS (1×, Life Technologies, cat. no. 10010-023)
- α-Amanitin (Sigma-Aldrich, cat. no. A2263) **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations.
- DTT, liquid (0.1 M, part of the SuperScript III first-strand synthesis system; Life Technologies, cat. no. 18080-051) **! CAUTION** DTT is toxic and corrosive, and it is an irritant. Handle solutions containing DTT with care and dispose of waste according to institutional regulations. DTT also has a short half-life, so the number of freeze–thaw cycles should be minimal.
- DTT, powder (Thermo Fisher Scientific, cat. no. R0861) **! CAUTION** DTT is toxic and corrosive, and it is an irritant. Handle solutions containing DTT with care and dispose of waste according to institutional regulations. DTT also has a short half-life, so it is optimal to prepare solutions fresh and keep them on ice before use.
- NP-40, molecular biology grade (Life Technologies, cat. no. 28324) **! CAUTION** NP-40 is an irritant. Handle solutions containing NP-40 with care and dispose of waste according to institutional regulations.
- Triton X-100, molecular biology grade (Sigma-Aldrich, cat. no. T9284) **! CAUTION** Triton X-100 is harmful, and it is an irritant. Triton X-100 is hazardous to the environment. Handle solutions containing Triton X-100 with care and dispose of waste according to institutional regulations.
- Tween 20, molecular biology grade (Sigma-Aldrich, cat. no. 274348)
- Sucrose, molecular biology grade (Sigma-Aldrich, cat. no. S0389)
- Glycerol, molecular biology grade (Sigma-Aldrich, cat. no. G5516)
- Urea, molecular biology grade (Sigma-Aldrich, cat. no. U6504)
- Sodium acetate, RNase-free (3 M; Life Technologies, cat. no. AM9740)
- GlycoBlue (15 mg/ml; Life Technologies, cat. no. AM9515)
- NaCl, RNase-free (5 M; Life Technologies, cat. no. AM9760G)
- NaOH, molecular biology grade (Sigma-Aldrich, cat. no. 72068)
- EDTA, RNase-free (0.5 M; Life Technologies, cat. no. AM9260G) **! CAUTION** EDTA is an irritant. Handle solutions containing EDTA with care and dispose of waste according to institutional regulations.
- Tris-HCl, RNase-free (1 M, pH 7.0; Life Technologies, cat. no. AM9850G)
- Tris-HCl, RNase-free (1 M, pH 8.0; Life Technologies, cat. no. AM9855G)
- HEPES, RNase-free (1 M, pH 7.5; Teknova, cat. no. H1035)
- SUPERase.In (20 U/µl; Life Technologies, AM2696)
- Protease inhibitor mix cComplete, EDTA-free (Roche, cat. no. 11873580001) **! CAUTION** This mix is an irritant. Handle solutions containing the protease inhibitor mix with care and dispose of waste according to institutional regulations.
- QIAzol lysis reagent (QIAGEN, cat. no. 79306, or part of the miRNeasy micro kit, QIAGEN cat. no. 217084) **! CAUTION** The QIAzol lysis reagent is toxic and corrosive. Use personal protective equipment when handling this kit, and be sure to dispose of waste according to institutional regulations.

- Chloroform, molecular biology grade (Sigma-Aldrich, cat. no. 288306) **! CAUTION** Chloroform is volatile and toxic. Chloroform is an irritant. Handle solutions containing chloroform with care and dispose of chloroform waste according to institutional regulations.
- Isopropanol, molecular biology grade (Sigma-Aldrich, cat. no. 278475) **! CAUTION** Isopropanol is highly flammable and volatile. Isopropanol is an irritant. Handle solutions containing isopropanol with care and dispose of waste according to institutional regulations.
- Ethanol, molecular biology grade (VWR, cat. no. V1016) **! CAUTION** Ethanol is highly flammable and volatile. Ethanol is an irritant. Handle solutions containing ethanol with care and dispose of waste according to institutional regulations.
- EZ-Link Biotin-HPDP (Thermo Fisher Scientific, cat. no. 21341) **! CAUTION** EZ-Link Biotin-HPDP is light sensitive, so exposure to light should be kept minimal. **▲ CRITICAL** Optimization may be required if using a different 4sU isolation strategy.
- *N,N*-Dimethylformamide (DMF, Sigma-Aldrich, cat. no. 227056) **! CAUTION** DMF is flammable, toxic, and an irritant. Always handle solutions with DMF with care while wearing personal protective equipment in a fume hood and dispose of waste according to institutional regulations.
- Chloroform–isoamyl alcohol (24:1, BioUltra cat. no. 25666) **! CAUTION** Chloroform is toxic if inhaled and a potential carcinogen. Always handle solutions containing chloroform with care while wearing personal protective equipment in a chemical fume hood and dispose of waste according to institutional regulations.
- μ MACS streptavidin kit (Miltenyi Biotec, cat. no. 130-074-101) **▲ CRITICAL** Optimization may be required if using a different 4sU isolation strategy.
- miRNeasy micro kit (QIAGEN, cat. no. 217084)
- RNase-free DNase set (QIAGEN, cat. no. 79254)
- Qubit RNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32855)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)
- RiboMinus Eukaryotic Kit v2 (Thermo Fisher, cat. no. A15020) or riboPOOL probes targeting *Drosophila* ribosomal RNAs (siTOOLS Biotech) with Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific, cat. no. 65001)
- *Escherichia coli* poly(A) polymerase (Clontech/Takara, cat. no. 2180) **▲ CRITICAL** Optimization may be required if using an *E. coli* poly(A) polymerase from a different supplier for poly(A) tailing (Box 1).
- NEB poly(A) buffer (New England Biolabs, cat. no. M0276S)
- Yeast poly(A) polymerase (Thermo Fisher, cat. no. 74225Z25KU) **▲ CRITICAL** It is crucial to use yeast poly(A) polymerase for poly(I) tailing. Optimization may be required if using a yeast poly(A) polymerase from a different supplier (Box 1).
- Adenosine triphosphate (ATP, 10 mM, New England Biolabs, cat. no. P0756S)
- Inosine triphosphate (ITP, Sigma, cat. no. I0879)
- Direct RNA sequencing kit (Oxford Nanopore Technologies, cat. no. SQK-RNA002) **▲ CRITICAL** Use the latest version of the Direct RNA sequencing kit from ONT. Nano-COP has not been tested with sequencing kits from other suppliers.
- ONT_RTA_C₁₀_oligoA: /5PHOS/GGCTTCTTCTTGCTCTTAGGTAGTAGGTTTC (order from IDT as 100 nmol DNA oligo with RNase Free HPLC Purification)
- ONT_RTA_C₁₀_oligoB: GAGGCGAGCGGTCAATTTTCCCTAAGAGCAAGAAGAAGCCCCCCCC CCCC (order from IDT as 100 nmol DNA oligo with RNase Free HPLC Purification)
- T4 DNA ligase (2,000,000 units/ml, New England Biolabs, cat. no. M0202M)
- SuperScript III (Invitrogen, cat. no. 18080044)
- Agencourt RNAClean XP beads (Beckman Coulter, cat. no. A63987)
- oGAB11 RNA control oligonucleotide: 5'-agucacuuagcgauguacacugacugug-3' (order from IDT as 100 nmol RNA oligo with RNase Free HPLC Purification)
- Low Range ssRNA ladder (New England Biolabs, cat. no. N0364S)
- ss10 DNA ladder (Simplex Sciences)
- TBE-urea gels, 15% (wt/vol) (ThermoFisher, cat. no. EC68852BOX)
- TBE-Urea Sample Buffer (2 \times , ThermoFisher, cat. no. LC6876)
- SYBR Gold nucleic acid gel stain (10,000 \times concentrate; Thermo Fisher, cat. no. S-11494) **! CAUTION** SYBR Gold nucleic acid gel stain is flammable. Nucleic acid stains are usually mutagenic. Use personal protective equipment when handling nucleic acid gel stains, and dispose of waste according to institutional regulations.
- miRNeasy mini kit (QIAGEN, cat. no. 217004)

- SuperScript III First-Strand Synthesis System (Invitrogen, cat. no. 18080051)
- SsoFast EvaGreen Supermix (BioRad, cat. no. 1725200)

Equipment

- Tissue culture flasks, 175 cm² (VWR, cat. no. 353112) for human cells or 75 cm² (Corning, cat. no. 430641U) for *Drosophila* S2 cells
- Bottles with 0.2 µm vacuum filter (500 ml, Corning, cat. no. 430758)
- RNase/DNase-free PCR tubes (0.2 ml, Corning, cat. no. 3745)
- RNase/DNase-free microcentrifuge tubes (1.5 ml, Life Technologies, cat. no. AM12450)
- Needles (22G, BD, cat. no. 305155) **! CAUTION** Handle sharps with care. Dispose of sharps according to institutional regulations.
- Steriflip-GP Sterile Centrifuge Tube Top Filter Unit (Millipore Sigma, cat. no. SCG00525)
- Syringes (1 ml, BD, cat. no. 309628)
- Syringe filters (0.2 µm, VWR, cat. no. 28145-477)
- Phase-lock gel heavy tubes (2 ml, QuantaBio, cat. no. 10847-802)
- Barrier pipette tips (VWR, cat. nos. BT20 89140-932, BT200 89140-936, BT1250 89168-754)
- Countess cell counter (Life Technologies, cat. no. C10227)
- Countess cell counting chamber slides (Life Technologies, cat. no. C10228)
- Refrigerated microcentrifuge 5424R for 1.5 ml microcentrifuge tubes (VWR, cat. no. 97058-914)
- Benchtop centrifuge for 50 ml conical tubes (Eppendorf, cat. no. 5810R)
- Vortexer (VWR, cat. no. 58816-121)
- Magnetic rack (Thermo Scientific, cat. no. 21359)
- µMACS magnetic separator for 4sU pulldown (comes with µMACS Streptavidin Starting Kit, Miltenyi Biotec, cat. no. 130-091-287)
- Thermal cycler (Fisher Scientific, cat. no. E950030020)
- Thermomixer (Core Life Sciences, cat. no. H5000-HC)
- NanoDrop 2000 UV-visible spectrophotometer (Thermo Scientific, cat. no. ND-2000)
- Qubit 2.0 fluorometer (Life Technologies, cat. no. Q32866)
- Mini-Cell polyacrylamide gel box, XCell SureLock (ThermoFisher, cat. no. EI0001)
- Electrophoresis power supply (VWR, cat. no. 93000-744)
- Black gel box (LI-COR, cat. no. 929-97301)
- MinION nanopore device (Oxford Nanopore Technologies) or PromethION nanopore device (Oxford Nanopore Technologies)
- FLO-MIN106 R.9.4.1 flow cell for MinION or FLO-PRO002 flow cell for PromethION (Oxford Nanopore Technologies)
- Computer for connecting and running the MinION device, with the following minimum specifications:
 - Windows 7, 8, or 10/macOS Sierra, High Sierra, or Mojave / Linux Ubuntu 16.04 or 18.04
 - 16 GB RAM
 - i7 or Xeon with 4+ cores
 - 1 TB internal SSD
 - USB3 port
- Hard-Shell 384-Well PCR Plates (BioRad, cat. no. HSP3821)
- Microseal 'B' PCR Plate Sealing Film (BioRad, cat. no. MSB1001)
- BioRad C1000 Touch Thermal Cycler Chassis (BioRad, cat. no. 1841100)
- BioRad CFX384 Optical Reaction Module for Real-Time PCR Systems (BioRad, cat. no. 1845385)

Software and data

- Linux/Unix operating system and high-performance computing platform
- MinKNOW (Oxford Nanopore Technologies, release 18.12.9 or above)
- Minimap2 (ref. ⁴⁰)
- Python v3.7.4 or above
- Jupyter notebook (<https://jupyter.org/>)
- Pysam⁴⁸ (<https://github.com/pysam-developers/pysam>)
- Pybedtools⁴⁹ (<https://daler.github.io/pybedtools/>)
- Reference genome in FASTA format
- Annotation files for reference genome (.gtf and .bed)
- Optional: Illumina short-read total RNA-seq from the cell line of interest to annotate intron stringency (Box 3)

Reagent setup

Human cell culture medium

Mix 500 ml RPMI 1640 medium with 50 ml FBS and 5 ml 10,000 units/ml penicillin and 10,000 µg/ml streptomycin. Can be stored at 4 °C for 4 months. Heat to 37 °C before adding to cell culture.

Drosophila cell culture medium

Mix 450 ml Schneider's *Drosophila* medium with 50 ml heat-inactivated FBS and 5 ml of 5,000 units/ml penicillin and 5,000 µg/ml streptomycin. Filter sterilize with a 0.2 µm vacuum filter. Can be stored at 4 °C for 4 months. Heat to 25 °C before adding to cell culture.

4sU solution

Prepare a fresh 50 mg/ml solution in DNase/RNase-free H₂O immediately before use. 4sU is light sensitive, so exposure to light should be kept minimal.

Protease inhibitor mix (50×)

Dissolve one tablet of protease inhibitor in 1 ml of pre-cooled RNase-free H₂O. Prepare it before use and store aliquots for up to 1 year at -20 °C. **!CAUTION** The protease inhibitor mix is an irritant. Handle solutions containing the protease inhibitor mix with care and dispose of waste according to institutional regulations.

α-Amanitin solution (1 mM)

Dissolve 1 mg of α-amanitin in 1 ml of DNase/RNase-free H₂O. Prepare the solution before use and store aliquots for up to 1 year at -20 °C. **!CAUTION** α-Amanitin is toxic. Handle α-amanitin solution with care and dispose of waste according to institutional regulations.

Cytoplasmic lysis buffer

Contains 0.15% (vol/vol) NP-40, 10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25 µM α-amanitin, 10 units SUPERase.In, and 1× protease inhibitor mix. For 12 cellular fractionation reactions, mix 42 µl of 10% (vol/vol) NP-40, 28 µl of 1 M Tris-HCl (pH 7.0), 84 µl of 5 M NaCl, 56 µl of 1× protease inhibitor mix (50×), 70 µl of 1 mM α-amanitin, 7 µl of SUPERase.In (20 units/µl), and 2,513 µl of RNase-free H₂O. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **!CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care, and dispose of waste according to institutional regulations. NP-40 and the protease inhibitor mix are irritants.

Sucrose buffer

Contains 10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25% (wt/vol) sucrose, 25 µM α-amanitin, 20 units SUPERase.In, and 1× protease inhibitor mix. For 12 cellular fractionation reactions, mix 67.2 µl of 1 M Tris-HCl (pH 7.0), 201.6 µl of 5 M NaCl, 3,360 µl of 50% (wt/vol) filter-sterilized sucrose, 134.4 µl of 1× protease inhibitor mix (50×), 168 µl of 1 mM α-amanitin, 16.8 µl of SUPERase.In (20 units/µl), and 2,772 µl of DNase/RNase-free H₂O. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **!CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

Nuclei wash buffer

Contains 0.1% (vol/vol) Triton X-100, 1 mM EDTA, 25 µM α-amanitin, 40 units SUPERase.In, and 1× protease inhibitor mix in 1× PBS. For 12 cellular fractionation reactions, mix 22.4 µl of 0.5 M EDTA solution, 112 µl of 10% (vol/vol) Triton X-100, 224 µl of 1× protease inhibitor mix (50×), 280 µl of 1 mM α-amanitin, 28 µl of SUPERase.In (20 units/µl), and 10,533.6 µl of 1× PBS. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **!CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. Triton X-100 is harmful and is an irritant. Triton X-100 is hazardous to the environment. Handle solutions containing Triton X-100 with care and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

Glycerol buffer

Contains 20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 0.5 mM EDTA, 50% (vol/vol) glycerol, 0.85 mM DTT, 25 μ M α -amanitin, 10 units SUPERase.In, and 1 \times protease inhibitor mix. For 12 cellular fractionation reactions, mix 56 μ l of 1 M Tris-HCl (pH 8.0), 42 μ l of 5 M NaCl, 14 μ l of 0.1 M EDTA, 1,400 μ l of 100% (vol/vol) filter-sterilized glycerol, 23.8 μ l of 0.1 M filter-sterilized DTT, 56 μ l of 1 \times protease inhibitor mix (50 \times), 70 μ l of 1 mM α -amanitin, 7 μ l of SUPERase.In (20 units/ μ l), and 1,131.2 μ l of RNase-free H₂O. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **!CAUTION** α -Amanitin is toxic. Handle solutions containing α -amanitin with care and dispose of waste according to institutional regulations. DTT is toxic and corrosive. DTT, EDTA and the protease inhibitor mix are irritants.

Nuclei lysis buffer

Contains 1% (vol/vol) NP-40, 20 mM HEPES (pH 7.5), 300 mM NaCl, 1 M urea, 0.2 mM EDTA, 1 mM DTT, 25 μ M α -amanitin, 10 units SUPERase.In, and 1 \times protease inhibitor mix. For 12 cellular fractionation reactions, mix 280 μ l of 10% (vol/vol) NP-40, 56 μ l of 1 M HEPES (pH 7.5), 5.6 μ l of 0.1 M EDTA, 168 μ l of 5 M NaCl, 280 μ l of 10 M filter-sterilized urea, 28 μ l of 0.1 M filter-sterilized DTT, 56 μ l of 1 \times protease inhibitor mix (50 \times), 70 μ l of 1 mM α -amanitin, 7 μ l of SUPERase.In (20 units/ μ l), and 1,849.4 μ l of RNase-free H₂O. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **!CAUTION** α -Amanitin is toxic. Handle solutions containing α -amanitin with care and dispose of waste according to institutional regulations. DTT is toxic and corrosive. DTT, EDTA, NP-40, and the protease inhibitor mix are irritants. **▲CRITICAL** Use fresh made 10 M filter-sterilized urea.

Chromatin resuspension solution

Contains 25 μ M α -amanitin, 20 units SUPERase.In, and 1 \times protease inhibitor mix in 1 \times PBS. Mix 5 μ l of 1 mM α -amanitin, 4 μ l of 1 \times protease inhibitor mix (50 \times), 0.5 μ l of SUPERase.In (20 units/ μ l), and 190.5 μ l of 1 \times PBS. This volume is sufficient for performing three RNA extraction reactions with the chromatin samples from four cellular fractionation reactions in each. Prepare this solution fresh with RNase-free reagents and store on ice before use. **!CAUTION** α -Amanitin is toxic. Handle solutions containing α -amanitin with care, and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

10 \times biotinylation buffer

Contains 100 mM Tris-HCl (pH 7.5) and 10 mM EDTA. Mix 70 μ l of 1M Tris-HCl (pH 7.0), 30 μ l of 1M Tris-HCl (pH 8.0), 20 μ l of 500 mM EDTA, and 880 μ l of DNase/Rnase-free H₂O. Prepare this solution fresh with Rnase-free reagents. **!CAUTION** EDTA is an irritant. Handle solutions containing EDTA with care and dispose of waste according to institutional regulations.

EZ-Link Biotin-HPDP solution

Prepare a 1 mg/ml solution of EZ-Link Biotin-HPDP in DMF. Heat to 37 $^{\circ}$ C with rotation in the dark for 1 h to mix. Store aliquots for up to 1 year at -20° C. EZ-Link Biotin-HPDP is light sensitive, so exposure to light should be kept minimal. **!CAUTION** DMF is flammable, toxic, and an irritant. Always handle solutions with DMF with care while wearing personal protective equipment in a fume hood and dispose of waste according to institutional regulations.

1 \times biotinylation wash buffer

Contains 100 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 M NaCl, and 0.1% (vol/vol) Tween 20. Mix 2.8 ml of 1M Tris-HCl (pH 7.0), 1.2 ml of 1 M Tris-HCl (pH 8.0), 0.8 ml of 500 mM EDTA, 8 ml of 5 M NaCl, 40 μ l of 100% (vol/vol) Tween 20, and 27.2 ml DNase/Rnase-free H₂O. Prepare this solution fresh with Rnase-free reagents. This volume is sufficient for four samples. Divide buffer into two tubes: one containing 25 ml and the other containing 15 ml. Keep the tube with 25 ml buffer at room temperature (20–25 $^{\circ}$ C) and move the tube with 15 ml buffer to 65 $^{\circ}$ C until use **!CAUTION** EDTA is an irritant. Handle solutions containing EDTA with care and dispose of waste according to institutional regulations.

DTT elution buffer

Prepare 100 mM DTT solution in Rnase-free H₂O. Mix by vortexing until fully resuspended. Prepare fresh before use and leave on ice until ready to use. A volume of 250 μ l per sample is sufficient.

! CAUTION DTT is toxic and corrosive, and it is an irritant. Handle solutions containing DTT with care and dispose of waste according to institutional regulations.

In-house 10× poly(A) buffer (for poly(A) tailing only)

Contains 25 mM MnCl₂, 10 mM DTT, and 5 mg/ml BSA. Mix 1 μl of 25 mM MnCl₂, 4 μl of 100 mM DTT, 4 μl of 50 mg/ml BSA, and 31 μl of nuclease-free water. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **! CAUTION** DTT is toxic, corrosive, and an irritant.

5× yeast poly(A) polymerase buffer (for poly(I) tailing only)

Contains 100 mM Tris-HCl pH 7.0, 3 mM MnCl₂, 0.1 mM EDTA, 1 mM DTT, and 0.5 mg/ml BSA. Mix 25 μl of Tris-HCl (pH 7.0), 0.75 μl of MnCl₂, 0.5 μl of 50 mM EDTA, 2.5 μl of 100 mM DTT, 2.5 μl of 50 mg/ml BSA, and 218.75 μl of nuclease-free water. This is the same buffer as the one provided with the enzyme, but we have found that the efficiency of the reaction is increased by preparing fresh buffer each time. Prepare this solution fresh with RNase-free reagents and keep on ice before use. **! CAUTION** DTT is toxic, corrosive, and an irritant.

ITP solution (for poly(I) tailing only)

Prepare 10 mM ITP solution in RNase-free H₂O. Mix by vortexing until fully resuspended. Sterilize through a 0.2 μm filter. Prepare aliquots to freeze and store at −20 °C for up to 1 year to limit the number of freeze-thaw cycles and potential for contamination.

RTA_C10 RTA (for poly(I) tailing only)

Mix ONT_RT_A_C10_oligoA and ONT_RT_A_C10_oligoB at a concentration of 1.4 μM each in 10 mM Tris-HCl (pH 7.5) and 50 mM NaCl. Anneal oligonucleotides by heating to 95 °C for 2 min and then slowly cooling to room temperature in a thermal cycler. Store aliquots for up to 1 year at −20 °C.

Procedure

4sU labeling and cellular fractionation ● Timing 4 h

▲ CRITICAL 4sU labeling and cellular fractionation has been optimized for human K562 and B lymphoblast cells grown in suspension and for semi-adherent *Drosophila* S2 cells. Protocol optimization may be required if other cell types are used.

▲ CRITICAL Cell fractionation is performed on ice or at 4 °C, with buffers freshly prepared on the same day. All buffers are pre-cooled on ice before use. Use RNase-free reagents and equipment.

▲ CRITICAL To obtain enough RNA for 4sU selection and nanopore sequencing, at least 120 million human cells or 600 million *Drosophila* cells need to be harvested for cellular fractionation.

Since cell fractionation volumes are optimized for 10 million human K562 and BL1184 cells and 50 million *Drosophila* cells, it is critical to perform 10–12 reactions per sample for nano-COP. This number may need to be adjusted for other cell lines. Performing cellular fractionation with too many cells per reaction leads to inefficient lysis and increased levels of cytoplasmic material in the chromatin fraction. In the steps below, ‘reaction’ refers to each cell fractionation of 10 million human cells or 50 million *Drosophila* cells, and ‘sample’ refers to all the combined reactions for one condition.

▲ CRITICAL We found that a 4sU labeling time of 8 min provides optimal recovery of 4sU-labeled chromatin-associated RNA with a manageable number of cells for labeling and fractionation experiments. Reducing the labeling time will recover a larger proportion of RNAs in the process of transcription, but may require additional cells and cellular fractionation reactions to obtain enough material for direct RNA nanopore sequencing. Increasing the labeling time will recover more material, but may lead to a larger proportion of RNAs that have already completed transcription. Depending on your experimental goals, it may be worth adjusting the 4sU labeling time to capture an appropriate proportion of RNAs in the process of transcription.

- 1 *Prepare reagents for fractionation.* Prepare all cellular fractionation buffers (cytoplasmic lysis buffer, sucrose buffer, nuclei wash buffer, glycerol buffer, nuclei lysis buffer, and chromatin resuspension solution) as detailed in the ‘Reagent setup’ section before proceeding with Step 5. Leave on ice before using.
- 2 If working with human cells, pre-warm 1× PBS and three 50 ml conical tubes (per sample) to 37 °C. If working with *Drosophila* cells, pre-warm 1× PBS to room temperature.

- 3 Cut 1,000 μl filtered pipette tips 2–3 mm from the bottom for resuspending some fragile and viscous solutions. At least 6 pipette tips per sample are sufficient.
- 4 Set up a microcentrifuge for 1.5 ml Eppendorf tubes to 4 °C.
- 5 Label cells with 500 μM 4sU and harvest.

▲ CRITICAL step It is important to have consistent incubation times across flasks. To prevent uneven labeling times, it may be helpful for a second experimentalist to support the following labeling and harvesting steps (Step 5A or B(i–vii)).

 - (A) **Human K562 or B lymphoblast cells**
 - (i) Grow cells in three 175 cm^2 flasks in 50 ml human cell culture medium in a 37 °C incubator with 5% CO_2 until they reach a cell density of 0.8×10^5 cells/ml, for a total of 120×10^6 cells.
 - (ii) Add 500 μl of 50 mM 4sU to each flask and mix by tilting.
 - (iii) Move flasks to a 37 °C incubator for 7 min 30 s.
 - (iv) Take out flasks and immediately pour the contents of each flask (containing the labeled cells) into three pre-warmed 50 ml conical tubes.
 - (v) Centrifuge at 500g for 2 min at 25 °C.
 - (vi) Aspirate the supernatant and resuspend cells in 10 ml of 1× PBS pre-warmed at 37 °C.
 - (vii) Centrifuge at 500g for 2 min at 25 °C.
 - (B) **Drosophila S2 cells**
 - (i) Grow S2 cells in six 75 cm^2 flasks in 10 ml of *Drosophila* cell culture medium at 25 °C until they are 90% confluent ($\sim 100 \times 10^6$ cells/flask, for a total of 600×10^6 cells).
 - (ii) Add 100 μl of 50 mM 4sU to each flask.
 - (iii) Incubate at room temperature for 7 min 30 s.
 - (iv) Lift the cells by gentle pipetting and transfer the content of two flasks to each of the three pre-warmed 50 ml conical tubes.
 - (v) Centrifuge at 500g for 2 min at 25 °C.
 - (vi) Aspirate the media and resuspend cells in 10ml of 1× PBS.
 - (vii) Centrifuge at 500g for 2 min at 25 °C.
- 6 Remove the supernatant by aspiration.

▲ CRITICAL STEP It is important to remove the supernatant completely at this step. If the supernatant is not completely removed the cytoplasmic lysis buffer (Step 7) will be diluted, which affects the cell lysis efficiency.
- 7 Add 800 μl of cytoplasmic lysis buffer to each 50 ml conical tube. Use a cut 1,000 μl pipette tip to resuspend the cells by pipetting the sample up and down ten times.

▲ CRITICAL STEP It is important to fully resuspend the cells at this step for high lysis efficiency.
- 8 Incubate the cell lysate on ice for 5 min.
- 9 Pipette 500 μl of sucrose buffer into 12 RNase-free microcentrifuge tubes, for each of the cellular fractionation reactions.
- 10 By using a cut 1,000 μl pipette tip, layer 200 μl of cell lysate onto the 500 μl of sucrose buffer in each microcentrifuge tube.
- 11 Collect cell nuclei by centrifugation at 16,000g for 10 min at 4 °C.
- 12 Remove the supernatant.

▲ CRITICAL STEP The supernatant represents the cytoplasmic fraction. If you want to check RNA (Box 2 and Fig. 5) or protein levels in the cytoplasm, keep the supernatant from one reaction and move it to a separate tube rather than discarding it. To use it for RNA analysis, mix the supernatant by pipetting up and down. Combine 200 μL of supernatant with 700 μL of QIAzol lysis reagent in a new tube. Mix briefly by vortexing, incubate at room temperature for 5 min and store at -80 °C for up to 6 months. If you wish to compare absolute RNA quantities across fractions (rather than spliced:unspliced ratios), repeat these collection instructions for the entire sample. To use for protein analysis, follow steps 1 and 2 in Box 1 from ref. ³⁸.

? TROUBLESHOOTING
- 13 Wash nuclei with 800 μl of nuclei wash buffer. Resuspend by gentle pipetting up and down approximately five or six times using a cut 1,000 μl pipette tip.

▲ CRITICAL STEP The pellet may not resuspend completely. It is okay to proceed to the next step without completely resuspending the sample.
- 14 Collect washed nuclei by centrifugation at 1,150g for 1 min at 4 °C and remove the supernatant.

▲ CRITICAL STEP To remove cytoplasmic mature RNAs, it is important to remove the supernatant completely.

- 15 Add 200 μ l of glycerol buffer to each reaction. Resuspend the washed nuclei by gentle pipetting up and down five or six times with a cut 1,000 μ l pipette tip. Transfer the content of each tube to a new 1.5 ml RNase-free microcentrifuge tube.
 - ▲ **CRITICAL STEP** For certain cell types, including K562 and B lymphoblast cells, the pellet may not resuspend completely. It is okay to proceed to the next step without completely resuspending the pellet.
- 16 Add 200 μ l of nuclei lysis buffer to each reaction and mix by pulsed vortexing (with vortexer at medium setting, mix contents of microcentrifuge tube five times for 1 s each). Incubate the mixture on ice for 2 min.
- 17 Centrifuge the mixture at 18,500g for 2 min at 4 °C.
- 18 Remove the supernatant completely.
 - ▲ **CRITICAL STEP** It is important to completely remove the supernatant containing nucleoplasmic RNAs. If you want to check RNA (Box 2) or protein levels in the nucleoplasm, keep the supernatant and move it to a separate tube rather than discarding it. To use for RNA analysis, mix sample by pipetting and combine 200 μ l of supernatant with 700 μ l of QIAzol lysis reagent. Mix briefly by vortexing, incubate at room temperature for 5 min and store at –80 °C for up to 6 months. If you wish to compare absolute RNA quantities across fractions (rather than spliced:unspliced ratios), repeat these collection instructions for the entire sample. To use for protein analysis, follow Box 1 steps 1 and 2 from ref. ³⁸.
 - ? **TROUBLESHOOTING**
- 19 Use 50 μ l of chromatin resuspension solution to resuspend the chromatin from one reaction and transfer it to a new 1.5 ml RNase-free microcentrifuge tube. Using the same 50 μ l of resuspension buffer, repeat with three other reactions such that four chromatin pellets are combined into one tube. Repeat with the remaining chromatin pellets to have a total of three tubes with four chromatin samples in each.
 - ▲ **CRITICAL STEP** In order for the purification of chromatin-associated RNA (Steps 23–35) to have the expected yield, it is important to combine a minimum of three to four chromatin samples.
 - ▲ **CRITICAL STEP** If you want to check RNA (Box 2) or protein levels in the chromatin sample, keep one chromatin pellet and move it to a separate 1.5 ml RNase-free microcentrifuge tube and resuspend in 50 μ l of chromatin resuspension solution. To use for RNA analysis, combine with 700 μ l of QIAzol lysis reagent and homogenize completely using a 1 ml syringe and a 22G needle. Incubate at room temperature for 5 min and store at –80 °C for up to 6 months. To use for protein analysis, follow Box 1 steps 1 and 2 from ref. ³⁸.
 - ? **TROUBLESHOOTING**
- 20 Add 1 ml of QIAzol lysis reagent to each of the three microcentrifuge tubes with chromatin samples.
- 21 Mix thoroughly by vortexing. If the chromatin pellet does not easily resuspend, as is frequently the case with K562 and B lymphoblast cells, use a 1 ml syringe and a 22G needle to homogenize the sample by moving it through the needle approximately ten times.
 - ▲ **CRITICAL STEP** It is important to fully homogenize the sample during this step to enable complete RNA extraction. When using a syringe and needle to mix the sample, work slowly and carefully to avoid spilling.
- 22 Incubate the sample at room temperature for 5 min.
 - **PAUSE POINT** Store sample at –80 °C for up to 6 months until purification of chromatin-associated RNA.

Purification of chromatin-associated RNA ● Timing 2 h

- 23 Prepare before starting:
 - Set the refrigerated microcentrifuge to 4 °C.
 - Prepare at least 10 ml of 75% (vol/vol) ethanol.
- 24 If samples were stored at –80 °C, thaw all three microcentrifuge tubes with the 4sU-labeled chromatin samples at 65 °C for 2 min. Briefly vortex and leave at room temperature for 2 min.
- 25 Add 200 μ l of RNase-free chloroform to each tube.
- 26 Vortex for 15 s and incubate at room temperature for 5 min.
- 27 Centrifuge at 12,000g for 15 min at 4 °C.
- 28 For each reaction, transfer the upper aqueous phase into a new 1.5 ml RNase-free Eppendorf tube.
- 29 Add 500 μ l RNase-free 100% (vol/vol) isopropanol to each aqueous sample. Mix by vortexing and incubate for 10 min at room temperature.

- 30 Centrifuge the samples at 12,000g for 10 min at 4 °C. Discard the supernatant.
▲ CRITICAL STEP At this point, the RNA pellet should be visible.
- 31 Wash once with 1 ml of 75% (vol/vol) ethanol and centrifuge at 7,500g for 10 min at 4 °C. Discard the supernatant.
- 32 Wash a second time with 1 ml of 75% (vol/vol) ethanol and centrifuge at 7,500g for 10 min at 4 °C. Discard the supernatant.
- 33 To completely remove the ethanol, spin the tube again at 1,000g for 1 min at 25 °C, remove the rest of the liquid with a 10–20 µl pipette tip, and air-dry pellet for 1–2 min.
- 34 Dissolve pellet in 17.5 µl nuclease-free water. Combine the three samples into a new tube so that it has ~50 µl total volume.
- 35 Measure RNA concentration using a Nanodrop spectrophotometer. Expected results: ~100–200 µg/sample for human K562, ~80–100 µg/sample for *Drosophila* S2 cells, ~65–75 µg/sample for human BL1184 lymphoblasts.
■ PAUSE POINT Store sample at –80 °C for up to 3 months until the next step.
? TROUBLESHOOTING

Biotinylation and selection of 4sU-labeled RNA ● Timing 6 h

- 36 Dilute up to 60 µg of chromatin-associated RNA in 7× nuclease-free water (i.e., 420 µl water for 60 µg RNA).
▲ CRITICAL STEP The total volume of this reaction is proportional (10×) to the starting amount of RNA. If using more than 60 µg, split into multiple reactions. If using less than 60 µg, reduce the volume of water accordingly. Generally 60 µg RNA will be more than enough to successfully capture enough material for direct RNA nanopore sequencing. However, recovery will vary with 4sU incubation times and cell lines.
- 37 Heat sample to 60 °C in a thermoblock with 800 r.p.m. rotation for 10 min.
- 38 Immediately place on ice for >2 min.
- 39 Add 1 µl 10× biotinylation buffer per 1 µg RNA to the sample (i.e., 60 µl for 60 µg RNA) and mix by pipetting.
- 40 Add 2 µl EZ-Link Biotin-HPDP solution per 1 µg RNA to the sample (i.e., 120 µl for 60 µg RNA) and mix by pipetting.
- 41 Incubate the sample in a thermoblock with 800 r.p.m. rotation at 24 °C for 90 min.
▲ CRITICAL STEP The biotin reaction is light sensitive, so it is best practice to cover samples with aluminum foil or perform this step in a dark room with the lights off.
- 42 Before the biotinylation reaction is complete, prepare 2 phase-lock tubes per sample by centrifuging at 14,000g for 1 min at room temperature.
- 43 After the biotinylation reaction is complete, transfer half (~300 µl) of the RNA–biotin sample to each prepared phase-lock tube.
- 44 Add 1 volume (~300 µl) of chloroform-isoamyl alcohol (24:1) to each tube and mix by manually shaking for >15 s.
▲ CRITICAL STEP Do not vortex!
- 45 Centrifuge at 16,000g for 5 min at room temperature.
- 46 Transfer each upper aqueous phase to a new 1.5 ml microcentrifuge tube.
- 47 Add 1/10 volume (~30 µl) RNase-free 5M NaCl to each tube and mix.
- 48 Add 1 volume (~300 µl) RNase-free isopropanol to each tube and mix.
- 49 Centrifuge at 20,000g for 30 min at 4 °C.
▲ CRITICAL STEP RNA pellets should be visible after this spin.
- 50 Remove the supernatant. Add 1 ml RNase-free 75% (vol/vol) ethanol and mix. Centrifuge at 20,000g for 10 min at 4 °C.
▲ CRITICAL STEP It is important to prepare the 75% (vol/vol) ethanol fresh before use to avoid evaporation.
- 51 Remove supernatant and air-dry pellet for ~1–2 min.
▲ CRITICAL STEP It is important to remove all supernatant without completely drying out the pellet. We have found that an additional quick spin (1,000g for 1 min) may be necessary to completely remove supernatant from the tube before leaving to air-dry.
- 52 Resuspend the RNA pellet in 200 µl nuclease-free water. If the sample was split into two reactions at Step 43, resuspend each RNA pellet in 100 µl of nuclease-free water and combine for a total volume of 200 µl.

- 53 Let the RNA dissolve at room temperature for 5 min.
- 54 Heat to 65 °C in a thermoblock for 10 min. Leave on ice for >5min.
- 55 Check concentration of RNA using a Nanodrop spectrophotometer.
▲ CRITICAL STEP The yield should be close to the starting amount of RNA (60 µg).
- 56 Combine 200 µl of the RNA–biotin sample with 100 µl µMACS streptavidin beads and mix by gentle pipetting.
- 57 Incubate samples in a thermomixer at 800 r.p.m. for 15 min at 24 °C.
- 58 Set up the µMACS magnetic separator according to the manufacturer’s instructions and prepare two RNase-free microcentrifuge tubes per sample (one for the flow-through and one for the eluate).
- 59 Five minutes before the end of the biotin–streptavidin incubation, place the µMACS column into the magnetic rack of the µMACS separator and equilibrate by passing 900 µl of the room temperature 1× biotinylation wash buffer through the column.
▲ CRITICAL STEP Initiate flow-through by pressing gently with a gloved finger on top of the column until a drop passes through the column.
- 60 Transfer the RNA–streptavidin beads mix to the µMACS column and collect the flow-through in the prepared microcentrifuge tube.
- 61 Reload the flow-through and collect again in the prepared microcentrifuge tube.
▲ CRITICAL STEP The flow-through represents RNA that was not labeled/biotinylated and can be discarded. If you wish to analyze the flow-through for troubleshooting purposes as described in Box 2, keep 200 µl and mix it with 700 µl QIAzol lysis reagent by vortexing. Mix briefly by vortexing, incubate at room temperature for 5 min and store at –80 °C for up to 6 months.
- 62 Wash the column three times with 900 µl of 65 °C 1× biotinylation wash buffer.
- 63 Wash the column three times with 900 µl room temperature 1× biotinylation wash buffer.
- 64 Elute RNA with 100 µl of 100 mM DTT.
- 65 Wait 5 min. Elute RNA again with an additional 100 µl of 100 mM DTT.
- 66 Mix 200 µl eluted RNA sample with 300 µl of 100% (vol/vol) ethanol.
- 67 Add sample mixture to an miRNeasy micro kit column and centrifuge at 10,000g for 15 s at room temperature.
- 68 Reload flow-through and centrifuge 10,000g for 15 s at room temperature. Discard flow-through.
- 69 Add 350 µl buffer RWT to the miRNeasy micro kit column and centrifuge at 10,000g for 15 s at room temperature. Discard flow-through.
▲ CRITICAL STEP RWT buffer should be initially resuspended in 45 ml isopropanol.
- 70 From the QIAGEN RNase-free Dnase set, mix 10 µl Dnase I and 70 µl buffer RDD by gentle pipetting. Add 80 µl Dnase solution to the column and incubate for 15 min at room temperature.
- 71 Add 500 µl buffer RWT and centrifuge at 10,000g for 60 s at room temperature.
▲ CRITICAL STEP RWT buffer should be initially resuspended in 45 ml of isopropanol.
- 72 Reload flow-through and centrifuge at 10,000g for 60 s at room temperature. Discard flow-through.
- 73 Add 500 µl RPE buffer to the column and centrifuge at 10,000g for 15 s at room temperature. Discard flow-through.
- 74 Add 500 µl Rnase-free 80% (vol/vol) ethanol and centrifuge at 10,000g for 1 min at room temperature. Discard flow-through.
- 75 Place column in a new 2 ml Rnase-free collection tube. Dry membrane by centrifuging at full speed for 5 min at room temperature.
- 76 Place column in a new 1.5 ml Rnase-free microcentrifuge tube.
- 77 Add 15 µl of 10 mM Tris pH 7.0.
- 78 Incubate at room temperature for 1 min.
- 79 Centrifuge at 10,000g for 1 min at room temperature to elute RNA.
- 80 Reload flow-through for a second elution. Repeat Steps 78 and 79.
- 81 Measure RNA concentration using the Qubit RNA HS assay kit. Expected results: if starting with 60 µg of chromatin-associated RNA, this step should yield 1–6 µg of 4sU–chr RNA for human K562 cells, 1–1.5 µg of 4sU–chr RNA for human BL1184 lymphoblast cells, or 3–10 µg of 4sU–chr RNA for *Drosophila* S2 cells.
■ PAUSE POINT Store sample for up to 3 months at –80 °C until the next step.
? TROUBLESHOOTING

rRNA depletion ● **Timing 3 h**

▲ **CRITICAL** Without this depletion step, rRNAs will dominate the sample and represent the vast majority of nano-COP reads after sequencing. Most in-house or commercial kits that remove rRNAs using biotinylated oligonucleotides will work for nano-COP libraries. We have had success with the Illumina Ribo-Zero Gold (standalone kit has been discontinued and is now only sold as part of TruSeq RNA library preparation kit), Thermo Fisher RiboMinus Eukaryotic Kit v2 kits for human and *Drosophila* cells (note that this kit is not designed for *Drosophila* and will only partially deplete rRNAs in *Drosophila* samples), and RiboPOOLS for *Drosophila* cells.

▲ **CRITICAL** To clean up RNA samples throughout the nano-COP protocol, we have had most success with isopropanol precipitation and believe that it is not subject to the same size selection biases as other purification strategies. Other methods that can be used to clean up RNA include bead-based purification (e.g., with Agencourt RNAClean XP beads) and column-based purification.

82 Perform rRNA depletion of up to 5 µg of 4sU-chr RNA following the manufacturer's instructions.

83 If necessary, add nuclease-free water to a volume of at least 100 µl.

84 Add 0.1× volume RNase-free 3 M sodium acetate to sample and mix.

85 Add 2 µl GlycoBlue to sample and mix by vortexing.

86 Add 1.25× volume 100% (vol/vol) isopropanol to the sample and mix by vortexing.

87 Incubate at -80 °C for >30 min.

88 Centrifuge at 20,000g for 30 min at 4 °C. Remove supernatant and discard.

89 Add 1 ml of 75% (vol/vol) ethanol and centrifuge at 20,000g for 5 min at 4 °C. Remove supernatant and discard.

90 Repeat Step 89.

91 Remove all supernatant and air-dry for ~1–2 min.

▲ **CRITICAL STEP** It is important to remove all supernatant without completely drying out the pellet. We have found that an additional quick spin (1,000g for 1 min) may be necessary to completely remove supernatant from the tube before leaving to air-dry.

92 Resuspend in 11 µl nuclease-free water.

93 Measure RNA concentration using the Qubit RNA HS assay kit. Expected results: ~0.5–1.5 µg/sample after starting with 5 µg 4sU-chr RNA. Concentrations may vary by species and kit used.

■ **PAUSE POINT** Store sample for up to 3 months at -80 °C until the next step.

3' end tailing ● **Timing ~3 h**

94 In a 0.2 ml RNase-free PCR tube, dilute 500 ng of 4sU-chr RNA in 10 µl nuclease-free water.

95 To denature RNA, heat sample to 80 °C for 2 min in a thermal cycler.

96 Prepare reaction mix as described in the tables below for either poly(A) or poly(I) tailing.

▲ **CRITICAL STEP** Before adding the poly(A) polymerase, it is important to mix the reaction until it is homogenous. Next, add the tailing enzyme and mix again. Incompletely mixed samples will negatively affect the tailing reaction.

(A) poly(A) tailing

(i) Combine the following:

Component	Amount (µl)	Final concentration
NEB poly(A) buffer	2	1×
In-house 10× poly(A) buffer	2	1×
ATP 10 mM	1	0.5 mM
RNA (500 ng)	10	—
SUPERase.In (20 units/µl)	1	1 unit/µl
Clontech poly(A) polymerase	1	—
Nuclease-free water	3	—

▲ **CRITICAL STEP** For poly(A) tailing, prepare the in-house 10× poly(A) buffer fresh each time with RNase-free reagents and keep on ice before use (see 'Reagent setup').

(B) **poly(I) tailing**

(i) Combine the following:

Component	Amount (μl)	Final concentration
5× yeast poly(A) polymerase buffer	4	1×
Tris-HCl 10 mM pH 7	2	0.5 mM
ITP 10 mM	1	0.5 mM
RNA (500 ng)	10	—
SUPERase.In (20 units/μl)	1	1 unit/μl
Yeast poly(A) polymerase	2	—

▲ CRITICAL STEP For poly(I) tailing, prepare the 5× yeast poly(A) polymerase buffer fresh each time with RNase-free reagents and keep on ice before use (see ‘Reagent setup’).

97 Incubate the tailing reaction at 37 °C in a thermal cycler for the following recommended times:

- poly(A) tailing: 7.5 min.
- poly(I) tailing: 30 min.

▲ CRITICAL STEP Incubation times for 3'-end tailing reactions may need to be optimized. See Box 1 for instructions to test incubation times and reagents.

98 To stop the tailing reaction, immediately add 0.5 μl of 500 mM EDTA and leave on ice.

99 Clean up by isopropanol precipitation (see Steps 83–91) and resuspend in 10 μl of 10 mM Tris-HCl (pH 7.0).

■ PAUSE POINT Store sample for up to 3 months at –80 °C until the next step.

Oxford Nanopore direct RNA library preparation and sequencing ● Timing 3 h

100 Prepare direct RNA sequencing library using the direct RNA sequencing kit from ONT following the manufacturer’s instructions, including the optional RT step.

▲ CRITICAL STEP ONT library kits and reagents are constantly adapting and improving. It is best practice to use the most up-to-date kits and protocols for sequencing nano-COP libraries. Adjustments we made to the direct RNA sequencing protocol from ONT are described in Box 4.

101 After completing the ONT library preparation, quantify 1 μl of reverse-transcribed and adapted RNA using the Qubit fluorometer DNA HS assay kit. Expected result: 0.5–3.5 ng/μl of RNA/cDNA hybrids. This is lower than the recovery amount that is recommended in the ONT protocol. However, we have found that it is sufficient material for sequencing.

? TROUBLESHOOTING

102 Prime and load a MinION or PromethION flow cell according to the manufacturer’s instructions.

103 Sequence on the MinION with local base calling using default parameters for at least 48 h.

▲ CRITICAL STEP Ensure that the MinION computer meets the hardware and software requirements for nanopore sequencing and that there is sufficient space for data storage.

? TROUBLESHOOTING

104 After sequencing is complete, flush the flow cell according to the manufacturer’s instructions and prepare for return to ONT.

Software installation, download, and preparation of annotation files

105 Install minimap2 by typing the following command on a terminal:

```
> curl -L https://github.com/lh3/minimap2/releases/download/v2.17/minimap2-2.17_x64-linux.tar.bz2 | tar -jxvf -./minimap2-2.17_x64-linux/minimap2
```

106 Download and install samtools by following the developers’ instructions: <http://www.htslib.org/download/>.

107 Go to Ensembl (<https://www.ensembl.org/>) and download the latest primary genome assembly from the species of interest in FASTA format.

108 *Prepare annotation files required for downstream analyses (Steps 108–114).* Ready-to-use annotation files for human (hg38) and *Drosophila* (dm6) are available at <https://github.com/churchmanlab/nano-COP/>.

Box 4 | Modifications to direct RNA sequencing protocol**Procedure**

We made the following adjustments to the direct RNA sequencing protocol for nano-COP libraries prepared with SQK-RNA002:

- Omit the RNA CS control from the initial ligation reaction.
 - ▲ CRITICAL STEP** This spike-in RNA is useful for troubleshooting library preparation and ensuring that nanopore sequencing is successful. Reads mapping to the spike-in should be mostly full-length and sequencing accuracy should be >85%. However, we have found that it takes up a large portion of the sequenced library and so we avoid using it when possible. When the RNA CS control is left out of the reaction, add an extra 0.5 µl of RNA sample or nuclease-free water.
- Use the appropriate adapter for the initial ligation reaction:
 - If poly(A) tailing was performed in Steps 96 and 97, use the RTA provided in the kit and follow the manufacturer's instructions.
 - If poly(I) tailing was performed in Steps 96 and 97, use the RTA-C₁₀ (not provided in the kit; see instructions for preparation in 'Reagent setup'), instead of the RTA provided in the kit for the initial ligation reaction. Use the same volume recommended for the standard RTA in the protocol (1 µl of 1.4 µM RTA-C₁₀ for SQK-RNA002).
- Incubate the initial ligation reaction with the RTA (RTA or RTA-C₁₀) for 15 min, rather than 10 min.

- Go to the UCSC Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>), select the genome and assembly of interest. From the Group menu, select 'Genes and Gene Predictions' and select the track (i.e., NCBI RefSeq) and table (i.e., RefSeq All) of interest. Select BED as the output format and click 'get output'.
- Under 'create one BED record per:', select 'Whole Gene' and click the 'get BED' button.
- Repeat Steps 109 and 110, but select 'Introns plus 0 bases at each end' and 'Exons plus 0 bases at each end' at Step 110.
- Parse the BED files to remove non-standard chromosomes and keep only protein-coding genes. For each BED file (genes/introns/exons), run the following command on the terminal:

```
> grep -v random <genes/introns/exons>.bed | grep -v Un | grep -v alt |
grep NM_ | sort -k1,1 -k2,2n > <genes/introns/exons>_parsed.bed
```

- Use the script `combine_annotation_files_nanoCOP.ipynb` (available at <https://github.com/churchmanlab/nano-COP>) to remove duplicate features and combine genes, introns and exons into one BED file: `<genome>_merge_parsed.bed`
- Sort the resulting BED file by gene name and coordinates by running the following command in the terminal:

```
> sort -k4,4 -k1,1n -k2,2n -k3,3n <genome>_merge_parsed.bed > <genome>_merge_parsed_sortByNameCoord.bed
```

Bioinformatic analysis

▲ CRITICAL The following steps are based on analyses developed using nanopore reads that were base called live with MinKNOW (18.12.9). Modifications may be required for different base calling methods or more recent versions of MinKNOW.

Alignment to reference genome (Bash)

- Concatenate all the fastq files in the `fastq_pass` directory:

```
> for i in path/to/fastq_pass/*.fastq; do cat $i » /path/to/fastq_pass/sample_name_pass.fastq; done
```

- Convert RNA sequences to DNA sequences:

```
> awk '{if(NR%4==2){gsub(/U/, "T", $0); print} else {print $0}}' /path/to/fastq_pass/sample_name_pass.fastq > /path/to/sample_name_pass_RNAtoDNA.fastq
```

- 117 Align to reference genome using minimap2. This is generally performed on a high-performance computing platform, but alignment with minimap2 can also be done within the EPI2ME platform (<https://community.nanoporetech.com/protocols/epi2me/>).

```
> /path/to/minimap2 -ax splice -uf -k14 /path/to/reference_genome.
fasta /path/to/sample_name_pass_RNAtoDNA.fastq > path/to/sample_
name_minimap2.sam
```

- 118 Extract unique reads from the alignment:

```
> grep ^@ /path/to/sample_name_minimap2.sam > /path/to/headers.sam
> grep -v ^@ /path/to/sample_name_minimap2.sam | sort > path/to/
alignment.sam
> awk '$3!="*" {print}' /path/to/alignment.sam | cut -f1 | sort | uniq -c
| awk '$1==1 {print $2}' > /path/to/uniq_names.txt
> grep -F -f /path/to/uniq_names.txt path/to/alignment.sam > path/to/
uniq_alignment.sam
> cat path/to/headers.sam /path/to/uniq_alignment.sam > path/to/head-
ers_uniq_temp.sam
> samtools view -bT /path/to/reference_genome.fasta path/to/header-
s_uniq_temp.sam > /path/to/headers_uniq_temp.bam
> samtools sort /path/to/headers_uniq_temp.bam -o /path/to/sample_
name_minimap2_uniq_sort.bam
> samtools index /path/to/sample_name_minimap2_uniq_sort.bam
> rm headers.sam
> rm alignment.sam
> rm uniq_names.txt
> rm uniq_alignment.sam
> rm headers_uniq_temp.sam
> rm headers_uniq_temp.bam
```

Quality control/sequencing statistics

▲ CRITICAL All the following analyses are performed using the script `main_nanoCOP_analyses.ipynb` available on the nano-COP github page: <https://github.com/churchmanlab/nano-COP>. For other analyses described in ref. ¹⁸, please see the scripts organized by figure in the same repository.

- 119 Obtain sequencing statistics (read length, aligned length, match percentage, etc.) using the `sample_name_pass.fastq` and `sample_name_minimap2_uniq_sort.bam` files as input and the commands within the section ‘Sequencing statistics’ of the script `main_nanoCOP_analyses.ipynb`.

? TROUBLESHOOTING

- 120 Map RNA 3' ends to genomic features and plot the results as a pie chart using `sample_name_minimap2_uniq_sort.bam` as input file and the annotation file `<genome>_merge_parsed_sortByNameCoord.bed` produced in Step 114. Use the commands within the section ‘Mapping RNA 3' ends to gene features’ of the script `main_nanoCOP_analyses.ipynb`.

▲ CRITICAL STEP The 3' end of the RNA is the first portion of the RNA that is sequenced through the nanopore. However, the sequence is flipped during base calling such that the 3' end of the read is represented as the 3' end of the RNA.

? TROUBLESHOOTING

Nano-COP analyses

- 121 *Determine distance between transcription and splicing (Steps 121–128).* To determine the physical distance between transcription and splicing, calculate the distance between the 3' end of each read (Pol II position) and the 3'SS of each intron using `sample_name_minimap2_uniq_sort.bam` as the input file and `<genome>_merge_parsed_sortByNameCoord.bed` produced in Step 114 as the annotation file. Use the commands within the section ‘Physical distance between

transcription and splicing' of the script `main_nanoCOP_analyses.ipynb`. Briefly, the required steps are as follows.

? TROUBLESHOOTING

- 122 Using BEDTools `intersect`, remove reads with transcript 3' ends near annotated poly(A) sites (150 nt upstream or any distance downstream) or 5' SSs (50 nt upstream or 10 nt downstream). This step removes reads that arise from RNAs that are not in the process of transcription.
- 123 Remove reads with more than 150 nt soft-clipped from the RNA 3' end during alignment, to avoid reads with inconclusive 3' end alignments.
- 124 Using BEDTools `intersect`, identify reads that overlap splice sites from constitutively spliced introns (see Box 3) on the same strand (option `s = True`).
- 125 Calculate the distance between the transcript 3' end and the 3'SS of each intron within the read.
- 126 To avoid biases from read length constraints, only keep reads where the read length is greater than the genetic distance between the read end and the intron 3' SS by at least 150 nt.
- 127 *Determine the splicing status of each intron.* For each read/intron pair, extract features of read CIGAR strings for the portion of the alignment mapping to the intron and 50 nucleotides surrounding the 5' and 3' SSs. Introns are determined to be 'spliced' if there is a splicing event that starts and ends within 50 nt of the 5' and 3' SSs and the size of this splicing event is within 10% of the annotated intron size. Introns are determined to be 'unspliced' if the read maps to the coordinates of an intron with no evidence of a splicing event and greater than 50% coverage in the 50 nt surrounding the splice site and at least 75% coverage within the region of intron it maps to. The thresholds selected take into account the higher error rate of nanopore sequencing, while remaining stringent enough for highly accurate splicing calls. Aligned reads that map to introns but do not meet these criteria are removed from subsequent analyses.
- 128 Plot the percentage of introns spliced as a function of the distance transcribed past the 3'SS (Fig. 2b).

? TROUBLESHOOTING

- 129 *Determine the order of splicing (Steps 129–133).* To determine the order of splicing between pairs of consecutive introns and plot the frequency of each intron in the pair being spliced first, follow these steps using `sample_name_minimap2_uniq_sort.bam` as the input file and `introns_parsed.bed` produced in Step 112 as the annotation file. Use the commands within the section 'Order of splicing in intron pairs' of the script `main_nanoCOP_analyses.ipynb`.
- 130 Using BEDTools `intersect`, extract reads that overlap two or more annotated introns on the same strand (option `s = True`). Ensure that the read length is greater than the genetic distance between the 3' end of the read and the 5' SS of the first intron transcribed in the pair to avoid biases from read length constraints.
- 131 Retrieve the splicing status of each intron in a read as described in Step 127 and save the information in a dictionary ('splicing dictionary').
- 132 Extract neighboring intron pairs in which the two introns have different splicing statuses (one intron is spliced and one is unspliced).
- 133 Plot the splicing frequency of the two introns in the pair as a bar plot.
- 134 *Plot splicing patterns (Steps 134–137).* To analyze and plot the splicing patterns of three or more consecutive introns, use the commands within the section 'Splicing coordination across multiple introns' of the script `main_nanoCOP_analyses.ipynb`.
- 135 Using the splicing dictionary created in Step 131, extract reads containing three, four, or more consecutive introns where at least one intron is spliced and one is unspliced.
- 136 Calculate the frequency of each possible isoform. For example, for intron triplets there are three possible isoforms where one intron is spliced and two are unspliced and three possible isoforms where two introns are spliced and one is unspliced.
- 137 Plot the frequency of each isoform as a bar plot.

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Possible solution
12,18,19, Box 2 (for troubleshooting of these steps by western blot, see ref. ³⁸)	Comparable splicing levels in the chromatin and cytoplasmic fractions or higher splicing levels than expected in the chromatin fraction	Inefficient cell lysis due to an excess of input per cell fractionation reaction	Reduce the number of cells per cell fractionation reaction. This number should not exceed 10 million K562 or BL1184 cells and 50 million <i>Drosophila</i> S2 cells
	Comparable splicing levels in the chromatin and cytoplasmic fractions or lower splicing levels than expected in the cytoplasmic fraction	Incomplete removal of the supernatant after washing of the cell nuclei (Step 14) Treatment during cell lysis is too harsh and results in dissociation of transcribing Pol II	Completely remove the supernatant after washing the cell nuclei Decrease the amount of cytoplasmic lysis buffer, but note that too little lysis buffer can also negatively affect cell lysis efficiency. Make sure to use P1000 cut tips during resuspension of pellets
35	Low recovery after RNA extraction	Not enough RNA present for efficient precipitation	Combine more chromatin pellets together at Step 19 for one RNA extraction
81	Low recovery after 4sU pulldown	Inefficient biotinylation reaction	Make a new stock of EZ-Link Biotin-HPDP solution, ensuring it is fully resuspended Increase 4sU pulse time
103	Low sequencing depth (<200,000 'passed' reads after base calling with MinION flow cell; <400,000 'passed' reads after base calling with PromethION flow cell)	Not enough transcription during pulse in cell line of interest	Ensure the buffers for the tailing reaction are made fresh with DTT that has not undergone multiple freeze-thaw cycles Test different incubation times or MnCl ₂ concentrations for the tailing reaction using oGAB11 (Box 1) to add >10 nt to all RNAs Do not load sample if there is an air bubble in the flow cell prior to loading Contact ONT customer support or follow the troubleshooting tips for sequencing runs on the ONT community website
		Inefficient tailing reaction	
119	Reads are shorter than expected	Air bubble in the flow cell or other manufacturing or user error during sample loading	Make all buffers and solutions using RNase-free reagents, especially in the 3'-end tailing reactions
120	A minority of 3' ends map to gene bodies	Presence of RNases in some reagents	Verify efficiency of cytoplasmic lysis by qRT-PCR (Step 12 and Box 2) or western blot ³⁸
121,128	High splicing levels (>20%) in the first window (100–200 nt) of distance transcribed by splicing plot, likely stemming from background signal	Inefficient poly(A) tailing (when using the standard RTA-T ₁₀ adapter)	Confirm that the poly(A) tailing reaction is working properly with an RNA oligonucleotide (see Box 1)
		Overabundance of non-nascent, spliced RNA Poor annotation of poly(A) sites in cell line of interest	Decrease 4sU pulse time Use publicly available datasets from experiments such as RNA-PET (for example from ⁵²) to better annotate poly(A) sites or generate a dataset without 3'-end tailing (Steps 94–99) and with the RTA provided in the direct RNA sequencing kit (Step 100) to identify unannotated loci where polyadenylated RNAs are retained on chromatin

Timing

Step 1, prepare reagents for cellular fractionation: 2.5 h
 Steps 2–22, 4sU labeling and cellular fractionation: 1.5 h
 Steps 23–35, purification of chromatin-associated RNA: 2 h

Steps 36–41, biotinylation of 4sU-labeled RNA: 2 h
 Steps 42–55, clean-up of biotinylated 4sU-labeled RNA: 1.5 h
 Steps 56–65, selection of 4sU-labeled RNA: 1.5 h
 Steps 66–81, purification of selected 4sU-labeled RNA: 1 h
 Step 82, rRNA depletion: 1 h if using RiboMinus kit (may depend on the kit used)
 Steps 83–93, clean-up of rRNA-depleted RNA: 2 h, or overnight and 1.5 h the next day
 Steps 94–98, 3'-end tailing: 40 min (poly(A) tailing) or 1 h (poly(I) tailing)
 Step 99, clean-up of 3'-end tailed RNA: 2 h, or overnight and 1.5 h the next day
 Steps 100–102, direct RNA preparation and sample loading for sequencing: 3 h
 Steps 103–104, direct RNA sequencing: up to 72 h. However, the nano-COP samples described in this protocol were sequenced for 48 h and most sequencing is collected within the first 12 h. Data can be analyzed immediately, even while the sequencing run is in progress
 Steps 105–137, bioinformatic analysis: 1–2 d. This is estimated based on the nano-COP datasets, analyses and computer setup described herein and so may vary

Anticipated results

The anticipated yields of RNA for each main section of the procedure are indicated in Table 2 for the three cell types on which we have performed nano-COP. These numbers may change if nano-COP is performed on other cell types or with a longer 4sU pulse.

The number of sequenced reads, uniquely mapped reads and the read length obtained are noted in Table 2. These statistics also vary by cell line and tend to be higher in cell lines that show less splicing, such as human BL1184 lymphoblasts, and lower in cell lines with more splicing, such as *Drosophila* S2 cells. Longer reads generally lead to higher proportions of uniquely mapped reads, so we expect that these statistics will continue to increase relative to what is indicated below as ONT sequencing and base calling accuracy improve.

Nano-COP is expected to yield long reads that span multiple introns and exons (Fig. 6a). The majority of RNA 3' read ends (>50%) from nano-COP should map to gene bodies (introns, exons, splice sites). This, along with the presence of unspliced introns (Fig. 6a), is indicative that the reads originate from nascent RNA currently undergoing transcription and processing. The proportion of read ends mapping to introns or exons will depend on the species and the corresponding gene structure. For instance, in human cells, where introns are long (typically >1 kb) and exons are short (typically <200 nt), the majority of read ends from nascent RNA are expected to map in introns (Fig. 6b). If poly(A) tailing is used, 10–20% of read ends will map to annotated poly(A) sites (Fig. 6b, left), while this number is substantially lower (<10%) if poly(I) tailing is used (Fig. 6b, right), although this may also vary from one cell type to another.

Table 2 | Anticipated RNA yield and sequencing statistics when performing nano-COP in human K562, BL1184 or *Drosophila* S2 cells

Step	Human K562	Human BL1184	<i>Drosophila</i> S2
35, Purification of chromatin-associated RNA	100–200 µg from 12 cellular fractionation reactions combined	65–75 µg from 12 cellular fractionation reactions combined	80–100 µg from 12 cellular fractionation reactions combined
81, Selection of 4sU-labeled RNA	2–10% of chromatin-associated RNA used (with 8 min 4sU pulse)	2–4% of chromatin-associated RNA used (with 8 or 10 min 4sU pulse)	5–15% of chromatin-associated RNA used (with 8 min 4sU pulse)
93, rRNA depletion	10–30% of 4sU-chr RNA used (depends on kit used)	30–45% of 4sU-chr RNA used (depends on kit used)	8–20% of 4sU-chr RNA used (depends on kit used)
103, Direct RNA sequencing	>200,000 'pass' read	>200,000 'pass' reads	>200,000 'pass' reads
119, Sequencing statistics	Median read length: 460–680 nt. Uniquely mapped reads: >75%	Median read length: 600–830 nt Uniquely mapped reads: >80%	Median read length: 410–510 nt Uniquely mapped reads: 10–60% (depends on rRNA depletion kit used)

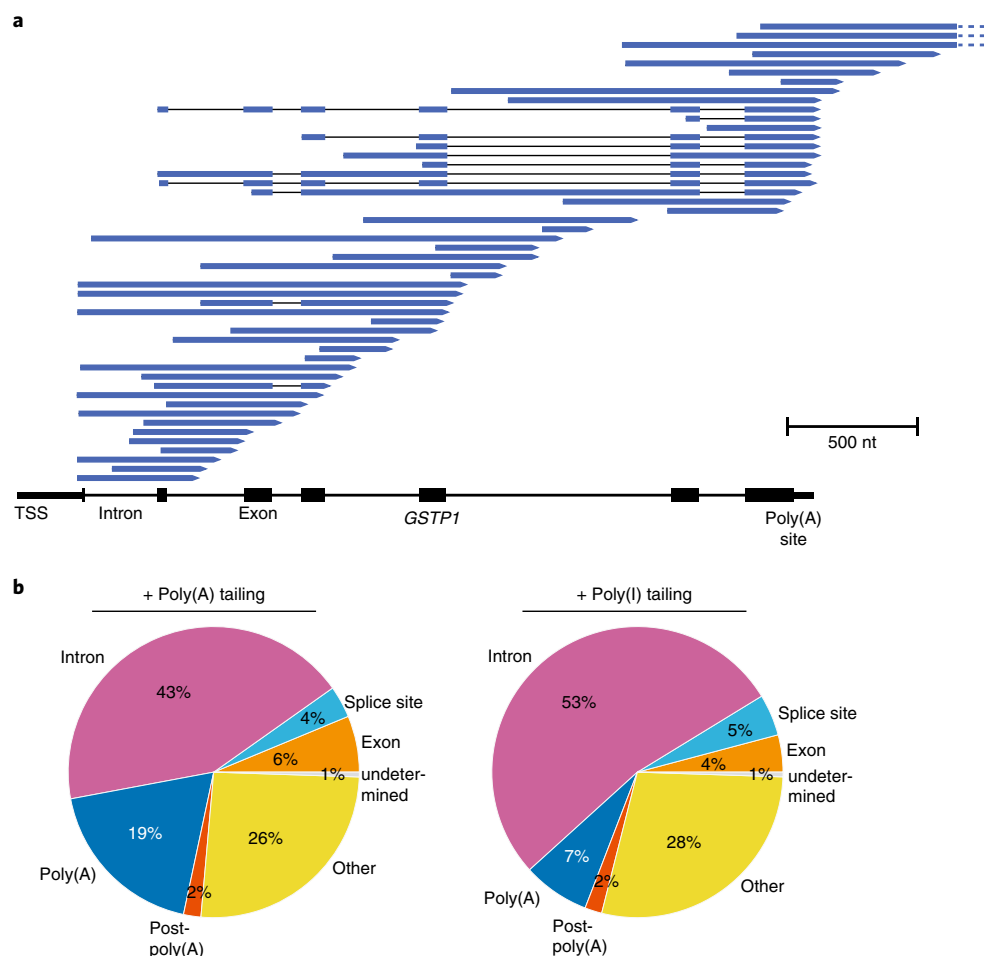


Fig. 6 | Nano-COP captures the nascent transcriptome. **a**, Representative nano-COP reads aligned to the *GSTP1* gene in human K562 cells. The gene structure is represented from the transcription start site (TSS) to the poly(A) site, with black boxes representing exons and lines representing introns. Within the reads, blue boxes represent read coverage, black lines represent skipped coverage due to splicing, and the start of the read (3' end of RNA) is represented with an arrow. Dashed lines represent reads that continue beyond the region displayed. **b**, Distribution of nano-COP 3' ends in human K562 cells with enzymatic poly(A) tail addition (left) and poly(I) tail addition (right). 'Poly(A)' sites are defined as regions within 50 nucleotides of the end coordinate of annotated genes or of RNA-PET annotations from cytoplasm and chromatin fractions in K562 ENCODE data (ENCODE Project Consortium, 2012). 'Post-poly(A)' sites are defined as the region between 50–550 nucleotides after the end of annotated genes. 'Splice sites' are defined as 50 nucleotides upstream and 10 nucleotides downstream of annotated 5' SSs. 'Undetermined' indicate reads that align to more than one category and 'other' represents read ends that do not align in the sense direction of annotated gene features (e.g., antisense transcripts, non-coding RNAs and intergenic transcription). Figure adapted from ref. ¹⁸.

Data availability

The accession numbers for the nanopore sequencing data presented in this paper are Gene Expression Omnibus (GEO): [GSE123191](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123191) (data from ref. ¹⁸) and [GSE154079](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154079). Supplementary Table 1 indicates the correct accession number for each sample. Source data are provided with this paper.

Code availability

All scripts for data analyses described in this paper are available at <https://github.com/churchmanlab/nano-COP>. The code for nanopolish-detect-polyI is available at <https://github.com/jts/nanopolish.git>.

References

1. Keohavong, P., Gattoni, R., LeMoullec, J. M., Jacob, M. & Stévenin, J. The orderly splicing of the first three leaders of the adenovirus-2 major late transcript. *Nucleic Acids Res.* **10**, 1215–1229 (1982).
2. Mariman, E. C., van Beek-Reinders, R. J. & van Venrooij, W. J. Alternative splicing pathways exist in the formation of adenoviral late messenger RNAs. *J. Mol. Biol.* **163**, 239–256 (1983).
3. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754–765 (1988).
4. Audibert, A., Weil, D. & Dautry, F. In vivo kinetics of mRNA splicing and transport in mammalian cells. *Mol. Cell. Biol.* **22**, 6706–6718 (2002).
5. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* **16**, 1128–1133 (2009).
6. Coulon, A. et al. Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife* **3**, e03939 (2014).
7. Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* **4**, 1144–1155 (2013).
8. Rabani, M. et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014).
9. Wachutka, L., Caizzi, L., Gagneur, J. & Cramer, P. Global donor and acceptor splicing site kinetics in human cells. *eLife* **8**, e45056 (2019).
10. Wan, Y. et al. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3467157> (2019).
11. Takahara, K. et al. Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptor-site mutation that results in abnormal Pro- α 1(V) N-propeptides and ehlers-danlos syndrome type I. *Am. J. Hum. Genet.* **71**, 451–465 (2002).
12. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28**, 2663–2676 (2014).
13. Dujardin, G. et al. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* **54**, 683–690 (2014).
14. De la Mata, M. et al. A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12**, 525–532 (2003).
15. Kessler, O., Jiang, Y. & Chasin, L. A. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol.* **13**, 6211–6222 (1993).
16. Schwarze, U., Starman, B. J. & Byers, P. H. Redefinition of exon 7 in the COL1A1 gene of type I collagen by an intron 8 splice-donor-site mutation in a form of osteogenesis imperfecta: influence of intron splice order on outcome of splice-site mutation. *Am. J. Hum. Genet.* **65**, 336–344 (1999).
17. Kim, S. W. et al. Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Res.* **45**, 9503–9513 (2017).
18. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* **77**, 985–998.e8 (2020).
19. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
20. Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
21. Eid, J. et al. Single polymerase molecules. *Science* **323**, 133–138 (2009).
22. Weirather, J. L. et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res.* **6**, 100 (2017).
23. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
24. Sonesson, C. et al. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat. Commun.* **10**, 3359 (2019).
25. Jia, J. et al. Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat. Plants* **6**, 780–788 (2020).
26. Danko, C. G. et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50**, 212–222 (2013).
27. Veloso, A. et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**, 896–905 (2014).
28. Dölken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
29. Windhager, L. et al. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* **22**, 2031–2042 (2012).
30. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
31. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**, 436–442 (2011).
32. Gregersen, L. H., Mitter, R. & Svejstrup, J. Q. Using Ttchem-seq for profiling nascent transcription and measuring transcript elongation. *Nat. Protoc.* **15**, 604–627 (2020).

33. Pai, A. A. et al. The kinetics of pre-mRNA splicing in the *Drosophila* genome and the influence of gene architecture. *eLife* **6**, 1–26 (2017).
34. Maier, K. C., Gressel, S., Cramer, P. & Schwalb, B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res.* **30**, 1332–1344 (2020).
35. Carrillo Oesterreich, F. et al. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**, 372–381 (2016).
36. Brody, Y. et al. The In vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* **9**, e1000573 (2011).
37. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).
38. Mayer, A. & Churchman, L. S. Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat. Protoc.* **11**, 813–833 (2016).
39. Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G. & Rutter, W. J. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* **170**, 447–449 (1970).
40. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
41. Han, F. & Lillard, S. J. In-situ sampling and separation of RNA from individual mammalian cells. *Anal. Chem.* **72**, 4073–4079 (2000).
42. Jackson, D. A., Iborra, F. J., Manders, E. M. & Cook, P. R. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 (1998).
43. Rädle, B. et al. Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *J. Vis. Exp.* <https://doi.org/10.3791/50195> (2013).
44. Tani, H. et al. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956 (2012).
45. Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C. & Simon, M. D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **15**, 221–225 (2018).
46. Kovaka, S., Fan, Y., Ni, B., Timp, W. & Schatz, M. C. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0731-9> (2020).
47. Payne, A. et al. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.02.03.926956> (2020).
48. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
49. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
50. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

Acknowledgements

We thank members of the Churchman lab, F. Winston, W. Timp, R. Workman, N. Sadowski, M. Marin, B. Smalec, M. Richardson, R. Ietswaart, and A. Markham for helpful discussions, advice, and assistance and C. Kaplan, J. Bridgers, B. Smalec, J. Falk and C. Patil for critical reading of the manuscript. This work was supported by the NIH (R21-HG009264, R01-HG010538, and R01-GM117333 to L.S.C.; F31-GM122133 to H.L.D.), an NSF Graduate Research Fellowship to H.E.M., the Fonds de Recherche du Québec–Santé and the Canadian Institutes of Health Research (Post-doctoral fellowship awards to K.C.). J.T.S. is supported by the Ontario Institute for Cancer Research through funds provided by the Government of Ontario and the Government of Canada through Genome Canada and Ontario Genomics (OGI-136).

Author contributions

H.L.D., K.C. and L.S.C. conceived and designed the study. H.L.D. established the nano-COP protocol and K.C. developed the poly(I) tailing approach. H.L.D., K.C. and H.E.M. performed experiments. H.L.D. and K.C. generated scripts and performed data analysis. P.S.T. and J.T.S. developed nanopore detect-poly(I). H.L.D., K.C., H.E.M. and L.S.C. wrote the manuscript. P.S.T. and J.T.S. reviewed and edited the manuscript.

Competing interests

J.T.S. receives research funding from Oxford Nanopore Technologies. J.T.S. and H.L.D. have received travel support to attend and speak at meetings organized by Oxford Nanopore Technologies. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41596-020-00469-y>.

Correspondence and requests for materials should be addressed to L.S.C.

Peer review information *Nature Protocols* thanks Shobbir Hussain, Jixian Zhai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 July 2020; Accepted: 20 November 2020;
Published online: 29 January 2021

Related links**Key reference using this protocol:**

Drexler, H. L., Choquet, K. & Churchman, L. S. *Mol. Cell* **77**, 985–998.e8 (2020): <https://doi.org/10.1016/j.molcel.2019.11.017>