

Mapping and quantifying nascent transcript start sites using TT-TSS-seq

Eleanor Elgood Hunt, Claudia Vivori, Richard Mitter, Vedis Agnadottir, and Folkert J. van Werven

The Francis Crick Institute, London, NW1 1AT, United Kingdom

Transcription initiation is a highly dynamic and tightly regulated process involving the coordinated action of transcription factors, chromatin remodelers, and RNA polymerase, which determine where and when transcription begins. Accurately mapping and quantifying transcription start sites (TSSs) from nascently transcribed RNAs remains a key area of interest, as it provides critical insights into transcription dynamics. Here, we combine transient transcriptome sequencing with transcription start site sequencing (TT-TSS-seq) to accurately map and quantify transcription initiation sites from nascent transcripts. Because transient metabolic labeling yields low-input RNA, we optimize the TSS-seq protocol to enhance sensitivity and accuracy. Specifically, we refine enzymatic reactions for decapping and RNA ligation and incorporate 5' oligonucleotides containing unique molecular identifiers (UMIs) and barcodes to enable accurate quantification and sample multiplexing. The TT-TSS-seq approach detects transcription initiation of unstable transcripts, such as enhancer RNAs. Moreover, we show that a large fraction of genes use multiple transcription initiation sites, yet often produce only a single stable transcript. Overall, TT-TSS-seq provides precise mapping and quantification of transcription initiation sites, offering new insights into transcriptional dynamics and expanding the toolkit for studying gene regulation.

[Supplemental material is available for this article.]

Transcription initiation is a critical regulatory step in which RNA polymerase II (RNAPII), guided by transcription factors and cofactors, is recruited to the promoter region and starts synthesizing RNA. The selection of transcription start sites (TSSs) plays a fundamental role in determining transcript diversity and gene expression patterns. Most gene loci harbor multiple TSSs, which are differentially regulated in response to cellular conditions, developmental cues, and environmental signals (Carninci et al. 2006; Wang et al. 2008; Brown et al. 2014; Lu and Lin 2019; Chia et al. 2021). Despite the prevalence of alternative TSS usage, the functional significance of this regulatory mechanism remains largely unknown. Understanding how transcription initiation is controlled and how different TSSs contribute to gene expression and cellular function is essential for uncovering new layers of gene regulation. Moreover, dysregulation of TSS selection has been linked to various diseases, including cancer and neurodevelopmental disorders, highlighting the need for further investigation into its underlying mechanisms (Thorsen et al. 2011; Demircioğlu et al. 2019).

For mapping TSSs, sequencing approaches have been employed with different biochemical strategies to capture the 5' ends of transcripts. Cap-trapping techniques, including Cap Analysis of Gene Expression (CAGE), are the most commonly used. These techniques involve chemical treatment to oxidize the 5' caps of RNA, enabling biotinylation and subsequent streptavidin pulldown to enrich for capped transcripts (Takahashi et al. 2012). Template-switching reverse transcription (TSRT), which involves reverse transcription using an enzyme which adds 1–3 non-templated bases (typically cytosine) at the 5' end, works well for low-input material reactions (Policastro et al. 2020). Oligo-capping

methods provide a high resolution but require a relatively large amount of input material. These processes involve the dephosphorylation of uncapped RNA, followed by enzymatic decapping to generate 5' monophosphate ends, which are essential for subsequent adaptor ligation (Arribere and Gilbert 2013; Pelechano et al. 2013).

Most TSS-based sequencing techniques use mature RNA, which enables the detection of TSSs of stably expressed transcripts but not of TSSs from unstable RNAs, such as long noncoding transcripts. Furthermore, using mature RNA prevents insights into transcription initiation dynamics. Therefore, nascent or newly transcribed RNA is required for detecting TSSs of unstable transcripts and quantitative analysis of transcription initiation. A few approaches have been developed to map and quantify TSS usage for nascent or newly synthesized RNA. For example, in NET-CAGE, RNA still associated with chromatin is selectively captured, enabling TSS mapping of actively transcribed genes (Hirabayashi et al. 2019). Isolation of RNAPII-associated nascent RNAs (e.g., POINT-5-seq), instead, provides insights into transcription initiation and early elongation (Sousa-Lúis et al. 2021). In vitro run-on assays, such as GRO-cap and PRO-cap, label and sequence newly synthesized RNA to determine TSSs (Danko et al. 2015; Mahat et al. 2016). Short-capped RNA sequencing (scaRNA-seq) selectively captures short, capped transcripts to identify sites of transcription initiation (Larke et al. 2021). Limitations of some of these approaches include the need for complex biochemical purifications or fractionations and the risk of contamination with steady-state RNAs.

Here, we aimed to combine transient metabolic labeling with the oligo-capping approach to map sites of transcription initiation at the nucleotide resolution (TT-TSS-seq). We first optimized the

Corresponding author: folkert.vanwerven@crick.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280726.125>. Freely available online through the *Genome Research* Open Access option.

© 2026 Elgood Hunt et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

chemical and enzymatic reaction conditions for TSS-seq and then applied it to transiently labeled RNA isolated from mouse embryonic stem cells (mESCs) (Gregersen et al. 2020; Chia et al. 2021).

Results

Optimization of the TSS-seq protocol

To develop TSS-seq for nascent or newly transcribed RNA, we improved the previously described TSS-seq protocol for relatively low RNA input (Fig. 1A; Chia et al. 2021). We chose an oligo-capping strategy because it enables the detection of TSSs at single-nucleotide resolution (Pelechano et al. 2013; Chia et al. 2021). To optimize the protocol for low amounts of RNA, we incorporated part of the iCLIP2 protocol for library preparation (Buchbender et al. 2020). First, cells are pulsed with 4-thiouridine (4sU), resulting in the labeling of newly synthesised RNA. RNA is treated with alkaline phosphatase to remove the 5'-phosphate groups from noncapped RNA. Next, the mRNA decapping enzyme (MDE) removes the 5'-terminal caps, exposing the 5'-phosphate group exclusively on once-capped transcripts. Subsequently, a single-stranded 5' RNA-DNA hybrid adaptor containing a barcode for sample multiplexing and a unique molecular identifier (UMI) is ligated. After the pooling of samples, polyadenylated (poly(A)⁺) RNA or transiently 4sU labeled RNA populations are purified (for poly(A)-TSS-seq or TT-TSS-seq, respectively). RNA is fragmented, and the 3' end fixed, before a preadenylated 3' adaptor is ligated using a truncated T4 RNA ligase that can only utilize preadenylated substrates. RNA is then reverse-transcribed and the cDNA preamplified before the first size selection is carried out to remove primer dimers and short inserts. A second amplification and size selection are performed, followed by quality control and sequencing.

Multiple steps of the TSS-seq protocol were optimized, including fragmentation, reverse transcription, RNA clean-up, PCR, and DNA size selection conditions. In addition, we increased the efficiency of the alkaline phosphatase, MDE reaction, and ligation reactions (Supplemental Table S1). We also denatured RNA before ligation to decrease the effects of RNA structure on the ligation reaction. Additionally, we included a control where the RNA was not treated with MDE. This yields a low library count and very few reads; however, it serves as a qualitative control when we optimized the protocol, suggesting that enzymatic reactions are specific and efficient.

We have robust QC analysis for TSS-seq libraries. In short, we used the same cycle number and conditions as recommended in the iCLIP2 protocol on which our protocol is based (Buchbender et al. 2020). For each library preparation, product formation was verified by gel electrophoresis, and the final pooled libraries were additionally assessed using a Bioanalyzer. The incorporation of UMIs allows for the removal of PCR duplicates, thereby ensuring accurate quantification.

To compare the optimized protocol to the original one, we first performed poly(A)-TSS-seq on RNA purified from yeast cells. The 5'-most nucleotides, referred to as "tags," of the reads were extracted for mapping of TSSs. The optimized protocol showed an increased fraction of tags located in the promoter-proximal region (−300 to +100 bp from annotated TSS) across different count thresholds (Fig. 1B,C). Additionally, the optimized protocol showed an increase in the number of tags (threshold of 3 or more) in the promoter-proximal region while decreasing the number of detected tags in the negative control, compared to the tags detected in the original conditions (Fig. 1D). TSS-seq with the op-

timized protocol showed an increase in the fraction of tags in the promoter-proximal region (94% vs. 65% with the original protocol) (Fig. 1E).

We determined the dinucleotide frequencies to assess whether differences occurred between the original and optimized protocols. The first nucleotide represents the −1 nucleotide, and the second nucleotide represents the +1 tag/TSS as identified by poly(A)-TSS-seq (Supplemental Fig. S1). We noted that the original protocol had the highest preference for AA₊₁, whereas the optimized protocol had an increased preference for CA₊₁ and TA₊₁. The CA₊₁ and TA₊₁ motif identified with the optimized protocol has also been reported in other studies (Policastro et al. 2020); however, we noted less enrichment for CG₊₁ and TG₊₁ motifs as reported in some studies (Zhang and Dietrich 2005; Lu and Lin 2019). Possibly, differences in threshold levels set for the analyses could affect the outcome of the dinucleotide motifs between data sets. Alternatively, CAGE and TSRT are known to be susceptible to 5' G artifacts, which could affect the correct calling of TSSs and inflate the CG₊₁ and TG₊₁ motifs (Policastro et al. 2020). In conclusion, the optimized TSS-seq protocol increased the accuracy of TSS detection in poly(A)⁺ RNA isolated from yeast. We recently applied the optimized TSS-seq protocol to study how DNA supercoiling affects alternative TSS usage (Elgood Hunt et al. 2026).

Multiplexing of samples

In the optimized TSS-seq protocol, we incorporated a barcode into the 5' adaptor to enable sample multiplexing. To assess whether the barcoded 5' adaptors gave comparable tag profiles, we carried out poly(A)-TSS-seq with yeast RNA. Dephosphorylation and decapping were carried out in a single reaction before samples were split, and twelve different barcoded 5' adaptors were ligated. These samples were then pooled together, and PCR was performed using primers with the same index.

The poly(A)-TSS-seq samples generated with barcoded 5' adaptors showed a comparable genomic distribution (Fig. 2A,B). Libraries produced using barcoded 5' adaptors showed a good correlation with each other (Pearson correlation coefficient: $r=0.92-0.98$) (Supplemental Fig. S2A). We determined the dinucleotide frequencies to assess whether there was a bias in adapter ligation. The first nucleotide represents the −1 nucleotide, and the second nucleotide represents the +1 tag/TSS identified by poly(A)-TSS-seq (Supplemental Fig. S2B). We observed enrichment for +1 As in the tags/TSSs and no clear differences in dinucleotide usage between the barcodes (Supplemental Fig. S2B). This suggested that the presence of the barcode does not bias towards the ligation of a specific 5' nucleotide. Thus, multiplexing samples, as described for the optimized TSS-seq protocol, is a feasible strategy for handling multiple samples and may offer potential advantages in reducing sample loss in the subsequent steps of the protocol.

Poly(A)-TSS-seq in mESCs and comparison to CAGE

To further validate the optimized poly(A)-TSS-seq protocol, it was similarly applied to RNA isolated from mESCs. The identified tags showed strong enrichment to annotated TSSs (Fig. 3A). For further analysis, we defined the promoter-proximal region as −500 to +500 bp from annotated TSSs. The proportion of promoter-proximal tags was determined at different count thresholds, and more than 90% of tags (using a threshold of three counts) were annotated to promoter-proximal regions (Fig. 3A,B). Furthermore, ~31,500 transcripts featured, at minimum, one TSS (Fig. 3A).

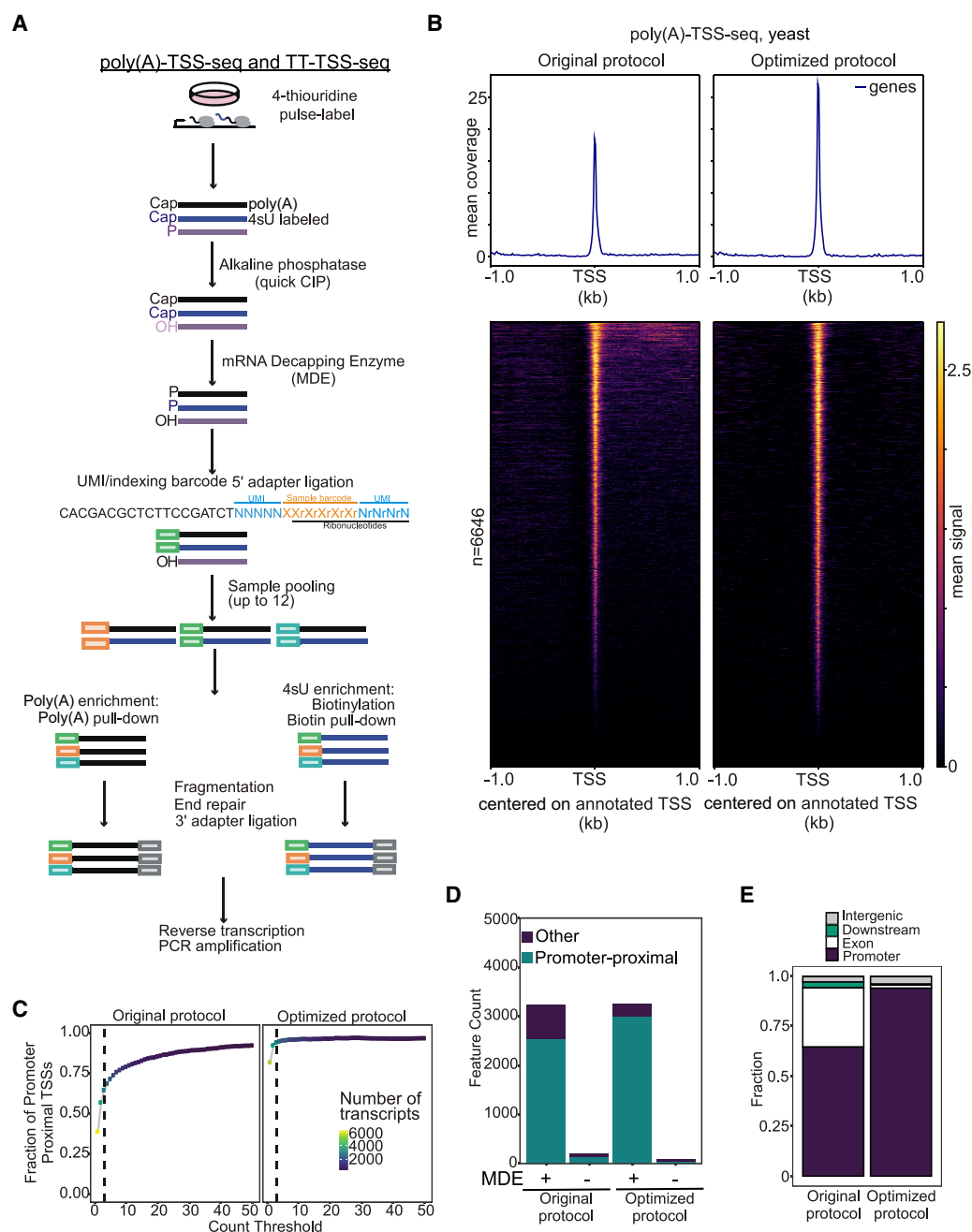


Figure 1. Optimization of the TSS-seq protocol with yeast RNA. (A) Schematic of the TSS-sequencing protocol. Alkaline phosphatase (quickCIP) removes the 5'-phosphate groups from noncapped fragments (shown in purple). The mRNA decapping enzyme (MDE) removes the 5'-terminal caps, exposing a 5'-phosphate group (labeled P). This enables the ligation of a 5' adaptor, which contains sample barcodes to allow sample multiplexing and UMIs. Nascent (4 thiouridine (4sU), labeled blue) or steady-state RNA (poly(A), labeled black) is then selected, before fragmentation, end-repair, 3' adaptor ligation, reverse transcription, and PCR amplification. (B) Density plot and heat map showing tag locations compared to annotated TSSs for poly(A)-TSS-seq. RNA isolated from yeast was subjected to poly(A)-TSS-seq using either the original or optimized conditions. The x-axis is centered on the annotated TSS. (C) Fraction (y-axis) of tags located in promoter-proximal regions (–300 to +100 bp from annotated ORF start codons) at different read thresholds (x-axis). (D) Bar plot showing the number of tags located in promoter-proximal regions with the original or optimized poly(A)-TSS-seq protocols (+MDE). As negative controls, samples that had not undergone the decapping reaction were included (–MDE). A threshold of $n = 3$ counts was applied. (E) Genomic locations of detected tags. Tag counts were normalized using the DESeq2 median-of-ratios approach using a threshold of $n = 3$ counts.

We compared poly(A)-TSS-seq data with two published CAGE data sets from similar cell lines, mESCs (Noguchi et al. 2017; Lloret-Llinares et al. 2018). CAGE is the gold standard for the 5'-end mapping. We observed substantial overlap in the TSSs detected between poly(A)-TSS-seq and the two CAGE data sets (Fig. 3C;

Supplemental Fig. S3). The poly(A)-TSS-seq data set showed stronger enrichment at promoter regions compared to both CAGE data sets (Fig. 3D). Despite this overlap, each method also identified unique peaks, which may reflect either technical differences or cell culture- and cell line-specific biases that we cannot distinguish

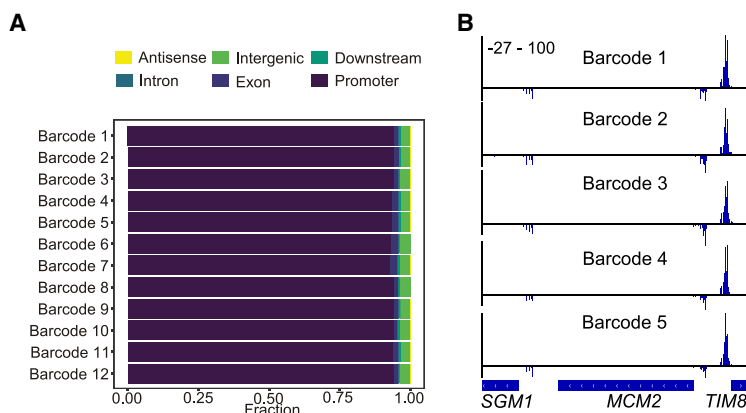


Figure 2. Testing of sample multiplexing in optimized TSS-seq protocol, using yeast RNA. (A) Genomic locations of detected TSSs using the different 5' adaptor barcode sequences (Supplemental Document S1). (B) Integrative Genomics Viewer (IGV; Robinson et al. 2011) showing an example locus of TSS-seq signals for barcodes 1 to 5.

(Supplemental Fig. S3). A closer assessment of example loci suggests that the data sets can also display distinct TSS patterns between poly(A)-TSS-seq and CAGE, which may reflect (from either method) in TSS detection (Fig. 3E). For example, additional tags were identified in the promoter with poly(A)-TSS-seq at the *Sox2* locus. Tags were detected within gene bodies in the two CAGE data sets at the *Srm2* locus, which could be indicative of increased background signal (Fig. 3E). We conclude that the optimized poly(A)-TSS-seq protocol is well-suited for studying more complex transcriptomes.

Transient 4sU labeling of RNA combined with TSS-seq (TT-TSS-seq)

Multiple methods have been described to enrich nascent or newly transcribed RNAs (Kruesi et al. 2013; Hirabayashi et al. 2019; Larke et al. 2021; Sousa-Luís et al. 2021). Each method offers distinct advantages while also presenting certain limitations (Wissink et al. 2019). Transient labeling of RNA with 4sU provides a way to capture RNAs from actively transcribing polymerases with advantages including high reproducibility and in vivo labeling of unperturbed cells (TT-seq or TT_{chem}-seq) (Schwalb et al. 2016; Gregersen et al. 2020). TT-seq methods have been widely used to study transcription dynamics.

To balance RNA yield without compromising transcription dynamics, we assessed the degree of 4sU incorporation following different labeling times and using different 4sU concentrations in mESCs (Supplemental Fig. S4A). With increased labeling time or 4sU concentration, increased 4sU incorporation occurred, as expected, and no signal was detected in the control (no 4sU) (Supplemental Fig. S4A). Similar to the TT_{chem}-seq protocol, we used 15 min of labeling with 1 mM 4sU for further analysis (Gregersen et al. 2020).

We carried out TSS-seq on purified RNA that was transiently labeled with 4sU (TT-TSS-seq). TT-TSS-seq requires a substantial amount of input material (100 µg) because nascent transcripts constitute only a small fraction of total RNA (~0.1%–0.5%), of which, in turn, a fraction contains capped nascent transcripts. In general, we observed low levels of ribosomal RNAs (rRNAs) present in the poly(A)-TSS-seq and TT-TSS-seq samples. Approximately 1% of poly(A)-TSS-seq and 8% of TT-TSS-seq reads mapped to rRNAs or tRNAs. This is because the TSS-seq library preparation selectively captures capped RNAs, which rRNAs and tRNAs lack.

We also used another method to enrich nascent RNA and selected transcripts with <300 nucleotides, as in short capped RNA-seq (Fig. 4A; Supplemental Fig. S4B; Larke et al. 2021). We used scaRNA-seq because it used a similar library preparation protocol and was relatively straightforward to adapt.

We compared the TT-TSS-seq profile to the profiles of poly(A)-TSS-seq and scaRNA-seq (Fig. 4B; Nechaev et al. 2010; Henriques et al. 2018; Larke et al. 2021). The poly(A)-TSS-seq and TT-TSS-seq detected tags showed strong enrichment at the annotated TSSs (Fig. 4C). In contrast, the tags detected with scaRNA-seq also showed signals downstream of annotated TSSs, suggesting the detection of RNA decay intermediates (Fig. 4C). For

each sample, the regions with multiple TSSs were clustered into transcript start regions (TSRs), and the correlation between TSR signals detected by each method was assessed. The TSR signals showed a good correlation between biological replicates (Pearson correlation coefficient: $r=0.86$ for poly(A)-TSS-seq, 0.89 for TT-TSS-seq, and 0.96 for scaRNA-seq) (Supplemental Fig. S4C). However, TSRs detected using scaRNA-seq showed little-to-no correlation with TT-TSS-seq (Pearson correlation coefficient: $r=0-0.22$), whereas TT-TSS-seq and poly(A)-TSS-seq showed some correlation (Pearson correlation coefficient: $r=0.41-0.58$).

We also assessed the dinucleotide frequency in the detected TSRs. A preference for pyrimidines at the -1 and purines at the $+1$ nucleotide (the TSS) has previously been identified (Vo Ngoc et al. 2017). The tags detected by poly(A)-TSS-seq on RNA isolated from mESCs showed enrichment for guanine (Fig. 4D). This is also consistent with the more GC-rich nature of mammalian promoters (Policastro et al. 2020). TSSs identified with TT-TSS-seq showed an increased preference for CA₊₁ and a decreased preference for GG₊₁ and CG₊₁ compared to poly(A)-TSS-seq (Fig. 4D). The CG₊₁ enrichment was consistent with previous work for CAGE (Frith et al. 2008). Other nascent TSS-sequencing techniques, such as GRO-Cap, also identified a preference for $+1$ A (Vo Ngoc et al. 2017; Luse et al. 2020). Furthermore, GRO-Cap identified a decreased preference for $+1$ G and C TSSs in unstable transcripts compared to stable ones (Core et al. 2014). In contrast, the TSSs detected here using scaRNA-seq did not show a preference for a dinucleotide combination (Fig. 3C).

We conclude that TT-TSS-seq is suitable for mapping and quantifying transcription initiation sites. The TSSs detected by scaRNA-seq showed poor correlation with TSSs detected using TT-TSS-seq and poly(TSS-seq and showed a lack of dinucleotide preference. A possible explanation is that other RNA species are present in the scaRNA-seq samples. For example, RNA degradation intermediates may be enriched during size selection, which would create more substrates for the alkaline phosphatase reaction, potentially resulting in incomplete dephosphorylation.

TT-TSS-seq detects TSSs of unstable RNAs

The TSSs identified using the TT-TSS-seq protocol showed less enrichment in the promoter-proximal region; however, it is unclear whether this reflects technical limitations or limitations in current TSS annotations (Fig. 5A). Therefore, we further investigated the

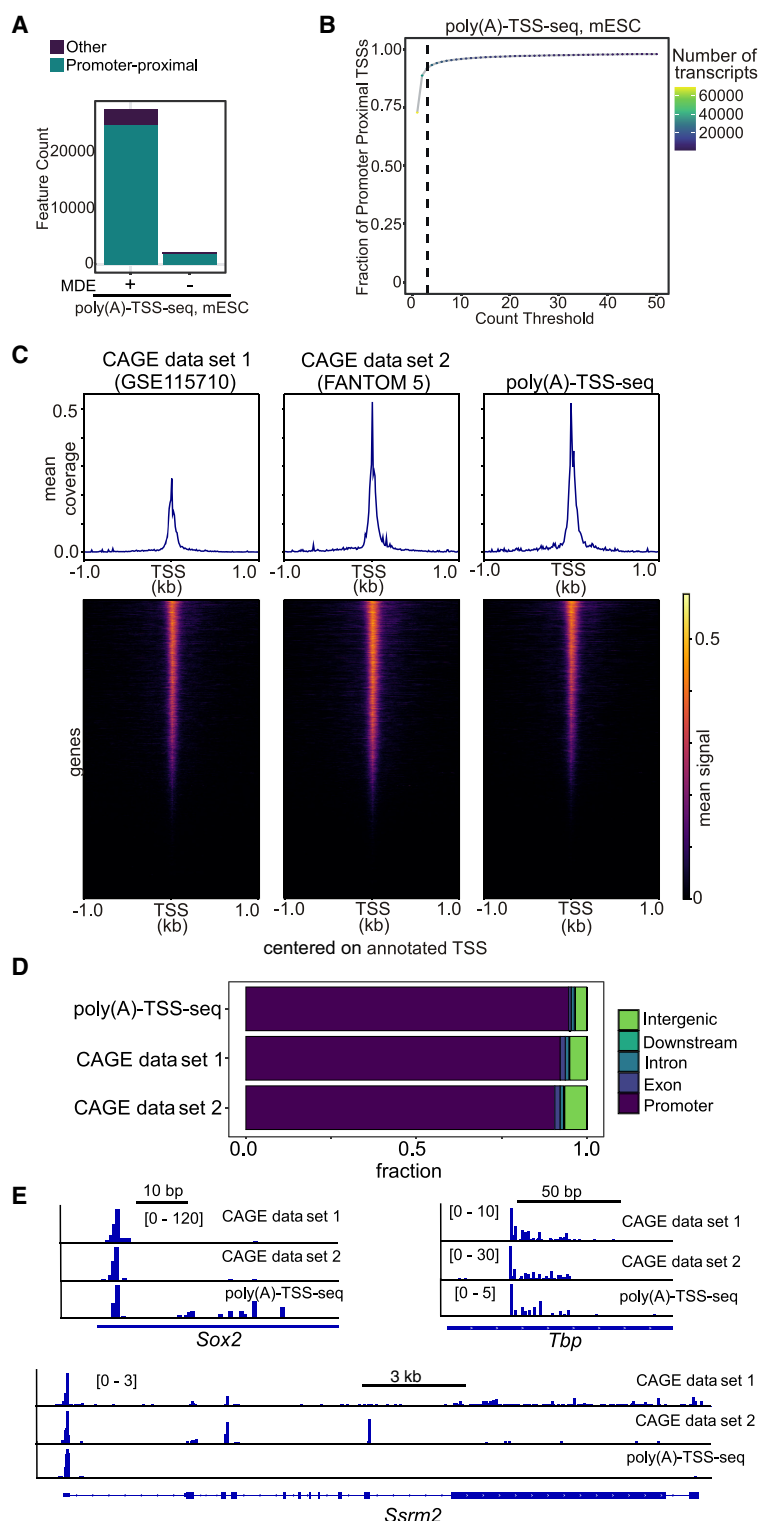


Figure 3. Poly(A)-TSS-seq protocol in mESCs and compared to CAGE. (A) Bar plot showing the number of transcripts with promoter-proximal tags in mESCs detected by poly(A)-TSS-seq (+MDE). As a negative control, a no-decapping reaction sample was included (–MDE). A threshold of $n = 3$ counts was applied. (B) Fraction of tags that are promoter-proximal (–500 to +500 bp from annotated TSSs) at different read count thresholds for mESC. A threshold of $n = 3$ read counts was used. (C) Density plots and heat maps showing tag locations detected with poly(A)-TSS-seq and two CAGE data sets using mESC RNA compared to annotated mouse TSSs. (D) Genomic locations of detected tags with poly(A)-TSS-seq and two CAGE data sets. (E) Example loci of poly(A)-TSS-seq and two CAGE data sets. Showing are *Sox2*, *Tbp*, and *Ssrn2*.

differences between the poly(A)-TSS-seq and TT-TSS-seq data sets. As before, the TSSs were clustered into TSRs, and the number of transcripts with detected TSRs was determined. The number of identified transcripts with TSRs was increased by using TT-TSS-seq compared to poly(A)-TSS-seq (Fig. 5B). Next, we carried out differential analysis on the TSRs detected by TT-TSS-seq and poly(A)-TSS-seq (Fig. 5C). Approximately 11,500 TSRs were detected at significantly increased levels and ~2500 at significantly decreased levels in TT-TSS-seq compared to poly(A)-TSS-seq. More transcripts with retained introns were associated with TSRs detected at significantly higher levels in TT-TSS-seq compared to poly(A)-TSS-seq (Fig. 5D).

Next, we assessed whether the differentially detected TSRs overlapped with enhancer regions. Over 2500 TSRs were detected at significantly higher levels using TT-TSS-seq overlapped with annotated enhancer regions (Fig. 5E). Indeed, it is known that most enhancer RNAs contain a 5' cap (Sartorelli and Laubert 2020). Finally, we determined whether using TT-TSS-seq increased the detection of alternative TSSs. We selected genes with one TSR that showed no significant difference in signal or were increased in poly(A)-TSS-seq compared to TT-TSS-seq. We found that nearly 8000 genes had at least one additional TSR detected at significantly higher levels by TT-TSS-seq (Fig. 5F).

For example, upstream of the *2410003Rik* locus displays an internal TSS detected by TT-TSS-seq but not by poly(A)-TSS-seq or RNA-seq and likely presents an unstable transcript within an intron (Fig. 6). Upstream of the *Lasp1* gene (~2 kb), we detected a TSS by TT-TSS-seq but not by poly(A)-TSS-seq. This upstream TSS was also detected by TT_{chem}-seq but not by RNA-seq, producing a transcript of ~100 bp. This transcript may represent an enhancer RNA, which will require further investigation. It is well established that TT-seq can detect unstable, nascent transcripts and thus provides a sensitive measure of TSSs (Schwalb et al. 2016; Gregersen et al. 2020).

Discussion

Here, we optimized the TSS-seq protocol to study TSSs from nascently transcribed RNAs. The TT-TSS-seq protocol identifies a broad spectrum of TSSs, including

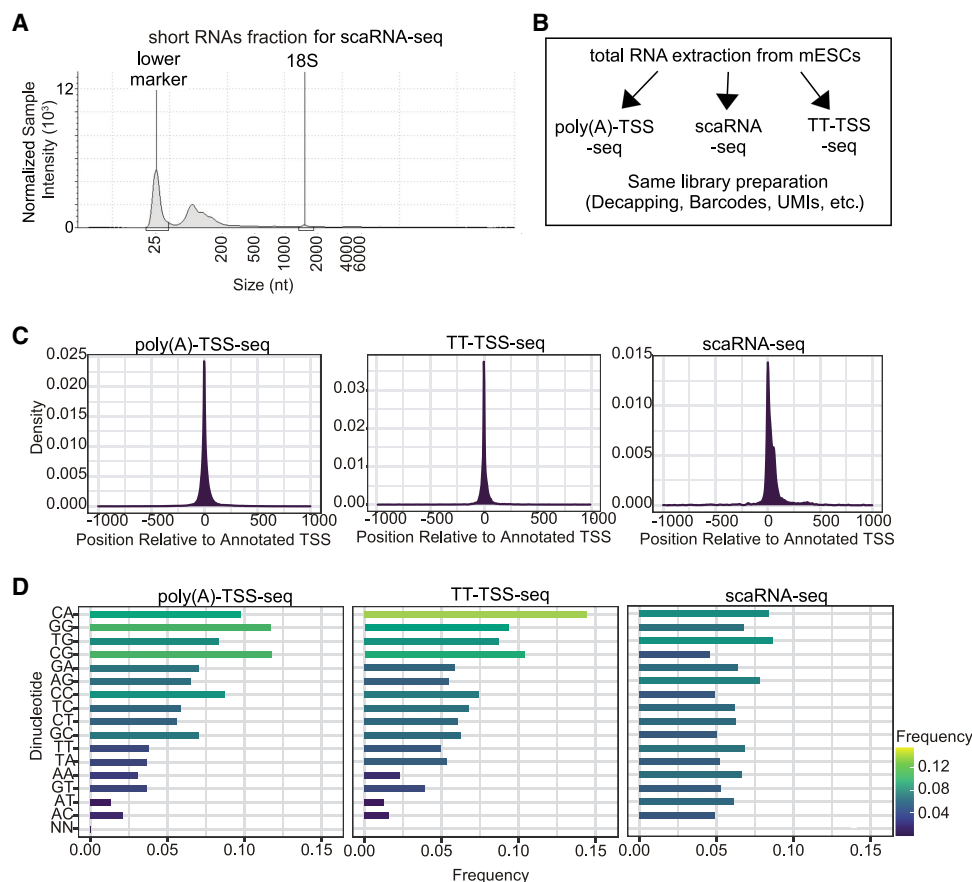


Figure 4. TT-TSS-seq protocol applied to RNA isolated from mESCs. (A) RNA electropherogram, as measured using TapeStation, showing the isolation of short RNAs for scaRNA-seq. The 25-nt peak represents the reference ladder. (B) Scheme of experimental set-up. (C) Density plots showing TSS locations detected using mESC poly(A)-TSS-seq, TT-TSS-seq, and scaRNA-seq (less than 300 nt) compared to annotated mouse TSSs. $n = 2$ biological repeats. (D) Dinucleotide frequencies analysis. The first nucleotide represents the -1 nucleotide, and the second nucleotide represents the identified $+1$ tag/TSS. A threshold of $n = 3$ counts was used.

alternative TSSs. Notably, TT-TSS-seq detects short-lived RNA species such as enhancer RNAs, long-noncoding RNAs, and transcripts predicted to undergo rapid decay. These findings highlight the sensitivity of TT-TSS-seq in capturing transcription initiation events with high accuracy. We propose that TT-TSS-seq is a powerful and reliable method for mapping transcription initiation sites across the mammalian genome, providing valuable insights into gene regulation and transcriptome dynamics.

How does TT-TSS-seq compare to other approaches? We did not identify publicly available data sets from related methods generated under comparable conditions and in the same cell line for direct comparison. Nonetheless, TT-TSS-seq can be evaluated against other methodologies based on two critical variables: how transcription by RNA polymerase is captured and how the TSS-seq library itself is constructed. In essence, TT-seq and its variant TT_{chem}-seq capture active transcription by RNA polymerase but do not provide information on transcriptional pausing (Wissink et al. 2019). NET-CAGE is another approach to measure TSS from nascent RNAs (Hirabayashi et al. 2019). Here, RNA is isolated from the chromatin fraction, and the CAGE approach for TSS detection is used. Thus, the NET-CAGE approach captures capped RNAs irrespective of whether they originate from active transcription events or from regulatory RNAs retained at their genomic loci.

Therefore, it does not exclusively capture transcription per se. Another method, POINT-5-seq, combines antibody-based purification of RNA polymerases with template switching for TSS library preparation (Sousa-Luís et al. 2021). It detects TSSs associated with stably bound RNA polymerases, including paused polymerase complexes. GRO-cap enables the mapping of TSSs from both stable and unstable RNAs (Core et al. 2014). It requires the isolation of nuclei, which can significantly perturb the cells. Thus, each method relies on a distinct combination of TSS detection and nascent RNA capture, which complicates side-by-side comparisons.

A potential challenge of TT-TSS-seq is the amount of RNA input required. Like TT-seq and TT_{chem}-seq, the yield of isolated 4sU+ RNA is $<1\%$, of which a fraction contains capped mRNAs. Hence, it is important to consider this when designing experiments with TT-TSS-seq.

We estimate that the protocol takes up to 5 days (depending on the planning and pause steps) from RNA extraction to library preparation, which is in the same range as the CAGE protocol (Takahashi et al. 2012). The costs are estimated at about 100 USD per sample for RNA isolation and library preparation, which is below the costs of library preparation for commercial kits. Further optimization and automated steps may reduce the timing and costs.

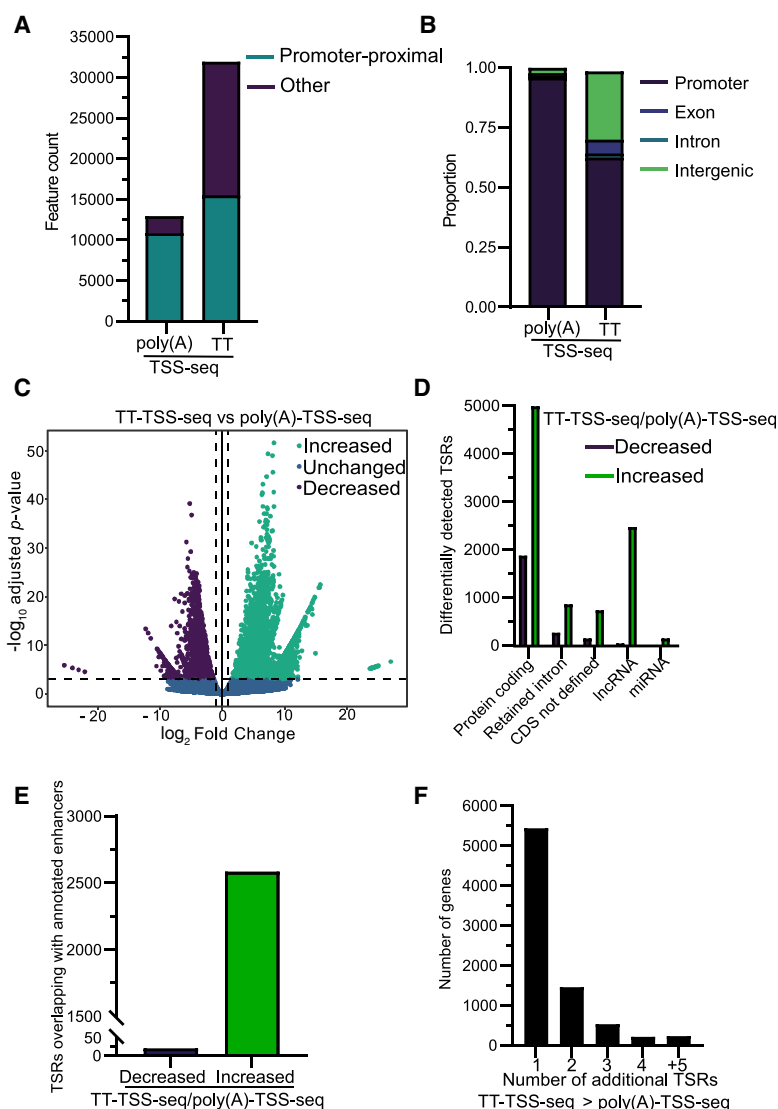


Figure 5. Comparing tags identified with poly(A)-TSS-seq and TT-TSS-seq. (A) Bar plot showing the number of transcripts with promoter-proximal TSSs. Normalized tags within 25 bp, and to a maximum distance of 250 bp, were merged to identify TSSs. (B) Bar graph showing the distribution of detected TSSs among genomic regions; -500 to $+500$ bp from annotated TSSs was used to define the promoter-proximal region. UMI-based PCR duplicate removal was carried out during TSS-sequencing analysis. Tag counts were normalized using the DESeq2 median-of-ratios approach using a threshold of 3 counts. (C) Volcano plot showing the TSSs detected at significantly different levels when using nascent compared to steady-state RNA. Normalized TSSs, located within 25 bp and to a maximum distance of 250 bp, were grouped into TSSs, and differential analysis was carried out (using DESeq2 with a false discovery threshold of 0.001 and a fold-change threshold of 2). (D) Number of TSSs, differentially detected between TT-TSS-seq versus poly(A)-TSS-seq associated with different transcript types (as annotated by Ensembl). “Protein-coding transcripts” harbor ORFs, “CDS not defined” are alternatively spliced isoforms of protein-coding transcripts with no identified ORF, and “retained intron transcripts” are alternatively spliced isoforms that harbor an intron and are predicted to be noncoding. (E) Number of TSSs, differentially detected between TT-TSS-seq and poly(A)-TSS-seq that overlap with enhancer regions (as annotated by FANTOM5) (Noguchi et al. 2017). (F) The number of genes with additional TSSs detected with a significantly increased signal by TT-TSS-seq compared to poly(A)-TSS-seq. Genes with a single TSS detected in the poly(A)-TSS-seq, which did not show a significant difference in signal or exhibited an increase compared to TT-TSS-seq, were used for the analysis. The number of additional TSSs detected at significantly higher levels by TT-TSS-seq is displayed.

Because TT-TSS-seq is performed in unperturbed cells and enables the capture of nascent RNA’s TSSs at single-nucleotide resolution, we consider it a powerful approach for investigating transcription initiation that is distinct from other approaches currently available.

Methods

Yeast growth conditions

S. cerevisiae BY4741 strain background (derived from S288C) were grown in liquid cultures (YPD [1.0% {w/v} yeast extract, 2.0% {w/v} peptone, 2.0% {w/v} glucose, and supplemented with uracil {2.4 mg/L} and adenine {1.2 mg/L}]) in an incubator/shaker (30°C, 300 rpm) until the exponential phase.

mESC culturing

mESCs (HM1) were grown in 2i/LIF (49% DMEM/F12–/–, 49% Neurobasal, 1× N2, 1× B27, 0.05 mM 2-mercaptoethanol, 2 mM L-glutamine [all of the previous from Gibco], recombinant mouse 50,000 units LIF [ESGRO], 1 μ M PD03259010 [Stemgent], 3 μ M CHIR 99021 [TOCRIS], and 1% penicillin/streptomycin). Cell culture surfaces were coated with 0.15% gelatin in PBS, for a minimum of 10 min at 37°C for mESC culture. mESCs were maintained at 37°C, 5% CO₂ and confirmed to be negative for mycoplasma. mESCs were passed every 2–3 days, depending on colony size, with media change occurring in the intermediate days. To pass cells, one wash with PBS was performed before incubating the cells with Accutase (Gibco) at 37°C for exactly 3 min.

RNA extraction

RNA was extracted from yeast as previously detailed (Chia et al. 2021). For RNA extraction from yeast grown in YPD, 24 optical density (OD) units for TSS-seq were collected by centrifugation and snap-frozen in liquid nitrogen. Per 10 OD units, RNA was extracted with 1 mL Tris-EDTA-SDS (TES) buffer (10 mM Tris-HCl [pH 7.5], 10 mM EDTA, 0.5% SDS) and 1 mL acid phenol:chloroform:isoamyl alcohol (125:24:1, Ambion), by shaking at 65°C for 45 min at 1400 rpm. After centrifugation at 4°C, max speed, 10 min, the aqueous phase was transferred to cold ethanol with 0.3 M sodium acetate. Precipitation was carried out overnight at 4°C, before centrifugation at 4°C, max speed, 20 min. The pellet was washed once with 80% v/v ethanol, dried, and reconstituted in DEPC-treated sterile water. rDNase (Machery-Nagel) treatment was carried out for 20 min at 37°C, before spin column purification (Machery-Nagel).

RNA was extracted from mammalian cells using TRIzol (Thermo Fisher Scientific). For TRIzol RNA extraction, 1 mL of TRIzol was added per 10-cm dish. Chloroform (0.2 volume, Thermo Fisher Scientific) was added, and the mixture was shaken

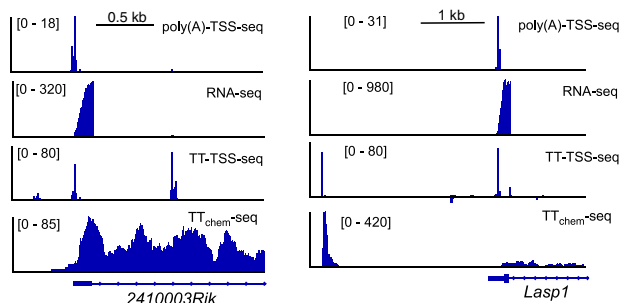


Figure 6. Example loci comparing TT-TSS-seq and poly(A)-TSS-seq. IGV showing example loci of TT-TSS-seq and poly(A)-TSS-seq in mESCs. Also included are TT_{chem}-seq and RNA-seq for the same cell line and the same growth conditions. Scale is indicated on the top.

for 30 sec before centrifugation was performed. The aqueous phase was mixed with 1.1 volumes of isopropanol (Thermo Fisher Scientific), and RNA was precipitated overnight at 4°C. After centrifugation, the pellet was washed with 85% (v/v) ethanol, dried, and reconstituted in DEPC-treated sterile water. rDNase (Qiagen) digestion was carried out at 37°C for 1 h before purification by spin column (Qiagen) or by phenol:chloroform extraction and ethanol precipitation.

Small RNA size selection for scaRNA-seq was achieved using Monarch RNA Clean up with a 1:3 ratio of sample: ethanol volume.

4sU RNA labeling and assessment

4sU labeling was carried out as detailed previously (Gregersen et al. 2020), with a few modifications. Cell culture media was removed, filtered, and 1 mM 4sU (Glentham Life Sciences) was added. The cells were returned to the incubator and labeled with 4sU for 15 min before medium aspiration and TRIzol addition, then 1/5 volume of chloroform was added, mixed, and then centrifuged at 12,000g for 15 min at 4°C. A 1.1 volume of isopropanol and 1 µL of GlycoBlue were mixed with the upper aqueous layer, and precipitation then occurred overnight at −20°C. The pellet was then washed once with 85% ethanol, dried, and resuspended.

RNA integrity was checked for a selection of samples using Bioanalyzer (Agilent Technologies). 4sU incorporation was assessed using slot blot. Here, 5 µg of total RNA was mixed with 3 µL biotin buffer (833 mM Tris-HCl [pH 7.4], and 83.3 mM EDTA) and 50 µL of 0.1 mg/mL MTSEA biotin-XX linker (Biotium, dissolved in dimethylformamide; Sigma-Aldrich) to a total volume of 200 µL, and incubated in the dark for 30 min. RNA clean-up was performed by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 [vol/vol/vol]), mixing, and centrifuging at 12,000g for 5 min at 4°C. The aqueous phase was mixed with 1.1 volumes of isopropanol, 1/10 volume of 5 M NaCl, and 1 µL of GlycoBlue and precipitated at −20°C for at least 2 h. The pellet was washed once with 85% ethanol, dried, and resuspended to 10 µL.

Samples were dropped onto the Hybond-N-membrane in the slot blot apparatus before UV crosslinking was performed using 0.2 J/cm² (254 nm) in a Stratalinker. The membrane was blocked in blocking buffer (10% [w/v] SDS, 1 mM EDTA, PBS) for 20 min, before being incubated with HRP-conjugated streptavidin (1:50,000 dilution of 1 mg/mL) in blocking buffer for 15 min at room temperature. The membrane was then washed two times in blocking buffer, two times in wash buffer 1 (1% [w/v] SDS, PBS), and two times in wash buffer 2 (0.1% [w/v] SDS, PBS) for 10 min for each wash. Streptavidin-HRP signal was visualized using Enhanced

Chemiluminescence (ECL) reagent. RNA levels were assessed by staining with 0.5 M sodium acetate and 0.5% w/v methylene blue for 10 min, before multiple washes with water.

TSS-seq library preparation

The optimized TSS-seq protocol was adapted from previously described protocols (Arribere and Gilbert 2013; Pelechano et al. 2013; Malabat et al. 2015; Mahat et al. 2016; Chia et al. 2021). The library preparation steps were adapted, as described, for the iCLIP2 protocol (Buchbender et al. 2020). DNase digestion was performed on total RNA (DNase I, Qiagen) for 1 h at 37°C. Following DNase treatment, RNA was purified using phenol:chloroform/isoamyl alcohol and precipitated overnight at −20°C using 1.1 volumes of isopropanol, 1/10 volume of 5 M NaCl and 1 µL of GlycoBlue. At least 69 µg total RNA were taken per sample.

Total RNA was dephosphorylated with quickCIP (NEB, 1.2 U/µg RNA) for 2 h at 37°C with RNasin Plus before heat inactivation of the enzyme at 80°C for 2 min. RNA was extracted using phenol:chloroform/isoamyl alcohol as before. RNA was treated with mRNA decapping enzyme (NEB, 0.7 U/µg RNA) for 2 h at 37°C with RNasin Plus. For the negative control, no decapping enzyme was included. RNA was purified using phenol:chloroform/isoamyl alcohol as before. RNA and the 5' adaptor (10 µM) were denatured for 2 min at 70°C, then cooled on ice for 2 min. Ligation was performed for 2 h at 25°C, then 16 h at 16°C using T4 RNA ligase I (30 U) with RNasin Plus. RNeasy XP (1.8× ratio, Beckman Coulter) was used to remove unligated adaptors, according to the manufacturer's instructions. The concentration of the samples was measured using the Qubit RNA BR assay, and samples were multiplexed. Poly(A)+ RNA was enriched using Dynabeads Oligo (dT)25 (Thermo Fisher Scientific), or 4sU+ RNA was purified as detailed in the TT_{chem}-seq protocol (Gregersen et al. 2020). Briefly, RNA was biotinylated as above, and reconstituted to 50 µL. Two hundred microliters of µMACS streptavidin MicroBeads (Milenyi) were added and mixed with the biotinylated RNA on a rotating wheel for 20 min at room temperature. µ Columns (Milenyi) were equilibrated with nucleic acid equilibration buffer before the samples were applied, and the flow-through non-4sU labeled RNA was collected. Columns were washed twice with 55°C wash buffer (100 mM Tris-HCl [pH 7.4], 10 mM EDTA, 1 M NaCl, and 0.1% v/v Tween 20) before 4sU+ RNA was eluted by the addition of 100 mM DTT (Sigma-Aldrich). Flow-through and eluted fractions were cleaned up using RNeasy MinElute kits (Qiagen).

Yeast RNA was fragmented for 3 min 15 sec and mammalian RNA for 5 min 15 sec at 70°C with alkaline fragmentation reagent (Ambion), to achieve 200- to 300-bp fragments. RNeasy MinElute Clean-up columns, with 1.5× volume of ethanol, were used to purify the samples. The 3' ends were fixed by quickCIP treatment (NEB, 30 U), with RNasin Plus, at 37°C for 1 h. Heat inactivation was carried out at 80°C for 2 min before the RNA was extracted using phenol:chloroform/isoamyl alcohol as before. RNA, PEG8000 (10%), and pre-adenylated 3' adaptor (2 µM) were denatured for 2 min at 70°C, then cooled on ice for 2 min. Ligation was carried out by T4 Rnl2tr (NEB, 200 U) for 2 h at 25°C, then 16 h at 16°C. RNeasy XP (1.8× ratio) was used to remove unligated adaptors. RNA, RT oligo (0.5 pmol), and dNTP (10 mM each) were denatured for 5 min at 65°C, then cooled on ice for 1 min. Reverse transcription was carried out by SuperScript III (2 µL, Invitrogen), with RNasin Plus with the following conditions: 25°C for 5 min, 42°C for 20 min, 50°C for 40 min, 80°C for 5 min. Template RNA was removed by RNase H (10 U) at 37°C for 30 min. AMPure XP beads (Agencourt, Beckman Coulter) were used to clean up the cDNA, and 3× beads and 1.7× isopropanol were added to the reverse

transcription reaction and incubated for 5 min. Beads were washed twice with 85% ethanol, dried, and eluted twice in nuclease-free water. cDNA preamplification was performed with i5_s and i7_s (300 nM each) primers and Phusion HF PCR Mastermix (Thermo Fisher Scientific). Six PCR cycles were carried out (98°C for 10 sec, 65°C for 30 sec, 72°C for 30 sec, with 3 min at 72°C final extension). ProNex beads (Promega) with a 1:2.95 ratio to sample were used to remove sequences less than 55 nt (including primer dimers). Final amplification was carried out with NEBNext i50 and i70 primers (500 nM each) and Phusion HF PCR Mastermix, for eight cycles. Amplification was checked by using Novex 6% TBE (Tris/Borate/EDTA) gel and staining with SYBR Green. ProNex beads (Promega) with a 1:2.3 ratio to sample were used for purification. Library concentration was measured using Agilent High Sensitivity D1000 ScreenTape. Libraries were sequenced with Illumina NovaSeq 6000 using 100-nt paired-end reads, with 50 million reads minimum per sample. A detailed step-by-step protocol is described in [Supplemental Document S1](#).

RNA-seq and TT_{chem}-seq

For RNA-seq, libraries were prepared using NEB Ultra II, with poly(A)⁺ purification, according to the manufacturer's protocol. Sequencing was performed on NovaSeq 6000 (Illumina) using 100-bp paired-end reads, to 60 million reads.

For TT_{chem}-seq, at least 10 ng of eluted RNA was used for library preparation with NEBNext Ultra II Directional Poly(A) mRNA, according to the manufacturer's protocol for fragmented RNA. Libraries were sequenced using NovaSeq 6000 (Illumina) as 100-bp paired-end reads, to 50 million reads.

Mapping of TSS-seq reads

Read demultiplexing was performed using Ultrplex with parameters “--phredquality 15 --min_length 0” (Wilkins et al. 2021). Adaptor trimming was performed using Cutadapt with parameters “--minimum-length 20” (Martin 2011). Bowtie 2 was used for pre-mapping to ribosomal and small RNAs (Langmead and Salzberg 2012). Genome mapping was performed using STAR, for mouse samples with Ensembl GRCm38 (mm10) release-89 annotation and for S288C samples with *S. cerevisiae* Ensembl R64-1-1 release-90 annotation (Dobin et al. 2013). UMI-tools was used for deduplication (Smith et al. 2017). BEDTools was used to identify the 5'-most nucleotide (tag) and for genome-wide comparative analyses (Quinlan and Hall 2010).

Data analysis TSS-seq

The majority of downstream analysis was performed using TSRExplorer and CAGER (Haberle et al. 2015; Policastro et al. 2021). The promoter-proximal region was defined as −300 to +100 bp from annotated yeast start codons (based on *S. cerevisiae* Ensembl R64-1-1 release-90 annotation for S288C), or −500 to +500 bp from annotated mouse TSSs (based on GRCm38 [mm10] release-89 annotation).

For the TSRExplorer analysis, a 5' tag threshold of 3 was used. DESeq2 median-of-ratios approach was used for normalization. For mammalian TSS analysis, the 5'-most tags were grouped into transcript start regions using a maximum distance of 25 bp and a maximum total width of 250 bp. The Pearson correlation coefficient between samples was calculated using the normalized counts of each TSR. DESeq2 was used to identify differentially expressed TSRs, using $\log_2(\text{fold-change}) \geq 1$, and false discovery rate (FDR) <0.001 (Love et al. 2014). Transcript biotype annotation from Ensembl release 112 was used (Harrison et al. 2024). Overlap between identified TSRs and CAGE-annotated enhancer regions

was carried out using the GRanges package (Noguchi et al. 2017). TSRs were associated with the closest gene on the same strand, prioritizing the promoter annotation, and the number of genes with multiple associated TSRs was calculated.

The yeast TSS annotation ($n = 6646$) was previously described (Park et al. 2014). For the heat maps, the mean depth in the ± 1 -kb area of each TSS was split into 10-bp bins for which the “vertical” mean of the bins was calculated. Sense read-depth profiles of transcript promoter regions ($\text{TSS} \pm 1\text{kb}$) were created using deepTools (Ramírez et al. 2016). Mouse TSSs were defined using a subset of protein-coding transcripts from standard chromosomes, Ensembl release 89 (GRCm38).

For Supplemental Figure S3, a union of all TSRs from CAGE and poly(A)-TSS samples was created and assessed for overlap between TSRs of individual samples. Overlaps were limited to those shared between all data sets and those seen privately in both replicates of a given data set.

RNA-seq and TT_{chem}-seq data analysis

RNA-seq and TT_{chem}-seq analysis read processing was done using nf-core/rnaseq (3.13.2) (Ewels et al. 2020). Adaptor and quality trimming were performed using Trim Galore! (<https://github.com/FelixKrueger/TrimGalore>) (minimum trimmed read threshold: 10,000). Genome mapping was performed using STAR with Ensembl GRCm38 (mm10) release-89 annotation (minimum uniquely mapped read threshold: 5%) (Dobin et al. 2013). Transcript assembly was performed using StringTie (Pertea et al. 2015).

Code used for bioinformatic analysis

We used TSRExplorer and deepTools for the data analysis (Ramírez et al. 2016; Policastro et al. 2021). Scripts for generating some of the figure panels are available in [Supplemental Material as Supplemental Code](#).

Statistical analysis

Information regarding any statistical tests used, number of samples, or number of biological replicate experiments is stated in the corresponding figure legends.

Publicly available data sets used in this study

CAGE data were obtained from Noguchi et al. (2017) and Lloret-Llinares et al. (2018).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE292786.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the Crick Genomics STP for sequencing the libraries. We also thank the members of the van Werven lab for the critical reading of the manuscript. This work was supported by the Francis Crick Institute (CC2043), which receives its core funding from

Cancer Research UK (CC2043), the UK Medical Research Council (CC2043), and the Wellcome Trust (CC2043).

Author contributions: F.J.v.W., E.E.H., and C.V. conceived the project. E.E.H. and C.V. performed experiments. E.E.H., C.V., and R.M. analyzed data. V.A. rewrote Supplemental Document S1. F.J.v.W. and E.E.H. wrote the manuscript with help from the other authors.

References

- Arribere JA, Gilbert WV. 2013. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res* **23**: 977–987. doi:10.1101/gr.150342.112
- Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, et al. 2014. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**: 393–399. doi:10.1038/nature12962
- Buchbender A, Mutter H, Sutandy FXR, Körtel N, Hänel H, Busch A, Ebersberger S, König J. 2020. Improved library preparation with the new iCLIP2 protocol. *Methods* **178**: 33–48. doi:10.1016/j.ymeth.2019.10.003
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engström PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–635. doi:10.1038/ng1789
- Chia M, Li C, Marques S, Pelechano V, Luscombe NM, van Werven FJ. 2021. High-resolution analysis of cell-state transitions in yeast suggests widespread transcriptional tuning by alternative starts. *Genome Biol* **22**: 34. doi:10.1186/s13059-020-02245-3
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320. doi:10.1038/ng.3142
- Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, Siepel A. 2015. Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12**: 433–438. doi:10.1038/nmeth.3329
- Demircioglu D, Cukuroglu E, Kundermans M, Nandi T, Calabrese C, Fonseca NA, Kahles A, Lehmann KV, Stegle O, Brazma A, et al. 2019. A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* **178**: 1465–1477.e17. doi:10.1016/j.cell.2019.08.018
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Elgood Hunt E, Vivori C, Albihlal WS, van Werven FJ. 2026. Topoisomerases regulate alternative transcription start site selection in yeast. *G3 (Bethesda)* doi:10.1093/g3journal/jkaf318
- Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**: 276–278. doi:10.1038/s41587-020-0439-x
- Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. 2008. A code for transcription initiation in mammalian genomes. *Genome Res* **18**: 1–12. doi:10.1101/gr.6831208
- Gregersen LH, Mitter R, Svejstrup JQ. 2020. Using TT_{chem}-seq for profiling nascent transcription and measuring transcript elongation. *Nat Protoc* **15**: 604–627. doi:10.1038/s41596-019-0262-3
- Haberle V, Forrest AR, Hayashizaki Y, Carninci P, Lenhard B. 2015. CAGER: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* **43**: e51. doi:10.1093/nar/gkv054
- Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Becker A, Bennett R, Berry A, Bhaj J, et al. 2024. Ensembl 2024. *Nucleic Acids Res* **52**: D891–D899. doi:10.1093/nar/gkad1049
- Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC, Adelman K. 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**: 26–41. doi:10.1101/gad.309351.117
- Hirabayashi S, Bhagat S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, Itoh M, Shirakawa K, Takaori-Kondo A, Takeuchi O, et al. 2019. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat Genet* **51**: 1369–1379. doi:10.1038/s41588-019-0485-9
- Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. 2013. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife* **2**: e00808. doi:10.7554/eLife.00808
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923
- Larke MSC, Schwesinger R, Nojima T, Telenius J, Beagrie RA, Downes DJ, Oudelaar AM, Truch J, Graham B, Bender MA, et al. 2021. Enhancers predominantly regulate gene expression during differentiation via transcription initiation. *Mol Cell* **81**: 983–997.e7. doi:10.1016/j.molcel.2021.01.002
- Lloret-Llinares M, Karadoulama E, Chen Y, Wojenski LA, Villafano GJ, Bornholdt J, Andersson R, Core L, Sandelin A, Jensen TH. 2018. The RNA exosome contributes to gene expression regulation during stem cell differentiation. *Nucleic Acids Res* **46**: 11502–11513. doi:10.1093/nar/gky817
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Lu Z, Lin Z. 2019. Pervasive and dynamic transcription initiation in *Saccharomyces cerevisiae*. *Genome Res* **29**: 1198–1210. doi:10.1101/gr.245456.118
- Luse DS, Parida M, Spector BM, Nilsson KA, Price DH. 2020. A unified view of the sequence and functional organization of the human RNA polymerase II promoter. *Nucleic Acids Res* **48**: 7767–7785. doi:10.1093/nar/gkaa531
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**: 1455–1476. doi:10.1038/nprot.2016.086
- Malabat C, Feuerbach F, Ma L, Saveanu C, Jacquier A. 2015. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* **4**: e06722. doi:10.7554/eLife.06722
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10. doi:10.14806/ej.17.1.200
- Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**: 335–338. doi:10.1126/science.1181421
- Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, Ishikawa-Kato S, Kaide K, Kaiho A, Kanamori-Katayama M, et al. 2017. FANTOM5 CAGE profiles of human and mouse samples. *Sci Data* **4**: 170112. doi:10.1038/sdata.2017.112
- Park D, Morris AR, Battenhouse A, Iyer VR. 2014. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res* **42**: 3736–3749. doi:10.1093/nar/gkt1366
- Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**: 127–131. doi:10.1038/nature12121
- Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290–295. doi:10.1038/nbt.3122
- Policastro RA, Raborn RT, Brendel VP, Zentner GE. 2020. Simple and efficient profiling of transcription initiation and transcript levels with STRIP-seq. *Genome Res* **30**: 910–923. doi:10.1101/gr.261545.120
- Policastro RA, McDonald DJ, Brendel VP, Zentner GE. 2021. Flexible analysis of TSS mapping data and detection of TSS shifts with TSRExploreR. *NAR Genom Bioinform* **3**: lqab051. doi:10.1093/nargab/lqab051
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Ramírez F, Ryan DP, Grünig B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165. doi:10.1093/nar/gkw257
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. doi:10.1038/nbt.1754
- Sartorelli V, Laubert SM. 2020. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**: 521–528. doi:10.1038/s41594-020-0446-0
- Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016. TT-seq maps the human transient transcriptome. *Science* **352**: 1225–1228. doi:10.1126/science.aad9841
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499. doi:10.1101/gr.209601.116
- Sousa-Luis R, Dujardin G, Zukher I, Kimura H, Weldon C, Carmo-Fonseca M, Proudfoot NJ, Nojima T. 2021. POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol Cell* **81**: 1935–1950.e6. doi:10.1016/j.molcel.2021.02.034
- Takahashi H, Kato S, Murata M, Carninci P. 2012. CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. *Methods Mol Biol* **786**: 181–200. doi:10.1007/978-1-61779-292-2_11

- Thorsen K, Schepeler T, Øster B, Rasmussen MH, Vang S, Wang K, Hansen KQ, Lamy P, Pedersen JS, Eller A, et al. 2011. Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. *BMC Genomics* **12**: 505. doi:10.1186/1471-2164-12-505
- Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SHC, Kadonaga JT. 2017. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* **31**: 6–11. doi:10.1101/gad.293837.116
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476. doi:10.1038/nature07509
- Wilkins OG, Capitanchik C, Luscombe NM, Ule J. 2021. Ultraplex: a rapid, flexible, all-in-one fastq demultiplexer. *Wellcome Open Res* **6**: 141. doi:10.12688/wellcomeopenres.16791.1
- Wissink EM, Vihervaara A, Tipples ND, Lis JT. 2019. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **20**: 705–723. doi:10.1038/s41576-019-0159-6
- Zhang Z, Dietrich FS. 2005. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res* **33**: 2838–2851. doi:10.1093/nar/gki583

Received April 1, 2025; accepted in revised form December 17, 2025.