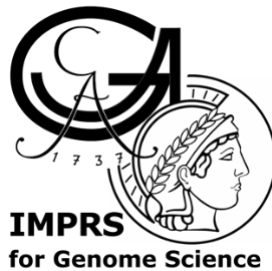


Bioinformatics analysis of multi-omics data elucidates U2 snRNP function in transcription

Dissertation

For the award of the degree
“Doctor of Philosophy” (Ph.D.)
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

Within the doctoral program



International Max Planck Research School for Genome Science
of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences
(GGNB)

submitted by

Sara Patrícia Monteiro Martins

from Porto, Portugal

Göttingen 2021

Thesis Committee

Prof. Dr. Patrick Cramer

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Department of Molecular Biology, Göttingen, Germany

Prof. Dr. Matthias Dobbelstein

University Medical Center Göttingen, Institute of Molecular Oncology, Göttingen, Germany

Dr. Juliane Liepe

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Research Group Quantitative and Systems Biology, Göttingen, Germany

Members of the Examination Board

Referee: Prof. Dr. Patrick Cramer

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Department of Molecular Biology, Göttingen, Germany

2nd Referee: Prof. Dr. Matthias Dobbelstein

University Medical Center Göttingen, Institute of Molecular Oncology, Göttingen, Germany

Further members of the Examination Board

Dr. Juliane Liepe

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Research Group Quantitative and Systems Biology, Göttingen, Germany

Prof. Dr. Tim Beißbarth

University Medical Center Göttingen, Department of Medical Bioinformatics, Göttingen, Germany

Prof. Dr. Reinhard Lührmann

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Department of Cellular Biochemistry, Göttingen, Germany

Dr. Marieke Oudelaar

Max Planck Institute for Biophysical Chemistry (MPI-bpc), Research Group Genome Organization and Regulation, Göttingen, Germany

Date of the oral examination: July 6, 2021

Acknowledgments

Throughout my Ph.D. studies, I have received a great deal of support and assistance.

I would first like to thank my supervisor, Prof. Dr. Patrick Cramer for the opportunity to work on such a challenging and interesting topic. A special thanks to his supervision and guidance throughout my Ph.D. studies. I learned a lot during these almost four years and has been an incredible experience to work in such a reputable laboratory.

I would also like to thank my thesis advisory committee members Prof. Dr. Matthias Dobbstein and Dr. Juliane Liepe for valuable comments and suggestions throughout my Ph.D. studies.

A special thanks to Prof. Dr. Tim Beißbarth, Prof. Dr. Reinhard Lührmann and Dr. Marieke Oudelaar for their interest in my work and for joining my thesis examination board.

I am grateful to all the International Max Planck Research School (IMPRS) for Genome Science community. A special thanks to the program coordinators Dr. Henriette Irmer and Frauke Bergmann and the former coordinator Dr. Katja Lidschreiber.

The present work was developed in collaboration with an excellent experimentalist from the lab: Dr. Livia Caizzi. She introduced me to the field of co-transcriptional splicing and help me grow as a researcher. Was a pleasure working with her, both regarding her professional and personal qualities.

I owe a big thanks to both Dr. Björn Schwalb and Dr. Michael Lidschreiber for all the help, supervision, and great discussions throughout my work. Furthermore, thanks to Dr. Björn Schwalb for introducing me to the bioinformatics analysis of transcriptomics data; and my gratitude to Dr. Michael Lidschreiber for reviewing parts of this document.

I would like to thank all my colleagues from Cramer's lab for the nice and stimulating working environment. A special thanks to Dr. Saskia Gressel that introduced me to the lab and the transcriptomics field. My gratitude to the bioinformatician Ph.D. candidates: Le Xiong, Eusra Mohammad, Gabriel Villamil, and Arjun Devadas for all the interesting discussions and mutual help. Additionally, a special thanks to Dr. Kristina Zumer, Dr. Anna Sawicka, and Kseniia Lysakovskaia for interesting discussions, support, and feedback during this work. A further big thanks to Kerstin Maier and Petra Rus for their great support on data sequencing; and Janine Blümel, Almuth Burgdorf and Kirsten Backs for keeping the lab running.

Thanks to my Ph.D. colleagues with whom I had the pleasure to organize our IMPRS for Genome Science 2018 retreat: Wanwan Ge, Christian Roth, Gesa Werner, and Nadine Übelmesser. Likewise, I am also thankful for the help provided by my colleagues while organizing the 2019 IMPRS for Genome Science Halloween party: Spiros Palikyras, Neil Singh, Ruoshi Zhang, Shu Zhang, and Sarah Henze; and Dr. Katja Lidschreiber which helped us keep on track during that adventure.

My gratitude to the Ph.D. and Postdoc MPI community and all the co-organizers of Friday's happy hour. It proportionated a lovely end of the week both as an organizer or as a guest.

A special thanks to my family, my friends, and my boyfriend for always being there for me and for the necessary distractions to rest my mind outside of my research. To my mother Fernanda Monteiro and my father José Rosas, my most sincere gratitude for believing in me and my dreams and always give me the strength to walk my path. Muito obrigada.

Summary

Transcription by RNA **polymerase II** (Pol II) is an important step in cell function and regulation. Pol II transcription has been shown to be coupled to pre-mRNA splicing, but the underlying mechanisms remain poorly understood. Co-transcriptional splicing requires the assembly of a functional spliceosome on nascent pre-mRNA, but whether and how this influences Pol II transcription remains unclear.

To investigate this, we used a human erythroleukemic cell line and performed **transient transcriptome sequencing** (TT-seq) and **mammalian native elongating transcript sequencing** (mNET-seq) upon fast inhibition of U2 snRNP function. We further studied how the **positive transcription elongation factor b** (P-TEFb) recruitment is related to the Pol II pause duration, using **chromatin immunoprecipitation and sequencing** (ChIP-seq) of the PTEF-b kinase **cyclin T1** (CycT1) upon U2 snRNP inhibition. I performed a bioinformatics analysis of the different datasets generated for this study and two additional published datasets. I also conducted a multi-omics analysis combining TT-seq and mNET-seq data to calculate and quantify transcription kinetic parameters such as Pol II productive initiation frequency, pause duration and elongation velocity. Here we show that inhibition of pre-mRNA branch site recognition by the spliceosome component U2 snRNP leads to a widespread and strong decrease in new RNA synthesis from human genes. We further show that inhibition of U2 snRNP function increases the duration of Pol II pausing in the promoter-proximal region, impairs recruitment of the pause release factor P-TEFb, and reduces Pol II elongation velocity at the beginning of genes. Our results indicate that efficient release of paused Pol II into active transcription elongation requires the formation of functional spliceosomes and that eukaryotic mRNA biogenesis relies on positive feedback from the splicing machinery to the transcription machinery. We further show that the fast U2 snRNP inhibition affects the expression of genes related to RNA synthesis and it is not related to stress response genes. Our new multi-omics approach for the calculation of Pol II elongation velocity can be applied to further study the impact on Pol II kinetics regarding different splicing and transcription factors. This is of great importance to unravel the mechanisms behind Pol II transcription and splicing and understand how the disruption of this regulation leads to several pathological cell phenotypes and diseases.

Table of contents

Acknowledgments.....	ii
Summary	iii
I. INTRODUCTION	
1. Transcription regulation and splicing	1
2. Motivation and aims of this thesis	2
3. Publication	4
4. Dissertation organization	4
II. STATE OF THE ART	
1. The transcription cycle.....	5
1.1. Transcription initiation	5
1.2. Promoter-proximal pausing	6
1.3. Productive elongation	7
1.4. RNA transcription termination	7
1.5. Pol II RPB1 CTD.....	9
1.6. Histone modifications and transcription regulation.....	9
2. Co-transcriptional processing	10
3. Transcription and splicing	11
3.1. The splicing cycle.....	11
3.2. Co-transcriptional splicing	13
4. Transcriptomics	15
4.1. RNA synthesis	15
4.2. DNA-protein interactions	16
4.3. DNA-RNA interactions	16
4.4. RNA-protein interactions	17
4.5. RNA-RNA interactions	17
4.6. Pol II occupancy	18
4.7. Chromatin related techniques	18
4.8. RNA-seq data analysis.....	19
4.9. Long-read sequencing.....	20
4.10. Single-cell sequencing.....	21
4.11. Multi-omics analysis.....	21
5. Overview of key software used in this analysis.....	22
5.1. Defining major transcribed isoforms using Salmon	22
5.2. Differential expression analysis using DESeq2.....	22
5.3. Gene ontology analysis with goseq	23
III. METHODS	
1. Creating annotation files.....	26
1.1. General annotation.....	26
1.2. Major isoform annotation	26
1.3. Intronless genes annotation	27
2. General analysis.....	28
2.1. TT-seq and RNA-seq.....	28
2.2. mNET-seq.....	31
2.3. ChIP-seq	32
3. Splicing ratio.....	32
4. Identification of splicing affected and unaffected transcripts.....	33
5. Initiation frequency, pause duration, and elongation frequency calculation	35
5.1. RNA amount per cell.....	35

5.2.	Detection of Pol II pause sites	36
5.3.	Productive initiation frequency	36
5.4.	Pause duration.....	37
5.5.	Elongation velocity estimation	38
6.	Differential expression analysis.....	39
7.	Gene Ontology enrichment analysis	40
8.	Statistical analysis.....	40
9.	Alternative splicing analysis	40
IV. RESULTS		
1.	Rapid splicing inhibition in human cells	41
2.	Inhibition of U2 snRNP function decreases RNA synthesis	44
3.	Inhibition of U2 snRNP function impairs early transcription elongation.....	50
4.	Inhibition of U2 snRNP function increases Pol II pause duration.....	53
5.	Inhibition of U2 snRNP function impairs recruitment of pause release factor P-TEFb.....	55
6.	Inhibition of U2 snRNP function alters Pol II elongation velocity	57
7.	U2 snRNP inhibition with Pla-B is independent of stress response	58
V. DISCUSSION AND OUTLOOK		
1.	Open questions regarding U2 snRNP function in transcription	61
2.	Possible mechanisms under U2 snRNP requirement for efficient Pol II elongation activation	63
3.	The impact of U2 snRNP inhibition on intronless genes.....	64
4.	Study limitations	65
5.	Future perspectives	66
5.1.	Annotation	66
5.2.	Kinetic modeling	66
5.3.	U2 snRNP-dependent Pol II elongation activation mechanism	67
5.4.	R-loops impact on fast U2 snRNP inhibition	68
5.5.	Cell type-specific U2 snRNP function	68
VI. SUPPLEMENTARY INFORMATION		
1.	Experimental methods	69
1.1.	Employed cell lines	69
1.2.	Cell culture treatments.....	70
1.3.	Total and 4sU-labeled RNA extraction and semiquantitative-PCR.....	70
1.4.	TT-seq and RNA-seq.....	70
1.5.	Morpholino inhibition.....	71
1.6.	mNET-seq.....	71
1.7.	RNA spike-ins for mNET-seq.....	72
1.8.	ChIP-seq	72
2.	Sequence of the spike-ins used on TT-seq and RNA-seq experiments.....	74
3.	Supplementary Figures	77
4.	Supplementary Tables.....	86
REFERENCES		95
APPENDIX		
Listed figures.....		I
Main figures		I
Supplementary figures.....		II
Listed tables		III
Main tables		III
Supplementary tables.....		III

List of items from the manuscript	IV
Main figures	IV
Supplementary figures.....	V
Tables	V
Chapters.....	V
Abbreviations and acronyms.....	VI
Curriculum vitae	XI



INTRODUCTION

This chapter highlights the importance of studying co-transcriptional splicing, the main goals, published research manuscript details, and the dissertation organization.

1. Transcription regulation and splicing

Transcription is the first step in the biological process of transcribing the genomic information into protein. Consequently, transcription regulation is a critical process that impacts the function of the cell, the tissue, and the whole organism. Transcription regulation involves a series of different processes, including splicing (Li and Liu, 2019; Lodish et al., 2000). The complex splicing process requires the assembly of a functional spliceosome on nascent pre-mRNA. Splicing occurs in most eukaryote genes in a co-transcriptional manner. It comprises the removal of introns as the nascent RNA emerges from RNA polymerase II (Pol II) to generate a functional mature protein-coding RNA (Bentley, 2014).

Splicing is known to have a significant impact on several processes, such as gene expression (Ding and Elowitz, 2019), cell cycle (Petasny et al., 2021), and immune activity (Su and Huang, 2021). Due to its vast importance in cell regulation, splicing has been associated with several pathologies, such as rare genetic diseases and cancer (Coltri et al., 2019; Tang et al., 2021). Splicing modifiers have been developed as therapeutic targets for many human disorders to correct for the mentioned aberrant splicing or to induce a reduction of the expression level of determined genes or isoforms (Tang et al., 2021).

The study field of systems biology integrates different omics laboratory techniques. Several transcriptomics techniques were developed and can be used to study transcription regulation and splicing. The processing and visualization of this data require the use of computational resources. Thus, bioinformatics analysis is an essential pillar of the systems biology field, which comprises an interface between biological and computational sciences, to unravel the fundamental biological questions.

2. Motivation and aims of this thesis

Splicing is a critical phenomenon in regulating gene expression, and many human diseases are linked to modifications in the splicing process, such as many molecular alterations observed in cancer (Coltri et al. 2019). Therefore, it is crucial to study the mechanisms involved in this process, especially how splicing and transcription are interlinked. Regarding the up-to-date knowledge, the effect of rapid inhibition of spliceosome function on transcription has not yet been investigated.

This study aims to understand the impact of the spliceosome U2 small nuclear ribonucleoprotein (snRNP) on transcription and transcription regulation. For that, we used a K562 cell line (Lozzio and Lozzio, 1975), an immortalized chronic myelogenous leukemia cell line with many available multi-omics sequencing data available at the ENCODE project portal (Sloan et al. 2016; Zhou et al. 2019). To study how transcription and splicing are interlinked, we used three orthogonal approaches to inhibit U2 snRNP function and monitor RNA synthesis and splicing genome-wide using **RNA sequencing** (RNA-seq) and **transient transcriptome sequencing** (TT-seq). To inhibit U2 snRNP function we used two SF3B1 inhibitors, pladienolide B (Pla-B) and spliceostatin A (SSA); and an antisense morpholino oligo (AMO) against U2 snRNA.

In addition, we studied the RNA polymerase II occupancy profile using **mammalian native elongation transcript sequencing** (mNET-seq) and **chromatin immunoprecipitation and sequencing** (ChIP-seq). We investigated the **positive transcription elongation factor b** (P-TEFb) recruitment through **cyclin T1** (CycT1) ChIP-seq. We then performed transcription kinetics calculations. Productive initiation frequency was calculated using TT-seq data. A multi-omics approach integrating TT-seq and mNET-seq data was performed to calculate Pol II pause duration and elongation velocity (**Figure 1**).

I performed a bioinformatics analysis of the different datasets generated for this study and the multi-omics approach for kinetic modeling. Two published datasets were also included (**Table S1**).

In particular, the following general questions were addressed:

- A) Does splicing inhibition affects new RNA synthesis?**
- B) Is co-transcriptional splicing related to Pol II productive transcription initiation frequency and promoter-proximal pausing duration?**
- C) Does co-transcriptional splicing influence Pol II elongation velocity?**

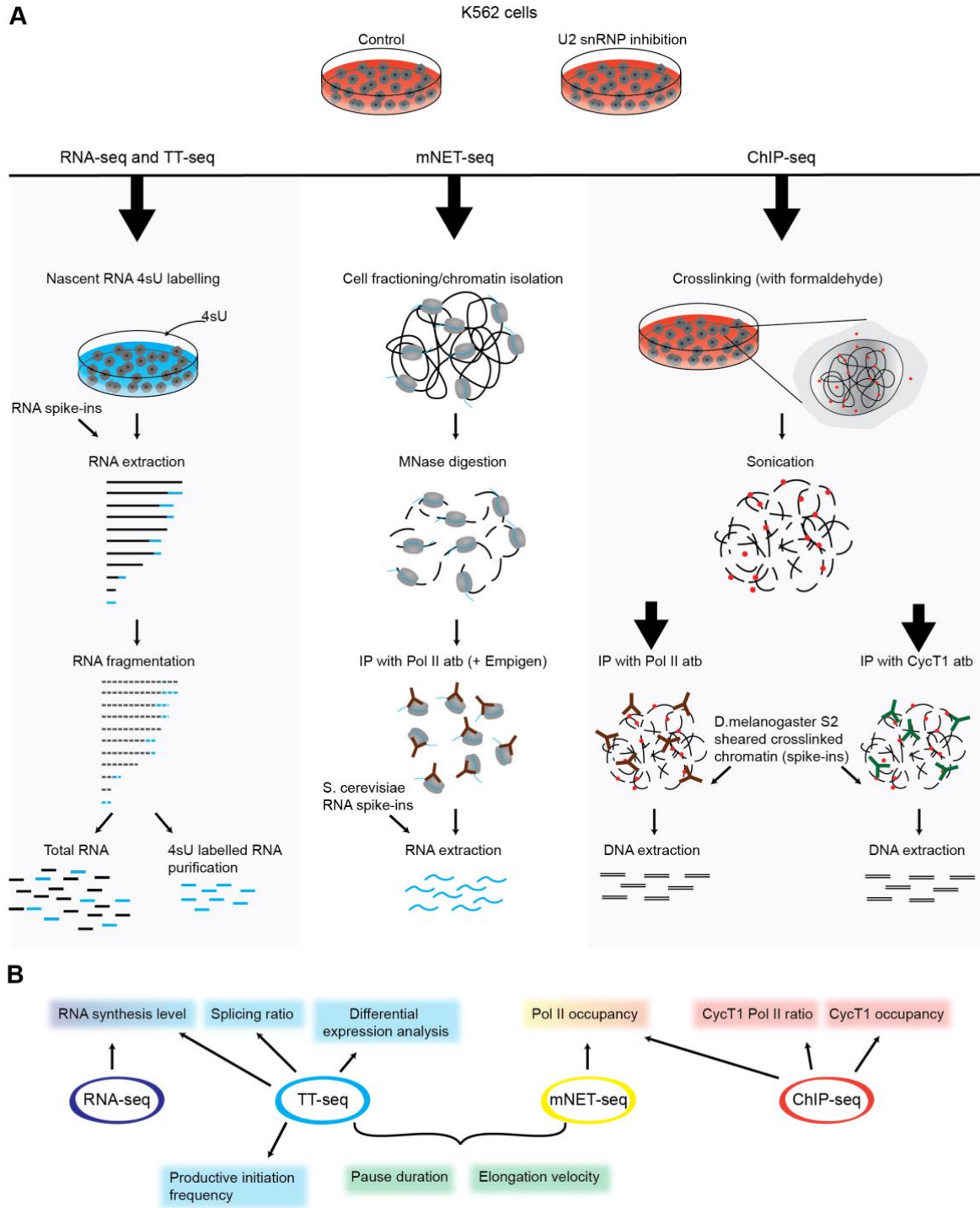


Figure 1. Study design.

(A) Overview of the experiments performed. RNA-seq, TT-seq, mNET-seq, and Pol II and CycT1 ChIP-seq were performed on K562 cells subjected to control or a U2 snRNA inhibitor treatment. (B) Overview of the bioinformatics analysis and the studied parameters with each methodology. Note that Pol II promoter-proximal pausing duration and elongation velocity calculations were obtained by a multi-omics approach combining TT-seq and mNET-seq data. atb: antibody

3. Publication

Part of this work has been published as a research manuscript:

Efficient RNA polymerase II pause release requires U2 snRNP function

Caizzi, L. *, **Monteiro-Martins, S. ***, Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., and Cramer, P. (2021). *Mol. Cell* 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

Author contributions:

Conceptualization, LC and PC. Methodology, LC and BS. Formal Analysis, SMM. Investigation, LC, KL, JS, AS and YC. Data Curation: SMM. Writing – Original Draft, LC, SMM and PC. Writing - Review & Editing, LC, SMM and PC. Visualization, SMM, LC and PC. Supervision, LC, BS, ML and PC. Project Administration, PC. Funding Acquisition, LC, AS and PC.

A detailed summary of items and chapters excerpted or adapted from the manuscript can be found in **Appendix** (see **page IV**).

4. Dissertation organization

The present document is structured in six (VI) chapters. The introductory chapter (I) highlights the importance of understanding the mechanisms behind transcription regulation and splicing and the motivation and aims of this study. The second chapter includes a literature overview for crucial topics essential to understanding the present work, such as RNA transcription cycle, splicing, co-transcriptional splicing, an overview of transcriptomics analysis methods, and a summary of critical software used in the present work.

A description of the bioinformatics methodology used can be found in the third chapter. The fourth chapter shows the main results regarding the impact of U2 snRNP inhibition. In chapter V, a brief discussion on the central questions addressed with this study and comment on the methodology limitations, and an insight into the remaining questions and future opportunities for this thematic are described. Chapter VI includes supplementary information to chapters IV and V, including a description of the experimental methodologies used, complementary figures, and tables.

II

STATE OF THE ART

I reviewed selected milestones in the transcription field to introduce the reader to the relevant information needed to understand this work. In section 1, I focus on the transcription cycle and co-transcriptional regulation. In section 2, I introduce the splicing concept and focus on co-transcriptional splicing. In section 3, I describe the most commonly used sequencing assays to study transcription and transcription regulation, including multi-omics approaches. In section 4, I summarily explain the concept of essential software used in the analysis.

1. The transcription cycle

Every cell of a living organism contains the same genetic code comprised of **deoxyribonucleic acid** (DNA) molecules. This genetic information is transcribed into **ribonucleic acid** (RNA) molecules according to the cell and tissue function. RNA molecules occur in different forms, and they can be transcribed into proteins or work in various regulatory processes in the cell (Li and Liu, 2019; Lodish et al., 2000). RNA transcription is carried out by enzymes named RNA polymerases (Pol). More than 50 years ago, three types of polymerases were discovered. Pol I synthesizes the **ribosomal RNA** (rRNA), which accounts for over 50% of the total RNA synthesized in a cell. Pol II transcribes **messenger RNAs** (mRNAs) that will be further transcribed into proteins and many **non-coding RNAs** (ncRNAs). Pol III is responsible for the synthesis of **transfer RNAs** (tRNAs), that are involved in decoding an mRNA sequence, and some other small ncRNAs (Cramer, 2019; Roeder, 2019).

The present work focuses on Pol II transcription, a highly regulated process that can be roughly divided into initiation, promoter-proximal pausing, elongation, and termination (Schier and Taatjes, 2020; Shandilya and Roberts, 2012). This whole process can take between a few seconds to several hours in human genes (Alpert et al., 2018).

1.1. Transcription initiation

The transcription initiation comprises DNA opening and the incorporation of the first nucleotide to start the mRNA chain. This first step is determined by the formation of a functional **pre-initiation complex** (PIC) which consists of eight factors: **transcription factor** (TF) IIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH, Pol II, and Mediator complex (**Figure 2A**). PIC assembles on DNA sequence regions called promoters, which serve as binding platforms for proteins to start transcription (Haberle and Stark, 2018). A critical and well-defined promoter sequence is the RNA transcription start position, denominated **transcription start site** (TSS) (Roy, 2006). Promoters can be regulated by regulatory sequences with approximately 200 **base pairs** (bp) designated enhancers, which can be located

down or upstream the promoter or as far as several thousand bp (Schaffner, 2006). A well-studied core promoter sequence is the TATA-box motif positioned around 30 bp upstream of the TSS (Haberle and Stark, 2018). In TATA-box containing genes, the first step of PIC formation is the recruitment of the TFIID complex **TATA-binding protein (TBP)** to the TATAA sequence. TBP is highly regulated both positively and negatively. TFIIA recruitment will stabilize the interaction between TBP and the DNA; TFIIB will also recognize and bind to elements flanking the TATA box. Pol II and TFIIF bind to the core promoter, followed by the Mediator complex, TFIIH, and TFIIE (**Figure 2A**). TFIIH is responsible for an ATP-dependent transition between closed and open PIC required for productive transcription initiation (Schier and Taatjes, 2020; Shandilya and Roberts, 2012). Recently, reconstitution of a human Mediator-PIC complex revealed interaction with TFIIH **cyclin-dependent kinase (CDK) 7**, which is responsible for Pol II **carboxy-terminal domain (CTD)** phosphorylation and promoter-proximal Pol II escape (addressed below on **sections 1.5 and 1.2**, respectively) (Rengachari et al., 2021). The Mediator complex CDK8 module impairs the binding to PIC, and the CDK8 kinase is responsible for facilitating Mediator dissociation from the CDK8 module and consequent binding to PIC (Osman et al., 2021).

Even though the TATA-box is an extensively studied motif, TATA-less genes are predominant in mammalian genomes, while TATA-box-containing genes are associated with a stress response. A more abundant core promoter motif in mammalian genomes is the **initiator (Int)**, which overlaps the TSS. Int is usually accompanied by a second motif located downstream the TSS, the **downstream promoter element (DPE)**. TFIID is suggested to bind DPE (Haberle and Stark, 2018; Schier and Taatjes, 2020; Shandilya and Roberts, 2012).

Another complex involved in transcription initiation is the SAGA co-activator complex. While TATA-containing promoters use more frequently the SAGA complex, TATA-less promoters are associated with TFIID recruitment. Nevertheless, SAGA and TFIID are redundantly recruited at several promoters (Haberle and Stark, 2018; Schier and Taatjes, 2020; Shandilya and Roberts, 2012).

1.2. Promoter-proximal pausing

Before continuing into productive elongation, Pol II often stalls 20-60 bp downstream of the TSS; an event defined as promoter-proximal pausing. This regulation step is suggested to work as a checkpoint before the assembly of elongation factors (Core and Adelman, 2019; Dollinger and Gilmour, 2021). Promoter-proximal Pol II pausing is mediated by the **negative elongation factor (NELF)** and the **DRB sensitivity inducing factor (DSIF)** proteins, which bind to Pol II after dissociation of TFIIB, TFIIF, and TFIIE (**Figure 2B**). NELF is recruited to the paused transcription complex by DSIF (Vos et al., 2018a). P-TEFb is part of the multi-subunit **super elongation complex (SEC)** and has a crucial role in releasing Pol II into productive elongation. P-TEFb comprises CDK9 and **cyclin T1 (CycT1)** or CycT2. CDK9 phosphorylates NELF resulting in its dissociation from the paused transcription elongation complex. DSIF subunit Spt5 is also phosphorylated by CDK9, causing a transition in DSIF function from being a pausing factor into a positive elongation factor. Besides NELF and DSIF, some other factors were found to be important for promoter-proximal pausing regulation, including TFIID, GAGA factor, and **Pol II-associated factor 1 (PAF)** complex (Dollinger and Gilmour, 2021). **Bromodomain-containing protein 4 (BRD4)** is also associated with

transcription elongation and pause release both by P-TEFb recruitment and histone acetylation (Itzen et al., 2014; Moon et al., 2005; Yang et al., 2005).

DNA and RNA sequences have been implicated in promoter-proximal pausing regulation, including the enrichment of promoter elements and pause motifs, along with GC-content. Furthermore, +1 nucleosome (see **section 1.6** for definition) positioning may also play a role in promoting Pol II pausing in humans (Dollinger and Gilmour, 2021). The **pre-exon junction complex** (pre-EJC), which assembles on RNA upstream of exon-exon boundaries (Le Hir et al., 2000), is also related to Pol II promoter pausing and splicing. The pre-EJC subunit Mago was shown to impact promoter-proximal pausing and be related to CDK9 occupancy in drosophila (Akhtar et al., 2019; Dollinger and Gilmour, 2021).

1.3. Productive elongation

The processing of Pol II into productive elongation is highly dependent on promoter-proximal pausing release and the formation of an active elongation complex (EC*) formed by Pol II, DSIF, PAF, and the transcription elongation factor SPT6 (Vos et al., 2018b). P-TEFb phosphorylates Pol II CTD, DSIF, and NELF. Upon phosphorylation, NELF is released, allowing the association of elongation factors, such as PAF and SPT6 (Dollinger and Gilmour, 2021) (**Figure 2C**). BRD4 stimulates P-TEFb to phosphorylate Pol II CTD, releasing paused Pol II into productive elongation (Itzen et al., 2014).

RTF1, a dissociable PAF complex subunit critical for chromatin transcription, stimulates Pol II elongation depending on DSIF, PAF, and P-TEFb, but not on SPT6 (Zhang et al., 2020). RTF1 and SPT6 bind opposite sides of Pol II (Vos et al., 2018b, 2020).

1.4. RNA transcription termination

The end of the transcriptional cycle is marked by Pol II dissociation from the DNA template. The most studied termination process depends on polyadenylation (poly(A)) sites. Most protein-coding genes contain a conserved poly(A)-site in eukaryotes, which includes an AAUAAA motif. The poly(A)-site is recognized by components of the cleavage and polyadenylation complex (CPA), and is cleaved by the CPSF73 endonuclease (Eaton and West, 2020; Shandilya and Roberts, 2012). Two different models have described poly(A)-site-dependent termination: allosteric/anti-terminator and torpedo models. In the allosteric/anti-terminator model, the transcription of a poly(A) will cause the dissociation of anti-termination factors, such as the SCAF4 and SCAF8 and a conformational change within Pol II that will lead to its termination. In the torpedo model, the termination is achieved after a nuclear 5' → 3' exonuclease, XRN2 in human (and Rat1 in yeast), degrades the Pol II-associated product of poly(A)-site cleavage and helps to dissociate Pol II from the DNA template. Recent findings support the existence of a combined model, where an allosteric switch decelerates Pol II beyond the poly(A)-site, followed by its termination by XRN2 (**Figure 2D**). However, not every transcript contains a poly(A)-site, as is the case of replication-dependent histones (RDH) which 3' end is cleaved by the CPSF73 endonuclease. In this case, CPSF73 is recruited by U7 snRNA, which is part of the histone cleavage complex (HCC). Similarly, snRNAs do not contain a poly(A)-site, and its 3' end processing has been described to be accomplished by the multi-subunit integrator complex (Eaton and West, 2020).

Recent findings show that early termination can occur in different stages of gene transcription, regulated by the integrator complex (Elrod et al., 2019; Tatomer et al., 2019), U1 snRNP (Berg et al., 2012; Chiu et al., 2018) and the exosome complex (Davidson et al., 2019). The integrator complex role in early termination was related to non-productive elongation regulation (Beckedorff et al., 2020; Lykke-Andersen et al., 2021).

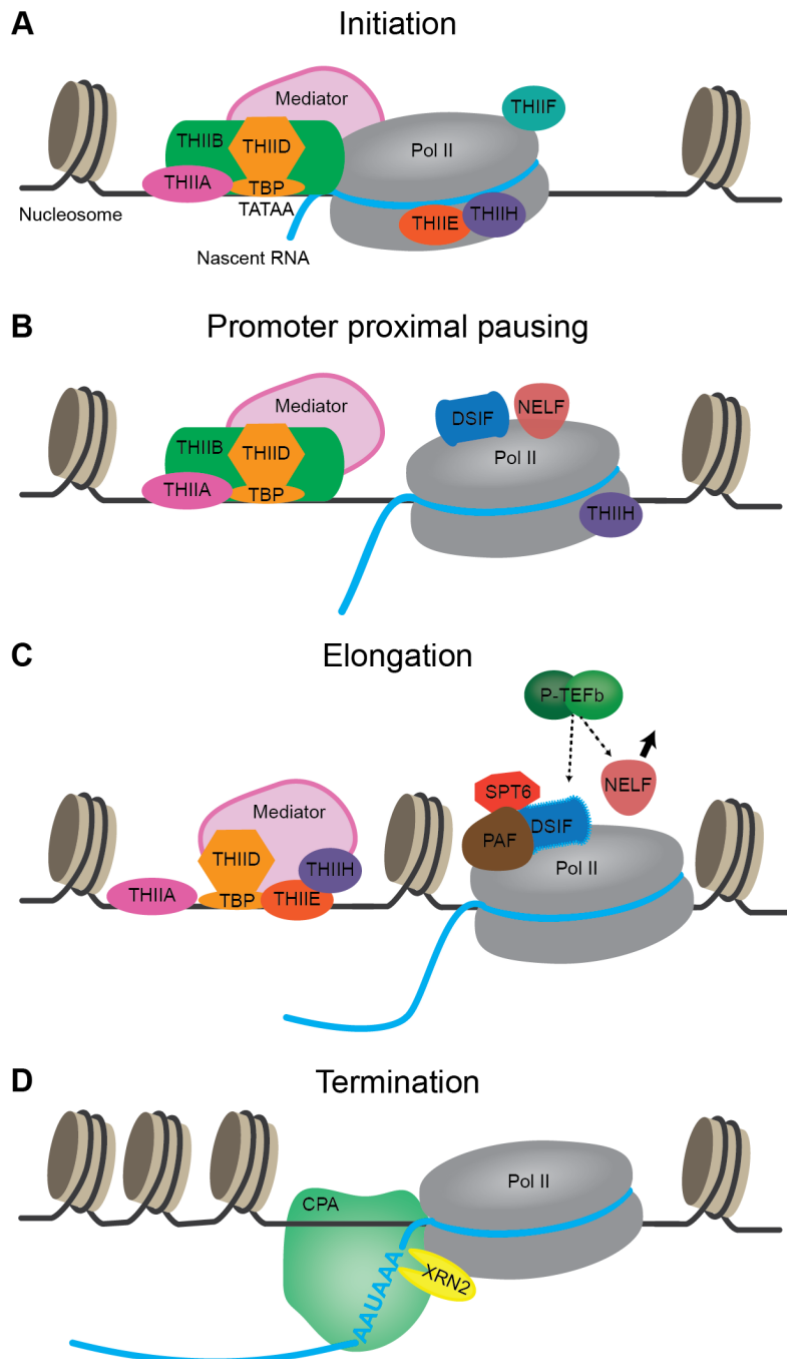


Figure 2. The transcription cycle.

(A) Transcription initiation of a TATA-box containing gene, showing PIC assembly. TFIID TBP is recruited to the TATAA sequence. TFIIA stabilizes the TBP-promoter interaction. TFIIB recognizes and binds to elements flanking the TATA box. Pol II, TFIIF, Mediator complex TFIIB and TFIIE bind the core promoter. (B) Promoter-proximal pausing with DSIF and NELF binding Pol II after dissociation of TFIIB, TFIIF, and TFIIE. (C) After P-TEFb phosphorylates DSIF and NELF, Pol II is released into productive elongation, leading to the transition of DSIF to a positive elongation factor and the release of NELF. PAF and SPT6 associate with Pol II and DSIF to form the EC*. Some PIC components remain associated at the promoter, forming a reinitiating scaffold. (D) Termination of a poly(A)-site containing gene: poly(A) is recognized by CPA and cleavage by XRN2. Adapted from (Eaton & West 2020; Shandilya & Roberts 2012).

1.5. Pol II RPB1 CTD

RPB1 is the largest subunit of RNA polymerase II. It comprises a tail-like domain that contains a unique CTD consisting of heptad repeats of YSPTSPS that is central in the regulation of Pol II transcription, including the co-transcriptional splicing process (Harlen and Churchman, 2017; Schier and Taatjes, 2020). The residues in the CTD are described based on their position in the heptapeptide: Tyr1 Ser2 Pro4 Thr4 Ser5 Pro6 and Ser7. The number of repeats varies between species, comprising 26 repeats in yeast and 52 repeats in human. The CTD residues can be subjected to several post-translational modifications, promoting or inhibiting proteins' binding to the CTD. The most study modification is phosphorylation (Harlen and Churchman, 2017).

During transcription pre-initiation, CTD is found in a hypo-phosphorylate form, where the Mediator is recruited to the promoter. Pol II CTD becomes phosphorylated at the Ser5 and Ser7 position by the THIIH kinase CDK7, to release Pol II from the PIC. Therefore, high peaks of Ser5P and Ser7P are found around the TSS and promoter. In metazoans, Tyr1P peaks are also correlated with paused Pol II. Elongating Pol II shows a decrease of Ser5P and an increase of Ser2P and Tyr4P (**Figure 3A**). Ser2 is phosphorylated by PTEF-b kinase CDK9 (Harlen and Churchman, 2017).

Besides its regulation regarding the residues post-translation modifications, it was also shown that CTD can undergo liquid-liquid phase separation and that its length is essential for Pol II clustering *in vivo* (Boehning 2018). A human CTD containing only 25 repeats out of the normal 52 will reduce the number of Pol II clusters in the nucleus and impacts the transcription activation of target genes and their putative enhancer elements (Boehning 2018, Sawicka 2021).

Pol II CTD phosphorylation is also shown to play a critical role in co-transcriptional splicing (Maita and Nakagawa, 2020), which I discuss further below (see **section 3.2**)

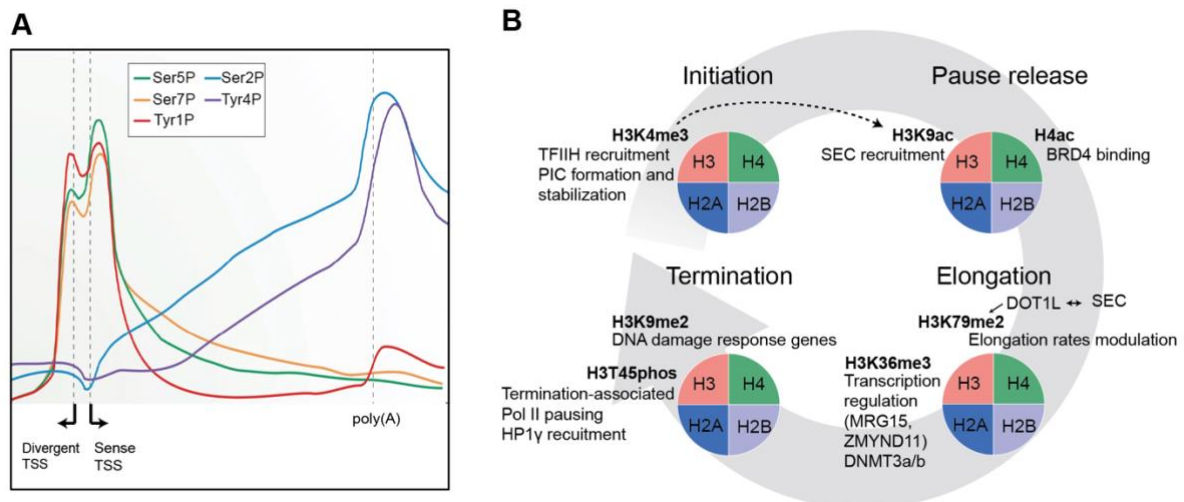


Figure 3. Transcription regulation.

(A) Pol II CTD phosphorylation profile showing a peak for Ser5, Ser7, and Tyr1 phosphorylation around TSS and increase signal of Ser2P and Tyr4P along the gene body. Adapted from (Harlen and Churchman, 2017). (B) Schematic representation of the main histone modifications and the correlation with the transcription cycle. H3K4me3 effect on H3K9ac is shown with a dotted arrow. Adapted from (Gates et al., 2017a).

1.6. Histone modifications and transcription regulation

Histone modifications have a substantial impact throughout the transcription cycle. The nucleosome is the basic unit of eukaryotic chromatin. It consists of a histone octamer: an eight-protein complex

formed by two copies each of H3, H4, H2A, and H2B, and 146 base pairs of DNA wrapped around the histones. Histones contain regions of polypeptide N-terminal to the central, globular helical histone fold that are exposed to broad post-translational modifications (Gates et al., 2017a).

Different histone modifications are observed during the transcription cycle (**Figure 3B**). During transcription initiation, trimethylation on histone H3 lysine 4 (H3K4me3) is associated with TFIID recruitment and PIC formation and stabilization (Vermeulen et al., 2007). Productive transcription elongation is associated with histone lysine acetylation (Church et al., 2017; Gao et al., 2020; Gates et al., 2017b). It was found that H3K4me3 triggers H3K9ac (Foulds et al., 2013) and H3K9ac was suggested to mediate Pol II pause release through SEC recruitment (Gates et al., 2017b). Likewise, H4ac promotes binding of BRD4 and is also related to pausing release (Gonzales-Cope et al., 2016; Zippo et al., 2009).

H3K36me3 is also associated with productive elongation and has an essential role in transcription regulation processes, such as preventing cryptic start sites and splicing (Kim et al., 2011; Kolasinska-Zwierz et al., 2009; Neri et al., 2017). H3K36me2 and H3K36me3 are correlated with *de novo* DNA methyltransferases DNMT3a and DNMT3b (Dukatz et al., 2019; Rondelet et al., 2016; Weinberg et al., 2019). H3K36me3 is also correlated with the MRG15 transcription factor (Xie et al., 2011) and the transcription repressor ZMYND11 (Guo et al., 2014; Wen et al., 2014). H3K79me2 and H3K79me3 were also associated with productive elongation. In humans, DOT1L is the enzyme responsible for H3K79 methylation and it has been associated with Pol II CTD and SEC (Kim et al., 2012). In addition, H3K79me2 levels are associated with elongation rates, suggesting that H3K79me2 can regulate gene expression through DOT1L regulation (Ljungman et al., 2019).

H3K9me2 is associated with transcription termination, specifically at genes that show termination-associated Pol II pausing. The enzyme responsible for H3K9 methylation, G9 methyltransferase, is required for Pol II to pause at the 3' end and for effective termination (Skourti-Stathaki et al., 2014). G9 H3K9 methylation is also responsible for the recruitment of the heterochromatin protein 1 γ (HP1 γ) that is implicated in splicing (Rachez et al., 2019; Saint-André et al., 2011; Vakoc et al., 2005). Additionally, H3T45phos was related to termination after exposure to DNA damaging agents (Lee et al., 2015).

2. Co-transcriptional processing

Mature functional mRNA undergoes processing mechanisms and checkpoints such as 5' capping, splicing, and 3' processing (de Almeida and Carmo-Fonseca, 2010; Bentley, 2014). The splicing mechanism will be further discussed in **section 3**.

The first mRNA processing event is the 5' capping, which occurs shortly after the transcript emerges from the RNA exit channel, and it protects the newly synthesized RNA against 5' exonucleases. A 7-methylguanosine cap is added to the 5' end of nascent transcripts. First, the triphosphate on the first nucleotide is converted to a diphosphate group. Then, a GMP molecule is attached, and the N7 atom of the guanosine base is methylated to produce a mature cap (Kachaev et al., 2020).

The better-characterized cap-binding complexes are the cap-binding complex (CBC) and the eukaryotic initiation factor 4F (eIF4F) that is related to translation initiation. CBC is a heterodimer formed by 20 and 80 kDa subunits, denominated CBP20 and CBP80, and is related to several

processes such as gene expression, RNA processing mechanisms such as splicing, and 3' end formation, transcript nuclear export, and translation (Gonatopoulos-Pournatzis and Cowling, 2014; Kachaev et al., 2020).

Promoter-proximal pausing and 5' capping were shown to be connected, suggesting a capping checkpoint model in which Pol II pausing in the promoter region will ensure the 5' capping of nascent RNA (Kachaev et al., 2020). DSIF subunit SPT5 and Pol II CTD Ser5 phosphorylation stimulate RNA capping (Doamekpor et al., 2014; Komarnitsky et al., 2000; Wen and Shatkin, 1999). Both P-TEFb and NELF interact with CBP, plus NELF was shown to have a role on CBC recruitment (Aoi et al., 2020; Lenasi et al., 2011; Narita et al., 2007; Schulze and Cusack, 2017). Moreover, CBC has an impact on promoter-proximal pausing release into productive elongation (Fujiwara et al., 2019; Lahudkar et al., 2014; Myers et al., 2002; Varshney et al., 2018).

Regarding 3' ending processing, most mRNAs contain a poly(A) tail. Therefore, the newly synthesized mRNAs undergo RNA cleavage (discussed in **section 1.4**) and poly(A) tail addition (Danckwardt et al., 2008). The nuclear poly(A) polymerase (PAP) adds between 150 to 250 A-nucleotides to the mRNA 3' end in a template-independent manner (Danckwardt et al., 2008). However, some mRNAs lack or have a shortened poly(A) tail. Additionally, poly(A) length has been associated with an impact on gene expression regulation (Danckwardt et al., 2008; Jalkanen et al., 2014). The poly(A) is bound by the **polyadenylate-binding nuclear protein 1** (PABPN1, also known as PABP2) in the nucleus. PABPN1 is then replaced in the cytoplasm by the **cytoplasmic polyadenylate-binding protein** (PABPC). PABPC interacts with eIF4G protein from eIF4F complex, inducing translation. PABPC also interacts with the **eukaryotic release factor 3** (ERF3) related to translation termination (Danckwardt et al., 2008).

3. Transcription and splicing

3.1. The splicing cycle

Splicing is an essential step of gene expression where the introns are removed from precursor RNAs, and the flanking exons are joined to generate mature transcripts. These reactions are catalyzed by the spliceosome, a multi-megadalton ribonucleoprotein complex. Most eukaryotes comprise a major spliceosome that splices U2-type introns and a less common minor spliceosome that splices a rarer class of U12-type introns (Will and Lührmann, 2011). Here I will focus on the common major spliceosome.

The major spliceosome is formed by five snRNPs: U1, U2, U4, U5, and U6, containing associated snRNAs with specific and shared proteins that recognize conserved RNA elements termed splice sites (SS). The 5' splice site (5' SS) is located in the exon-intron border (donor site), and the 3' splice site (3' SS) refers to the intron-exon border (acceptor site). Another important sequence is the branch site (BS), located from 18 to 40 nucleotides upstream of the 3' SS. The BS contains an adenosine followed by a polypyrimidine tract (Will and Lührmann, 2011).

Chemically, the splicing process occurs by two consecutive transesterification reactions. In the first reaction, the hydroxyl (HO) group of the BS adenosine carries out a nucleophilic attack on the phosphate (P) at the 5' SS. This generates a lariat intermediate and the release of the 5' exon with a HO group. In the second step, the P in the 3' SS is attacked by the 3' OH group of the 5' exon,

removing the intron and ligating the two exons together (Gomez-Escribano and Bibb, 2011) (**Figure 4A**).

The spliceosome assembly pathway begins with U1 snRNP binding to the 5' SS and the U2 auxiliary factor (U2AF) binding to the polypyrimidine tract and the 3' SS, resulting in the early spliceosome E complex. Then, U2 snRNP binds to the BS sequence forming the pre-spliceosome A complex. Afterwards, the tri-snRNP U4/U6·U5 is recruited to form the pre-catalytic spliceosome B complex. Structural arrangements, including several RNA-RNA interactions, lead to the destabilization of U1 and U4 snRNPs and to form the active spliceosome (B^{act} complex). Then the spliceosome is catalytically activated by the Prp2 ATPase, generating the B^* complex, which catalyzes the first transesterification reaction, leading to the formation of the C complex. Follows the C complex conversion into the activated form by Prp16 and catalyzes the second transesterification reaction. Finally, it results in the post catalytic P complex containing the matured RNA. Prp22 releases the matured RNA, and Prp43 disassembles the spliceosome (Kastner et al., 2019; Tellier et al., 2020; Will and Lührmann, 2011) (**Figure 4B**).

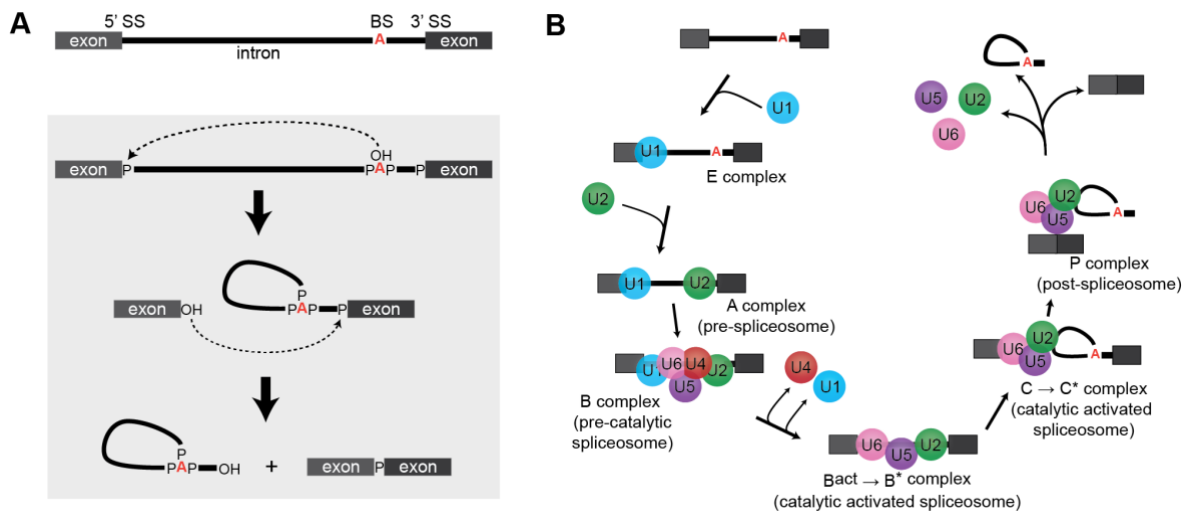


Figure 4. The splicing cycle.

(A) Schematic representation of the splicing chemical reaction. In the first transesterification reaction, the BS OH group attacks the P on the 5' SS, generating a lariat intermediate. In the second transesterification reaction, the 5' exon OH group attacks the 3' SS P, ligating the exons and releasing the intron. (B) Schematic representation of the spliceosome assembly and disassembly pathways. Complex E shows U1 snRNP (blue) bind to the 5' SS. U2 snRNP (green) binds the BS in complex A. The tri-snRNP U4/U6 (pink and red) and U5 (purple) are recruited to form complex B, which is further activated ($B^{\text{act}} \rightarrow B^*$) performing the first chemical reaction. This leads to the formation of complex C, which after activation performs the second catalytic step, resulting in P complex. Matured RNA is released, and the spliceosome is disassembled. "A" represents the BS adenosine. Adapted from (Tellier et al., 2020; Will and Lührmann, 2011).

3.2. Co-transcriptional splicing

Transcription and splicing are interlinked both physically and functionally (Custódio and Carmo-Fonseca, 2016; Herzel et al., 2017; Neugebauer, 2019; Saldi et al., 2016; Tellier et al., 2020; Wallace and Beggs, 2017). More than 30 years ago, Beyer and Osheim showed that the introns being spliced from nascent transcripts are still attached to Pol II on the DNA template. This finding was directly observed by electron microscopic visualization of a Miller chromatin spreading method (Miller and Beatty, 1969) in a drosophila embryo (Beyer and Osheim, 1988) (**Figure 5**).

Splicing and transcription interlink is observed throughout the transcription cycle. The spliceosome is related to transcription initiation by PIC recruitment and promoter selection (Damgaard et al., 2008; Fiszbein et al., 2019). Differences in promoter strength and alternative TSS influence alternative splicing (Anvar et al., 2018; Cramer et al., 1997; Fededa et al., 2005; Kadener et al., 2001). Mediator core component Med23 interacts with splicing factors, including the RNA binding protein hnRNP L, and co-immunoprecipitates (IP) with U1 and U2 snRNAs components (Huang et al., 2012).

Some studies have also found that splicing can influence transcription; for instance, the presence of introns can increase transcriptional efficiency (Brinster et al., 1988; Furger et al., 2002). Splicing inhibition reduces transcription from nearby promoters and impacts TSS usage (Fiszbein et al., 2019). Additionally, a 5' SS in close proximity with the promoter can enhance transcription in yeast and mammals (Furger et al., 2002).

U1 snRNA was found to associate with TFIIF and thus be involved in transcription initiation (Kwek et al., 2002). It was also observed that the insertion of a 5' SS or a complete intron into an intronless gene enhances mRNA production and stimulates early PIC assembly (Damgaard et al., 2008).

Furthermore, CBC has an impact on co-transcriptional splicing and spliceosome assembly (Görnemann et al., 2005; Laubinger et al., 2008; Pabis et al., 2013). Additionally, studies in the *FOS* gene show an interaction between CBC and U4/U6-U5 tri-snRNP (Pabis et al., 2013).

Promoter-proximally paused Pol II release into productive elongation was also observed to be correlated with splicing. In drosophila, a reduction in promoter-proximal pausing is observed after the knockdown of the pre-EJC component Mago (Akhtar et al., 2019). DSIF affects splicing in both *S. cerevisiae* and mammalian cells (Burckin et al., 2005; Lindstrom et al., 2003; Xiao et al., 2005) and can modulate spliceosome assembly (Maudlin and Beggs, 2019).

Inhibition of NUA1, a kinase involved with spliceosome activity, increases Pol II occupancy near the TSS and at the first exon-intron border and decreases nascent RNA synthesis (Cossa et al.,

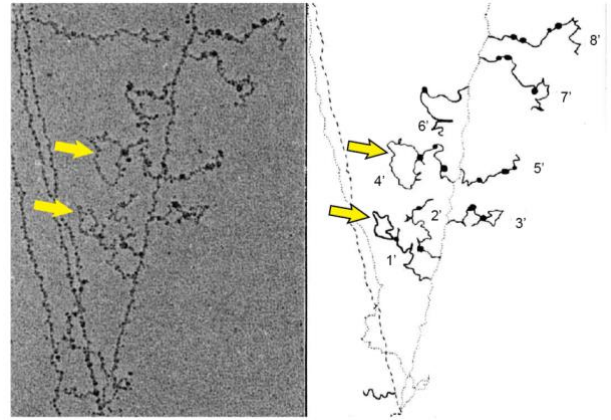


Figure 5. Co-transcriptional splicing.

“Miller spread” electron micrograph for a drosophila embryo gene (left) and its interpretive tracing with a DNA strand shown as a dashed line and the other as a dotted line (right). It is visible an DNA template with several engaged Pol II. Extending from either side of the DNA it’s possible to observe Pol II associated nascent transcripts with bound proteins (dark blobs). Yellow arrows mark introns spliced out co-transcriptionally. Adapted from (Beyer and Osheim, 1988)

2020). A high Pol II occupancy is also related to retained introns, particularly with Ser2P Pol II (Braunschweig et al., 2014).

U1 snRNP inhibition cause **premature cleavage and polyadenylation (PCPA)**; however, this is not observed upon U2 snRNP inhibition (Kaida et al., 2010). U1 role in protecting PCPA also impacts transcription, suggesting an additional U1 telescripting role separable from splicing (Berg et al., 2012; Oh et al., 2017; So et al., 2019).

Splicing occurs in the range of minutes and is limited by the rate of Pol II elongation (Wachutka and Gagneur, 2017), which was shown to affect splicing patterns (Aslanzadeh et al., 2018; Braberg et al., 2013; Fong et al., 2014; Howe et al., 2003; De La Mata et al., 2003; Maslon et al., 2019). Splicing factors stimulate transcription elongation, suggesting a spliceosome-dependent checkpoint for elongation. U2 snRNP stimulates elongation *in vitro* by binding the elongation factor HIV1 **Tat-specific factor 1 (TAT-SF1)** that interacts with P-TEFb (Fong and Zhou, 2001). HIV1-Tat interacts directly with CycT1 (Garber et al., 1998; Wei et al., 1998) and associates with TFIIF RAP30 protein and DSIF domain SPT5 (Kim et al., 1999). Moreover, the splicing factor SKIP enhances HIV1-Tat transactivation through P-TEFb interaction (Brès et al., 2005). Also, the **SR-splicing factor 2 (SRSF2 (or SC35))** influences promoter-proximal release into productive elongation and P-TEFb recruitment (Ji et al., 2013; Lin et al., 2008). Recently, it was shown that U1 snRNP interacts with elongating Pol II independent of nascent RNA on 5' SS, suggesting the formation of a growing intron loop on Pol II that supports scanning for the BS and the 3' SS (Zhang et al., 2021).

Furthermore, Pol II CTD recruits splicing factors (Emili et al., 2002; Misteli and Spector, 1999; Mortillaro et al., 1996; Nojima et al., 2018; Vincent et al., 1996) and stimulates splicing *in vitro* (Hirose et al., 1999) and *in vivo* (Fong and Bentley, 2001; McCracken et al., 1997). The CTD interacts with spliceosomal proteins and regulates spliceosome assembly and splicing (Harlen et al., 2016; Nojima et al., 2018). Phosphorylation on Ser5 was associated with the 5' SS in humans (Nojima et al., 2015). Mutation of the CTD hampering the phosphorylation on Ser2 impairs recruitment of the splicing factor U2AF65 and U2 snRNA, but not of U1 snRNA, impairing co-transcriptional splicing (Gu et al., 2013). Additionally, U2AF65 was found to bind to phosphorylated CTD *in vitro* (David et al., 2011). Moreover, hyperphosphorylated CTD is incorporated into Mediator condensates, preferentially if the condensates are formed by splicing factors (Guo et al., 2019).

Histone modifications are also correlated with splicing, having an essential role in alternative splicing modulation (Luco et al., 2010). Furthermore, SRSF1 and SRSF2 overexpression and splicing inhibition resulted in the decrease of nucleosome occupancy, while an increase is observed upon overexpression of the RNA binding protein hnRNP A1 (Keren-Shaul et al., 2013). Additionally, splicing regulator Hu proteins can induce local histone hyperacetylation around alternative exons regions (Zhou et al., 2011). Splicing inhibition or intron deletion decreases RNA transcription and H3K4me3 levels (Bieberstein et al., 2012). Additionally, splicing factors were found to be related to tri-methylation of H3K36 (de Almeida et al., 2011; Kim et al., 2011), particularly inhibition of SF3B1 and SF3B3 reduces H3K36m3 levels (Convertini et al., 2014; Yuan et al., 2009).

4. Transcriptomics

Several different techniques have been developed over the years to study RNA synthesis and processing. *In vitro* biochemical and single-molecule assays are used to determine the structure of Pol II, important TFs, and transcription and processing complexes. Fluorescent or bead-based assays are used in single-molecule studies; *in vivo* studies can also be accomplished using imaging and system biology (Wang and Greene, 2011; Wissink et al., 2019). Here I will focus on the transcription analysis using a systems biology approach with the resource of next-generation sequencing (NGS) techniques.

NGS platforms can sequence millions of small fragments of DNA at the same time. In summary, the first step is the library preparation, where the input DNA sample is fragmented, and common adapters are added to the small fragments. Then, depending on the sequencing technology, clonally clustered amplicons can be generated by different approaches, such as *in situ* polonies, emulsion PCR, and bridge PCR (Ansorge, 2009).

4.1. RNA synthesis

RNA synthesis studies can be conducted with simpler techniques to track individual transcripts, such as Northern blots or quantitative PCR. A high throughput analysis became possible with the introduction of microarray analysis. Microarray analysis uses microchips with probes to which the nucleic acid on the sample will hybridize. However, microarrays are limited to the detection of known sequences used as probes (Nelson, 2001).

NGS technologies provide a quantitative and qualitative study of different RNA types through RNA-seq. Short-read cDNA sequencing has been widely used to study transcription and gene expression; however, long-read cDNA and direct RNA sequencing approaches are an important alternative (discussed in **section 4.9**) (Stark et al., 2019). Single-cell RNA-seq (scRNA-seq) technologies, which provide a transcriptional profile of thousands of individual cells, are also an essential resource for transcription analysis from an individual cell point of view (discussed in **section 4.10**) (Chen et al., 2019; Yassen et al., 2020).

Some differences in RNA extraction and library preparation can be chosen concerning the biological questions and sample types (Stark et al., 2019). For instance, the use of a ribosomal RNA depletion protocol before library preparation results in greater intron enrichment than the poly(A)⁺ RNA enrichment method. However, available alternative splicing methods work better with the poly(A)⁺ RNA enrichment method (Sultan et al., 2014).

A more precise analysis of RNA synthesis and transcription processing can be accomplished with the study of newly synthesized RNA, which can be achieved by metabolic labeling, where the cells are labeled with uridine analogs, such as 4-thiouridine (4sU), 5'-Bromouridine (BrU), and 5-ethynyl uridine (EU) (Wissink et al., 2019).

In Bru-seq, the cells are exposed to BrU for 30 min, and the transcripts tagged with BrU are further separated from total RNA using magnetic beads coated with anti-BrdU antibodies. Then, the tagged transcripts are converted into a strand-specific cDNA library and subjected to deep sequencing. In BruChase-seq, a pulse-chase step is added before purification, permitting the study of the RNA stability (Paulsen et al., 2013). In 4sU-seq, cells are briefly treated with 4sU, and the labeled transcripts are isolated and sequenced (Garibaldi et al., 2017). To better study initiation frequencies and elongation rates, 4sUDRB-seq includes a treatment with DRB to inhibit RNA elongation before

labeling the cells with 4sU (Fuchs et al., 2014a). 4sU-seq comprises a 5' bias due to the brief 3' end labeling. To solve that, TT-seq includes a fragmentation step after labeling, which increases the coverage of short-lived introns (Schwalb et al., 2016). SLAM-seq (Herzog et al., 2017), TimeLapse-seq (Schofield et al., 2018), TUC-seq (Lusser et al., 2020; Riml et al., 2017) and TUC-seq DUAL (Gasser et al., 2020) also use 4sU labeling but rely on the chemical conversion of the 4sU into cytosine, which will create single nucleotide T > C mutations that are then sequenced by deep sequencing. The chemical conversion has the advantage of not requiring enrichment of the labeled RNA.

Several approaches had been developed to study different RNA molecules or processing processes, such as Start-seq (Nechaev et al., 2010) and CapSeq (Gu et al., 2012) to study TSS and capping. 3'-seq (Lianoglou et al., 2013), PAL-seq (Subtelny et al., 2014) and TAIL-seq (Chang et al., 2014) are used to explore the role of the 3' UTRs and the poly(A) tail.

4.2. DNA-protein interactions

Besides being a DNA-seq based approach, ChIP has an important impact on transcription study. This technique can measure the occupancy level of proteins bound to DNA *in vivo* (Aparicio et al., 2004). A ChIP assay starts with a cross-linking step (usually using formaldehyde) to stabilize DNA-protein complexes, followed by cell lysis, chromatin DNA fragmentation/shearing to obtain randomized fragments, IP with the antibody of interest, followed by de-cross-linking and DNA purification. A ChIP assay can be combined with different technics; for instance, it can be applied to a microarray (ChIP-chip) (Ren et al., 2000) or it can be followed by real-time PCR (ChIP-qPCR) to investigate target genes (Kim and Dekker, 2018). In 2007, the combination of ChIP with the increasing popularity of NGS allowed the development of ChIP-seq, which allows the overall study of DNA-protein interactions and histone modifications profiling (Barski et al., 2007; Johnson et al., 2007; Robertson et al., 2007). Different variations of ChIP-seq have been established since then. ChIP-exo and ChIP-nexus use an exonuclease, allowing near-nucleotide resolution (He et al., 2015; Rhee and Pugh, 2012).

Chromatin endogenous cleavage sequencing (ChEC-seq) was developed as an alternative to ChIP-seq. This approach comprises a DNA-binding protein of interest fused with **micrococcal nuclease (MNase)**, which will cleave the DNA around the binding site, removing the need for cross-linking (Zentner et al., 2015). Later, the same lab updated this approach to develop CUT&RUN. Instead of a fusion protein, it uses an antibody against the protein of interest and a pAG-MNase enzyme to isolate specific protein-DNA complexes (Skene and Henikoff, 2017). Techniques regarding chromatin-protein interactions are mentioned in section 4.7.

4.3. DNA-RNA interactions

DNA:RNA IP followed by **sequencing (DRIP-seq)** is used to study R-loops formation, which is involved in different regulatory mechanisms, such as gene expression. This technique uses a DNA-RNA structure-specific S9.6 antibody (Ginno et al., 2012). R-loops are formed by two antiparallel DNA strands and one RNA strand that is base-paired to one of the DNA strands while the other is unpaired, forming a DNA:RNA hybrid loop (Lieber and Chedin, 2001).

4.4. RNA-protein interactions

Ule *et al.* developed a **cross-linking** and **IP** (CLIP) approach to identify the interaction between proteins and RNA. This technique uses UV light to crosslink the proteins near the irradiated bases with the RNA; then, the samples are partially digested with RNase T1 followed by protein degradation by proteinase K (Ule *et al.*, 2005). The CLIP approach was further adapted to high throughput sequencing (CLIP-seq or HITS-CLIP) (Chi *et al.*, 2009; Licatalosi *et al.*, 2008). Many techniques have been derived from CLIP-seq to accomplish nucleotide resolution sequencing. PAR-CLIP includes the incorporation of 4sU, allowing for precise identification of the binding sites by scoring for T > C transitions in the sequenced cDNA (Hafner *et al.*, 2010). The use of RNase was removed on iCLIP. In this technique, after UV crosslinked, the RNA-protein complex is IP, digested by proteinase K, RNA is reverse transcribed, cDNA is truncated at the binding site and circularized, followed by PCR amplification and deep sequencing (König *et al.*, 2010). eCLIP is another nucleotide resolution CLIP approach which ligates the cDNA to single-stranded DNA adapters containing an in-line random-mer sequence to serve as unique identifiers against PCR duplicates (Van Nostrand *et al.*, 2016). irCLIP also provides nucleotide resolution by using on-bead nuclease digestion (Zarnegar *et al.*, 2016). Some techniques using CLIP were also developed for specific analysis, such as AGO-CLIP to detect argonaute protein binding sites (Grosswendt *et al.*, 2014) or mirCLIP to target miRNA (Imig *et al.*, 2015).

Another approach to detecting RNA-protein interaction is **RNA IP sequencing** (RIP-Seq), which, unlike CLIP-seq, does not use UV cross-linking. In this process, the RNA-protein complex is IP using antibodies against target proteins, followed by RNase digestion, RNA extraction, and reverse sequencing. Even though simpler than CLIP-seq, it relies on the antibody specificity to the target protein (Zhao *et al.*, 2010). **Targets of RNA-binding proteins identified by editing** (TRIBE) identifies the RNA-protein interaction of interest by coupling the protein to the catalytic domain of the drosophila RNA-editing enzyme ADAR and further identification of the fusion protein (McMahon *et al.*, 2016).

To identify ribosome-engaged mRNAs, two techniques were developed: **ribosome profiling** (Ribo-Seq, also known as ARTseq) and **targeted purification of polysomal mRNA** (TRAP-Seq) (Ingolia *et al.*, 2009; Jiao and Meyerowitz, 2010). In Ribo-seq, the ribosome-bound RNA is digested with RNase, followed by RNA extraction and rRNA depletion, reverse transcription, and sequencing (Ingolia *et al.*, 2009). TRAP-seq uses tagged ribosomal proteins that are further purified, and the complexed RNA is isolated, reverse transcribed, and sequenced (Jiao and Meyerowitz, 2010)

Protein interaction profile sequencing (PIP-seq) provides a protein-footprint based on RNA-protein interaction by using two sets of cells, subjected or not to UV cross-linking (Foley and Gregory, 2016).

4.5. RNA-RNA interactions

Another technique using CLIP was developed to study mRNA secondary structures: hiCLIP (Sugimoto *et al.*, 2015). Additionally, a set of different techniques using UV cross-linking were also established to focus on RNA-RNA interactions, such as *SPLASH* (Aw *et al.*, 2016), CLASH-seq (Kudla *et al.*, 2011), LIGR-seq (Sharma *et al.*, 2016), and PARIS (Lu *et al.*, 2018).

4.6. Pol II occupancy

Pol II occupancy can be studied with Pol II ChIP-seq (or the ChIP variants refereed previously in **section 4.2**). The need to increase sequencing resolution led to the development of techniques using nuclear run-on assays. In a nuclear run-on assay, the cells are put on ice to stop Pol II transcription; then, the nuclei are incubated with nucleoside analogs, followed by the extraction of the run-on transcripts (Hirayoshi and Lis, 1999). **Global run-on sequencing (GRO-seq)** uses 5-**bromouridine 5'-triphosphate (Br-UTP)** labeling after nuclei isolation, followed by hydrolyzation, purification, and isolation of the run-on transcripts RNAs that are further reverse-transcribed into a cDNA library and subjected to deep sequencing. GRO-seq captures the RNA bound to transcriptionally active Pol II (Core et al., 2008). **Precision run-on sequencing (PRO-seq)** includes the incorporation of biotin-11-NTPs which will result in the incorporation of a single nucleotide to the transcript and therefore improving the technique to a nucleotide resolution. Pol II initiation sites can be mapped using a cap selection (**PRO-cap**) (Kwak et al., 2013). **Chromatin run-on and sequencing (ChRO-seq)** works like a simplified version of PRO-seq that can also be applied to tissue samples. In ChRO-seq the nuclei isolation is replaced by an insoluble chromatin fractionation strategy (Chu et al., 2018). mNET-seq, adapted from yeast NET-seq (Churchman and Weissman, 2012), also produces nucleotide resolution Pol II occupancy. Elongating Pol II complexes are first isolated by fractionation, followed by an MNase digestion to break down all exposed DNA. Then, Pol II antibodies are used to IP the polymerase with attached RNAs. 3' linkers are ligated to RNA that is then purified and processed into cDNA sequencing libraries (Nojima et al., 2015). Antibodies directed to different Pol II CTD states and empigen detergent, which separates RNA processing complexes from Pol II, allow for a better understanding of the co-transcriptional process, such as splicing (Nojima et al., 2015, 2018; Schlackow et al., 2017).

4.7. Chromatin related techniques

Chromatin accessibility is an important factor in transcription analysis. In 2006, Crawford *et al.* developed DNase-chip, a technique based on chromatin digestion with Dnase I, followed by hybridization to microarrays (Crawford et al., 2006). Later with the use of NGS, DNase-seq was developed, generating nucleotide resolution sequencing data (Boyle et al., 2008). **Formaldehyde-assisted isolation of regulatory elements (FAIRE)** assays, on the other hand, use formaldehyde for chromatin cross-linking, following by DNA fragmentation/shearing with sonication and phenol-chloroform extraction (Giresi et al., 2007). MNase can be used for genomic DNA digestion and to study nucleosomes (Cusick et al., 1981). MNase -seq used MNase to degrade the naked DNA around the nucleosome, and the DNA from the nucleosome complexes is extracted and subjected to sequencing (Schones et al., 2008). A more recent approach, the **assay for transposase-accessible chromatin with sequencing (ATAC-Seq)**, uses hyperactive Tn5 to simultaneously cut and ligate adapters for high-throughput sequencing at regions of increased accessibility. ATAC-seq data can spot both the nucleosome-free regions (shorter reads) and nucleosome positions corresponding to DNA fragments protected by a nucleosome (longer read) (Buenrostro et al., 2015a).

Regarding the analysis of the spatial organization of chromatin in a cell, several techniques denominated **chromosome conformation capture (3C)** have been developed. These approaches comprise formaldehyde cross-linking, fragmentation, DNA extraction, and digestion with restriction enzymes. 3C techniques can be divided according to their scope. Techniques like 3C-seq

(Duan et al., 2012) can quantify interactions between two specific fragments (one-vs-one). 4C-seq (Zhao et al., 2006) and Capture-C (Hughes et al., 2014) can quantify interactions between one locus and all other genomic loci (one-vs-many). Hi-C (Lieberman-Aiden et al., 2009) allows the quantification of chromatin interactions genome-wide (all vs. all). Tiled-C (Oudelaar et al., 2020) accomplish a genome-wide interactions analysis with the possibility to precisely identify enhancer-promoter topologically associating domains (TADs). CHIP-loop (also called CHIP-3C) (Horike et al., 2005), ChIA-PET (Li et al., 2014) and HiChIP (Mumbach et al., 2016) combine a 3C approach with an IP step using an antibody against the protein of interest.

Regarding RNA-Chromatin Interactions, several techniques are available. ChIRP-seq (Chu et al., 2011) and CHART-seq (Simon et al., 2011) use biotinylated probes for the target RNA, followed by pull-down and elution of the target complexes. More recently, developed techniques allow the identification of RNA-chromatin interaction genome-wide, such as GRID-seq (Li et al., 2017), Char-seq (Bell et al., 2018) and RADICL-seq (Bonetti et al., 2020). These approaches use an oligonucleotide duplex bridge to ligate at one side the chromatin-associated RNA and at the other side the previously digested genomic DNA.

4.8. RNA-seq data analysis

Different approaches are available for transcriptomics analysis techniques, depending on the biological question to address and the specific method (**Table 1**). Regarding RNA synthesis approaches, the basic analysis can be summarized as quality control, read assembly, and RNA synthesis quantification.

The first step after sequencing is demultiplexing, where the sequences obtained from one lane are assessed to the respective samples, using the barcode information. Then, an evaluation of the raw sequences must be performed. A typical software used for this step is FASTQC (Andrews, 2010) that comprises both a graphical user interface (GUI) and a terminal version. Other tools for generic raw data quality analysis are available, such as FASTX-Toolkit (Hannon, 2010), NGS QC toolkit (Patel and Jain, 2012) and EasyQC (Rangamaran et al., 2018). Some critical aspects regarding quality check are sequence number, quality and length distribution, GC content, duplicate levels, and the presence of contaminants, such as adapters. Some tools were developed for read trimming and adapter removal, such as Trimmomatic (Bolger et al., 2014) and Cutadapt (Martin, 2011).

After read quality control, the most common process is the alignment of the sequenced reads to a genome or a transcriptome. Several software were designed to perform the alignment of cDNA reads to a reference genome. As an example, a commonly used alignment software is TopHat2 (Kim et al., 2013), which uses Bowtie (Langmead et al., 2009) or Bowtie2 (Langmead and Salzberg, 2012) as a first genome alignment, followed by splice junction identification. Bowtie2 is a fast aligner that uses Burrows-Wheeler transforms for indexing. It is mainly used for DNA sequencing analysis since it is not splice aware. Another common RNA-seq aligner, STAR (Dobin et al., 2013), uses a different strategy to account for spliced isoforms that do not rely on a continuous DNA aligner such as Bowtie. Instead, it uses a seed searching step that searches for the longest read that matches one or more genome reference sequences and does a sequential mapping until the read is completely mapped to the genome, considering the intronic mismatches. Then all the seeds are clustered, stitched together, and a local alignment scoring scheme is performed. In this study, STAR was used to align the TT-seq, RNA-seq, and mNET-seq sequenced cDNA reads to the genome. Bowtie2 was used to align CHIP-seq DNA reads.

After read alignment to a reference genome, transcriptome assembly can be performed, where the previously mapped reads will be assembled to transcripts and further gene expression quantification. Most of these tools use strategies based on graph theories. For example, Cufflinks (Trapnell et al., 2010) uses an overlap graph, where the mapped reads form the nodes, and two nodes will be connected regarding splicing pattern overlap. Other examples are StringTie (Pertea et al., 2015) and Bridger (Chang et al., 2015) that comprise the building of a splice graph. On the other hand, SLIDE (Li et al., 2011) uses a linear model for isoform assembly. A different way of achieving gene expression levels is by quantifying the number of reads that map to a defined sequence feature, such as full transcripts, exons, or introns, which can be achieved with HTSeq (Anders et al., 2015) or featureCounts (Liao et al., 2014). An alternative to transcript assembly is the direct coverage level study using the reads mapped to the genome. The GenomicRanges and GenomicAlignments R packages (Lawrence et al., 2013) allow the manipulation of the mapped reads to output coverage profiles as well as counting read files. The same could also be achieved in python language using PyRanges (Stovner and Sætrum, 2019).

Table 1. Overview of the developed techniques for transcriptomics analysis.

Representative summary of techniques for transcriptomics study

Analyses		Techniques	
RNA synthesis	total RNA	RNA-seq	
	newly synthesised RNA	4sU-seq, TT-seq, TimeLapse-seq, TUC-seq, TUC-seq DUAL, BrU-seq, BruChase-seq,	
	specific analysis	CapSeq, Start-seq, 3'-seq, PAL-seq, TAIL-seq	
DNA-protein interactions		ChIP-seq, ChIP-exo, ChIP-nexus, ChEC-seq, CUT&RUN	
DNA-RNA interactions	R-loop identification	DRIP-seq	
RNA-protein interactions		CLIP-seq, PAR-CLIP, iCLIP, irCLIP, RIP-seq, TRIBE, eCLIP	
	specific analysis	AGO-CLIP, mirCLIP	
	ribosome profiling	Ribo-seq, TRAP-seq	
	protein-footprint	PIP-seq	
RNA-RNA interactions		hiCLIP, SPLASH, CLASH-seq, LIGR-seq, PARIS	
Pol II occupancy		Pol II ChIP, GRO-seq, PRO-seq, PRO-cap, ChRO-seq, mNET-seq	
Chromatin related techniques	Chromatin accessibility		DNase-seq, Mnase-seq, ATAC-seq, Char-seq
	Chromatin interaction	3C	3C-seq, 4C-seq, Capture-C, Hi-C, Tiled-C, ChIP-loop
		3C + ChIP	ChIP-loop, ChIA-PET, HiChIP
RNA-Chromatin interactions		ChIRP-seq, CHART-seq, GRID-seq, Char-seq, RADICL-seq	

4.9. Long-read sequencing

Up to date, the most widely used long-read sequencing (LRS), also known as third-generation sequencing, is nanopore by **Oxford Nanopore Technologies (ONT)** and **single-molecule real-time (SMRT)** by **Pacific Biosciences (PacBio)**. In nanopore sequencing, a single-stranded DNA or RNA is passed through a 1.5 nm wide bioengineered protein channel embedded in an electrically resistant

polymer membrane (nanopore). An ionic current is passed through this nanopore by setting a voltage across this membrane. The current across the channel depends on which nucleotide is traversing, allowing for base determination (Wang et al., 2014). In SMRT, adapters are ligated to double-stranded DNA, creating a circular template. Then, the DNA template is added together with primers and polymerase to a cell containing millions of zero-mode waveguides wells. As the Pol incorporates labeled nucleotides, light is emitted, allowing for base determination (Roberts et al., 2013). However, LRS provides lower per read accuracy than short-read sequencing (Adewale, 2021; Amarasinghe et al., 2020). LRS has an important impact on sequencing analysis since it allows for insight into the structural variant study, base modifications, isoform quantification, and splicing analysis. In the light of these advantages, several studies have been performed using LRS to study transcription cycle and processing (Carrillo Oesterreich et al., 2016; Drexler et al., 2020, 2021; Ibrahim et al., 2018; Maier et al., 2020).

4.10. Single-cell sequencing

Different cell populations within the same tissue can have different gene expressions. scRNA-seq permits analyze a specific cell population. A limitation of scRNA-seq is that the amount of RNA present in a single cell is on the scale of a picogram, which is significantly lower than necessary for any NGS platform. Thus, RNA amplification is a critical step on scRNA-seq. Single-cell approaches to study newly synthesized RNA using 4sU labeling have been developed, such as scSLAM-seq (Erhard et al., 2019) and NASC-seq (Hendriks et al., 2019). Additionally, chromatin accessibility can be studied in a specific cell population using scATAC-seq (Buenrostro et al., 2015b). Single-cell sequencing can also be used for spatial transcriptomics analysis, allowing the analysis of specific cell populations (Nichterwitz et al., 2016; Wang and Greene, 2011).

4.11. Multi-omics analysis

The integration of different omics techniques can be of great importance to address transcription regulation and processing. In our lab, Gressel, Schwalb *et al.* used a multi-omics analysis to calculate Pol II pause duration using TT-seq and mNET-seq (Gressel et al., 2017, 2019). This approach was also used in the present work and is further described in the **Methods section 5.4**.

Gang *et al.* integrated data from CHIP-seq for TF and histone modifications with RNA-seq to further correlate TFs binding and histone modifications with gene expression level to further characterize these co-operations dynamics. They used a machine learning approach by constructing a support-vector machines (SVM) model to identify TF or histone alterations on gene expression level (Gang et al., 2018). Rodrigues *et al.* used a ratio between TRAP-seq and RNA-seq to calculate the number of transcripts bound to ribosomes to the total mRNA abundance (Rodrigues et al., 2020). Zapparoli *et al.* used a multi-omics approach integrating data from ChIRP-Seq, CHIP-Seq, and RNA-Seq to identify direct transcriptional targets of an intergenic lncRNA expressed in epithelial tissues (Zapparoli et al., 2020). A multi-omics approach was used in the present study to calculate Pol II elongation velocity using TT-seq and mNET-seq (see **Methods section 5.5**).

Multi-omics analyses are conducted in different fields of complex genetic diseases to correlate genetic and epigenetic features. Some of these approaches take advantage of machine learning analysis to unravel the causal regulatory interactions (Nicora et al., 2020). For example, for a gene-wise characterization of cancer-related DNA variants, Kim *et al.*, created a comprehensive Bayesian

prior model using TF ChIP-seq, ChIA-PET, and gene expression QTL (eQTL) from breast cancer cells. They further tested the transcriptional causality of the previously found interactions with microarray and RNA-seq (Kim et al., 2015).

5. Overview of key software used in this analysis

The bioinformatics analysis implemented in this work made use of some published software. I will briefly describe the principles behind the essential software used.

5.1. Defining major transcribed isoforms using Salmon

Salmon uses a dual-phase inference algorithm for alignment and quantification of raw reads to a reference transcriptome. This software can be used with both raw or previously mapped reads as input. In the present analysis, I used raw reads. First, indexing the transcripts in the reference transcriptome is performed, followed by a quasi-mapping approach (by default), determining the best mapping for each read. Then, Salmon uses an online (or streaming) inference algorithm to estimate fragment abundances using a variant of a stochastic collapsed variational Bayesian inference. It also calculates per-fragment conditional probabilities, which will be used to estimate auxiliary models and bias terms and to update the abundance estimates.

In simplified terms, the online inference algorithm will focus on each read separately and calculates the alignment's abundance and collapses reads into equivalent classes. Afterwards, an offline inference (also denominated batch inference) algorithm refines the previous findings based on the different read classes defined (Patro et al., 2017).

5.2. Differential expression analysis using DESeq2

A typical analysis performed using feature counts is the gene expression analysis commonly performed with DESeq2 (Love et al., 2014) or edgeR (Robinson et al., 2009) which use a negative binomial distribution. In the present work, I used DESeq2 (1.24.0, RRID:SCR_015687) (Love et al., 2014) applied to different input counts from TT-seq analysis. DESeq2 is available as a R/Bioconductor R package and fits a generalized linear model for each gene following a negative binomial distribution. It is based on the analysis of logarithmic fold changes (LFC) of a gene expression value. It accounts for the variability between replicates which is a critical step for the statistical inference of differential expression (DE) between small sample sizes, such as control vs. treatment (Love et al., 2014).

In summary, it uses an Empirical Bayes shrinkage method in which the prior LFC distribution is estimated from the data and shrank towards zero. In that way, it solves the problem found with high LFC on genes with low counts. After fitting a generalized linear model to each gene, a Wald test is used to test if the shrink LFC divided by its standard error differs significantly from zero, by comparison to a standard normal distribution. It also removes outliers using Cooks' distance which measures the effect of deleting a given observation (Love et al., 2014).

5.3. Gene ontology analysis with goseq

A widely used approach to highlight biological processes on a group of selected genes is the categorization of genes into specific ontologies. Gene ontology (GO) knowledgebase is divided into three categories: biological process, cellular component, and molecular function. GO has a hierarchical graph structure, where each GO term is a node with related nodes being connected. Parent terms are more generalized, while child terms are more specialized. However, one child term can have more than one parent term. GO analysis permits us to investigate how likely it is by chance that a gene group in the sample is represented in a GO term that contains n genes (Young et al., 2010). In the present work, I used goseq R/Bioconductor package (1.36.0, RRID:SCR_017052) (Young et al., 2010) upon gene expression analysis to identify the enriched GOs in DE genes. Goseq accounts for transcript length bias by the estimation of a **probability weighting function (PWF)** from the data. First, the data is classified in a binary fashion: 1 for DE genes and 0 for the remaining genes in the sample. Then, a monotonic cubic spline interpolation is fitted to the gene length. Therefore, the PWF will quantify the probability of DE gene changes as a function of its length. The categories significance is calculated using an approximation obtained by approximating the generalized hypergeometric Wallenius distribution for the mean of the values from the PWF for each gene in the set and dividing by the mean of the values from the PWF for genes outside the set (Young et al., 2010).

III

METHODS

In this chapter are listed the bioinformatics methods used for data analysis.

The methods described from section 1 to 5 have been adapted from:

Caizzi, L. *, **Monteiro-Martins, S.** * et al. Efficient RNA polymerase II pause release requires U2 snRNP function. (2021). Mol. Cell 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

The laboratory experiments were performed by Dr. Livia Caizzi (MPI-bpc, Dept. of Molecular Biology), except for mNET-seq experiments which were performed by both Dr. Livia Caizzi and Kseniia Lysakovskaia (MPI-bpc, Dept. of Molecular Biology) and can be found in the Supplementary information section 1.

The majority of the analysis was conducted using a computer cluster supported by the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG). A list of the used software can be found below (**Table 2**).

Table 2. Software resources used in the present work.

Resource	Source	Link/identifier
Operating systems		
macOS High Sierra 10.13.6	Apple Inc.	https://support.apple.com/mac
Scientific Linux release 7.9 (Nitrogen) (cluster)	Linux distribution	https://scientificlinux.org/downloads/
Software/tools		
Bowtie 2 (2.3.5)	(Langmead et al., 2009)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml RRID:SCR_005476
Cutadapt (1.18)	(Martin, 2011)	https://cutadapt.readthedocs.io/en/v1.18/installation.html RRID:SCR_011841
FastQC (0.11.6)	(Andrews, 2010)	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ RRID:SCR_014583
HTSeq (0.6.1.p1)	(Anders et al., 2015)	https://htseq.readthedocs.io/en/master/ RRID:SCR_005514
Picard tools (2.20.5)	(Broad Institute, 2019)	http://broadinstitute.github.io/picard/ RRID:SCR_006525
Salmon (0.13.1)	(Patro et al., 2017)	https://combine-lab.github.io/salmon/ RRID:SCR_017036

Samtools (1.3.1)	(Li et al., 2009)	https://sourceforge.net/projects/samtools/files/samtools/1.3.1/ RRID:SCR_002105
STAR (2.6.0)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR , RRID:SCR_015899
R versions and R packages		
R (3.5.1 and 3.6.2)	(R Core Team, 2018)	
Bioconductor (3.8) (BiocManager 1.30.12)	(Morgan, 2019)	https://www.bioconductor.org https://CRAN.R-project.org/package=BiocManager RRID:SCR_006442
RStudio (1.0.153)	(RStudio Team, 2017)	http://www.rstudio.com/ RRID:SCR_000432
biomaRt (2.40.5)	(Durinck et al., 2009)	https://bioconductor.org/packages/biomaRt/ RRID:SCR_019214
Biostrings (2.50.2)	(Pagès H, Aboyoun P, Gentleman R, 2019)	https://bioconductor.org/packages/Biostrings . RRID:SCR_016949
DESeq2 (1.24.0)	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html RRID:SCR_015687
doParallel (1.0.16)	(Microsoft Corporation and Weston, 2020)	https://cran.r- project.org/web/packages/doParallel/index.html
dplyr (1.0.5)	(Wickham et al., 2020)	https://dplyr.tidyverse.org/reference/dplyr-package.html RRID:SCR_016708
foreach (1.5.1)	(Wallig et al., 2018)	https://cran.r-project.org/web/packages/foreach/index.html
GenomicAlignments (1.18.1)	(Lawrence et al., 2013)	https://bioconductor.org/packages/release/bioc/html/Genomi cAlignments.html
GenomicRanges (1.34.0)	(Lawrence et al., 2013)	http://www.bioconductor.org/packages/2.13/bioc/html/Geno micRanges.html RRID:SCR_000025
goseq (1.36.0)	(Young et al., 2010)	https://www.bioconductor.org/packages/release/bioc/html/g oseq.html RRID:SCR_017052
Gviz (1.28.3)	(Hahne and Ivanek, 2016)	https://bioconductor.org/packages/release/bioc/html/Gviz.ht ml
LSD (4.1-0)	(Schwalb et al., 2020)	https://CRAN.R-project.org/package=LSD
refGenome (1.7.7)	(Kaisers, 2019)	https://CRAN.R-project.org/package=refGenome
rMATS (4.1.0)	(Shen et al., 2014)	http://rnaseq-mats.sourceforge.net/ RRID:SCR_013049
rtracklayer 1.42.2	(Lawrence et al., 2009)	https://bioconductor.org/packages/release/bioc/html/rtrackla yer.html
spliceSites (1.31.0)	(Kaisers et al., 2017)	https://www.bioconductor.org/packages//2.13/bioc/html/spli ceSites.html
Figures and document production and edition		
Microsoft Word for Mac version 16.17	Microsoft Corporation	https://apps.apple.com/us/app/microsoft- word/id462054704?mt=12
Mendeley Desktop version 1.19.4	Elsevier	http://www.mendeley.com/ RRID:SCR_002750

Adobe Illustrator CS6 | Adobe Inc.
version 16.0.4

<http://www.adobe.com/products/illustrator.html>
RRID:SCR_010279

The main scripts used for this analysis are stored on GitHub:

https://github.com/cramerlab/Efficient-RNA-polymerase-II-pause-release-requires-U2-snRNP-function_2021

1. Creating annotation files

1.1. General annotation

I downloaded an NCBI RefSeq hg38 genome assembly annotation gtf file from <https://genome.ucsc.edu/cgi-bin/hgTables>, using the following selection: clade Mammal, genome Human, assembly Dex.2013 GRCh38/h38, group Gene and Gene Predictions, track NCBI RefSeq, table RefSeq Curated. Using the biomaRt R package (2.40.5, RRID:SCR_019214) (Durinck et al., 2009), I included a column with the feature “gene_name” using the Ensembl gene name annotation. I additionally calculated and added some column features, such as transcript length, exon number, exon order, intron number, intron order, and UTR length.

1.2. Major isoform annotation

In order to reduce the bias when using splice junction regions for analysis and plotting, I created an annotation comprising only major transcript isoforms using Salmon (0.13.1, RRID:SCR_017036) (Patro et al., 2017). First, I created the index using the RefSeq transcriptome (UCSC RefSeq GRCh38, downloaded in April 2019) using the default quasi-mapping mode. Then, the isoforms were quantified using RNA-seq reads for 1 h DMSO or 1 μ M Pla-B treatments, with automatic library type detection (-l A) and selective alignment (-validateMappings).

The output quantification files were further analyzed using the R/Bioconductor environment.

I removed the entries with transcripts **per million** (TPM) inferior to 0.5 in every sample and non-curated RefSeq records. Then, I merged the information from the quantification files from Salmon with the general RefSeq annotation file by transcript identifier.

For each gene, the major transcript isoform was determined as the one with maximum mean TPM over all samples. For each condition (DMSO and Pla-B treated samples), the percentage of each transcript isoform P was calculated, according to

$$P_t^{condition} = \frac{mean(T_{tg}^{condition})}{mean(\sum_{t=1}^n T_{tg}^{condition})}$$

with TPM T_{tg} for n transcripts t in gene g .

I excluded major transcript isoforms from further analysis if they had a $P < 70\%$ in both conditions. Furthermore, I used the findOverlaps function from GenomicRanges R package (1.34.0, RRID:SCR_000025) (Lawrence et al., 2013) to identify overlapping major isoforms, which were discarded from further analysis. To ensure that the definition of a splice junction site was not masked by the expression of an overlapping gene, I excluded major isoforms overlapping with other annotated gene isoforms (general annotation file). Isoforms located on chromosomes X, Y, and M

were removed as well. The final major isoform annotation included 6,694 isoforms containing 65,976 exons and 59,282 introns with a median transcript length of around 23 kbp, and most transcripts containing 4 exons (**Figure 6**).

A total of 5,535 major transcript isoforms of protein-coding genes with RPK ≥ 50 from TT-seq solvent control (DMSO) were included in the analysis.

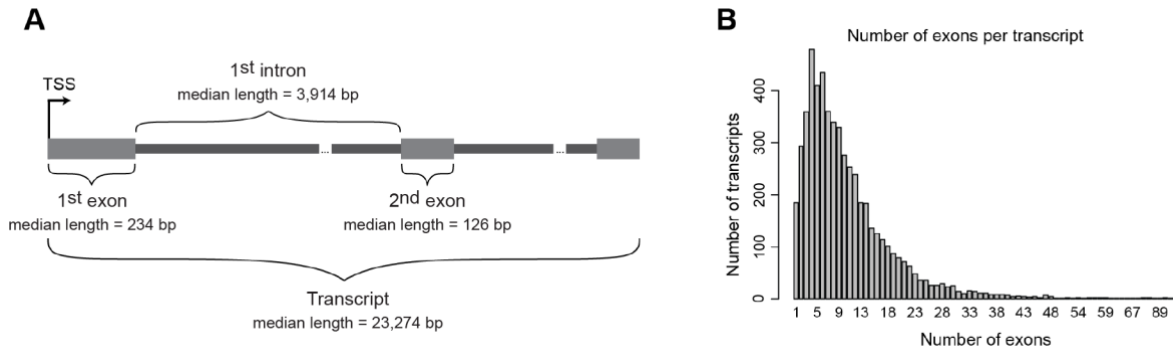


Figure 6. Major isoform annotation metrics.

(A) Schematic representation of a transcript in the major isoform annotation showing median length for all major transcript isoforms (23,274 bp), first and second exon (234 and 126 bp), and first intron (3,914 bp). (B) Bar plot showing the number of exons per major isoform transcript.

1.3. Intronless genes annotation

First, I defined single-exon genes as RefSeq annotated genes comprising one transcript isoform with one single exon (using the general annotation file). Then, I filtered the major isoform annotation for the defined single-exon genes, accounting for a total of 314 intronless genes. To avoid effects from neighboring intron-containing genes, I used the `findOverlaps` function from GenomicRanges R package (1.34.0, RRID:SCR_000025) (Lawrence et al., 2013) to find overlaps between intronless genes with an added length of 1 kb at both 5' and 3' ends and other intron-containing annotated transcripts (strand independent). Overlapping genes were removed, leaving the intronless genes annotation with a remaining 180 genes.

Moreover, because a long 3' UTR has been recently reported to be alternatively spliced in an intronless gene (Le François et al., 2018), I investigated if the intronless gene UTR length has an impact on new RNA synthesis upon 1 h Pla-B treatment (described below). The analysis was performed for 116 expressed protein-coding (see below **section 2.1**) intronless genes. 5' UTR and 3' UTR length was calculated for each transcript. If alternative start or stop codons were annotated, the one closer to the 5' or 3' end was selected, respectively.

A higher impact on RNA synthesis was observed on the third and fourth quartile for both 5' UTR and 3' UTR (**Figure 7**). The same trend was not observed for intron-containing genes (**Figure S1**). Therefore, only intronless genes with UTR ≤ 100 bp were included. Thus, a total of 51 expressed protein-coding intronless genes were included in the final analysis.

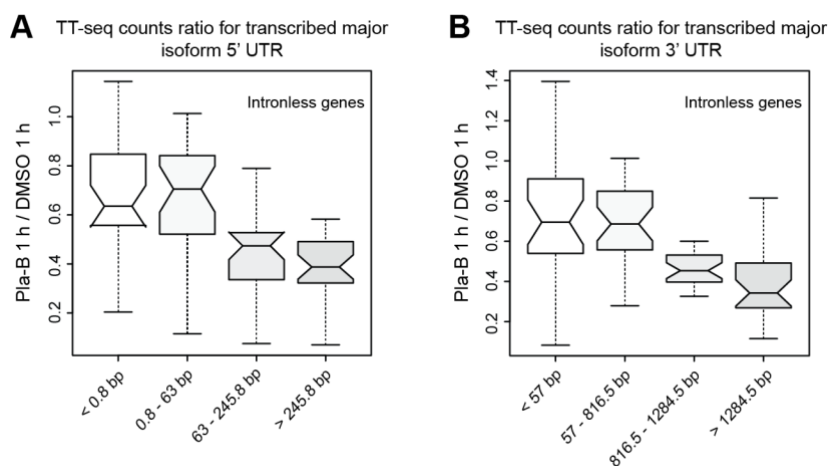


Figure 7. UTR length impact on Pla-B response of intronless genes.

(**A** and **B**) Ratio of 1 μ M Pla-B to DMSO antisense bias-corrected TT-seq read counts for 5' UTR (**A**) and 3' UTR (**B**) in four different intronless gene UTR length quartiles. See also **Figure S1**.

2. General analysis

2.1. TT-seq and RNA-seq

Paired-end 75 and 150 bp reads with an additional 6 bp of barcodes were obtained for each group of samples. First, I performed a quality control analysis on sequence reads using FastQC (0.11.6, RRID:SCR_014583) (Andrews, 2010).

Reads were aligned to the hg38 (GRCh38) genome assembly (Human Genome Reference Consortium) using STAR (2.6.0, RRID:SCR_015899) (Dobin et al., 2013), with the following specifications: `outFilterMismatchNmax 2`, `outFilterMultimapScoreRange 0` and `alignIntronMax 500000`. The human genome was combined with the RNA spike-ins annotation information used for normalization (see below). BAM files were then filtered with Samtools (1.3.1, RRID:SCR_002105) (Li et al., 2009), to remove alignments with MAPQ smaller than 7 (`-q 7`), and only proper pairs (`-f2`) were selected.

I used MarkDuplicates from Picard tools (2.20.5, RRID:SCR_006525) (Broad Institute, 2019) to mark duplicated reads that originated from the same DNA fragment. To make sure the duplicate numbers were not influencing the results, I conducted a parallel analysis removing the duplicated reads. Since the removal of duplicated reads did not change the overall results, we decided not to remove the duplicated aligned sequences for further analysis. Moreover, removing duplicated reads would bias the calibration strategy using RNA spike-ins, since the relation between the amount of initial RNA to the number of reads would be masked.

Read counts for transcripts, exons, and introns were calculated with HTSeq (0.6.1.p1, RRID:SCR_005514) (Anders et al., 2015) using the major isoform annotation. Further data processing was carried out using the R/Bioconductor environment.

Sequence coverage as rle lists

Sequence coverage lists were calculated using the GenomicAlignments R package (1.18.1) (Lawrence et al., 2013) across the bam files to define the number of transcribed bases (tb). Coverage lists were stored as run-length-encoded (rle) lists, which are lists of run-length-encoded vectors. Rle represents the vector as a set of distinct runs with their own value, i.e. sequentially how many times (length) a value is repeated (Lawrence et al., 2013). For the aligned reads with a negative

inner distance between mate pairs, I calculated the number of tb using the coverage function from GenomicAlignments (1.18.1) (Lawrence et al., 2013). In the case of aligned reads with a positive inner distance between mate pairs, I used the findOverlaps function from GenomicRanges (1.34.0, RRID:SCR_000025) (Lawrence et al., 2013) to investigate if these inner mates overlapped with annotated introns (general annotation file). In this case, the inner mates not entirely overlapping introns were summed to the coverage calculated on the paired reads. I calculated coverage tables with coverage means using the sequence coverage rle lists for each feature (transcripts, exons, and introns) as a mean coverage between feature start and end.

Using spike-ins HTSeq counts to calculate antisense-bias ratio c_j , sequencing depth σ_j and cross-contamination ϵ_j .

After treatment, RNA spike-ins were added to lysed cells adjusted to cell number, which allow us to further infer the input amount of RNA per sample. The used spike-ins are derived by the external RNA controls consortium (ERCC) spike-in mix (Baker et al., 2005). During library preparation, the same amount of input RNA (100 ng) was used for each sample to correct for different treatments and biological variations bias. In addition, this analysis comprises the 4sU metabolic labeling of RNA for the analysis of newly synthesized RNA molecules. To address the ratio of labeled RNA to total RNA in the cells half of the internal control spike-ins were labeled *in vitro* with 4sU (**Table 3**). The spike-in normalization strategy allows for the correction of antisense-bias ratio c_j , sequencing depth σ_j and cross-contamination ϵ_j .

Table 3. Overview of the RNA Spike-ins used in this work.

The spike-ins used in this analysis were derived by the ERCC spike-in mix.

Spike-ins	ERCC identifier	Length	4sU labeled
Spike 2	ERCC-00043	1022	yes
Spike 4	ERCC-00136	1032	yes
Spike 5	ERCC-00145	1041	no
Spike 8	ERCC-00092	1123	yes
Spike 9	ERCC-00002	1060	no
Spike 12	ERCC-00170	1022	no

- **Antisense-bias ratio**

We used a strand-specific library preparation strategy/kit, which uses a nucleotide analog for strand selection. cDNA synthesis on genes overlapping on sense and antisense strands may induce the formation of spurious chimeric RNA (Peng et al., 2015). Therefore, the sense coverage is constituted by true sense reads and a small number of spurious antisense reads. To correct for that, I calculated antisense-bias ratios using spike-in information on sense and antisense strand.

The mapped reads to the spike-ins sequences were counted with HTSeq on both sense and antisense strands. Antisense-bias ratios c_j were calculated for each sample j according to

$$c_j = \text{median}_i \left(\frac{k_{ij}^{\text{antisense}}}{k_{ij}^{\text{sense}}} \right)$$

with read counts k_{ij} for all available spike-ins i in sample j . Antisense-bias ratios c_j were inferior to 0.005 for all samples.

- **Sequencing depth and spike-in normalization**

The sequencing depth σ_j for each sample can be inferred by the number of reads mapped to spike-in sequences. Sequencing depths σ_j were calculated for each sample j according to

$$\sigma_j = \text{median}_i \left(\frac{k_{ij}}{l_i} \right)$$

with read counts k_{ij} for all available spike-ins i in sample j for the RNA-seq samples and for the labeled spike-ins i in sample j for the TT-seq samples.

The normalized transcribed bases Ntb_{ij} were calculated for each sample j according to

$$Ntb_{ij} = \frac{tb_{ij}}{\sigma_j}$$

for each genomic position i in sample j .

- **Cross-contamination**

The total RNA found in the TT-seq samples is defined as the cross-contamination ϵ_j rate and was calculated for each sample j according to

$$\epsilon_j = \frac{\text{median}_i \left(\frac{k_{ij}}{l_i} \right)}{\sigma_j}$$

with read counts k_{ij} for all the labeled spike-ins i in sample j for the TT-seq samples and was set to 1 for the RNA-seq samples. Cross contamination ϵ_j rates were inferior to 0.02 for all TT-seq samples upon DMSO or Pla-B treatment and were inferior to 0.5 for TT-seq samples from SSA and U2 AMO treatments.

Expression level

HTSeq read counts were corrected for antisense bias using the previously calculated antisense bias ratios c_j . The number corrected of read counts ck_{ij} for transcript i in sample j was calculated as

$$ck_{ij} = \frac{k_{ij}^{\text{sense}} - c_j \cdot k_{ij}^{\text{antisense}}}{1 - c_j^2}$$

with read counts k_{ij} for all transcripts i .

Read counts per kilobase (RPK) were calculated by dividing the corrected read counts by transcript length and multiplied by 100. The expressed major isoform transcripts were defined as possessing more than 50 RPK in two summarized replicates of TT-seq solvent control (DMSO) for Pla-B 1 h treatment. An RPK of 50 corresponds to approximately a coverage of 5 per sample due to an average fragment size of 200 and the sum of 2 replicates.

2.2. mNET-seq

Paired-end 42 bp reads with additional 6 bp of barcodes were obtained for each sample. Reads were trimmed for adapter content with Cutadapt (1.18, RRID:SCR_011841) (Martin, 2011) with `-O 12 -m 25 -a TGGGAATTCTCGG -A GATCGTCGGACT`. After adapter removal, a FastQC (0.11.6, RRID:SCR_014583) (Andrews, 2010) sample quality report was generated.

To normalize the mNET-seq data with *S. cerevisiae* RNA spike-ins, I generated a combined genome using the Ensembl genome assembly for both human hg38 (GRCh38) and *S. cerevisiae* sacCer3 (R64-1-1) (**Table S4**). Reads were aligned using STAR (2.6.0, RRID:SCR_015899) (Dobin et al., 2013), with the following specifications: `outFilterMismatchNmax 2`, `outFilterMultimapScoreRange 0` and `alignIntronMax 500000`. Bam files were filtered with Samtools (1.3.1, RRID:SCR_002105) (Li et al., 2009) to remove alignments with MAPQ smaller than 7 (`-q 7`) and only proper pairs (`-f2`) were selected. Similar to TT-seq and RNA-seq, I performed the analysis in parallel with and without duplicate removal using MarkDuplicates from Picard tools (2.20.5, RRID:SCR_006525) (Broad Institute, 2019).

Read counts for transcripts were calculated with HTSeq (0.6.1.p1, RRID:SCR_005514) (Anders et al., 2015) using the major isoform annotation. Further data processing was carried out using the R/Bioconductor environment.

Sequence coverage as rle lists

To identify Pol II occupancy, I created rle lists for the mapped reads using the GenomicAlignments R package (1.18.1) (Lawrence et al., 2013). Since the Pol II occupancy is given by the last sequenced nucleotide, the polymerase occupancy (po) was calculated summing the mapped fragments' endpoints at each genomic position.

Antisense-bias ratio

Using mapped reads (bam files) information, the antisense bias ratio was determined using positions in regions without antisense annotation with a coverage of at least 100 according to the defined major isoforms. Antisense-bias ratio values were inferior to 0.01 in all samples.

Normalization with *S. cerevisiae* RNA spike-ins

To include a spike-in normalization to the mNET-seq data, *S. cerevisiae* fragmented mRNA was added to the samples after IP, accounting for approximately 5% of total RNA.

Normalization factors nf_j were calculated using the median of ratios method (Love et al., 2014) on antisense corrected HTSeq counts for 5639 *S. cerevisiae* transcripts with an RPK of 100 or higher in two summarized replicates of mNET-seq solvent control (DMSO).

The normalized polymerase occupancy Npo_{ij} were calculated for each sample j according to

$$Npo_{ij} = \frac{po_{ij}}{nf_j}$$

using the calculated normalization factors nf_j for each genomic position i in sample j .

Datasets for mNET-seq against CTD ser5P in HeLa cells upon 4 h DMSO or Pla-B treatments were obtained from (Schlackow et al., 2017), GEO:GSE81662. Empigen treatment during IP was not performed in this publicly available dataset.

2.3. ChIP-seq

Paired-end 42 or 75 bp reads with additional 6 bp of barcodes were obtained for each of the samples. A combined genome was generated using the Ensembl genome assembly for both human hg38 (GRCh38) and *D. melanogaster* dm6 (BDGP6.28) (**Table S4**). Reads were mapped using Bowtie 2 (2.3.5, RRID:SCR_005476) (Langmead & Salzberg 2012). Bam files were filtered with Samtools (1.3.1, RRID:SCR_002105) (Li et al., 2009) to remove alignments with MAPQ smaller than 7 ($-q 7$) and only proper pairs ($-f 2$) were selected. To make sure the duplicate level did not affect our results, I performed the analysis in parallel with and without duplicate removal using MarkDuplicates from Picard tools (2.20.5, RRID:SCR_006525) (Broad Institute, 2019). Further data processing was carried out using the R/Bioconductor environment.

Sequence coverage as rle lists

I calculated ChIP-seq sequence coverage rle lists using the GenomicAlignments R package (1.18.1) (Lawrence et al., 2013) across mapped fragments for every genomic position. I also created a dedicated annotation around the TSS ($-100, +200$ bp) to compare the CycT1/Pol II ChIP-seq coverage ratio.

Normalization with *D. melanogaster* RNA spike-ins

To correct for global changes in DNA levels upon treatments, we included a *D. melanogaster* spike-in normalization as previously described (Egan et al., 2016). An antibody against *D. melanogaster*-specific histone variant H2Av was added to all samples. *D. melanogaster* S2 sheared cross-linked-chromatin was added before IP, being 0.2 % of the total chromatin. Normalization factors were obtained by dividing the total *D. melanogaster* mapped reads number for each sample by the number of mapped reads of the sample with the lowest reads number.

Normalized ChIP-seq coverages were obtained for each sample by dividing the ChIP-seq coverage by the respective normalization factor.

3. Splicing ratio

I performed an exon-based splice junction analysis using TT-seq. In order to reduce bias due to library preparation, I only included exons from major isoforms containing a first exon > 100 bp. Since we are aiming at exon-intron junctions, I excluded major isoforms with a single annotated exon.

A window of ± 4 bp around the splice junction (2 bp in the exon and 2 bp in the intron) was defined to investigate spliced and unspliced reads employing the findOverlaps function from the GenomicRanges R package (1.34.0, RRID:SCR_000025) (Lawrence et al., 2013). This analysis was based on exon coordinates; therefore, I defined a window between the start of the exon - 2 bp and the start of the exon + 1 bp for 3' SS. For analysis of the 5' SS, I defined a window between the end of exon -1 bp and the end of the exon + 2 bp (as exemplified in **Figure 8**). Unspliced reads were defined as the ones overlapping at least 3 bp of the defined window and total reads (spliced and unspliced) the ones spanning at least 2 bp of the defined window. Afterwards, I calculated the number of spliced reads by the difference between total and unspliced reads. The splicing ratio was

calculated by dividing the number of spliced reads by the total amount of spliced and unspliced reads (**Figure 8**).

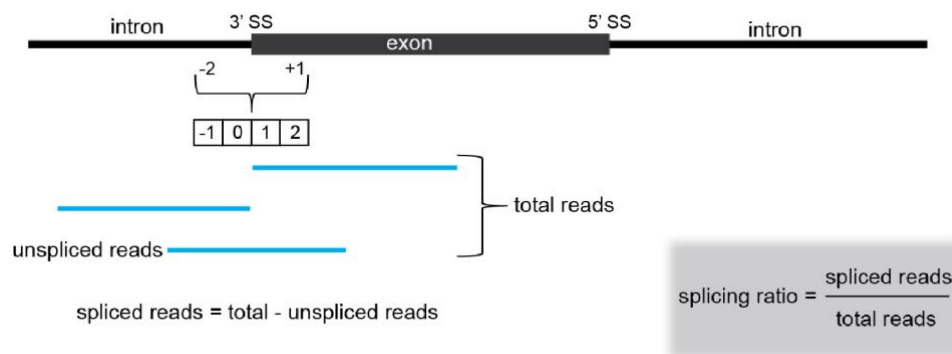


Figure 8.
Splicing ratio calculation.

Schematic representation of the definition of total, unspliced and spliced reads for 3' SS.

Because our preliminary analysis showed that the transcription levels were decreased upon Pla-B treatment, we considered only exon-based splice junctions that were covered by at least 30 total reads (spliced and unspliced) in Pla-B, SSA, or U2 AMO treated samples.

The number of exon-based splice junctions included for each treatment can be found below (**Table 4**).

Table 4. Number of exon-based splice junctions included for each treatment

Sample	Exon-based 5' SS	Exon-based 3' SS	First exons	Int. exons	Last exons
Pla-B 1 h	15,551	14,840	2,309	13,242	1,598
Pla-B 4 h	10,041	9,550	1,658	8,383	1,167
100 nM Pla-B 1 h	9,759	9,845	1,226	8,533	1,312
30 ng/mL SSA 1 h	16,459	16,178	1,989	14,4470	1,708
75 μ M U2 AMO 1 h	8,928	8,916	1,245	7,683	1,233

To confirm the use of this strategy, I conducted a parallel analysis based on CIGAR strings for TT-seq transcribed bases between DMSO or 1 μ M Pla-B treated cells for 1 h. CIGAR stands for compact idiosyncratic gapped alignment report, being the representation of spliced alignments found on aligned sequence files (sam or bam files). I performed this analysis with the spliceSites R package (Kaisers et al., 2017) which focuses on align-gaps (identified by the N CIGAR tag). First, a GapSites alignments table was created with readTabledBamGaps function from the spliceSites R package. Then, the GapSites were annotated using the getSplICEtable from refGenome R package (Kaisers, 2019). I removed GapSites with less than 5 alignments throughout all samples or which did not include a major isoform annotated transcript. A total of 648 major isoform transcripts were evaluated. I used the major isoform annotation to classify the GapSites as 5' SS or 3' SS splice junctions by matching the GapSite genomic location with exon start and/or end. A total of 7,081 5' SS junctions and 7,141 3' SS junctions were identified.

4. Identification of splicing affected and unaffected transcripts

To identify transcripts affected or unaffected by splicing, I performed a differential expression analysis on the previously calculated exon-based splice junctions (**Figure 9**).

I used the DESeq2 R package (1.24.0, RRID:SCR_015687) (Love et al., 2014) to test for significant changes of splicing ratios between conditions, with design \sim assay + condition + assay:condition, where assay corresponds to the number of spliced and total amount of spliced and unspliced reads, and condition corresponds to DMSO and Pla-B or DMSO and SSA or Ctr AMO and U2 AMO treatments.

The previously calculated sequencing depths σ_j were introduced as normalization size factors.

The analysis was focused on the 5' SS of first exons and 3' SS of second exons in order to flank the first intron. To select splicing affected exons, DESeq2 results (using altHypothesis = "less") were filtered with $\text{padj} < 0.05$ and $\log_2\text{FC} < -1.5$. To select splicing unaffected exons, DESeq2 results (using altHypothesis = "lessAbs" and lfcThreshold = 1.5) were filtered with $\text{padj} < 0.05$.

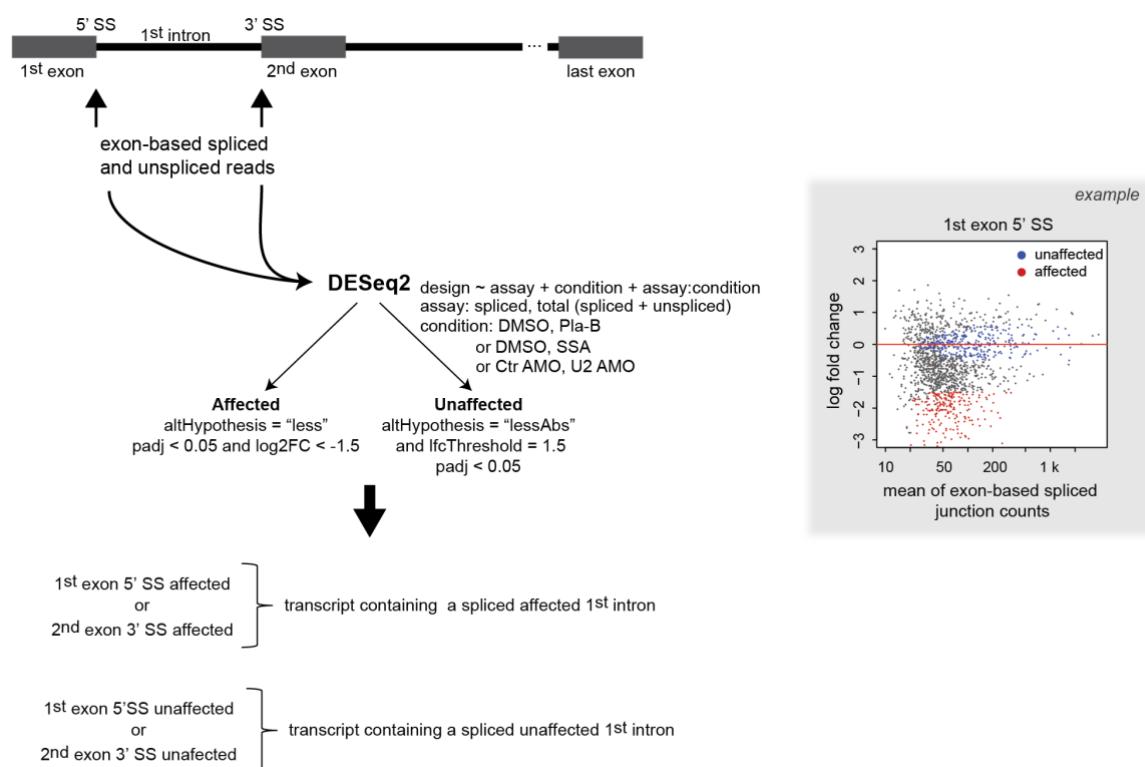


Figure 9. Schematic representation of the analysis for the selection of splicing affected and unaffected transcripts.

Exon-based spliced and unspliced read counts for the 5' SS corresponding to the first exon and the 3' SS of the second exon were subjected to DESeq2 analysis. Affected and unaffected exon-based junctions were defined. Transcripts containing a spliced affected or unaffected first intron were defined if included an affected or unaffected first exon on 5' SS and/or a second exon on 3' SS. The example for differential expression MA plot for the first exon 5' SS was created with exon-based spliced junctions calculate for DMSO or 1 μM Pla-B treatment for 1 h.

The analysis was conducted for 1 h and 4 h DMSO or 1 μM Pla-B treatment, 1 h DMSO or 30 ng/mL SSA treatment, and 1 h 75 μM Ctr AMO or 75 μM U2 AMO treatment. The number of exon-based junctions significantly affected can be found in **Table 5**, and the number of exon-based junctions significantly unaffected can be found in **Table 6**. Transcripts containing a splicing affected first intron were defined as containing a first exon 5' SS junction significantly affected and/or a second exon significantly 3' SS junction affected (**Table 5**). Transcripts containing a splicing unaffected first intron were defined as containing a first exon 5' SS junction significantly unaffected and/or a second exon significantly 3' SS junction unaffected (**Table 6**). There were not

identified transcripts containing a first exon 5' SS junction affected and second exon significantly 3' SS junction unaffected nor the other way around.

Table 5. Significantly splicing affected transcripts.

Number of exon-based splice junctions significantly affected for the 5' SS of the first exons and 3' SS of the second exon. Number of transcripts identified as having a first intron affected by splicing and number of exons included in the affected transcripts.

Sample	First exon 5' SS	Second exon 3' SS	Transcripts	Exons from affected transcripts
Pla-B 1 h	220	186	329	4,326
Pla-B 4 h	536	428	784	9,771
SSA 1 h	398	415	606	7,078
AMO 1 h	180	276	381	5,058

Table 6. Significantly splicing unaffected transcripts.

Number of exon-based splice junctions significantly unaffected for the 5' SS of first exons and 3' SS of the second exon. Number of transcripts identified as having a first intron unaffected by splicing and number of exons included in the unaffected transcripts.

Sample	First exon 5' SS	Second exon 3' SS	Transcripts	Exons from unaffected transcripts
Pla-B 1 h	316	287	355	2,756
Pla-B 4 h	187	157	194	1,082
SSA 1 h	402	372	449	2,842
AMO 1 h	147	145	177	1,164

5. Initiation frequency, pause duration, and elongation frequency calculation

To study genome-wide transcription kinetics we used data from TT-seq and mNET-seq experiments on cells treated with DMSO or 1 μ M Pla-B for 1 h. For elongation velocity calculation we also included a published mNET-seq data on HeLa cells for 4 h (Schlackow et al., 2017), GEO:GSE81662. I derived the values of productive initiation frequency from TT-seq and used a multi-omics approach integrating data from TT-seq and mNET-seq to calculate pause duration and elongation velocity.

5.1. RNA amount per cell

To infer the productive initiation, pause duration, and elongation frequency units, such as $\text{cell}^{-1} \text{min}^{-1}$, min, and bp min^{-1} , respectively, I calculated the RNA amount per cell factor k [cell^{-1}]. The calculation of k is possible using the known sequence and mixture of the used spike-ins and their molecular weight, and assuming a perfect RNA extraction.

The RNA amount per cell was calculated using the RNA spike-ins as previously described (Gressel et al., 2017, 2019).

The number of spike-in molecules per cell N [cell^{-1}] was calculated as

$$N = \frac{m}{M_n} N_A$$

with 50×10^{-9} g of spike-ins m , 5×10^7 cells n the Avogadro number N_A and the molar-mass of the spike-ins M calculated as

$$M = A_n \times 329.2 + (1 - \tau) \times U_n \times 306.2 + C_n \times 305.2 + G_n \times 345.2 + \tau \times U_n \times 322.26 + 159$$

Where A_n , U_n , C_n and G_n are the number of each respective nucleotide within each spike-in polynucleotide (**Table 3**) calculated based on spike-in sequence (**Supplementary information section 2**). τ is set to 0.1 in case of a labeled spike-in and 0 otherwise ($\tau \times U_n$ corresponds to the number of 4sU nucleotides). 159 corresponds to the molecular weight of a 5' triphosphate added to facilitate *in vitro* transcription by T7 polymerase.

Table 7. Number of each nucleotide (A, C, G, and U) and calculated molar-mass for each spike-in sequence

Spike-ins	A	C	G	U	M
Spike 2	354	127	207	297	318331.0
Spike 4	314	189	243	268	327586.2
Spike 5	286	207	258	264	327385.0
Spike 8	233	265	294	287	347569.7
Spike 9	226	272	276	263	333378.4
Spike 12	323	114	217	296	306827.0

The conversion factor to RNA amount per cell k [cell^{-1}] was then calculated as

$$k = \text{mean} \left(\text{median}_i \left(\frac{tb_i}{L_i \cdot N} \right) \right)$$

for all labeled spike-ins i with length L_i .

5.2. Detection of Pol II pause sites

Pause sites were defined as described (Gressel et al., 2017, 2019) for intron-containing major isoforms exceeding 10 kbp in length. Normalized sequence coverage lists generated for mNET-seq (see **section 2.2**) were used to address the Pol II signal strength. The pause site was calculated separately for DMSO and Pla-B treated samples using mNET-seq coverage rle lists within a window from the TSS to the end of the first exon.

For intronless genes, the pause site was calculated within a 100 bp window starting at the TSS.

The pause site m was determined via the maximization of the function

$$\rho_i = \max_m P_{im}$$

where ρ_i needed to exceed 5 times the median of the signal strength P_{im} for all non-negative antisense bias-corrected mNET-seq coverage values (Nojima et al., 2015).

5.3. Productive initiation frequency

The initiation frequency was calculated as previously described (Gressel et al., 2017, 2019) for intron-containing major isoforms exceeding 10 kbp in length.

The productive initiation frequency I_i refers to the transcribed bases by Pol II that initiated and exit the promoter-proximal pausing site. Therefore, I_i was calculated for each major isoform i for both DMSO and Pla-B treated samples using spike-in normalized TT-seq coverage means (see **coverage**

means on section 2.1) corrected for antisense bias and cross-contamination for major exons, excluding the first exon, as

$$I_i = \frac{1}{k} \cdot \frac{tb_i^{condition}}{t \cdot L_i}$$

with labeling time $t = 10$ and length L . For intronless genes, the productive initiation frequency was calculated within a window from the defined pause site on DMSO treated samples + 50 bp until the end of the transcript. Genes with a calculation window inferior to 100 bp were excluded from the analysis.

5.4. Pause duration

Using the mNET-seq information on Pol II occupancy and the TT-seq information on productive initiation frequency, we defined the pause duration $d_i^{condition}$ as the time a Pol II needs to pass through a defined pause window located around the pause site.

For each condition (DMSO and Pla-B treated samples), I calculated $d_i^{condition}$ in a window of ± 100 bp around the pause site m as previously described (Gressel et al., 2017, 2019) for intron-containing major isoforms exceeding 10 kbp in length.

For intronless genes, I calculated the pause duration $d_i^{condition}$ in a window of ± 50 bp around the pause site m .

A spike-in corrected and antisense bias-corrected mNET-seq mean coverage for a window of ± 100 bp around the pause site (or ± 50 bp for intronless genes) was calculated for both conditions as $P_i^{condition}$. The pause duration $d_i^{condition}$ was calculated using the previously calculated initiation frequency $I_i^{condition}$ and $P_i^{condition}$ as

$$d_i^{condition} = s \cdot \frac{\sum_{\pm 100} P_i^{condition}}{I_i^{condition}}$$

To adjust $d_i^{condition}$ to an absolute scale, Gressel *et al.* used a Pol II response window calculated by comparing CDK9 inhibition (Gressel et al., 2017) or heat shock (HS) derived elongation velocities (Gressel et al., 2019). For this analysis, I created a calibration factor s by comparing the median pause durations of our DMSO condition with the pause duration for expressed protein-coding genes obtained with control condition from the HS experiment on K562 cells (Gressel et al., 2019).

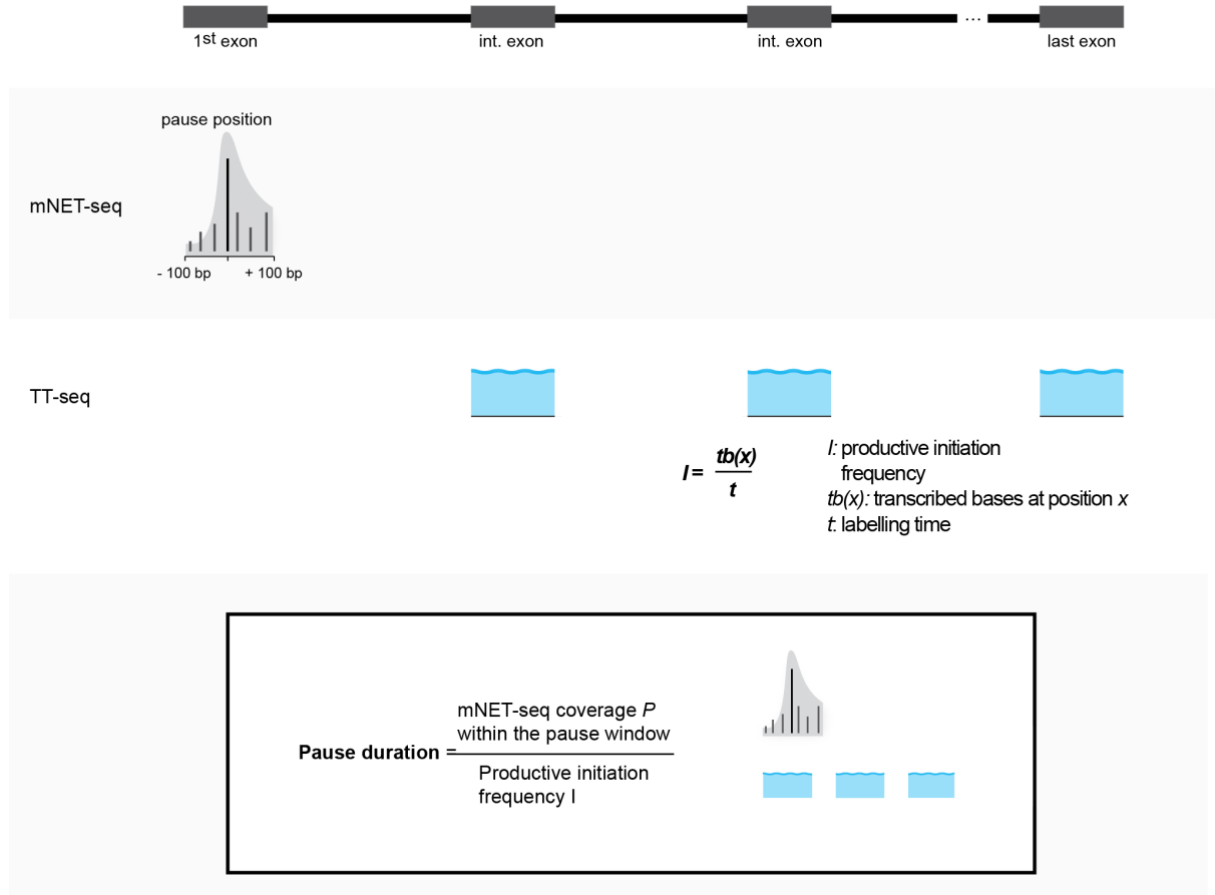


Figure 10. Schematic simplified representation of the pause duration calculation.

Pol II pause site identified using mNET-seq coverage rle lists within the first exon as the highest signal peak (being 5 times higher than the median). Initiation frequency calculated using TT-seq mean coverage on all exons except the first, so that it's contemplating the number transcribed bases by a productive elongating Pol II. Using the mNET-seq mean coverage around the pause site and the calculated initiation frequency we can infer the pause duration.

5.5. Elongation velocity estimation

The elongation velocity estimation is based on the principle that the number of transcribed bases from Pol II enzymes released into productive elongation can be inferred from TT-seq and the Pol II occupancy can be measured with mNET-seq. Therefore, we can estimate Pol II elongation velocity by the ratio between the productive initiation frequency calculated with TT-seq and the Pol II occupancy calculated with mNET-seq (see **Figure 21A**).

For each condition (DMSO and Pla-B treated samples), each feature (exon and intron) in all intron-containing major isoforms exceeding 10 kbp in length, or intronless genes, the elongation velocity $v_i^{condition}$ was calculated as

$$v_i^{condition} = \frac{s}{t \cdot k} \cdot \frac{\sum_{feature} tb_i^{condition}}{\sum_{feature} P_i^{condition}}$$

with calibration factor s (please see **section 5.4**). For meta-gene plotting, the formula was altered with a smoothing approach as follows:

$$v_i^{condition} = \frac{s}{t \cdot k \cdot 200} \cdot \frac{\sum_{\pm 100} tb_i^{condition}}{\sum_{\pm 5} w_i \cdot P_i^{condition}}$$

with calibration factor s and a weighting vector $w_i = (1,2,3,4,5,6,5,4,3,2,1)$.

Datasets for mNET-seq against CTD ser5P in HeLa cells upon 4 h DMSO and Pla-B treatments were obtained from (Schlackow et al., 2017), GEO:GSE81662. Empigen treatment during IP was not performed in this publicly available dataset.

6. Differential expression analysis

To investigate significant changes in gene expression between samples treated with DMSO or 1 μ M Pla-B for 1 h or 4 h, or DMSO or 30 ng/mL SSA or Ctr AMO and U2 AMO, I used the DESeq2 R package (1.24.0, RRID:SCR_015687) (Love et al., 2014) with design ~condition on TT-seq antisense bias-corrected HTSeq counts. The previously calculated sequencing depths σ_j using spike-ins were introduced as normalization size factors. I also performed this analysis on published TT-seq data from K562 cells subjected to 42 °C HS for 30 min (Gressel et al., 2019). The increased intronic signal due to the U2 snRNP inhibition could create a bias on differential expression analysis. Thus, I used the sum of the HTSeq exon counts for each transcript to perform the differential expression analysis.

To select upregulated transcripts, DESeq2 results were filtered with $padj < 0.05$ and $\log_2FC > 0$, downregulated transcripts were defined as having $padj < 0.05$ and $\log_2FC < 0$.

The number of defined up and downregulated transcripts for each treatment can be found below (**Table 8**).

Table 8. Number of up and downregulated transcripts upon DESeq2 analysis on TT-seq summed exons counts.

Sample	Upregulated	Downregulated
<i>DMSO vs Pla-B 1 h</i>	11	2,894
<i>DMSO vs Pla-B 4 h</i>	9	3,705
DMSO vs SSA 1 h	3	3,721
Ctr AMO vs U2 AMO	24	2,665
Control vs 42 °C HS 30 min (Gressel et al., 2019)	182	507

7. Gene Ontology enrichment analysis

I performed a GO enrichment analysis on the output from the differential expression analysis (see **section 6**) using goseq (1.36.0, RRID:SCR_017052) (Young et al., 2010). First, I created a binary vector defining genes upregulated (as 1) and all other genes used in the analysis (as 0). Then, I created a weighting table for each transcript, depending on its length, using the fitting the probability weighting function (pwf) function. Goseq was run on the pwf weighting table using the default method (Wallenius approximation). The top 10 GO categories overrepresented amongst DE transcripts were then selected. I conducted this analysis for all U2 snRNP inhibition TT-seq experiments and the published data for HS 30 min treatment (Gressel et al., 2019) as described in **section 6**.

8. Statistical analysis

To investigate the correlation between paired data I used the non-parametric Spearman's correlation coefficient test.

Statistical significance was investigated using the non-parametric Wilcoxon signed-rank test to compare treatment and control samples, and the non-parametric Wilcoxon rank-sum test to compare different groups of transcripts.

9. Alternative splicing analysis

I used rMATS (4.1.0, RRID:SCR_013049) (Shen et al., 2014) to search for differential alternative splicing events on RNA-seq data for 1h DMSO or 1 μ M Pla-B treatment. rMATS retrieved a total of 61,515 alternative splicing events. However, there were not identified any alternative splicing events with false discovery rate (FDR) $\leq 1\%$ and modulus of delta percent-spliced-in ($|\Delta\psi|$) $\geq 5\%$.

IV

RESULTS

Here are presented the main findings regarding the role of U2 snRNP inhibition in transcription.

The results presented in section 1 to 6 have been adapted from:

Caizzi, L. *, **Monteiro-Martins, S. *** et al. Efficient RNA polymerase II pause release requires U2 snRNP function. (2021). Mol. Cell 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

1. Rapid splicing inhibition in human cells

To specifically inhibit spliceosome assembly and function at an early stage genome-wide in human cells, we used the splicing inhibitor Pla-B (Kotake et al., 2007). Pla-B binds the spliceosomal U2 snRNP component SF3B (Cretu et al., 2018; Effenberger et al., 2016; Yokoi et al., 2011), which is essential for BS recognition (Gozani et al., 1996, 1998; Krämer et al., 1999), and inhibits usage of the BS (Cretu et al., 2018; Effenberger et al., 2016), thereby impairing co-transcriptional splicing (Drexler et al., 2020) and perturbing occupancy of genes with transcriptionally engaged Pol II (Nojima et al., 2015, 2018; Schlackow et al., 2017). Pla-B binds to a hinge of the SF3B subunit 1 (SF3B1), blocking SF3B1 in an open conformation that prevents the accommodation and stabilization of the U2 snRNA/BS duplex (Cretu et al., 2018). Consequently, Pla-B leads to a stalled ‘A-like’ spliceosomal complex containing U1 and U2 snRNPs, impairing further spliceosome assembly *in vitro* (Cretu et al., 2018; Effenberger et al., 2016). Consistent with these *in vitro* studies, Pla-B treatment arrests spliceosome assembly *in vivo*, causing cellular mobilization of the later-stage spliceosome component U5, but not of U1 and only partially of U2 (Tresini et al., 2015).

To monitor changes in RNA synthesis after spliceosome inhibition, we performed TT-seq and RNA-seq in K562 cells treated with 1 μ M Pla-B or with dimethyl sulfoxide (DMSO) as solvent control (**Figure 11A**). Both Pla-B and DMSO were added to cell media at 1:20,000 dilutions to prevent side effects of DMSO (Verheijen et al., 2019). RNA labeling with 4-thiouridine (4sU) was carried out for 10 minutes (min). After 1 h of Pla-B treatment, unspliced RT-PCR products and intron retention were observed (**Figure S2A-B**).

TT-seq and RNA-seq data were generated from two independent biological replicates. To define the treatment time, a preliminary analysis was performed on TT-seq shallow sequencing using 75 bp paired read for samples treated with 1 μ M Pla-B or DMSO for 30 min, 1 h, and 4 h. Total reads sequencing (RNA-seq) was performed for 1 h treatment. An initial quality control check was performed with FastQC (Andrews, 2010). On average, we obtained 40 million paired-end reads per

sample for TT-seq and 25 million for RNA-seq, with a minimum value of 49% of deduplicated reads for TT-seq and 39% for RNA-seq. The percentage of deduplicated reads defines the percentage of distinct reads, formed by singletons (reads found one single time) and distinct duplicated reads (distinct reads duplicated n times) (**Table S5**).

TT-seq and RNA-seq data were mapped against the human hg38 (GRCh38) genome assembly (**Methods section 2.1**). On average, the samples showed 27% duplicated mapped paired reads, which had no impact on results output (**Methods section 2.1**) (**Figure S2F-K, Table S5**). The experiments were highly reproducible (Spearman correlation of 0.9 and 1), with the mapped reads showing a uniform fragment size distribution (**Figure 11B-G**). Since 30 min Pla-B treatment did not show an evident impact on new RNA synthesis, it was not used for further analysis (**Figure S2C-G**). TT-seq and RNA-seq data were globally normalized using spike-ins (**Methods section 2.1**). Because splicing inhibitors may vary in activity at different introns and can cause intron retention or alternative splicing (Corrionero et al., 2011; Teng et al., 2017), we included in our analysis only major isoforms with 70 % or higher prevalence per gene in both DMSO and Pla-B conditions (**Methods section 1.2**). For further analysis, we considered 5,535 major isoforms of protein-coding genes that showed RPK ≥ 50 in TT-seq 1 h solvent control replicates.

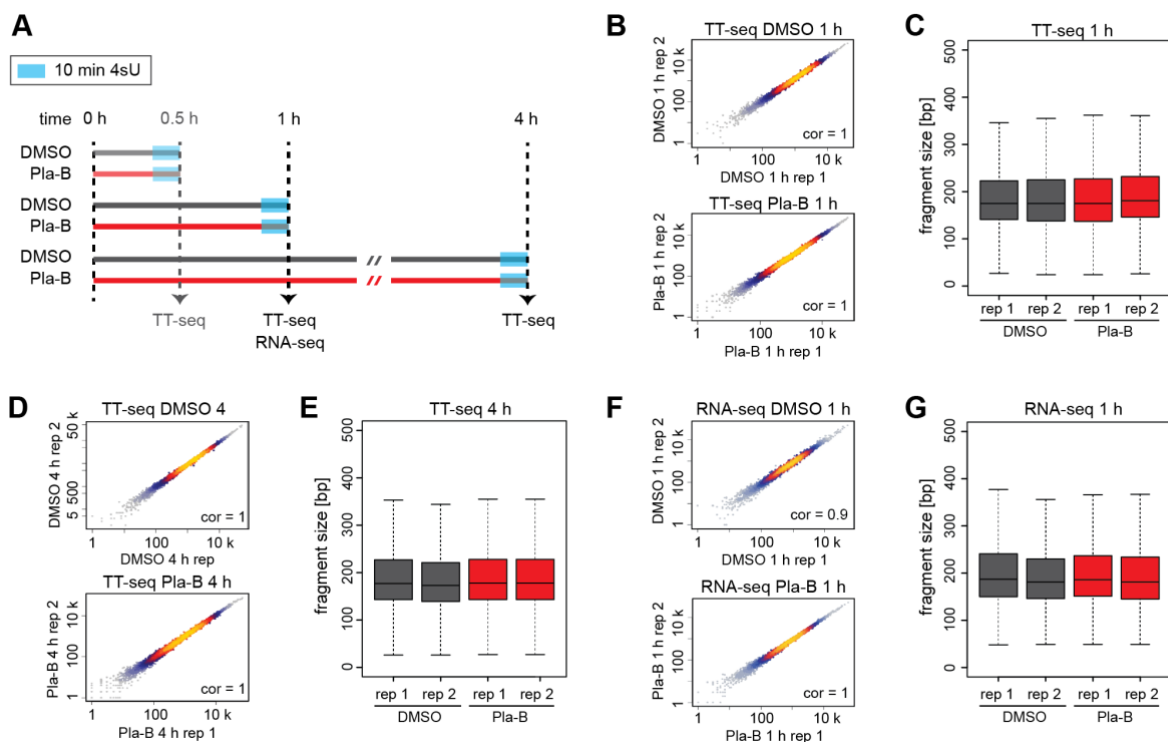


Figure 11. U2 snRNP inhibition with 1 μ M Pla-B

(A) Experimental design. TT-seq was performed on K562 cells after 30 min, 1 h, and 4 h treatments with DMSO (solvent control) or 1 μ M Pla-B using a 10 min 4sU labeling time. 30 min Pla-B treatment was not used for further analysis since this shorter treatment time did not show an evident impact on RNA synthesis. (B, D and F) Scatter plots comparing TT-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with DMSO or 1 μ M Pla-B for 1 h (B) and 4 h (D) and RNA-seq for 1 h treatment (F). Spearman correlation of 0.9 and 1. (C, E and G) Fragment size distribution for paired mapped reads for TT-seq upon 1 h (C) and 4 h (E) DMSO or 1 μ M Pla-B treatment and for RNA-seq samples treated with DMSO or 1 μ M Pla-B for 1 h (G). See also **Figure S2A-E**.

We analyzed our TT-seq data for the occurrence of reads spanning exon-intron junctions (5' SS) and intron-exon junctions (3' SS) (**Methods section 3**) for a total of 15,551 5' SS and 14,840 3' SS after 1 h treatment, and for 10,041 5' SS and 9,550 3' SS after 4 h treatment. The splicing ratio was calculated by dividing the number of spliced reads by the total amount of spliced and unspliced reads (**Methods section 3**). In accordance with our observations (**Figure S2A-B**), the splicing ratio decreased slightly but significantly (Wilcoxon signed-rank test, p -value $< 2.2 \times 10^{-16}$) already after 1 h of Pla-B treatment (**Figure 12A, Figure S3A-D**). Splicing inhibition was readily visible at the first introns (**Figure 12B, Figure S3B**), and was observed at all introns genome-wide upon 4 h of Pla-B treatment (**Figure 12C-D**), consistent with earlier observations (Effenberger et al., 2014; Nojima et al., 2015). These results show that splicing was rapidly inhibited genome-wide under our experimental conditions.

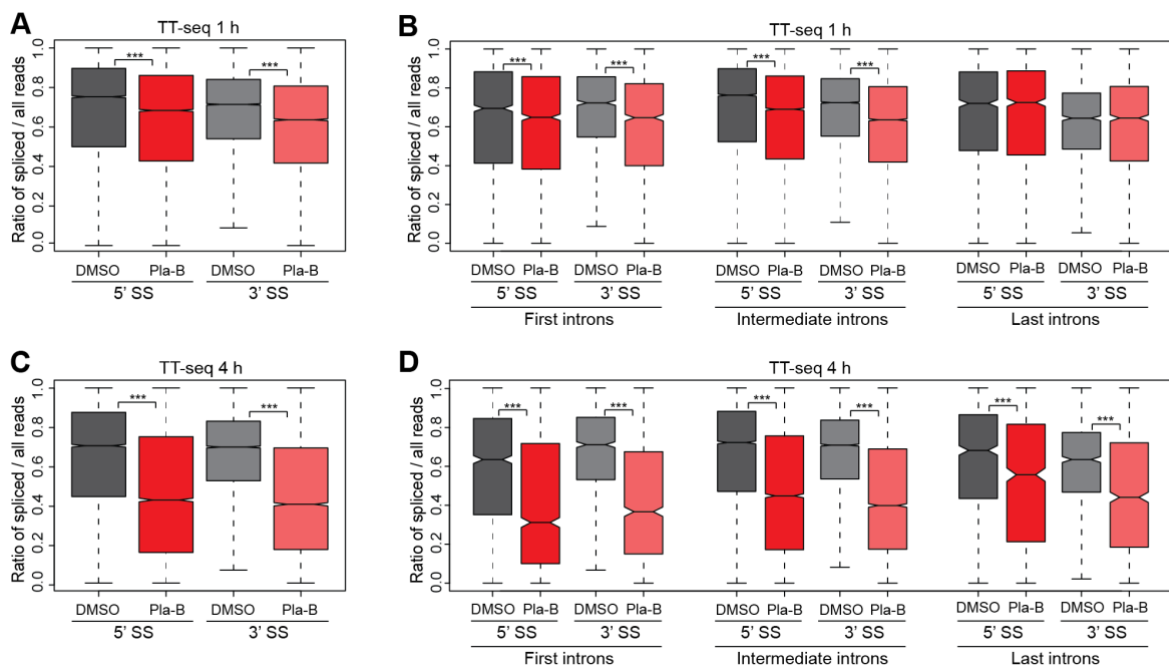


Figure 12. Rapid splicing inhibition in human cells

(A) Ratio of spliced reads over total unspliced and spliced reads upon 1 h of DMSO or 1 μ M Pla-B treatments. (B) Ratio of spliced reads over total unspliced and spliced reads upon 1 h of DMSO or 1 μ M Pla-B treatments in first, intermediate (non-first and non-last), and last introns. 4,698 major isoforms containing at least 4 exons were considered in the analysis. (C) Ratio of spliced reads over total unspliced and spliced reads as in (A) upon 4 h of DMSO or 1 μ M Pla-B treatments. (D) Ratio of spliced reads over total unspliced and spliced reads as in (B) upon 4 h of DMSO or 1 μ M Pla-B treatments. Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are not drawn. (***) p -value $< 2.2 \times 10^{-16}$ by Wilcoxon signed-rank test. See also **Figure S3**.

2. Inhibition of U2 snRNP function decreases RNA synthesis

Following these findings, we concentrated our analysis on studying the earliest effects of splicing inhibition. We sequenced each sample treated with Pla-B for 1 h at a depth of up to 100 million paired-end reads and used the concatenated shallow and deep sequencing reads for further analysis with an average of 134 million paired-end reads per sample (**Table S5**). Metagene analysis showed that TT-seq coverage across the gene body was strongly reduced (**Figure 13A-B, J**). The level of this reduction was not related to the expression level of genes (**Figure 13C**), arguing against a direct effect of Pla-B on Pol II transcription. TT-seq coverage was still strongly reduced when the experiment was repeated with only 100 nM Pla-B (**Figure S4, Table S5**), a concentration that was recently reported to globally diminish co-transcriptional splicing (Drexler et al., 2020). A less pronounced decrease in RNA synthesis was observed at intronless protein-coding genes (**Figure 13D**), supporting a splicing-independent role of SF3B1 (Van Nostrand et al., 2016; Wang et al., 2019). The drastic effects of Pla-B on RNA synthesis were only observed by TT-seq, and not by RNA-seq (**Figure 13E-F**), which only detected the splicing defect through an accumulation of unspliced introns (**Figure 13G**).

Metagene plots of the TT-seq signal upon 4 h of Pla-B treatment showed a change in the slope toward the 3' regions of genes (**Figure 13H-J**), indicating defects in Pol II elongation or processivity.

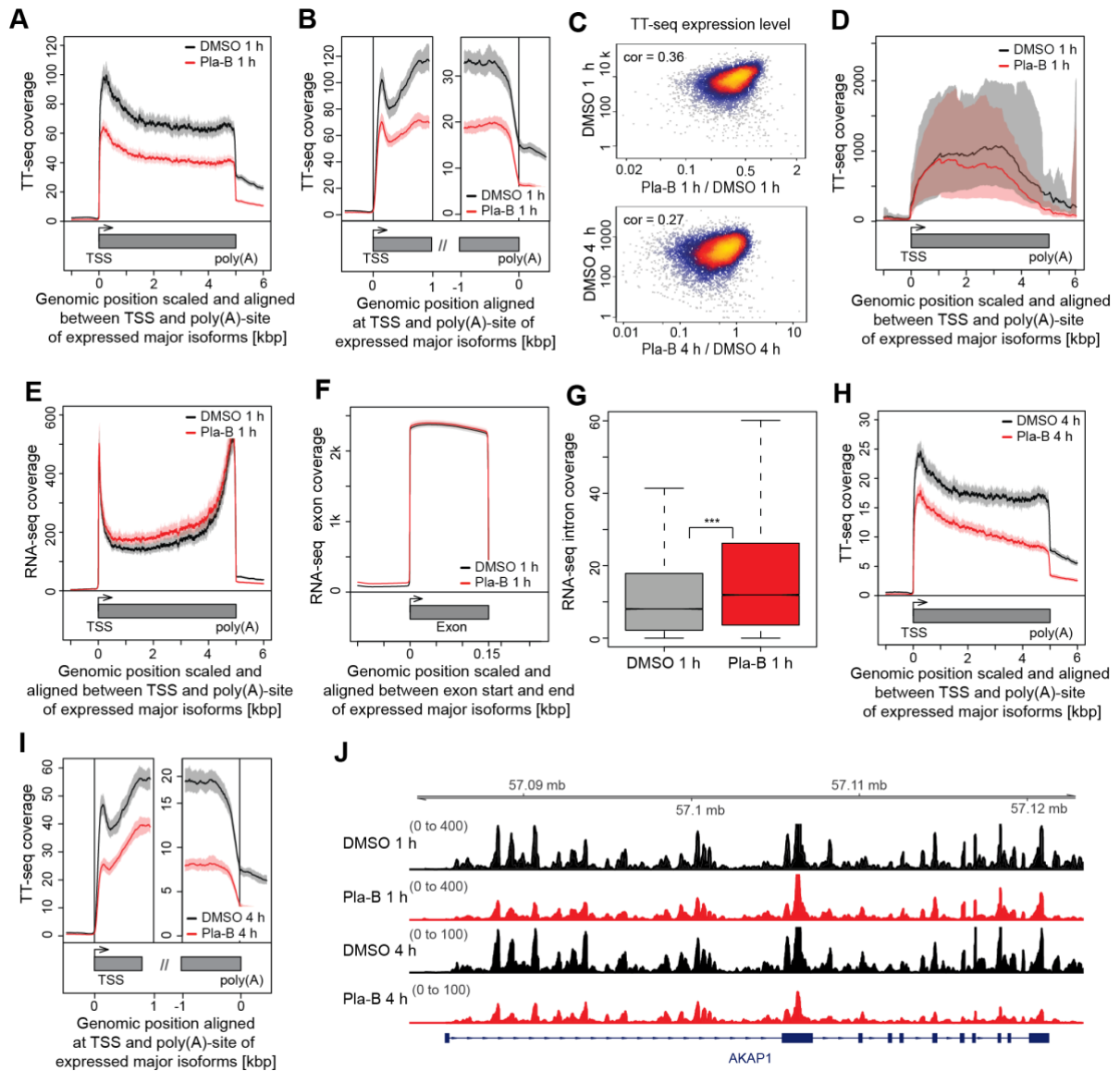


Figure 13. Inhibition of the U2 snRNP factor SF3B1 decreases RNA synthesis.

(A and H) Metagenesis comparing TT-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h (A) and 4 h (H). TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. See also **Figure S2H-I, K**. (B and I) Metagenesis focusing on the expression on TSS and poly(A)-site comparing DMSO and 1 μ M Pla-B treated cells for 1 h (B) and 4 h (I). TT-seq coverage was aligned at TSS and poly(A)-site for 5,465 major isoforms that exceed 1 kbp in length. (C) Scatter plot comparing DMSO TT-seq RPK and ratio of Pla-B to DMSO TT-seq RPK upon 1 h (top) and 4 h (bottom) of DMSO or 1 μ M Pla-B treatments. Spearman correlation of 0.36 for 1 h and 0.27 for 4 h. (D) Intronless genes meta-analysis comparing TT-seq signal between cells treated with DMSO and 1 μ M Pla-B for 1 h. TT-seq coverage was averaged for 51 intronless genes scaled between TSS and poly(A)-site. (E) Metagenesis comparing RNA-seq signal between cells treated with DMSO and 1 μ M Pla-B for 1 h. RNA-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. See also **Figure S2J**. (F) Meta-analysis comparing RNA-seq signal over exons between cells treated with DMSO and 1 μ M Pla-B for 1 h. RNA-seq coverage was averaged for 46,243 exons scaled between exon start and end site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). (G) Boxplot comparison of antisense bias-corrected counts on 52,593 introns between RNA-seq samples treated with DMSO and 1 μ M Pla-B for 1 h. Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are not drawn. (***) p-value < 2.2×10^{-16} by Wilcoxon signed-rank test. (J) TT-seq coverage track of AKAP1 gene upon 1 h or 4 h of DMSO or 1 μ M Pla-B treatments (GViz R package).

Since long human genes can take 1 h or longer to be transcribed, RNA synthesis activity in the 3' regions of long genes should be less affected upon 1 h of Pla-B inhibition, and more affected after 4 h. To investigate this, we divided the major isoforms into different length classes and checked the metagene profile of short (first quartile Q1) and long (fourth quartile Q4) for both 1 h and 4 h DMSO or 1 μ M Pla-B treatment (**Figure 14A-B**). In addition, we plotted the ratio of Pla-B to DMSO on the last exons after 1 h and 4 h of treatments. As expected, major isoforms longer than 52 kbp showed a smaller change in transcription after 1 h of treatment when compared with shorter major isoforms (**Figure 14A, C left**). In contrast, we observed that RNA synthesis is strongly defective in major isoforms of all lengths after 4 h of Pla-B treatment (**Figure 14B, C right**). Long major isoforms are more affected than short major isoforms after 4 h of treatment, explaining the slope towards the 3' region observed in the metagene plot (**Figure 13H**). To investigate whether the effects of Pla-B on transcription are related to splicing inhibition, we defined two groups of transcripts in which splicing of the first intron was affected or unaffected, based on DESeq2 (Love et al., 2014) analysis (**Methods section 4**). Indeed, RNA synthesis was significantly more decreased for genes where splicing of the first intron was affected for both 1 h and 4 h treatment (**Figure 14D**). Together, these results indicate that splicing has a direct positive effect on transcription.

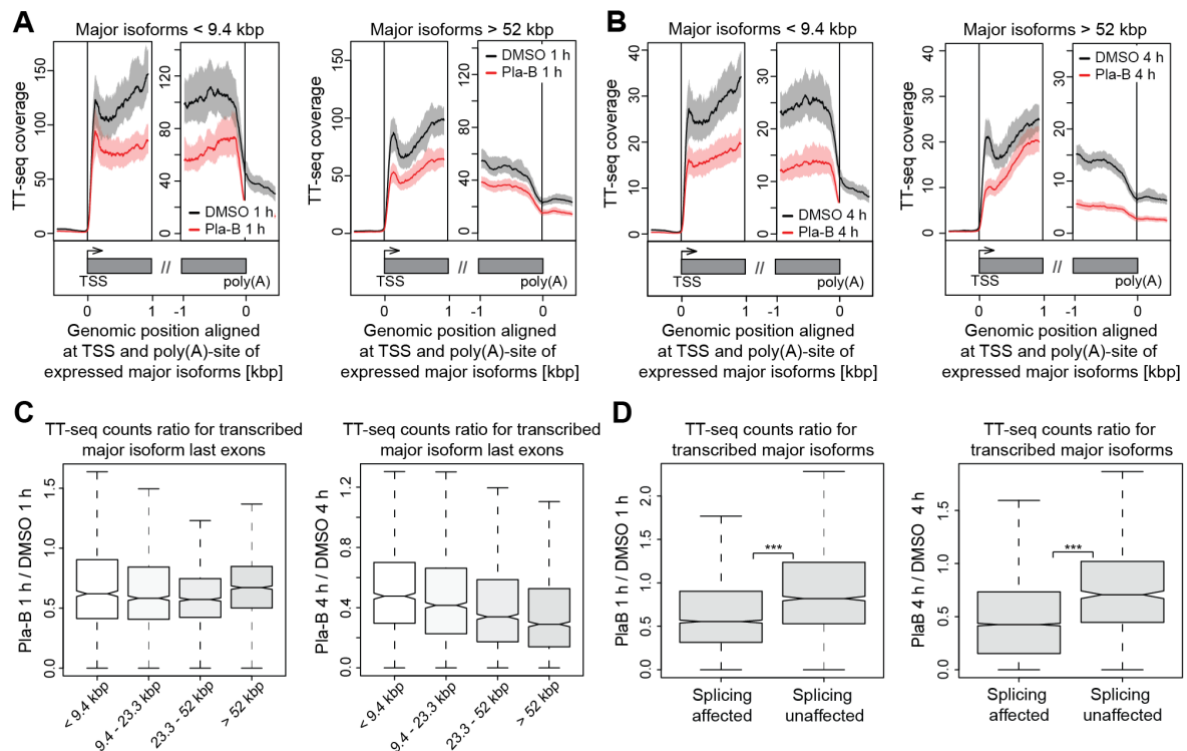


Figure 14. Pla-B effect on transcription is length-dependent and is related to splicing inhibition.

(**A** and **B**) Metagene analysis for 1 h (**A**) and 4 h (**B**) DMSO or 1 μ M Pla-B treated cells comparing major isoforms belonging to the Q1 (left) and Q4 (right) of length. TT-seq coverage aligned at TSS and poly(A)-site for 1,314 (Q1) and 1,384 (Q4) major isoforms that exceed 1 kbp in length (**C**) Ratio of 1 μ M Pla-B to DMSO antisense bias-corrected TT-seq read counts for last exons upon 1 h (left) and 4 h (right) treatments in four different major isoforms length quartiles. (**D**) Ratio of 1 μ M Pla-B to DMSO antisense bias-corrected TT-seq read counts for splicing affected and unaffected major isoforms upon 1 h (left) and 4 h (right) treatments. Outliers are not drawn. (***) p-value < 2.2×10^{-16} by Wilcoxon sum rank test.

To further support our findings, we performed TT-seq in K562 cells upon splicing inhibition using a different chemical inhibitor, SSA. Similar to Pla-B, SSA targets SF3B1 and inhibits splicing *in vitro* and *in vivo*, interfering with spliceosome assembly after A complex formation (Corrionero et al., 2011; Kaida et al., 2007; Martins et al., 2010; Roybal and Jurica, 2010). RT-PCR analysis showed increased intron retention already after 1 h of 30 ng/mL SSA (**Figure S5A**). TT-seq data were generated from two independent biological replicates, with the mapped reads showing a spearman correlation of 1 and with a uniform size distribution (**Figure 15A-B**). Between 44 and 53 million paired-end reads were obtained per sample, with a minimum value of 57% of deduplicated reads (**Table S5**). In order to investigate global changes on RNA-synthesis, the data was normalized with spike-ins (**Methods section 2.1**). Sequence reads were mapped against the hg38 (GRCh38) genome assembly, with a maximum of 29% duplicated paired reads per sample, which had no impact on results output (**Methods section 2.1**) (**Figure S5B, Table S5**). We found that SSA inhibited splicing genome-wide after 1 h of treatment (**Figure 15C**). As for the Pla-B treatment, metagene analysis showed that new RNA synthesis was impaired genome-wide (**Figure 15D-E, G**) and that the decrease in RNA synthesis was significantly higher for genes where splicing was more affected (**Methods section 4**) (**Figure 15F**).

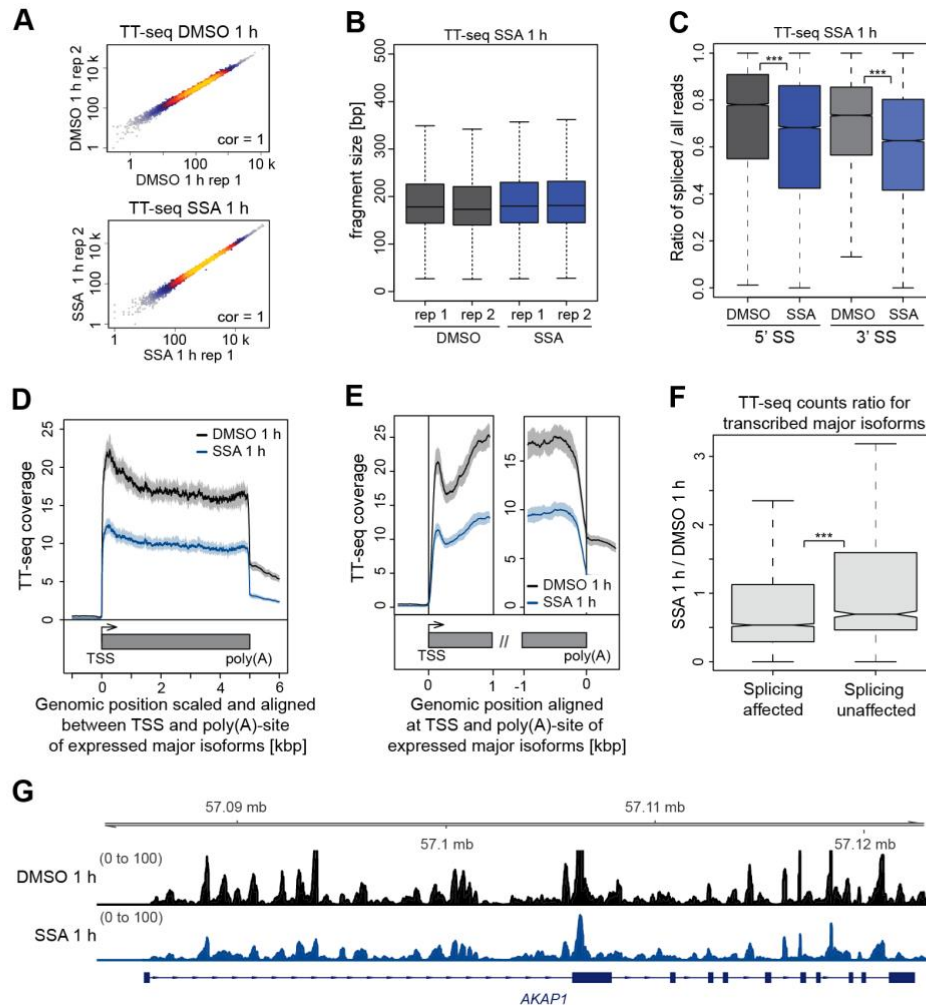


Figure 15. SSA treatment decreases new RNA synthesis.

(A) Scatter plots comparing TT-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with DMSO or 30 ng/mL SSA for 1 h. Spearman correlation of 1. (B) Fragment size distribution for paired mapped reads for TT-seq upon 1 h DMSO or 30 ng/mL SSA treatment. (C) Ratio of spliced reads over total unspliced and spliced reads upon 1 h DMSO or 30 ng/mL SSA. See also **Figure S5A**. (D) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 30 ng/mL SSA for 1 h. TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. See also **Figure S5B**. (E) Metagene analysis focusing on the expression on TTS and poly(A)-site comparing DMSO and 30 ng/mL SSA treated cells for 1 h. TT-seq coverage was aligned at TSS and poly(A)-site for 5,465 major isoforms that exceed 1 kbp in length. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). (F) Ratio of 30 ng/mL SSA to DMSO antisense bias-corrected TT-seq read counts for splicing affected and unaffected major isoforms upon 1 h. Outliers are not drawn. (***) p -value $< 2.2 \times 10^{-16}$ by Wilcoxon sum rank test. (G) TT-seq coverage track of *AKAP1* gene upon 1 h DMSO or 30 ng/mL SSA (GViz R package).

To exclude that the decrease in transcription after Pla-B or SSA treatments was due to a stress response pathway triggered by the chemical compounds, we performed TT-seq after treating cells with an antisense morpholino oligo targeting the U2 snRNA (U2 AMO) or with a control oligo (Ctr AMO). We used a U2 AMO that specifically blocks the RNA-RNA interactions between U2 snRNAs and pre-mRNA (Matter and König, 2005). RT-PCR showed that splicing was inhibited after treatment of cells with 75 μ M U2 AMO for 1 h (**Figure S6A**), indicating rapid splicing inhibition. Between 39 and 49 million paired-end reads were obtained per sample, with a minimum value of 49% of deduplicated reads (**Table S5**). TT-seq data were generated from two biological replicates after treatment with U2 AMO or Ctr AMO and globally normalized using spike-ins

(**Methods section 2.1**). Sequence reads were mapped against the hg38 (GRCh38) genome assembly, with a maximum of 35% duplicated paired reads per sample, which had no impact on results output (**Methods section 2.1**) (**Figure S6B**). The data were highly reproducible, with a Spearman correlation of 1, and a uniform fragment size distribution (**Figure 16A-B**). The data showed that 1 h treatment with U2 AMO causes a decrease in new RNA synthesis across the gene body (**Figure 16D-E, 5G**) and inhibits splicing genome-wide (**Figure 16C**). Again, the decrease in newly synthesized RNA was significantly higher for genes where splicing was more affected (**Methods section 4**) (**Figure 16F**).

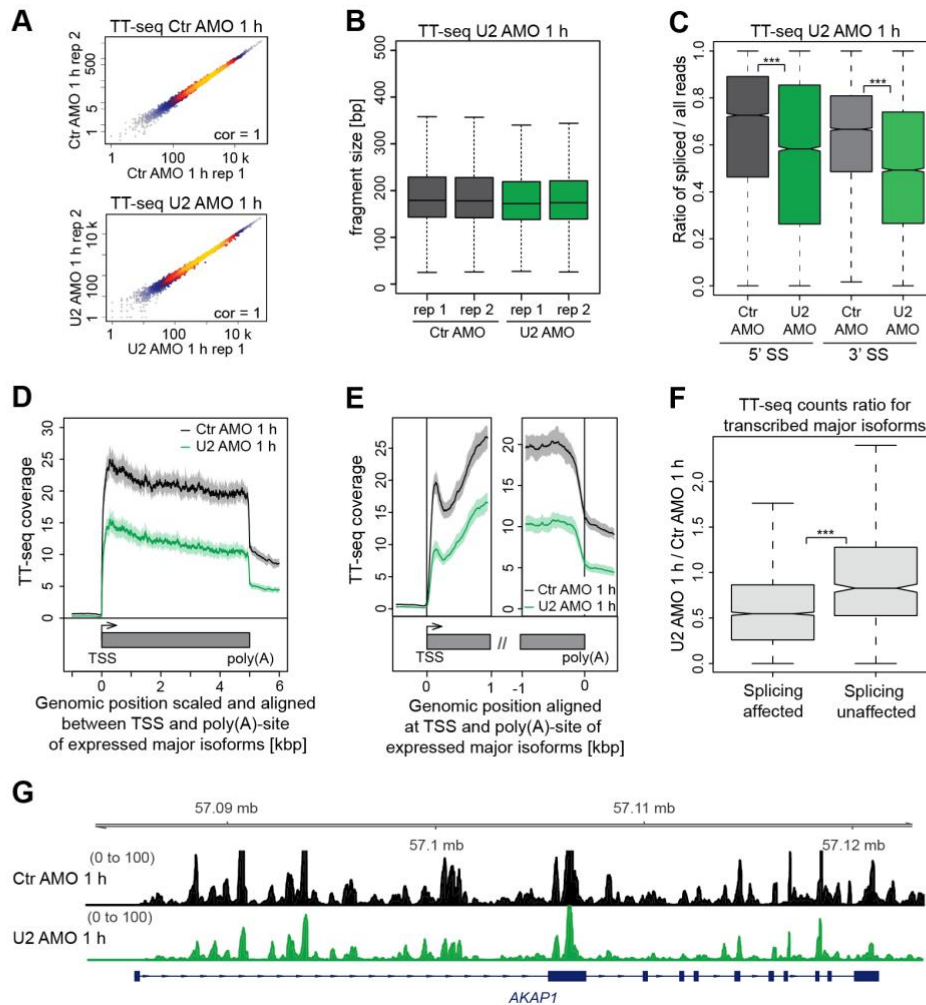


Figure 16. AMO U2 treatment decreases new RNA synthesis.

(**A**) Scatter plots comparing TT-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with 75 μM Ctr AMO or 75 μM U2 AMO for 1 h. Spearman correlation of 1. (**B**) Fragment size distribution for paired mapped reads for TT-seq upon 1 h 75 μM Ctr AMO or 75 μM U2 AMO. (**C**) Ratio of spliced reads over total unspliced and spliced reads upon 1 h 75 μM Ctr AMO or 75 μM U2 AMO. See also **Figure S6A**. (**D**) Metagene analysis comparing TT-seq signal between cells treated with 75 μM Ctr AMO or 75 μM U2 AMO for 1 h. TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. **Figure S6B**. (**E**) Metagene analysis focusing on the expression on TTS and poly(A)-site comparing DMSO and 30 ng/mL SSA treated cells for 1 h. TT-seq coverage was aligned at TSS and poly(A)-site for 5,465 major isoforms that exceed 1 kbp in length. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). (**F**) Ratio of 75 μM Ctr AMO to 75 μM U2 AMO (antisense bias-corrected TT-seq read counts for splicing affected and unaffected major isoforms upon 1 h. Outliers are not drawn. (***) p-value $< 2.2 \times 10^{-16}$ by Wilcoxon sum rank test. (**G**) TT-seq coverage track of *AKAP1* gene upon 1 h 75 μM Ctr AMO or 75 μM U2 AMO (GViz R package).

3. Inhibition of U2 snRNP function impairs early transcription elongation

Our results indicated that the observed reduced RNA synthesis activity resulted from a decreased number of Pol II enzymes entering productive and processive elongation of genes. To investigate this, we performed mNET-seq of total Pol II in K562 cells treated for 1 h with 1 μ M Pla-B or DMSO. The detergent emipigen was used to prevent co-IP of spliceosome-associated RNAs (Nojima et al., 2018a; Schlackow et al., 2017). mNET-seq data were generated from two independent biological replicates. To introduce a reference (spike-ins) normalization were included approximately 5% *S. cerevisiae* fragmented mRNA after IP. We obtained, on average 170 to 214 million reads per sample, with a minimum value of 47% of deduplicated reads (**Table S6**). mNET-seq reads were mapped to a combined human hg38 (GRCh38) and *S. cerevisiae* sacCer3 (R64-1-1) genome (**Methods section 2.2**). Between 114 to 138 million reads were uniquely mapped to the combined human and yeast genome. From these reads, approximately 20% mapped to the *S. cerevisiae* genome and 80% to the human genome (**Table S6**). A metagene analysis over *S. cerevisiae* expressed transcripts before and after normalization confirms the use of this strategy (**Figure S7A-C**). Data were highly reproducible, with the mapped reads showing a Spearman correlation of 1 and a uniform fragment size distribution (**Figure 17A-B**). On average, the samples showed 54% duplicated paired reads, which had no impact on results output (**Methods section 2.2**) (**Figure S7D, Table S6**).

Metagene plots of the mNET-seq signals over major isoforms revealed a sharp peak of Pol II occupancy just downstream of the TSS, indicative of promoter-proximally paused Pol II (**Figure 17C**, black line). Upon 1 h of Pla-B treatment, the peak for paused Pol II strongly increased, and the signal in the gene body decreased (**Figure 17C**, red line, **Figure 17D**). We observed the same trend when we used published mNET-seq data obtained for ser5-phosphorylated (ser5P) Pol II in HeLa cells after 4 h of Pla-B treatment (Nojima et al., 2015) (**Figure 17E**). At intronless genes, Pol II occupancy increased also, but to a lower extent (**Figure 17F**). These changes in the mNET-seq signal indicate an increase in promoter-proximal Pol II pausing and a defect in the release of Pol II into elongation.

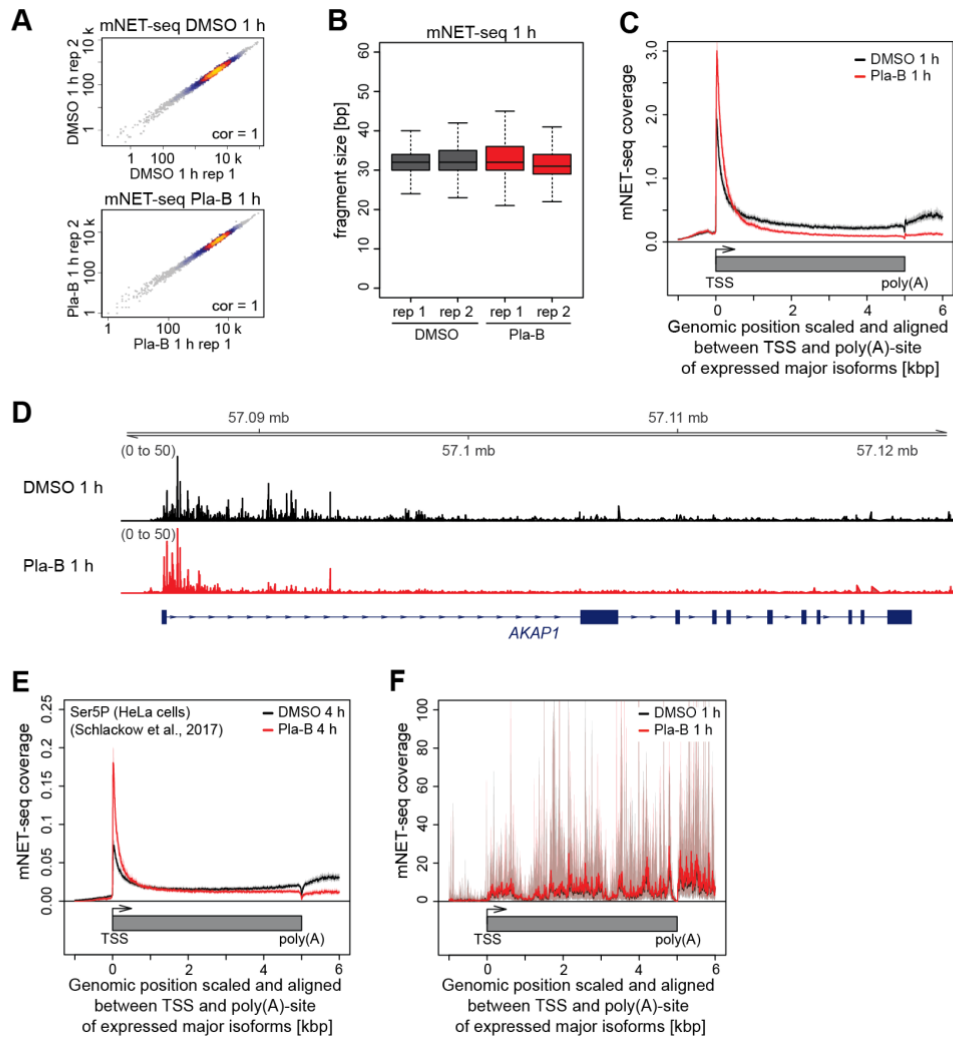


Figure 17. Inhibition of U2 snRNP function impairs early transcription elongation.

(A) Scatter plots comparing mNET-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with DMSO and 1 μ M Pla-B for 1 h. Spearman correlation of 1. (B) Fragment size distribution for paired mapped reads for mNET-seq upon 1 h DMSO or 1 μ M Pla-B treatment. (C) Metagene analysis comparing the mNET-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h. mNET-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. See also **Figure S7D**. (D) mNET-seq coverage track of *AKAP1* gene upon 1 h of DMSO or 1 μ M Pla-B treatments (GViz R package). (E) Metagene analysis comparing the averaged mNET-seq signal for CTD ser5P between HeLa cells treated with DMSO and 1 μ M Pla-B for 4 h. mNET-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Datasets for mNET-seq against CTD ser5P in HeLa cells upon 4 h DMSO and 1 μ M Pla-B treatments were obtained from (Schlackow et al., 2017). (F) Intronless genes meta-analysis comparing mNET-seq signal between cells treated with DMSO control and 1 μ M Pla-B for 1 h. TT-seq coverage was averaged for 51 intronless genes scaled between TSS and poly(A)-site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap).

A more detailed analysis of the mNET-seq data showed that Pol II accumulated not only in the promoter-proximal region but also further downstream in the 5' region of the first intron (**Figure 18A-B**). At about 1.3 kbp downstream of the first 5' SS, the mNET-seq signal however dropped below the level observed without Pla-B treatment (**Figure 17C**, **Figure 18B**). These results are consistent with a strong defect in productive and processive Pol II elongation. Changes in Pol II occupancy were by far strongest at the beginning of the first intron (**Figure 18B**) and were independent of the length of the first intron (**Figure 18C**). These results argue against a simple

model that Pol II is paused around the 3' end of the first intron due to the accumulation of elongating Pol II complexes upstream all the way back to the promoter. Instead, our findings are consistent with a role of the functional spliceosome in facilitating the release of promoter-proximally paused Pol II into active elongation.

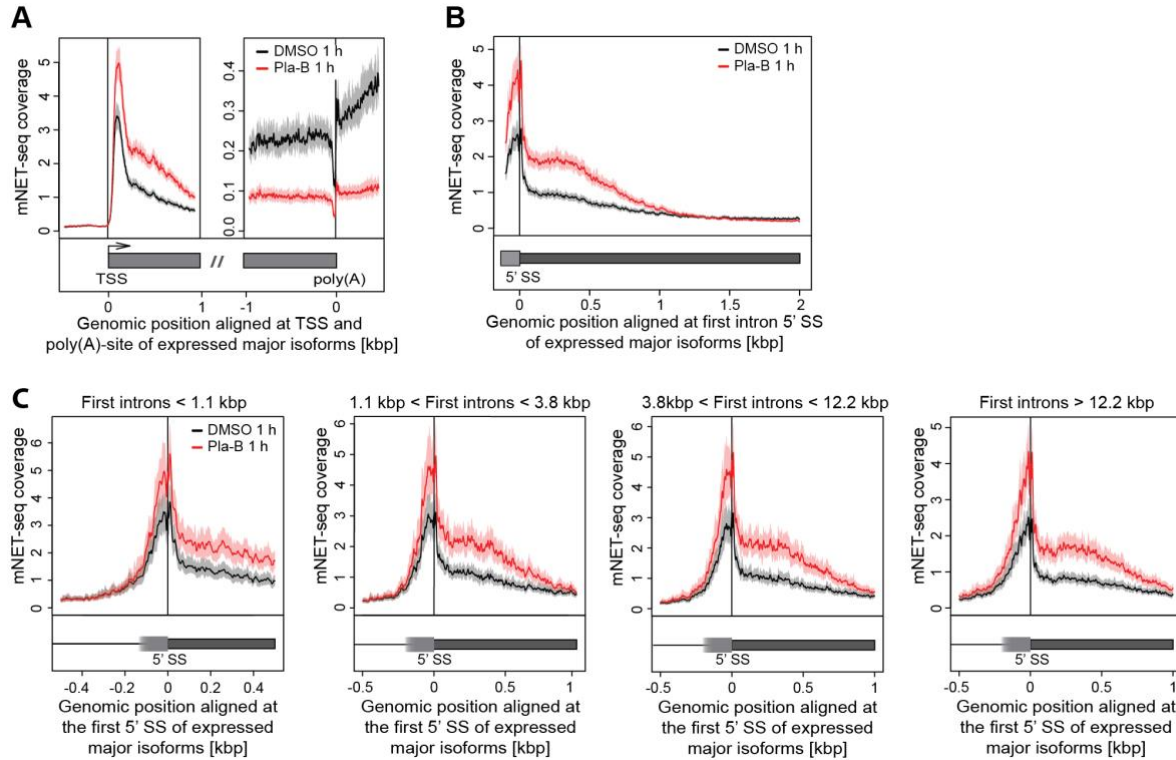


Figure 18. mNET-seq data reveals a defect in productive and processive Pol II elongation upon Pla-B treatment.

(A) Metagenesis analysis comparing the mNET-seq signal upon 1 h DMSO or 1 μ M Pla-B treated cells aligned at TSS (left) and poly(A)-site (right). mNET-seq signal was averaged for 5,465 major isoforms that exceed 1 kbp in length. (B) Metagenesis analysis comparing mNET-seq signal upon 1 h DMSO or 1 μ M Pla-B treated cells aligned at the first 5' SS. mNET-seq signal was averaged for 3,449 first introns that exceed 2 kbp in length. (C) First intron meta-analysis comparing the averaged mNET-seq signal between cells treated with DMSO and 1 μ M Pla-B for 1 h. First introns were divided into 4 quartiles based on the first intron length. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap).

4. Inhibition of U2 snRNP function increases Pol II pause duration

We next investigated whether Pla-B treatment increased the duration of promoter-proximal pausing by Pol II. We defined Pol II pause sites as the positions of the major mNET-seq peaks with a signal of at least five times above the median coverage in the first exon of the intron-containing major isoforms or in the first 100 bp of the intronless genes (**Methods section 5.2**, (Gressel et al., 2017)). The distribution of genomic positions of these pause sites was not altered upon Pla-B treatment (**Figure 19A**). We then used our previously described approach that combines multi-omics with kinetic modeling (Gressel et al., 2017, 2019) to estimate the pause duration d as the ratio of the mNET-seq signal over the productive initiation frequency I in a pause window of ± 100 bp (intron-containing major isoforms) or ± 50 bp (intronless genes) around the Pol II pause site (**Methods section 5.3 and 5.4**).

This quantitative kinetic analysis showed that the pause duration d was strongly increased genome-wide after Pla-B treatment (**Figure 19B**). Thus, Pol II pauses longer in the promoter-proximal region when U2 snRNP function is compromised. Further, the productive initiation frequency I decreased after Pla-B treatment (**Figure 19C**), confirming the transcription defect. Plotting d and I for all expressed intron-containing major mRNA isoforms exceeding 10 kbp in length showed a shift to longer pause durations and to lower productive initiation frequencies upon 1 h of Pla-B treatment (**Figure 19D-E**). These observations can be explained by the existence of the previously described ‘pause-initiation limit’ in mammalian cells that sets a limit to the maximum Pol II initiation frequency at a given pause duration (Gressel et al., 2017; Shao and Zeitlinger, 2017). Consistent with our previous results, splicing affected genes showed significantly more increased pause durations and more decreased initiation frequencies than splicing unaffected genes or intronless genes (**Methods section 4**) (**Figure 19F-G**, p -values < 0.001 , Wilcoxon rank-sum test). In summary, Pla-B treatment leads to an increase of the duration of promoter-proximal Pol II pausing, and to lower productive initiation frequencies at many genes.

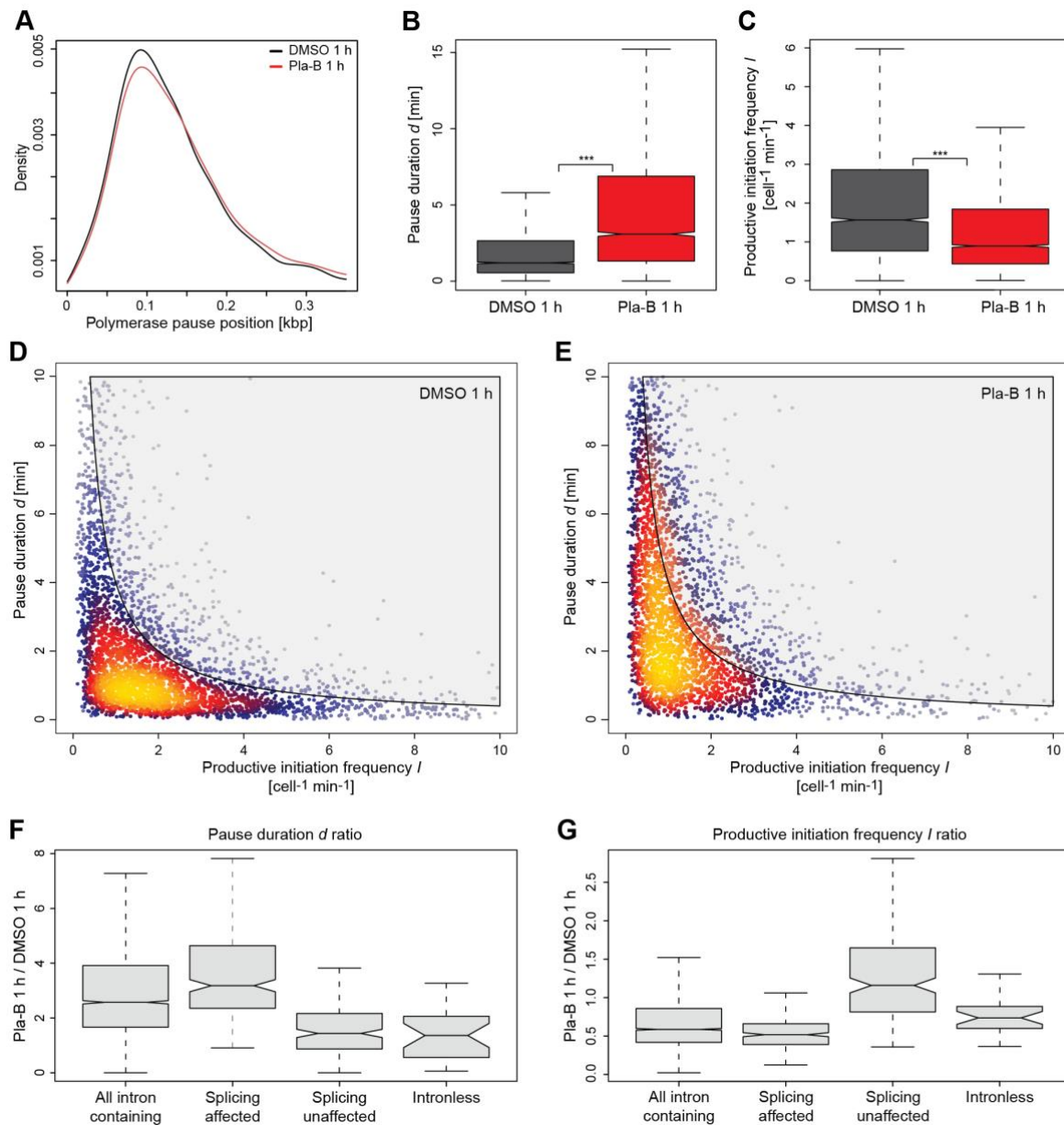


Figure 19. Inhibition of U2 snRNP function increases Pol II pause duration.

(A) Polymerase pause position calculated for 3,626 intron-containing major isoforms that exceed 10 kbp in length and had an identified pause position (**Methods section 5.2**) for both DMSO and 1 μ M Pla-B treated samples. (B and C) Pause duration d (B) and productive initiation frequency I (C) calculated for 3,636 intron-containing major isoforms that exceed 10 kbp in length and had an identified polymerase pause position (**Methods section 5**) for both 1 h DMSO and 1 μ M Pla-B. Outliers are not drawn. (***) p-value $< 2.2 \times 10^{-16}$ by Wilcoxon signed-rank test. (D and E) Correlation between the productive initiation frequency I and pause duration d upon 1 h DMSO (D) and 1 μ M Pla-B (E) treated cells. The grey shaded area depicts impossible combinations of I and d considering the footprint of the transcribing elongation complexes a limit distance between the active sites of initiating and paused Pol II (Ehrensberger et al., 2013). (F and G) Ratio of Pla-B to DMSO for pause duration d (F) and productive initiation frequency I (G) upon 1 h of DMSO or 1 μ M Pla-B treatments calculated for 3,636 intron-containing, 329 splicing affected, 355 splicing unaffected, and 29 intronless transcripts. Note that not all genes fit the criteria for pausing and productive initiation frequency calculation (**Methods section 5**).

5. Inhibition of U2 snRNP function impairs recruitment of pause release factor P-TEFb

We next investigated the reasons for the observed increase in the duration of promoter-proximal Pol II pausing. Release of paused Pol II into active elongation requires P-TEFb, which contains the kinase CDK9 and the Cyclin T1 (CycT1) (Fujinaga et al., 2004; Marshall and Price, 1995; Marshall et al., 1996; Vos et al., 2018b; Wei et al., 1998). We, therefore, investigated whether increased Pol II proximal-promoter pausing after Pla-B treatment may be accompanied by a defect in P-TEFb recruitment. We used ChIP-seq to measure genome occupancy with total Pol II and with the P-TEFb subunit CycT1 in K562 cells treated for 1 h with 1 μ M Pla-B or DMSO. ChIP-seq data were generated from two independent biological replicates. To correct for global changes in DNA levels upon treatments, *D. melanogaster* S2 sheared cross-linked-chromatin was added before IP (**Methods section 2.3**). ChIP-seq data were mapped against a combined human hg38 (GRCh38) and *D. melanogaster* dm6 (BDGP6.28), genome (**Methods section 2.3**). Data were highly reproducible (Spearman correlation of 0.99) with a homogeneous fragment size distribution (**Figure 20A-B, Figure S8A-B**). On average, 34 million paired-end reads were obtained per sample with a minimum value of 92% of deduplicated reads (**Table S7**). From these reads, on average 99.7% mapped to human genome and 0.3% mapped to *D. melanogaster* genome (**Table S8**). A maximum of 8% duplicated mapped paired reads were found, which had no impact on the results output (**Figure S8E-F, Table S7**).

Analysis of these ChIP-seq data showed that Pla-B treatment resulted in increased Pol II occupancy just downstream of the TSS of protein-coding genes (**Figure 20C**). This is consistent with the increase in promoter-proximally paused Pol II that we had observed by mNET-seq (**Figure 17C**). In contrast, occupancy with the P-TEFb subunit CycT1 was decreased in the same region (**Figure 20D-E**), indicating that the recruitment of P-TEFb to promoter-proximally paused Pol II is impaired. Consistent with our previous evidence, intronless genes showed similar effects, but they were much less pronounced (**Figure 20F-G**). These data indicate that U2 snRNP inhibition impairs P-TEFb recruitment to the promoter-proximal region, thereby impairing the release of paused Pol II, increasing Pol II pause duration, and decreasing the productive initiation frequency at affected genes.

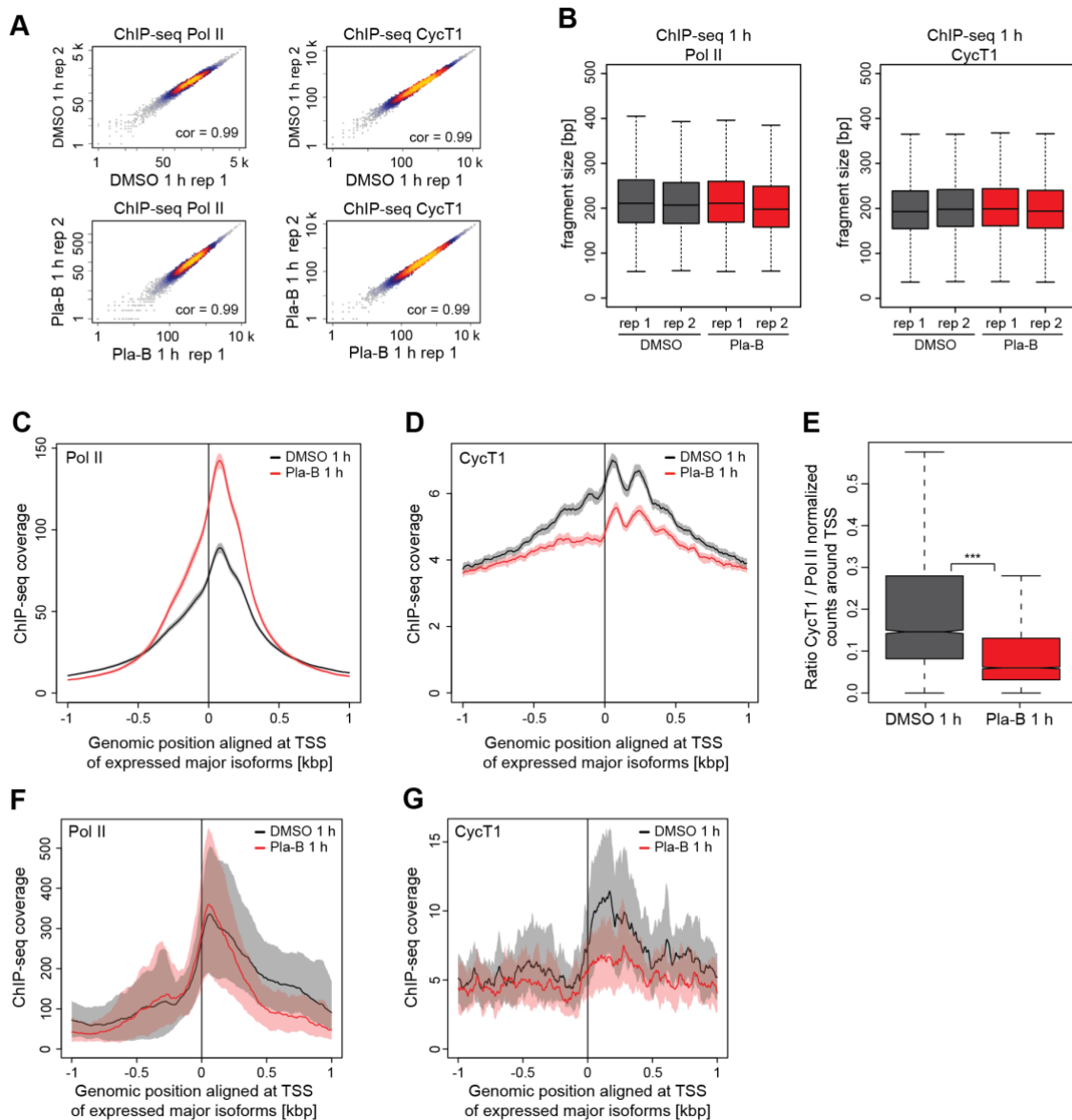


Figure 20. Inhibition of U2 snRNP function impairs P-TEFb recruitment.

(A) Scatter plots comparing Pol II (left) and CycT1 (right) ChIP-seq replicates using counts for major isoforms for samples treated with DMSO and 1 μ M Pla-B. Spearman correlation of 0.99. See also **Figure S8A**. (B) Fragment size distribution for paired mapped reads for Pol II (left) and CycT1 (right) ChIP-seq upon 1 h DMSO or 1 μ M Pla-B treatment. See also **Figure S8B**. (C and D) Metagenesis analysis comparing ChIP-seq for Pol II (C) and CycT1 (D) between cells treated with DMSO and 1 μ M Pla-B for 1 h. ChIP-seq coverage was aligned at the TSS and averaged for 5,465 major isoforms that exceed 1 kbp in length. See also **Figure S8C-F**. (E) Ratio of CycT1 to Pol II normalized counts around TSS (from -100 to +200 bp) comparing 1 h DMSO and 1 μ M Pla-B treated samples. Outliers are not drawn. (***) p -value $< 2.2 \times 10^{-16}$ by Wilcoxon signed-rank test. (F and G) Intronless genes meta-analysis comparing ChIP-seq signal between cells treated with DMSO and 1 μ M Pla-B for 1 h. ChIP-seq coverage for Pol II (F) and CycT1 (G) was averaged for a total of 51 intronless genes that exceed 1 kbp in length. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap).

6. Inhibition of U2 snRNP function alters Pol II elongation velocity

To investigate in more detail how U2 snRNP inhibition influences Pol II elongation downstream of the promoter-proximal region and downstream of the first intron, we estimated the Pol II elongation velocity at all major isoforms exceeding 10 kbp in length. We obtained the elongation velocity v from the ratio of the number of Pol II enzymes released into elongation, as measured by TT-seq at each genomic location, over the Pol II occupancy, as measured by mNET-seq (**Methods section 5.5**) (**Figure 21A**). Metagene analysis of the cells treated with DMSO showed that the elongation velocity is low in the region close to the TSS, and then increases, reaching a maximum of around 3 kbp/min (**Figure 21B**, black line). This agrees with previous estimates for Pol II velocity (Fuchs et al., 2014b; Gressel et al., 2017; Jonkers et al., 2014; Saponaro et al., 2014; Veloso et al., 2014), supporting our approach.

After 1 h of Pla-B treatment, we observed a decrease in the elongation velocity in the promoter-proximal region and at the beginning of the first intron, whereas regions further downstream were less affected (**Figure 21B-C**). This is consistent with an unaltered elongation velocity for Pol II enzymes that were released before Pla-B treatment.

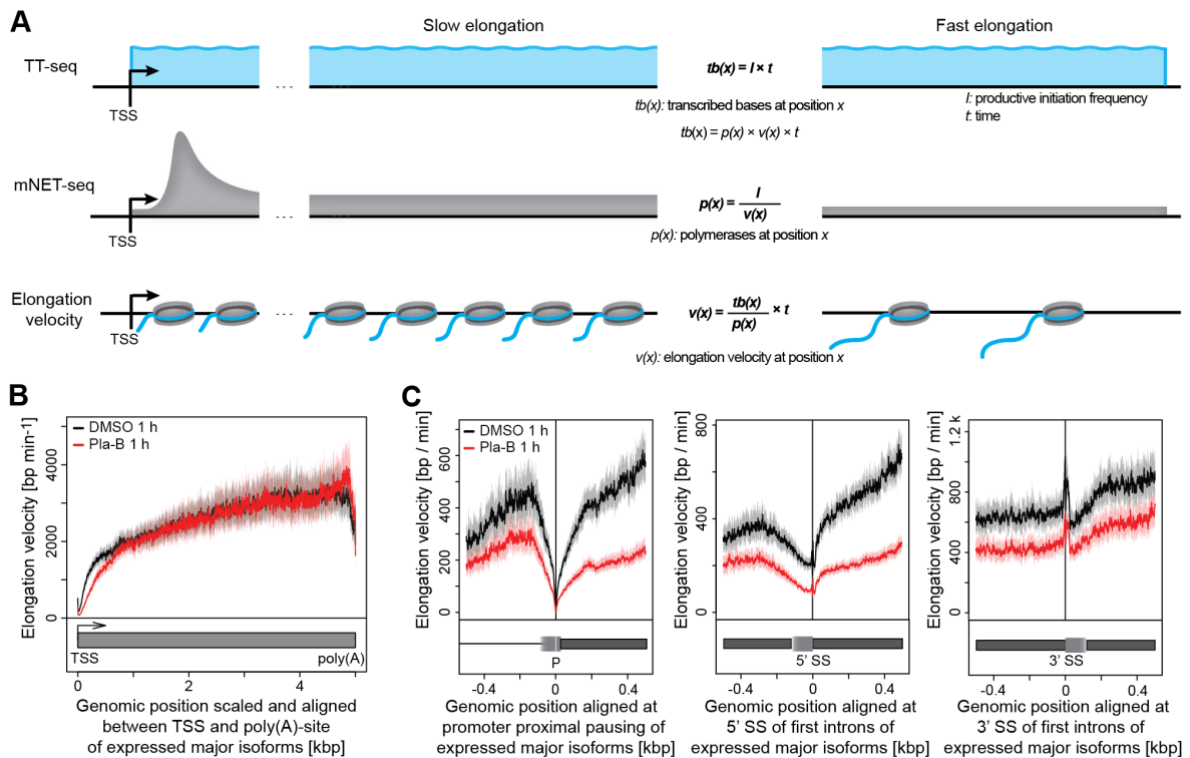


Figure 21. Inhibition of U2 snRNP function alters Pol II elongation velocity.

(A) Schematic representation of Pol II elongation velocity v . The TT-seq signal (top) corresponds to the productive initiation frequency I . The mNET-seq signal (middle) corresponds to the ratio of the productive initiation frequency I to the elongation velocity v . Therefore, v (bottom) can be calculated by dividing the TT-seq signal over the mNET-seq signal. (B) Metagene analysis comparing the elongation velocity in cells treated with DMSO or 1 μ M Pla-B for 1 h. Elongation velocity was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. (C) Metagene analysis comparing the elongation velocity in cells treated for 1 h with DMSO or 1 μ M Pla-B. Elongation velocity was averaged for 3,626 first exons aligned at the pause site defined on the DMSO treated samples (left). Elongation velocity was averaged for 5,350 first introns and aligned at the 5' SS (middle) or the 3' SS (right). Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap).

In accordance with this and with our previous observations, short genes were already affected throughout their entire length after 1 h of treatment (**Figure 22A**). We also found that the average elongation velocity in the first exon was much lower than in the first intron (**Figure 22B**), reflecting Pol II pausing in the promoter-proximal region. Elongation velocity was also affected upon Pla-B inhibition in intronless genes but to a much lower extent (**Figure 22C**). After 4 h of Pla-B treatment, the elongation velocity decreased in all genes, as seen from a corresponding multi-omics analysis with the use of published mNET-seq data (**Figure 22D**). In summary, these results argue against a direct effect of Pla-B on released and processive Pol II elongation complexes. Instead, the results are consistent with a direct effect of spliceosome function on Pol II release from the promoter-proximal region into efficient downstream elongation.

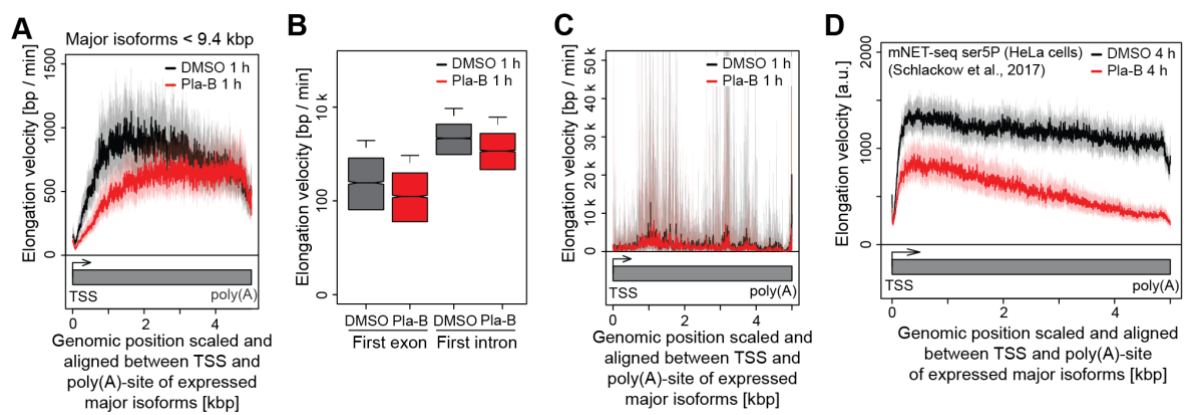


Figure 22. Inhibition of U2 snRNP function alters Pol II elongation velocity.

(A) Metagene analysis comparing the elongation velocities between cells treated with DMSO or 1 μ M Pla-B for 1 h for major isoforms belonging to the lowest quartile of major isoforms length (< 9.4 kbp). Elongation velocity was averaged for 1,384 major isoforms scaled between TSS (left) and poly(A)-site (right). (B) Elongation velocities for the first exon and first intron in DMSO and 1 μ M Pla-B treated cells for 1 h. Elongation velocity distribution was calculated for 5,350 first exons and first introns (unique exons were excluded). (C) Intronless genes meta-analysis comparing the elongation velocity in cells treated with 1 h DMSO or 1 μ M Pla-B. TT-seq coverage was averaged for 51 intronless genes scaled between TSS and poly(A)-site. (D) Metagene analysis comparing the elongation velocities calculated in cells treated with DMSO or 1 μ M Pla-B for 4 h. Elongation velocity was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Datasets for mNET-seq against CTD ser5P in HeLa cells upon 4 h DMSO and 1 μ M Pla-B treatments were obtained from (Schlackow et al., 2017), GEO: GSE81662.

7. U2 snRNP inhibition with Pla-B is independent of stress response

To test whether the transcription response is similar to the one obtained under stress conditions, such as heat shock (HS), I performed a differential gene expression analysis using the DESeq2 R package (Love et al., 2014) on TT-seq data from K562 cells subjected to 42 $^{\circ}$ C HS for 30 min (Gressel et al., 2019) and on our TT-seq data upon U2 snRNP inhibition with Pla-B, SSA and U2 AMO (**Figure 23A, C-F**). TT-seq data analysis revealed that 507 transcripts are significantly downregulated, and 182 transcripts are significantly upregulated upon 30 min of HS in K562 cells. Among the upregulated ones, a GO enrichment analysis shows that within the top 10 GO terms, 6 include “response to stress” (GO:0006950) as a parental term, in agreement with (Gressel et al.,

2019) (**Figure 23A, B, Figure S9**). On the contrary, Pla-B treatment causes mainly downregulation in both 1 h (2,894 transcripts downregulated versus 11 transcripts upregulated) and 4 h (3,705 transcripts downregulated versus 9 transcripts upregulated) (**Figure 23C-D**). Out of the upregulated ones, only 1 upon 1 h (BST2) and 3 upon 4 h (ARL6IP5, MSMO1, SQLE) overlap with the 182 transcripts activated upon HS (**Table S9, Table S10**), showing that different genes are affected and also that the response observed with Pla-B is much faster and more global. Similar to Pla-B, 1 h SSA treatment induces mostly downregulation of gene expression (3,705 downregulated and 9 upregulated transcripts), without any upregulated transcript overlapping with the 30 HS upregulated transcripts (**Figure 23E, Table S11**). These findings indicate that both chemical compounds, Pla-B and SSA, are not responsible for activating a stress pathway. To corroborate this, the treatment with 75 μ M Ctr and 75 μ M U2 AMO follows the same profile as the one from Pla-B and SSA. In this case, Ctr AMO and U2 AMO would induce the same stress pathways. U2 AMO induces upregulation of 24 transcripts and downregulation of 2,665 (**Figure 23F**). Only two transcripts matched the ones upregulated upon 30 min HS (AZIN1 and BAG2) (**Table S12**). The differential gene expression analysis was performed only on the exonic counts for each transcript, since the intronic counts would cause a bias due to the increased intronic coverage (**Methods section 6**).

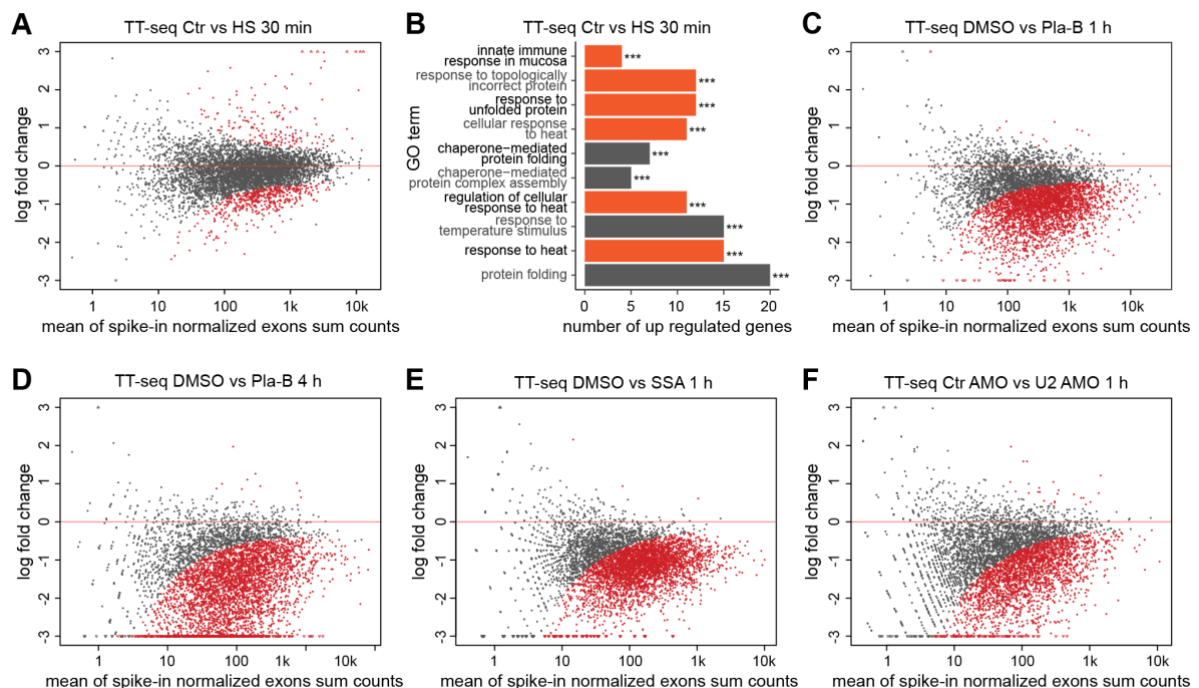


Figure 23. U2 snRNP inhibition with Pla-B is independent of stress response

(A) Differential expression analysis between Control and HS treatments for 30 min on K562 cells (507 downregulated and 182 upregulated transcripts) (data from (Gressel et al., 2019), GSE123980). (B) Top 10 GO terms on the biological process category of upregulated transcripts on TT-seq upon control (Ctr) or 42 °C treatment for 30 min (Gressel et al., 2019)). GO terms belonging to “response to stress” (GO:0006950) are marked in orange. See also **Figure S9**. (C) Differential expression analysis between DMSO and 1 μ M Pla-B treatments for 1 h on K562 cells (2,894 downregulated and 11 upregulated transcripts). See also **Table S9**. (D) Differential expression analysis between DMSO and 1 μ M Pla-B treatments for 4 h on K562 cells (3,705 downregulated and 9 upregulated transcripts). See also **Table S10**. (E) Differential expression analysis between DMSO and 30 ng/mL SSA treatments for 1 h on K562 cells (3,721 downregulated and 3 upregulated transcripts). See also **Table S11**. (F) Differential expression analysis between 75 μ M Ctr and 75 μ M U2 AMO treatments for 1 h on K562 cells (2,665 downregulated and 24 upregulated transcripts). See also **Table S12**. Differential expression analysis was performed with DESeq2 (padj < 0.05 and 0 < log₂FC < 0).

To further characterize the downregulated genes upon U2 snRNP inhibition, a GO enrichment analysis was performed using the previously defined downregulated transcripts for both chemical compound treatments, Pla-B and SSA, and for U2 AMO. As expected, U2 snRNP inhibition for 1 h shows downregulation of genes related to RNA biosynthesis (**Figure 24A, C-D, Figure S10, Figure S12, Figure S13**). Differences in the downregulated genes between the different inhibition methods can be explained by the fact that in this experiment, cells were not synchronized to reduce the bias these methodologies may cause on the overall physiology of the cell (Davis et al., 2001). Upon 4 h treatment with DMSO or 1 μ M Pla-B, the overall RNA synthesis reduction impacts DNA metabolic process and DNA replication, as well as downregulation of genes related to positive regulation of gene expression and regulation of metabolic biosynthesis (**Figure 24B, Figure S11**).

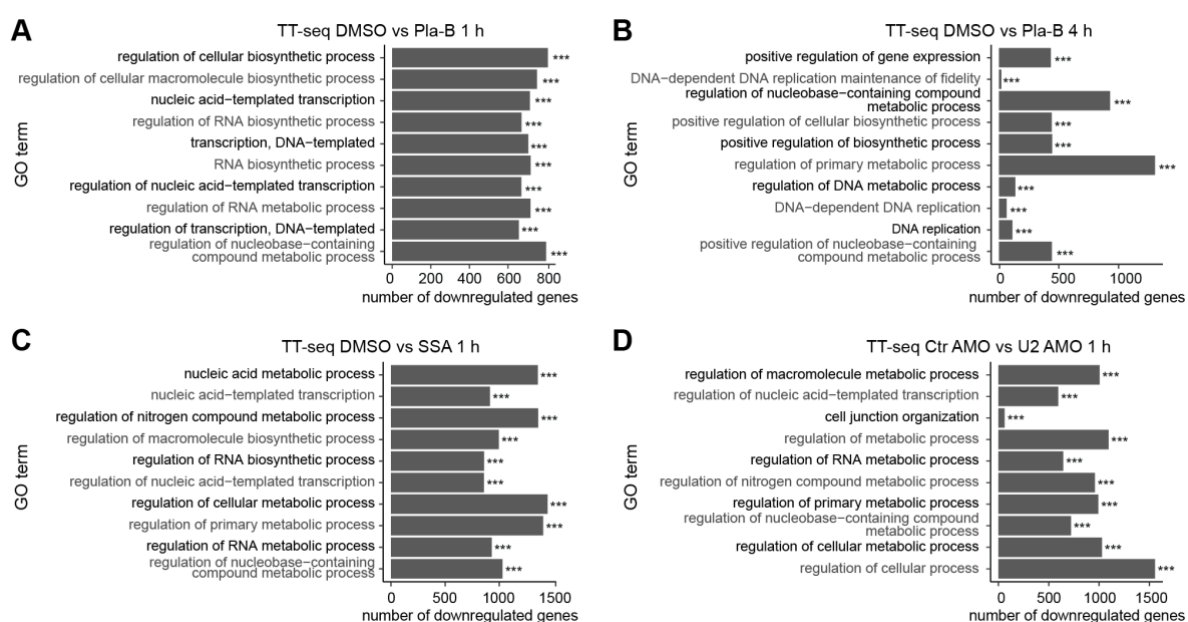


Figure 24. GO enrichment analysis.

(A and B) Top 10 biological process GO terms for downregulated transcripts upon DMSO or 1 μ M Pla-B treatments for 1 h (A) and 4 h (B). See also **Figure S10** and **Figure S11**. (C) Top 10 biological process GO terms for downregulated transcripts upon DMSO or 30 ng/mL SSA treatments for 1 h. See also **Figure S12**. (D) Top 10 GO terms on the biological process for downregulated transcripts upon 75 μ M Ctr and 75 μ M U2 AMO treatments for 1 h. See also **Figure S13**.



DISCUSSION AND OUTLOOK

Here I highlight the major findings, as well as the limitations of this study and the remaining open questions as future perspectives.

The discussion present in sections 2 to 4 is adapted from:

Caizzi, L. *, **Monteiro-Martins, S. *** et al. Efficient RNA polymerase II pause release requires U2 snRNP function. (2021). Mol. Cell 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

1. Open questions regarding U2 snRNP function in transcription

We rapidly inactivated U2 snRNP function in human cells and monitored RNA synthesis, Pol II activity, productive initiation frequency, pause duration, and elongation velocity genome-wide.

Question A) Does splicing inhibition affects new RNA synthesis?

We observed by three independent U2 snRNP inhibition approaches that splicing was overall inhibited already upon 1 h treatment and that this impacts the amount of newly synthesized RNA. Even though the RNA synthesis was overall reduced, we further observed that transcripts whose first intron was significantly affected on 5' SS and/or 3' SS had a significantly higher reduction than the transcripts with splicing unaffected first introns. We further show that the inhibition strategies used are not related to stress response and that upon 1 h U2 snRNP inhibition is observed downregulation of genes related to RNA biosynthesis.

Question B) Is co-transcriptional splicing related to Pol II productive transcription initiation frequency and promoter-proximal pausing duration?

With our analysis, we account for the number of polymerases that initiate and are further released from promoter-proximal pause site to productive elongation. We witnessed an overall reduction in productive initiation frequency upon 1 h Pla-B treatment, which matches the overall decrease in transcription. Once again, the transcripts with a first intron significantly affected on 5' SS and/or 3' SS had a significant decrease in initiation frequency when compared with the transcripts with splicing unaffected first introns.

We calculated the effective pause between two Pol II initiation events that successfully lead to productive elongation. After 1 h Pla-B treatment, the promoter-proximal pausing duration

significantly increased, being once again significantly higher on splicing affected transcripts. The position of the pause sites was, however, not affected by the U2 snRNP inhibition.

Question C) Does co-transcriptional splicing influence Pol II elongation velocity?

The kinetics multi-omics analysis revealed a U2 snRNP impact on productive initiation frequency and promoter-proximal pausing duration. In agreement with these findings, our elongation velocity calculation shows a decrease in Pol II velocity upon Pla-B treatment. The decreased velocity was more visible at the beginning of the transcript upon 1 h Pla-B treatment. However, a decrease in the velocity throughout the full transcript body was observed for shorter genes upon 1 h treatment and in all genes upon 4 h treatment. This finding can be explained by an unaltered velocity on the Pol II enzymes that were already released into productive elongation at the time of the Pla-B treatment.

2. Possible mechanisms under U2 snRNP requirement for efficient Pol II elongation activation

We found that U2 snRNP function is important for the efficient release of paused Pol II into genes for active transcription elongation. Although previous reports had already shown that splicing inhibition has consequences for Pol II occupancy and transcription, our data now define these consequences genome-wide and at the level of transcription kinetics.

Our results suggest that U2 snRNP function is important for the establishment of complete, RNA processing-competent, active Pol II elongation complexes in the promoter-proximal region of transcribed human genes (**Figure 25**). We refer to this process as ‘U2 snRNP-dependent Pol II elongation activation’.

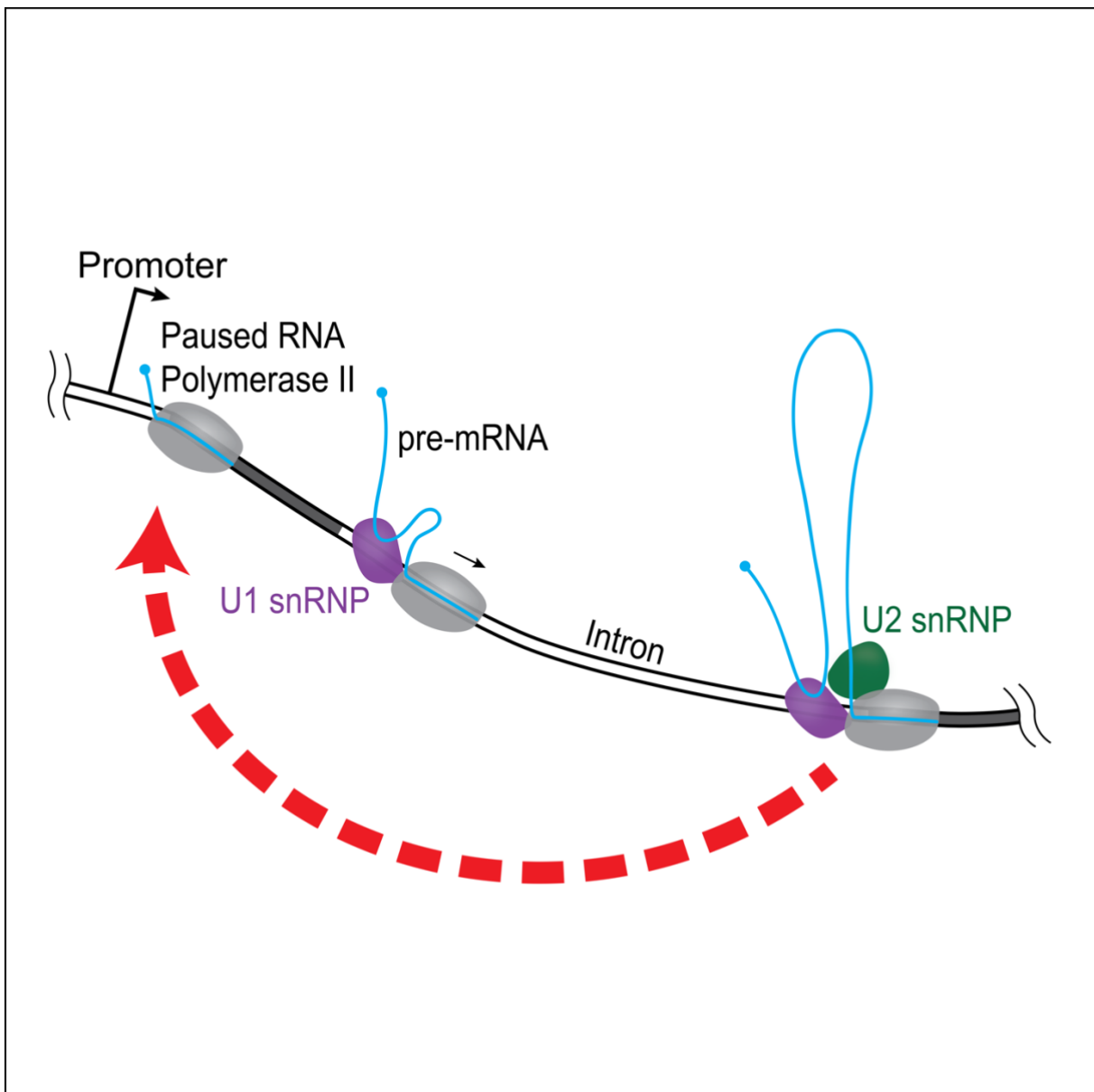


Figure 25. U2 snRNP impact on Pol II release into productive elongation.

Our results show that upon U2 snRNP inhibition there is an increase in paused Pol II at the promoter-proximal site and a decrease in the elongation velocity. This suggests an impact of the formation of the spliceosomal complex A on the release of promoter-proximal paused Pol II into productive elongation.

The mechanism underlying U2 snRNP-dependent Pol II elongation activation remains to be elucidated. However, recent structural studies suggest a possible model for this process. It was recently shown that the Pol II-U1 snRNP complex is compatible with the presence of the transcription elongation factors DSIF, SPT6, and PAF1 complex (PAF) on the Pol II surface and that U1 snRNP can remain bound to Pol II during transcription elongation. This interaction does neither require nascent RNA nor a 5' SS (Zhang et al., 2021).

Modeling further showed that U2 snRNP can be accommodated in addition to U1 snRNP to form the spliceosomal A complex on the Pol II surface, whereas conversion to the splicing-active B complex would separate Pol II and the spliceosome (Zhang et al., 2021). How such separation may activate further release of paused Pol II far upstream remained unclear. It was also shown that U2 snRNP inhibition with SSA reduces the availability of U1 snRNP by retaining it on the pre-mRNAs. The reduced U1 availability was also related to premature cleavage and polyadenylation (Yoshimoto et al., 2021). Here we speculate that U2 snRNP inhibition arrests co-transcriptional spliceosome assembly at the stage of A complex formation on Pol II (**Figure 25**), and impairs the separation of transcribing Pol II and the spliceosome.

Previous results suggest that U2 snRNP-dependent Pol II elongation activation may involve the elongation factor TAT-SF1. TAT-SF1 can bind the U2 snRNP factor SF3B1 and stimulate elongation *in vitro* (Chen et al., 2009; Fong and Zhou, 2001; Loerch et al., 2019). TAT-SF1 also associates with Pol II and SPT5 (Kim et al., 1999). TAT-SF1 further interacts with CycT1 (Fong and Zhou, 2001) and co-IP with U2 snRNP components *in vivo* (Abramczuk et al., 2017) and *in vitro* (Agafonov et al., 2011). The yeast TAT-SF1 homolog Cus-2 interacts with PRP11, a subunit of yeast SF3a (Yan et al., 1998) and with SF3b1 (Talkish et al., 2019). Consistent with our results, mutation of the BS causes Cus-2 dependent Pol II accumulation over introns (Chathoth et al., 2014). It was recently shown that TAT-SF1 binds SF3B1 in the open conformation, and that displacement of TAT-SF1 must occur to allow for stable U2 snRNP-BS interaction (Zhang et al., 2020). U2 snRNP-dependent Pol II pause release may additionally involve the splicing factor SRSF2 (or SC35) that is required for ATP-dependent interaction between U2 snRNP and the BS (Fu and Maniatis, 1992) and for normal Pol II occupancy and P-TEFb recruitment (Ji et al., 2013; Lin et al., 2008). In conclusion, we provide strong evidence that efficient Pol II elongation activation requires U2 snRNP function, and speculate that this is due to co-transcriptional assembly of functional spliceosomes.

3. The impact of U2 snRNP inhibition on intronless genes

Our results further showed that inhibition of U2 snRNP function also decreased RNA synthesis from intronless genes, albeit to a lower extent. In terms of evolution, intronless genes are younger genes, mostly eukaryotic specific (Shabalina et al., 2010). Although U1 snRNP is poorly detectable at selected intronless genes (Listerman et al., 2006), it can be recruited to transcription units of splicing-deficient reporter genes (Spiluttini et al., 2010). Moreover, direct binding of U1 snRNP to the Pol II elongation complex does not require the presence of a 5' SS in nascent pre-mRNA (Zhang et al., 2021). Furthermore, U2 snRNP can also be recruited to intronless genes (Van Nostrand et al., 2016; Wang et al., 2019) and has a role in 3' RNA processing at intronless histone genes (Friend et al., 2007; Kyburz et al., 2006).

Thus, our results may be explained with the presence of low levels of U1 and U2 snRNPs at intronless genes. Moreover, the pre-EJC subunit Mago impact in promoter-proximal pausing is also observed in an intronless gene (*Sry-delta*) in drosophila (Akhtar et al. 2019). However, we cannot entirely exclude indirect effects, since splicing inhibitors such as Pla-B and SSA may induce alternative splicing (Corrionero et al., 2011; Teng et al., 2017). Nevertheless, the rapid inhibition of U2 snRNP function and fast transcription readout in our experiments is expected to avoid such indirect effects.

Structural studies of spliceosomes isolated from mammalian cell nuclei showed the assembly of a 21 MDa macromolecular complex (Müller et al., 1998). The 3D structure for this macromolecular complex was solved at 20-Å resolution by cryo-EM in 2004 (Azubel et al., 2004) and comprised four active native spliceosomes connected by pre-mRNA (Azubel et al., 2004, 2006). Recently, it was observed that the supraspliceosome assembles in all RNAs independently of intron number or length, including intronless genes. Furthermore, SSA does not inhibit supraspliceosome assembly, being observed that affinity-purified supraspliceosomes are enriched with unspliced mRNA after SSA inhibition. Upon Pla-B inhibition, the unspliced mRNAs accumulate in the nucleus; however, SRF2 (or SC35), U1-70k, and PRP8 splicing factors are still recruited (Sebbag-Sznajder et al., 2020). This overall recruitment of a supraspliceosome to transcribing mRNAs may explain the overall effect on newly RNA synthesis upon U2 snRNP inhibition, including intronless genes.

4. Study limitations

Our study is based on K562 cells, a human immortalized myelogenous leukemia cell line. We cannot exclude that we miss in our analysis signals for some RNA isoforms that are not covered by the RefSeq annotation. This might impact the definition of our major isoforms and intronless genes and the assessment of genomic positions including TSS, 5' SS, 3' SS, and poly(A)-site. However, it is not expected to alter any of our conclusions because we selected only major isoforms with $\geq 70\%$ prevalence per gene and we excluded intronless genes with UTRs longer than 100 bp. Pla-B could in principle have effects on transcription, but this is unlikely because orthogonal *in vivo* approaches argue against this. Pla-B treatment will also inhibit the formation of later stages of spliceosome assembly. Therefore, we cannot entirely exclude that Pla-B has a minor effect on cell viability. With respect to the TT-seq method, we cannot entirely exclude the presence of a small amount of unlabeled RNAs. However, we used spike-ins to estimate and correct for minor cross-contamination by pre-existing RNAs. mNET-seq and ChIP-seq involve IP and the quality of the results relies on the specificity of the antibody used. ChIP-seq of splicing factors is technically challenging and there are few studies available, because of the indirect and transient nature of the binding. We performed ChIP-seq of SF3B1 and TAT-SF1 to investigate the binding sites, but unfortunately, we did not obtain high-quality data. I also checked the available SF3B1 eCLIP data obtained in (Van Nostrand et al., 2016) (ENCODE accession: ENCSR133QEA). The antibody used for this experiment was SF3B1-MBL, the same one we tried in our ChIP-seq experiment. To investigate SF3B1 binding to RNA, I downloaded the processed data, available from ENCODE (ENCFF887ARJ). These data were subjected to peak calling, input normalization, and irreproducible discovery as described in the eCLIP processing pipeline v2.2 from Yeo lab. Unfortunately, only 37 peaks were identified over the entire genome, indicating an issue with the quality of the data. The same data were used in (Wang et al., 2019) to demonstrate SF3B1 binding

to intronless genes, and it is not clear, from their Methods, how the processing and normalization were performed. Due to doubts, we did not use the published data.

Even though was not under the scope of this work, it would be interesting to address the impact of alternative splicing, since SSA and Pla-B have been found to modulate alternative splicing (Corrionero et al., 2011; Teng et al., 2017). To allow for better detection of introns, we used an rRNA depletion protocol before library generation. However, alternative splicing detection tools work much better on poly(A)+ RNA-seq (Sultan et al., 2014). Still, I tried to detect alternative splicing events with rMATS (4.1.0) (Shen et al., 2014). 61,515 alternative splicing events were identified for 1 h Pla-B treatment. However, at a typical FDR $\leq 1\%$ no alternative splicing events were identified. We, therefore, did not include this in our analysis, since this approach is not suitable for alternative splicing analysis.

5. Future perspectives

5.1. Annotation

As previously mentioned there are some limitations regarding this analysis concerning the available annotation. LRS can be a big challenge in the study of alternative isoforms or the detection of splice sites. With nanopore LRS it is possible to directly sequence full RNA transcripts (Wang et al., 2014). Moreover, recent techniques show that it is possible to distinguish newly synthesized RNA from the total RNA using labeling strategies applied to nanopore sequencing (Drexler et al., 2021; Maier et al., 2020). Unfortunately, these techniques still output low-quality reads when compared with NGS RNA-seq related approaches (Adewale, 2021; Amarasinghe et al., 2020). The use of a combination of both NGS and LRS, such as the one used on the FLAIR pipeline (Tang et al., 2020) can help address both techniques' limitations. The use of machine learning approaches for the correct identification of newly synthesized RNAs can also be of great help to study co-transcriptional processing such as splicing (Maier et al., 2020).

The annotations available nowadays need to be updated and eventually tuned for different cell types. Use of the genomic **state annotation** STAN R package (Zacher et al., 2014), which is based on Hidden Markov Models to model different types of genomic data, may be introduced to improve the annotation. *De novo* transcriptome assemblers are also powerful tools that could solve the annotation problem, however, these software still show some flaws, inaccurately defining gene isoforms (Freedman et al., 2021; Hölzer and Marz, 2019). A possible approach could be the implementation of a *de novo* transcriptome aligner with massive experimental confirmation of the different isoforms, using machine learning approaches to train the data against the experimentally confirmed isoforms. **Long-read capture sequencing** (IrcaptureSeq) is an interesting approach to annotate poorly characterized mRNA isoforms of target genes (Ray et al., 2020).

5.2. Kinetic modeling

Several studies show a level of promoter-proximal Pol II turnover, suggesting that several Pol II would not exit the pause site into productive elongation (Erickson et al., 2018; Krebs et al., 2017; Steurer et al., 2018). An important improvement on the kinetic analysis would be to account for Pol II turnover and early termination. This could be achieved by the implementation of mathematical modelling together with experimental procedures. Pause duration and elongation

frequency kinetics approaches could be calibrated by integrating TT-seq with a technique to study RNA 3' end such as 3'-seq (Lianoglou et al., 2013) or TAIL-seq (Chang et al., 2014). The use of 3'-seq upon 4sU labeling, such as performed in (Wu et al., 2020), would give the positions of newly synthesized RNAs 3' ends which could be further integrated into the kinetic modelling to account for early terminating transcripts. The integrator complex has been associated with early termination (Beckedorff et al., 2020; Elrod et al., 2019; Lykke-Andersen et al., 2021; Tatomer et al., 2019). Therefore, integrating TT-seq and 3'-seq data using integrator depletion could also be used to validate the kinetics dynamics modelling. A recent study showed that intron removal is a highly variable process with stochastic splice site selection, suggesting that the spliceosome may remove introns in chunks, instead of a whole intron unit (Wan et al., 2021). Therefore, an extensive look on Pol II elongation velocity around annotated and putative splice sites could help understand intron removal dynamics.

5.3. U2 snRNP-dependent Pol II elongation activation mechanism

To investigate the mechanism of U2 snRNP-dependent Pol II elongation activation, future studies should test whether and how splicing and elongation factors contribute to P-TEFb recruitment and promoter-proximal release. DNA-protein interactions studies such as ChIP-seq or CUT&RUN (discussed in the **State of the Art section 4.2**) and RNA-protein interactions studies such as eCLIP (discussed in the **State of the Art section 4.4**) against key factors involved in promoter-proximal pausing and pause-release, such as P-TEFb, NELF and BRD4; and factors involved in splicing such as SF3B1 and TAT-SF1 would help understand how this mechanism works.

As mentioned before, Pol II CTD is related to transcription regulation, including promoter-proximal pausing and splicing. A careful analysis of Pol II CTD phosphorylation should be performed after U2 snRNP inhibition. A better understanding of Pol II occupancy profiles regarding CTD phosphorylation could be accomplished by using mNET-seq with dedicated Pol II antibodies. Previous analysis of Pol II Ser5 phosphorylation upon Pla-B treatment showed a higher Ser5P peak on the 5' SS for DMSO control but not for Pla-B treated cells (Nojima et al., 2015) suggesting that the Ser5P is related to spliceosome assembly. Further studies should include different CTD phosphorylation analyses, such as Ser2P. Ser2 is required for efficient splicing *in vivo*, being associated with the recruitment of the splicing factor U2AF65 and U2 snRNA (Gu et al., 2013). Additionally, Ser2 was shown to be decreased upon SSA treatment (Koga et al., 2015). Further studies could help understand the Pol II CTD role on the relation between promoter-proximal pausing and U2 snRNP inhibition.

Appropriate RNA folding is essential for the spliceosome assembly and the catalytic reaction (Warf and Berglund, 2010). Therefore, RNA-RNA interactions assays (discussed in the **State-of-the-Art section 4.5**) should be performed upon fast U2 snRNP inhibition to understand if and how it affects the RNA structure.

Histone modifications, nucleosome positioning, and chromatin accessibility are important factors in the splicing process (Gómez Acuña et al., 2013). Histone marks are related to exon exclusion (Agirre et al., 2021), and chromatin accessibility was predicted to be a regulator of intron retention (Petrova et al., 2021). Thus, the study of different histone marks by DNA-protein interaction techniques and further analysis on nucleosome positioning and chromatin accessibility (example: ATAC-seq) upon a rapid U2 snRNP inhibition could give important insight on the mechanism for Pol II elongation activation dependent on U2 snRNP.

5.4. R-loops impact on fast U2 snRNP inhibition

Alterations in the splicing process have been linked to R-loop formation. R-loop formation is related to spliceosome recruitment (Bonnet et al., 2017) and point mutations in the U2 snRNP component U2AF1 lead to R-loop accumulation (Chen et al., 2018; Nguyen et al., 2018). On the other hand, R-loop formation is related to DNA instability and stress response (Tam and Stirling, 2019). In our analysis, we performed a fast U2 snRNP inhibition and we did not observe an impact on stress response genes. However, a closer look at R-loop formation upon fast U2 snRNP inhibition would be important to understand this mechanism. That could be accomplished by performing DRIP-seq (Ginno et al., 2012) upon a short U2 snRNP inhibition. A preprint from Frédéric Chédin's lab (Castillo-Guzman et al., 2020), using Pla-B inhibition for 2 and 4 h followed by DRIP-seq, shows a reduction of R-loops formation upon treatment. Taking into consideration that we observed an overall decrease in transcription upon Pla-B inhibition, a normalization by expression level would be essential to understand the impact on R-loop formation upon treatment.

5.5. Cell type-specific U2 snRNP function

In the present work, we used a human cancer cell line (K562). An interesting analysis would be the comparison of different cell types, including human primary cells isolated from healthy tissues, to check if the observed effect is cell type-specific and not influenced by cancer type mutations. Since response to stimuli may have a heterogeneous response on gene expression and splicing across cells (Shalek et al., 2013), the use of scRNA-seq could help clarify U2 snRNP function. In particular, the study of newly synthesized RNA and chromatin accessibility in a single cell approach (see **State of the art section 4.10**) upon fast U2 snRNP inhibition could be of great importance to further understand this mechanism.

VI

SUPPLEMENTARY INFORMATION

This chapter contains supplementary information to the Methods and Results chapters, including the experimental methods, Spike-ins sequence, and all the supplementary figures and tables.

1. Experimental methods

The laboratory experiments were performed by Dr. Livia Caizzi (MPI-bpc, Dept. of Molecular Biology), except for mNET-seq experiments which were performed by both Dr. Livia Caizzi and Kseniia Lysakovskaia (MPI-bpc, Dept. of Molecular Biology).

The experimental methods presented in this section have been published:

Caizzi, L. *, **Monteiro-Martins, S.** * et al. Efficient RNA polymerase II pause release requires U2 snRNP function. (2021). Mol. Cell 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

A list of reagents and resources used on the experimental methods can be found in **Table S2**.

1.1. Employed cell lines

Human K562 cells (female) were obtained from DSMZ (DSMZ no.: ACC-10, RRID:CVCL_0004) and cultured in antibiotic-free RPMI 1640 medium (Thermo Fisher Scientific, 31870–074) supplemented with 10 % heat-inactivated fetal bovine serum (Thermo Fisher Scientific, 10500–064) and 2 mM GlutaMAX (Thermo Fisher Scientific, 35050087) at 37 °C and 5 % CO₂. K562 cells were authenticated at the DSMZ Identification Service according to standards for STR profiling (ASN-0002). Biological replicates were grown independently. Cells were verified to be free of mycoplasma contamination using Plasmoguard Mycoplasma Detection Kit (InvivoGen, rept1). *D. melanogaster* S2 (Schneider-2) cells (used for ChIP-seq spike-ins) were obtained from DSMZ (DSMZ no.: ACC-130, RRID:CVCL_Z232) and cultured in Schneider's drosophila medium (Biowest, L0207) supplemented with 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific, 10500–064) at 27°C without CO₂. Yeast *S. cerevisiae* cells genotype BY4741 (used for mNET-seq spike-ins) were obtained from Euroscarf (ACC-Y00000) and cultured in YP medium supplemented with 2% glucose to OD₆₀₀ 0.5 at 30°C.

1.2. Cell culture treatments

Cells were plated the day before the experiment at a confluency of 300,000 cell/mL to obtain approximately 500,000-700,000 cells/mL on the next day. For Pla-B experiments, cells were treated with dimethyl sulfoxide (DMSO, 1:20,000 dilution, Sigma-Aldrich, D2438) as solvent control or with 1 μ M Pla-B from a DMSO-resuspended 20 mM stock (Santa Cruz, sc-391691). For SSA experiments, cells were treated with DMSO (1:17,000 dilution, Sigma-Aldrich, D2438) as solvent control or with SSA from a DMSO-resuspended 1 mM stock (gift from Vladimir Pena).

1.3. Total and 4sU-labeled RNA extraction and semiquantitative-PCR

Total RNA was extracted using QIAzol (Qiagen, 79306) manufacturer's instruction. 4sU-labeled RNA was extracted and purified as described for TT-seq experiment (see below) except for the sonication of 4sU-labeled RNA which was omitted to avoid introns fragmentation. Isolated RNA was treated with DNase (Qiagen, 79254) to avoid genomic DNA contamination. 400 ng of total or 140 ng of purified 4sU-labeled RNA were reverse transcribed (Thermo Fisher Scientific, EP0751) using random hexamer primers. PCR was performed with Phusion High-Fidelity DNA Polymerase (NEB, M0530L) using the following parameters and primers:

DNAJB1: 30 seconds (sec) at 98 °C, 28 cycles of 5 sec at 98 °C – 10 sec at 68 °C – 15 sec at 72 °C, and 5 minutes (min) at 72 °C.

RIOK3: 30 seconds at 98 °C, 28 cycles of 5 sec at 98°C – 10 sec at 59 °C – 15 sec at 72 °C, and 5 min at 72 °C.

DNAJB1_RT-PCR_agarose gel_F: GAACCAAATCACTTTCCCAAGGAAGG

DNAJB1_RT-PCR_agarose gel_R: AATGAGGTCCCCACGTTTCTCGGGTGT

RIOK3_RT-PCR_agarose gel_F: GCTGAAGGACCATTATTACTGGAG

RIOK3_RT-PCR_agarose gel_R: TTCTTGCTGTGTTCTTTCTCCCACA

Amplified cDNA was loaded in a 2 % agarose gel. Band intensity was analyzed with ImageJ (Fiji).

1.4. TT-seq and RNA-seq

TT-seq was performed as described (Schwalb et al., 2016; Wachutka et al., 2019) with minor modifications. Specifically, 5×10^7 cells from two biological replicates were either treated for with 1 μ M of Pla-B (Santa Cruz, sc-391691) or 30 ng/mL SSA (gift from Vladimir Pena) or DMSO (Sigma-Aldrich, D2438) control at times indicated. Cells were exposed the last 10 min of the treatment time to 500 μ M of 4-thiouridine (4sU, Carbosynth, NT06186) at 37 °C and 5 % CO₂. 300 ng of RNA spike-ins mix were added to each sample after cell lysis in 10 mL of QIAzol (Qiagen, 79306). Spike-ins sequences and production are described in (Wachutka et al., 2019). 150 μ g of RNAs were sonicated to obtain fragments of < 6 kb using AFA micro tubes in a S220 Focused-ultrasonicator (Covaris). The quality of RNAs and the size of fragmented RNAs were checked using a Fragment Analyzer. 1 μ g of each of the sonicated RNAs was stored at -80 °C as total RNA (RNA-seq). 4sU-labeled RNAs were purified from 300 μ g of each of the fragmented RNAs. Biotinylation and purification of 4sU-labeled RNAs was performed as described (Dölken et al., 2008; Wachutka et al., 2019). 100 ng of input RNA was used for strand-specific library preparation according to the Ovation Universal RNA-seq System (NuGEN, 0343-32). Libraries were prepared using random hexamer priming only. The size-selected libraries were analyzed on a Fragment Analyzer before

sequencing on the Illumina NextSeq500 (2 × 75 and 2 × 150 base paired-end for shallow and deep sequencing, respectively).

1.5. Morpholino inhibition

3.2×10^6 of K562 cells were transfected by electroporation with 7.5 nmol/100 μ l (75 μ M) of antisense morpholino oligo against a control region (Ctr AMO, CCTCTTACCTCAGTTACAATTTATA) or U2 snRNA (U2 AMO, TGATAAGAACAGATACTACACTTGA) in a transfection volume of 100 μ l using the NEON system (Invitrogen, MPK10025). The following transfection parameters were used: pulse: 1,350, pulse width: 10, pulse number: 4. Transfected cells were transferred to 5 mL of fresh culture medium, and harvested at the indicated times. For semi-quantitative PCR analysis, RNA was isolated and treated as previously described (see “Total and 4sU-labeled RNA extraction and semiquantitative-PCR”). For TT-seq experiment, cells were exposed the last 10 min of the 1 h transfection time to 500 μ M of 4-thiouridine. RNA was isolated and treated as previously described (see “TT-seq and RNA-seq”).

1.6. mNET-seq

Experiments were performed as previously described (Gressel et al., 2019; Nojima et al., 2016; Nojima et al., 2015; Schlackow et al., 2017) with minor modifications. All buffers for the cellular fractionation and immunoprecipitation (IP) were supplied with protease inhibitor cocktail (Sigma-Aldrich, P8340) and phosphatase inhibitors (Sigma-Aldrich, 4906837001). In detail, two biological replicates of 1×10^8 K562 cells were treated for 1 h with 1 μ M of Pla-B (Santa Cruz, sc-391691) or DMSO (Sigma-Aldrich, D2438) solvent. Cellular fractionation was performed according to the previously published protocol (Conrad and Ørom, 2017) using 1×10^7 cells per one reaction. Isolated chromatin was subjected to micrococcal nuclease (MNase, NEB, M0247S) digestion at 37 °C and 1,400 rpm for 2 min. Afterwards, soluble chromatin fractions were pooled together for each sample and diluted 8-fold with IP buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 0.05 % (vol/vol) NP-40, 0.3 % (vol/vol) empigen BB (Sigma-Aldrich, 30326)). For each sample, 30 μ g of RNA polymerase II (Pol II) antibody (Diagenode, C15200004, RRID:AB_2728744) were coupled to Dynabeads M-280 Sheep Anti-Mouse IgG (Invitrogen, 11201D) prior to IP step. Diluted chromatin was mixed with Pol II antibody-beads complexes and subjected to IP performed on a rotating wheel at 4 °C for 1 h. Beads were washed seven times with IP buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 0.05 % (vol/vol) NP-40, 0.3 % (vol/vol) empigen BB (Sigma-Aldrich, 30326)) and one time with PNKT buffer (1 × T4 PNK buffer (NEB, M0236L), 0.1 % vol/vol Tween-20). For the RNA phosphorylation, beads were resuspended in PNK reaction mixture containing 1 × T4 PNK buffer (NEB, M0236L), 0.1 % vol/vol Tween-20, 1 mM ATP (Cell Signaling Technology, 9804S), T4 Polynucleotide Kinase (phosphatase minus) (NEB, M0236L) and incubated at 37 °C and 800 rpm for 10 min. Beads were washed once with IP buffer and resuspended in TRIzol reagent (Invitrogen, 15596026). At this point, 5 ng of RNA spike-ins were added to each sample. RNA was extracted from the beads and precipitated with GlycoBlue coprecipitant in 100 % ethanol at -20 °C. Next day, RNA was size-selected (25-110 nt) using a denaturing 6 % (wt/vol) polyacrylamide gel containing 7 M urea. RNA was extracted from the gel by incubation with the elution buffer (1 M NaOAc pH 5.5, 1 mM of EDTA pH 8.0) and precipitated with GlycoBlue coprecipitant (Invitrogen, AM9515)

in 100 % ethanol at -20 °C. RNA size distribution and concentration were estimated using Fragment Analyzer. RNA samples were used for the library preparation in equal amounts. Libraries were prepared using TruSeq Small RNA Library Kit (RS-200-0048) according to the manual and as described (Nojima et al., 2015). Purity and size distribution of the libraries were estimated using Fragment Analyzer. Libraries were sequenced on Illumina NEXTseq 550 (2 × 42 base paired-end).

1.7. RNA spike-ins for mNET-seq

50 mL of *S. cerevisiae* cells (wild type strain BY4741) were grown overnight at 30°C. Cells were spin down, and pellet was washed 1 × with 25 mL milli-Q water. Cells from 25 mL volume pellet were lysed in 1.5 mL of TSNTE buffer (2 % Triton X-100, 1 % SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). One volume of acidic phenol (CarlRoth, A980.2) and 1 volume glass beads acid washed (Sigma-Aldrich, G8772) were added to 1 volume of the lysed cells and TSNTE buffer mixture. Cell lysate was further homogenized in FastPrep (8 × 40 sec with 1 min on ice between each step). Cell lysate was centrifuged at high speed for 5 min. One volume of chloroform was added to 1 volume of supernatant. 0.3 M of sodium acetate was added and RNA was precipitated overnight in 100 % ethanol. RNA pellet was washed with 75 % ethanol and resuspended in 900 µl of RNase-free water. Total RNA quality was checked in Fragment Analyzer. 200 µg of total RNA were subjected to DNase digestion (TURBO DNA-free DNase treatment, Thermo Fisher Scientific, AM1907) following manufacturer's instructions. RNA was purified with chloroform and precipitated overnight in 100 % ethanol at -20 °C. RNA pellet was washed with 75 % ethanol, and pellet was resuspended in 50 µl of RNase-free water. Quality of total RNA DNase-treated was checked in Fragment Analyzer. 75 µg of total RNA DNase treated were subjected to mRNA purification using Dynabeads mRNA Purification kit (Invitrogen, cat n. 61006) following manufacturer's instructions. Two rounds of mRNA purification were performed to eliminate possible rRNA contaminants. Approximately 0.4 % of original amount of total RNA was recovered. 250 ng of purified mRNA was fragmented for 5 min at 94 °C using NEBNext® Magnesium RNA Fragmentation Module Protocol (NEB, E6150). 0.3M sodium acetate and 1 µl of Glycoblue was added to the fragmented mRNA. mRNA was precipitated overnight in 100 % ethanol. RNA was washed with 75 % ethanol, and pellet was resuspended in 10 µl of RNase-free water. End-repair was performed following NEB manufacturer's instruction. In details, reaction was incubated for 30 min at 37 °C (9 µl of fragmented mRNA, 5.5 µl of RNase-free water, 0.5 µl of RNase OUT, 2 µl of T4 PNK buffer 10X (NEB, B0201), 1 µl of T4 PNK enzyme (NEB, M0201). 2 µl of 10 mM ATP were added to the mixture and the reaction was further incubated for other 30 min at 37 °C. 0.3 M sodium acetate and 1 µl of Glycoblue was added to the reaction and RNA was precipitated overnight in 100 % ethanol. RNA was washed in 75 % ethanol and pellet resuspended in 20 µl of RNase-free water. RNA spike-ins quality and size distribution were checked in Fragment Analyzer (miRNA kit) and in 6 % urea gel, as described for mNET-seq protocol. Size distribution ranged between 30 nt and 150 nt, similar to mNET-seq samples.

1.8. CHIP-seq

6×10^7 cells from two biological replicates were treated for 1 h with 1 µM of Pla-B (Santa Cruz, sc-391691) or DMSO (1:20,000 dilution, Sigma-Aldrich, D2438). Formaldehyde (16 % concentrate stock methanol-free, Thermo Fisher Scientific 28908) was directly added to the media to a final

concentration of 1 % and incubated for 8 min. 125 mM Glycine (final concentration) was added to quench the reaction for 5 min. Cells were spin down and pellet was washed twice with PBS at 4 °C. Protease (Merk Millipore, 5892970001) and phosphatase inhibitors (Merk Millipore, 4906845001) were added to all the buffers. A pellet from 3×10^7 cells was lysed for 10 min on ice with Farnham Lysis buffer (5 mM Pipes pH 8, 85 mM KCl, 0.5 % NP-40). Pellet was centrifuged for 5 min at 1,700 g at 4 °C. Pellet was resuspended with 1 mL of sonication buffer (10 mM Tris-HCl 7.5 pH, 1 mM EDTA, 0.4 % SDS) and incubate on ice for 10 min and transfer to AFA milliTube. Sonication was performed with a S220 Focused-ultrasonicator (Covaris) with the following parameters: duty cycle 5 %, peak incident power 140 W, cycle per burst 200, processing time 1,080 sec, degassing mode continuous, water run level 8. Sheared chromatin was centrifuged at 10,000 g for 15 min at 4 °C. 15 μ l of samples were de-cross-linked overnight at 65 °C and size distribution was checked in 1 % agarose gel. 15 μ g of RNA polymerase II (Pol II) antibody (Diagenode, C15200004, RRID:AB_2728744), 20 μ g of Cyclin T1 (CycT1) antibody (Cell Signaling, D1B6G, RRID:AB_2799973) or 1 μ g of histone H2Av (Active Motif, 39715) were coupled to Dynabeads Protein G (Thermo Fisher Scientific, 10009D) for 2 h at room temperature for each sample. 100 μ g of chromatin was used for each IP. 200 ng of *Drosophila* S2 sheared cross-linked-chromatin (Covaris S200 parameters: duty cycle 5 %, peak incident power 140 W, cycle per burst 200, processing time 1,800 sec, degassing mode continuous, water run level 8) were added to 100 μ g of chromatin as spike-ins control. Chromatin was diluted with IP buffer (50 mM Hepes pH 7.9, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Sodium-deoxycholate) to obtain a 0.05 % final concentration of SDS. 1 % of diluted chromatin was kept as input at 4 °C. Diluted chromatin was mixed with antibody-beads complexes and subjected to IP performed on a rotating wheel at 4°C overnight. Beads were washed 5 times with IP wash buffer (100 mM Tris HCl pH 7.5, 500 mM LiCl, 1 % NP-40, 1 % Sodium-deoxycholate) and one time with TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). Immuno-bound chromatin was eluted at 70 °C for 10 min with elution buffer (0.1M NaHCO₃, 1 %SDS) and de-cross-linked overnight at 65 °C. After RNase A treatment at 37 °C for 1.5 h and proteinase K treatment at 45 °C for 2 h, DNA was extracted with one volume phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, P2069) and precipitated for 30 min at -80 °C with 200 mM NaCl and 100 % ethanol. Pellet was washed with 70 % ethanol and resuspended in TE buffer. DNA quality and size distribution were checked on Fragment Analyzer. 10 ng of DNA was used for library preparation according to NEBNext® Ultra™ II DNA Library Prep Kit (NEB, E7645S). Purity and size distribution of the libraries were estimated using Fragment Analyzer. Size-selected libraries were sequenced on Illumina NEXTseq 550 (2 × 75 and 2 × 42 base paired-end for Pol II and CycT1, respectively).

2. Sequence of the spike-ins used on TT-seq and RNA-seq experiments.

Spike-ins are derived from ERCC Spike-in Mix with an addition of 5' triphosphate to facilitate in vitro transcription by T7 polymerase. The triphosphate (GGG) is not included in the sequence.

Related to Methods section 5.1.

Spike 2

AATACCTTTACAAATGCTTTAACAAGAGGAAATTGTGTTTTTGCCAATTTAAGACCTAATTTAATAGTTA
 AACCATTAACCTTAGTTGTTCCAAGGCATAATATAGAGAGTGAGATACAGGATGAGCTATTTACAGGGAG
 TTATTACAGTATGCAGTTGCCAAGGCAGTTGCTGATTTAGATTTAGATGAAGATTTAAAGGTTGTTGTCTCT
 GTTAATGTCCAGAGGTTCCAATAACCAATTTAAATAAAAAGAAAACCTTTCCAATACTTCTATGCCTCAG
 CAAAGTTAGCTATAAACAGAGCTTTAAATGAATATCCTTCAAAGAGAAAGGTAAGAAAAGAGAAATAT
 AGAGCTTTCATCCATTAGTTGGATTTAGGGATGTTAGATTGGAGTATCCTCCATATCTACAAATTGCTTT
 GGATGTCCCAACTATGGAGAATTTGGAATTTTTGTTACAAACAATCCAAATAGCGACCACATCATCTTA
 GAGGCTGGAACACCACTAATTAAGGTTTGGTTTAGAGGTTATTGAAATAATGAGAGAATATTTTGAT
 GGCTTTATTGTTGCTGATTTAAAAACCTTAGACACTGGAAGGGTTGAGGTAAGATTGGCATTGGAAGCAA
 CAGCTAATGCAGTGGCAATAAGTGGAGTAGCACCAAAATCAACAATAATTAAGCTATCCACGAATGTC
 AAAAATGTGGTTTAATCAGCTATTTGGATATGATGAACGCTCTGAACCTCAAAAATTATATGATTTCATT
 AAAATTAAGCCAGATGTTGTTATCTTGCATAGAGGGATTGATGAGGAGACATTTGGAATTAAGGGA
 ATGGAAATTTAAGGAAAACCTGCTTATTAGCAATTGCTGGAGGAGTTGGTGTGGAGAATGTTGAAGAGCT
 TTTAAAAGAATATCAAAATTAATCGTTGGTAGAGCAATTACAAAATCAAAAGACCCAGGAAGAGTAAT
 TAGGATTTTATAAACAAGATGGGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Spike 4

TTTCGACGTTTTGAAGGAGGGTTTTAAGTAATGATCGAGATTGAAAAACCAAAAATCGAAACGGTTGAA
 ATCAGCGACGATGCCGAATTTGGTAAGTTTGTCTAGAGCCACTTGAGCGTGGATATGGTACAACCTCTG
 GGTAACCTTACGTCGTATCCTCTTATCCTCACTCCCTGGTGCCGCTGTAACATCAATCCAGATAGATG
 GTGTACTGCACGAATTCTCGACAATTGAAGGCGTTGTGGAAGATGTTACAACGATTATCTTACACATTA
 AAAGCTTGCATTGAAAATCTACTCTGATGAAGAGAAGACGCTAGAAATTGATGTACAGGGTGAAGGAAAC
 TGTAACGGCAGCTGATATTACACACGATAGTATGATGTAGAGATCTTAAATCCTGATCTTCATATCGCGACT
 CTTGGTGAGAATGCGAGTTTCCGAGTTTCGCTTACTGCTCAAAGAGGACGTGGGTATACGCCTGCTGACG
 CAAACAAGAGAGGCGATCAGCCAATCGGCGTGATTCCGATCGATTCTATCTATACGCCAGTTTCCCGTGT
 ATCTTATCAGGTAGAGAACACTCGTGTAGGCCAAGTTGCAAATATGATAAACTTACACTTGTATGTTGG
 ACTGATGGAAGCACTGGACCGAAAAGAAGCAATTGCGCTTGGTTCAAAGATTTTAACTGAACACCTTAAT
 ATATTGCTGTTTAACTGACGAAGCTCAACATGCTGAAATCATGGTTGAAGAAGAAGATCAAAAA
 GAGAAAGTTCTTGAAATGACAATTGAAGAATTGGATCTTCTGTTCTTACAACCTGCTTAAAGCGTG
 CGGGTATTAACACGGTTCAAGAGCTTGCGAACAAGACGGAAGAAGATATGATGAAAGTTCGAAATCTA
 GGACGCAAATCACTTGAAGAAGTGAAGCGAGACTAGAAGAAGTGGACTCGGACTTCGCAAAGACGA
 TTGACTAGTTTCCCTTGTGAACTAGGATTTTCCCGGGTACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Spike 5

ACTGTCCTTTCATCCATAAGCGGAGAAAGAGGGAATGACATTGTTCTTACACGGCACAAGCAGACAAAA
 TCAACATGGTCATTTAGAAATCGGAGGTGTGGATGCTCTCTATTTAGCGGAGAAATATGGTACACCTCTT
 TACGTATATGATGTGGCTTTAATACGTGAGCGTGCTAAAAGCTTTAAGCAGGCGTTTATTTCTGCAGGGC
 TGAAAGCACAGGTGGCATATGCGAGCAAAGCATTCTCATCAGTCGCAATGATTACAGCTCGCTGAGGAAG
 AGGGACTTTCCTTAGATGTCGTATCCGGAGGAGAGCTATATACGGCTGTTGCAGCAGGCTTTCCGGCAGA
 ACGCATCCACTTTCATGGAAACAATAAGAGCAGGGAAGAAGTGCAGGATGGCGCTTGAGCACCCGCATCGG
 CTGCATTGTGGTGGATAATTTCTATGAAATCGCGCTTCTTGAAGACCTATGTAAAGAAACGGGTCACTCC
 ATCGATGTTCTTCTTCGGATCACGCCCCGAGTAGAAGCGCATAACGCATGACTACATTACAACGGGCCAG
 GAAGATTCAAAGTTTGGTTTCGATCTTACATAACGGACAACTGAACGGGCCATTGAACAAGTATTACAA
 TCGGAACACATTCAGCTGCTGGGTGTCCATTGCCATATCGGCTCGCAAATCTTTGATACGGCCGGTTTTG

TGTTAGCAGCGGAAAAAATCTTCAAAAAACTAGACGAATGGAGAGATTCATATTCATTTGTATCCAAGG
TGCTGAATCTTGGAGGAGGTTTCGGCATTTCGTTATACGGAAGATGATGAACCGCTTCATGCCACTGAATA
CGTTGAAAAAATTATCGAAGCTGTGAAAGAAAATGCTTCCCGTTACGGTTTTGACATTCCGGAAAATTTGG
ATCGAACCGGGCCGTTCTCTCGTGGGAGACGCAGGCACAACCTTTTATACGGTTGGCTCTCAAAAAGAA
GTGGATAAGCTGTACAATCGTTTCATCATTTCGGCGTGCGAATTAATAAAAAAAAAAAAAAAAAAAAAA
A

Spike 8

AGATGTATATATGATGTCCTTGGACGGGGTGGCGCAGTATTACTGCAAGAGAGCGGACAGATTAGTGTG
TTGGAGCCGACACATCAAAGGTTTCGTCGGGGACCGATCTGCAGCCTACGGGACATTTATCCGTA AAAAG
CATGGCGCTGTTTCGTA CTTATCGGAGGCCAGGTATCGTCGCGGCGAGTCTCCCGACGACGGAGATGG
GCGTTACTATCTGGGCCGCTCTCGTACTCTGTTACTTGGCACAGATGCGAGCCCTCGTAATGTGCATCAGC
TAAGGGCGATATTATAATGCGACGTTTGTACGGATTTCGTTACTAACGTGTTGGACGCTAGTGAATATGT
GTCGTTGGTTAGCCTACCCATGGCTTTCGCGGGCAGACATGCTTAGACTCTTCAAACTTCGGTGAAGT
TCACTCAAGCCGCGGAGCGCCGTCGTAATTCAGGGATGGCGGTACCCGTGCCCGTCCGATTTCGTAG
CAACCTGCATCACGATTTTGTCTTCGGGCGACTTATCAGATACGGTAATGTAAATACCTGGCATTGGGGC
ACTTCTTTCGTTTAAAGCGGGAAAGATCGCGAGGGCCCCGCTATTTGCGATACTTCCCATGTCCGGTGCCGTC
GCCTCTATGTACTCGGAGACGTTAATGCAGAGGCTAAGGACAATTTACCATGACTCGGTAATCCGTTTCGT
CAAGCAGGTAGCTCGAGTCTCCCGACGACACGTAGTGGGTTTGTAACGATCGATAACCGAGTCTTTTTGT
CTAGTAGAACCAACCAACCATTAAGGAGTTCACTAGCACATCTTTGCGACCCGATCGTCCGTGTGTCGCG
TAATACTTTTGTATGACGAGACATACGCTCAAGCCCTGGGTAGCTAGTCGCGGAGGCACGTTACCGCGC
ACAACCCCTATTCGTTTACATGTACATCGCATCTGAGGTAGTACACTTCCGGCGTACGTGAGTATTTGCG
CGTAATAAGCGCGTGTATTAGCTGATCCCCTCTCGTATCGAGGTTAAGGCAGATTAGTGCCAGTAATTGC
GTTTTTTTGTTCGTTGTCGAGAACGCGATTTGCTCCGAAAGCTTTAAGCCGTGGAAAAAAAAAAAAAAAA
AAAAAAAAA

Spike 9

TCCAGATTACTTCCATTTCCGCCAAGCTGCTCACAGTATACGGGCGTCGGCATCCAGACCGTCGGCTGA
TCGTGGTTTTACTAGGCTAGACTAGCGTACGAGCACTATGGTCAGTAATTCCTGGAGGAATAGGTACCA
AGAAAAAACGAACCTTTGGGTTCCAGAGCTGTACGGTCGCACTGAACTCGGATAGGTCTCAGAAAAAC
GAAATATAGGCTTACGGTAGGTCCGAATGGCACAAAGCTTGTTCCGTTAGCTGGCATAAGATTCCATGC
CTAGATGTGATACACGTTTCTGGAAACTGCCTCGTCATGCGACTGTTCCCGGGGTCAGGGCCGCTGGTA
TTTGCTGTAAAGAGGGGCGTTGAGTCCGTCCGACTTCACTGCCCCCTTTCAGCCTTTTGGGTCCTGTATCC
CAATTCTCAGAGGTCCC GCGTACGCTGAGGACCACCTGAAACGGGCATCGTCGCTCTTCGTTGTTTCGTC
GACTTCTAGTGTGGAGACGAATTGCCAGAATTATTA ACTGCGCAGTTAGGGCAGCGTCTGAGGAAGTTT
GCTGCGTTTTGCTTACCGCGGGAAGGAGACATAACGATAGCGACTCTGTCTCAGGGGATCTGCATA
TGTTTGCAGCATACTTTAGGTGGCCCTTGGCTTCCCTCCG CAGTCAAACC GCGCAATTATCCCCGTCCT
GATTTACTGGACTCGCAACGTGGGTCCATCAGTTGTCCGTATAACCAAGACGTCTAAGGGCGGTGTACACC
CTTTTGAGCAATGATTGCACAACCTGCGATCACCTTATACAGAATTATCAATCAAGCTCCCCGAGGAGCG
GACTTGTAAAGACCGCCGCTTTCGCTCGGGTCTGCGGGTTATAGCTTTTCAGTCTCGACGGGCTAGCACA
CATCTGGTTGACTAGGCGCATAGTCGCCATTCACAGATTTGCTCGGCAATCAGTACTGGTAGGCGTTAGA
CCCCGTGACTCGTGGCTGAACGGCCGTACAACCTCGACAGCCGGTGCTTGCCTTTTACCCTTAAAAAAA
AAAAAAAAAAAAAAAAA

Spike 12

TATTGGTGGAGGGGCACAAGTTGCTGAAGTTGCGAGAGGGGCGATAAGTGAGGCAGACAGGCATAATA
TAAGAGGGGAGAGAATTAGCGTAGATACTTTCCAATAGTTGGTGAAGAAAATTTATATGAGGCTGTTA
AAGCTGTAGCAACTCTTCCACGAGTAGGAATTTTAGTTTTAGCTGGCTCTTTAATGGGAGGGAAGATAAC
TGAAGCAGTTAAAGAATTAAGGAAAAGACTGGCATTCCCGTGATAAGCTTAAAGATGTTTGGCTCTGT
TCCTAAGGTTGCTGATTTGGTTGTTGGAGACCCATTGCAGGCAGGGGTTTTAGCTGTTATGGCTATTGCT
GAAACAGCAAAATTTGATATAAATAAGGTTAAAGGTAGGGTGCTATAAAGATAATTTAATAATTTTTGA
TGAAACCGAAGCGTTAGCTTTGGGTTATGAACTCCATGATTTTCATTTAATTTTTTCTATTAATTTTCT
CCTAAAAAGTTTCTTTAACATAAATAAGGTTAAAGGGAGAGCTCTATGATTGCTTCAAAAATACAAAG

ATTATTGATGTATATACTGGAGAGGTTGTTAAAGGAAATGTTGCAGTTGAGAGGGATAAAAATATCCTTTG
TGGATTTAAATGATGAAATTGATAAGATAATTGAAAAAATAAAGGAGGATGTTAAAGTTATTGACTTAA
AAGGAAAATATTTATCTCCAACATTTATAGATGGGCATATACATATAGAATCTTCCCATCTCATCCCATC
AGAGTTTGAGAAAATTTGTATTA AAAAGCGGAGTTAGCAAAGTAGTTATAGACCCGCATGAAATAGCAAA
TATTGCTGGAAAAGAAGGAATTTTGTATGTTGAATGATGCCAAAATTTTAGATGTCTATGTTATGCTTC
CTTCCTGTGTTCCAGCTACAAACTTAGAAACAAGTGGAGCTGAGATTACAGCAGAGAATATTGAAGAAC
TCATTCTTTAGATAATGTCTTAGGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. Supplementary Figures

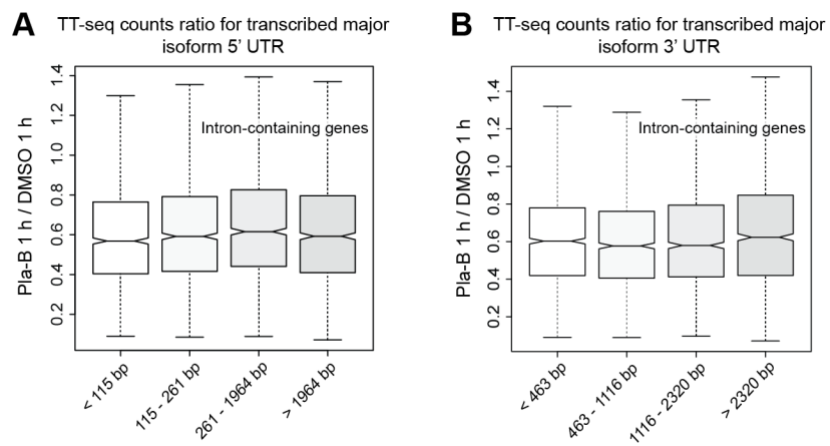


Figure S1. UTR length impact on intron-containing genes. Related to Figure 7.

(**A** and **B**) Ratio of 1 μ M Pla-B to DMSO antisense bias-corrected TT-seq read counts for 5' UTR (**A**) and 3' UTR (**B**) in four different major isoforms UTR length quartiles.

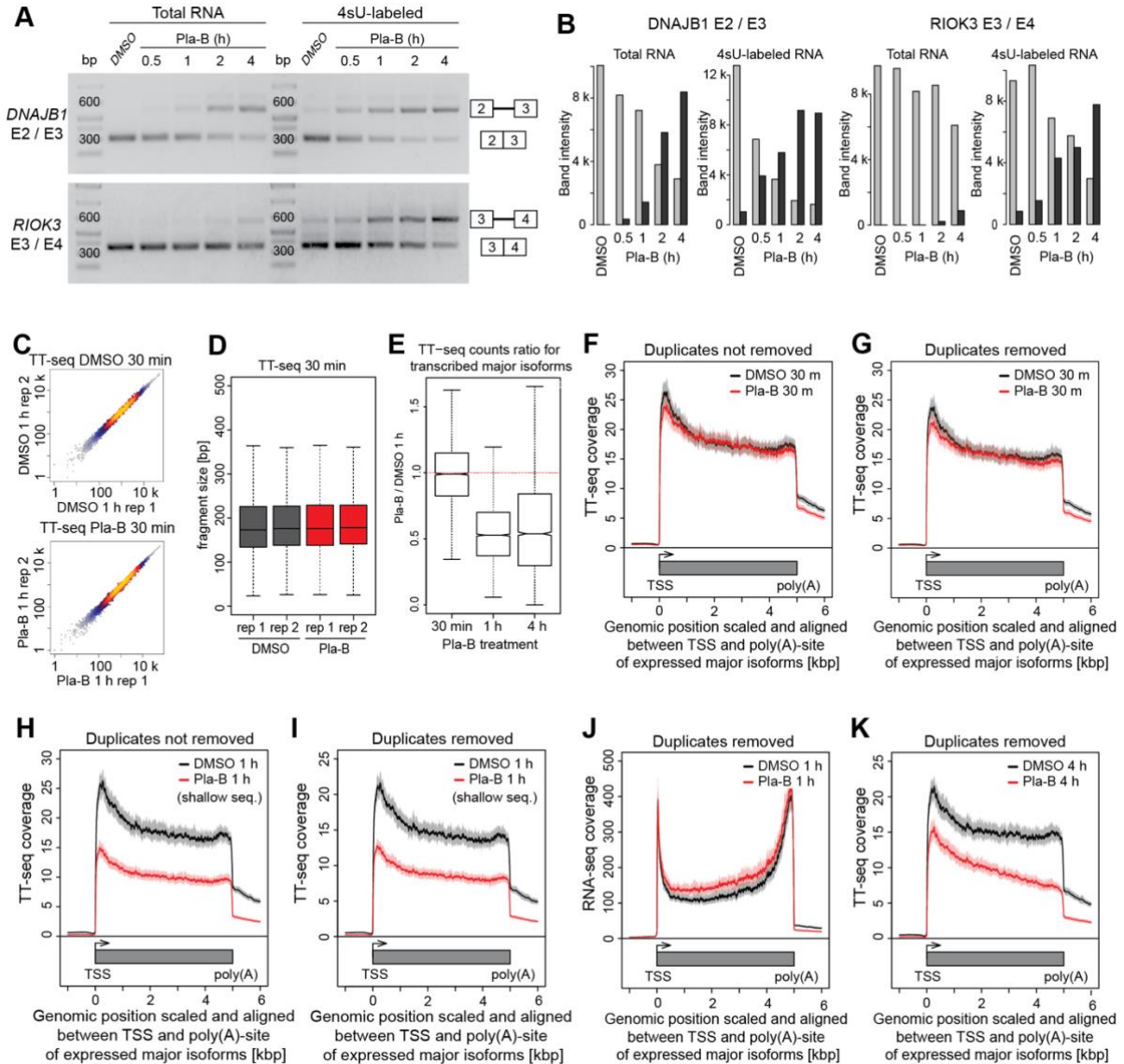


Figure S2. TT-seq monitors the decrease of RNA synthesis activity upon Pla-B treatment. Related to Figure 11, Figure 12 and Figure 13.

(A) Agarose gel showing spliced and unspliced RT-PCR products for total and 4sU-labeled enriched RNAs upon 30 min, 1 h, 2 h, and 4 h of 1 μ M Pla-B and 4 h DMSO treatments for regions spanning the exons 2 and 3 of *DNAJB1* (unspliced: 597 bp, spliced: 302 bp) and exons 3 and 4 of *RIOK3* (unspliced: 650 bp, spliced: 352 bp). See also Table S3. (B) Agarose gel band intensity (ImageJ) for spliced and unspliced RT-PCR products for total and 4sU-labeled RNAs. See also Table S3. (C) Scatter plots comparing TT-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with DMSO or 1 μ M Pla-B (batch 2) for 30 min. Spearman correlation of 1. (D) Fragment size distribution between samples. (E) Ratio of 1 μ M Pla-B to DMSO antisense bias-corrected TT-seq read counts for 30 min, 1 h, and 4 h treatment. (F and G) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 1 μ M Pla-B for 30 min without removing duplicated paired reads (F) and with duplicates removal (G). (H and I) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h for shallow sequenced data without removing duplicated paired reads (H) and with duplicate removal (I). (J) Metagene analysis comparing RNA-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h with duplicated removal (as in comparison to Figure 13E without duplicate removal). (K) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 1 μ M Pla-B for 4 h with duplicated removal (as in comparison to Figure 13H without duplicate removal). Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also Table S5.

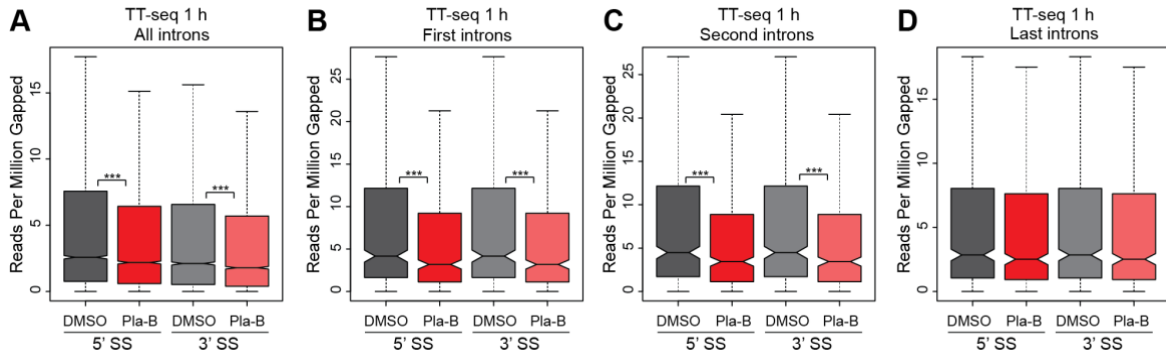


Figure S3. Alternative splicing analysis using CIGAR strings. Related to Figure 12.

Reads per million gapped alignments (RPMG) were calculated using the spliceSites R package upon 1 h DMSO or 1 μ M Pla-B treatment. (A) RPMG for all annotated introns for a total of 7,081 5' SS and 7,141 3' SS splice junctions. (B) RPMG for 549 spliced junctions on both 5' SS and 3' SS of first introns. (C) RPMG for 549 spliced junctions on both 5' SS and 3' SS of first introns (not last). (D) RPMG for 538 spliced junctions on both 5' SS and 3' SS of second introns (not last). (E) RPMG for 586 spliced junctions on both 5' SS and 3' SS of last introns. For B, C and D were only used identified gap alignments that comprised two consecutive exons.

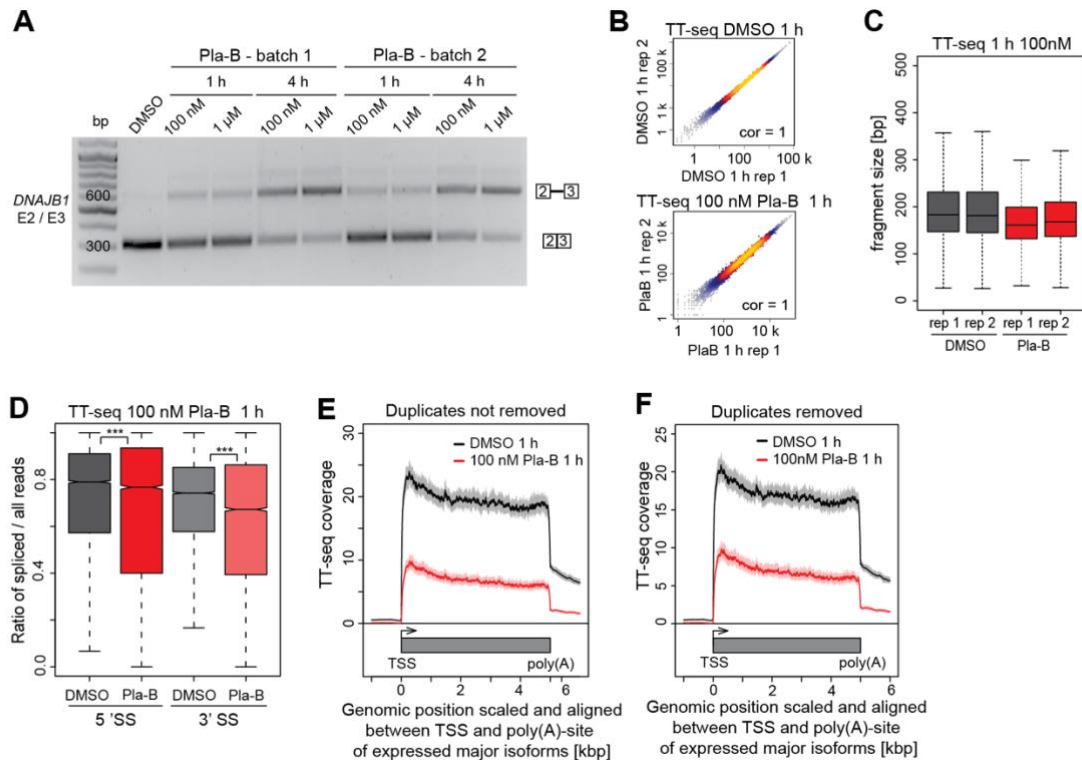


Figure S4 .TT-seq upon U2snRNP inhibition with 100nM Pla-B. Related to Results section 2

(A) Agarose gel showing spliced and unspliced RT-PCR products for region spanning exons 2 and 3 of DNAJB1 of total RNA (unspliced: 597 bp, spliced: 302 bp). Pla-B concentration and batch are indicated for each time point (batch 1: G0516, batch 2: B0420). See also **Table S3**. (B) Scatter plots comparing TT-seq replicates using antisense bias-corrected counts for major isoforms for samples treated with DMSO or 100 nM Pla-B (batch 2) for 1 h. Spearman correlation of 1. (C) TT-seq fragment size distribution for paired mapped reads upon 1 h DMSO or 100 nM Pla-B treatment. (D) Ratio of spliced reads over total unspliced and spliced reads upon 1 h DMSO or 100 nM Pla-B (batch 2). (***) p-value $< 2.2 \times 10^{-16}$ by Wilcoxon signed-rank test. (E and F) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 100 nM Pla-B (batch 2) without removing duplicated paired reads (E) and with duplicates removal (F). TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also **Table S5**.

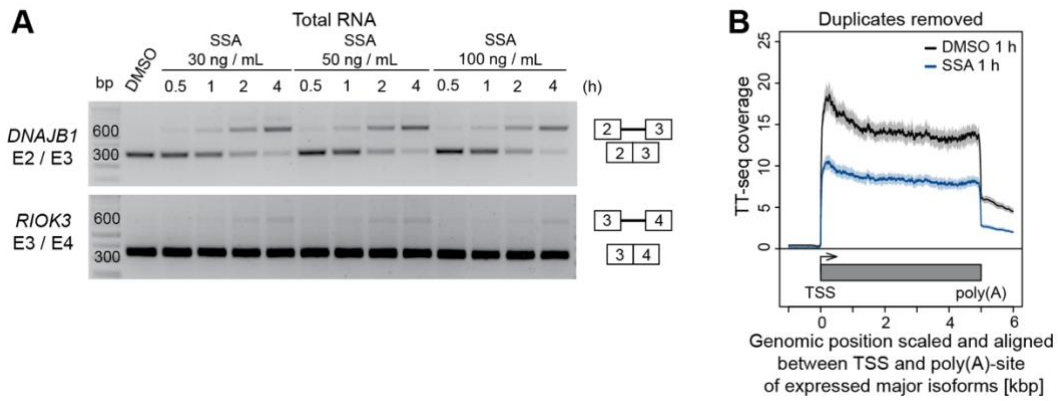


Figure S5. TT-seq monitors decrease of RNA synthesis activity upon SSA treatment. Related to Figure 15.

(A) Agarose gel showing spliced and unspliced RT-PCR products for total RNA upon 30 min, 1 h, 2 h and 4 h of 75 μ M U2 AMO treatment and 4 h 75 μ M Ctr AMO for regions spanning the exons 2 and 3 of *DNAJB1* (unspliced: 597 bp, spliced: 302 bp) and exons 3 and 4 of *RIOK3* (unspliced: 650 bp, spliced: 352 bp). See also **Table S3**. (B) Metagene analysis comparing TT-seq signal between cells treated with DMSO or 30 ng/mL SSA for 1 h with paired read duplicates removal (as in comparison to **Figure 15D** without duplicate removal). TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also **Table S5**.

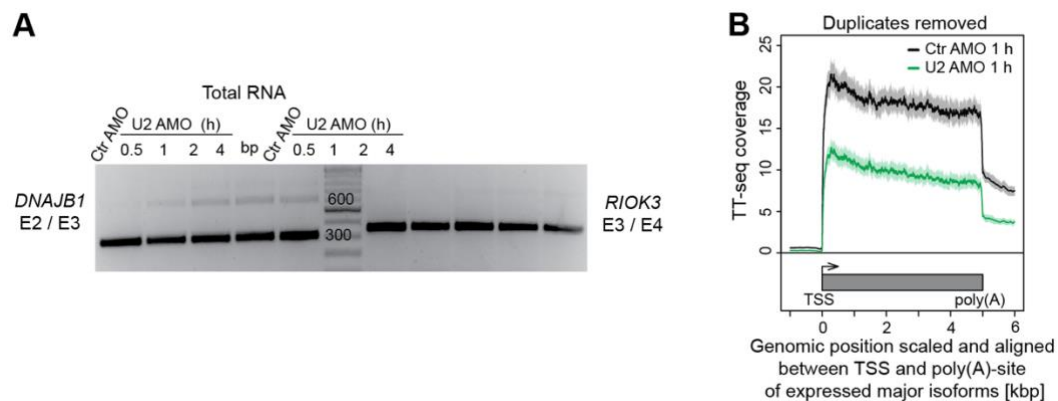


Figure S6. TT-seq monitors decrease of RNA synthesis activity upon U2 AMO treatment. Related to Figure 16.

(A) Agarose gel showing spliced and unspliced RT-PCR products for total RNA upon 30 min, 1 h, 2 h and 4 h of 30 ng/mL SSA treatment and 4 h DMSO for regions spanning the exons 2 and 3 of *DNAJB1* (unspliced: 597 bp, spliced: 302 bp) and exons 3 and 4 of *RIOK3* (unspliced: 650 bp, spliced: 352 bp). See also **Table S3**. (B) Metagene analysis comparing TT-seq signal between cells treated with 75 μ M Ctr AMO or 75 μ M U2 AMO for 1 h with paired read duplicates removal (as in comparison to **Figure 16D** without duplicate removal). TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also **Table S5**.

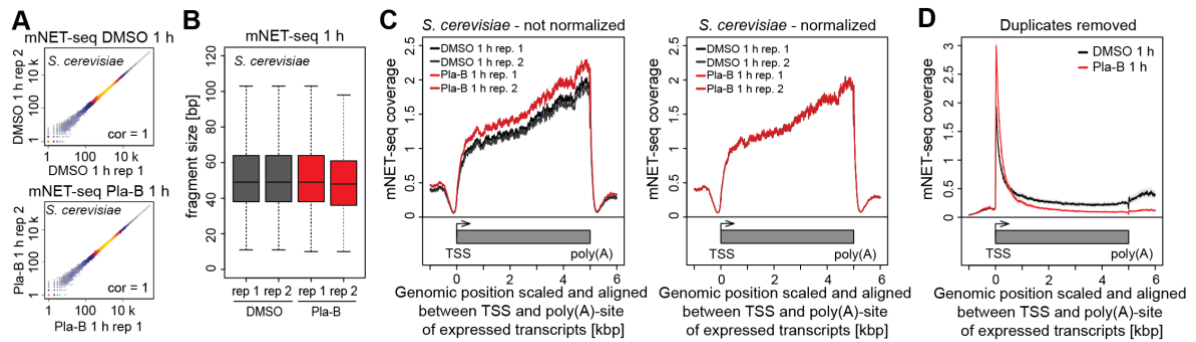


Figure S7. mNET-seq *S. cerevisiae* spike-ins normalization. Related to Figure 17.

(A) Scatter plots comparing mNET-seq replicates using antisense bias-corrected counts for *S. cerevisiae* transcripts for samples treated with DMSO and 1 μ M Pla-B for 1 h. Spearman correlation of 1. (B) mNET-seq fragment size distribution for paired reads mapped to the *S. cerevisiae* genome upon 1 h DMSO or 1 μ M Pla-B treatment. (C) Metagene analysis used to verify the *S. cerevisiae* spike-ins normalization method, comparing mNET-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h before spike-ins normalization (left) and after spike-ins normalization (right). Replicates are shown separately. mNET-seq coverage was averaged for 5639 *S. cerevisiae* expressed transcripts between TSS and poly(A)-site. (D) Metagene analysis comparing mNET-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h with paired read duplicates removal (as in comparison to **Figure 17C** without duplicate removal). TT-seq coverage was averaged for 5,535 major isoforms scaled between TSS and poly(A)-site. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also **Table S6**.

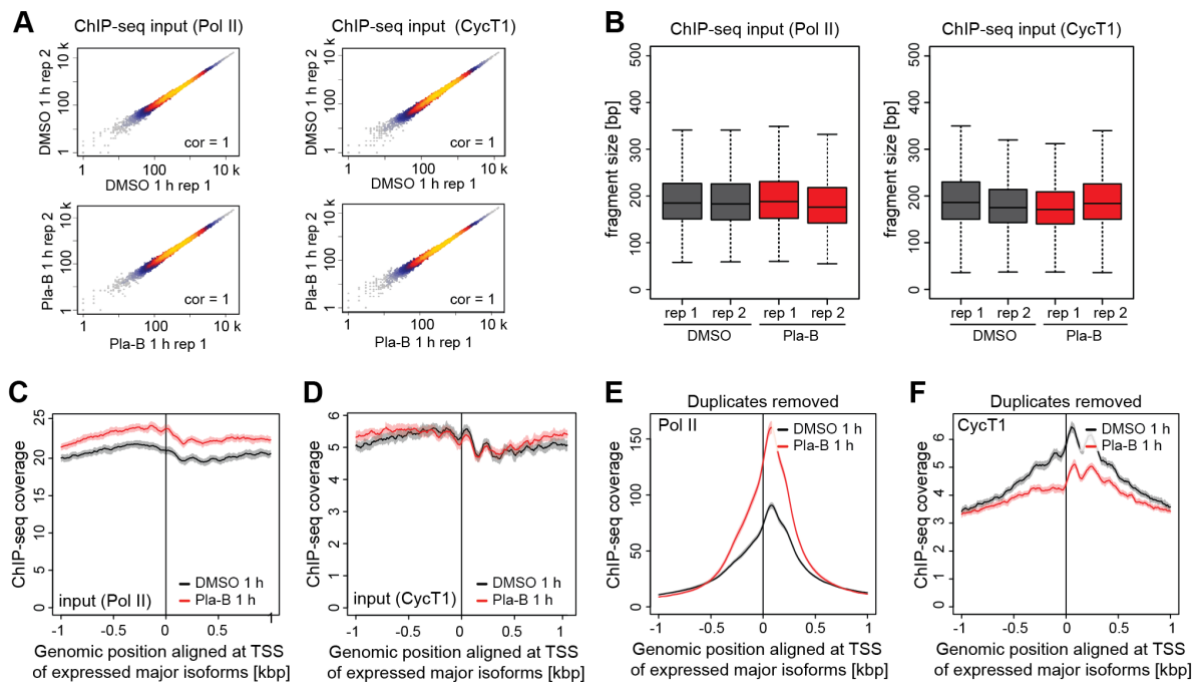


Figure S8. ChIP-seq input and duplicate removal. Related to Figure 20.

(A) Scatter plots comparing input used on Pol II IP (left) and input used on CycT1 (right) ChIP-seq replicates using counts for major isoforms for samples treated with DMSO and 1 μ M Pla-B. Spearman correlation of 1. (B) Fragment size distribution for paired mapped reads for input for Pol II IP experiment (left) and input for CycT1 (right) ChIP-seq upon 1 h DMSO or 1 μ M Pla-B treatment. (C and D) Metagene analysis comparing ChIP-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h for input used on Pol II IP (C) and input used on CycT1 IP (D). ChIP-seq coverage was aligned at the TSS and averaged for 5,465 major isoforms that exceed 1 kbp in length. (E and F) Metagene analysis comparing ChIP-seq signal between cells treated with DMSO or 1 μ M Pla-B for 1 h with paired read duplicates removal (as in comparison to **Figure 20C-D** without duplicate removal). ChIP-seq coverage for Pol II (E) and CycT1 (F) was aligned at the TSS and averaged for 5,465 major isoforms that exceed 1 kbp in length. Solid lines represent the averaged signal and the shaded area represents 95 % confidence interval of the mean (bootstrap). See also **Table S7**.

TT-seq Ctr vs HS 30 m

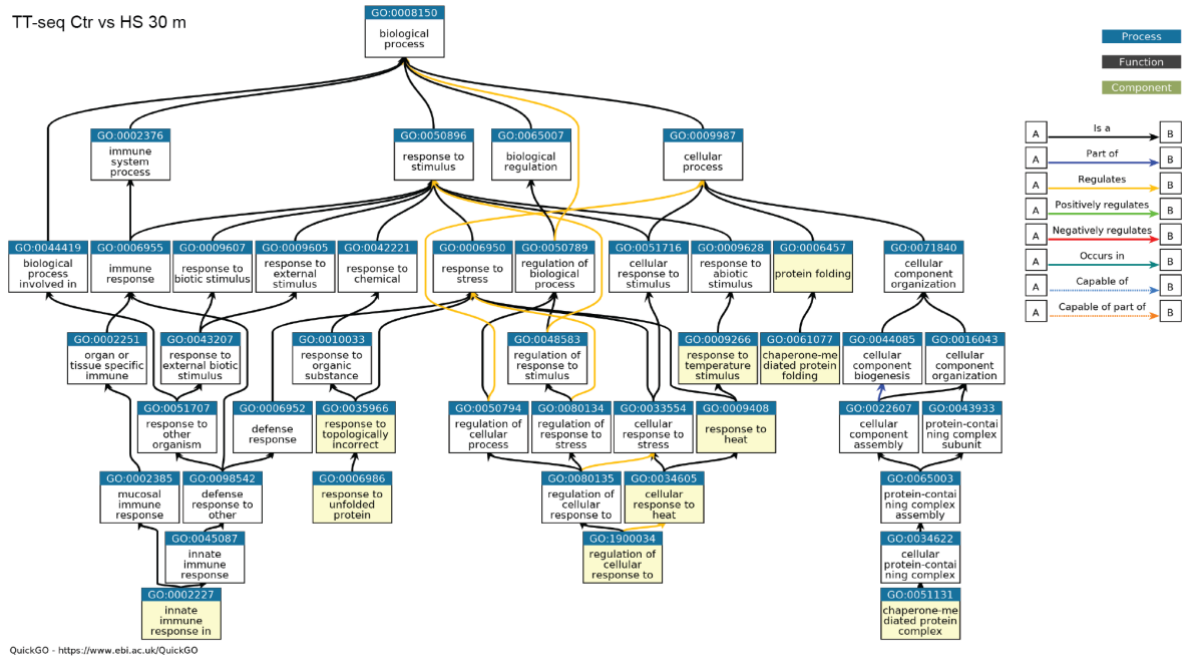


Figure S9. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon Ctr or HS treatment for 30 min (Gressel et al., 2019) (QuickGO). Related to Figure 23B.

TT-seq DMSO vs Pla-B 1 h

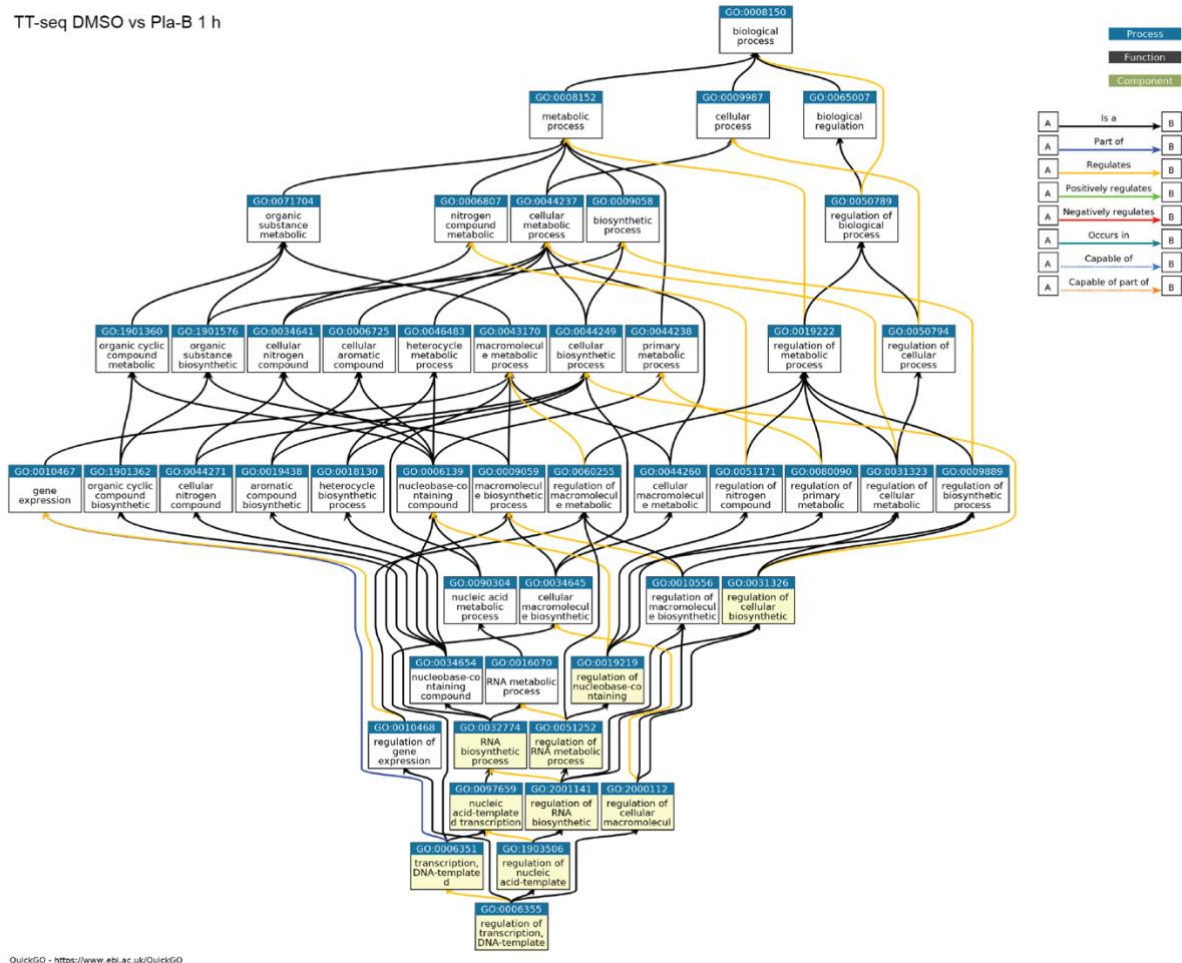
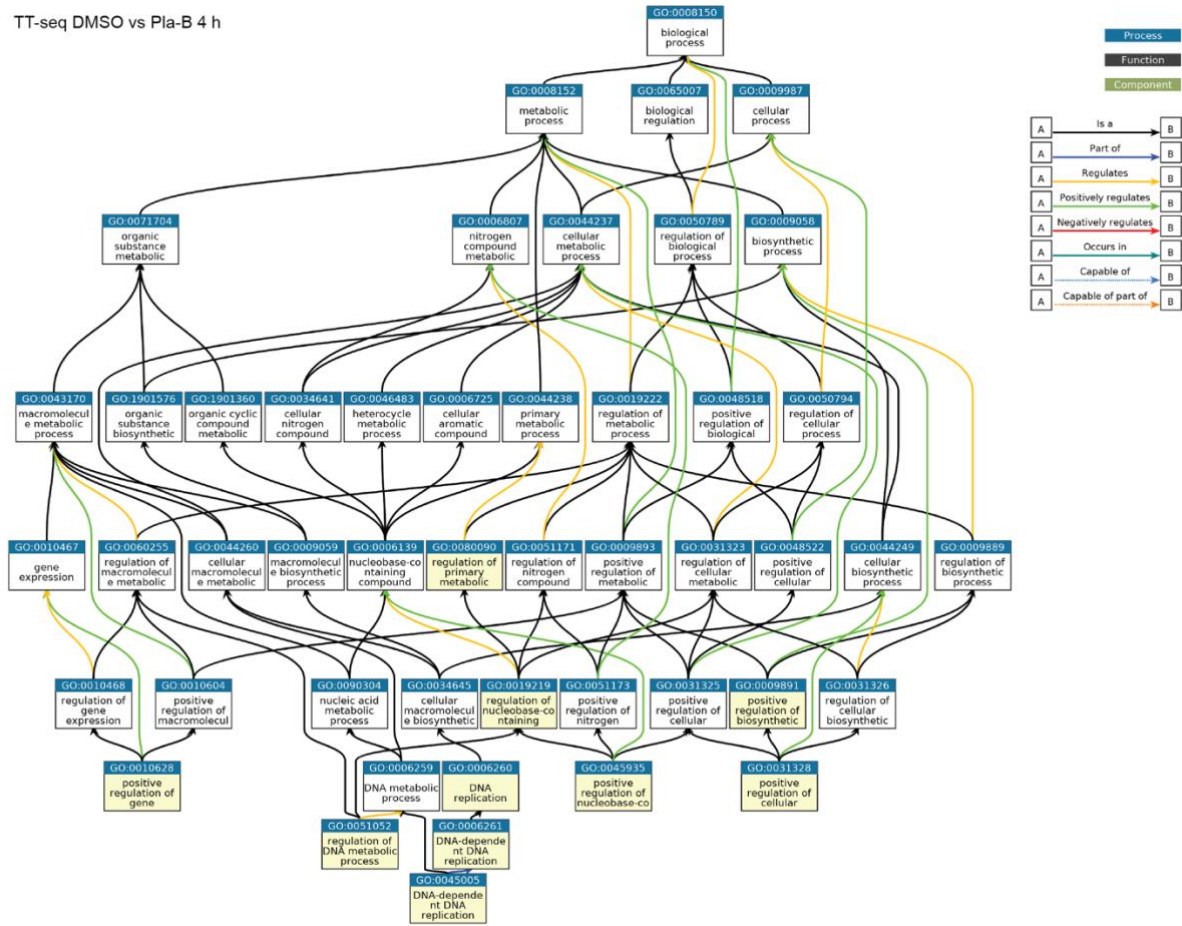


Figure S10. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 1 μ M Pla-B treatment for 1 h (QuickGO). Related to Figure 24A.

TT-seq DMSO vs Pla-B 4 h



QuickGO - <https://www.ebi.ac.uk/QuickGO>

Figure S11. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 1 μM Pla-B treatment for 4 h (QuickGO). Related to Figure 24B.

TT-seq DMSO vs SSA 1 h

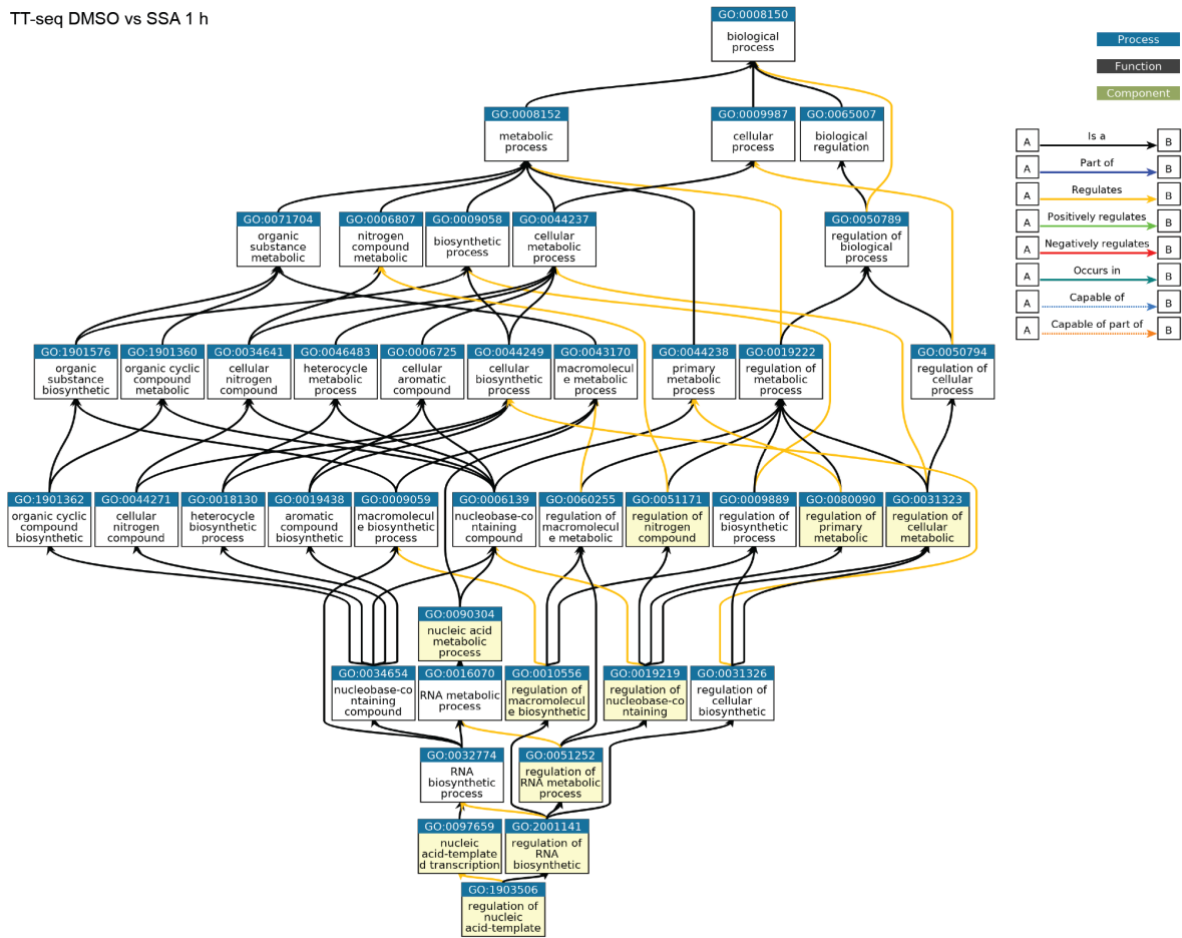


Figure S12. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 30 ng/mL SSA treatment for 1 h (QuickGO). Related to Figure 24C.

TT-seq Ctr AMO vs U2 AMO 1 h

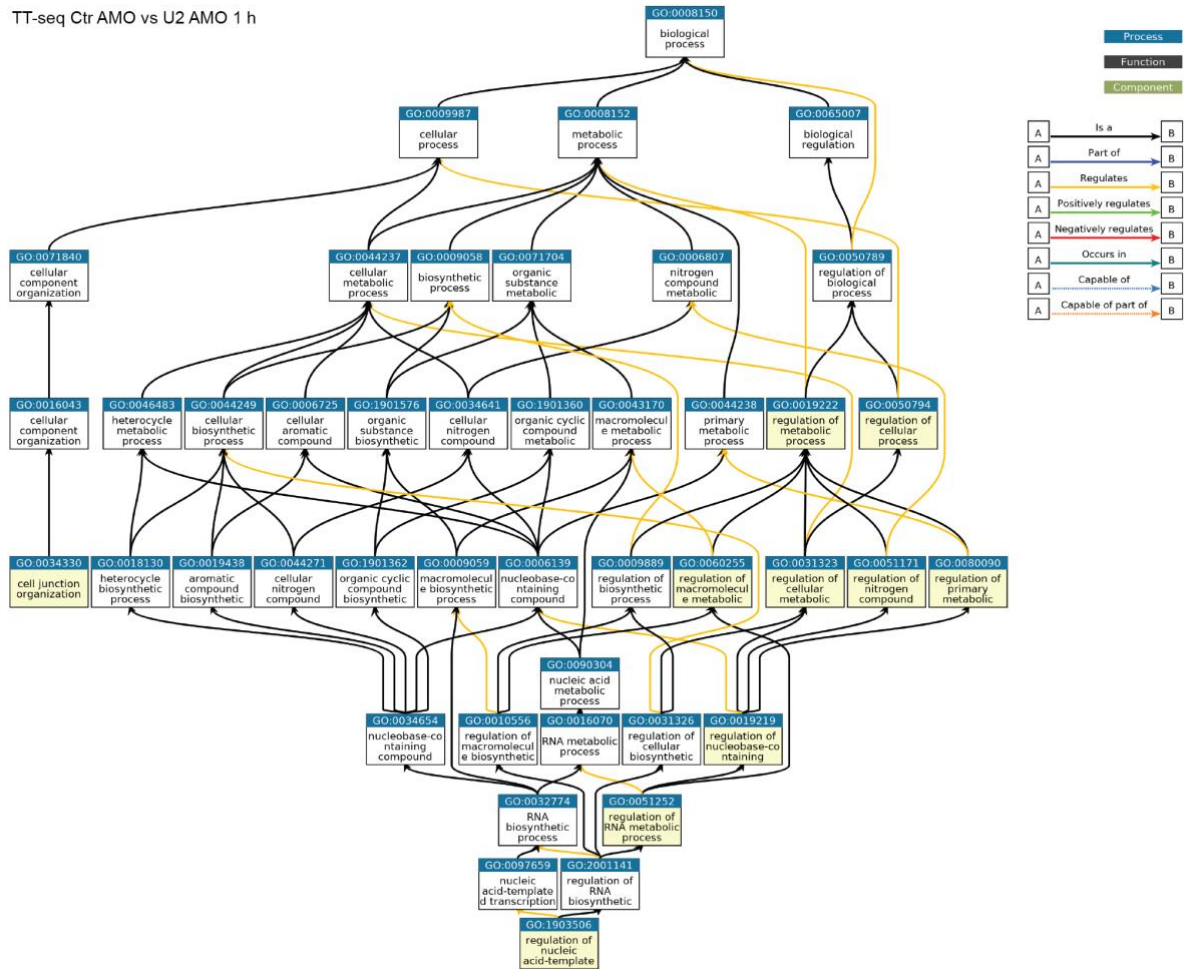


Figure S13. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon 75 μ M Ctr AMO or 75 μ M U2 AMO treatment for 1 h (QuickGO). Related to Figure 24D.

4. Supplementary Tables

Table S1. Overview of the datasets used in the present study. Related to Introduction section 2.

Method	Treatment	Cell type	GEO series	Source
TT-seq	30 min 1 μ M Pla-B or DMSO	K562	GSE148433	this study
TT-seq	1 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
RNA-seq	1 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
TT-seq	4 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
TT-seq	1 h 100 nM Pla-B or DMSO	K562	GSE148433	this study
TT-seq	30 ng/mL SSA or DMSO	K562	GSE148433	this study
TT-seq	75 μ M U2 AMO or 75 μ M Ctr AMO	K562	GSE148433	this study
mNET-seq	1 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
ChIP-seq Pol II	1 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
ChIP-seq CycT1	1 h 1 μ M Pla-B or DMSO	K562	GSE148433	this study
mNET-seq	4 h 1 μ M Pla-B or DMSO	HeLa	GSE81662	(Schlackow et al., 2017)
TT-seq	30 min 42°C HS (or 37°C Ctr)	K562	GSE123980	(Gressel et al., 2019)

Table S2. Reagents and resources used for the experimental methods for the present work. Related to Supplementary information section 1.

Reagent or resource	Source	Identifier
Antibodies		
Mouse monoclonal against RNA polymerase II	Diagenode	Cat#C15200004, RRID:AB_2728744
Rabbit monoclonal against Cyclin T1	Cell Signaling	Cat#D1B6G, RRID:AB_2799973
Chemicals, Peptides, and Recombinant Proteins		
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat#D2438
Pladienolide B	Santa Cruz	Cat#sc-391691
Spliceostatin A	Gift from Vladimir Pena	-
4-thiouridine	Carbosynth	Cat#NT06186
Empigen	Sigma-Aldrich	Cat#30326
Formaldehyde 16% concentrate stock methanol-free	Thermo Fisher Scientific	Cat#28908
Critical Commercial Assays		
Plasmo Test Mycoplasma Detection Kit	InvivoGen	Cat#rep-pt1
CellTiter 96® AQueous One Solution Cell Proliferation Assay	Promega	Cat#G3580
NEON system	Invitrogen	Cat#MPK10025
Ovation Universal RNA-seq System	NuGEN	Cat#0343-32
TruSeq Small RNA Library Kit	Illumina	Cat#RS-200-0048
NEBNext® Ultra™ II DNA Library Prep Kit	NEB	Cat#E7645S
Experimental Models: Cell Lines		
Human: K562 (female)	DSMZ	ACC-10, RRID:CVCL_0004
Human: HeLa S3 (female)	GBF (Helmholtz Center for Infection Research, Brunswick)	RRID:CVCL_0058
Saccharomyces cerevisiae strain: BY4741	Euroscarf	ACC-Y00000
Drosophila melanogaster: Schneider 2 (S2)	DSMZ	ACC-130 RRID:CVCL_Z232
Oligonucleotides		
Ctr AMO: CCTCTTACCTCAGTTACAATTTATA	GeneTools	-
U2 AMO: TGATAAGAACAGATACTACACTTGA	GeneTools	-
Recombinant DNA		
pCMV-GLuc 2 control plasmid	NEB	Cat#N8081
pGL4.10[luc2] vector	Promega	Cat#E6651
pSV40-CLuc control plasmid	NEB	Cat#N0318S
pCMV-GLuc 2 intron inserted plasmid	This manuscript	-

Table S3. Primers used for RT-PCR. Related to Figure S2, Figure S4, Figure S5 and Figure S6.

Primer sequences used for total and 4sU-labeled RNA semiquantitative-PCRs

Gene	Primer sequence
DNAJB1_RT-PCR_agarose gel_F	GAACCAAATCACTTTCCCAAGGAAGG
DNAJB1_RT-PCR_agarose gel_R	AATGAGGTCCCCACGTTTCTCGGGTGT
RIOK3_RT-PCR_agarose gel_F	GCTGAAGGACCATTATTACTGGAG
RIOK3_RT-PCR_agarose gel_R	TTCTTGCTGTGTTCTTTCTCCACA

Table S4. Names and length of Human, *S. cerevisiae* and *D. melanogaster* genomes. Related to Methods section 2.2 and 2.3

chr: chromosome

Human		<i>S. cerevisiae</i>		<i>D. melanogaster</i>	
chr name	chr length	chr name	chr length	chr name	chr length
chr1	248956422	I	230218	2L	23513712
chr2	242193529	II	813184	2R	25286936
chr3	198295559	III	316620	3L	28110227
chr4	190214555	IV	1531933	3R	32079331
chr5	181538259	V	576874	4	1348131
chr6	170805979	VI	270161	X	23542271
chr7	159345973	VII	1090940	Y	3667352
chr8	145138636	VIII	562643	mitochondrion _genome	19524
chr9	138394717	IX	439888		
chr10	133797422	X	745751		
chr11	135086622	XI	666816		
chr12	133275309	XII	1078177		
chr13	114364328	XIII	924431		
chr14	107043718	XIV	784333		
chr15	101991189	XV	1091291		
chr16	90338345	XVI	948066		
chr17	83257441	Mito	85779		
chr18	80373285				
chr19	58617616				
chr20	64444167				
chr21	46709986				
chr22	50818468				
chrX	156040895				
chrY	57227415				

Table S5. Sequencing statistics for the TT-seq data used in this analysis. Related to Results section 1 and 2 and Figure S2, Figure S4, Figure S5 and Figure S6.

Two replicates are indicated as “rep 1” and “rep 2”. Paired-end libraries (2 x 75 bp if not stated otherwise) were sequenced on the Illumina NextSeq500. TT-seq samples treated with DMSO or 1 μ M Pla-B for 1 h were subjected to a shallow sequencing (2 x 75 bp) and further sequenced deeper with 2 x 150 bp reads. The two were concatenated (75 bp + 150 bp) for further analysis. The percentage of deduplicated reads corresponds to the percentage of distinct reads, formed by singletons (reads found one single time) and distinct duplicated reads (distinct reads duplicated n times). Uniquely mapped reads number correspond to the paired reads mapped to the genome with STAR (Dobin et al., 2013) and filtered with Samtools (Li et al., 2009). Duplicated paired reads were calculated with Picard tools (Broad Institute, 2019).

Samples	Treatment	Sequenced reads	Deduplicated sequenced reads (% read 1, read 2)	Uniquely mapped reads	Duplicated paired reads	GEO number
TT-seq 30 min	DMSO rep 1	35 514 609	71.71, 63.44	33 943 592	19.27	-
	DMSO rep 2	40 2846 911	67.03, 58.45	40 918 852	22.46	-
	Pla-B rep 1	42 925 965	65.97, 58.02	41 055 445	25.12	-
	Pla-B rep 2	42 811 737	65.53, 57.26	40 688 441	25.31	-
TT-seq 1 h (75 bp)	DMSO rep 1	36 784 363	66.10, 58.07	35 092 648	25.61	GSM4471381
	DMSO rep 2	54 075 738	56.03, 48.65	51 362 916	36.35	GSM4471382
	Pla-B rep 1	45 519 747	60.16, 53.32	43 291 101	32.08	GSM4471383
	Pla-B rep 2	47 769 843	59.25, 52.71	45 243 934	32.99	GSM4471384
TT-seq 1 h (150 bp)	DMSO rep 1	79 334 466	53.07, 46.01	100 388 896	41.78	GSM4471385
	DMSO rep 2	91 357 858	45.98, 38.91	125 106 525	53.66	GSM4471386
	Pla-B rep 1	800 34 812	50.69, 43.88	106 774 758	45.41	GSM4471387
	Pla-B rep 2	100 620 427	47.73, 41.70	127 100 081	47.73	GSM4471388
TT-seq 1 h (75 bp + 150 bp)	DMSO rep 1	116 118 829	50.43, 43.27	100 388 896	41.78	-
	DMSO rep 2	145 433 596	41.87, 35.74	125 106 525	53.66	-
	Pla-B rep 1	125 554 559	47.38, 41.27	106 774 758	45.41	-
	Pla-B rep 2	148 390 270	45.01, 39.50	127 100 081	47.73	-
TT-seq 4 h	DMSO rep 1	41 527 090	66.51, 57.62	39 622 355	24.54	GSM4471393
	DMSO rep 2	38 030 584	66.98, 58.40	36 288 697	24.16	GSM4471394
	Pla-B rep 1	37 943 066	60.48, 52.99	36 054 517	29.75	GSM4471395
	Pla-B rep 2	44 403 933	58.88, 51.22	41 654 987	30.31	GSM4471396
RNA-seq 1 h	DMSO rep 1	25 555 667	54.15, 51.46	21 726 274	31.86	GSM4471389
	DMSO rep 2	25 352 910	41.89, 39.37	21 661 759	48.99	GSM4471390
	Pla-B rep 1	26 470 018	53.83, 51.19	22 763 649	32.35	GSM4471391
	Pla-B rep 2	25 278 414	50.47, 47.84	21 336 150	39.36	GSM4471392

TT-seq 100 nM Pla-B 1 h	DMSO rep 1	44 612 305	71.31, 61.83	40 971 048	18.89	GSM4970206
	DMSO rep 2	59 984 807	67.21, 58.27	54 863 640	23.84	GSM4970207
	Pla-B rep 1	52 893 216	36.20, 31.97	48 186 952	65.06	GSM4970208
	Pla-B rep 2	46 519 479	48.34, 41.07	42 198 478	46.03	GSM4970209
TT-seq SSA 1 h	DMSO rep 1	48 437 487	67.33, 58.47	44 895 714	24.73	GSM4970210
	DMSO rep 2	43 831 288	65.1, 57.23	40 724 390	27.91	GSM4970211
	SSA rep 1	53 256 104	63.84, 57.33	49 148 708	28.73	GSM4970212
	SSA rep 2	50 337 889	65.36, 57.88	46 615 663	26.97	GSM4970213
TT-seq U2 AMO 1 h	Ctr AMO rep 1	48 849 741	64.95, 55.49	44 676 201	25.04	GSM4970214
	Ctr AMO rep 2	43 150 243	70.5, 61.38	39 428 950	20.01	GSM4970215
	U2 AMO rep 1	39 322 274	57.07, 48.79	35 628 696	35.04	GSM4970216
	U2 AMO rep 2	42 766 764	60.19, 51.54	38 293 135	30.52	GSM4970217

Table S6. Sequencing statistics for the mNET-seq data used in this analysis. Related to Results section 3 and Figure S7.

Two replicates are indicated as “rep 1” and “rep 2”. Paired-end libraries (2 x 42 bp) were sequenced on the Illumina NextSeq500. The percentage of deduplicated reads corresponds to the percentage of distinct reads, formed by singletons (reads found one single time) and distinct duplicated reads (distinct reads that are duplicated n times). Uniquely mapped reads number correspond to the paired reads mapped to the human hg38 (GRCh38) and *S. cerevisiae* sacCer3 (R64-1-1) combined genome with STAR (Dobin et al., 2013) and filtered with Samtools (Li et al., 2009). Duplicated paired reads were calculated with Picard tools (Broad Institute, 2019).

Treatment		DMSO rep 1	DMSO rep 2	Pla-B rep 1	Pla-B rep 2
Sequenced reads		213 855 281	195 923 363	185 903 379	169 665 589
Deduplicated sequenced reads (% read 1, read 2)		48.92, 49.27	51.71, 51.68	46.93, 47.44	47.41, 47.87
Uniquely mapped reads	Total	137 746 586	125 218 396	122 226 432	113 651 005
	Human	113 284 917	101 935 180	94 567 048	85 705 434
	Human (%)	82.24	81.41	77.37	75.41
	<i>S. cerevisiae</i>	24 461 669	23 283 216	27 659 384	27 945 571
	<i>S. cerevisiae</i> (%)	17.76	18.59	22.63	24.59
Duplicated paired reads		54.07	51.56	55.33	55.16
GEO number		GSM4471397	GSM4471398	GSM4471399	GSM4471400

Table S7. Sequencing statistics for the ChIP-seq data used in this analysis. Related to Results section 5 and Figure S8.

Two replicates are indicated as “rep 1” and “rep 2”. Paired-end libraries (2 x 75 bp for Pol II and 2 x 42 bp for CycT1) were sequenced on the Illumina NextSeq500. Pol II and CycT1 IP were performed in distinct experiments and therefore were generated two input samples. The percentage of deduplicated reads corresponds to the percentage of distinct reads, formed by singletons (reads found one single time) and distinct duplicated reads (distinct reads duplicated n times). Uniquely mapped reads number correspond to the paired reads mapped to the human hg38 (GRCh38) and *D. melanogaster* dm6 (BDGP6.28) combined genome with STAR (Dobin et al., 2013) and filtered with Samtools (Li et al., 2009). Duplicated paired reads were calculated with Picard tools (Broad Institute, 2019).

Sample	Treatment	Sequenced reads	Deduplicated sequenced reads (% read 1, read 2)	Duplicated paired reads	GEO number
input 1	DMSO rep 1	42 143 914	95.38, 95.48	2.91	GSM4970218
	DMSO rep 2	41 068 397	95.70, 95.70	2.93	GSM4970219
	Pla-B rep 1	43 973 742	95.25, 95.33	3.32	GSM4970220
	Pla-B rep 2	55 890 334	95.08, 95.14	3.31	GSM4970221
Pol II	DMSO rep 1	19 135 434	92.93, 93.61	7.81	GSM4970222
	DMSO rep 2	19 431 013	93.37, 93.92	6.77	GSM4970223
	Pla-B rep 1	29 352 773	92.01, 92.59	8.09	GSM4970224
	Pla-B rep 2	10 024 957	94.23, 94.91	5.98	GSM4970225
input 2	DMSO rep 1	31 787 818	94.20, 94.37	2.69	GSM4970226
	DMSO rep 2	32 0544 51	94.71, 94.64	2.35	GSM4970227
	Pla-B rep 1	40 785 308	93.89, 93.88	2.75	GSM4970228
	Pla-B rep 2	42 160 671	94.01, 93.89	2.78	GSM4970229
CycT1	DMSO rep 1	30 000 866	93.36, 93.04	3.66	GSM4970230
	DMSO rep 2	31 438 774	93.07, 93.17	3.57	GSM4970231
	Pla-B rep 1	40 730 438	92.46, 92.64	3.23	GSM4970232
	Pla-B rep 2	28 048 798	93.27, 93.43	3.61	GSM4970233

Table S8. Reads distribution per genome. Related to Results section 5.

Uniquely mapped reads number correspond to the paired reads mapped to the genome with STAR (2.6.0, RRID:SCR_015899) (Dobin et al., 2013) and filtered with Samtools (1.3.1, RRID:SCR_002105) (Li et al., 2009).

Sample	Treatment	Uniquely mapped reads				
		Total	Human	Human (%)	D. melanogaster	D. melanogaster (%)
Input 1	DMSO rep 1	38494817	38458645	99.91	36172	0.09
	DMSO rep 2	37717106	37668833	99.87	48273	0.13
	Pla-B rep 1	40165741	40120100	99.89	45641	0.11
	Pla-B rep 2	51112255	51067169	99.91	45086	0.09
Pol II	DMSO rep 1	14152608	14077899	99.47	74709	0.53
	DMSO rep 2	13297544	13245849	99.61	51695	0.39
	Pla-B rep 1	22928355	22823066	99.54	105289	0.46
	Pla-B rep 2	5915592	5896421	99.68	19171	0.32
Input 2	DMSO rep 1	28595714	28517919	99.73	77795	0.27
	DMSO rep 2	28796195	28731252	99.77	64943	0.23
	Pla-B rep 1	36576410	36500999	99.79	75411	0.21
	Pla-B rep 2	37926653	37842232	99.78	84421	0.22
CycT1	DMSO rep 1	26304890	26191607	99.57	113283	0.43
	DMSO rep 2	27615374	27503670	99.60	111704	0.40
	Pla-B rep 1	35460620	35326434	99.62	134186	0.38
	Pla-B rep 2	24344919	24259236	99.65	85683	0.35

Table S9. Overview of upregulated genes between 1 h DMSO or 1 μ M Pla-B treatment. Related to Results section 7 and Figure 23C.

Gene description from Ensembl annotation. BST2 gene was also upregulated genes upon 30 min HS.

Transcript identifier	Gene name	Gene description
NM_012094	PRDX5	peroxiredoxin 5 [Source:HGNC Symbol;Acc:HGNC:9355]
NM_001003	RPLP1	ribosomal protein lateral stalk subunit P1 [Source:HGNC Symbol;Acc:HGNC:10372]
NM_005332	HBZ	hemoglobin subunit zeta [Source:HGNC Symbol;Acc:HGNC:4835]
NM_194434	VAPA	VAMP associated protein A [Source:HGNC Symbol;Acc:HGNC:12648]
NM_001863	COX6B1	cytochrome c oxidase subunit 6B1 [Source:HGNC Symbol;Acc:HGNC:2280]
NM_198589	BSG	basigin (Ok blood group) [Source:HGNC Symbol;Acc:HGNC:1116]
NM_000041	APOE	apolipoprotein E [Source:HGNC Symbol;Acc:HGNC:613]
NM_001172667	FAM178B	FAM178B family with sequence similarity 178 member B [Source:HGNC Symbol;Acc:HGNC:28036]
NM_004335	BST2	bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:HGNC:1119]
NM_017854	TMEM160	transmembrane protein 160 [Source:HGNC Symbol;Acc:HGNC:26042]
NM_013335	GMPPA	GDP-mannose pyrophosphorylase A [Source:HGNC Symbol;Acc:HGNC:22923]

Table S10. Overview of upregulated genes between 4 h DMSO or 1 μ M Pla-B treatment. Related to Results section 7 and Figure 23D.

Gene description from Ensembl annotation. ARL6IP5, MSMO1 and SQLE genes were also upregulated upon 30 min HS treatment.

Transcript identifier	Gene name	Gene description
NM_194434	VAPA	VAMP associated protein A [Source:HGNC Symbol;Acc:HGNC:12648]
NM_002167	ID3	inhibitor of DNA binding 3, HLH protein [Source:HGNC Symbol;Acc:HGNC:5362]
NM_001305066	TMEM59	transmembrane protein 59 [Source:HGNC Symbol;Acc:HGNC:1239]
NM_021970	LAMTOR3	late endosomal/lysosomal adaptor, MAPK and MTOR activator 3 [Source:HGNC Symbol;Acc:HGNC:15606]
NM_001330663	HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1 [Source:HGNC Symbol;Acc:HGNC:5007]
NM_005648	ELOC	elongin C [Source:HGNC Symbol;Acc:HGNC:11617]
NM_006407	ARL6IP5	ADP ribosylation factor like GTPase 6 interacting protein 5 [Source:HGNC Symbol;Acc:HGNC:16937]
NM_006745	MSMO1	methylsterol monooxygenase 1 [Source:HGNC Symbol;Acc:HGNC:10545]
NM_003129	SQLE	squalene epoxidase [Source:HGNC Symbol;Acc:HGNC:11279]

Table S11. Overview of upregulated genes between 1 h DMSO or 30 ng/mL SSA treatment. Related to Results section 7 and Figure 23E.

Gene description from Ensembl annotation.

Transcript identifier	Gene name	Gene description
NM_017828	COMMD4	COMM domain containing 4 [Source:HGNC Symbol;Acc:HGNC:26027]
NM_005332	HBZ	hemoglobin subunit zeta [Source:HGNC Symbol;Acc:HGNC:4835]
NM_194434	VAPA	VAMP associated protein A [Source:HGNC Symbol;Acc:HGNC:12648]

Table S12. Overview of upregulated genes between 1 h 75 μ M Ctr AMO or 75 μ M U2 AMO treatment. Related to Results section 7 and Figure 23F.

Gene description from Ensembl annotation. AMO, AZIN1 and BA62 genes were also upregulates upon 30 min HS treatment.

Transcript identifier	Gene name	Gene description
NM_005252	FOS	Fos proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3796]
NM_014117	C16orf72	chromosome 16 open reading frame 72 [Source:HGNC Symbol;Acc:HGNC:30103]
NM_000981	RPL19	ribosomal protein L19 [Source:HGNC Symbol;Acc:HGNC:10312]
NM_194434	VAPA	VAMP associated protein A [Source:HGNC Symbol;Acc:HGNC:12648]
NM_014280	DNAJC8	DnaJ heat shock protein family (Hsp40) member C8 [Source:HGNC Symbol;Acc:HGNC:15470]
NM_024700	SNIP1	Smad nuclear interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:30587]
NM_001354602	TMEM269	transmembrane protein 269 [Source:HGNC Symbol;Acc:HGNC:52381]
NM_001009899	USF3	upstream transcription factor family member 3 [Source:HGNC Symbol;Acc:HGNC:30494]
NM_005908	MANBA	mannosidase beta [Source:HGNC Symbol;Acc:HGNC:6831]
NM_005275	GNL1	G protein nucleolar 1 (putative) [Source:HGNC Symbol;Acc:HGNC:4413]
NM_001527	HDAC2	histone deacetylase 2 [Source:HGNC Symbol;Acc:HGNC:4853]
NM_002489	NDUFA4	NDUFA4, mitochondrial complex associated [Source:HGNC Symbol;Acc:HGNC:7687]
NM_015713	RRM2B	ribonucleotide reductase regulatory TP53 inducible subunit M2B [Source:HGNC Symbol;Acc:HGNC:17296]
NM_001301668	AZIN1	antizyme inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:16432]
NM_001206	KLF9	Kruppel like factor 9 [Source:HGNC Symbol;Acc:HGNC:1123]
NM_033198	PIGS	phosphatidylinositol glycan anchor biosynthesis class S [Source:HGNC Symbol;Acc:HGNC:14937]
NM_000998	RPL37A	ribosomal protein L37a [Source:HGNC Symbol;Acc:HGNC:10348]
NM_015508	TIPARP	TCDD inducible poly(ADP-ribose) polymerase [Source:HGNC Symbol;Acc:HGNC:23696]
NM_001319236	RPL34	ribosomal protein L34 [Source:HGNC Symbol;Acc:HGNC:10340]
NM_018321	BRX1	BRX1, biogenesis of ribosomes [Source:HGNC Symbol;Acc:HGNC:24170]
NM_001964	EGR1	early growth response 1 [Source:HGNC Symbol;Acc:HGNC:3238]
NM_004282	BAG2	BCL2 associated athanogene 2 [Source:HGNC Symbol;Acc:HGNC:938]
NM_006716	DBF4	DBF4 zinc finger [Source:HGNC Symbol;Acc:HGNC:17364]
NM_001354697	SECISBP2	SECIS binding protein 2 [Source:HGNC Symbol;Acc:HGNC:30972]

REFERENCES

- Abramczuk, M.K., Burkard, T.R., Rolland, V., Steinmann, V., Duchek, P., Jiang, Y., Wissel, S., Reichert, H., and Knoblich, J.A. (2017). The splicing Co-Factor barricade/Tat-SF1 is required for cell cycle and lineage progression in *Drosophila* neural stem cells. *Dev.* *144*, 3932–3945.
- Adewale, B.A. (2021). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr. J. Lab. Med.* *9*, 1340.
- Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Luhrmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* *31*, 2667–2682.
- Agirre, E., Oldfield, A.J., Bellora, N., Segelle, A., and Luco, R.F. (2021). Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat. Commun.* *12*, 1–16.
- Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H., and Roignant, J.Y. (2019). Promoter-proximal pausing mediated by the exon junction complex regulates splicing. *Nat. Commun.* *10*, 1–17.
- de Almeida, S.F., and Carmo-Fonseca, M. (2010). Cotranscriptional RNA checkpoints. *Epigenomics* *2*, 449–455.
- de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* *18*, 977–983.
- Alpert, T., Herzel, L., and Neugebauer, K.M. (2018). Perfect timing : splicing and transcription rates in living cells. *8*, 1–18.
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* *21*, 1–16.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq - a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. *N. Biotechnol.* *25*, 195–203.
- Anvar, S.Y., Allard, G., Tseng, E., Sheynkman, G.M., de Klerk, E., Vermaat, M., Yin, R.H., Johansson, H.E., Ariyurek, Y., den Dunnen, J.T., et al. (2018). Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol.* *19*, 46.
- Aoi, Y., Smith, E.R., Shah, A.P., Rendleman, E.J., Marshall, S.A., Woodfin, A.R., Chen, F.X., Shiekhattar, R., and Shilatifard, A. (2020). NELF regulates a promoter-proximal step distinct from RNA Pol II pause-release. *Mol. Cell* *78*, 261-274.e5.
- Aparicio, O., Geisberg, J. V., and Struhl, K. (2004). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr. Protoc. Cell Biol.* *23*, Chapter 17:Unit 17.7.
- Aslanzadeh, V., Huang, Y., Sanguinetti, G., and Beggs, J.D. (2018). Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Res.* *28*, 203–213.
- Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.L., Tapsin, S., Chan, Y.S., Tan, C.P., Sim, A.Y.L., et al. (2016). In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* *62*, 603–617.
- Azubel, M., Wolf, S.G., Sperling, J., and Sperling, R. (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol. Cell* *15*, 833–839.

- Azubel, M., Habib, N., Sperling, R., and Sperling, J. (2006). Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. *J. Mol. Biol.* *356*, 955–966.
- Baker, S.C., Bauer, S.R., Beyer, R.P., Brenton, J.D., Bromley, B., Burrill, J., Causton, H., Conley, M.P., Elespuru, R., Fero, M., et al. (2005). The external RNA controls consortium: A progress report. *Nat. Methods* *2*, 731–734.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823–837.
- Beckedorff, F., Blumenthal, E., Ferreira, L., Gaidosh, G., Shilatifard, A., Beckedorff, F., Blumenthal, E., Ferreira, L., Aoi, Y., Cingaram, P.R., et al. (2020). The human integrator complex facilitates transcriptional elongation by endonucleolytic cleavage of nascent transcripts. *CellReports* *32*, 107917.
- Bell, J.C., Jukam, D., Teran, N.A., Risca, V.I., Smith, O.K., Johnson, W.L., Skotheim, J.M., Greenleaf, W.J., and Straight, A.F. (2018). Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* *7*.
- Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* *15*, 163–175.
- Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* *150*, 53–64.
- Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* *2*, 754–765.
- Bieberstein, N.I., Oesterreich, F.C., Straube, K., and Neugebauer, K.M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Rep.* *2*, 62–68.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.
- Bonetti, A., Agostini, F., Suzuki, A.M., Hashimoto, K., Pascarella, G., Gimenez, J., Roos, L., Nash, A.J., Ghilotti, M., Cameron, C.J.F., et al. (2020). RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* *11*, 1–14.
- Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Géli, V., de Almeida, S.F., and Palancade, B. (2017). Introns protect eukaryotic genomes from transcription-associated genetic instability. *Mol. Cell* *67*, 608–621.e6.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* *132*, 311–322.
- Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F., et al. (2013). XFrom structure to systems: High-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* *154*, 775–788.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* *24*, 1774–1786.
- Brès, V., Gomes, N., Pickle, L., and Jones, K.A. (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev.* *19*, 1211–1226.
- Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E., and Palmiter, R.D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* *85*, 836–840.
- Broad Institute (2019). Picard Toolkit (Broad Institute).
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015a). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* *2015*, 21.29.1–21.29.9.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y.,

- and Greenleaf, W.J. (2015b). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* *523*, 486–490.
- Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M. (2005). Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.* *12*, 175–182.
- Carrillo Oesterreich, F., Herzelt, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* *165*, 372–381.
- Castillo-Guzman, D., Hartono, S., Sanz, L., and Chédin, F. (2020). SF3B1-targeted splicing inhibition triggers global alterations in transcriptional dynamics and R-loop metabolism. *BioRxiv* 2020.06.08.130583.
- Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* *53*, 1044–1052.
- Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., and Huang, X. (2015). Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* *16*, 30.
- Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. *Mol. Cell* *53*, 779–790.
- Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* *10*, 317.
- Chen, L., Chen, J.Y., Huang, Y.J., Gu, Y., Qiu, J., Qian, H., Shao, C., Zhang, X., Hu, J., Li, H., et al. (2018). The augmented R-Loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell* *69*, 412-425.e6.
- Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009). DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev.* *23*, 2765–2777.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* *460*, 479–486.
- Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018). Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. *Mol. Cell* *69*, 648-663.e7.
- Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* *44*, 667–678.
- Chu, T., Rice, E.J., Booth, G.T., Salamanca, H.H., Wang, Z., Core, L.J., Longo, S.L., Corona, R.J., Chin, L.S., Lis, J.T., et al. (2018). Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat. Genet.* *50*, 1553–1564.
- Church, M., Smith, K.C., Alhussain, M.M., Pennings, S., and Fleming, A.B. (2017). Sas3 and Ada2(Gcn5)-dependent histone H3 acetylation is required for transcription elongation at the de-repressed FLO1 gene. *Nucleic Acids Res.* *45*, 4413–4430.
- Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). *Curr. Protoc. Mol. Biol.* *1*.
- Coltri, P.P., dos Santos, M.G.P., and da Silva, G.H.G. (2019). Splicing and cancer: challenges and opportunities. *Wiley Interdiscip. Rev. RNA* e1527.
- Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagiseti, C., De La Grange, P., Horbinski, C., Fondufe-Mittendorf, Y.N., Webb, T.R., and Stamm, S. (2014). Sudemycin e influences alternative splicing and changes chromatin modifications. *Nucleic Acids Res.* *42*, 4947–4961.
- Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* *33*, 960–982.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing

- and divergent initiation at human promoters. *Science* (80-.). 322, 1845–1848.
- Corrionero, A., Miñana, B., and Valcárcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev.* 25, 445–459.
- Cossa, G., Roeschert, I., Prinz, F., Baluapuri, A., Silveira Vidal, R., Schüle-Völk, C., Chang, Y.C., Ade, C.P., Mastrobuoni, G., Girard, C., et al. (2020). Localized inhibition of Protein Phosphatase 1 by NUA1 promotes spliceosome activity and reveals a MYC-sensitive feedback control of transcription. *Mol. Cell* 77, 1322-1339.e11.
- Cramer, P. (2019). Eukaryotic transcription turns 50. *Cell* 179, 808–812.
- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* 94, 11456–11460.
- Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat. Methods* 3, 503–509.
- Cretu, C., Agrawal, A.A., Cook, A., Will, C.L., Fekkes, P., Smith, P.G., Lührmann, R., Larsen, N., Buonamici, S., and Pena, V. (2018). Structural basis of splicing modulation by antitumor macrolide compounds. *Mol. Cell* 70, 265-273.e8.
- Cusick, M.E., Herman, T.M., DePamphilis, M.L., and Wassarman, P.M. (1981). Structure of chromatin at deoxyribonucleic acid replication forks: pre-nucleosomal deoxyribonucleic acid is rapidly excised from replicating simian virus 40 chromosomes by micrococcal nuclease. *Biochemistry* 20, 6648–6658.
- Custódio, N., and Carmo-Fonseca, M. (2016). Co-transcriptional splicing and the CTD code. *Crit. Rev. Biochem. Mol. Biol.* 51, 395–411.
- Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H., and Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol. Cell* 29, 271–278.
- Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 27, 482–498.
- David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* 25, 972–982.
- Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. *Cell Rep.* 26, 2779-2791.e5.
- Davis, P.K., Ho, A., and Dowdy, S.F. (2001). Biological methods for cell-cycle synchronization of mammalian cells. *Biotechniques* 30, 1322–1331.
- Ding, F., and Elowitz, M.B. (2019). Constitutive splicing and economies of scale in gene expression. *Nat. Struct. Mol. Biol.* 26, 424–432.
- Doamekpor, S.K., Sanchez, A.M., Schwer, B., Shuman, S., and Lima, C.D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes Dev.* 28, 1323–1336.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dollinger, R., and Gilmour, D.S. (2021). Regulation of promoter proximal pausing of RNA polymerase II in metazoans. *J. Mol. Biol.*
- Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985-998.e8.
- Drexler, H.L., Choquet, K., Merens, H.E., Tang, P.S., Simpson, J.T., and Churchman, L.S. (2021). Revealing nascent RNA processing dynamics with nano-COP. *Nat. Protoc.* 16, 1343–1375.

- Duan, Z., Andronescu, M., Schutz, K., Lee, C., Shendure, J., Fields, S., Noble, W.S., and Anthony Blau, C. (2012). A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods* 58, 277–288.
- Dukat, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., and Jeltsch, A. (2019). H3K36me_{2/3} binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. *J. Mol. Biol.* 431, 5063–5074.
- Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191.
- Eaton, J.D., and West, S. (2020). Termination of transcription by RNA polymerase II : BOOM ! *Trends Genet.* xx, 1–12.
- Effenberger, K.A., Anderson, D.D., Bray, W.M., Prichard, B.E., Ma, N., Adams, M.S., Ghosh, A.K., and Jurica, M.S. (2014). Coherence between cellular responses and in vitro splicing inhibition for the anti-tumor drug pladienolide b and its analogs. *J. Biol. Chem.* 289, 1938–1947.
- Effenberger, K.A., Urabe, V.K., Prichard, B.E., Ghosh, A.K., and Jurica, M.S. (2016). Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA* 22, 350–359.
- Egan, B., Yuan, C.-C., Craske, M.L., Labhart, P., Guler, G.D., Arnott, D., Maile, T.M., Busby, J., Henry, C., Kelly, T.K., et al. (2016). An alternative approach to ChIP-Seq normalization enables detection of genome-wide changes in histone H3 lysine 27 trimethylation upon EZH2 inhibition. *PLoS One* 11, e0166438.
- Ehrensberger, A.H., Kelly, G.P., and Svejstrup, J.Q. (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* 154.
- Elrod, N.D., Henriques, T., Huang, K., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., Elrod, N.D., Henriques, T., Huang, K., Tatomer, D.C., et al. (2019). The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* 76, 738-752.e7.
- Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., and Ingles, C.J. (2002). Splicing and transcription-associated proteins PSF and p54nrb/NonO bind to the RNA polymerase II CTD. *RNA* 8, 1102–1111.
- Erhard, F., Baptista, M.A.P., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C.S., Theis, F.J., Saliba, A.E., and Dölken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* 571, 419–423.
- Erickson, B., Sheridan, R.M., Cortazar, M., and Bentley, D.L. (2018). Dynamic turnover of paused pol II complexes at human promoters. *Genes Dev.* 32, 1215–1225.
- Fededa, J.P., Petrillo, E., Gelfand, M.S., Neverov, A.D., Kadener, S., Nogués, G., Pelisch, F., Baralle, F.E., Muro, A.F., and Kornblihtt, A.R. (2005). A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol. Cell* 19, 393–404.
- Fiszbein, A., Krick, K.S., Begg, B.E., and Burge, C.B. (2019). Exon-mediated activation of transcription starts. *Cell* 179, 1551-1565.e17.
- Foley, S.W., and Gregory, B.D. (2016). Protein interaction profile sequencing (PIP-seq). *Curr. Protoc. Mol. Biol.* 2016.
- Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* 15, 1783–1795.
- Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929–933.
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676.

- Foulds, C.E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T.L., Malovannaya, A., Hamilton, R.A., Gates, L.A., Zhang, Z., Li, C., et al. (2013). Proteomic analysis of coregulators bound to ER α on DNA and nucleosomes reveals coregulator dynamics. *Mol. Cell* *51*, 185–199.
- Le François, B., Zhang, L., Mahajan, G.J., Stockmeier, C.A., Friedman, E., and Albert, P.R. (2018). A novel alternative splicing mechanism that enhances human 5-HT1A receptor RNA stability is altered in major depression. *J. Neurosci.* *38*, 8200–8210.
- Freedman, A.H., Clamp, M., and Sackton, T.B. (2021). Error, noise and bias in de novo transcriptome assemblies. *Mol. Ecol. Resour.* *21*, 18–29.
- Friend, K., Lovejoy, A.F., and Steitz, J.A. (2007). U2 snRNP Binds Intronless Histone Pre-mRNAs to Facilitate U7-snRNP-Dependent 3' End Formation. *Mol. Cell* *28*, 240–252.
- Fu, X.D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 1725–1729.
- Fuchs, G., Voichkek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014a). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* *15*, R69.
- Fuchs, G., Hollander, D., Voichkek, Y., Ast, G., and Oren, M. (2014b). Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome Res.* *24*, 1572–1583.
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. *Mol. Cell. Biol.* *24*, 787–795.
- Fujiwara, R., Damodaren, N., Wilusz, J.E., and Murakami, K. (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 22573–22582.
- Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., and Proudfoot, N.J. (2002). Promoter proximal splice sites enhance transcription. *Genes Dev.* *16*, 2792–2799.
- Gang, W.U., Liying, F.A.N., Edmonson, M.N., Shaw, T., Boggs, K., Easton, J., Rusch, M.C., Webb, T.R., Zhang, J., and Potter, P.M. (2018). Inhibition of SF3B1 by molecules targeting the spliceosome results in massive aberrant exon skipping. *RNA* *24*, 1056–1066.
- Gao, Y., Chen, L., Han, Y., Wu, F., Yang, W.S., Zhang, Z., Huo, T., Zhu, Y., Yu, C., Kim, H., et al. (2020). Acetylation of histone H3K27 signals the transcriptional elongation for estrogen receptor alpha. *Commun. Biol.* *3*, 1–10.
- Garber, M.E., Wei, P., KewalRamani, V.N., Mayall, T.P., Herrmann, C.H., Rice, A.P., Littman, D.R., and Jones, K.A. (1998). The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* *12*, 3512–3527.
- Garibaldi, A., Carranza, F., and Hertel, K.J. (2017). Isolation of newly transcribed rna using the metabolic label 4-thiouridine. In *Methods in Molecular Biology*, (Humana Press Inc.), pp. 169–176.
- Gasser, C., Delazer, I., Neuner, E., Pascher, K., Brillet, K., Klotz, S., Trixl, L., Himmelstoß, M., Ennifar, E., Rieder, D., et al. (2020). Thioguanosine conversion enables mRNA-lifetime evaluation by RNA sequencing using double metabolic labeling (TUC-seq DUAL). *Angew. Chemie - Int. Ed.* *59*, 6881–6886.
- Gates, L.A., Foulds, C.E., and O'Malley, B.W. (2017a). Histone marks in the 'driver's seat': functional roles in steering the transcription cycle. *Trends Biochem. Sci.* *42*, 977–989.
- Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J., et al. (2017b). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* *292*, 14456–14472.

- GINNO, P.A., LOTT, P.L., CHRISTENSEN, H.C., KORF, I., and CHÉDIN, F. (2012). R-Loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45, 814–825.
- GIRESI, P.G., KIM, J., MCDANIELL, R.M., IYER, V.R., and LIEB, J.D. (2007). FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885.
- GOMEZ-ESCRIBANO, J.P., and BIBB, M.J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* 4, 207–215.
- GÓMEZ ACUÑA, L.I., FISZBEIN, A., ALLÓ, M., SCHOR, I.E., and KORNBLIHT, A.R. (2013). Connections between chromatin signatures and splicing. *Wiley Interdiscip. Rev. RNA* 4, 77–91.
- GONATPOULOS-POURNATZIS, T., and COWLING, V.H. (2014). Cap-binding complex (CBC). *Biochem. J.* 457, 231–242.
- GONZALES-COPE, M., SIDOLI, S., BHANU, N. V., WON, K.J., and GARCIA, B.A. (2016). Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *BMC Genomics* 17, 95.
- GÖRNMANN, J., KOTOVIC, K.M., HUJER, K., and NEUGEBAUER, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* 19, 53–63.
- GOZANI, O., FELD, R., and REED, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* 10, 233–243.
- GOZANI, O., POTASHKIN, J., and REED, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.* 18, 4752–4760.
- GRESSEL, S., SCHWALB, B., DECKER, T.M., QIN, W., LEONHARDT, H., EICK, D., and CRAMER, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* 6, e29736.
- GRESSEL, S., SCHWALB, B., and CRAMER, P. (2019). The pause-initiation limit restricts transcription activation in human cells. *Nat. Commun.* 10, 3603.
- GROSSWENDT, S., FILIPCHYK, A., MANZANO, M., KLIRONOMOS, F., SCHILLING, M., HERZOG, M., GOTTWEIN, E., and RAJEWSKY, N. (2014). Unambiguous identification of miRNA: target site interactions by different types of ligation reactions. *Mol. Cell* 54, 1042–1054.
- GU, B., EICK, D., and BESAUDE, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.* 41, 1591–1603.
- GU, W., LEE, H.C., CHAVES, D., YOUNGMAN, E.M., PAZOUR, G.J., CONTE, D., and MELLO, C.C. (2012). CapSeq and CIP-TAP identify pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* 151, 1488–1500.
- GUO, R., ZHENG, L., PARK, J.W., LV, R., CHEN, H., JIAO, F., XU, W., MU, S., WEN, H., QIU, J., et al. (2014). BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol. Cell* 56, 298–310.
- GUO, Y.E., MANTEIGA, J.C., HENNINGER, J.E., SABARI, B.R., DALL'AGNESE, A., HANNETT, N.M., SPILLE, J.H., AFAYAN, L.K., ZAMUDIO, A. V., SHRINIVAS, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548.
- HABERLE, V., and STARK, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* 19, 621–637.
- HAFNER, M., LANDTHALER, M., BURGER, L., KHORSHID, M., HAUSSER, J., BERNINGER, P., ROTHBALLER, A., ASCANO, M., JUNGKAMP, A.C., MUNSCHAUER, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- HAHNE, F., and IVANEK, R. (2016). Visualizing genomic data using Gviz and bioconductor. In *Methods in Molecular Biology*, (Humana Press Inc.), pp. 335–351.
- HANNON, G.J. (2010). FASTX-Toolkit.

- Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* *18*, 263–273.
- Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep.* *15*, 2147–2158.
- He, Q., Johnston, J., and Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* *33*, 395–401.
- Hendriks, G.J., Jung, L.A., Larsson, A.J.M., Lidschreiber, M., Andersson Forsman, O., Lidschreiber, K., Cramer, P., and Sandberg, R. (2019). NASC-seq monitors RNA synthesis in single cells. *Nat. Commun.* *10*, 1–9.
- Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* *18*, 637–650.
- Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., Von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* *14*, 1198–1204.
- Le Hir, H., Moore, M.J., and Maquat, L.E. (2000). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* *14*, 1098–1108.
- Hirayoshi, K., and Lis, J.T. (1999). Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases. *Methods Enzymol.* *304*, 351–362.
- Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* *13*, 1234–1239.
- Hölzer, M., and Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience* *8*.
- Horike, S.I., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* *37*, 31–40.
- Howe, K.J., Kane, C.M., and Ares, M. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA* *9*, 993–1006.
- Huang, Y., Li, W., Yao, X., Lin, Q., Jiang, Y., Yin, J., Wen, L., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol. Cell* *45*, 459–469.
- Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* *46*, 205–212.
- Ibrahim, F., Maragkakis, M., Alexiou, P., and Mourelatos, Z. (2018). Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat. Struct. Mol. Biol.* *25*, 302–310.
- Imig, J., Brunschweiler, A., Brümmer, A., Guennewig, B., Mittal, N., Kishore, S., Tsikrika, P., Gerber, A.P., Zavolan, M., and Hall, J. (2015). MiR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction. *Nat. Chem. Biol.* *11*, 107–114.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* (80-.). *324*, 218–223.
- Itzen, F., Greifenberg, A.K., Böskén, C.A., and Geyer, M. (2014). Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Res.* *42*, 7577–7590.
- Jalkanen, A.L., Coleman, S.J., and Wilusz, J. (2014). Determinants and implications of mRNA poly(A) tail size - Does this protein make my tail look big? *Semin. Cell Dev. Biol.* *34*, 24–32.
- Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR

- proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* *153*, 855–868.
- Jiao, Y., and Meyerowitz, E.M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* *6*.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* (80-.). *316*, 1497–1502.
- Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* *2014*.
- Kachaev, Z.M., Lebedeva, L.A., Kozlov, E.N., and Shidlovskii, Y. V. (2020). Interplay of mRNA capping and transcription machineries. *Biosci. Rep.* *40*.
- Kadener, S., Cramer, P., Nogués, G., Cazalla, D., De La Mata, M., Fededa, J.P., Werbajh, S.E., Srebrow, A., and Kornblihtt, A.R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J.* *20*, 5759–5768.
- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* *3*, 576–583.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* *468*, 664–668.
- Kaisers, W. (2019). refGenome: gene and splice site annotation using annotation data from “Ensembl” and “UCSC” genome browsers.
- Kaisers, W., Ptok, J., Schwender, H., and Schaal, H. (2017). Validation of splicing events in transcriptome sequencing data. *Int. J. Mol. Sci.* *18*.
- Kastner, B., Will, C.L., Stark, H., and Lührmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* *11*.
- Keren-Shaul, H., Lev-Maor, G., and Ast, G. (2013). Pre-mRNA splicing is a determinant of nucleosome organization. *PLoS One* *8*.
- Kim, T.H., and Dekker, J. (2018). ChIP–quantitative polymerase chain reaction (ChIP–qPCR). *Cold Spring Harb. Protoc.* *2018*, 354–355.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, R36.
- Kim, J.B., Yamaguchi, Y., Wada, T., Handa, H., and Sharp, P.A. (1999). Tat-SF1 protein associates with RAP30 and human SPT5 proteins. *Mol. Cell. Biol.* *19*, 5960–5968.
- Kim, K., Yang, W., Lee, K.S., Bang, H., Jang, K., Kim, S.C., Yang, J.O., Park, S., Park, K., and Choi, J.K. (2015). Global transcription network incorporating distal regulator binding reveals selective cooperation of cancer drivers and risk genes. *Nucleic Acids Res.* *43*, 5716–5729.
- Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 13564–13569.
- Kim, S.K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.M., Kim, Y.S., Kim, D., et al. (2012). Human histone H3K79 methyltransferase DOT1L methyltransferase binds actively transcribing RNA polymerase II to regulate gene expression. *J. Biol. Chem.* *287*, 39698–39709.
- Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. *Nucleic Acids Res.* *43*, 8258–8267.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* *41*, 376–381.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* *14*, 2452–

2460.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). ICLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* *17*, 909–915.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.* *3*, 570–575.

Krämer, A., Grüter, P., Gröning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J. Cell Biol.* *145*, 1355–1368.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* *67*, 411-422.e4.

Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 10010–10015.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* (80-.). *339*, 950–953.

Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O’Gorman, W., Kimura, H., Proudfoot, N.J., and Akoulitchev, A. (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat. Struct. Biol.* *9*, 800–805.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3’ End Processing and Splicing. *Mol. Cell* *23*, 195–205.

De La Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* *12*, 525–532.

Lahudkar, S., Durairaj, G., Uprety, B., and Bhaumik, S.R. (2014). A novel role for Cet1p mRNA 5’-triphosphatase in promoter proximal accumulation of RNA polymerase II in *Saccharomyces cerevisiae*. *Genetics* *196*, 161–176.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Rättsch, G., and Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 8795–8800.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* *25*, 1841–1842.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* *9*.

Lee, J.H., Kang, B.H., Jang, H., Kim, T.W., Choi, J., Kwak, S., Han, J., Cho, E.J., and Youn, H.D. (2015). AKT phosphorylates H3-threonine 45 to facilitate termination of gene transcription in response to DNA damage. *Nucleic Acids Res.* *43*, 4505–4516.

Lenasi, T., Peterlin, B.M., and Barboric, M. (2011). Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *J. Biol. Chem.* *286*, 22758–22768.

- Li, J., and Liu, C. (2019). Coding or noncoding, the converging concepts of RNAs. *Front. Genet.* *10*, 496.
- Li, G., Cai, L., Chang, H., Hong, P., Zhou, Q., Kulakova, E. V., Kolchanov, N.A., and Ruan, Y. (2014). Chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing technology and application. *BMC Genomics* *15*, S11.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
- Li, J.J., Jiang, C.R., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 19867–19872.
- Li, X., Zhou, B., Chen, L., Gou, L.T., Li, H., and Fu, X.D. (2017). GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* *35*, 940–950.
- Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* *27*, 2380–2396.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923–930.
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* *456*, 464–469.
- Lieber, M.R., and Chedin, F. (2001). R-Loop. In *Encyclopedia of Genetics*, (Elsevier), pp. 1731–1732.
- Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (80-.). *326*, 289–293.
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* *15*, 819–826.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* *23*, 1368–1378.
- Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* *13*, 815–822.
- Ljungman, M., Parks, L., Hulbatte, R., and Bedi, K. (2019). The role of H3K79 methylation in transcription and the DNA damage response. *Mutat. Res. - Rev. Mutat. Res.* *780*, 48–54.
- Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). The Molecules of Life. In *Molecular Cell Biology*. 4th Edition, (New York: W. H. Freeman), p.
- Loerch, S., Leach, J.R., Horner, S.W., Maji, D., Jenkins, J.L., Pulvino, M.J., and Kielkopf, C.L. (2019). The pre-mRNA splicing and transcription factor Tat-SF1 is a functional partner of the spliceosome SF3b1 subunit via a U2AF homology motif interface. *J. Biol. Chem.* *294*, 2892–2902.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lozzio, C.B., and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome. *Blood* *45*, 321–334.
- Lu, Z., Gong, J., and Zhang, Q.C. (2018). PARIS: psoralen analysis of RNA interactions and structures with high throughput and resolution. In *Methods in Molecular Biology*, (Humana Press Inc.), pp. 59–84.

- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* (80-.). 327, 996–1000.
- Lusser, A., Gasser, C., Trixl, L., Piatti, P., Delazer, I., Rieder, D., Bashin, J., Riml, C., Amort, T., and Micura, R. (2020). Thiouridine-to-cytidine conversion sequencing (TUC-Seq) to measure mRNA transcription and degradation Rates. *Methods Mol. Biol.* 2062, 191–211.
- Lykke-Andersen, S., Žumer, K., Molska, E.Š., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell* 81, 514-529.e6.
- Maier, K.C., Gressel, S., Cramer, P., and Schwalb, B. (2020). Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res.* 30, 1332–1344.
- Maita, H., and Nakagawa, S. (2020). What is the switch for coupling transcription and splicing? RNA polymerase II C-terminal domain phosphorylation, phase separation and beyond. *Wiley Interdiscip. Rev. RNA* 11, e1574.
- Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* 270, 12335–12338.
- Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* 271, 27176–27183.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10.
- Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., De Almeida, S.F., and Carmo-Fonseca, M. (2010). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. *Nat. Struct. Mol. Biol.* 18, 1115–1123.
- Maslon, M.M., Braunschweig, U., Aitken, S., Mann, A.R., Kilanowski, F., Hunter, C.J., Blencowe, B.J., Kornblihtt, A.R., Adams, I.R., and Cáceres, J.F. (2019). A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.* 38, e101244.
- Matter, N., and König, H. (2005). Targeted “knockdown” of spliceosome function in mammalian cells. *Nucleic Acids Res.* 33, 1–5.
- Maudlin, I.E., and Beggs, J.D. (2019). Spt5 modulates cotranscriptional spliceosome assembly in *Saccharomyces cerevisiae*. *RNA* 25, 1298–1310.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–360.
- McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. *Cell* 165, 742–753.
- Microsoft Corporation, and Weston, S. (2020). doParallel: foreach parallel adaptor for the “parallel” package.
- Miller, O.L., and Beatty, B.R. (1969). Visualization of nucleolar genes. *Science* (80-.). 164, 955–957.
- Misteli, T., and Spector, D.L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* 3, 697–705.
- Moon, K.J., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* 19, 523–534.
- Morgan, M. (2019). BiocManager: access the bioconductor project package repository.
- Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8253–

8257.

- Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. (1998). A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. *J. Mol. Biol.* 283, 383–394.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922.
- Myers, L.C., Lacomis, L., Erdjument-Bromage, H., and Tempst, P. (2002). The yeast capping enzyme represses RNA polymerase II transcription. *Mol. Cell* 10, 883–894.
- Narita, T., Yung, T.M.C., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y., and Handa, H. (2007). NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol. Cell* 26, 349–365.
- Nechaev, S., Fargo, D.C., Santos, G. Dos, Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* (80-.). 327, 335–338.
- Nelson, N.J. (2001). Microarrays have arrived: Gene expression tool matures. *J. Natl. Cancer Inst.* 93, 492–493.
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543, 72–77.
- Neugebauer, K.M. (2019). Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.* 11.
- Nguyen, H.D., Leong, W.Y., Li, W., Reddy, P.N.G., Sullivan, J.D., Walter, M.J., Zou, L., and Graubert, T.A. (2018). Spliceosome mutations induce R loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes. *Cancer Res.* 78, 5363–5374.
- Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storrval, H., Cao, M., Sandberg, R., Deng, Q., and Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat. Commun.* 7.
- Nicora, G., Vitali, F., Dagliati, A., Geifman, N., and Bellazzi, R. (2020). Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. *Front. Oncol.* 10, 1030.
- Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540.
- Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA polymerase II phosphorylated on CTD serine 5 interacts with the spliceosome during co-transcriptional splicing. *Mol. Cell* 72, 369-379.e4.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514.
- Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I., et al. (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat. Struct. Mol. Biol.* 24, 993–999.
- Osman, S., Mohammad, E., Lidschreiber, M., Stuetzer, A., Bázsó, F.L., Maier, K.C., Urlaub, H., and Cramer, P. (2021). The Cdk8 kinase module regulates interaction of the Mediator complex with RNA polymerase II. *J. Biol. Chem.* 100734.
- Oudelaar, A.M., Beagrie, R.A., Gosden, M., de Ornellas, S., Georgiades, E., Kerry, J., Hidalgo, D., Carrelha, J., Shivalingam, A., El-Sagheer, A.H., et al. (2020). Dynamics of the 4D genome during in

- vivo lineage specification and differentiation. *Nat. Commun.* *11*, 1–12.
- Pabis, M., Neufeld, N., Steiner, M.C., Bojic, T., Shav-Tal, Y., and Neugebauer, K.M. (2013). The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* *19*, 1054–1063.
- Pagès H, Aboyou P, Gentleman R, D.S. (2019). Biostings: efficient manipulation of biological strings.
- Patel, R.K., and Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* *7*.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* *14*, 417–419.
- Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Tsan, Y.C., Chang, C.W., Tarrier, B., Washburn, J.G., Lyons, R., et al. (2013). Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 2240–2245.
- Peng, Z., Yuan, C., Zellmer, L., Liu, S., Xu, N., and Liao, D.J. (2015). Hypothesis: artifacts, including spurious chimeric RNAs with a short homologous sequence, caused by consecutive reverse transcriptions and endogenous random primers. *J. Cancer* *6*, 555–567.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* *33*, 290–295.
- Petasny, M., Bentata, M., Pawellek, A., Baker, M., Kay, G., and Salton, M. (2021). Splicing to keep cycling: the importance of pre-mRNA splicing during the cell cycle. *Trends Genet.* *37*, 266–278.
- Petrova, V., Song, R., Consortium, D., Nordström, K.J.V., Walter, J., Wong, J.J.-L., Armstrong, N.J., Rasko, J.E.J., and Schmitz, U. (2021). Chromatin accessibility determines intron retention in a cell type-specific manner. *BioRxiv* 2021.02.17.431609.
- R Core Team (2018). R: a language and environment for statistical computing.
- Rachez, C., Legendre, R., Costallat, M., Varet, H., Yi, J., Kornobis, E., and Muchardt, C. (2019). HP1 γ binding pre-mRNA at intronic repeats increases splicing fidelity and regulates alternative exon usage. *BioRxiv* 686790.
- Rangamaran, V.R., Uppili, B., Gopal, D., and Ramalingam, K. (2018). EasyQC: tool with interactive user interface for efficient next-generation sequencing data quality control. *J. Comput. Biol.* *25*, 1301–1311.
- Ray, T.A., Cochran, K., Kozlowski, C., Wang, J., Alexander, G., Cady, M.A., Spencer, W.J., Ruzycki, P.A., Clark, B.S., Laeremans, A., et al. (2020). Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nat. Commun.* *11*, 1–20.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* (80-.). *290*, 2306–2309.
- Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., and Cramer, P. (2021). Structure of human Mediator–RNA polymerase II pre-initiation complex. *Nature* 1–8.
- Rhee, H.S., and Pugh, B.F. (2012). ChiP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr. Protoc. Mol. Biol.* *0 21*.
- Riml, C., Amort, T., Rieder, D., Gasser, C., Lusser, A., and Micura, R. (2017). Osmium-mediated transformation of 4-thiouridine to cytidine as key To study RNA dynamics by sequencing. *Angew. Chemie - Int. Ed.* *56*, 13479–13483.
- Roberts, R.J., Carneiro, M.O., and Schatz, M.C. (2013). The advantages of SMRT sequencing. *Genome Biol.* *14*, 405.

- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rodrigues, D.C., Mufteev, M., Weatheritt, R.J., Djuric, U., Ha, K.C.H., Ross, P.J., Wei, W., Piekna, A., Sartori, M.A., Byres, L., et al. (2020). Shifts in ribosome engagement impact key gene sets in neurodevelopment and ubiquitination in Rett syndrome. *Cell Rep.* 30, 4179–4196.e11.
- Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol.* 26, 783–791.
- Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J. Struct. Biol.* 194, 357–367.
- Roy, A.L. (2006). Core Promoters. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, (Springer Berlin Heidelberg), pp. 338–341.
- Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic Acids Res.* 38, 6664–6672.
- RStudio Team (2017). RStudio: integrated development environment for R.
- Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* 18, 337–344.
- Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635.
- Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Söding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157, 1037–1049.
- Schaffner, W. (2006). Enhancer. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, (Springer Berlin Heidelberg), pp. 493–500.
- Schier, A.C., and Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 465–488.
- Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* 65, 25–38.
- Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* 15, 221–225.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132, 887–898.
- Schulze, W.M., and Cusack, S. (2017). Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nat. Commun.* 8, 1–14.
- Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* (80-.). 352, 1225–1228.
- Schwalb, B., Tresch, A., Torkler, P., Dümcke, S., Demel, C., Ripley, B., and Venables, B. (2020). LSD: lots of superior depictions.
- Sebbag-Sznajder, N., Brody, Y., Hochberg-Laufer, H., Shav-Tal, Y., Sperling, J., and Sperling, R. (2020). Dynamic supraspliceosomes are assembled on different transcripts regardless of their intron number and splicing state. *Front. Genet.* 11, 409.
- Shabalina, S.A., Ogurtsov, A.Y., Spiridonov, A.N., Novichkov, P.S., Spiridonov, N.A., and Koonin, E. V. (2010). Distinct patterns of expression and evolution of intronless and intron-containing mammalian

- genes. *Mol. Biol. Evol.* *27*, 1745–1749.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* *498*, 236–240.
- Shandilya, J., and Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *BBA - Gene Regul. Mech.* *1819*, 391–400.
- Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* *49*, 1045–1051.
- Sharma, E., Sterne-Weiler, T., O’Hanlon, D., and Blencowe, B.J. (2016). Global mapping of human RNA-RNA interactions. *Mol. Cell* *62*, 618–626.
- Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl. Acad. Sci. U. S. A.* *111*, E5593–E5601.
- Simon, M.D., Wang, C.I., Kharchenko, P. V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 20497–20502.
- Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* *6*.
- Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* *516*, 436–439.
- So, B.R., Di, C., Cai, Z., Venters, C.C., Guo, J., Oh, J.M., Arai, C., and Dreyfuss, G. (2019). A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* *76*, 590-599.e4.
- Spiluttini, B., Gu, B., Belagal, P., Smirnova, A.S., Nguyen, V.T., Hébert, C., Schmidt, U., Bertrand, E., Darzacq, X., and Bensaude, O. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* *123*, 2085–2093.
- Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* *20*, 631–656.
- Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., Van Cappellen, W.A., et al. (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* *115*, E4368–E4376.
- Stovner, E.B., and Sætrom, P. (2019). PyRanges: efficient comparison of genomic intervals in Python. *Bioinformatics* *36*, 918–919.
- Su, Z., and Huang, D. (2021). Alternative splicing of pre-mRNA in the control of immune activity. *Genes (Basel)*. *12*, 574.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* *508*, 66–71.
- Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D’Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). HiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* *519*, 491–494.
- Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., and Yaspo, M.L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* *15*.
- Talkish, J., Igel, H., Hunter, O., Horner, S.W., Jeffery, N.N., Leach, J.R., Jenkins, J.L., Kielkopf, C.L., and Ares, M. (2019). Cus2 enforces the first ATP-dependent step of splicing by binding to yeast SF3b1 through a UHM–Ulm interaction. *RNA* *25*, 1020–1037.

- Tam, A.S., and Stirling, P.C. (2019). Splicing, genome stability and disease: splice like your genome depends on it! *Curr. Genet.* 65, 905–912.
- Tang, A.D., Soulette, C.M., van Baren, M.J., Hart, K., Hrabeta-Robinson, E., Wu, C.J., and Brooks, A.N. (2020). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat. Commun.* 11, 1–12.
- Tang, Z., Zhao, J., Pearson, Z.J., Boskovic, Z. V., and Wang, J. (2021). RNA-targeting splicing modifiers: drug development and screening assays. *Molecules* 26, 2263.
- Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.S., Jiang, J.Z., Jonathan, M., Huang, K.L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* 33, 1525–1538.
- Tellier, M., Maudlin, I., and Murphy, S. (2020). Transcription and splicing: a two-way street. *Wiley Interdiscip. Rev. RNA* 11, 1–25.
- Teng, T., Tsai, J.H., Puyang, X., Seiler, M., Peng, S., Prajapati, S., Aird, D., Buonamici, S., Caleb, B., Chan, B., et al. (2017). Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. *Nat. Commun.* 8.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A.A., Van Ijcken, W.F.J., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H.J., et al. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* 523, 53–58.
- Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 37, 376–386.
- Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1 γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell* 19, 381–391.
- Varshney, D., Lombardi, O., Schweikert, G., Dunn, S., Suska, O., and Cowling, V.H. (2018). mRNA Cap methyltransferase, RNMT-RAM, promotes RNA Pol II-dependent transcription. *Cell Rep.* 23, 1530–1542.
- Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 24, 896–905.
- Verheijen, M., Lienhard, M., Schrooders, Y., Clayton, O., Nudischer, R., Boerno, S., Timmermann, B., Selevsek, N., Schlapbach, R., Gmuender, H., et al. (2019). DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro. *Sci. Rep.* 9, 1–12.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131, 58–69.
- Vincent, M., Lauriault, P., Dubois, M.F., Lavoie, S., Bensaude, O., and Chabot, B. (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res.* 24, 4649–4652.
- Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018a). Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* 560, 601–606.
- Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018b). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* 560, 607–612.
- Vos, S.M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* 27,

668–677.

Wachutka, L., and Gagneur, J. (2017). Measures of RNA metabolism rates: Toward a definition at the level of single bonds. *Transcription* 8, 75–80.

Wallace, E.W.J., and Beggs, J.D. (2017). Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. *RNA* 23, 601–610.

Wallig, M., Microsoft Corporation, and Weston, S. (2018). `foreach`: provides `foreach` looping construct for R.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*.

Wang, F., and Greene, E.C. (2011). Single-molecule studies of transcription: From one RNA polymerase at a time to the gene expression profile of a cell. *J. Mol. Biol.* 412, 814–831.

Wang, K., Yin, C., Du, X., Chen, S., Wang, J., Zhang, L., Wang, L., Yu, Y., Chi, B., Shi, M., et al. (2019). A U2-snRNP-independent role of SF3b in promoting mRNA export. *Proc. Natl. Acad. Sci. U. S. A.* 116, 7837–7846.

Wang, Y., Yang, Q., and Wang, Z. (2014). The evolution of nanopore sequencing. *Front. Genet.* 5, 449.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35, 169–178.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* 92, 451–462.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286.

Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* 13, 1774–1779.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., et al. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* 508, 263–268.

Wickham, H., François, R., Henry, L., and Müller, K. (2020). `dplyr`: a grammar of data manipulation.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, 1–2.

Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.* 20, 705–723.

Wu, G., Schmid, M., Rib, L., Polak, P., Meola, N., Sandelin, A., and Jensen, T.H. (2020). A two-layered targeting mechanism underlies nuclear RNA sorting by the human exosome. *Cell Rep.* 30, 2387-2401.e5.

Xiao, Y., Yang, Y.H., Burckin, T.A., Shiue, L., Hartzog, G.A., and Segal, M.R. (2005). Analysis of a splice array experiment elucidates roles of chromatin elongation factor Spt4-5 in splicing. *PLoS Comput. Biol.* 1, 0276–0288.

Xie, L., Pelz, C., Wang, W., Bashar, A., Varlamova, O., Shadle, S., and Impey, S. (2011). KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J.* 30, 1473–1484.

Yan, D., Perriman, R., Igel, H., Howe, K.J., Neville, M., and Ares, M. (1998). CUS2, a Yeast Homolog of Human Tat-SF1, Rescues Function of Misfolded U2 through an Unusual RNA Recognition Motif. *Mol. Cell. Biol.* 18, 5000–5009.

Yang, Z., Yik, J.H.N., Chen, R., He, N., Moon, K.J., Ozato, K., and Zhou, Q. (2005). Recruitment of P-

- TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* *19*, 535–545.
- Yasen, A., Aini, A., Wang, H., Li, W., Zhang, C., Ran, B., Tuxun, T., Maimaitinijati, Y., Shao, Y., Aji, T., et al. (2020). Progress and applications of single-cell sequencing techniques. *Infect. Genet. Evol.* *80*, 104198.
- Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M., et al. (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. *FEBS J.* *278*, 4870–4880.
- Yoshimoto, R., Chhipi-Shrestha, J.K., Schneider-Poetsch, T., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Hayashizaki, Y., Mayeda, A., Nakagawa, S., et al. (2021). Spliceostatin A interaction with SF3B limits U1 snRNP availability and causes premature cleavage and polyadenylation. *Cell Chem. Biol.*
- Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* *11*, R14.
- Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., and Reinberg, D. (2009). Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/set2 complex required for H3 Lys-36 trimethylation activity in vivo. *J. Biol. Chem.* *284*, 15701–15707.
- Zacher, B., Lidschreiber, M., Cramer, P., Gagneur, J., and Tresch, A. (2014). Annotation of genomics data using bidirectional hidden Markov models unveils variations in Pol II transcription cycle. *Mol. Syst. Biol.* *10*, 768.
- Zapparoli, E., Briata, P., Rossi, M., Brondolo, L., Bucci, G., and Gherzi, R. (2020). Comprehensive multi-omics analysis uncovers a group of TGF- β -regulated genes among lncRNA EPR direct transcriptional targets. *Nucleic Acids Res.* *48*, 9053–9066.
- Zarnegar, B.J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y., and Khavari, P.A. (2016). IrCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* *13*, 489–492.
- Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015). ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* *6*, 1–13.
- Zhang, S., Aibara, S., Vos, S.M., Agafonov, D.E., Lührmann, R., and Cramer, P. (2021). Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* (80-.). *371*, 305–309.
- Zhang, Z., Will, C.L., Bertram, K., Dybkov, O., Hartmuth, K., Agafonov, D.E., Hofele, R., Urlaub, H., Kastner, B., Lührmann, R., et al. (2020). Molecular architecture of the human 17S U2 snRNP. *Nature* *583*, 310–313.
- Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* *40*, 939–953.
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* *38*, 1341–1347.
- Zhou, H.L., Hinman, M.N., Barron, V.A., Geng, C., Zhou, G., Luo, G., Siegel, R.E., and Lou, H. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* *108*, E627–E635.
- Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* *138*, 1122–1136.

APPENDIX

Listed figures

Main figures

Figure 1. Study design.	3
Figure 2. The transcription cycle.	8
Figure 3. Transcription regulation.	9
Figure 4. The splicing cycle.	12
Figure 5. Co-transcriptional splicing.	13
Figure 6. Major isoform annotation metrics.	27
Figure 7. UTR length impact on Pla-B response of intronless genes.	28
Figure 8. Splicing ratio calculation.	33
Figure 9. Schematic representation of the analysis for the selection of splicing affected and unaffected transcripts.	34
Figure 10. Schematic simplified representation of the pause duration calculation.	38
Figure 11. U2 snRNP inhibition with 1 μ M Pla-B.	42
Figure 12. Rapid splicing inhibition in human cells.	43
Figure 13. Inhibition of the U2 snRNP factor SF3B1 decreases RNA synthesis.	45
Figure 14. Pla-B effect on transcription is length-dependent and is related to splicing inhibition.	46
Figure 15. SSA treatment decreases new RNA synthesis.	48
Figure 16. AMO U2 treatment decreases new RNA synthesis.	49
Figure 17. Inhibition of U2 snRNP function impairs early transcription elongation.	51
Figure 18. mNET-seq data reveals a defect in productive and processive Pol II elongation upon Pla-B treatment.	52
Figure 19. Inhibition of U2 snRNP function increases Pol II pause duration.	54
Figure 20. Inhibition of U2 snRNP function impairs P-TEFb recruitment.	56
Figure 21. Inhibition of U2 snRNP function alters Pol II elongation velocity.	57
Figure 22. Inhibition of U2 snRNP function alters Pol II elongation velocity.	58
Figure 23. U2 snRNP inhibition with Pla-B is independent of stress response.	59
Figure 24. GO enrichment analysis.	60
Figure 25. U2 snRNP impact on Pol II release into productive elongation.	63

Supplementary figures

Figure S1. UTR length impact on intron-containing genes. Related to Figure 7.....	77
Figure S2. TT-seq monitors the decrease of RNA synthesis activity upon Pla-B treatment. Related to Figure 11, Figure 12 and Figure 13.	78
Figure S3. Alternative splicing analysis using CIGAR strings. Related to Figure 12.....	79
Figure S4 .TT-seq upon U2snRNP inhibition with 100nM Pla-B. Related to Results section 2...	79
Figure S5. TT-seq monitors decrease of RNA synthesis activity upon SSA treatment. Related to Figure 15.	80
Figure S6. TT-seq monitors decrease of RNA synthesis activity upon U2 AMO treatment. Related to Figure 16.	80
Figure S7. mNET-seq <i>S. cerevisiae</i> spike-ins normalization. Related to Figure 17.....	81
Figure S8. ChIP-seq input and duplicate removal. Related to Figure 20.....	81
Figure S9. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon Ctr or HS treatment for 30 min (Gressel et al., 2019) (QuickGO). Related to Figure 23B.	82
Figure S10. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 1 μ M Pla-B treatment for 1 h (QuickGO). Related to Figure 24A.....	82
Figure S11. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 1 μ M Pla-B treatment for 4 h (QuickGO). Related to Figure 24B.....	83
Figure S12. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon DMSO or 30 ng/mL SSA treatment for 1 h (QuickGO). Related to Figure 24C.....	84
Figure S13. Graph comparing the top 10 enriched GO terms for upregulated transcripts upon 75 μ M Ctr AMO or 75 μ M U2 AMO treatment for 1 h (QuickGO). Related to Figure 24D.	85

Listed tables

Main tables

Table 1. Overview of the developed techniques for transcriptomics analysis.	20
Table 2. Software resources used in the present work.....	24
Table 3. Overview of the RNA Spike-ins used in this work.....	29
Table 4. Number of exon-based splice junctions included for each treatment	33
Table 5. Significantly splicing affected transcripts.	35
Table 6. Significantly splicing unaffected transcripts.....	35
Table 7. Number of each nucleotide (A, C, G, and U) and calculated molar-mass for each spike-in sequence.....	36
Table 8. Number of up and downregulated transcripts upon DESeq2 analysis on TT-seq summed exons counts.....	39

Supplementary tables

Table S1. Overview of the datasets used in the present study. Related to Introduction section 2.	86
Table S2. Reagents and resources used for the experimental methods for the present work. Related to Supplementary information section 1.	87
Table S3. Primers used for RT-PCR. Related to Figure S2, Figure S4, Figure S5 and Figure S6.	88
Table S4. Names and length of Human, <i>S. cerevisiae</i> and <i>D. melanogaster</i> genomes. Related to Methods section 2.2 and 2.3.....	88
Table S5. Sequencing statistics for the TT-seq data used in this analysis. Related to Results section 1 and 2.....	89
Table S6. Sequencing statistics for the mNET-seq data used in this analysis. Related to Results section 3.	91
Table S7. Sequencing statistics for the ChIP-seq data used in this analysis. Related to Results section 5.	91
Table S8. Reads distribution per genome. Related to Results section 5.	92
Table S9. Overview of upregulated genes between 1 h DMSO or 1 μ M Pla-B treatment. Related to Results section 7 and Figure 23C.	92
Table S10. Overview of upregulated genes between 4 h DMSO or 1 μ M Pla-B treatment. Related to Results section 7 and Figure 23D.	93
Table S11. Overview of upregulated genes between 1 h DMSO or 30 ng/mL SSA treatment. Related to Results section 7 and Figure 23E.	93
Table S12. Overview of upregulated genes between 1 h 75 μ M Ctr AMO or 75 μ M U2 AMO treatment. Related to Results section 7 and Figure 23F.....	94

List of items from the manuscript

Below are listed the figures, tables and sections that are excerpts or adaptations from the manuscript:

Caizzi, L. *, **Monteiro-Martins, S.** * et al. Efficient RNA polymerase II pause release requires U2 snRNP function. (2021). Mol. Cell 81, 1920-1934.e9

doi: <https://doi.org/10.1016/j.molcel.2021.02.016>

(* joint first authorship)

A detailed list of author contributions can be found on **Introduction chapter I (page 4)**.

Main figures

Figure 11A is adapted from Figure 1C

Figure 11B corresponds to Figure S1C-D

Figure 12A-D corresponds to Figure 1E-H

Figure 13A corresponds to Figure 2A

Figure 13B corresponds to Figure S2A

Figure 13C corresponds to Figure S2B

Figure 13D corresponds to Figure S2G

Figure 13E-G corresponds to Figure S2H-J

Figure 13H corresponds to Figure 2B

Figure 13I corresponds to Figure 2C

Figure 14A-B corresponds to Figure S2L-M

Figure 14C-D corresponds to Figure 2D-E

Figure 15A corresponds to Figure S3C

Figure 15C-D corresponds to Figure 3A-B

Figure 15F corresponds to Figure 3F

Figure 15G is adapted from Figure 3C

Figure 16A corresponds to Figure S3E

Figure 16C-D corresponds to Figure 3D-E

Figure 16F corresponds to Figure 3G

Figure 16G is adapted from Figure 3C

Figure 17A corresponds to Figure S4A

Figure 17C corresponds to Figure 4A

Figure 17D corresponds to Figure 4D

Figure 17E corresponds to Figure S4B

Figure 17F corresponds to Figure S4C

Figure 18A-B corresponds to Figure 4B-C

Figure 18C corresponds to Figure S4D

Figure 19A corresponds to Figure S5

Figure 19B-G corresponds to Figure 5A-F

Figure 20A corresponds to Figure S6A

Figure 20C-E corresponds to Figure 6A-C

Figure 20F-G corresponds to Figure S6B-C

Figure 21A-B corresponds to Figure 7A-B

Figure 21C corresponds to Figure S7A

Figure 22A-C corresponds to Figure S7B-D

Figure 22D corresponds to Figure 7C

Supplementary figures

Figure S2A corresponds to Figure 1D

Figure S2B corresponds to Figure S1B

Figure S4A corresponds to Figure S2C

Figure S4B corresponds to Figure S2D

Figure S4D-E corresponds to Figure S2E-F

Figure S5A corresponds to Figure S3B

Figure S6A corresponds to Figure S3D

Tables

Table 2 and **Table S2** are adapted from the “key resources table”

Table S3 corresponds to Table S1

Chapters

Some chapters and sections were adapted from the referred manuscript, namely:

Summary

State of the art (section 3.2)

Methods (sections 1 to 5)

Results (sections 1 to 6)

Discussion (sections 2 to 4)

Supplementary information (section 1 and the Supplementary Figures and Tables listed above)

Abbreviations and acronyms

3C	chromosome conformation capture
4sU	4-thiouridine
A	adenosine
Å	angstroms
ac	acetylation
AGO-CLIP	argonaute-cross-linking and immunoprecipitation
ARTseq	active mRNA translation sequencing (also known as Ribo-seq)
ATAC-Seq	assay for transposase-accessible chromatin with sequencing
ATP	adenosine triphosphate
bp	base pairs
Br-UTP	5-bromouridine 5'-triphosphate
BRD4	bromodomain-containing protein 4
BrU	5'-Bromouridine
BS	branch site
CBC	cap-binding complex
CBP20	nuclear cap-binding protein subunit 2
CBP80	nuclear cap-binding protein subunit 1
CDK	cyclin dependent kinase
cDNA	complementary DNA
Char-seq	chromatin-associated RNA sequencing
CHART-seq	capture hybridization analysis of RNA targets
ChEC-seq	chromatin endogenous cleavage sequencing
ChIA-PET	chromatin interaction analysis with paired-end tag
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation followed by microarray analysis
ChIP-seq	chromatin immunoprecipitation and sequencing
ChIRP-seq	chromatin isolation by RNA purification
ChRO-seq	chromatin run-on and sequencing
CIGAR	compact idiosyncratic gapped alignment report
CLASH-seq	cross-linked, ligated, and selected hybrids sequencing
CLIP	cross-linking and immunoprecipitation
CLIP-seq	cross-linking and immunoprecipitation followed by sequencing (also known as HITS-CLIP)
CPA	cleavage and polyadenylation
CPSF	cleavage and polyadenylation specific factor
CTD	carboxy-terminal domain
Ctr	control
Ctr AMO	antisense morpholino with a control oligo
CycT1	cyclin T1
DE	differential expression
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid

Appendix

Dnase I	deoxyribonuclease I
DNMT3a	DNA methyltransferase 3 alpha
DNMT3b	DNA methyltransferase 3 beta
DOT1L	DOT1 like histone lysine methyltransferase
DPE	downstream promoter element
DRB	5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside
DRIP-seq	DNA:RNA immunoprecipitation followed by sequencing
DSIF	DRB sensitivity inducing factor
EC*	active elongation complex
eCLIP	enhanced cross-linking and immunoprecipitation
eIF4F	eukaryotic initiation factor 4F
eQTL	expression quantitative trait locus
ERCC	External RNA Control Consortium
ERF3	eukaryotic release factor 3
EU	5-ethynyl uridine
FAIRE	formaldehyde-assisted isolation of regulatory elements
FDR	false discovery rate
GC-content	guanine-cytosine content
GO	gene ontology
GRID-seq	global RNA interactions with DNA by deep sequencing
GRO-seq	global run-on sequencing
gtf	gene transfer format
GUI	graphical user interface
GWDG	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
H	histone
h	hour(s)
HCC	histone cleavage complex
hiCLIP	RNA hybrid and individual-nucleotide resolution ultraviolet crosslinking and immunoprecipitation
HITS-CLIP	high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation (also known as CLIP-seq)
hnRNP	heterogeneous nuclear ribonucleoprotein
HO	hydroxyl
HS	heat shock
iCLIP	individual-nucleotide resolution cross-linking and immunoprecipitation
Int	initiator
IP	immunoprecipitation/immunoprecipitated
irCLIP	infrared cross-linking and immunoprecipitation
kDa	kilodalton
LFG	logarithmic fold changes
LIGR-seq	ligation of interacting RNA followed by high-throughput sequencing
log	logarithmic
IrCaptureSeq	long-read capture sequencing
LRS	long-read sequencing
m	minutes

MDa	megadalton
me	methylation
Med23	Mediator complex subunit 23
miR-CLIP	microRNA cross-linking and immunoprecipitation
mL	milliliter
MNase	micrococcal nuclease
mNET-seq	mammalian native elongating transcript sequencing
MRG15	MORF4-related gene on chromosome 15
mRNA	messenger RNA
NASC-seq	new transcriptome alkylation-dependent single-cell RNA sequencing
ncRNA	non-coding RNA
NELF	negative elongation factor
ng	nanogram
NGS	next generation sequencing
ONT	Oxford Nanopore Technologies
P	phosphate
P-TEFb	positive transcription elongation factor b
PABPC	cytoplasmic polyadenylate-binding protein
PABPN1	polyadenylate-binding nuclear protein 1 (also known as PABP2)
PacBio	Pacific Biosciences
padj	adjusted p-value
PAF	Pol II-associated factor 1
pAG-MNase	fusion of protein A and protein G to micrococcal nuclease
PAL-seq	poly(A)-tail length profiling by sequencing
PAP	nuclear poly(A) polymerase
PAR-CLIP	photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation
PARIS	psoralen analysis of RNA interactions and structures
PCPA	premature cleavage and polyadenylation
PCR	polymerase chain reaction
phos	phosphorylation
PIC	pre-initiation complex
PIP-seq	protein interaction profile sequencing
Pla-B	pladienolide B
Pol	RNA polymerase
poly(A)	polyadenylation
pre-EJC	pre-exon junction complex
Pro	proline
PRO-seq	precision run-on sequencing
PWF	probability weighting function
qPCR	real time PCR
RADICL-seq	RNA and DNA interacting complexes ligated and sequenced
RAP30	RNA polymerase II-associating protein 30
RDH	replication-dependent histones

Appendix

Ribo-seq	ribosome profiling (also known as ARTseq)
RIP-seq	RNA immunoprecipitation sequencing
rle	run-length-encoded
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNase	ribonuclease
RNP	ribonucleoprotein
rRNA	ribosomal RNA
SAGA	Spt-Ada-Gcn5 acetyltransferase
SC35	Serine/arginine-rich splicing factor SC35 (also known as SRF2)
SCAF	SR-related CTD associated factor
scATAC-seq	single cell assay for transposase-accessible chromatin using sequencing
scRNA-seq	single cell RNA sequencing
scSLAM-seq	single cell thiol(SH)-linked alkylation for the metabolic sequencing
SEC	super elongation complex
Ser	serine
SLAM-seq	thiol(SH)-linked alkylation for the metabolic sequencing
SMRT	single-molecule real-time
snRNP	small nuclear ribonucleoprotein
SPLASH	sequencing of psoralen cross-linked, ligated, and selected hybrids
SRSF2	SR-splicing factor 2 (also known as SC35)
SS	splice site
SSA	spliceostatin A
STAN	state annotation
STAR	spliced transcripts alignment to a reference
SVM	support-vector machines
TADs	topologically associating domains
TAT-SF1	HIV1 tat-specific factor 1
TBP	TATA binding protein
TF	transcription factor
Thr	threonine
TRAP-Seq	targeted purification of polysomal mRNA
TRIBE	targets of RNA-binding proteins identified by editing
tRNA	transfer RNA
TSS	transcription start site
TT-seq	transient transcriptome sequencing
TUC-seq	thiouridine-to-cytidine conversion sequencing
Tyr	tyrosine
U2 AMO	antisense morpholino oligo targeting the U2 snRNA
U2AF	U2 auxiliary factor
UTR	untranslated region
UV	ultraviolet
ZMYND11	zinc finger, MYND-type containing 11

$\Delta\psi$ delta percent-spliced-in
 μM micromolar