Multi-omics studies of transcription regulatory mechanisms during human cell fate transition

Dissertation

For the award of the degree

"Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

Within the doctoral program

International Max Planck Research School for Molecular Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Kseniia Lysakovskaia

from Moscow, Russian Federation

Göttingen, 2023

Thesis Committee

Prof. Dr. Patrick Cramer
Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Göttingen, Germany

Dr. Ufuk Günesdogan
Göttingen Center for Molecular Biology, Department of Developmental Biology, Göttingen, Germany

Prof. Dr. Gregor Eichele
Max Planck Institute for Multidisciplinary Sciences, Department of Genes and Behavior, Göttingen, Germany

Members of the Examination Board

Referee: Prof. Dr. Patrick Cramer
Max Planck Institute for Multidisciplinary Sciences, 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

Prof. Dr. Bernd Wollnik
University Medical Center Göttingen, Institute of Human Genetics, Göttingen, Germany

Prof. Dr. Gregor Eichele
Max Planck Institute for Multidisciplinary Sciences, Department of Genes and Behavior, Göttingen, Germany

Dr. Ufuk Günesdogan
Göttingen Center for Molecular Biology, Department of Developmental Biology, Göttingen, Germany

Dr. Marieke Oudelaar
Max Planck Institute for Multidisciplinary Sciences, Research Group Genome Organization and Regulation, Göttingen, Germany

Date of the oral examination: February 20, 2023

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Prof. Dr. Patrick Cramer for giving me an opportunity to pursue such an exciting and challenging research project. Prof. Dr. Patrick Cramer constantly provided guidance and constructive feedback regarding my research projects throughout the PhD studies. For me, it was a pleasure to work in his laboratory with the fascinating scientific environment and fruitful discussions.

Moreover, I am grateful to Dr. Michael Lidschreiber for the supervision, numerous discussions and feedback regarding the current collaborative project. Additionally, I would like to thank Michael for reading parts of this dissertation.

I would like to extend my thanks to my colleague Arjun Devadas for a great collaboration on the work presented here and a lot of discussions and speculations within our research project. Also, thank you for being a very nice and supportive office mate and friend.

I am thankful to my rotation student André Fischer for his help with the experiments and providing a possibility to improve my supervision and teaching skills.

Furthermore, I would like to thank my Thesis Advisory Committee members Dr. Ufuk Günesdogan and Prof. Dr. Eichele for their guidance and useful suggestions throughout my PhD. Additionally, I would like to thank Prof. Dr. Matthias Dobbelstein for reviewing my dissertation. I am also grateful to the examination board members Prof. Dr. Patrick Cramer, Prof. Dr. Matthias Dobbelstein, Prof. Dr. Bernd Wollnik, Prof. Dr. Eichele, Dr. Ufuk Günesdogan, Dr. Marieke Oudelaar for agreeing to attend and assess my upcoming dissertation defense.

Also, I owe many thanks to the brilliant coordinators of Molecular Biology IMPRS Graduate Program - Dr. Steffen Burkhardt and Kerstin Grüniger for the administrative work and the continuous great support throughout the time of my studies.

Special thanks to the lab members of the laboratory for being always available to help and share your expertise. Here, I would like to mention Dr. Saskia Gressel who introduced me to the lab techniques and inspired me to continue the scientific path. I am thankful to Dr. Livia Caizzi and Dr. Anna Sawicka for a huge support and advice on the various scientific aspects during my PhD time in the laboratory. Additionally, I would like to thank Dr. Kristina Žummer, Dr. Svetlana Dodonova, Dr. Jinmi Choi and Dr. Björn Schwalb for

help and scientific discussions. I am thankful to Kerstin Maier, Petra Rus, Kirsten Backs, Janine Blümel and Almuth Burgdorf for their technical assistance throughout my PhD. I am also grateful to my lab mates Taras Velychko, Gaurika Garg, Dr. Sara Monteiro-Martins, Dr. Isaac Fianu, Aiturgan Zheenbekova, Dr. Marco Dombrowski, Dr. Sara Osman, Dr. Marc Böhning, Dr. Felix Wagner, Dr. Shen Han and Aleyna Eray for creating a wonderful research atmosphere, speculations on scientific topics and tons of help.

Finally, I would like to express my gratitude to my family and friends for the incredible support and being always there for me during this challenging time. Moreover, I would like to say thank you to my boyfriend who went through this PhD journey with me and constantly believed in my strength during these dark times of the war between our countries.

SUMMARY

RNA polymerase II (Pol II) transcription is an essential process in living organisms that requires accurate regulation and coordination in order to produce protein-coding and many types of non-coding transcripts. One of the highly regulated transcriptional steps is promoter-proximal Pol II pausing that occurs shortly after initiation of transcription. Paused Pol II can either be released into productive elongation of transcription or terminated prematurely. Promoter-proximal Pol II pausing was demonstrated to tune gene expression in the cell fate transition processes ensuring controlled and rapid transcriptional response. However, the exact mechanisms underlying these processes remain to be further elucidated in a quantitative manner.

Here, we investigate promoter-proximal Pol II regulation throughout the cell type switching by employing a reprogrammable human cell transdifferentiation as a model system. In order to assess the kinetics of Pol II transcription, we apply a multi-omics approach combining a measurement of new RNA synthesis by TT-seq and profiling of transcriptionally engaged Pol II by mNET-seq. We additionally introduce the estimation of promoter-proximal Pol II stability at terminal time points of transdifferentiation using a high-resolution ChIP-nexus technique with the time series of initiation inhibition treatment. We show the global changes in the transcription kinetic parameters including productive initiation frequency (pIF) and apparent pause duration (aPD) for the defined gene groups across the time course of transdifferentiation. Strikingly, the pausing-related factors occupancies analyzed by ChIP-seq exhibit the highest correlation with the pIF and partial correlation with the aPD parameters. Our stability measurements from the ChIP-nexus data indicate that Pol II is highly dynamic and relatively short-lived in the pausing region. Importantly, we suggest distinct mechanisms of promoter-proximal Pol II regulation for the upregulated (iMac) and downregulated (pre-B) gene groups based on our Pol II kinetics and stability estimations during the transdifferentiation. The productive transcription of pre-B genes appears to be downregulated through an increase of the prematurely terminating Pol II fraction rather than stable Pol II pausing. Also, the upregulation of the iMac genes is mediated by an increase of the actual transcription initiation rate along with modulating the productive RNA synthesis via prematurely terminating Pol II fraction in the promoter-

proximal region. Taken together, our results illuminate the importance of the Pol II premature termination in transcriptional regulation of the cell stage-specific gene groups throughout the cell fate transition process. We note that our findings can be transferred to the biologically relevant systems for the future medical applications and drug development.

PUBLICATIONS

Parts of this dissertation are being prepared for the publication:

Mechanisms of transcriptional kinetics during human cell fate transition.

Lysakovskaia K*, Devadas A*, Schwalb B, Choi J, Lidschreiber M and Cramer P.

(* co-shared authorship)

Author contributions: **K.L.**, A.D. and M.L. planned the project. **K.L.** designed and performed the experiments. A.D. designed and conducted bioinformatical analysis. **K.L.** and A.D. interpreted the data. B.S. and M.L. advised on bioinformatical analysis. J.C. provided TT-seq data prior to publication. **K.L.** wrote the original draft of the manuscript with input from all authors. **K.L.** and A.D. visualized data. P.C. and M.L. conceived, planned and supervised the project. P.C. acquired funding.

Contribution to the additional publications that are not included in the dissertation:

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.

(* co-shared authorship)

Molecular Cell 2021 May 6. doi: 10.1016/j.molcel.2021.02.016

Author contributions: L.C. and P.C. conceived and planned the project. L.C. and B.S. designed the methodology. S.M.-M. conducted bioinformatical analysis and curated the data. L.C., **K.L.**, J.S., A.S., and Y.C. performed the experiments. L. C., S. M.-M. and P.C. wrote the manuscript with input from all authors. L.C., S. M.-M. and P.C. visualized the data. L.C., B.S., M.L., and P.C. supervised the project. L.C., A.S., and P.C. acquired funding.

Evidence for additive and synergistic action of mammalian enhancers during cell fate determination.

Choi J, **Lysakovskaia K**, Stik G, Demel C, Söding J, Tian T V, Graf T and Cramer P.

eLife 2021 Mar 26. doi: 10.7554/eLife.65381.

Author contributions: J.C. and P.C. conceived and planned the project. J.C. and P.C. designed the methodology. J.C. performed the experiments, except when stated otherwise. J.C. conducted bioinformatical analysis. **K.L.** set up and performed CRISPR validation and gene reporter assays. G.S. carried out CRISPR-based enhancer deletion and provided Hi-C data prior to publication. C.D. carried out TU annotation. J.S. advised on data analysis. T.T.V. and T.G. provided BLaER1 cell line and instructions for transdifferentiation. J.C. and P.C. interpreted the data and wrote the manuscript, with input from T.G. P.C. supervised the project and acquired funding.

TABLE OF CONTENTS

I. INTRODUCTION

1. GENERAL OVERVIEW OF POL II TRANSCRIPTION

Living organisms store the genetic information in the form of deoxyribonucleic acid (DNA) that serves as a template for ribonucleic acid (RNA) synthesis in the process of transcription, a crucial step in establishing cellular identity and functions. Produced RNA molecules represent a pool of protein-coding or messenger RNA (mRNA) molecules and non-coding RNAs (ncRNAs) that perform transcriptional regulatory or functional role in the cell (J. Li & Liu, 2019; Morris & Mattick, 2014). In view of this, recent definition of a gene states that it is a DNA region transcribed into mRNA or ncRNA molecules (Salzberg, 2018). The process of transcription is carried out by the multi-subunit complexes called DNA-dependent RNA polymerases (Pol) (Cramer, 2002). These enzymes were first purified from the eukaryotic organisms in 1969 and named Pol I, II and III (Roeder & Rutter, 1969). Hereinafter, they were shown to transcribe distinct classes of RNA: Pol I transcribes three largest species of ribosomal RNA (rRNA), Pol II produces mRNA and many types of ncRNA, Pol III synthesizes the smallest rRNA species and transfer RNA (tRNA) (Cramer, 2019; Sentenac, 1985). In this section, I will focus on Pol II transcription cycle and its key regulatory mechanisms.

1.1. POL II TRANSCRIPTION CYCLE

Transcription by Pol II is a precisely regulated and coordinated process ensuring where, when and at which level specific gene is expressed. This is controlled by the DNA regulatory sequences that can be found within or outside the gene and represented by promoter elements comprising core promoter, site proximal to the core promoter and distal elements such as enhancers, silencers and insulators (Maston et al., 2006). Core promoter positions the assembly of pre-initiation complex (PIC) containing Pol II and the general transcription factors (GTFs) (Juven-Gershon et al., 2008; Orphanides et al., 1996; Thomas & Chiang, 2006). Proximal region of the promoter and distal elements recruit transcriptional

activators or repressors that can either directly interact with PIC or, in most of the cases, bring additional co-regulators in order to control gene expression (Fuda et al., 2009; Kornberg, 1999; Maston et al., 2006). The Pol II transcription cycle is tightly regulated and consistent of the following steps: promoter opening, PIC assembly on the core promoter, transcription initiation, promoter-proximal Pol II pausing, productive elongation, transcription termination and recycling (**Figure 1**) (Fuda et al., 2009).

The certain position where transcription begins is defined as a transcription start site (TSS) and located at the 5' end of the gene (Haberle & Stark, 2018). Prior to transcription initiation, the promoter is cleared of nucleosomes in order to make it accessible for Pol II and GTFs (Fuda et al., 2009). Nucleosomes are removed or displaced from the promoter by the chromatin-remodeling complexes allowing transcription to be initiated (**Figure 1**, step 1) (Lorch & Kornberg, 2017). Given this, the active promoters are associated with the nucleosome-depleted region (NDR) flanked by strongly positioned and phased nucleosomes downstream and upstream of TSS, that are referred as +1 and -1 nucleosomes correspondingly (Jiang & Pugh, 2009; Schones et al., 2008). Following nucleosome remodeling, PIC assembly occurs at the core promoter including sequential binding of six GTFs that recruit Pol II and prime transcription initiation (**Figure 1**, step 2) (Thomas & Chiang, 2006). Briefly, TFIID recognizes the core promoter elements and stabilizes further binding of TFIIA and TFIIB, then Pol II is recruited by TFIIF resulting in the core PIC formation (Orphanides et al., 1996; Sainsbury et al., 2015). The whole complex is completed after sequential binding of TFIIE and TFIIH (Y. He et al., 2013; Sainsbury et al., 2015). After PIC assembly, double-stranded DNA is melted resulting in the open PIC formation and Pol II initiates synthesis of the first few nucleotides followed by Pol II escape from the promoter and GTFs dissociation (**Figure 1**, step 3) (Aibara et al., 2021; Dienemann et al., 2019; Y. He et al., 2016; Sainsbury et al., 2015). Once synthesized 5' end of RNA emerges from Pol II, it undergoes capping in order to protect RNA from exonuclease degradation (Martinez-Rucobo et al., 2015). Another essential RNA processing step occurring co-transcriptionally is called splicing and includes removal of long non-coding RNA parts called introns followed by ligation of coding exons (Herzel et al., 2017). After promoter escape, Pol II encounters an important regulatory step called promoter-proximal pausing that occurs 30-60 bp downstream of the TSS (**Figure 1**, step 4) (Adelman & Lis, 2012; Core & Adelman, 2019). Paused Pol II is associated with the nascent RNA and capable of restarting

transcription, however, it requires additional cues in order to proceed into productive elongation or terminate prematurely (section 2) (Adelman & Lis, 2012; Core & Adelman, 2019; Kamieniarz-Gdula & Proudfoot, 2019). Release of Pol II into elongation is mediated by phosphorylation of Pol II carboxy-terminal domain and pausing factors resulting in paused complex rearrangement and further binding of the elongation factors (**Figure 1**, step 5 and 6) (Jonkers & Lis, 2015). Most of the protein-coding genes contain polyadenylation signal (PAS) that is a hallmark of transcription termination. PAS at the 3' end of synthesized RNA is recognized by cleavage and polyadenylation (CPA) complex containing cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF) and cleavage factors (CF) I and II. After PAS recognition by the CPSF30 and WDR33 components of CPSF, cleavage of pre-mRNA is performed by the CPSF73 endonuclease. The 3' end of pre-mRNA is immediately polyadenylated, whereas associated with Pol II unprotected 5' end of the nascent RNA is co-transcriptionally degraded by 5'-3' XRN2 exonuclease. Following allosteric conformational changes, Pol II is slowed down and displaced from the DNA by XRN2 resulting in transcription termination (**Figure 1**, step 7) (Cortazar et al., 2019; Eaton & West, 2020). Released Pol II can be recycled and enter the new cycle of transcription (**Figure 1**, step 8) (Fuda et al., 2009).

1.2. Carboxy-terminal domain of Pol II

The largest Pol II subunit RPB1 possess a carboxy-terminal domain (CTD), which consists of multiple repeats of the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ (Eick & Geyer, 2013). Pol II enters transcription initiation with unphosphorylated CTD that further binds Mediator protein complex. After the PIC assembly, Mediator facilitates phosphorylation by TFIIH cyclin-dependent kinase (CDK) 7 of Ser5 and Ser7 Pol II CTD leading to promoter escape (Boeing et al., 2010; Meyer et al., 2010; Rengachari et al., 2021; Wong et al., 2014). Along with transcription cycle, other CDKs phosphorylate CTD, which then serves as a binding platform for various factors involved in regulation of transcriptional steps and RNA processing (Buratowski, 2009; Sansó & Fisher, 2013; Zaborowska et al., 2016).

1.3. Histone modifications

As discussed before, nucleosomes are removed or displaced from the promoter region in order to allow transcription initiation. In general, nucleosomes represent a barrier for transcription process (Petesch & Lis, 2012). Structurally, nucleosome is consistent of histone octamer core containing two of each histone (H2A, H2B, H3 and H4) and 145-147 bp of the DNA wrapped around it (Luger et al., 1997; McGhee & Felsenfeld, 1980). Accessible N-tails of histones can be post-translationally modified including acetylation, methylation, phosphorylation, ubiquitinylation, sumoylation, ADP ribosylation, deamination (Kouzarides, 2007). The additional modifications were also described recently (Kebede et al., 2015). The gene transcriptional state is dependent on the histone modifications that recruit numerous transcriptional repressors or activators creating an additional layer of gene expression regulation (Lawrence et al., 2016).

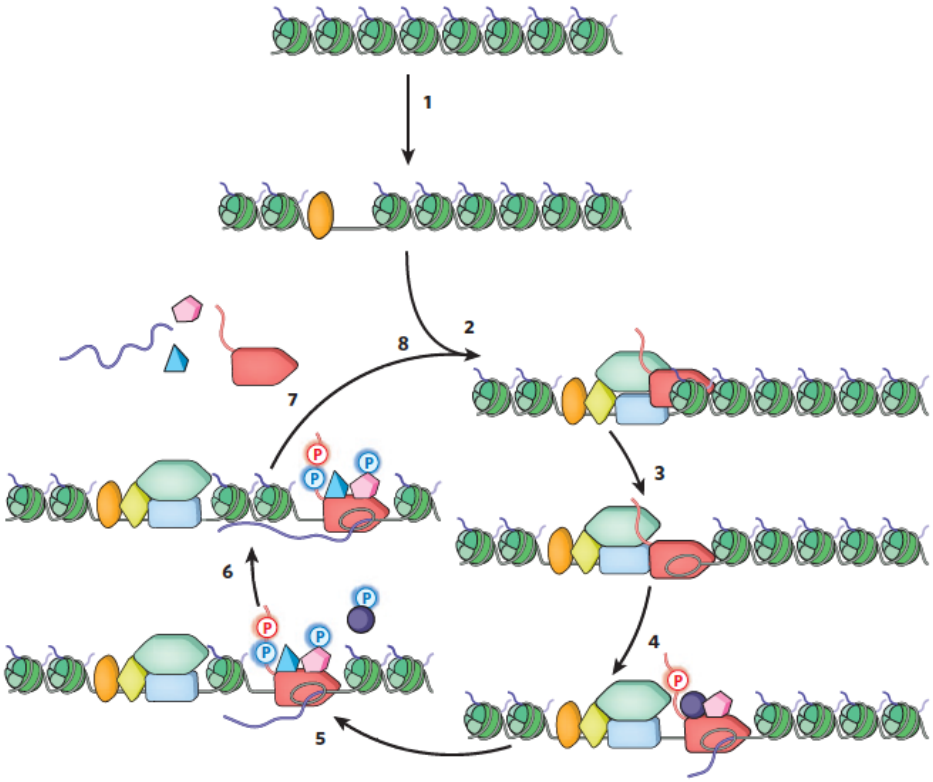


Figure 1. Transcription cycle of Pol II.
Taken from (Fuda et al., 2009). Step 1: promoter opening. Chromatin remodelers are recruited by transcriptional activator (orange) to clear promoter of the nucleosomes. Step 2: PIC assembly. Second activator (yellow) facilitates GTFs (light blue) and co-activators (green) recruitment followed by PIC assembly including Pol II (red) at the promoter. Step 3: initiation. DNA is melted and an open complex is formed. Step 4: promoter clearance and pausing. Pol II escapes the promoter, transcribes

20-50 bp and pauses downstream of TSS. Pol II pausing is stabilized by DSIF (pink) and NELF (purple) factors. Ser5 residue of Pol II CTD is phosphorylated. Step 5: pause release. P-TEFb (blue) phosphorylates NELF, DSIF and Ser2 residues of Pol II CTD. Upon phosphorylation NELF dissociates from the complex, while DSIF is converted into elongation factor and travels with Pol II. Pol II is released into productive elongation or terminates transcription prematurely. Step 6: productive elongation. Elongation factors are recruited to Pol II and aid transcription through the nucleosomes. Step 7: termination. After the gene is transcribed, Pol II is terminated and displaced from the DNA. Nascent RNA is released. Step 8: recycling. Liberated Pol II can enter another round of transcription.

2. PROMOTER-PROXIMAL POL II PAUSING

As discussed above, promoter-proximal pausing is a regulatory step representing an accumulation of Pol II occurring after transcription initiation and promoter escape. In this section, I will start with the history of Pol II pausing and then focus on the molecular mechanisms and functions of this process.

2.1. DISCOVERY OF PROMOTER-PROXIMAL POL II PAUSING

The phenomenon of Pol II promoter-proximal pausing was initially discovered on the heat-shock inducible gene *hsp70* in *Drosophila*. Accumulation of the transcriptionally engaged Pol II associated with the short nascent RNA was observed around the *hsp70* promoter in the unstimulated conditions. Upon the heat shock stimuli, Pol II was released into the gene body indicating a signal-dependent target gene transcription activation (Gilmour & Lis, 1986; Rougvie & Lis, 1988). Later on, halted Pol II was found in the beginning of the other heat shock response and housekeeping genes in *Drosophila* and referred as Pol II pausing representing a rate-limiting step in early elongation of transcription (Rougvie & Lis, 1990). Similarly, paused Pol II around the promoter region was observed in the other organisms. Engaged Pol II was detected in the 5' end moiety of the β-globin locus post the gene activation in mature chicken erythrocytes (Gariglio et al., 1981). The first evidence of Pol II paused downstream of TSS, or promoter-proximally, appeared from the studies performed in mammalian systems including crucial regulatory genes human *c-myc* and *Fos* (Krumm et al., 1992; Plet et al., 1995; Strobl & Eick, 1992). Additional studies of the HIV long terminal repeat (LTR) promoters showed a trans-activator dependent relief of

transcriptional block within the specific sequence followed by the transcript elongation (Kao et al., 1987).

2.2. Promoter-proximal Pol II pausing is a general phenomenon

Along with the development of the new techniques allowing performing genome-wide studies of transcriptional regulation *in vivo*, Pol II was found to accumulate near promoters of many human genes using chromatin immunoprecipitation (ChIP) on chip (ChIP-on-chip) experiments (Guenther et al., 2007; T. H. Kim et al., 2005). Following studies in *Drosophila* demonstrated that promoter-proximally enriched Pol II was transcriptionally engaged across different genes suggesting a key general role of pausing in transcriptional regulation (Muse et al., 2007; Zeitlinger et al., 2007). With the establishment of global nuclear run-on (GRO-seq) and precision nuclear run-on (PRO-seq) sequencing techniques promoter-proximally paused complexes of Pol II were shown to be capable of resuming nascent RNA synthesis genome-wide (Core et al., 2008; Kwak et al., 2013). All above listed studies were mostly focused on the protein-coding genes. Importantly, promoter-proximal pausing of Pol II was also detected for the non-coding transcripts including upstream antisense RNAs (uaRNAs), enhancer RNAs (eRNAs) and other long non-coding RNAs (lncRNAs) (Bunch et al., 2016; Flynn et al., 2011; Gressel et al., 2019; Henriques et al., 2018; Sigova et al., 2013).

2.3. Molecular mechanism of promoter-proximal Pol II pausing

Paused Pol II is stabilized in the promoter-proximal region with the aid of two factors called 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) (Yamaguchi et al., 2013). DSIF is a heterodimer protein comprising SPT5 and SPT4 subunits (Ivanov et al., 2000; Wada et al., 1998; Yamaguchi et al., 2013). SPT5 was shown to interact with nascent RNA capping enzymes and stimulate their function (Pei & Shuman, 2002; Wen & Shatkin, 1999). Indeed, the carboxy-terminal domain of DSIF is positioned near RNA exiting Pol II explaining its functions in capping regulation (Bernecky et al., 2017; Vos, Farnung, Boehning, et al., 2018; Vos, Farnung, Urlaub, et al., 2018). Another pausing factor NELF consists of four subunits

including NELF-A, NELF-B, NELF-C/D (isoforms) and NELF-E. NELF was demonstrated to bind DSIF/Pol II complex and nascent RNA resulting in transcription repression and Pol II pausing (Yamaguchi et al., 2002). Apart from the stabilizing the paused complex, NELF contains several RNA-binding motifs suggesting that nascent RNA recognition can be important for the complex formation (Pagano et al., 2014; Vos et al., 2016; Vos, Farnung, Urlaub, et al., 2018). Additional factor GDOWN1 was found to stabilize Pol II by blocking TFIIF-mediated elongation stimulation and inhibiting premature termination in the promoter-proximal region (Cheng et al., 2012). Nucleosomes were also shown to serve as a barrier for Pol II enhancing promoter-proximal pausing (Jimeno-González et al., 2015). Notably, Pol II pausing can be caused by transcribed sequence or unfavorable conformation of the DNA-RNA hybrid (Nechaev et al., 2010; Vos, Farnung, Boehning, et al., 2018; Vos, Farnung, Urlaub, et al., 2018).

2.4. Fate of the promoter-proximally paused Pol II

2.4.1. Release into productive elongation of transcription

The release of paused Pol II complex into productive elongation is mediated by the positive transcription elongation factor b (P-TEFb) consisting of CDK9 and predominant cyclin T1 subunits (Marshall & Price, 1995; Price, 2000). The CDK9 kinase was shown to phosphorylate many targets including DSIF, NELF and Pol II CTD (J. B. Kim & Sharp, 2001; Marshall et al., 1996; Sansó et al., 2016). P-TEFb phosphorylation of NELF causes its dissociation from the paused complex (Fujinaga et al., 2004; Wu et al., 2003). DSIF is phosphorylated by P-TEFb at its SPT5 subunit and converted into the elongation factor promoting resumption of elongation by Pol II (J. B. Kim & Sharp, 2001; Yamada et al., 2006). Additionally, P-TEFb introduces phosphorylation at Ser2 position of Pol II CTD which is the signature of Pol II elongation state (Zaborowska et al., 2016). Importantly, PTEF-b was shown to perform Ser2 phosphorylation efficiently on the Ser7 pre-phosphorylated Pol II CTD (Czudnochowski et al., 2012). This suggests that P-TEFb specifically phosphorylates Pol II that have undergone transcription initiation and CDK7-mediated promoter escape indicating a highly regulated checkpoint between initiation and pausing (Buratowski, 2009). Following transition into productive elongation, Ser2 phosphorylation of Pol II CTD is taken

over by CDK12 and CDK13 kinases (Bartkowiak et al., 2010; Tellier et al., 2020; Zaborowska et al., 2016). Entering the elongation phase of transcription, Pol II approaches the +1 nucleosome barrier followed by elongation factors recruitment including polymerase-associated factor 1 (PAF1), factor facilitating transcription through the nucleosomes (FACT) and SPT6 (L. Core & Adelman, 2019; Kwak & Lis, 2013; Weber et al., 2014). In parallel to Pol II encountering +1 nucleosome, PAF1 associates with Pol II at the binding site of NELF complex after its P-TEFb-mediated dissociation (Adelman et al., 2006; Van Oss et al., 2016; Vos, Farnung, Boehning, et al., 2018). To aid Pol II transcription through the nucleosomes, FACT is recruited to the elongation complex via Pol II or histone interactions (Belotserkovskaya et al., 2004; Jeronimo et al., 2021). Among all its targets, P-TEFb additionally phosphorylates CTD-linker facilitating SPT6 binding (Vos, Farnung, Boehning, et al., 2018). Apart from the elongation factor functions, SPT6 possesses a histone chaperone activity assisting Pol II to overcome the nucleosome barriers (Ardehali et al., 2009; Bortvin & Winston, 1996). Importantly, productive elongation is achieved by both elongation factors stimulatory functions and efficient nucleosome remodeling in order to facilitate Pol II passage through the chromatin (Kwak & Lis, 2013).

Interestingly, recent *in vivo* studies shed light onto the new functions of the pausing and elongation factors using the novel rapid protein depletion techniques. Rapid depletion of NELF resulted in Pol II accumulation at the +1 nucleosome and global defects of the proper promoter-proximal pause release upon heat-shock stimulus indicating that NELF is a pause stabilizing factor and not necessary repressing transcription elongation (Aoi et al., 2020). Additionally, DSIF subunit SPT5 rapid depletion led to Pol II promoter-proximal stability loss followed by the Pol II core subunit degradation suggesting a critical function of SPT5 in the gene expression control (Aoi et al., 2021).

2.4.2. P-TEFb RECRUITMENT MECHANISMS

Normally P-TEFb is found in either active or inactive state as a part of different complexes in the cell (Lin et al., 2010; Moon et al., 2005; Nguyen et al., 2001; Z. Yang et al., 2001, 2005). P-TEFb can be associated with the 7SK small nuclear RNA (snRNA) forming a large complex with repressed kinase activity (Nguyen et al., 2001; Z. Yang et al., 2001). 7SK snRNA serves as a scaffold for other proteins that inhibit P-TEFb activity: HEXIM1 or

HEXIM2, LARP7, MePCE (Q. Li et al., 2005; Quaresma et al., 2016). The majority of P-TEFb in human cells is present in inactive form (Zhou et al., 2012). The release of P-TEFb from the inactive complex is mediated by several transcription factors including HIV Tat and BRD4 (Barboric et al., 2007; Schröder et al., 2012). Another factor responsible for P-TEFb activation and its recruitment to the paused Pol II is super elongation complex (SEC) (N. He et al., 2010; Lin et al., 2010; Luo et al., 2012). The subunits of SEC usually co-exist with translocation partners of mixed lineage leukemia (MLL) proteins that are suggested to recruit SEC to their target genes in cancer conditions (Lin et al., 2010; Smith et al., 2011). Additionally, Mediator complex subunit MED26 can recruit SEC by interaction with its components facilitating Pol II pause release (Lens et al., 2017; Takahashi et al., 2011). Previous studies revealed a variety of activator complexes that bind SEC and guide it to the target genes (F. X. Chen et al., 2018). Notably, recent work demonstrated that BRD4-mediated P-TEFb recruitment predominate for most of the genes, whereas SEC is responsible for the pause release at the genes activated in the heat shock conditions (B. Zheng et al., 2021).

2.4.3. Premature termination of the paused Pol II complex

Paused Pol II in the promoter-proximal region can undergo an alternative fate – premature termination (Brannan et al., 2012; Hay et al., 1982; Kamieniarz-Gdula & Proudfoot, 2019; Wagschal et al., 2012). Over the past years, several studies provided evidence of premature termination in promoter-proximal window. Among them, termination and 3’ end RNA processing factors occupancy was detected near 5’ ends of the genes (Brannan et al., 2012; Wagschal et al., 2012). Additionally, detection of the short capped RNAs and determination of RNA cleavage sites in promoter-proximal region supported this hypothesis (Almada et al., 2013; Nechaev et al., 2010). More resent findings suggested a rapid turnover of promoter-proximal Pol II independent on the transcriptional activity level using single-molecule footprinting and fluorescence recovery after photobleaching (FRAP) imaging in live cells (Krebs et al., 2017; Steurer et al., 2018). Short residence time of paused Pol II (up to 10 min) was further reported using genome-wide techniques combined with initiation inhibition or hydrogen peroxide treatment in human cells (Erickson et al., 2018; Nilson et al., 2017). Interestingly, previous work demonstrated a

small fraction of the genes with stably paused Pol II similarly employing initiation inhibition and Pol II ChIP-seq in human cells (F. Chen et al., 2015). GRO-seq experiments detecting the nascent RNA additionally revealed 6.9 min Pol II half-live in mouse embryonic stem cells (mESCs) under initiation inhibition time series (Jonkers et al., 2014). The broader range of the paused Pol II stability time was observed in *Drosophila* using detection of the short TSS-associated RNAs (tssRNAs) and Pol II ChIP experiments. While average of the Pol II half-life was estimated between 5-10 min, shorter and longer values were reported (Henriques et al., 2013). Further work corrected these estimations showing average of the half-life within 10-20 min in *Drosophila* employing transcription initiation inhibition followed by high-resolution method ChIP-nexus allowing distinguishing both PIC-associated and promoter-proximal Pol II (Shao & Zeitlinger, 2017). Interestingly, recently published study reported approximately 6 min average half-life time and a high premature termination rate (~80%) of promoter-proximal Pol II in *Drosophila* using a novel Start-TimeLapse-seq (STL-seq) capturing a short capped RNAs (scRNAs) near TSS (Zimmer et al., 2021).

Previously published findings strongly suggested an involvement of the Integrator complex in premature termination of transcription (Beckedorff et al., 2020; Elrod et al., 2019; Tatomer et al., 2019). Integrator binds paused Pol II complex associated with DSIF and NELF and docks its endonuclease module to the RNA exit site. After endonuclease cleavage of the nascent RNA, Pol II can be released from the DNA and terminate prematurely (Fianu et al., 2021). Additionally, Integrator associates with the protein phosphatase 2A (PP2A), which dephosphorylates Pol II CTD and SPT5 preventing transition of the paused complex into productive elongation (K. L. Huang et al., 2020; H. Zheng et al., 2020). Initially, Integrator was shown to be involved in snRNA 3' end processing (Baillat et al., 2005). Further studies reported that Integrator functions at other classes of ncRNAs including uaRNAs, eRNAs and other lncRNAs (Barra et al., 2020; Lai et al., 2015; Liu et al., 2022; Lykke-Andersen et al., 2021; Nojima et al., 2018). Specific subsets of protein-coding genes were also found to be regulated by Integrator complex (Beckedorff et al., 2020; Dasilva et al., 2021; Elrod et al., 2019; K. L. Huang et al., 2020; Tatomer et al., 2019). Interestingly, a number of studies proposed that Integrator attenuates gene transcription by premature termination in the promoter-proximal region (Elrod et al., 2019; Fianu et al., 2021; K. L. Huang et al., 2020; Lykke-Andersen et al., 2021). In contrast, Integrator was suggested to remove paused non-productive Pol II complexes promoting recruitment of elongation-

competent Pol II proceeding into productive transcription elongation (Beckedorff et al., 2020). Notably, recent work demonstrated Integrator complex as a global regulator of Pol II promoter-proximal termination independent on the RNA biotype utilizing a rapid depletion of the cleavage subunit INTS11 in combination with the nascent RNA sequencing techniques (Stein et al., 2022).

In parallel, transcription termination factor 5'-3' XRN2 exonuclease (section 1) was additionally found to localize at 5' ends of the genes and interact with decapping factors suggesting its role in premature termination (Brannan et al., 2012). Interestingly, 3' end RNA processing and termination factor PCF11 and factors of CPA machinery were detected in the proximity of TSS indicating alternative mechanisms of premature termination (Kamieniarz-Gdula et al., 2019; G. Martin et al., 2012; Nojima et al., 2015). Confirming these findings, depletion experiments of termination factors (PCF11 and XRN2) and CPA machinery components revealed an increased level of nascent RNA in the promoter-proximal region (Kamieniarz-Gdula et al., 2019; Nojima et al., 2015). Recent work similarly reported an increase of nascent transcription signal in the proximity of promoter in human cells containing exonuclease-dead mutant of XRN2 (Cortazar et al., 2022).

2.4.4. Backtracking and arrest of paused Pol II complex

Apart from the factor-regulated Pol II promoter-proximal pausing, Pol II can move backwards in the process called backtracking caused by the certain DNA sequences or transcriptional barriers such as nucleosomes (Bondarenko et al., 2006; Izban & Luse, 1992; Reines, 1992). Backtracking results in a displacement of 3' RNA end from the Pol II active site and impeding the following nucleotide addition leading to transcriptional arrest. Reactivation of Pol II elongation is achieved by the transcription factor IIS (TFIIS) that recognizes arrested complex and induces a cleavage of backtracked RNA creating a new 3' end in the active site of Pol II (Cheung & Cramer, 2011; Kettenberger et al., 2003). Importantly, TFIIS appeared to have an important role in transcription elongation resumption upon the heat shock stimuli in *Drosophila* (Adelman et al., 2005). Additionally, depletion of TFIIS in *Drosophila* cells led to an increase in the start site-associated RNA length at the genes with stalled Pol II complex indicating a presence of backtracked RNA requiring TFIIS-mediated reactivation of Pol II transcription (Nechaev et al., 2010). More

resent study demonstrated an importance of TFIIS-mediated rescue from the Pol II backtracked state for the efficient promoter-proximal pause release in the normal conditions and during the stress response in human cells (Sheridan et al., 2019).

2.5. Current view on promoter-proximal Pol II pausing functions

Promoter-proximal Pol II pausing is a widespread regulatory process occurring at many genes with different levels of transcriptional activity. Notably, majority of promoter-proximal Pol II pausing signal was detected on the transcriptionally active genes meaning that pausing is the way to tune gene expression rather than a switch between active/inactive state. Promoter-proximal pausing can additionally maintain a basal level of transcriptional activity in order to "boost" expression of the specific gene(s) in response to the physiological or environmental stimuli (Core & Adelman, 2019; Gilchrist et al., 2012).

Promoter-proximal Pol II pausing can serve as a quality control step in order to ensure that the nascent transcript is capped (Tome et al., 2018). Also, promoter-proximally paused Pol II complex can be involved in the recruitment of 3' end RNA processing factors (Glover-Cutter et al., 2008). In many cases, promoter-proximal pausing was shown to regulate various cellular signaling pathways including environmental stress, immune response signals, nuclear receptors and steroid hormone stimuli, developmental and differentiation signals (Liu et al., 2015). One of the well-studied extracellular stress signals is the heat shock (HS) response. First evidence of the promoter-proximal Pol II pausing appeared from the studies on the *Drosophila* heat shock response genes (section 2.1). Further studies identified that promoter activation and pause release during the heat shock response in *Drosophila* can be regulated by coordination of two transcriptional factors: GAGA-associated factor (GAF) and heat shock factor (HSF) (Duarte et al., 2016; Lis et al., 2000). Similarly, hypoxia-inducible factor 1α (HIF1α) was shown to mediate P-TEFb recruitment to the target genes with pre-occupied promoter-proximal Pol II leading to their activation in response to hypoxia (Galbraith et al., 2013). Early studies of proinflammatory signaling revealed that a key immune response regulator nuclear factor κB (NF-κB) can recruit P-TEFb and stimulate paused Pol II release at the interleukin 8 (IL-8) gene in response to the tumor necrosis factor α (TNFα) treatment (Barboric et al., 2001). Other

proinflammatory genes additionally appeared to accumulate promoter-proximally paused Pol II and maintain a minimal transcriptional activity in macrophage cells (Adelman et al., 2009; Escoubet-Lozach et al., 2011). Following bacterial lipopolysaccharide (LPS) treatment triggered an activation of the immune response signaling genes via P-TEFb recruitment to the paused Pol II (Adelman et al., 2009). Besides that, several studies suggested promoter-proximal Pol II regulation during nuclear receptor and hormone signaling including glucocorticoid, retinoid, estrogen and androgen receptor signaling pathways (Flajollet et al., 2013; Gupte et al., 2013; Hah et al., 2011; D. K. Lee et al., 2001; Mitra et al., 2012). Additionally, promoter-proximal Pol II pausing regulation during the cell type switching was extensively investigated over the past years. This question will be discussed in details in the next section. Thus, promoter-proximal pausing of Pol II represents a highly regulated and a rate-limiting transcriptional step indispensable for cell survival.

3. Promoter-proximal Pol II pausing in the context of cell fate transition

Although initial studies of the promoter-proximal Pol II pausing regulation mostly focused on the extracellular stimuli response, the evidence of Pol II pausing involvement in the cell fate transition processes such as development and cell differentiation appeared and posed a big question in the field. Establishment of the genome-wide techniques illuminated potential mechanisms on how the cell type specific genes are switched “on” and “off” during these processes (Gaertner & Zeitlinger, 2014; Liu et al., 2015). In this chapter, I will describe the findings highlighting promoter-proximal Pol II regulation during the cell fate determination.

3.1. Developmental regulation

First Pol II ChIP-on-chip experiments were performed in *Drosophila* embryo cells and could identify promoter-proximal Pol II enrichment prevailing at the developmental control and signal response genes (Muse et al., 2007; Zeitlinger et al., 2007). Transcriptional status of the gene (active, stalled, no Pol II) was defined based on the stalling index representing a ratio of Pol II signal around TSS versus Pol II signal in the internal transcript

region. Notably, this analysis detected stalled Pol II on the developmental genes that are poised and prepared for activation in the further developmental stage (Zeitlinger et al., 2007). These findings were further supported with Pol II ChIP-seq and mRNA-seq experiments using muscle development in *Drosophila* as a model. Poised Pol II was considered to have similarly calculated high stalling index and minimal transcription level from the mRNA-seq signal. In general, poised Pol II was recruited to the promoters prior to the gene activation or marking the gene promoters post activation dependent on the certain developmental stage (Gaertner et al., 2012). The following studies suggested that predominant regulation occurs at the promoter-proximal Pol II pausing step for the target genes involved in *Drosophila* axis patterning using GRO-seq experiments (Saunders et al., 2013). Notably, stalled or paused Pol II was observed at the genes activated in the synchronous manner rather than at the stochastically activated genes during *Drosophila* embryogenesis (Boettiger & Levine, 2009). These findings suggested that promoter-proximal Pol II pausing is a way to rapidly coordinate the gene expression in order to ensure a proper organism development (Boettiger & Levine, 2009; Lagha et al., 2013).

3.2. Cell differentiation and reprogramming

A number of studies revealed that precise transcriptional regulation during hematopoietic cell differentiation is achieved via hematopoietic transcription factors control of promoter-proximal Pol II pausing and pause release (Bai et al., 2010; Bottardi et al., 2011, 2013; Elagib et al., 2008, 2013; Meier et al., 2006; Song et al., 2010). In this context, the erythroid cell fate-specific factor TIF1α was shown to recruit P-TEFb and FACT to the paused Pol II at the target genes involved in erythropoiesis promoting their activation (Bai et al., 2010). Additionally, megakaryocytic differentiation induction caused transcription factor GATA1-mediated dissociation of the inhibitory protein HEXIM1 from P-TEFb resulting in the Pol II pause release followed by elongation at the target genes (Elagib et al., 2008). Furthermore, GATA1 mutation resulted in the disruption of calpain 2 protease-dependent P-TEFb activation through its release from the 7SK inhibitory complex leading to megakaryopoiesis dysregulation (Elagib et al., 2013). Another example of transcriptional regulator is Ikaros factor, which was shown to regulate promoter-proximal Pol II release during hematopoietic differentiation (Bottardi et al., 2011, 2013). Notably, recent work

suggested that target gene regulation involved in mouse erythropoiesis occurs primarily via initiation rather than Pol II pause release using scRNA-seq (Larke et al., 2021).

One of the first genome-wide studies of Pol II pausing was performed in human embryonic stem cells (hESCs) and revealed that both transcriptionally active and inactive genes exhibit marks of initiation including chromatin signatures and promoter-proximal Pol II occupancy. These pre-occupied promoters of the transcriptionally inactive genes were proposed to facilitate future differentiation process (Guenther et al., 2007). Further work suggested a role of the c-Myc in P-TEFb recruitment followed by Pol II pause release occurring at ~30% of the all actively transcribed genes in mESCs (Rahl et al., 2010). Later on, P-TEFb recruiting complex SEC was demonstrated to activate some of developmental genes in a uniform manner under retinoic acid (RA)-induced differentiation of mESCs (Lin et al., 2011). In contrast, GRO-seq experiments in mESCs detected promoter-proximally paused Pol II only in transcriptionally active genes enriched in metabolism, cell cycle and signaling pathways but not in lineage-specific or developmental control terms (Jonkers et al., 2014; Min et al., 2011; Williams et al., 2015). Additionally, NELF depletion resulted in severe cell proliferation, growth and metabolism defects as well as resistance to the differentiation stimuli indicating Pol II pausing control of the genes involved in the basic cellular functions (Williams et al., 2015). Interestingly, recent study reported that promoter-proximally paused Pol II is enriched at the pluripotency genes in mammalian somatic cells. Reprogramming factor KLF4 appeared to aid BRD4-dependent P-TEFb recruitment to the pluripotency gene promoters upon induction of somatic cells reprogramming into induced pluripotent stem cells (iPSCs) (L. Liu et al., 2014).

All the described above findings pose a big question whether promoter-proximal Pol II pausing plays a role in the regulation during the cell type switching in mammalian cells. This has been broadly studied over the past years; however, a quantitative estimation of the early transcriptional mechanisms during this process remains to be investigated.

4. AIMS OF THE STUDY

In the current study, we address the question of promoter-proximal Pol II pausing regulation in the context of cell fate transition and how this is achieved in a quantitative

way. As a model system, we employ a human cell transdifferentiation including a highly efficient transition from precursor leukemia B-cells into functional macrophage-like cells by estrogen-inducible C/EBPα overexpression (Choi et al., 2021; Rapino et al., 2013; Stik et al., 2020). The cells are collected at the different time points after transdifferentiation induction for the further analysis (**Figure 2A**). Previously, promoter-proximal pausing of Pol II was estimated using only the Pol II occupancy profiling techniques by calculating the relative ratio of the Pol II signal in the promoter-proximal region to the gene body that was termed as pausing index or traveling ratio (Reppas et al., 2006; Zeitlinger et al., 2007). Prior studies performed calculations of this ratio utilizing DNA sequencing or microarray methods including Pol II ChIP-on-chip (Muse et al., 2007; Zeitlinger et al., 2007) and ChIP-seq (F. Chen et al., 2015; Day et al., 2016; Gaertner et al., 2012; Johnson et al., 2007; L. Liu et al., 2014; Rahl et al., 2010) or RNA sequencing methods such as GRO-seq (L. Core et al., 2008; Jonkers et al., 2014; Min et al., 2011; Williams et al., 2015), PRO-seq (Kwak et al., 2013; Schaaf et al., 2013), NET-seq (Churchman & Weissman, 2011; Mayer et al., 2015) and mNET-seq (Nojima et al., 2015, 2016). However, these data alone do not define transcriptional kinetics since high Pol II occupancy can be interpreted as a signal from few slow or many fast polymerases (Ehrensberger et al., 2013). Therefore, to assess the Pol II promoter-proximal pausing changes, we utilize transient transcriptome sequencing (TT-seq) that measures newly synthesized RNA and mammalian nascent elongating transcript sequencing (mNET-seq) that provides an occupancy of transcriptionally engaged Pol II (**Figure 2B**) (Nojima et al., 2015, 2016; Schwalb et al., 2016). Combination of these two genome-wide methods allows estimating the kinetic parameters (productive initiation frequency and apparent pause duration) and was reported before as multi-omics approach (Gressel et al., 2017, 2019). We additionally introduce Pol II stability measurements in the promoter-proximal region in order to account for the premature termination in our system. Promoter-proximal Pol II stability is estimated according to the previously published method including transcription initiation inhibition time series and monitoring of the paused Pol II signal using high resolution ChIP-nexus (Q. He et al., 2015; Shao & Zeitlinger, 2017).

We thus focus at the question of how the promoter-proximal Pol II kinetics changes throughout the time course of transdifferentiation and what are the functional implications of the target genes involved in these changes. We additionally ask whether the pausing-related factor occupancies are correlated with the calculated kinetic parameters and to

what extent. Moreover, we aim to elucidate how promoter-proximal Pol II stability is reflected with potentially observed pausing kinetics changes during transdifferentiation. Finally, we aim to propose what are the potential mechanisms of the cell-type specific transcription changes across the transdifferentiation time course.

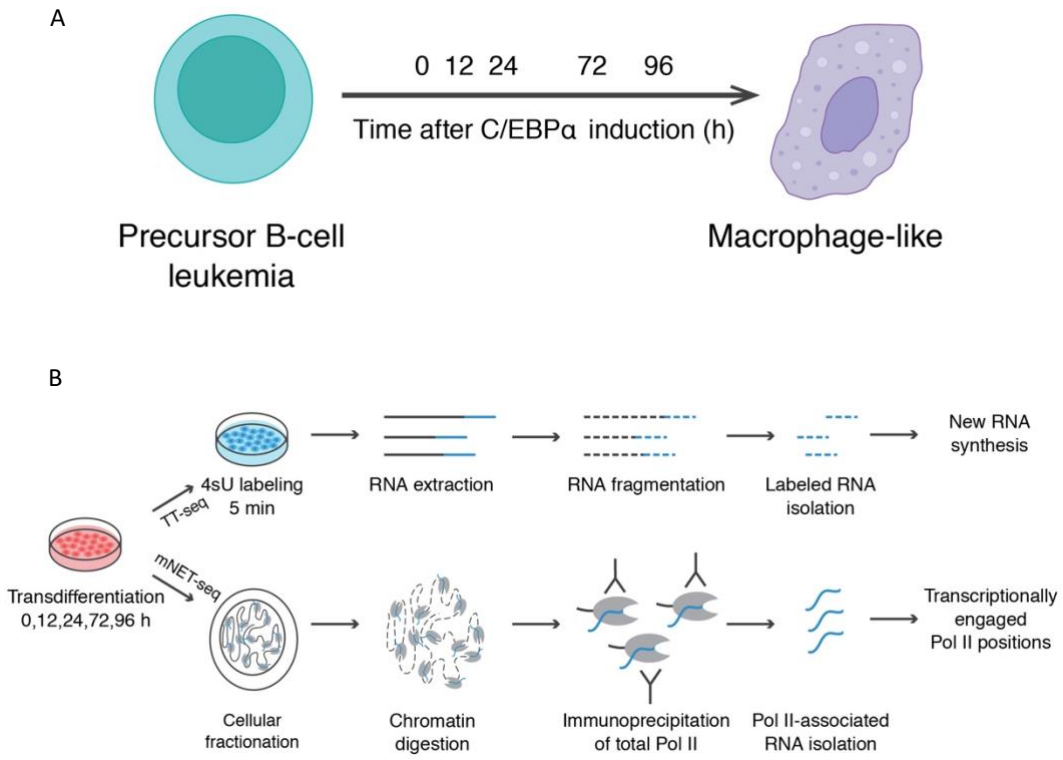


Figure 2. Experimental design of the study.

A. Overview of the transdifferentiation in human cells. Precursor leukemia B-cells undergo transdifferentiation into macrophage-like cells upon estrogen-inducible C/EBPα overexpression (Rapino et al., 2013). Time points of the cells collection after induction are displayed. **B**. Schematic representation of the multi-omics approach (Gressel et al., 2017, 2019). TT-seq provides information about newly produced RNA (Schwalb et al., 2016). TT-seq data was previously collected across the transdifferentiation time course (Choi et al., 2021). mNET-seq allows for transcriptionally engaged Pol II positions mapping (Nojima et al., 2015, 2016).

II. MATERIALS AND METHODS

This section is adapted from Lysakovskaia, Devadas* et al., in preparation*

(co-shared authorship)*

*Author contributions: **K.L.**, A.D. and M.L. planned the project. **K.L.** designed and performed the experiments. A.D. designed and conducted bioinformatical analysis. **K.L.** and A.D. interpreted the data. B.S. and M.L. advised on bioinformatical analysis. J.C. provided TT-seq data prior to publication. **K.L.** wrote the original draft of the manuscript with input from all authors. **K.L.** and A.D. visualized data. P.C. and M.L. conceived, planned and supervised the project. P.C. acquired funding.*

1. CELL LINE AND TRANSDIFFERENTIATION

We utilized previously obtained engineered human BLaER1 cell line stably expressing C/EBPα fused to the estrogen receptor hormone binding domain and GFP (Choi et al., 2021; Rapino et al., 2013). BLaER1 cell line is originated from precursor leukemia B-cells that can be efficiently transdifferentiated into functional macrophage-like cells upon estrogen induction (Choi et al., 2021; Rapino et al., 2013). Cells were cultured in the growth media consisted of RPMI 1640 (Thermo Fisher Scientific, 31870-074) supplemented with 10% FBS (Thermo Fisher Scientific, 10500-064), 4 mM GlutaMAX (Thermo Fisher Scientific, 35050087), 25 mM HEPES (Thermo Fisher Scientific, 15630080) and 100 U/mL Penicillin-Streptomycin (Thermo Fisher Scientific, 15140122) at 37°C and 5% CO_2. BLaER1 cells were regularly examined and tested negative for the mycoplasma contamination using Plasmo Test Mycoplasma Detection Kit (InvivoGen, rep-pt1).

2. TREATMENTS

To induce the transdifferentiation, BLaER1 cells were brought to a density of $0.4*10^6$ cells/mL and mixed with 100 nM β-Estradiol (Sigma-Aldrich, E2758-250MG), 10 ng/mL recombinant human IL-3 (PeproTech, 200-03) and 10 ng/mL recombinant human M-CSF (PeproTech, 300-25) in the growth media. For the 0 h control, BLaER1 cells were treated with the same concentration of solvents (ethanol and water). Cells were harvested at

different time points: 0, 12, 24, 72, 96 h after induction for RNA extraction; 0, 12, 24, 72, 96 h after induction for mNET-seq; 0, 24, 96 h after induction for ChIP-seq; 0, 96 h after induction for ChIP-nexus. For the transcription initiation inhibition, BLaER1 cells were collected at 0 or 96 h of transdifferentiation, brought to a density of $1*10^6$ cells/mL and treated with 5 μM triptolide (TRP) (Sigma-Aldrich, T3652) or DMSO (Sigma-Aldrich, D2438) as a solvent control.

3. TOTAL RNA EXTRACTION AND RT-QPCR

To verify the efficient cell type transition from B-like into macrophage-like cells, BLaER1 cells were transdifferentiated and harvested at 0, 12, 24, 72, 96 h after the induction. To test the initiation inhibition in macrophage-like cells, BLaER1 cells were transdifferentiated up to 96 h, treated with TRP (Sigma-Aldrich, T3652) or DMSO (Sigma-Aldrich, D2438) as a solvent control as described above and harvested. The cell pellets were resuspended in QIAzol Lysis Reagent (Qiagen, 79306) and incubated at RT for 5 min. For the TRP treatment experiments, synthetic RNA spike-ins were added to each sample in the amount of 8-10 ng per $5*10^6$ cells. RNA spike-ins were produced as described previously (Schwalb et al., 2016). Total RNA was extracted according to the manufacturer's instructions (Qiagen). To eliminate genomic DNA contamination, total RNA was treated with the TURBO DNA-free Kit (Thermo Fisher Scientific, AM1907) according to the manufacturer's instructions. cDNA synthesis was performed using Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific, EP0753) according to the manufacturer's instructions. Quantitative polymerase chain reaction (qPCR) was conducted with SYBR Select Master Mix (Thermo Fisher Scientific, 4472919) according to the manufacturer's instructions. Primer sequences used for qPCR are listed in the **Table S1** and **S2** (Supplementary materials).

For TRP treatment, primers were designed for intronic regions in order to capture nascent RNA response. To test the initiation inhibition efficiency with TRP, intronic primers were designed for the beginning and end of the long genes (>15 kbp) expecting the beginning of the genes being more affected with the short treatment times. The final TRP concentration for the downstream experiments was selected based on these results.

4. mNET-seq

mNET-seq protocol was performed as previously described with minor modifications (Nojima et al., 2015, 2016). In short, two independent biological replicates of BLaER1 cells were subjected to transdifferentiation and collected at 0, 12, 24, 72, 96 h post induction. Cells were further used in the amount of $2*10^8$ per replicate per time point. All the buffers were supplemented with protease inhibitor cocktail (Sigma-Aldrich, P8340) and phosphatase inhibitors (Millipore Sigma, 4906837001). After washing with DPBS (Thermo Fisher Scientific, 14190169), cellular fractionation of $2*10^7$ cells per reaction was performed according to the previously published protocol (Conrad & Ørom, 2017). Isolated chromatin was subjected to the Micrococcal Nuclease (MNase; NEB, M0247S) digestion at 37°C and 1,400 rpm for 90 sec followed by stopping the reaction with 25 mM EGTA (Bioworld, 40520008-1). The solution was clarified by centrifugation at 4°C and 13,000 g for 5 min and the supernatants corresponding to the same sample were pooled and utilized for immunoprecipitation (IP). Supernatant was diluted eightfold 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)). RNA Polymerase II antibody (MBL Life science, MABI0601) was coupled to Dynabeads M-280 Sheep Anti-Mouse IgG (Thermo Fisher Scientific, 11201D) according to the manufacturer's instructions and added to the digested chromatin at 30 μg per $2*10^8$ cells. IP was performed at 4°C and 8 rpm on a rotating wheel for 1 h. Afterwards, beads were washed seven times with IP buffer and one time with PNKT buffer (1× T4 PNK buffer (NEB, M0236L), 0.1% vol/vol Tween-20). For 5' RNA phosphorylation, beads were resuspended in PNK reaction mix (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. After the reaction, beads were washed one time with IP buffer and mixed with QIAzol Lysis Reagent (Qiagen, 79306) by vortexing for 1 min. RNA was extracted according to the manufacturer's instructions (Qiagen). Precipitation of RNA was performed with GlycoBlue Coprecipitant (Thermo Fisher Scientific, AM9515) in 100% ethanol overnight. RNA was size-selected in a range of 25-110 nt using a denaturing 6% (wt/vol) polyacrylamide gel with 7M urea. Then, RNA was extracted from the gel in elution buffer (1 M NaOAc pH 5.5, 1 mM of EDTA pH 8.0) on a

rotating wheel and precipitated with GlycoBlue Coprecipitant (Thermo Fisher Scientific, AM9515) in 100% ethanol overnight. RNA libraries were prepared with NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, E7300S) according to the manufacturer's instructions. Libraries were size-selected with 4% E-Gel High-ReSolution Agarose Gels (Thermo Fisher Scientific, G501804) and purified using QIAquick Gel Extraction Kit (Qiagen, 28706X4) according to the manufacturer's instructions. Concentration and fragment size distribution of the libraries were estimated using Fragment Analyzer (Agilent). Libraries were sequenced on Illumina NEXTseq 550 using 75 cycles paired-end mode.

5. ChIP-seq

ChIP-seq protocol was performed as described with minor modifications (Caizzi et al., 2021). Two biological replicates of $3*10^7$ BLaER1 cells were collected at 0, 12, 24 h after transdifferentiation induction. For single crosslinking (Pol II, NELF-E ChIP-seq), cells were fixed in the growth media with 1% methanol-free formaldehyde (Thermo Fisher Scientific, 28908) at room temperature for 10 min. For double crosslinking (CDK9, Cyclin T1, INT11 ChIP-seq), cells were washed once with DPBS (Thermo Fisher Scientific, 14190169) and fixed in DPBS first with 2 mM DSG (Thermo Fisher Scientific, 20593) at room temperature for 20 min, then with 1% methanol-free formaldehyde (Thermo Fisher Scientific, 28908) at room temperature for 10 min. For quenching, 125 mM glycine (Sigma-Aldrich, 50046) was added to the cells and incubated at room temperature for 5 min. Fixed cells were spun down and washed twice with ice-cold DPBS (Thermo Fisher Scientific, 14190169). All the buffers were supplemented with protease inhibitor cocktail (Sigma-Aldrich, P8340) and phosphatase inhibitors (Millipore Sigma, 4906837001). Cell pellet was resuspended in Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% vol/vol NP-40) and incubated on ice for 10 min. Isolated nuclei were washed once with ice-cold DPBS (Thermo Fisher Scientific, 14190169) and resuspended in sonication buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.4% SDS) followed by incubation on ice for 10 min. Suspension was transferred into 1 ml AFA milliTUBE (Covaris, 520130) and subjected to sonication using S220 Focused-ultrasonicator (Covaris) with the following parameters: 5% duty cycle, 140 W peak incident power, 200 cycles per burst, 960 sec processing time, 4-7°C bath temperature, continuous degassing

mode, 8 water level. Sonicated chromatin was centrifuged at 4°C and 10,000 g for 15 min and supernatant was transferred to a new tube. DNA was quantified and analyzed on 1% agarose gel to confirm 200-500 bp fragment size distribution. Antibodies were coupled to Dynabeads Protein G (Thermo Fisher Scientific, 10004D) according to the manufacturer's instructions. The following antibodies were used: RNA Polymerase II NTD antibody (Cell Signaling, 14958), Cyclin T1 antibody (Cell Signaling, 81464), CDK9 antibody (Abcam, ab239364), NELF-E antibody (Bethyl Laboratories, A301-913A), INT11 antibody (Bethyl Laboratories, A301-274A), *Drosophila* H2Av antibody (Active Motif, 61686). Immunoprecipitation (IP) was performed with 50-100 μg chromatin per sample. *Drosophila* S2 spike-in chromatin was added in the amount of 122-400 ng per sample. 1% of each sample was kept as input and stored at 4°C. Chromatin was diluted with IP buffer (56.25 mM Tris-HCl pH 7.5, 157.5 mM NaCl, 1 mM EDTA, 1.125% vol/vol Triton X-100, 0.1125% vol/vol sodium deoxycholate) to obtain 0.1-0.05% SDS concentration. Chromatin was mixed with antibody-bead complexes and incubated on a rotating wheel at 4°C overnight. Next day, beads were washed five times with LiCl wash buffer (100 mM Tris-HCl pH 7.5, 500 mM LiCl, 1% vol/vol NP-40, 1% wt/vol sodium deoxycholate) and one time with TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA). DNA was eluted from the beads at 70°C for 10 min and decrosslinked at 65°C overnight along with input samples. Then DNA was subsequently treated with RNase A (Thermo Fisher Scientific, EN0531) and proteinase K (Thermo Fisher Scientific, AM2546) followed by precipitation in 100% EtOH. DNA concentration was measured with Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and IP enrichment over input was analyzed with RT-qPCR. Equal amounts of DNA were used for the library preparation with NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) according to the manufacturer's instructions. Concentration and fragment size distribution of the libraries were estimated using Fragment Analyzer (Agilent). Libraries were sequenced on Illumina NEXTseq 550 using 75 cycles paired-end mode.

6. ChIP-nexus

Two biological replicates of BLaER1 cells were collected at 0 and 96 h post transdifferentiation induction and treated with 5 μM triptolide (Sigma-Aldrich, T3652) or

DMSO (Sigma-Aldrich, D2438) as described in Treatments section. $3*10^7$ cells per condition were single-crosslinked as described in ChIP-seq section. Cell lysis, nuclei isolation, chromatin shearing and quantification, IP were performed as described in ChIP-seq section. For IP, the following antibodies were used: RNA Polymerase II NTD antibody (Cell Signaling, 14958), *Drosophila* H2Av antibody (Active Motif, 61686). After IP, all downstream steps were performed exactly as in published protocol with minor modifications listed below (Krueger et al., 2019). Amplified DNA libraries were size-selected using 4% E-Gel High-ReSolution Agarose Gels (Thermo Fisher Scientific, G501804) and purified using QIAquick Gel Extraction Kit (Qiagen, 28706X4) according to the manufacturer's instructions. Concentration and fragment size distribution of the libraries were estimated using Fragment Analyzer (Agilent). Libraries were sequenced on Illumina NEXTseq 550 using 75 cycles paired-end mode.

7. Western Blotting of the whole cell lysate

Two biological replicates of BLaER1 cells were collected at 0 and 96 h of transdifferentiation and treated with triptolide (Sigma-Aldrich, T3652) or DMSO (Sigma-Aldrich, D2438) as described in Treatments section. $3*10^7$ cells per each condition were collected by centrifugation at RT and 300 g for 5 min. Cell pellet was resuspended in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% vol/vol NP-40, 0.5% wt/vol sodium deoxycholate, 0.1% vol/vol SDS) supplemented with 500 U/mL Benzonase (Sigma-Aldrich, E1014), 2 mM MgCl2, protease inhibitor cocktail (Sigma-Aldrich, P8340) and phosphatase inhibitors (Millipore Sigma, 4906837001). The lysate was incubated on ice for 20 min with occasional mixing and centrifuged at 4°C and 15,000 rpm for 15 min. The supernatant was transferred to a new tube and protein concentration was measured using Bradford assay (Bio-rad, 5000006) according to manufacturer's instructions. 7-10 μg of the protein were loaded in NuPAGE LDS Sample Buffer (Thermo Fisher Scientific, NP0007) supplemented with 400 mM DTT and subjected to SDS-PAGE (Bio-rad) followed by the transfer to the PVDF membrane (Bio-rad, 1704156). The membrane was blocked using 5% milk in PBS containing 0.05% (v/v) Tween and incubated using 2% milk in PBS containing 0.05% (v/v) Tween with the primary antibodies: RNA Polymerase II NTD antibody (Santa-cruz, sc-55492), GAPDH antibody (Sigma-Aldrich, G8795). Next, the membranes were washed in PBS containing

0.05% (v/v) Tween and incubated with HRP-coupled secondary anti-mouse antibody (Abcam, ab5870). After the washing in PBS containing 0.05% (v/v) Tween, the membranes were developed with Pierce ECL Plus Western Blotting Substrate (Thermo Scientific, 32109) on INTAS imager (GmbH) according to manufacturer's instructions.

III. RESULTS

This section is adapted from Lysakovskaia, Devadas* et al., in preparation*

(co-shared authorship)*

*Author contributions: **K.L.**, A.D. and M.L. planned the project. **K.L.** designed and performed the experiments. A.D. designed and conducted bioinformatical analysis. **K.L.** and A.D. interpreted the data. B.S. and M.L. advised on bioinformatical analysis. J.C. provided TT-seq data prior to publication. **K.L.** wrote the original draft of the manuscript with input from all authors. **K.L.** and A.D. visualized data. P.C. and M.L. conceived, planned and supervised the project. P.C. acquired funding.*

1. TRANSDIFFERENTIATION AS A MODEL TO STUDY POL II PAUSING DURING CELL FATE TRANSITION

1.1. INTRODUCTION AND VALIDATION OF THE TRANSDIFFERENTIATION SYSTEM

To investigate the dynamics of Pol II promoter-proximal regulation in the context of cell fate determination, we used a previously described transdifferentiation system as a model (Choi et al., 2021; Rapino et al., 2013; Stik et al., 2020). In this system, human precursor leukemia B-cells (BLaER1) can be reprogrammed into functional macrophage-like cells by estrogen-inducible C/EBPα overexpression within 96-168 h (Rapino et al., 2013). For our experiments, we selected the time points with the most prominent changes in gene expression after transdifferentiation induction based on publicly available RNA-seq data (**Figure 2A**) (Choi et al., 2021). Since nascent RNA could be obtained at the latest 96 h post transdifferentiation induction in the previous study, we carried out our experiments until this time point (Choi et al., 2021). Importantly, the initial work demonstrated that BLaER1 cells acquire macrophage phenotype already after 96 h post estrogen induction (Rapino et al., 2013). To confirm the efficiency and synchronicity of transdifferentiation, we harvested the cells after 0, 12, 24, 72 and 96 h post estrogen induction and performed total RNA extraction followed by RT-qPCR analysis of the cell type specific markers (**Figure 2A** and **S1A**). We observed 70-90% of transdifferentiated cells at 96 h according to the key B-cell

and macrophage specific markers' expression (**Figure S1A**), in agreement with previously performed gene expression and FACS analysis (Choi et al., 2021; Rapino et al., 2013).

1.2. TRANSCRIBED UNITS AND PAUSE ANNOTATION

Next, we aimed to estimate Pol II kinetic parameters across the time course of transdifferentiation using a previously described multi-omics approach (Gressel et al., 2017, 2019). For this, we conducted mNET-seq experiments with the antibody recognizing total (phosphorylated and unphosphorylated) Pol II CTD (Materials and methods) at 0, 12, 24, 72 and 96 h post transdifferentiation induction for two independent biological replicates (Pearson correlation coefficient = 0.96-0.99) (**Figure S1B**). As discussed before, mNET-seq provides genome-wide occupancy of Pol II associated with the nascent transcript (Nojima et al., 2015, 2016). Furthermore, we used published TT-seq data (Choi et al., 2021) for the same time points providing a measure of new RNA synthesis (Schwalb et al., 2016). According to both TT-seq and mNET-seq data, we observed near complete downregulation of B-cell specific and upregulation of macrophage specific gene transcription confirming our RT-qPCR data (**Figure S1A and S1C**). TT-seq allows estimation of the productive initiation frequency (pIF) (**Figure 2B**) (Gressel et al., 2017, 2019; Schwalb et al., 2016), which we defined as the number of Pol II enzymes that initiate transcription and are successfully released from the promoter-proximal pausing into productive elongation. Combination of both mNET-seq and TT-seq gives an estimation of apparent pause duration (aPD) (Gressel et al., 2017, 2019), which we defined as the average time that Pol II resides in the promoter-proximal region prior to the transition into productive elongation. We use the terms 'productive' and 'apparent' since an unknown fraction of the actual initiated polymerases might terminate prematurely in the promoter-proximal region (Kamieniarz-Gdula & Proudfoot, 2019).

In order to define robust and unique transcription start sites (TSSs) and hence pause positions for every gene, we created an annotation containing only the major transcribed isoforms for each gene using RNA-seq data (**Figure 3A**, Supplementary methods). Along with the eRNAs annotated previously using GenoSTAN and TT-seq data (Choi et al., 2021), we obtained 36,136 unique transcribed units (TUs) including 8,765 mRNAs, 9,904 lncRNAs and 8,165 eRNAs that formed the main transcript types in our final

annotation (**Figure 3A**). Using our mNET-seq data, we extracted the positions of promoter-proximally paused polymerases for all genes that showed significant mNET-seq signal enrichment above background within 250 bp downstream of the TSS and classified these genes as paused (Supplementary methods). We further calculated the signal within +/-100 bp around this mNET-seq derived pause position and referred to this region as the pause window. A total of 10,807 annotated TUs across various biotypes were determined to be paused at at least one time point during transdifferentiation (**Figure 3A**). We additionally calculated the fraction of paused TUs among all the expressed ones at the terminal time points of transdifferentiation. Notably, we obtained significant numbers of paused TUs at both 0 h (n = 8,321) and 96 h (n = 8,976) (**Figure S2A**). These numbers provide an opportunity for robust analysis of the Pol II pausing kinetics in the transdifferentiation system. The pause position was found to be around 51 bp for most of the TUs (**Figure 3B**), which is consistent with the previous findings in human cells (Gressel et al., 2019; Larke et al., 2021; Tome et al., 2018). Additionally, we observed that the vast majority of annotated TUs did not exhibit a significant change in the pause positions at the different cell stages during transdifferentiation (**Figure 3C**).

1.3. Distinct patterns of RNA synthesis changes throughout transdifferentiation

To examine the changes in RNA synthesis, we performed differential expression (DE) analysis based on TT-seq data (|log2FC| > 1.5, padj < 0.05; Materials and methods; **Figure S2B**). We identified 10,748 DE TUs among which 5,438 appeared in the pause annotation including 2,393 mRNAs, 893 lncRNAs and 1,885 eRNAs (**Figure 3A**). During the time course of transdifferentiation, RNA synthesis of protein-coding DE TUs changed in four distinct directions: downregulated (n = 938), bending (n = 316), peaking (n = 487) and upregulated (n = 416) (**Figure 3D**). Analysis of all DE TUs demonstrated similar patterns of RNA synthesis change (**Figure S2C**). Hereafter, we refer to the downregulated TUs as pre-B genes, since they demonstrated the expression maximum at 0 h of transdifferentiation representing a precursor B-cell leukemia stage (**Figure 3D** and **S2C**). The upregulated TUs were termed iMac genes, since their expression maximum was observed at 96 h of transdifferentiation indicating a macrophage-like or induced macrophage cell stage (**Figure 3D** and **S2C**). For the downstream analysis, we used this subset of clustered protein-coding

DE TUs unless stated otherwise. Moreover, we focused our analysis on the terminal cell stage gene groups pre-B and iMac genes. To identify the biological functions of the selected groups, we performed gene ontology (GO) analysis (Supplementary methods). We found that pre-B genes were enriched for biogenesis, metabolic and cellular processes – the hallmarks of actively transcribing cancer cells; whereas iMac genes were enriched for the immune and macrophage-specific functions (**Figure 3E**), adding to the validity of the classification of our gene groups.

To summarize, the transdifferentiation appears to be a well-suited system to study Pol II pausing during cell fate transition due to its simplicity, high efficiency and significant number of paused-called genes providing a possibility to statistically analyze transcriptional kinetic parameters.

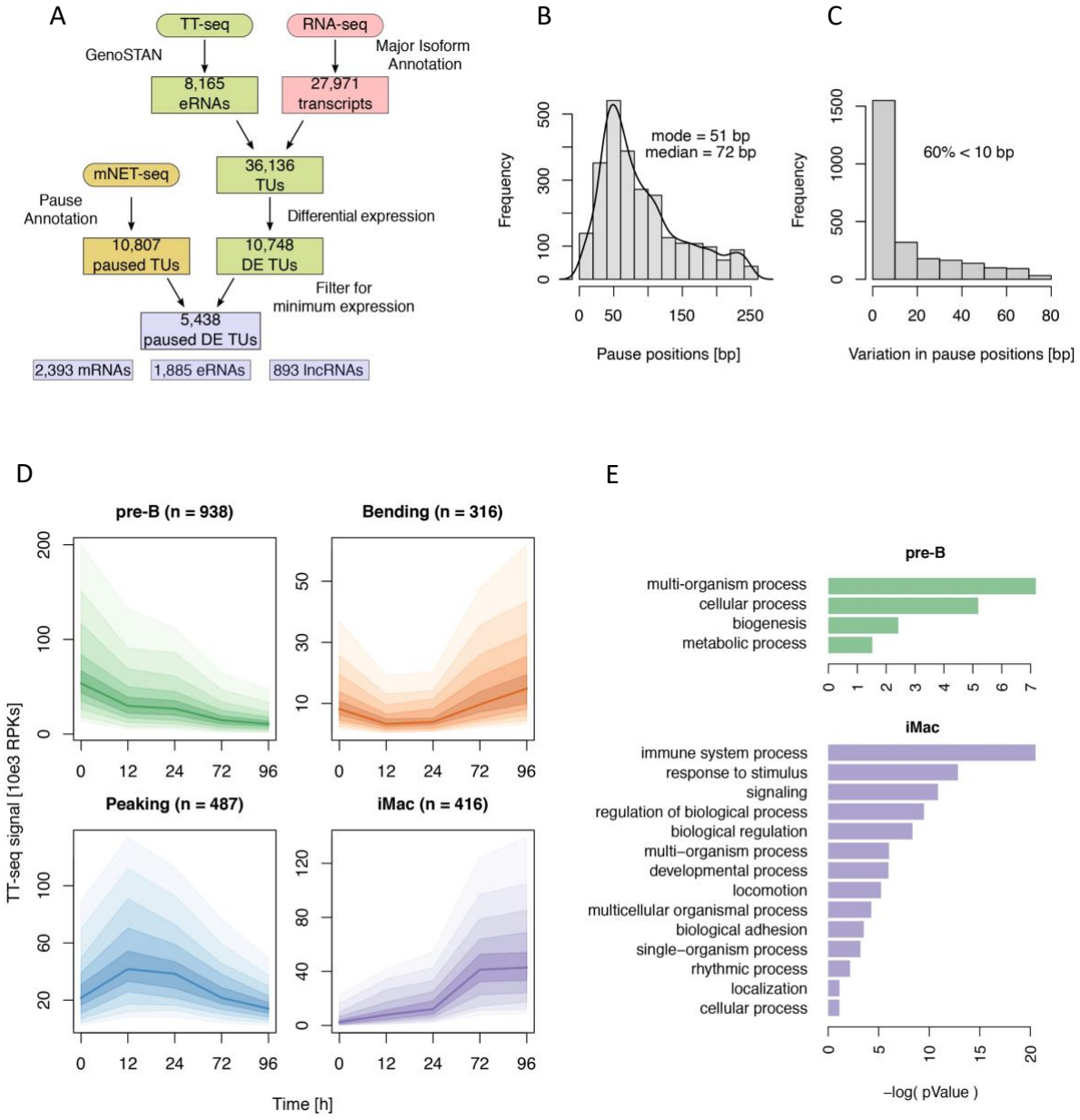


Figure 3. Transdifferentiation as a system to study cell fate transition.

A. Schematic representation of TU annotation. **B.** Histogram showing the distribution of Pol II promoter-proximal pause positions. Data is shown for 10,807 TUs that were called as paused across all time points of transdifferentiation based on the mNET-seq. **C.** Histogram representing the variation in the Pol II promoter-proximal pause positions called at different time points of transdifferentiation. Data is shown for 10,807 TUs that were called as paused across all time points of transdifferentiation based on the mNET-seq. **D.** Representation of the clustering of 2,157 protein-coding DE TUs changed during transdifferentiation time course based on the TT-seq. Expression changes occurred in 4 defined patterns named pre-B (downregulated), bending, peaking and iMac (upregulated) correspondingly. **E.** Gene ontology results for pre-B and iMac genes.

2. TRANSCRIPTIONAL KINETICS OF PRE-B GENES INDICATES THEIR PROMOTER-PROXIMAL REGULATION

Next, we looked in detail at the transcriptional kinetics of the pre-B genes (n = 938) which were downregulated throughout the transdifferentiation (**Figure 4**). We observed a dramatic decrease of TT-seq signal all over the gene from 0 h to 96 h, while the mNET-seq signal was mainly diminished in the gene body in comparison to the promoter-proximal region indicating the accumulation of transcriptionally engaged Pol II towards the macrophage-like stage of transdifferentiation (**Figure 4A**). This is consistent with the previously published findings reporting that Pol II occupancy marks promoter-proximal regions post gene activation during development or differentiation in *Drosophila* and cell reprogramming or embryonic development in mammals (Gaertner et al., 2012; B. Liu et al., 2020; L. Liu et al., 2014; Zeitlinger et al., 2007). We further estimated pIF for the pre-B genes and observed a significant gradual decrease (median $FC_{0\text{-}96}$ = 0.25, p-value < 2.2e-16) that reflected changes in the TT-seq signal from 0 h to 96 h of transdifferentiaiton (**Figure 4B**). Quantification of the Pol II pause window coverage (PWC) using mNET-seq signal showed a slight, significant decrease from 0 h to 12 h of transdifferentiation (median $FC_{0\text{-}12}$ = 0.80, p-value = 1e-3), however, remained unchanged comparing with 96 h of transdifferentiation (median $FC_{0\text{-}96}$ = 0.94, p-value = 4e-3) (**Figure 4B**). Next, we calculated the aPD and detected a significant gradual increase (median $FC_{0\text{-}96}$ = 4.46, p-value < 2.2e-16) from 0 h to 96 h of transdifferentiation suggesting Pol II retention downstream of the TSS (**Figure 4B**). The observed reciprocal behavior of pIF and aPD was reported previously, proposing that paused Pol II limits the transcription initiation (Gressel et al., 2017; Shao & Zeitlinger, 2017). Additionally, we visualized these results for individual exemplary genes in the genome browser plots (**Figure 4C**).

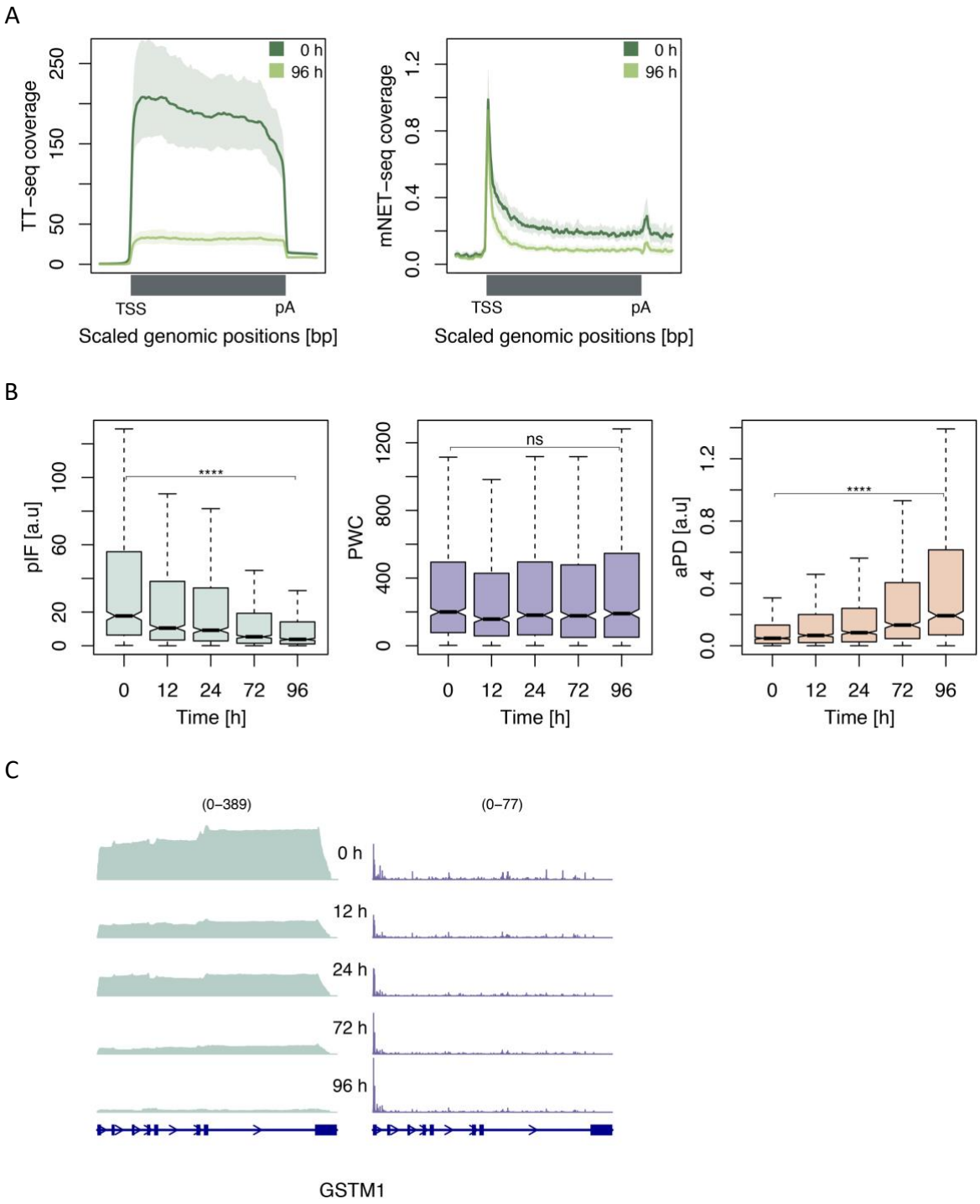


Figure 4. Kinetic parameters for pre-B genes suggest their promoter-proximal regulation.
A. Metagene profiles comparing TT-seq and mNET-seq signals between 0 h (dark green) and 96 h (light green) post transdifferentiation induction for pre-B genes. Both TT-seq and mNET-seq signals were averaged for 938 TUs (pre-B genes) and scaled between TSS and poly(A)-site. Solid lines represent the averaged signal, shaded regions show 95% confidence interval of the mean estimated from the bootstrap method. **B.** Boxplots representing PWC and kinetic parameters (pIF and aPD) estimates for pre-B genes across the time course of transdifferentiation. **C.** Representative example of the pre-B gene (GSTM1) TT-seq and mNET-seq data collected at 0, 12, 24, 72 and 96 h post transdifferentiation induction. Each dataset is averaged for two independent biological replicates.

The presence of transcriptionally engaged Pol II in the pause window and the aPD increase at later time points of transdifferentiation suggested a promoter-proximal pause regulation for pre-B genes. Thus, we aimed to confirm whether the observed changes in aPD are dependent on the P-TEFb that was shown to release paused Pol II complex into productive elongation (Fujinaga et al., 2004; Marshall et al., 1996; Marshall & Price, 1995; Vos, Farnung, Boehning, et al., 2018). We selected the time points that exhibited the most prominent changes in kinetic parameters 0, 24 and 96 h post transdifferentiation induction (**Figure 4B**). We carried out the ChIP-seq experiments for P-TEFb subunits including CDK9 and predominant cyclin T1 for two independent biological replicates (Pearson correlation coefficient = 0.98-0.99) (**Figure S3A** and **S3B**). A gradual decrease in both CDK9 and cyclin T1 occupancy was observed during the transdifferentiation, which was in accordance with increase in aPD (**Figure 5** and **S3E**). Additionally, we investigated whether observed increase in aPD correlated with gene occupancy of NELF multi-subunit factor, which was shown to stabilize paused Pol II complex in the promoter-proximal region (Vos, Farnung, Urlaub, et al., 2018; Yamaguchi et al., 2002, 2013). For this, we conducted the ChIP-seq experiment for the NELF-E subunit at 0, 24 and 96 h of transdifferentiation for two independent biological replicates (Pearson correlation coefficient = 0.91-0.93) (**Figure S3C**). Unexpectedly, we detected a decrease in NELF-E occupancy throughout the transdifferentiation along with aPD increase (**Figure 5**). We furthermore analyzed the total Pol II occupancy at 0, 24 and 96 h post transdifferentiation induction via ChIP-seq for two independent biological replicates (Pearson correlation coefficient = 0.99) (**Figure S3D**) and observed a slight signal reduction in the pausing region across the time course (**Figure 5** and **S3F**). Consistent with our mNET-seq results, Pol II ChIP-seq coverage in the gene body was decreased (**Figure 4A**, **5** and **S3F**). Interestingly, we detected a correlation between Pol II and NELF-E ChIP-seq signal in the promoter-proximal region (Pearson correlation coefficient = 0.74-0.89) (**Figure S3G**). We therefore propose that NELF appears to stabilize the paused Pol II complex during the transcriptionally active state of pre-B genes at 0 h of transdifferentiation. These observations agreed with the previous studies showing that loss of NELF results in the both promoter-proximal and gene body Pol II signal reduction indicating defects of the Pol II pause release into productive elongation of transcription (Core et al., 2012; Gilchrist et al., 2008, 2010; Williams et al., 2015). Additionally, recent work demonstrated accumulation of paused Pol II at the +1 nucleosome and pause release impairment at the target genes in the

heat-shock conditions under rapid depletion of NELF-C subunit (Aoi et al., 2020). To summarize, our data show that the pre-B gene group exhibits a promoter-proximal Pol II regulation across the time course of transdifferentiation based on the kinetics calculations. Furthermore, the detected that increase in aPD is correlated with the reduction of the levels of the pause release factor P-TEFb from 0 h to 96 h of transdifferentiation. We finally suggest that NELF complex is required for the promoter-proximal Pol II pausing stabilization for the pre-B genes at the stage when they are transcriptionally active and released into productive elongation.

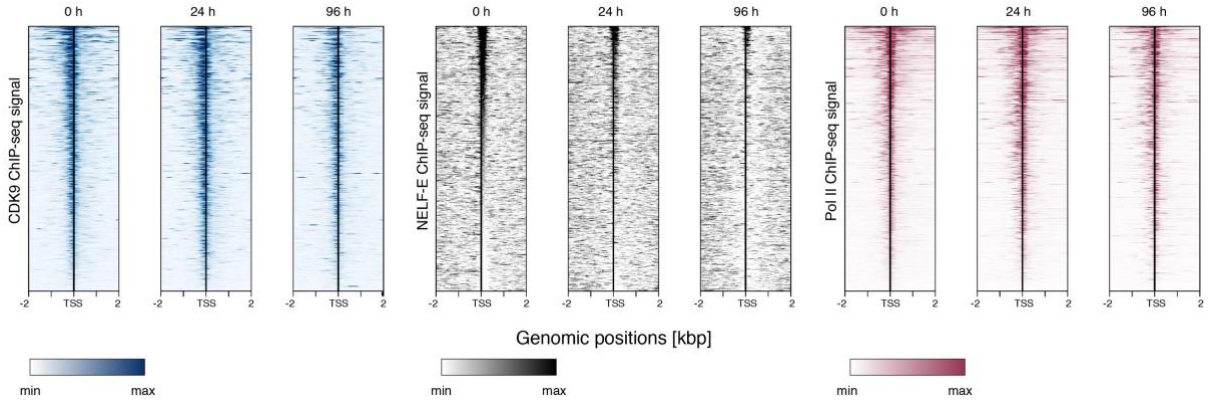


Figure 5. Occupancy of the pausing and pause release factors follows the activity of pre-B gene group during transdifferentiation.
Heatmaps displaying ChIP-seq signal ranked by intensity at 96 h for 938 TUs (pre-B genes) in the range of +/-2 kbp and quantified in the pause window (NELF-E) or in the +/-1 kbp region (CDK9, Pol II). Occupancy of the CDK9, NELF-E and Pol II factors is shown for 0, 24 and 96 h of transdifferentiation.

3. iMac genes display two groups that differ in transcription kinetics and biological functions

We further estimated the transcription kinetic parameters for the iMac gene group (n = 416) that activate their expression towards the macrophage-like cell state (**Figure 6**). pIF showed a significant gradual increase (median $FC_{0\text{-}96}$ = 5.2, p-value < 2.2e-16) (**Figure 6A**). Correspondingly to iMac gene group upregulation and pIF increase, PWC of Pol II similarly increased across the time course of transdifferentiation (median $FC_{0\text{-}96}$ = 4.5, p-value < 2.2e-

16) (**Figure 6A**). Strikingly, aPD did not show significant changes at the terminal time points of transdifferentiation (median $FC_{0\text{-}96}$ = 0.87, p-value = 0.06) (**Figure 6A**). We then analyzed our collected ChIP-seq data for the P-TEFb components CDK9 and cyclin T1 and observed a gain in the signal for the iMac genes groups throughout the transdifferentiation (**Figure 6B** and **S4A**). Additionally, total Pol II ChIP-seq signal accumulated in the promoter-proximal region already at 0 h and increased all over the gene at 96 h of transdifferentiation suggesting a pattern of active transcription elongation (**Figure 6B** and **S4B**). These observations indicated a possible promoter-proximal regulation of iMac genes, however with no obvious correlation with estimated aPD values (**Figure 6A**). Interestingly, NELF-E occupancy was clearly decreased from 0 h to 96 h of the transdifferentiation for iMac genes that resembled the pattern of NELF-E ChIP-seq for pre-B genes (**Figure 5** and **6B**). Since we argued about NELF-mediated regulation of the promoter-proximal Pol II pausing during the active release into productive elongation, we compared the NELF-E ChIP-seq signal for both pre-B and iMac gene groups across the transdifferentiation on the same graph (**Figure S4C**). Notably, the NELF-E levels appeared to be dependent on the transcriptional activity within a certain cell stage or a time point rather than between the different time points. In other words, NELF-E ChIP-seq signal was always higher for the gene group that was actively transcribed at the specific transdifferentiation cell state confirming our initial hypothesis (**Figure S4C**). We additionally ranked pIFs for our analyzed 2,157 TUs from lowest to highest value and plotted against corresponding NELF-E ChIP-seq signal. We obtained a positive correlation of these datasets supporting our findings (**Figure S4D**). Moreover, the decrease trend in NELF-E ChIP-seq signal for both pre-B and iMac gene groups can be explained by the overall reduction of transcriptional activity in the macrophage-like cells that become quiescent by the 168 h of transdifferentiation (Rapino et al., 2013). Indeed, the expression of the gene encoding NELF-B subunit is gradually decreased across the transdifferentiation with the minimum value at 96 h (**Figure S4E**).

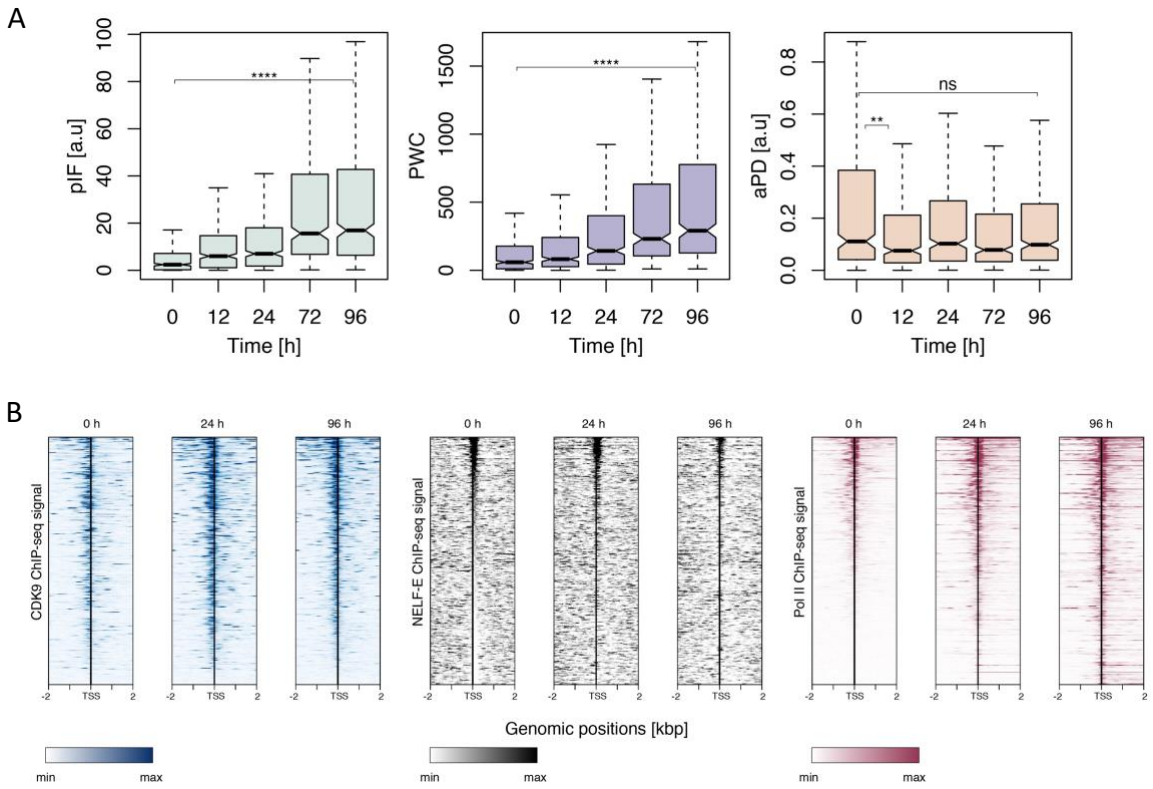


Figure 6. Pausing kinetics of iMac genes is poorly correlated with the pausing and pause release factors occupancy.

A. Boxplots representing PWC and kinetic parameters (pIF and aPD) calculations for the iMac genes across the time course of transdifferentiation. **B.** Heatmaps displaying ChIP-seq signal ranked by intensity at 0 h for 416 TUs (iMac genes) in the range of +/- 2 kbp and quantified in the pause window (NELF-E) or in the +/-1 kbp region (CDK9, Pol II). Occupancy of the CDK9, NELF-E and Pol II factors is shown for 0, 24 and 96 h of transdifferentiation.

Apart from our CDK9 and total Pol II ChIP-seq observations suggesting promoter-proximal Pol II regulation for the iMac gene group, a number of studies strongly proposed that Pol II stalling or pausing is involved in the developmental or cell fate transition processes (Introduction, section 3). We therefore used k-means clustering (k=2) with regards to the kinetic parameters for the iMac genes and obtained two distinct gene groups demonstrating different dynamics of pIF and aPD across the transdifferentiation (**Figure 7A**). Clustering with k>2 resulted in additional groups with similar patterns of pIF and aPD indicating the presence of only two prominent kinetics scenarios (**Figure S5A**). We named the obtained gene groups iMac I (n = 193) and iMac II (n = 223). Both of these groups

exhibited a gradual increase in pIF during the transdifferentiation, however, iMac II genes showed a greater pIF upregulation (median FC_{0-96} = 11.07, p-value < 2.2e-16) than iMac I genes (median FC_{0-96} = 2.72, p-value < 2.2e-16) (**Figure 7A**). Interestingly, we observed a reciprocal behavior of aPD only for the iMac II genes (median FC_{0-96} = 0.38, p-value = 8.37e-11), whereas aPD for the iMac I genes was slightly increased (median FC_{0-96} = 1.94, p-value = 8.8e-5) (**Figure 7A**). Previously, our laboratory reported that human genes have a pause-initiation limit according to which Pol II promoter-proximal pausing restricts the new transcription initiation, however, other combinations of pIF and aPD are also possible (Gressel et al., 2017). According to this principle, we suggested that for iMac II gene group the observed aPD decrease is required to allow increase of pIF resulting in the reciprocal patterns of these parameters across the transdifferentiation (**Figure 7A**). We detected similar reciprocal behavior of kinetic parameters in the pre-B gene group where decrease in pIF can be mediated by corresponding increase in aPD during the transdifferentiation (**Figure 4**). However, we did not observe this anti-correlation for the iMac I gene group that increased both parameters from 0 h to 96 h of transdifferentiation (**Figure 7A**). In agreement with pause-initiation limit theory, the increase of pIF and aPD can occur simultaneously if aPD remains short enough not to restrict pIF, as is the case with our iMac I gene group (Gressel et al., 2019). Metagene profiles of the TT-seq signal for the iMac gene groups reflected estimated pIF at 0 and 96 h of transdifferentiation (**Figure 7B**). The mNET-seq metagene profiles illustrated clear increase of transcriptionally active Pol II in the pausing window for both groups from 0 to 96 h, while the gene body coverage increased remarkably higher for iMac II genes implying more Pol II release into productive elongation in comparison to iMac I genes at 96 h of transdifferentiation (**Figure 7B**). We additionally visualized the TT-seq and mNET-seq data in the genome browser plots and could detect similar patterns for the individual candidates of the iMac I and iMac II gene groups (**Figure S5B**).

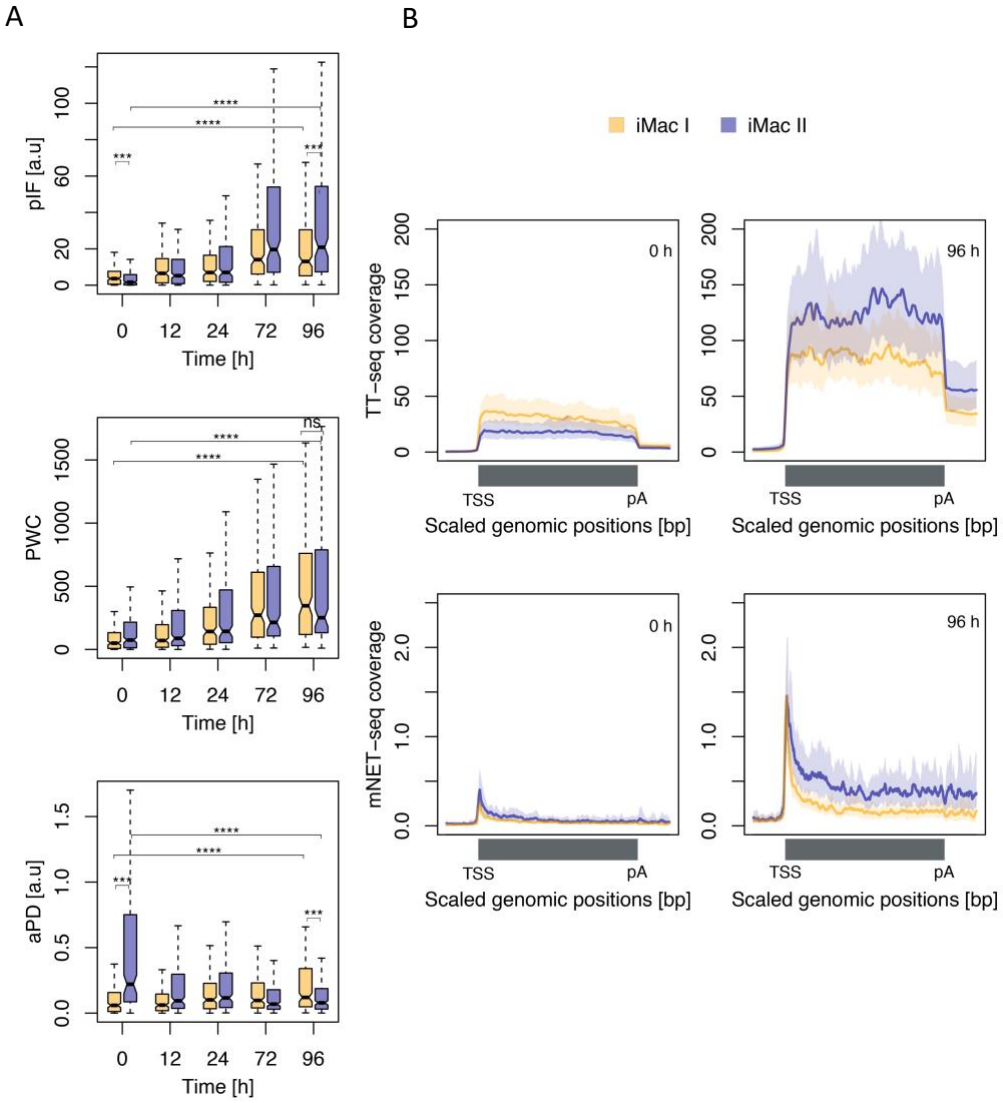


Figure 7. iMac genes include two groups with distinct kinetic parameters.
A. Comparison of the PWC and kinetic parameters (pIF and aPD) calculations between the iMac I and iMac II genes across the time course of transdifferentiation. **B.** Metagene profiles comparing TT-seq and mNET-seq signals for iMac I and iMac II genes at 0 h and 96 h post transdifferentiation induction. Both TT-seq and mNET-seq signals were averaged for 193 (iMac I genes) and 223 (iMac II genes) TUs and scaled between TSS and poly(A)-site. Solid lines represent the averaged signal, shaded regions show 95% confidence interval of the mean estimated from the bootstrap method.

Taken together, our genome-wide data suggested that both iMac gene groups get activated during the transdifferentiation, however, iMac II genes more efficiently release Pol II into productive elongation compared to iMac I genes, which appeared to acquire transcriptionally engaged Pol II in the promoter-proximal region that is less efficiently

proceeding into productive elongation. Therefore, we aimed to examine the differences in the potential regulation mechanisms and implications for the biological functions of iMac I and iMac II genes. We first hypothesized that iMac I genes accumulate promoter-proximal Pol II in order to keep transcriptionally "prepared" state for the further activation in response to stimuli as it was previously shown for the proinflammatory genes (Adelman et al., 2009; Escoubet-Lozach et al., 2011; Hargreaves et al., 2009; Yu et al., 2020). Thus, we compared the GO terms (Supplementary methods) for both iMac gene groups and obtained a higher number of macrophage-specific immune functions for iMac II genes in comparison to iMac I genes that demonstrated minor enrichment in immune system processes terms (**Figure 8A** and **S6A**). Moreover, the total number of genes that displayed enrichment in GO terms was twice as high for iMac II genes compared to iMac I (**Figure S6B**). We additionally performed the pathway analysis for these gene groups using STRING database (version 11.5; Supplementary methods). Interestingly, we found a larger number of interactions for iMac II gene group (average node degree = 4.21) comparing to iMac I gene group (average node degree = 1.65). The STRING reactome pathway analysis for iMac II provided a significant enrichment in macrophage-related signaling cascades including toll-like receptor cascades, cytokine signaling in immune system, interleukins signaling and others (**Table S4**). The same analysis for iMac I gene group exhibited an enrichment only in three reactome pathways including general terms as immune or innate immune response process (**Table S3**). To conclude, we could observe some fraction of the immune response genes among iMac I with pre-loaded Pol II in the promoter-proximal region, which is consistent with previous findings (Adelman et al., 2009; Escoubet-Lozach et al., 2011; Hargreaves et al., 2009; Yu et al., 2020), however, the majority of the GO terms enrichment was detected among iMac II genes that were actively elongating and demonstrated more functional importance for the macrophage-like cells (Watanabe et al., 2019; Wynn et al., 2013). Next, we looked into collected ChIP-seq data for the pausing-related factors to examine how they are related to the observed differences in kinetics of iMac gene groups. Strikingly, the levels of pause release factor P-TEFb components CDK9 and cyclin T1 did not exhibit a significant difference between iMac I and iMac II gene groups across the time course of transdifferentiation (**Figure 8B** and **S6C**). Similarly, occupancy of the pause-stabilizing factor NELF-E subunit remained unchanged between two groups of iMac genes during the transdifferentiation (**Figure 8B**). We then proposed that recruitment of these factors for iMac genes is

independent on the Pol II fraction released into productive elongation, however, their activity might be regulated through the additional mechanisms.

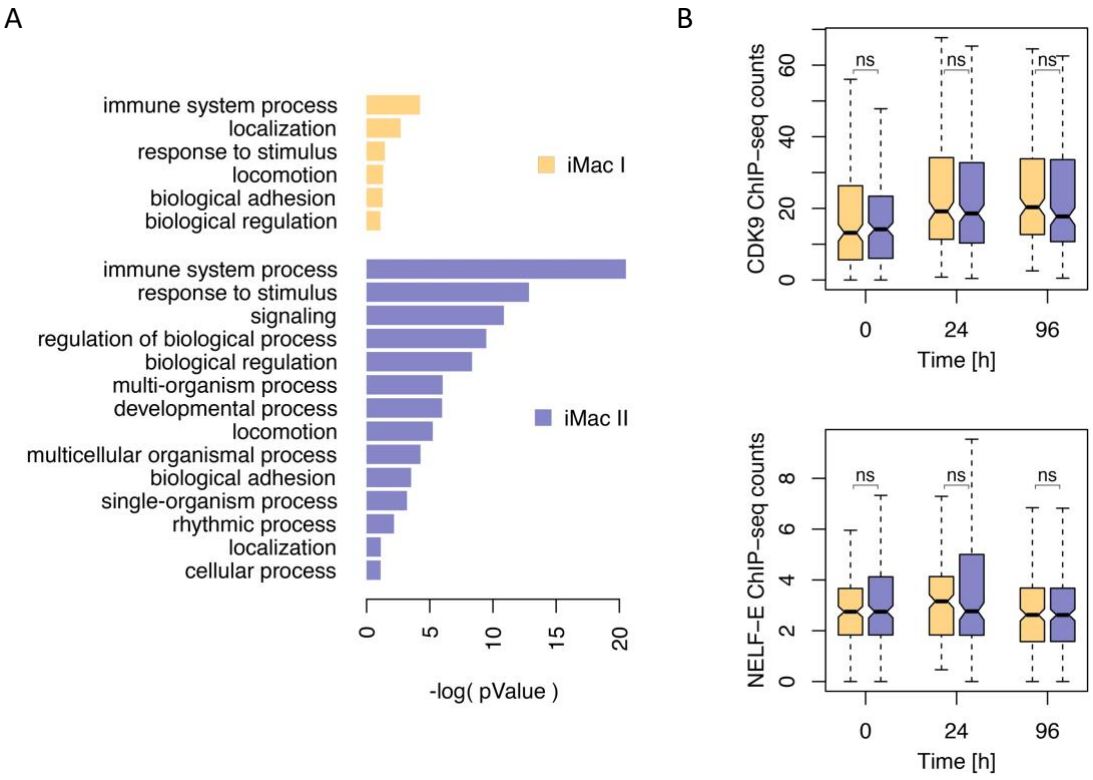


Figure 8. iMac gene groups differ in biological functions but not in the pausing and pause release factors occupancy.

A. Gene ontology terms compared between iMac I (n = 193) and iMac II (n = 223) genes. **B.** Boxplots representing comparison of the CDK9 and NELF-E ChIP-seq signals for the iMac I and iMac II genes at 0, 24 and 96 h post transdifferentiation induction.

To summarize, our findings pose a question about differences in promoter-proximal regulation in the individual iMac gene groups. Notably, Pol II signal detected in the promoter-proximal region can be interpreted as paused Pol II, transcriptionally arrested Pol II or Pol II undergoing a processivity gap (Ehrensberger et al., 2013). Since we observed a pIF increase for iMac I genes from 0 h to 96 h of transdifferentiation (median $FC_{0\text{-}96}$ = 1.94, p-value = 8.8e-5), we excluded the possibility of transcriptional arrest. Therefore, we hypothesized that Pol II transcribing iMac I genes can undergo constant cycles of initiation followed by a processivity gap or premature termination allowing transition into productive

elongation only for a fraction of Pol II. Since a variety of recent studies proposed that Integrator complex attenuates transcription of Pol II in the promoter-proximal region (Elrod et al., 2019; Fianu et al., 2021; K. L. Huang et al., 2020; Lykke-Andersen et al., 2021; Tatomer et al., 2019), we aimed to compare the Integrator complex occupancy between iMac I and iMac II gene groups. We thus conducted ChIP-seq experiments for the cleavage subunit of the complex INTS11 at 0, 24 and 96 h of transdifferentiation for two independent biological replicates (Pearson correlation coefficient = 0.92-1) (**Figure S7A**). Notably, we could not observe significant differences among the iMac gene groups (**Figure 9**). Interestingly, INTS11 ChIP-seq signal exhibited a general decrease for both pre-B and iMac gene groups across the transdifferentiation time course, however, the levels if INTS11 were always higher for the actively transcribed genes within the same time point (**Figure 9** and **S7B**). We also obtained a positive correlation between pIFs of our analyzed 2,157 TUs and corresponding INTS11 ChIP-seq signal (**Figure S7C**). Similar results were observed for NELF-E ChIP-seq that demonstrated a clear correlation with the gene activity at the same transdifferentiation cell stage and with pIFs in general (**Figure S4C** and **S4D**). We additionally confirmed that as for NELF, the expression of the genes encoding Integrator subunits was downregulated towards 96 h of transdifferentiation due to the acquiring transcriptional quiescence (**Figure S7D**). Resembling our observations for NELF and P-TEFb, we suggested that Integrator complex is recruited to both transcriptionally active iMac gene groups at 96 h of transdifferentiation, however, its cleavage activity can be modulated through the distinct mechanisms.

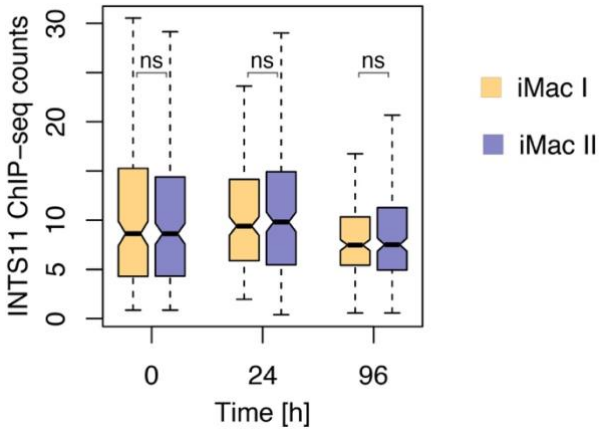


Figure 9. Integrator complex occupancy across the time course of transdifferentiation.
Boxplots representing ChIP-seq signal of INTS11 integrator subunit at 0, 24 and 96 h of transdifferentiation for iMac I (n = 193) and iMac II (n = 223) genes.

4. Initiation inhibition reveals short half-lives of Pol II in the promoter-proximal window

Our data indicated alternative mechanisms of promoter-proximal Pol II transcription regulation in the promoter-proximal region rather than factor-dependent pausing in some of the gene groups involved in the transdifferentiation process. In this regard, the downregulated pre-B gene group exhibited a Pol II accumulation in the promoter-proximal region and decreased the Pol II signal in the gene body towards the macrophage-like cell stage of transdifferentiation (**Figure 4**). However, the paused Pol II stabilizing factor NELF decreased the occupancy following Pol II pattern (**Figure 5**). Since loss of NELF was shown to decrease stability of the promoter-proximal Pol II (Henriques et al., 2013), we hypothesized that observed transcriptional kinetics of pre-B genes could suggest a processivity gap for Pol II in the pausing region rather than a highly stably paused Pol II (Ehrensberger et al., 2013). Moreover, we observed Pol II occupancy gain mostly in the pausing window in comparison to the gene body along with slight increase in productive transcription for the upregulated iMac I gene group during transdifferentiation (**Figure 7**). Notably, the iMac I genes demonstrated the same levels of the pause-related P-TEFb and

NELF factors as iMac II genes that showed a clear pattern of Pol II release into productive elongation (**Figure 8**). These results led us to hypothesize that observed transcriptional kinetics of the iMac gene groups can suggest a higher fraction of Pol II released into productive elongation of transcription for the iMac II genes in comparison to the iMac I genes that might exhibit a processivity gap of promoter-proximal Pol II resulting in the lower productive RNA synthesis (Ehrensberger et al., 2013).

We thus sought to measure the half-lives of Pol II in the pausing window employing high-resolution technique ChIP-nexus of the total Pol II under transcription initiation inhibition in the terminal cell stages of transdifferentiation - pre-B leukemia and macrophage-like cells. This approach was introduced previously using *Drosophila* Kc167 cells and provided an estimation of promoter-proximal Pol II stability in the range of 5 to 20 min for most of the analyzed genes (Shao & Zeitlinger, 2017). In order to block transcription initiation, we used triptolide (TRP), which is a known inhibitor of the general transcription factor TFIIH (Titov et al., 2011). It was reported that TRP covalently binds XPB subunit of TFIIH blocking its ATP-dependent DNA translocase activity *in vitro* (Titov et al., 2011). Additionally, TRP treatment was shown to block initiating Pol II and inhibit RNA synthesis in living cells (F. Chen et al., 2015; Henriques et al., 2013; Jonkers et al., 2014; Krebs et al., 2017; Shao & Zeitlinger, 2017; Vispé et al., 2009). Our downstream experiments were conducted for the cells at 0 and 96 h post transdifferentiation induction with a TRP treatment time series (**Figure 10A**). We first tested the TRP concentration and the treatment time for precursor leukemia B-cells and macrophage-like cells by total RNA extraction and RT-qPCR with the intronic primers in order to catch the nascent RNA response (Materials and methods, **Figure S8A**). The working TRP concentration for nascent RNA synthesis inhibition appeared to be lower for 0 h than for 96 h cells, presumably, due to the biological differences between these cell types. Therefore, we applied the optimized concentration of 5 μM (Materials and methods) based on the results from 96 h cells in order to be consistent with the conditions between the time points (**Figure S8A**). Additionally, since TRP was shown to promote Pol II proteasomal degradation *in vivo* after prolonged treatment times (Manzo et al., 2012; Wang et al., 2011), we performed western blotting analysis of the whole cell lysate to monitor Pol II protein levels after TRP treatment. We confirmed that total Pol II signal was not reduced after TRP treatment up to 30 min that we selected as the

final time point for inhibition at 0 h and 96 h of transdifferentiation (Materials and methods, **Figure S8B**).

Next, we carried out ChIP-nexus experiments using the cells at 0 h and 96 h of transdifferentiation treated with 5 μM TRP or DMSO solvent control for 6 and 30 min for two independent biological replicates (Pearson correlation coefficient$_0$ = 0.91-0.99 and Pearson correlation coefficient$_{96}$ = 0.94-1) (**Figure 10A, S8C** and **S8D**). Interestingly, we observed a global loss of Pol II signal at both upstream and downstream TSS regions that occurred gradually with the time of TRP treatment for the analyzed 2,157 TUs at both time points of transdifferentiation (**Figure 10B**). Additionally, we could resolve both ChIP-nexus peaks corresponding to Pol II blocked as a part of the PIC and promoter-proximal Pol II at the different TRP treatment time points. In detail, Pol II ChIP-nexus signal was initially reduced in the promoter-proximal window and displaying a small accumulation in the region corresponding to the PIC at 6 min of TRP treatment (**Figure 10B**). At 30 min of TRP treatment, we detected a complete shift of Pol II ChIP-nexus signal to the PIC indicating initiation block and release into productive elongation or premature termination of all polymerases that resided in the pausing region (**Figure 10B**). The observed changes in Pol II ChIP-nexus signal under TRP treatment were almost identical for both 0 h and 96 h of transdifferentiation (**Figure 10B**). Similar pattern was detected previously using conventional Pol II ChIP-seq experiments combined with TRP treatment in human cells; however, the resolution of this method is too low to reliably distinguish initiating and promoter-proximal pausing Pol II signals and therefore calculate paused Pol II stability without biases (F. Chen et al., 2015; Erickson et al., 2018). We then estimated the half-life of Pol II in the promoter-proximal region by fitting the Pol II occupancy measurements from the time series of TRP treatment to an exponential decay model (**Figure S8E**). Since we initially collected additional Pol II ChIP-nexus data for the 96 h cells with 15 min DMSO or TRP treatment, we included this time point into our calculations. However, based on our further estimations, we could omit this treatment time point for the ChIP-nexus experiment for 0 h cells (**Figure S8F**). We obtained the distribution of the promoter-proximal Pol II half-lives mainly between 5 and 10 min for our analyzed 2,157 TUs for both time points of transdifferentiation (**Figure 10C**). The median half-life was 6.5 and 6.6 min for 0 h and 96 h transdifferentiation time points respectively (**Figure 10C**). Our estimates agreed with the

previously published studies for mouse, *Drosophila* and human using genome-wide nascent RNA sequencing techniques (Jonkers et al., 2014; Zimmer et al., 2021). Another confirmation appeared from the study suggesting that the process of Pol II turnover varies in the range from 5 to 10 min in human cells (Erickson et al., 2018). In contrast to the previous study using ChIP-nexus with TRP treatment time series to calculate Pol II stability in *Drosophila* (Shao & Zeitlinger, 2017), we could not detect any stable pausing behavior in our data.

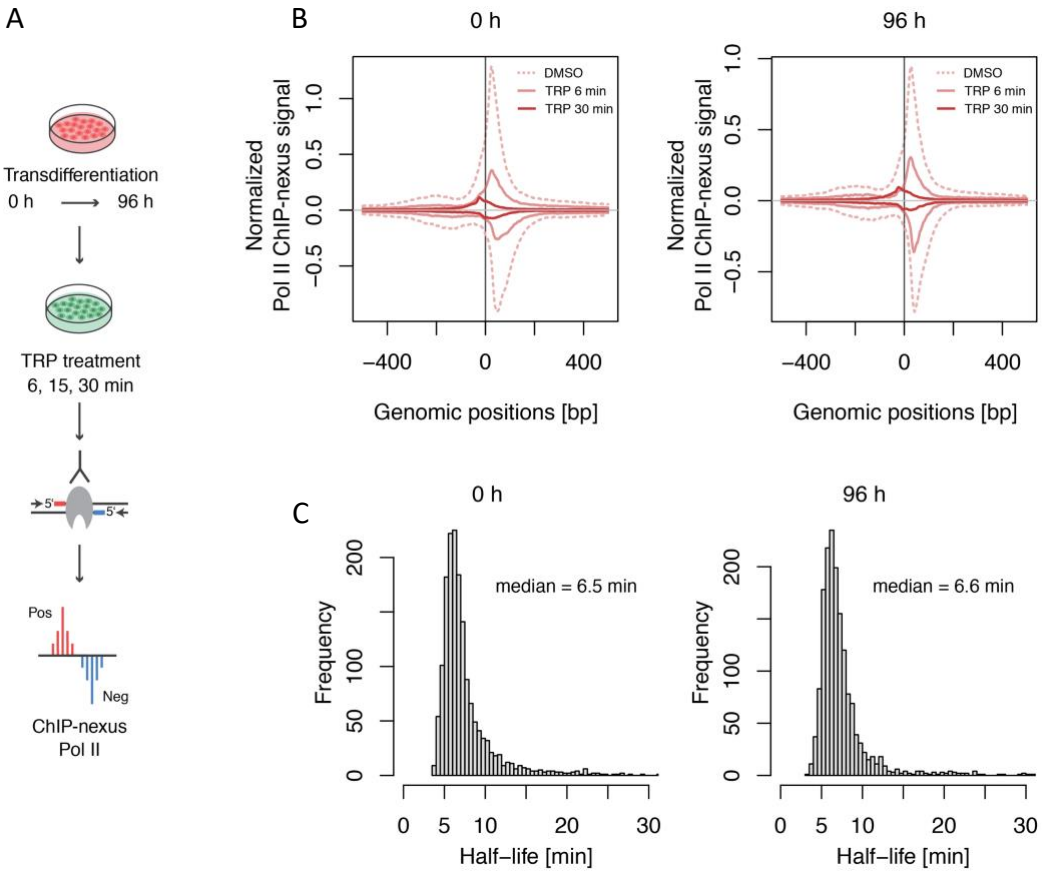


Figure 10. Global high turnover of Pol II in the promoter-proximal window under initiation inhibition.

A. Schematic representation of TRP treatment at 0 and 96 h time point of transdifferentiation followed by ChIP-nexus experiment. **B.** Metagene profiles for the ChIP-nexus signal under 6 and 30 min of DMSO or TRP treatment at 0 h and 96 h of transdifferentiation. The data is shown for 2,157 expressed protein-coding DE TUs. **C.** Histogram displaying calculated Pol II half-lives in the promoter-proximal region at 0 h (median = 6.5 min) and 96 h (median = 6.6 min) of transdifferentiation. The data is shown for 2,157 expressed paused protein-coding DE TUs.

We furthermore ranked the analyzed TUs based on the calculated promoter-proximal Pol II half-lives into five groups ordered from lowest to highest value. Next, we plotted the aPD values estimated for the TUs distributed in these five groups (**Figure 11**). Notably, Pol II pausing index (ratio of the promoter-proximal Pol II signal to the Pol II gene body signal) calculated based on the ChIP-seq data was reported to correlate with Pol II stability in the pausing region (Shao & Zeitlinger, 2017). Strikingly, we did not detect a clear correlation between Pol II half-lives and corresponding aPD values for both transdifferentiation time points (**Figure 11**). We observed the highest aPD for the more stable Pol II and the lowest aPD for less stable Pol II based on our calculations, however, the intermediate Pol II stability groups did not demonstrate significant differences due to the narrow range of half-lives distribution in general (**Figure 10** and **11**). These results indicate the importance of considering promoter-proximal Pol II stability measurements while calculating Pol II pausing kinetics.

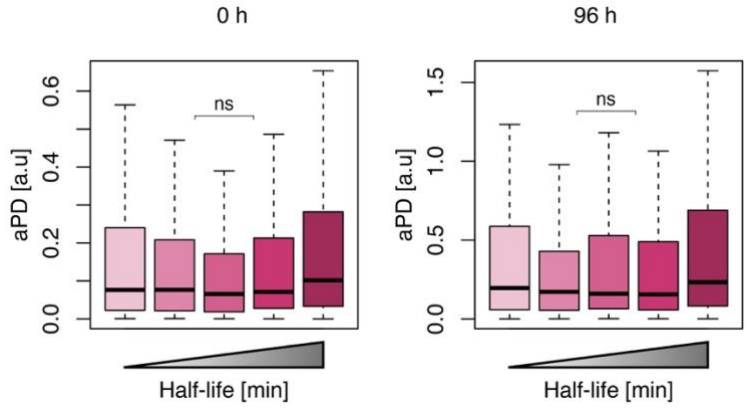


Figure 11. Half-lives of promoter-proximal Pol II demonstrate a poor correlation with estimated aPD.

Boxplots representing correlation of half-life quintiles sorted from lowest to highest and estimated aPD values correspondingly. The data is shown for 2,157 expressed paused protein-coding DE TUs.

5. Pre-B genes are presumably downregulated through the early termination mechanism rather than stable pausing

Next, we investigated a potential mechanism of the pre-B gene group downregulation throughout the time course of transdifferentiation (**Figure 3D**). Our prior data demonstrated an increase of aPD along with decrease in pause release factor P-TEFb levels towards the 96 h of transdifferentiation suggesting an accumulation of stable Pol II in the promoter-proximal region (**Figure 4** and **5**). We first plotted the total Pol II ChIP-nexus signal from the 6 and 30 min of DMSO/TRP treatment conditions for the pre-B gene group at both 0 and 96 h of transdifferentiation. We observed a similar global loss from the promoter-proximal region and further shift to the PIC position of the Pol II ChIP-nexus signal for both transdifferentiation time points as for all analyzed TUs (**Figure 10** and **12A**). We then calculated the half-lives of Pol II in the promoter-proximal region for pre-B genes, and, strikingly, could not detect significant differences at both time points of transdifferentation (median $FC_{0\text{-}96}$ = 1.02) (**Figure 12B**). The half-lives of promoter-proximal Pol II appeared to be relatively short for pre-B gene group in both cell states – during active release of Pol II into productive elongation resulting in high transcriptional activity at 0 h and during accumulation of Pol II in the pausing region along with downregulation of active transcription at 96 h (**Figure 4** and **12B**). Notably, the observed PWC was not significantly different for pre-B gene group at 0 h and 96 h of transdifferentiation indicating a different fate of Pol II in the pausing region at both time points (**Figure 4B**). Thus, we concluded that observed rapid loss of promoter-proximal Pol II can be a result of either actively released and elongating Pol II at 0 h of transdifferentiation or early transcription termination in the pausing region at 96 h of transdifferentiation. Our results are consistent with the recent findings demonstrating that transcription repression upon osmotic stress conditions in human cells is primarily regulated via increase in early termination rather than increase in pausing of promoter-proximal Pol II (Zimmer et al., 2021).

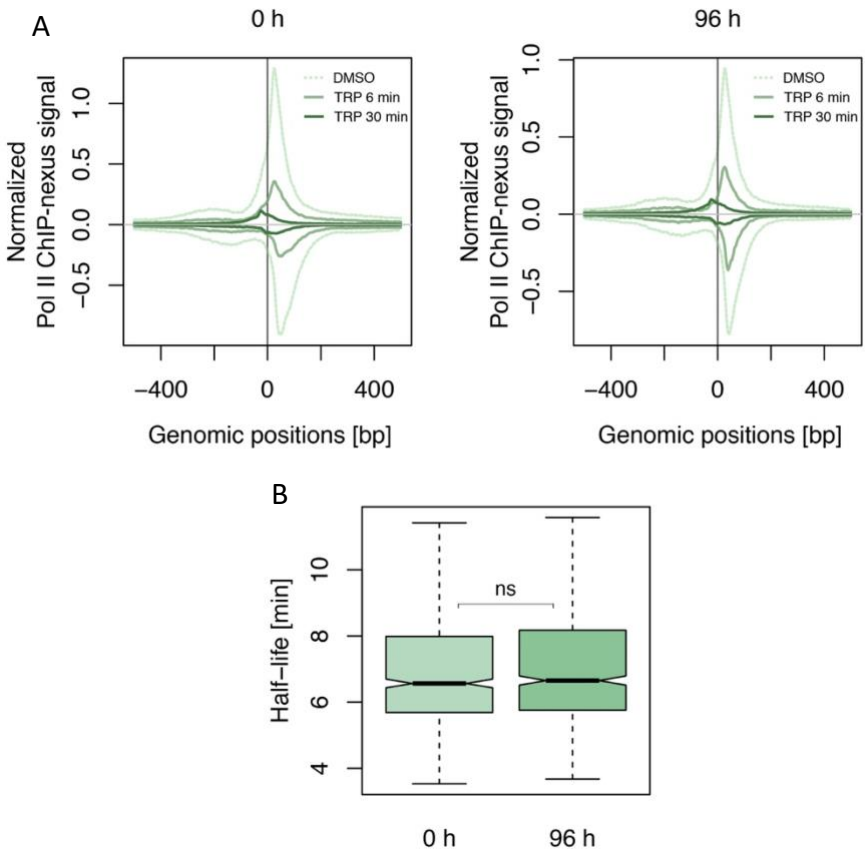


Figure 12. Pre-B genes exhibit similar half-lives of promoter-proximal Pol II independent on aPD. **A.** Metagene profiles for the ChIP-nexus signal under 6 and 30 min of DMSO or TRP treatment at 0 h and 96 h of transdifferentiation. The data is shown for 938 TUs (pre-B genes). **B.** Boxplots showing the half-lives calculations for 938 TUs (pre-B genes) at 0 and 96 h of transdifferentiation.

6. Early transcription termination contributes to distinct transcription kinetics of iMac genes in the macrophage-like cell state

We further focused on the Pol II stability in iMac gene groups that are upregulated towards the macrophage-like state of transdifferentiation (**Figure 3D**). We aimed to validate our previous hypothesis regarding distinct scenarios of promoter-proximal Pol II kinetics for iMac I and iMac II genes. Previously, we observed the increase of pIF for both gene groups throughout the transdifferentiation time course, however, iMac II genes exhibited aPD decrease and a clear pattern of pause release, while iMac I genes demonstrated a gain of promoter-proximal Pol II without proceeding into productive elongation at 96 h resulting in

aPD increase and suggesting potential early termination mechanism (**Figure 7** and **8**). The metagene analysis of the Pol II ChIP-nexus data showed similar pattern of rapid promoter-proximal Pol II reduction followed by further shift to the PIC region under the time series of TRP treatment for both iMac I and iMac II gene groups (**Figure 13A** and **13B**). We then estimated the half-lives of promoter-proximal Pol II in both gene groups at 0 and 96 h of transdifferentiation. Interestingly, we could not obtain significant differences in half-lives (median $FC_{0\text{-}96}$ = 0.96) at both transdifferentiation time points for iMac I genes indicating rapid turnover of promoter-proximal Pol II independent of the aPD increase (**Figure 13C** and **7A**). In contrast, we observed a significant decrease in promoter-proximal Pol II half-life for iMac II genes from 0 to 96 h of transdifferentiation (median $FC_{0\text{-}96}$ = 0.95, p-value = 5e-2). This result indicated increased stability of promoter-proximal Pol II at 0 h of transdifferentiation that reflected our aPD calculations for the iMac II gene group (**Figure 7A**). Additionally, the promoter-proximal Pol II stability decrease was caused by the active release into productive elongation at 96 h of transdifferentiation in case of iMac II gene group (**Figure 13C**, **7A** and **7B**). Moreover, we compared the Pol II stability in the promoter-proximal region for iMac I and iMac II groups at 96 h and could not observe significant differences (**Figure 13C**). We therefore confirmed our hypothesis regarding early transcription termination for iMac I gene group, since they displayed accumulation of promoter-proximal Pol II along with its rapid turnover without a strong increase of pIF as it was observed for the actively elongating iMac II gene group (**Figure 13C**, **7A** and **7B**).

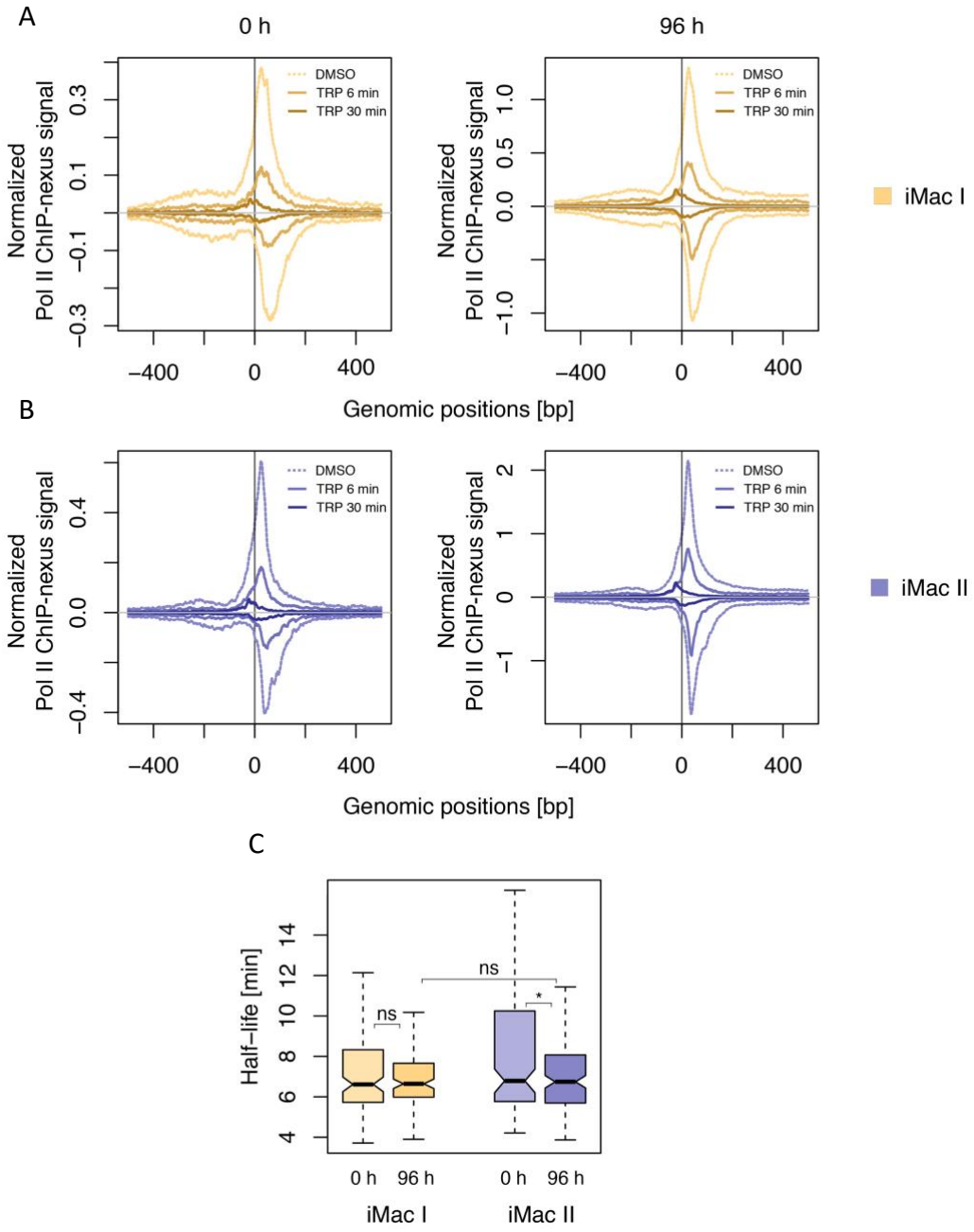


Figure 13. Distinct Pol II stability during iMac I and iMac II genes activation.

A and **B.** Metagene profiles for the ChIP-nexus signal under 6 and 30 min of DMSO or TRP treatment at 0 h and 96 h of transdifferentiation. The data is shown for 193 TUs (iMac I genes) (**A**) and 223 TUs (iMac II genes) (**B**). **C.** Boxplots showing the half-lives calculations for 193 TUs (iMac I genes) and 223 TUs (iMac II genes) at 0 and 96 h of transdifferentiation.

7. PROMOTER-PROXIMAL TERMINATION AND PAUSE RELEASE MODULATE TRANSCRIPTION KINETICS OF PRE-B AND IMAC GENE GROUPS

Finally, we sought to calculate total turnover rate of Pol II in the promoter-proximal region and the fraction of Pol II undergoing premature termination of transcription for our analyzed gene groups. In order to estimate total turnover rate of promoter-proximal Pol II, we used our half-life measurements and steady-state Pol II signal in the pausing region from the ChIP-nexus data (Supplementary methods). We additionally assumed that a total turnover of Pol II in the promoter-proximal region represents a total fraction of Pol II that is released from the PIC therefore providing a proxy for the actual initiation rate measurement. Notably, we obtained a significantly lower total Pol II turnover in the pausing region for the pre-B gene group at 96 h in comparison to 0 h of transdifferentiation that we explained by the general downregulation of transcriptional activity towards the macrophage-like state (median $FC_{0\text{-}96}$ = 0.89, p-value = 1e-4) (**Figure 14A** and **4**). Expectedly, we detected a significant increase in the total promoter-proximal Pol II turnover for iMac I (median $FC_{0\text{-}96}$ = 2.64, p-value = 1.3e-9) and iMac II (median $FC_{0\text{-}96}$ = 2.87, p-value = 1e-5) gene groups due to overall increase of pIF and gain of Pol II in the pausing region from 0 to 96 h of transdifferentiation (**Figure 14A** and **7**). We then estimated the relative fractions of Pol II that proceed into productive elongation and prematurely terminate by comparing the total Pol II turnover rate and pIF (Supplementary methods). We observed a significant increase in the fraction of prematurely terminating Pol II for the pre-B gene group from 0 to 96 h of transdifferentiation confirming our previous hypothesis (p-value = 2.2e-16) (**Figure 14B** and **12**). Although the total Pol II turnover was decreased towards 96 h of transdifferentiation for the pre-B gene group, its value appeared to be significantly higher than the total Pol II turnover for iMac gene groups at 0 h of transdifferentiaiton (p-value = 3e-5) (**Figure 14B**). These results indicate that downregulation of pre-B genes is primarily caused by increase in the promoter-proximal termination during the transdifferentiation. Furthermore, we did not detect significant changes in the premature termination fraction of the iMac I gene group comparing 0 and 96 h of transdifferentiaion (p-value = 0.16) (**Figure 14B**). Notably, premature termination fraction of iMac II gene group exhibited significant decrease from 0 to 96 h of transdifferentiation due to the active Pol II release into productive elongation in the macrophage-like state (p-value = 7.9e-5) (**Figure 14B** and **7**).

These observations suggest that Pol II pre-occupies promoter-proximal region of the iMac gene groups and undergoes constant cycles of initiation and early termination at 0 h of transdifferentiation. Upon the transdifferentiation induction, iMac I genes gain promoter-proximal Pol II that is not productively released into elongation and prematurely terminates to the same extent. In contrast, iMac II genes acquire transcriptional activity that is caused by increased pause release and corresponding decrease in premature termination towards 96 h of transdifferentiation. Moreover, we concluded that levels of pause release and pause-stabilizing factors are not dependent on the fate of promoter-proximal Pol II since they demonstrated similar quantities for iMac I and iMac II gene groups despite the differences in premature termination fractions (**Figure 8** and **14B**).

We additionally analyzed how the estimated Pol II premature termination fraction is related to the INTS11 occupancy. Interestingly, the lowest premature termination fraction values were correlated with the lowest INTS11 occupancy and vice versa, however, the intermediate premature termination fraction values did not display a pronounced correlation trend with INTS11 occupancy (**Figure S9**). This result suggests the alternative ways of regulating Integrator complex cleavage activity in order to attenuate Pol II transcription in the promoter-proximal region.

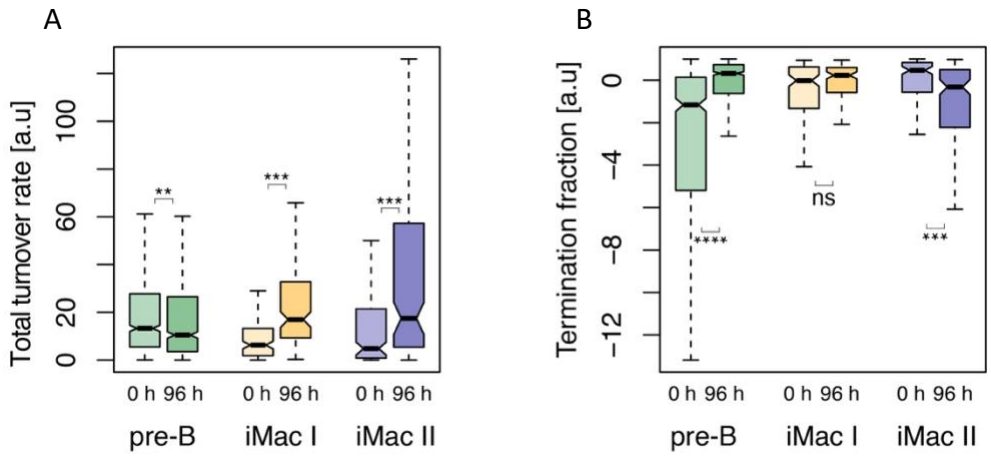


Figure 14. Premature transcription termination as a regulatory mechanism.
A. Boxplots representing comparison of the total Pol II turnover rate at 0 h and 96 h of transdifferentiation for 938 TUs (pre-B genes), 193 TUs (iMac I genes) and 223 TUs (iMac II genes). **B.** Boxplots representing comparison of the Pol II premature termination fraction at 0 h and 96 h of transdifferentiation for 938 TUs (pre-B genes), 193 TUs (iMac I genes) and 223 TUs (iMac II genes).

To summarize, we show that Pol II is highly dynamic in the pausing region undergoing constant cycles of initiation and early termination or promoter-proximal pause release rather than representing a stable complex. Our data suggests distinct transcription regulatory mechanisms for the downregulation of pre-B and upregulation of iMac genes during the transdifferentiation process. Whereas downregulation of pre-B gene expression appears to be primarily regulated via modulating premature termination, upregulation of iMac genes is governed by increase of the actual transcription initiation rate and the corresponding control of productive transcription by the different fractions of prematurely terminating Pol II for individual gene groups. In these regards, iMac II genes exhibit an increase of the actual initiation rate and decrease of the termination fraction of promoter-proximal Pol II allowing an efficient Pol II pause release into productive elongation towards the macrophage-like stage. In contrast, along with increase of the actual initiation rate iMac I genes demonstrate similar fraction of promoter-proximal Pol II termination at 0 h and 96 h of transdifferentiation resulting in a lesser fraction of the Pol II released into productive elongation in comparison to iMac II genes throughout the transdifferentiation. Taken

together, our observations shed light onto the role of early termination in transcriptional regulation during the cell type switching.

IV. DISCUSSION

In this chapter, I will focus on the novel aspects of our work and discuss the employed technical improvements that we applied in order to study transcription kinetics during cell fate transition. Furthermore, I will describe the potential limitations that can be encountered in our study and how to resolve them in the future. Finally, I will discuss the possible directions for the follow-up research in the context of our topic.

1. NOVEL FINDINGS OF OUR STUDY

In our research work we aimed to elucidate transcriptional regulatory mechanisms occurring during the cell fate transition process by using reprogrammable transdifferentiaiton system (Rapino et al., 2013) and various genome-wide sequencing techniques including TT-seq (Schwalb et al., 2016), mNET-seq (Nojima et al., 2015, 2016), ChIP-seq (Barski et al., 2007; Johnson et al., 2007) and ChIP-nexus combined with initiation inhibition (Q. He et al., 2015; Shao & Zeitlinger, 2017). In particular, we sought to estimate transcriptional kinetics with the focus on promoter-proximal Pol II regulation across the time course of human cell transdifferentiation. Uncovering these mechanisms is of prime importance since the cell type switching underlies processes such as cellular differentiation and organism development that are vital for the heathy functioning.

For the first time, we quantitatively estimated transcription kinetic parameters in the cell fate transition model system using previously established multi-omics approach comprising combination of the two nascent sequencing techniques TT-seq and mNET-seq (**Figure 2**) (Gressel et al., 2017, 2019). We obtained sufficient number of the genes for these calculations therefore being able to perform our statistical analysis with confidence (**Figure 3**). Furthermore, we showed the global changes in promoter-proximal Pol II kinetic parameters namely apparent pause duration (aPD) and productive initiation frequency (pIF) for the downregulated (pre-B) and upregulated (iMac I and iMac II) genes throughout the time course of transdifferentiation (**Figure 4**, **6** and **7**). Our initial results agreed with the previously published studies reporting changes in the promoter-proximal Pol II pausing at the target genes involved in development and differentiation in *Drosophila* as well as embryo development and somatic cell reprogramming in mammals using conventional ChIP-

seq experiments (Gaertner et al., 2012; B. Liu et al., 2020; L. Liu et al., 2014; Zeitlinger et al., 2007). Similar to these findings, we demonstrated accumulation of the transcriptionally engaged Pol II in the promoter-proximal region along with the signal decrease in the gene body for the downregulated genes during the transdifferentiation. We additionally complemented these observations with kinetic parameters estimations showing significant changes, specifically, the decrease in pIF and corresponding increase in aPD (**Figure 4**). For the upregulated genes, we could distinctly detect two kinetic behaviors using the clustering of the calculated parameters pIF and aPD. One of the groups (iMac II) exhibited expected reciprocal patterns wherein pIF was gradually increased and aPD was decreased accordingly (**Figure 6** and **7**). Notably, the other gene group (iMac I) displayed increase in both parameters. We interpreted this scenario with the pause-initiation limit concept wherein this gene group increase pIF slightly allowing for the simultaneous promoter-proximal Pol II gain leading to aPD increase (**Figure 7**) (Gressel et al., 2017, 2019).

Moreover, we assessed the occupancy of the pause release factor P-TEFb constituents during transdifferentiaiton and found that it reflects observed differences in aPD for the two analyzed gene groups (pre-B and iMac II) exhibiting the anticipated promoter-proximal regulation with reciprocal patterns of pIF and aPD (**Figure 5**, **7** and **8**). - Similar correlation was detected previously providing an evidence for Pol II pause release impairment upon U2 snRNP spliceosome component inhibition in human cells (Caizzi et al., 2021). Interestingly, the upregulated iMac I gene group, which increased both pIF and aPD parameters exhibited gain in the P-TEFb occupancy to the same extent as iMac II gene group raising a question regarding the mechanism of promoter-proximal regulation for each of these gene groups (**Figure 8**). We furthermore examined how paused Pol II stabilizing factor NELF occupancy is correlated with the estimated kinetic parameters throughout the transdifferentiation (**Figure 5**, **6** and **8**). Remarkably, we did not find evidence for its repressing function during the downregulation of the gene expression in contrast to the *in vitro* studies (Yamaguchi et al., 1999, 2002). We observed that NELF occupancy positively correlates with the transcriptional activity of our analyzed gene groups indicating its role in the proper establishment of Pol II pausing prior to release into productive elongation (**Figure 5**, **S4C** and **S4D**). These results were consistent with the *in vivo* findings, which demonstrated that NELF ablation leads to the overall decrease in Pol II signal, pause release defects and corresponding reduction of the mRNA levels (F. Chen et al., 2015; L. J. Core et

al., 2012; Gilchrist et al., 2008, 2010; Muse et al., 2007; Williams et al., 2015). Recent *in vivo* work additionally reported that Pol II promoter-proximal pausing can occur at the + 1 nucleosome in the absence of NELF, however, Pol II struggles to proceed into productive elongation at the target genes upon heat-shock conditions (Aoi et al., 2020). Similarly, we detected a positive correlation between the Integrator complex occupancy and transcriptional activity of the analyzed gene groups suggesting its functions in modulating premature termination levels during the active state (**Figure S7B** and **S7C**). Indeed, Integrator complex was shown to occupy the actively transcribed genes attenuating non-productive promoter-proximal Pol II and, thus, stimulating the release of the elongation-capable Pol II complexes (Beckedorff et al., 2020).

Furthermore, we explored the functional implications of our analyzed gene groups with distinct kinetic behaviors. Notably, downregulated during the transdifferentiaiton pre-B genes exhibited significant enrichment in biogenesis, metabolic, cellular and multi-organism processes since the initial transdifferentiation stage cells represent transcriptionally active cancer precursor leukemia B-cell line (**Figure 3**). A number of studies reported similar gene ontology results for the paused genes in various cancer cell lines, undifferentiated precursors and embryonic stem cells (ESCs) in mammals (Day et al., 2016; J. Liu et al., 2017; Mantsoki et al., 2018; Williams et al., 2015). On the other hand, iMac genes that are activated towards the 96 h of transdifferentiation displayed most enrichment of macrophage-specific functions including immune system process, stimulus response, biological adhesion and locomotion (**Figure 3** and **8**). These results agreed with the previous findings showing promoter-proximal Pol II pausing regulation of the proinflammatory genes upon the immune response stimuli in macrophage cells (Adelman et al., 2009; Escoubet-Lozach et al., 2011; Hargreaves et al., 2009; Yu et al., 2020). With the separate analysis of iMac gene groups, we detected enrichment of only immune functions in iMac I group that displayed accumulation of promoter-proximal Pol II without proceeding into productive elongation and corresponding aPD increase representing a classical example of the immune genes prepared for activation (**Figure 7** and **8**). Interestingly, iMac II genes that decreased aPD showing more pause release towards the macrophage stage exhibited generally higher enrichment in immune response and macrophage-specific functions along with signaling, cellular and biological processes functions (**Figure 8**). We also noted a higher enrichment of macrophage-specific signaling pathways within iMac II gene group in comparison to the

iMac I that demonstrated lesser number of signaling cascades that were enriched in innate immune response processes. We therefore propose that iMac II genes are more functionally important at the 96 h of transdifferentiation, whereas, iMac I gene expression is needed to be maintained at the basal level of transcription. We explain the observed productive elongation for the iMac II gene group by the general transcriptionally active cell state at 96 h prior to the transcriptional quiescence at 168 h post transdifferentiation induction (Rapino et al., 2013).

We finally introduced the measurements of promoter-proximal Pol II stability throughout the transdifferentiation time course in order to elucidate the potential mechanisms of the analyzed gene groups. In this regard, we applied time series of transcription initiation inhibition followed by high resolution ChIP-nexus technique at the 0 h and 96 h post transdifferentiaiton induction (**Figure 10**). This approach allowed resolving PIC-associated and promoter-proximal Pol II and was employed in *Drosophila* cells for the paused Pol II stability estimation (Q. He et al., 2015; Shao & Zeitlinger, 2017). For the first time in human cells, we could track ChIP-nexus Pol II signal under transcription initiation inhibition with triptolide (TRP) and provide an estimate of promoter-proximal Pol II stability. Notably, we showed a rapid loss of ChIP-nexus signal already after 6 min of TRP treatment and its complete shift to the PIC region after 30 min of TRP treatment. These observations resulted in a relatively short promoter-proximal Pol II half-life giving us a narrow distribution mostly within a range of 5-10 min across our analyzed gene groups (**Figure 10**). Our measurements were consistent with the previous estimates of promoter-proximal Pol II half-life utilizing conventional ChIP-seq in human (Erickson et al., 2018) and nascent RNA sequencing techniques including GRO-seq and STL-seq in mouse, human and *Drosophila* systems (Jonkers et al., 2014; Zimmer et al., 2021). Interestingly, Pol II half-lives did not exhibit a defined correlation with aPD indicating the importance of considering Pol II stability while calculating this parameter (**Figure 11**). Further analysis revealed similar half-lives of promoter-proximal Pol II for the pre-B gene group at 0 h and 96 h post transdifferentiation induction suggesting a high Pol II turnover even when productive transcription is downregulated and Pol II accumulates in the promoter-proximal region (**Figure 12** and **4**). This behavior can be achieved during the constant cycles of transcription initiation followed by early termination of Pol II in the promoter-proximal region. Next, we showed that the half-life of promoter-proximal Pol II is not changed for the iMac I gene

group, however, it displays a significant decrease for the iMac II gene group between 0 and 96 h of transdifferentiation (**Figure 13**). We explain these observations by the greater difference in aPD demonstrating corresponding significant decrease for the iMac II genes from 0 to 96 h of transdifferentiation (**Figure 7**). Notably, the promoter-proximal Pol II half-life appeared to be the same for iMac I and iMac II gene groups in the macrophage-like stage indicating a potential early termination mechanism for iMac I group based on the previously calculated pIF and aPD parameters (**Figure 13** and **7**). To perform calculations supporting our hypotheses, we introduced and estimated additional parameters such as the total turnover rate and the termination fraction of Pol II in the promoter-proximal region. In our model, we assumed that total Pol II turnover is a proxy for the actual initiation rate of transcription since all Pol II complexes released from the PIC enter the promoter-proximal pausing window wherein they can be either released into productive elongation or terminate prematurely. We observed a significant decrease in the total Pol II turnover rate for the pre-B gene group presumably due to the downregulation of their transcriptional activity towards the 96 h of transdifferentiation. Nevertheless, we detected a significantly higher premature termination fraction of Pol II for the pre-B gene group at 96 h of transdifferentiation suggesting a constant drop-off rather than an accumulation of the stably paused Pol II in the promoter-proximal region (**Figure 14** and **4**). Moreover, we demonstrated two separate mechanisms for the individual groups of the iMac genes. Expectedly, we obtained a significant increase of the total Pol II turnover rate for both gene groups causing the overall upregulation of the RNA synthesis from 0 to 96 h of transdifferentiation. We furthermore showed that premature termination fraction of Pol II remains the same for the iMac I gene group at 0 h and 96 h of transdifferentiation (**Figure 14**). In contrast, the premature termination fraction of Pol II was significantly decreased for the iMac II genes resulting in promoter-proximal Pol II release into productive transcription elongation towards 96 h of transdifferentiaiton (**Figure 14** and **7**). We additionally confirmed our initial hypothesis that iMac I gene group exhibited a higher fraction of premature termination within the 96 h cell stage leading to the observed lower levels of productive transcription in comparison to the iMac II gene group (**Figure 14**). Notably, our proposed mechanisms were partially consistent with the recent study using STL-seq allowing capturing dynamics of promoter-proximal Pol II associated short capped RNAs (scRNAs) turnover (Zimmer et al., 2021). They utilized a hyperosmotic stress conditions in human cells

and showed that caused repression of transcription is primarily mediated by premature termination resembling downregulation mechanism of our pre-B gene group during transdifferentiation (**Figure 4**, **12** and **14**) (Zimmer et al., 2021). Remarkably, this study additionally reported predominant regulation by the pause release at the target genes upon hormone stimulus leading to both activation and repression of transcription in *Drosophila* cells (Zimmer et al., 2021). In contrast, our iMac gene groups are presumably upregulated through the initiation rate increase along with modulation of the pause release via premature termination mechanism (**Figure 7**, **13** and **14**). These differences in mechanisms can occur due to the distinct biology of the two organisms or the stimulus nature, therefore, STL-seq performed in human cells using transcription activating stimuli would resolve this question.

Interestingly, our data demonstrated that occupancies of pause stabilizing factor NELF and pause release factor P-TEFb display similar levels for both iMac gene groups (**Figure 8**). We therefore suggest that recruitment of these factors is independent on the fate of promoter-proximal Pol II and an additional layer of their activity regulation might exist. These findings are supported with the previously published work proposing that accumulation of P-TEFb and NELF positively correlates with promoter-proximal Pol II density but not with Pol II gene body entrance indicating pause release regulation via these factors' activity rather than recruitment (Day et al., 2016). Moreover, the occupancy of cleavage subunit INTS11 of potential premature termination Integrator complex did not demonstrate a clear correlation trend with the Pol II premature termination fraction in general (**Figure S9**). We similarly propose that cleavage activity of the Integrator can be regulated by additional factors, however, it remains to be further investigated.

2. TECHNICAL ASPECTS OF OUR WORK

2.1. TECHNICAL IMPROVEMENTS

In the current study, we introduced several technical improvements in comparison to the previous findings regarding promoter-proximal Pol II regulation during the cell fate transition and promoter-proximal Pol II stability estimations in human cells. Prior reported methods of analyzing promoter-proximal Pol II pausing involved in the regulation of the cell

type switching mostly included Pol II occupancy profiling techniques (Day et al., 2016; Gaertner et al., 2012; Jonkers et al., 2014; L. Liu et al., 2014; Min et al., 2011; Muse et al., 2007; Williams et al., 2015; Zeitlinger et al., 2007) or separate measurements of TSS and gene body associated nascent RNA levels (Larke et al., 2021). However, these data alone do not determine transcriptional kinetics (Introduction) (Ehrensberger et al., 2013; Gressel et al., 2017, 2019). We therefore for the first time applied the novel multi-omics approach providing a quantitative estimation of the transcriptional kinetic parameters (Gressel et al., 2017, 2019) using human cell transdifferentiation as a model system for the cell fate transition. Apart from mNET-seq, alternative techniques detecting occupancy of the transcriptionally engaged Pol II such as PRO-seq can be employed in multi-omics approach (Jaeger et al., 2020). PRO-seq appears to have an advantage due to the lower amount of cell input for the experiment and possibility of the spike-ins usage for the further normalization. However, it should be considered that PRO-seq does not capture the backtracked promoter-proximal Pol II complex positions in comparison to the mNET-seq due to the technical specifications (Kwak et al., 2013; Nojima et al., 2015, 2016).

As mentioned before, we utilized the high-resolution ChIP-nexus method combined with the time series of transcription initiation inhibition with TRP for the promoter-proximal Pol II stability measurement in human system that was previously performed only in Drosophila cells (Shao & Zeitlinger, 2017). For this purpose, ChIP-nexus is well-suited since it can monitor separate signal from PIC and pause region associated Pol II in comparison to the conventional ChIP-seq technique that can introduce biases due to capturing Pol II signal from both complexes (Q. He et al., 2015; Johnson et al., 2007). Interestingly, application of the standard ChIP-seq protocol with TRP treatment in human cells revealed a stably paused genes that kept Pol II ChIP-seq signal after the treatment (F. Chen et al., 2015). Notably, we did not detect any stable Pol II pausing pattern in our results (Results, section 4). Alternative reason for stably paused Pol II detection could be incomplete transcription initiation inhibition due to the low TRP concentration usage or prolonged treatment time (F. Chen et al., 2015). Therefore, utilization of the higher TRP concentrations and shorter treatment time would be beneficial. Considering this point, our results were consistent with the prior studies using similar treatment conditions (Results, section 4) (Erickson et al., 2018).

2.2. TECHNICAL LIMITATIONS

As described in the previous sections, we employed transdifferentiation system of human precursor leukemia B-cells into macrophage-like cells based on the estrogen-inducible C/EBPα overexpression (Rapino et al., 2013). This system is well-suited to study cell fate transition due to its simplicity and high efficiency (Results, section 1). Nevertheless, comparing our results to promoter-proximal Pol II regulation of the essential process occurring in the living organism such as monocyte-to-macrophage differentiation during early inflammation would provide a biological relevance (Italiani et al., 2014; J. Yang et al., 2014). Additionally, investigating the transcriptional Pol II kinetics of the genes from the previously published dataset on transcriptional profiling during differentiation of human primary monocytes into macrophages in our system would be beneficial (Martinez et al., 2006). As an alternative, the recently obtained mRNA sequencing data from the differentiation of THP-1 human monocytic leukemia cell line into macrophage-like cells can be overlapped and compared with our analyzed gene groups (Green et al., 2020). However, the above-mentioned *in vitro* differentiation systems appear to have limitations for studying populations of the specialized tissue-resident macrophages. In this respect, isolation of the primary monocytes from peripheral blood mononuclear cells (PBMCs) is typically time-consuming and results in a low yield. Macrophage cell lines including human THP-1 or mouse RAW264.7 can solve these issues; however, their usage cannot fully recapitulate differentiation *in vivo* (Chanput et al., 2014; C. Z. W. Lee et al., 2018). Notably, recently developed approach wherein macrophages are first derived from iPSCs and then cultured in the conditions corresponding to the certain tissue type appears to be the most favorable for reproducing tissue-specific *in vivo* environment (C. Z. W. Lee et al., 2018; S. Li et al., 2022) and can be considered for the further studies.

Additional technical limitation can be caused with the major isoform annotation required for the following kinetic parameters calculation (Supplementary methods, Results section 1). Since only TUs with either one RNA isoform or with one major RNA isoform that was prevailing more than 70% for the gene were kept (Supplementary methods), we filtered out a substantial number of the key genes involved in the B-cell or macrophage specific functions. Here, defining the TSS positions with scRNA-seq or GRO-cap techniques in order to assess the correct RNA isoforms and, therefore, extend our analysis to the missing gene sets can be considered for the future studies (L. J. Core et al., 2014; Larke et al., 2021). For

the follow-up research, the full-length RNA isoforms can be identified with the long-read sequencing technologies such as Pacific Biosciences (PacBio) or Oxford Nanopore Technologies (ONT) platforms (Au et al., 2013; Garalde et al., 2018; Soneson et al., 2019; Workman et al., 2019). Importantly, stability of expressed RNA isoforms in the cells can be additionally examined with the recently introduced nano-ID method (Maier et al., 2020).

Another essential aspect that can be noted as a limitation is the usage of TRP compound blocking transcription initiation for our promoter-proximal Pol II stability measurements (Results, section 4). A number of studies demonstrated TRP as a potential therapeutic drug since it possesses anti-cancer, anti-inflammatory and immunosuppressive properties (S. R. Chen et al., 2018). Particularly, cancer-killing properties include cell growth and proliferation defects, transcriptional kinases-mediated proteasomal degradation of Pol II, global inhibition of RNA synthesis and DNA damage (Manzo et al., 2012; Wang et al., 2011). Taken together, these features might interfere our results with the prolonged TRP treatment times since we utilized cancer-derived BLaER1 cell line. Although we optimized treatment conditions and verified the absence of Pol II degradation, rapid protein depletion techniques allowing targeted degradation without potential off-target effects can be further considered (Jaeger & Winter, 2021; Mayor-Ruiz & Winter, 2019).

3. FUTURE PERSPECTIVES

In the current work, we suggested several mechanisms of promoter-proximal Pol II regulation wherein early termination of Pol II transcription appears to have an essential role. As described before, we utilized the occupancy-based technique ChIP-nexus of Pol II with the time series of TRP treatment in order to estimate promoter-proximal Pol II stability. Importantly, additional confirmation of Pol II transcriptional activity such as detection of associated nascent RNA would complement these findings. Here, method capturing scRNA associated with promoter-proximal Pol II and mapping the transcript TSSs called Start-seq can be employed (Nechaev et al., 2010). Previous studies reported that Start-seq signal is co-localized with Pol II ChIP-seq signal allowing estimating steady-state transcriptionally engaged Pol II in the pausing region (Henriques et al., 2013). Later on, Start-seq was used in combination with TRP treatment to estimate the half-lives of promoter-proximal Pol II at coding and non-coding transcripts (Henriques et al., 2018). Since the above-mentioned

reports applied Start-seq in *Drosophila* cells, consequently, recent work published a modified version of this protocol that can be performed in human cells (Larke et al., 2021). Having said that, alternative confirmation of our Pol II half-life measurements on the RNA level would support our findings. Moreover, the novel approach STL-seq was demonstrated to measure the dynamics of promoter-proximal transcript turnover estimating the amount of prematurely terminated transcripts without perturbation and can be also applied in our system in the future (Zimmer et al., 2021). Furthermore, one of the most important follow-up questions would be which key factors are responsible for the modulation of early termination at our distinctly regulated gene groups and how this occurs mechanistically. Importantly, some of the termination and 3' end RNA processing factors occupancy was detected in the beginning of the genes suggesting their role in premature termination (Brannan et al., 2012; Kamieniarz-Gdula et al., 2019; Kamieniarz-Gdula & Proudfoot, 2019; G. Martin et al., 2012; Nojima et al., 2015; Wagschal et al., 2012). Moreover, increased level of nascent RNA transcription in the promoter-proximal region observed upon loss of the termination factors PCF11 and XRN2 as well as CPA machinery components provided more evidence for this hypothesis (Cortazar et al., 2022; Davidson et al., 2019; Kamieniarz-Gdula et al., 2019; Lykke-Andersen et al., 2021; Nojima et al., 2015). Additionally, ablation of the subunit of exoribonucleolytic exosome complex revealed increase in the nascent RNA levels in the TSS proximity of different transcript types suggesting the early terminated RNA degradation (Lykke-Andersen et al., 2021). Taken together, levels of these factors can be assessed and compared between our distinctly regulated gene groups with occupancy-based techniques. Also, evidence of premature termination can be provided using detection of termination signatures of Pol II CTD that was shown to appear in the promoter-proximal region previously (Kamieniarz-Gdula et al., 2019; Schlackow et al., 2017). However, ChIP-seq and related techniques can be limited with availability and specificity of the antibodies against these factors. Despite this, the new protein candidates involved in regulatory functions cannot be detected with these methods. In this respect, the novel approach including proximity-labeling followed by mass spectrometry (MS) was described in the recent years (Ummethum & Hamperl, 2020). Briefly, the procedure involves a covalent binding of a biotin tag to the proteins surrounding the target of interest with further MS analysis of the labeled fraction. The advantage of this method appears to be the detection of the protein interactions in the native conditions since the labeling occurs inside the cell

(Shkel et al., 2022; Ummethum & Hamperl, 2020). Remarkably, the proximity-labeling with MS can be additionally performed employing proximity-labeling enzyme APEX/APEX2 fused to the dead Cas9 mutant (dCas9) associated with the single guide RNA (sgRNA) targeting the complex to the specific DNA loci (X. D. Gao et al., 2018; Ummethum & Hamperl, 2020). This approach was further improved to examine non-repetitive single loci (Myers et al., 2018). Having said that, this novel technique can be utilized in the described transdifferentiation system to analyze the protein interaction network in the promoter-proximal region of the separate gene candidates from our defined groups. The obtained proteomics results would indicate potential factors that might be involved in the observed distinct promoter-proximal regulation behaviors. Here, the novel regulators or the factors regulating activity of the observed in our work pause-related and premature termination protein complexes can be discovered. Following these results, the functional studies of the identified candidates can be further performed with the advanced and specific rapid protein depletion methods (Jaeger & Winter, 2021; Mayor-Ruiz & Winter, 2019).

Another important aspect that can be pursued in the follow-up research is studying the transcriptional kinetics during cell fate transition on the single-cell level. A variety of prior reports demonstrated evidences of transcription as a stochastic and discontinuous process implying that it occurs in bursts (Raj & van Oudenaarden, 2008; Tunnacliffe & Chubb, 2020). Interestingly, visualization of the nascent transcription of α-globin locus revealed distinct transcriptional bursting patterns during erythroid differentiation in mouse cells (Jeziorska et al., 2022). Apart from the fluorescent reporter-based labeling of nascent RNA and single-molecule RNA FISH (smFISH), several techniques assessing transcriptional bursting genome-wide such as single-cell RNA-seq were introduced (Tunnacliffe & Chubb, 2020). For instance, detection of the spliced and unspliced transcript levels in single cells can predict cell fate decisions (La Manno et al., 2018) or application of allelic single-cell RNA-seq can determine transcriptional bursts parameters such as burst size and frequency and their regulation by core promoters and enhancer elements correspondingly (Larsson et al., 2019). Furthermore, the dynamics of transcriptional bursting kinetics can be extracted using metabolic labeling of the newly synthesized RNA in single cells (Erhard et al., 2019; Hendriks et al., 2019). Therefore, one of the future directions would be studying the transdifferentiation on the single-cell level with the above-mentioned methods in order to

elucidate how the transcriptional bursting parameters are related to our observed kinetics changes.

In our work, we mostly focused on the protein-coding transcripts and their kinetics during transdifferentiation time course. Previous study in our laboratory utilized the same system and demonstrated global changes in the eRNA synthesis of the putative enhancer elements paired with the target genes involved in the transdifferentiation. Moreover, potential different modes of enhancer action including additive and synergistic behaviors were suggested (Choi et al., 2021). For the follow-up work, transcriptional kinetics analysis of the annotated eRNAs putatively paired with our defined gene groups can be considered. Notably, we performed the initial analysis in this direction and observed possible differences in the kinetic parameters of the gene groups paired with or without putative enhancers. However, these results are preliminary and require verification by employing CRISPR/Cas9-mediated enhancer(s) deletion for representative candidates followed by multi-omics analysis throughout transdifferentiation. In this regard, a few studies suggested that enhancer elements can facilitate promoter-proximal Pol II pause release at the specific DNA loci (W. wei Gao et al., 2018; W. Liu et al., 2013). Interestingly, eRNAs themselves showed the ability to bind pause-stabilizing NELF factor and promote its dissociation from the paused Pol II complex followed by the pause release (Gorbovytska et al., 2022; Schaukowitch et al., 2014). In contrast, recent work reported that enhancer-mediated target gene regulation prevails at the transcription initiation level rather than promoter-proximal Pol II pausing at the α- and ß-globin loci during erythroid differentiation in mouse cells (Larke et al., 2021). Taken together, these findings motivate to further elucidate enhancer or eRNA functions with respect to our observed transcriptional kinetics changes across the transdifferentiation. Another important point here is the general analysis of promoter-proximal Pol II stability for enhancers and other non-coding transcribed regions. Importantly, Pol II stability in the pausing region was previously estimated for enhancer elements and appeared to be lower in comparison to protein-coding genes based on the estimations from the Start-seq combined with TRP treatment in *Drosophila* (Henriques et al., 2018). Therefore, it would be advantageous to assess promoter-proximal Pol II stability for enhancers and other non-coding loci in human cells using our collected ChIP-nexus data with TRP treatment time series. Moreover, obtained results can be related to the corresponding analyzed gene groups at the different time points of transdifferentiaiton.

Finally, our results enlighten an importance of transcriptional kinetics regulation in order to modulate expression of the target genes involved in the transdifferentiation process. Therefore, as discussed before, transferring our experimental approach in order to study kinetics of promoter-proximal Pol II regulation in the biologically relevant system would be important for the future research. The closest process would be monocyte-to-macrophage differentiation occurring during early inflammation that was extensively studied in the past years (Chanput et al., 2014; Green et al., 2020; Italiani et al., 2014; C. Z. W. Lee et al., 2018; S. Li et al., 2022; Martinez et al., 2006; J. Yang et al., 2014). Since, different types of macrophages encompass both cancer progression and tumor killing properties, further investigation in this direction would be of a prime importance for future medical applications and anti-cancer drug development (Mantovani et al., 2022).

V. SUPPLEMENTARY INFORMATION

1. SUPPLEMENTARY FIGURES

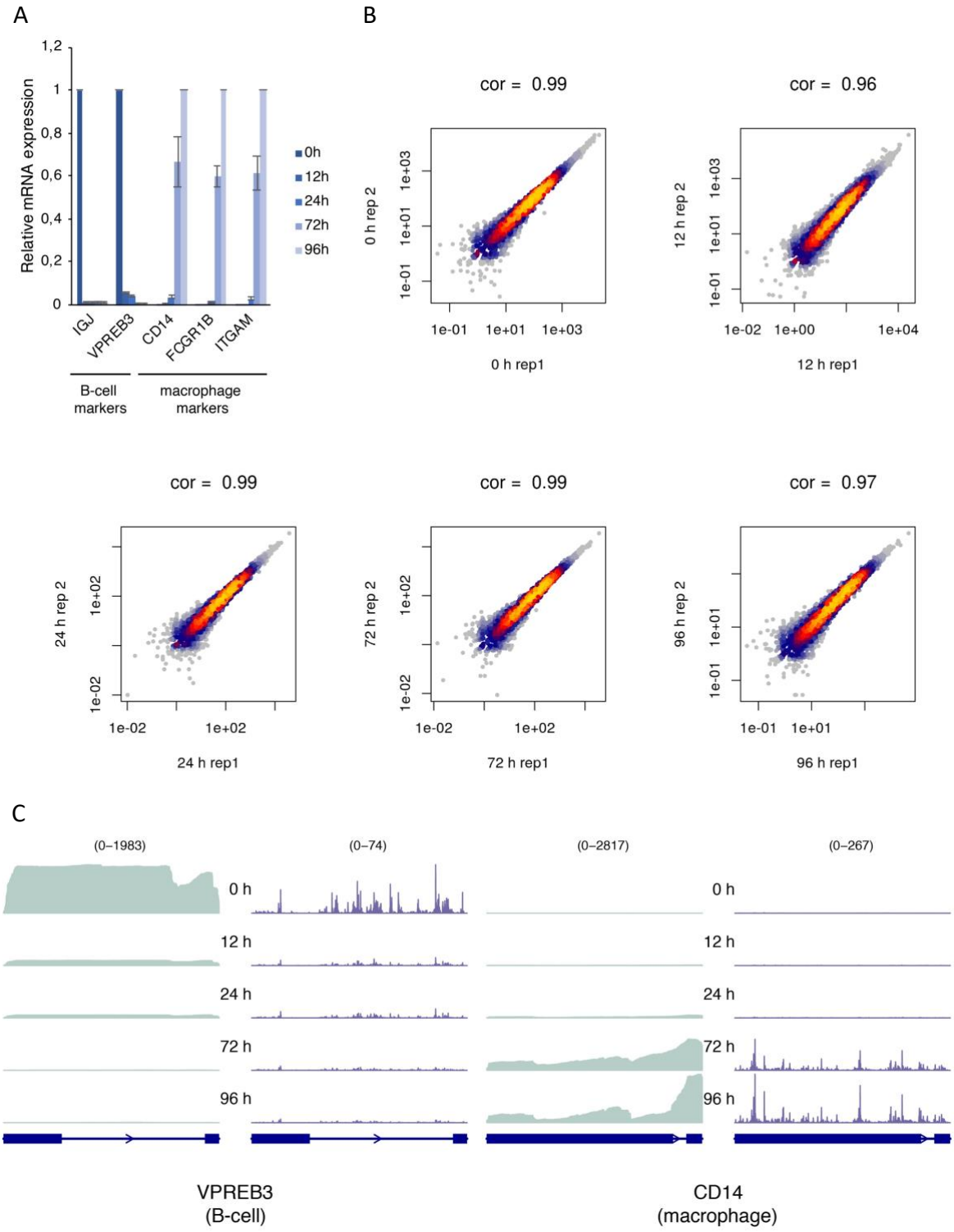
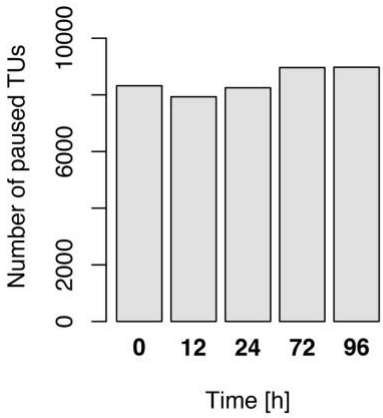


Figure S1. (legend is on the next page)

Figure S1. Validation of the transdifferentiation system.

A. RT-qPCR results showing the relative mRNA expression of the B-cell and macrophage markers throughout the transdifferentiation time course (0, 12, 24, 72 and 96 h post induction). The signal was normalized with GAPDH expression. The error bars represent standard deviation of the data for two independent biological replicates. **B.** Pearson correlation between two independent biological replicates of the mNET-seq data collected at 0, 12, 24, 72 and 96 h post transdifferentiation induction. **C.** Representative examples of the B-cell (VPREB3) and macrophage (CD14) markers TT-seq and mNET-seq data collected at 0, 12, 24, 72 and 96 h post transdifferentiation induction. Each dataset is averaged for two independent biological replicates.

A



B

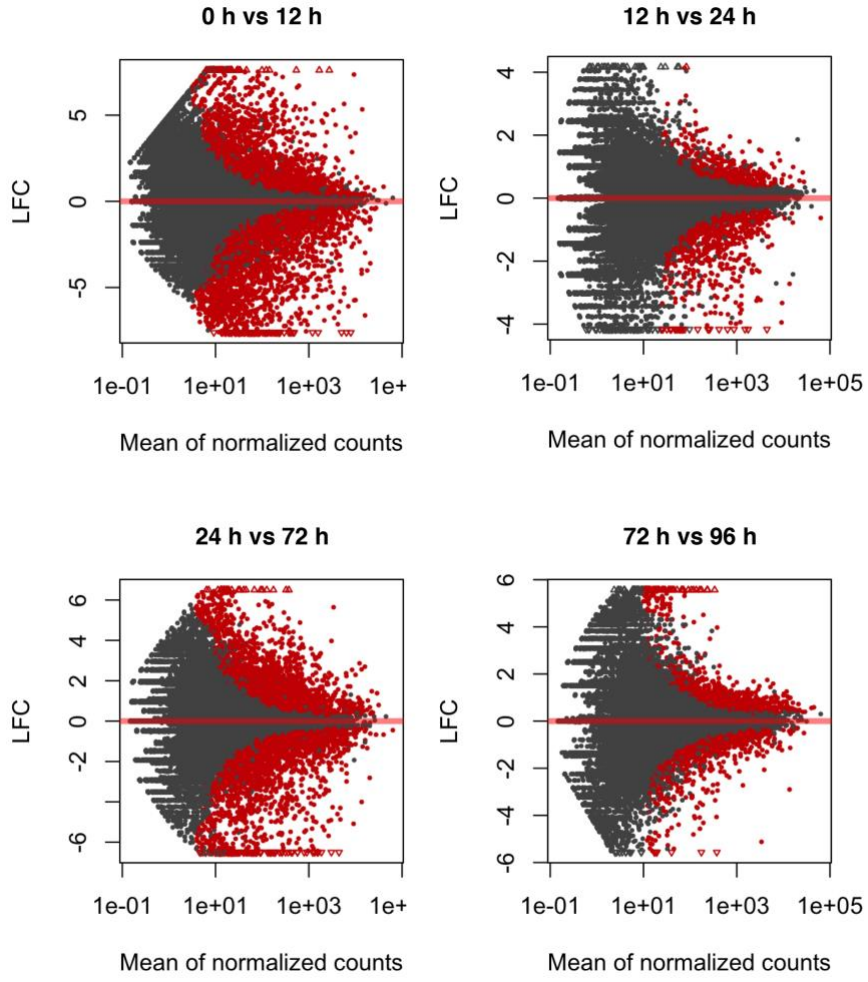


Figure S2. (legend is on the next page)

C

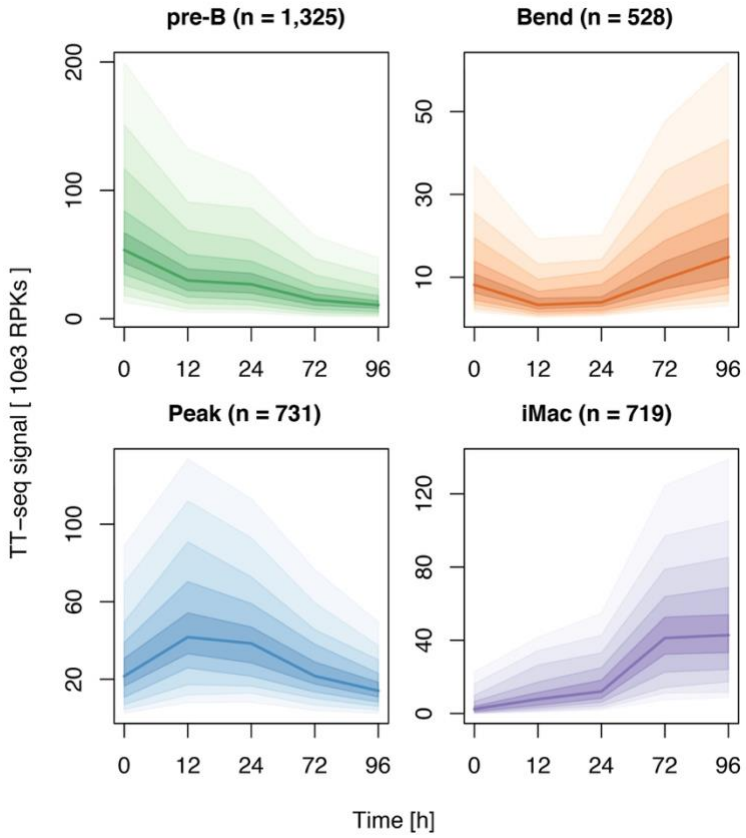


Figure S2. Distinct patterns of the differentially expressed paused TUs throughout transdifferentiation time course.

A. Number of annotated TUs with the called paused Pol II peak across the time course (0, 12, 24, 72, 96 h) of transdifferentiation. **B**. Log fold change of the TT-seq expression upon transdifferentiation induction versus mean of normalized counts. Significantly up- or downregulated TUs are colored in red (padj < 0.05). **C.** Representation of the clustering of 3,303 DE TUs changed during transdifferentiation time course based on the TT-seq. Expression changes occurred in 4 defined patterns named pre-B (downregulated), bending, peaking and iMac (upregulated) correspondingly.

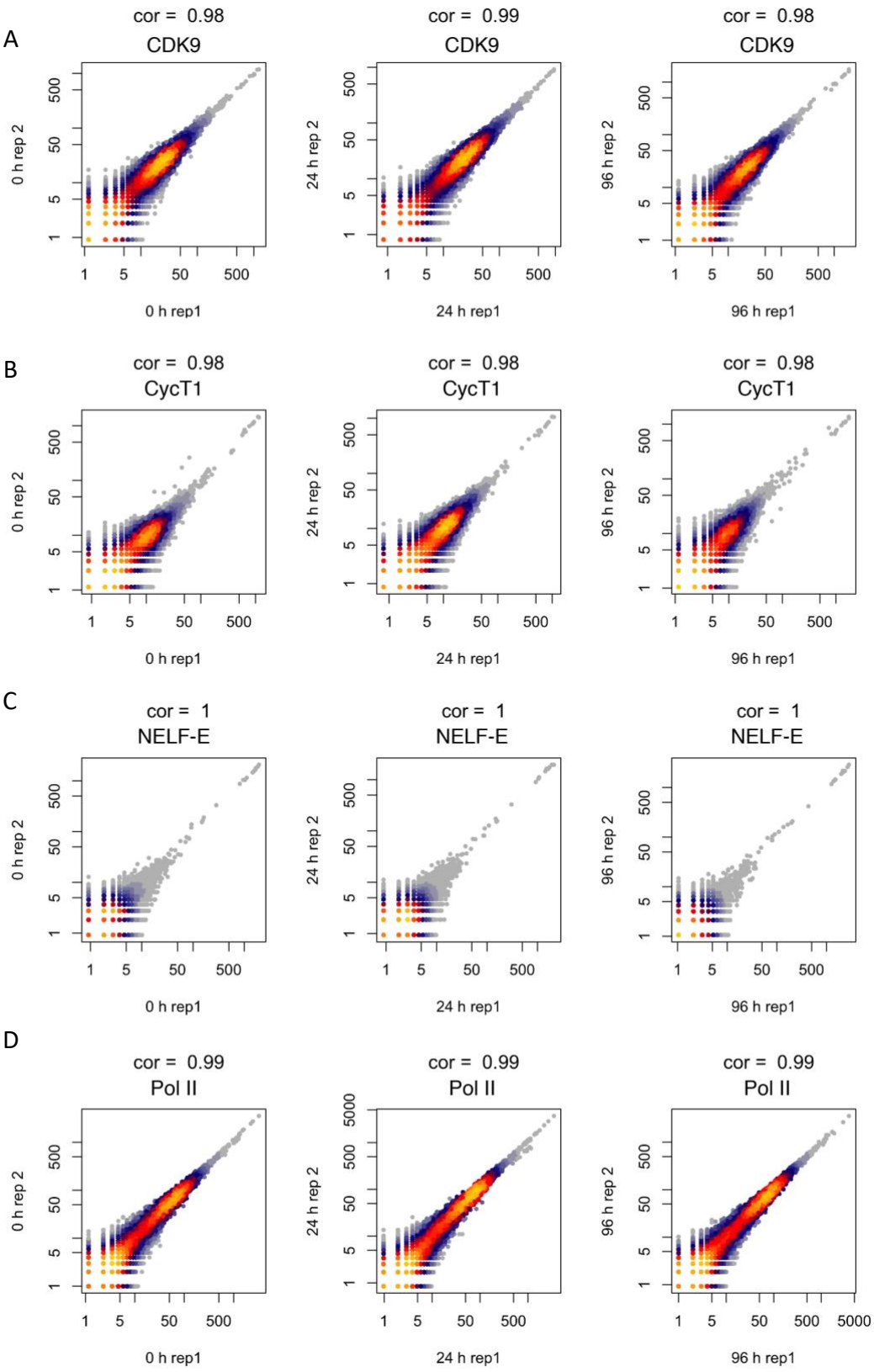


Figure S3. (legend is on the next page)

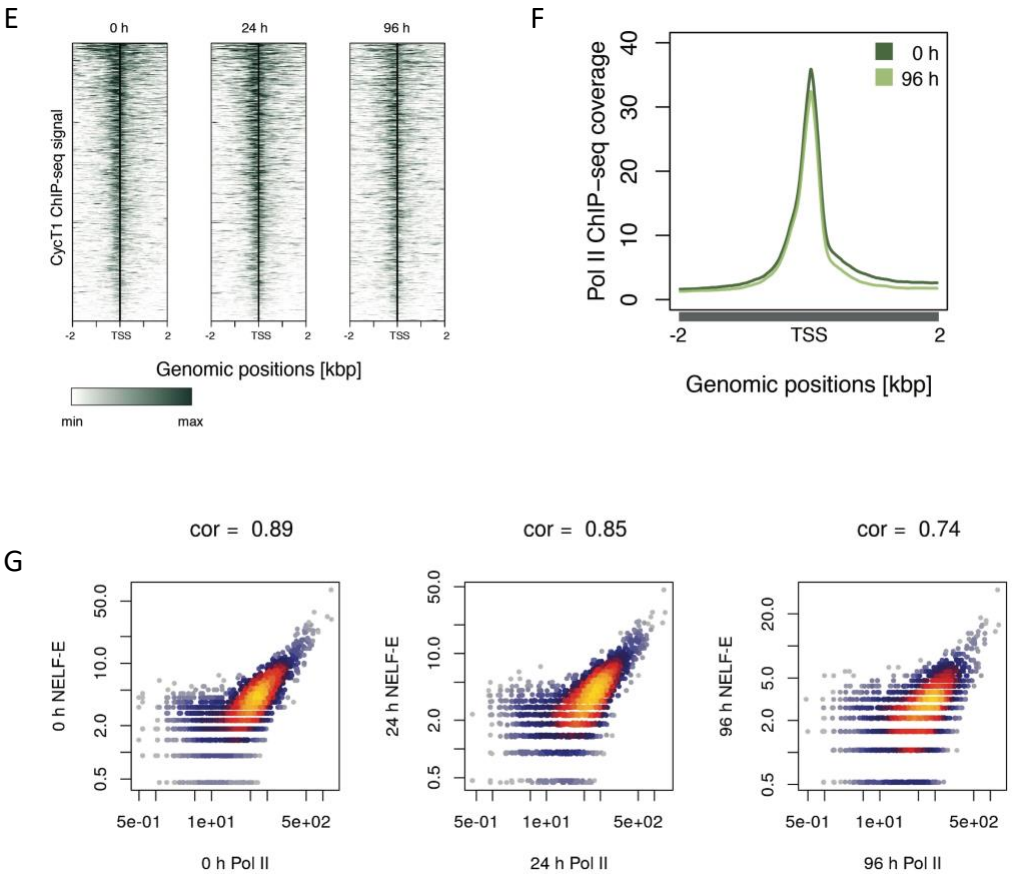


Figure S3. Occupancy of the pausing-related factors follows the activity of pre-B gene group during transdifferentiation.

A, B, C and **D.** Pearson correlation between two independent biological replicates of the CDK9, cyclin T1, NELF-E and Pol II ChIP-seq data collected at 0, 24 and 96 h post transdifferentiation induction. **E.** Heatmaps displaying cyclin T1 ChIP-seq signal ranked by intensity at 96 h for 938 TUs (pre-B genes) in the range of +/-2 kbp and quantified in the +/-1 kbp region. The data is shown for 0, 24 and 96 h of transdifferentiation. **F.** Metagene profiles comparing Pol II ChIP-seq signal between 0 h (dark green) and 96 h (light green) post transdifferentiation induction for 938 TUs (pre-B genes). ChIP-seq signal was averaged for two biological replicates and displayed in the range of -2 to 2 kbp with respect to the TSS. **G.** Pearson correlation between NELF-E and Pol II ChIP-seq data at 0, 24 and 96 h of transdifferentiation. Each dataset is averaged for two independent biological replicates.

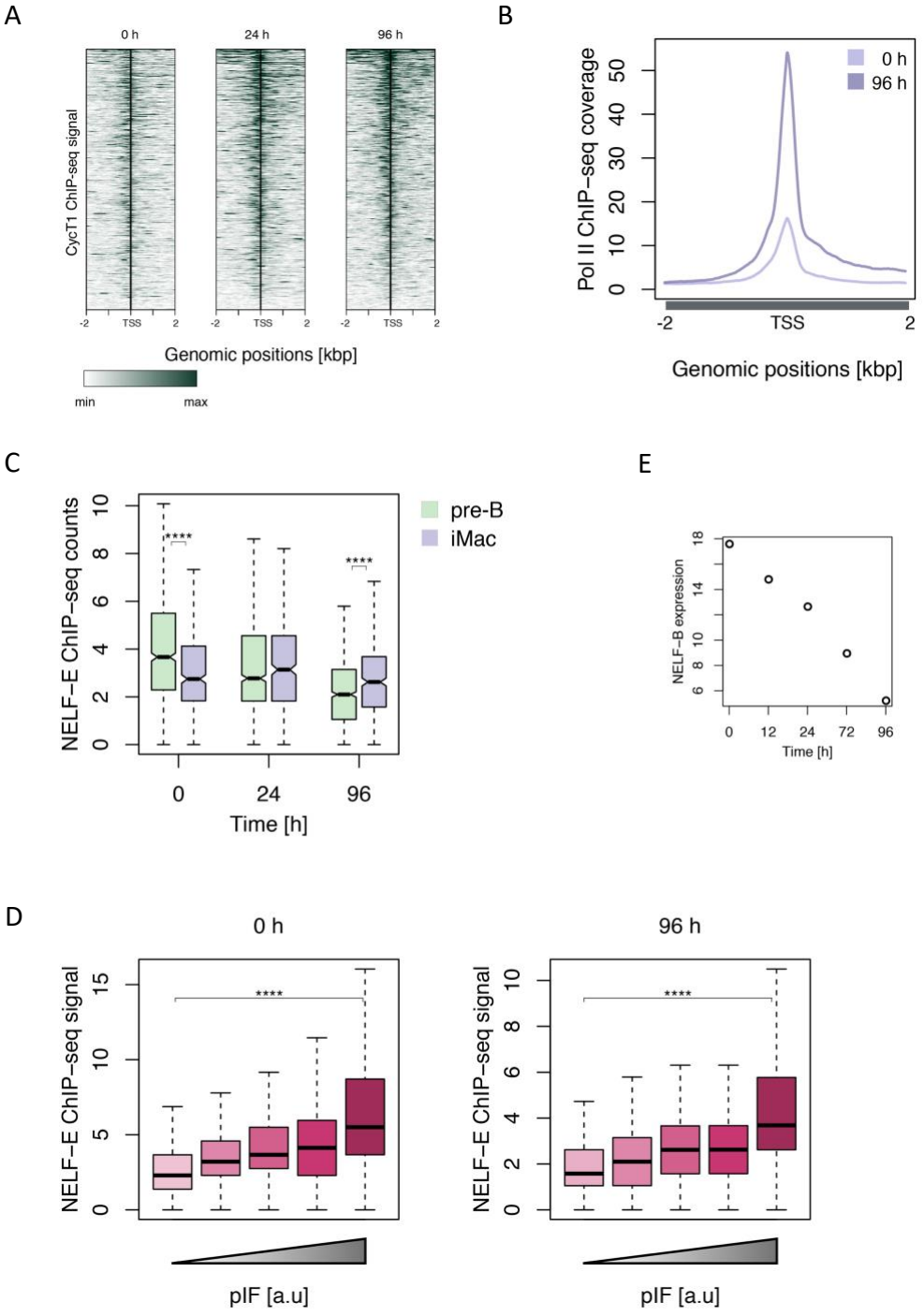


Figure S4. Occupancy of the pausing-related factors correlates with the transcriptional activity during transdifferentiation.

A. Heatmaps displaying cyclin T1 ChIP-seq signal ranked by intensity at 0 h for 416 TUs (iMac genes) in the range of +/-2 kbp and quantified in the +/-1 kbp region. The data is shown for 0, 24 and 96 h of transdifferentiation. **B.** Metagene profiles comparing Pol II ChIP-seq signal between 0 h (light purple) and 96 h (dark purple) post transdifferentiation induction for 416 TUs (iMac genes). ChIP-seq signal was averaged for two biological replicates and displayed in the range of -2 to 2 kbp with respect to the TSS. **C.** Boxplots representing NELF-E ChIP-seq signal comparison of pre-B (938 TUs) and iMac (416 TUs) gene groups at 0, 24 and 96 h post transdifferentiation induction. **D.** Boxplots

showing NELF-E ChIP-seq signal plotted versus pIF groups ranked from lowest to highest for 2,157 analyzed TUs. The data is shown for 0 and 96 h post transdifferentiation induction. **E.** NELF-B expression based on the TT-seq data across the time course of transdifferentiation (0, 12, 24, 72 and 96 h).

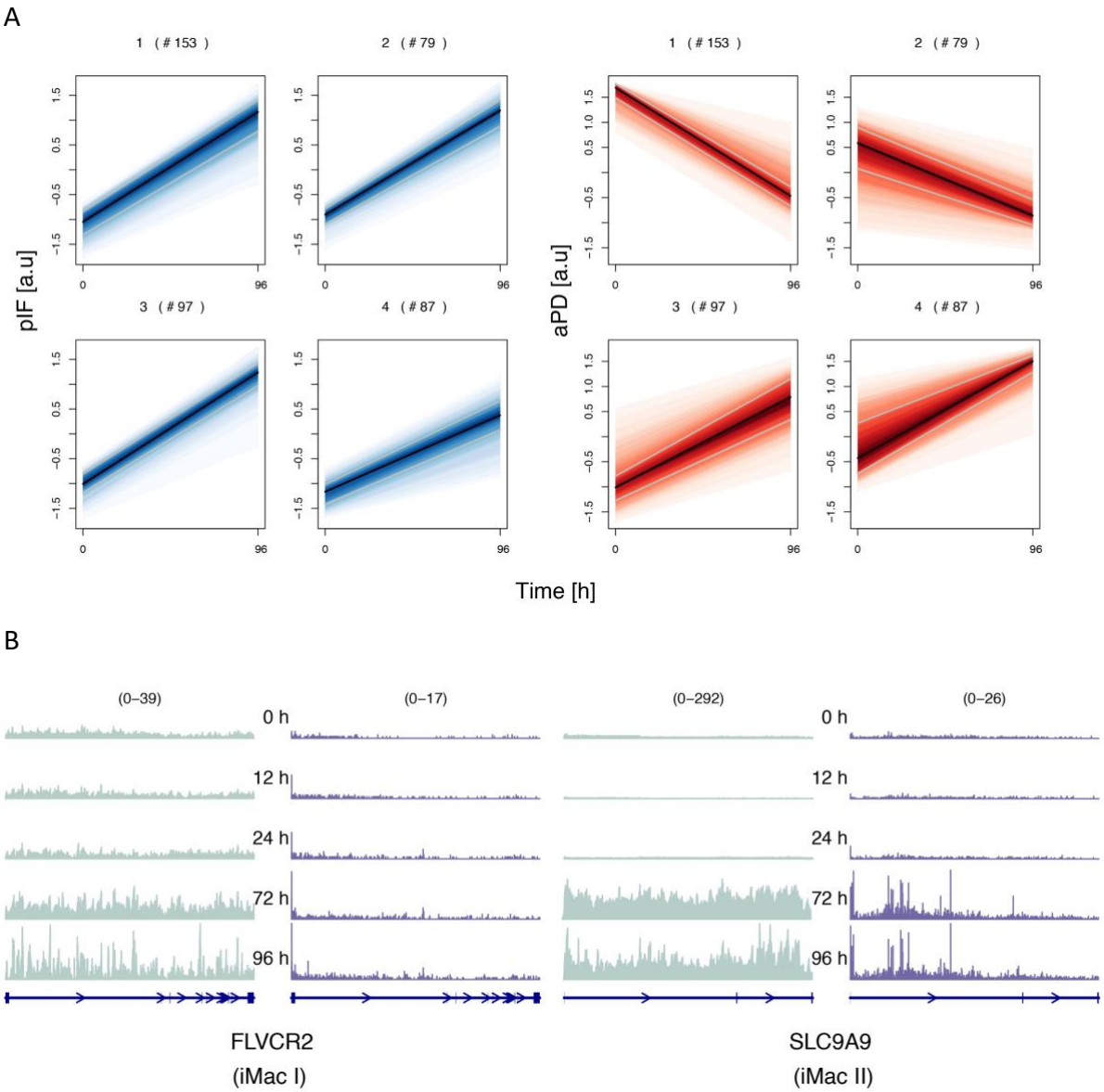


Figure S5. iMac gene groups demonstrate two distinct kinetics scenarios.

A. Representation of the k-means (n = 4) clustering of the kinetic parameters pIF and aPD for the iMac gene group (416 TUs). **B.** Representative examples of the iMac I (FLVCR2) and iMac II (SLC9A9) genes TT-seq and mNET-seq data collected at 0, 12, 24, 72 and 96 h post transdifferentiation induction. Each dataset is averaged for two independent biological replicates.

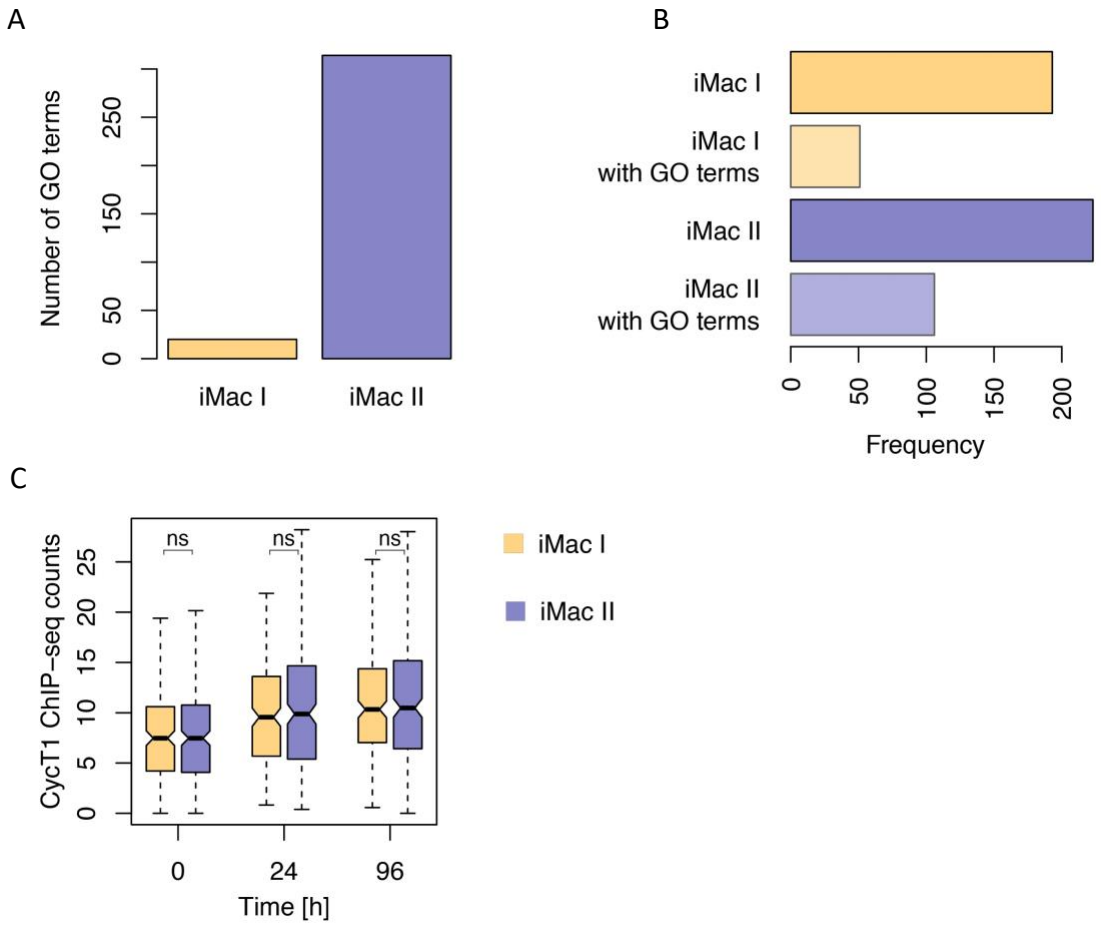


Figure S6. iMac gene groups differ in biological functions but not in the pausing-related factors occupancy.

A. Comparison of the gene ontology terms numbers between iMac I (193 TUs) and iMac II (223 TUs) gene groups. **B.** Representation of the iMac I and iMac II gene fractions enriched in gene ontology terms. **C.** Boxplots displaying comparison of the cyclin T1 ChIP-seq signal for the iMac I and iMac II genes at 0, 24 and 96 h post transdifferentiation induction.

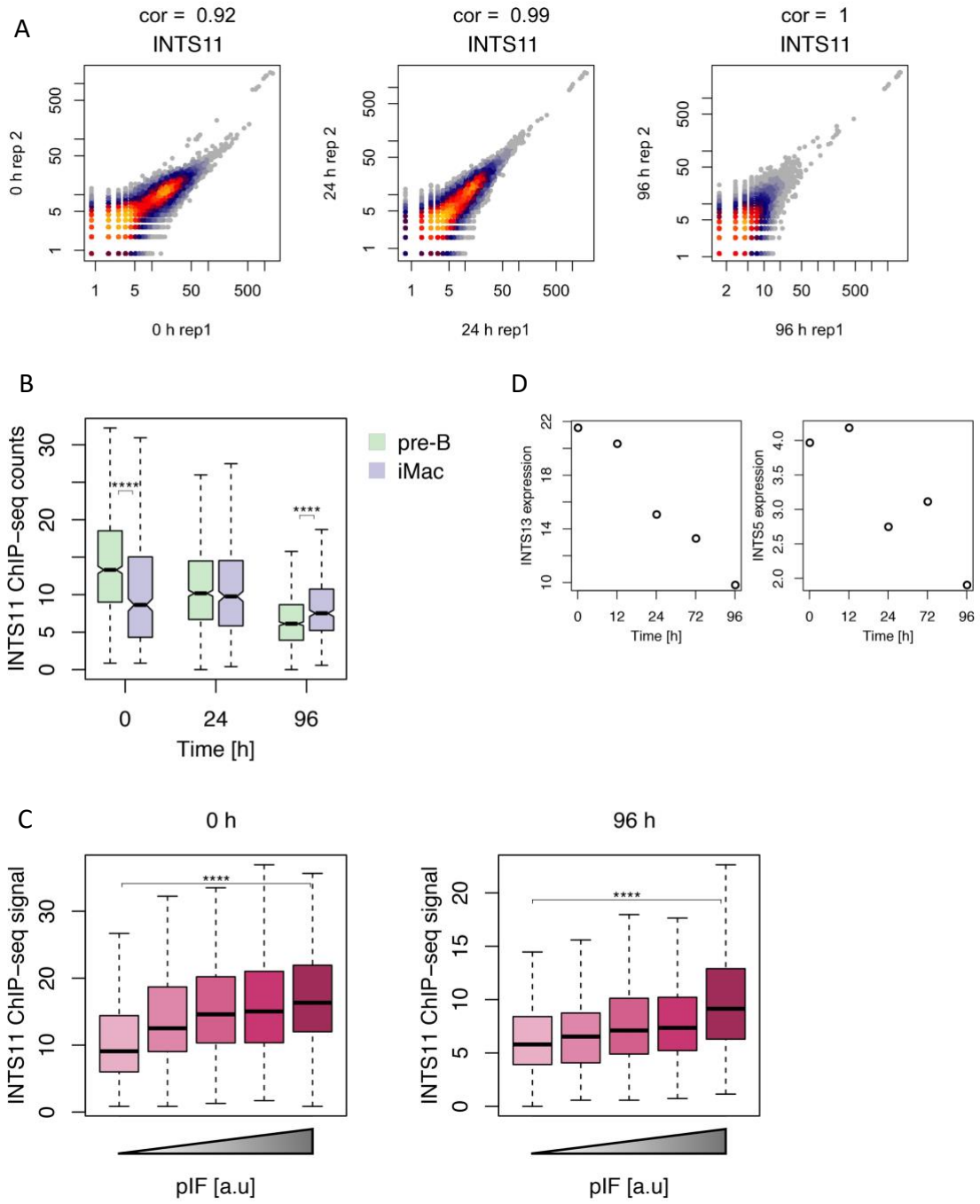


Figure S7. INTS11 occupancy correlates with the transcriptional activity during transdifferentiation.

A. Pearson correlation between two independent biological replicates of the INTS11 ChIP-seq data collected at 0, 24 and 96 h post transdifferentiation induction. **B.** Boxplots representing INTS11 ChIP-seq signal comparison of pre-B (938 TUs) and iMac (416 TUs) gene groups at 0, 24 and 96 h post transdifferentiation induction. **C.** Boxplots showing INTS11 ChIP-seq signal plotted versus pIF groups ranked from lowest to highest for 2,157 analyzed TUs. The data is shown for 0 and 96 h post transdifferentiation induction. **D.** INTS13 and INTS5 expression based on the TT-seq data across the time course of transdifferentiation (0, 12, 24, 72 and 96 h).

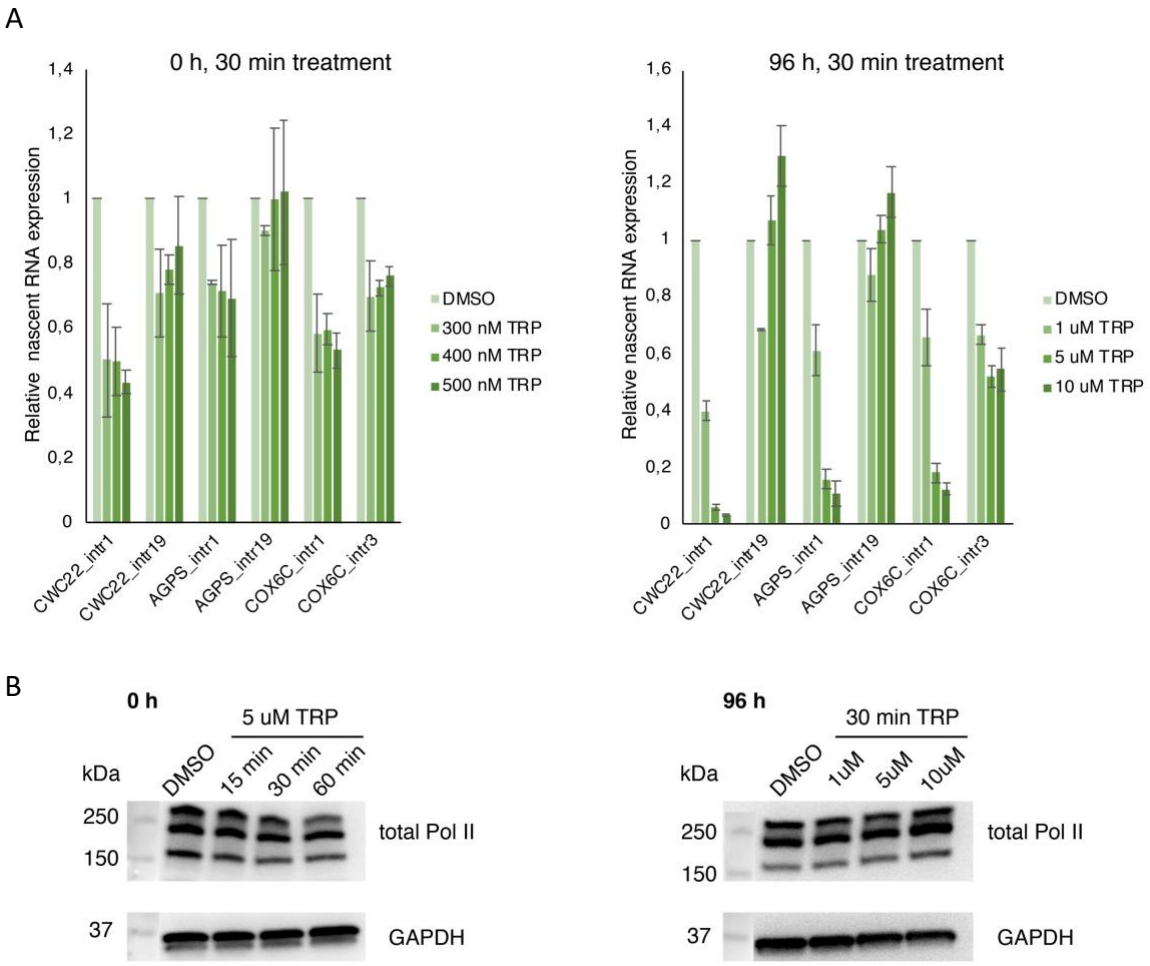


Figure S8. (legend is on the next page)

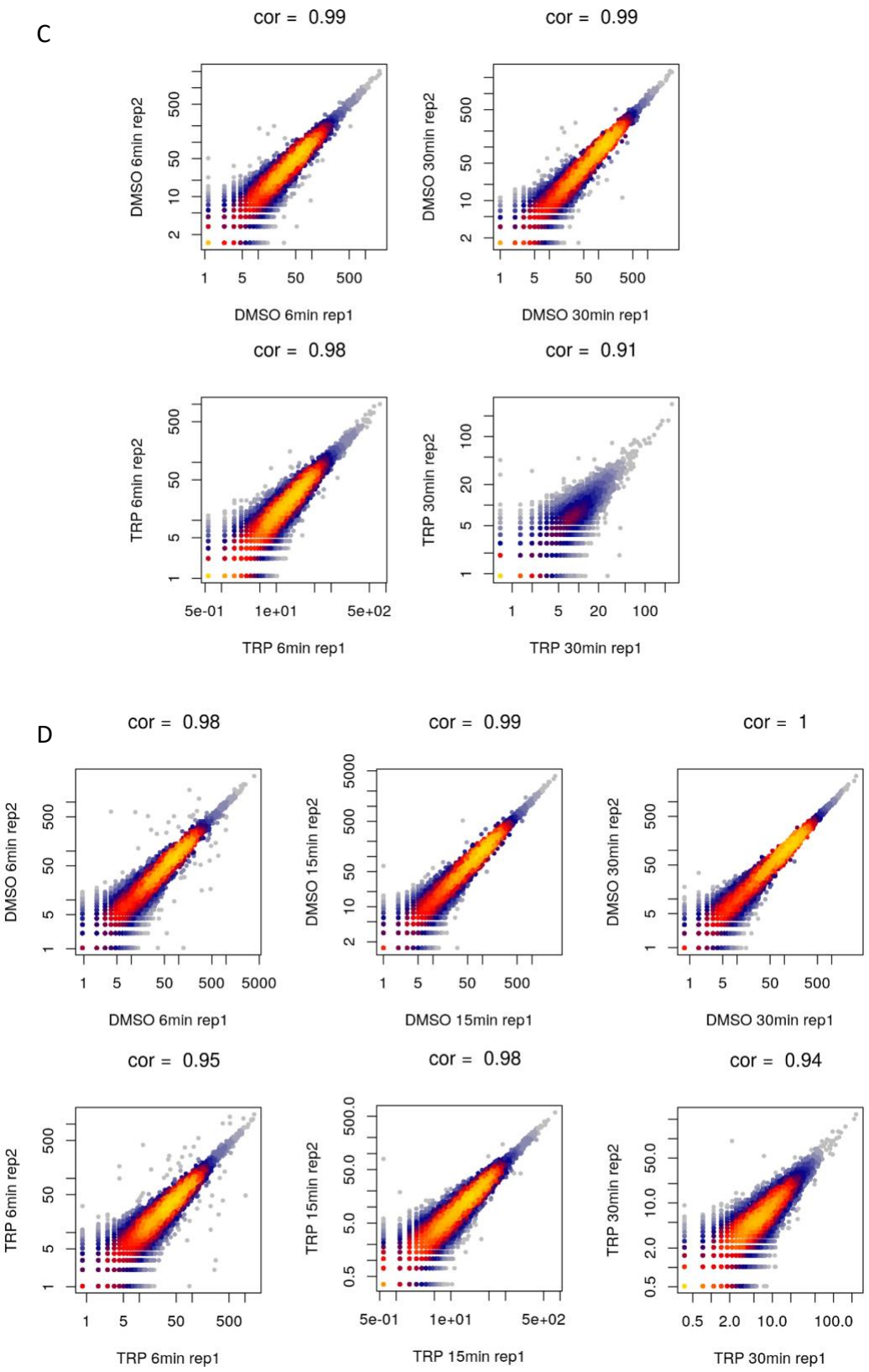


Figure S8. (legend is on the next page)

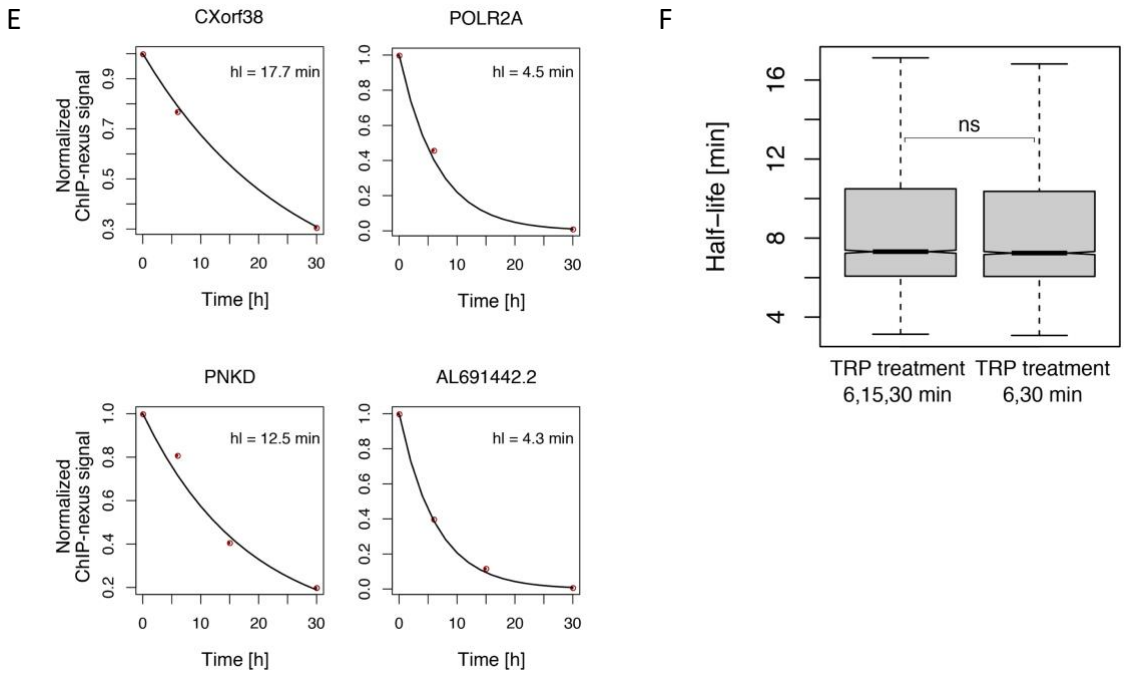


Figure S8. Global high turnover of Pol II in the promoter-proximal window under initiation inhibition.

A. RT-qPCR results showing the spike-in normalized nascent RNA expression in the beginning and the end of the genes CWC22, AGPS and COX6C under 30 min DMSO/TRP treatment at 0 h and 96 h of transdifferentiation. TRP concentration values are shown in the legend. The error bars represent standard deviation of the data for two independent biological replicates. **B.** Western blotting analysis of the whole cell lysate obtained from DMSO/TRP-treated cells at 0 h and 96 h of transdifferentiation. The treatment conditions are displayed in the legend. Total Pol II and GAPDH levels were detected. **C.** Pearson correlation between two independent biological replicates of the ChIP-nexus data collected at 0 h post transdifferentiation induction and treated with 5 µM of TRP or DMSO solvent control for 6 and 30 min. **D.** Pearson correlation between two independent biological replicates of the ChIP-nexus data collected at 96 h post transdifferentiation induction and treated with 5 µM TRP or DMSO solvent control for 6, 15 and 30 min. **E.** Representative gene examples with different half-lives of promoter-proximal Pol II. The measurements of Pol II occupancy under the time course of TRP treatment were fit to an exponential decay model. **F.** Boxplots showing comparison of promoter-proximal Pol II half-life calculations using ChIP-nexus data from either 6, 15 and 30 min of DMSO/TRP treatment time course or 6 and 30 min of DMSO/TRP treatment time course.

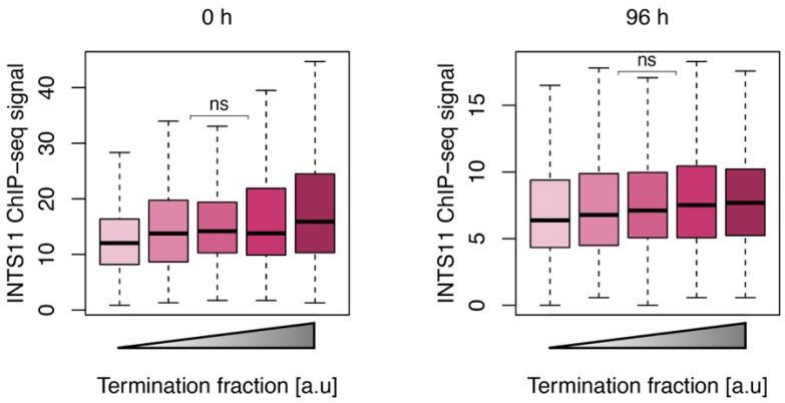


Figure S9. INTS11 occupancy demonstrates a poor correlation with estimated Pol II premature termination fraction.

Boxplots showing INTS11 ChIP-seq signal plotted versus Pol II premature termination fraction groups ranked from lowest to highest for 2,157 analyzed TUs. The data is shown for 0 and 96 h post transdifferentiation induction.

2. SUPPLEMENTARY TABLES

Table S1. Oligonucleotide sequences used in the RT-qPCR analysis of the cell-stage specific markers expression across the time course of transdifferentiation.

Designation	Source or reference	Sequence
CD19 forward	(Rapino et al., 2013)	GATGCAGACTCTTATGAGAAC
CD19 reverse	(Rapino et al., 2013)	TCAGATTTCAGAGTCAGGTG
IGJ forward	(Rapino et al., 2013)	TGTTCATGTGAAAGCCCAAG
IGJ reverse	(Rapino et al., 2013)	TCGGATGTTTCTCTCCACAA
VPREB3 forward	(Rapino et al., 2013)	GGGGACCTTCCTGTCAGTTT
VPREB3 reverse	(Rapino et al., 2013)	ACCGTAGTCCCTGATGGTGA
CD14 forward	(Rapino et al., 2013)	GATTACATAAACTGTCAGAGGC
CD14 reverse	(Rapino et al., 2013)	TCCATGGTCGATAAGTCTTC
FCGR1B forward	(Rapino et al., 2013)	CCTTGAGGTGTCATGCGTG
FCGR1B reverse	(Rapino et al., 2013)	AAGGCTTTGCCATTTCGATAGT
ITGAM forward	(Rapino et al., 2013)	GGGGTCTCCACTAAATATCTC
ITGAM reverse	(Rapino et al., 2013)	CTGACCTGATATTGATGCTG
GAPDH forward	(Choi et al., 2021)	TCTCTGCTCCTCCTGTTCGAC
GAPDH reverse	(Choi et al., 2021)	GGCGCCCAATACGACCAAAT

Table S2. Oligonucleotide sequences used in the RT-qPCR analysis of the TRP treatment conditions optimization at 0 h and 96 h of transdifferentiation.

Designation	Source or reference	Sequence
AGPS (intron 1) forward	This study	GGACGAAGACTGGAGAGGTGAATCT
AGPS (intron 1) reverse	This study	ATCCAGGTCTATGTTCCCAGCCATT
AGPS (intron 19) forward	This study	CAGCAGACGCTGGATTTATCCCTTT
AGPS (intron 19) reverse	This study	AAATGTGGACATGGAGGCCAATCTG
COX6C (intron 1) forward	This study	CCATTACTAGTGTGCCGCAGTTTGT
COX6C (intron 1) reverse	This study	CCACAAAGCCTAAGCTGTCTCTTGG
COX6C (intron 3) forward	This study	AGGGAAGAGCCAGGGTAGTTTTCTT
COX6C (intron 3) reverse	This study	CAGCAACCAAAGGAACCACCTATCC
CWC22 (intron 1) forward	This study	GGCCCAACAGTCTGTCTGTCTTTTT
CWC22 (intron 1) reverse	This study	CATAGCATGCCCCTAAAGGATGAGC
CWC22 (intron 19) forward	This study	ATAGTTCCTCTCGGGCTTTCCTTGT
CWC22 (intron 19) reverse	This study	ACGAAACTCCATACTGAGGGACAGG

Table S3. Reactome pathway analysis with STRING (version 11.5) of the iMac I gene group activated towards 96 h of transdifferentiation.

Term description	Observed gene count	Background gene count	Strength	FDR
Neutrophil degranulation	25	473	0.74	2.24e-08
InnateImmune System	33	1025	0.53	1.41e-06
Immune System	45	1956	0.38	1.88e-05

Table S4. Reactome pathway analysis with STRING (version 11.5) of the iMac II gene group activated towards 96 h of transdifferentiation.

Term description	Observed gene count	Background gene count	Strength	FDR
Immune System	74	1956	0.55	3.11e-19
Neutrophil degranulation	31	473	0.79	2.92e-12
Cytokine Signaling in Immune system	32	681	0.64	3.66e-09
Innate Immune System	39	1025	0.55	5.07e-09
Signaling by Interleukins	24	440	0.71	7.41e-08
Interleukin-4 and Interleukin-13 signaling	13	107	1.05	1.76e-07
Interleukin-10 signaling	9	45	1.27	1.58e-06
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	11	129	0.9	8.29e-05
Caspase activation via Death Receptors in the presence of ligand	5	16	1.46	0.00055
MyD88 deficiency (TLR2/4)	4	10	1.57	0.0025
IRAK4 deficiency (TLR2/4)	4	11	1.53	0.0030
Cell surface interactions at the vascular wall	9	138	0.78	0.0047
Adaptive Immune System	21	743	0.42	0.0105
Regulation of TLR by endogenous ligand	4	19	1.29	0.0135
TRAIL signaling	3	8	1.54	0.0248
TRIF-mediated programmed cell death	3	9	1.49	0.0277

Toll-like Receptor Cascades	8	152	0.69	0.0345

3. SUPPLEMENTARY METHODS - QUANTIFICATION AND STATISTICAL ANALYSES

This section is adapted from Lysakovskaia, Devadas* et al., in preparation*

(co-shared authorship)*

Author contributions: ***K.L.****, A.D. and M.L. planned the project.* ***K.L.*** *designed and performed the experiments. A.D. designed and conducted bioinformatical analysis.* ***K.L.*** *and A.D. interpreted the data. B.S. and M.L. advised on bioinformatical analysis. J.C. provided TT-seq data prior to publication.* ***K.L.*** *wrote the original draft of the manuscript with input from all authors.* ***K.L.*** *and A.D. visualized data. P.C. and M.L. conceived, planned and supervised the project. P.C. acquired funding.*

3.1. MAJOR ISOFORM ANNOTATION

Salmon version 1.3.0 (Patro et al., 2017) was used to quantify the counts for each gene isoform and to select the major isoforms from our RNA-seq dataset. An isoform from GENCODE GRCh28.p5 annotation was defined to be major if it is present in an amount more than 70% of the total mean transcripts per million (TPM) in at least one of the time points in our analysis (0, 12, 24, 72, 96 h post transdifferentiation induction) and if no other isoform of the same gene has this property at any other time points. Major isoforms for genes in chromosome M were discarded from further analysis. Final major isoform annotation contained 27,971 annotated TUs. 6,987 of them were protein coding, 9,824 were annotated as lncRNAs, 4,183 were processed pseudogenes that together made up the main part of the annotation. Apart from this, our annotation also contained 8,165 eRNAs defined and annotated from the TT-seq data (Choi et al., 2021).

3.2. TT-SEQ DATA PROCESSING AND NORMALIZATION

TT-seq BAM files (Choi et al., 2021) were processed in R/Bioconductor environment. Read pairs were discarded from further analysis if their inserts spanned a region at most 500 bp longer than the major isoform that they map to. Expressed genes were defined as those having 10 RPK mapped to them in at least one of the time points of data collection (0, 12, 24, 72, 96 h post transdifferentiation induction). Read counts were

generated using custom R scripts and spike-in normalization of the data was performed as described previously (Schwalb et al., 2016). From the RNA spike-ins we also calculated and corrected for antisense bias and cross contamination as described previously (Schwalb et al., 2016). DESeq2 algorithm (Love et al., 2014) was used to calculate size factors for normalizing the data, which was utilized for all further analysis.

3.3. Estimation of productive initiation frequency

Productive initiation frequency (pIF) was calculated from the normalized non-first exonic coverages for multi-exon transcripts and from normalized coverage from 1000 bp downstream of the TSS until the end of the genes for single exon transcripts. Coverage normalized on the respective lengths of the region used was defined as the pIF.

3.4. mNET-seq data processing and normalization

Paired-end reads of 75 bp length were collected for the samples. Reads were quality checked using FastQC (Andrews, 2010). Reads were mapped to the human genome (GRCh38) using STAR aligner (Dobin et al., 2013) which reported a consistent 95% or more uniquely mapped reads on an average for all samples. Further data processing was performed in R/Bioconductor environment using custom scripts. Reads were trimmed to the last 3’ base to identify the Pol II positions and the results were saved as RLE lists. Counts were calculated for the annotated transcripts using a custom R script for regions of interest depending on the analysis. DESeq2 (Love et al., 2014) size factors were calculated from the counts and used for normalization. From the mNET-seq data, we calculated and corrected for antisense bias and cross contamination as described previously (Caizzi et al., 2021; Schwalb et al., 2016).

3.5. Detection of pause sites

mNET-seq data was used to define and call pause peaks for the annotated TUs. A TU is said to be paused if there is a clear maxima in the mNET-seq signal profile within the first 250 bp downstream of the TSS. Further, the maximal value was required to be at least

five times greater than the median of the signals in this window to be called a pause site. 11,567 TUs were annotated with a pause position in at least one time point during transdifferentiation. Bias due to variations in the position of pause sites were reduced by further limiting the analysis to only those transcripts that did not change pause positions significantly during transdifferentiation (s.d. ≤ 75 bp). In the end, a total of 10, 807 TUs were called as paused with defined pause positions.

3.6. Estimation of apparent pause duration

Apparent pause duration (aPD) of a transcript was defined as the ratio of mNET-seq signal at a pause window of 200 bp around the called pause sites to the respective pIF as described previously (Gressel et al., 2017, 2019). Removing the ones which had no signal at one or more time points from the data, pause duration was calculated for 8,254 TUs at all time points and this subset was used for further analysis related to pause regulation.

3.7. Clustering

Clustering was performed using a custom R script that utilizes a bootstrapped k-means clustering algorithm. Genes were first clustered on the basis of trends in pIF during transdifferentiation. To further identify differences in pause regulation, we further clustered the groups obtained on the basis of calculated pause duration as well. To minimize bias, clustering was performed with multiple values for k before settling on the ones described in the results.

3.8. Gene ontology and protein-protein interaction analysis

Gene ontology analysis for groups of genes was performed using DAVID (D. W. Huang et al., 2009). GOTERM_BP_1 data from DAVID was used for analysis and creating the plots using a custom R script. For looking at the interactions between proteins coded by different groups of genes we used the multiple proteins functionality of STRING database (Szklarczyk et al., 2019). It was also used to generate the interaction networks, subcellular localization and reactome pathways.

3.9. ChIP-seq data processing and normalization

Paired-end reads of 75 bp length were collected for the samples. Reads were quality checked using FastQC (Andrews, 2010). Reads were mapped to the human genome (GRCh38) using Bowtie2 aligner (Langmead & Salzberg, 2012) which reported a consistent 95% uniquely mapped reads on an average for all samples. Further data processing was performed in R/Bioconductor environment using custom scripts. Duplicate reads were defined as those with the same start and end of mapping positions and were discarded from the data. Reads with an insert size of more than 500 bp were also discarded and the remaining reads which mapped uniquely were converted to and saved as RLE lists. Counts were calculated for the annotated transcripts using a custom R script for regions of interest depending on the factor that was immunoprecipitated. DESeq2 (Love et al., 2014) size factors were calculated from the counts at the regions of interest and used for normalization.

3.10. ChIP-nexus data processing and normalization

ChIP-nexus data was processed as described in (Q. He et al., 2015) with minor modifications as follows. Reads were checked for adapter content using CutAdapt 2.3 (M. Martin, 2011) and any regions with adapter were removed. Further to remove the region with barcodes, 9 bp from the 5' end of the reads were trimmed. Reads were then mapped to a combined human and *Drosophila* genome and samples reported a 74% uniquely mapped reads on average consistent with those from the original ChIP-nexus experiment (Q. He et al., 2015; Shao & Zeitlinger, 2017). Further data processing was done as described (Q. He et al., 2015). Duplicates were removed based on mapping locations and mapped reads were further trimmed to their 3' end to extract the position of Pol II and saved as RLE lists for further analysis. *Drosophila* spike-ins were used for normalization.

3.11. Estimation of promoter-proximal Pol II half-life

Pol II ChIP-nexus signals at the defined pause sites +/- 20 bp was used to calculate the turnover rates at the pause site. The normalized coverages for control and TRP-treated samples were fit to an exponential decay model to estimate the decay constant (k) as described (Shao & Zeitlinger, 2017) Half-lives were then calculated from the decay constant as log 2/k.

3.12. Estimation of total Pol II turnover rate and premature termination fraction

Pol II half-lives calculated from the ChIP-Nexus data were used in combination with mNET-seq and TT-seq data to estimate the potential fractions of premature termination and pause release. Half-lives ($t_{1/2}$) along with Pol II signal at the pause (p_0) in the steady state provide the total Pol II turnover rate (r):

$$r = \frac{dp}{dt} = log_2 \frac{p_0}{t_{1/2}}$$

Comparison of r with the pIF can provide a quantitative proxy of pause release and premature termination fractions that can be analyzed across different time points of transdifferentiation:

$$pause\ release\ fraction = \frac{pIF}{r}$$

$$premature\ termination\ fraction = 1 - \frac{pIF}{r}$$

3.13. Visualization and plots

All plots including metagene profiles, heatmaps, GO plots and boxplots were created using corresponding custom scripts in R/Bioconductor environment. In the metagene plots, except for the ones from ChIP-nexus data, the 95% confidence interval was defined by calculating the minimum and maximum of bootstrapped means estimated from random sampling of data with replacement (n = 500). For the ChIP-nexus and ChIP-seq data, the 95% confidence interval was not shown in order to increase visibility and separation of mean signals.

VI. REFERENCES

Adelman, K., Kennedy, M. A., Nechaev, S., Gilchrist, D. A., Muse, G. W., Chinenov, Y., & Rogatsky, I. (2009). Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(43), 18207–18212. https://doi.org/10.1073/PNAS.0910177106

Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics 2012 13:10*, *13*(10), 720–731. https://doi.org/10.1038/nrg3293

Adelman, K., Marr, M. T., Werner, J., Saunders, A., Ni, Z., Andrulis, E. D., & Lis, J. T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Molecular Cell*, *17*(1), 103–112. https://doi.org/10.1016/j.molcel.2004.11.028

Adelman, K., Wei, W., Ardehali, M. B., Werner, J., Zhu, B., Reinberg, D., & Lis, J. T. (2006). Drosophila Paf1 Modulates Chromatin Structure at Actively Transcribed Genes . *Molecular and Cellular Biology*, *26*(1), 250–260. https://doi.org/10.1128/mcb.26.1.250-260.2006

Aibara, S., Schilbach, S., & Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature*, *594*(7861), 124–128. https://doi.org/10.1038/s41586-021-03554-8

Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B., & Sharp, P. A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, *499*(7458), 360–363. https://doi.org/10.1038/nature12349

Andrews, S. (2010). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aoi, Y., Smith, E. R., Shah, A. P., Rendleman, E. J., Marshall, S. A., Woodfin, A. R., Chen, F. X., Shiekhattar, R., & Shilatifard, A. (2020). NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release. *Molecular Cell*, *78*(2), 261-274.e5. https://doi.org/10.1016/j.molcel.2020.02.014

Aoi, Y., Takahashi, Y. hei, Shah, A. P., Iwanaszko, M., Rendleman, E. J., Khan, N. H., Cho, B. K., Goo, Y. A., Ganesan, S., Kelleher, N. L., & Shilatifard, A. (2021). SPT5 stabilization of promoter-proximal RNA polymerase II. *Molecular Cell*, *81*(21), 4413-4424.e5. https://doi.org/10.1016/j.molcel.2021.08.006

Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., & Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO Journal*, *28*(8), 1067–1077. https://doi.org/10.1038/emboj.2009.56

Au, K. F., Sebastiano, V., Afshar, P. T., Durruthy, J. D., Lee, L., Williams, B. A., Van Bakel, H., Schadt, E. E., Reijo-Pera, R. A., Underwood, J. G., & Wong, W. H. (2013). Characterization of the human ESC transcriptome by hybrid sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50). https://doi.org/10.1073/pnas.1320101110

Bai, X., Kim, J., Yang, Z., Jurynec, M. J., Akie, T. E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., Zhou, Y., Grunwald, D. J., Lin, S., Cantor, A. B., Orkin, S. H., & Zon, L. I. (2010). TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell*, *142*(1), 133–143. https://doi.org/10.1016/j.cell.2010.05.028

Baillat, D., Hakimi, M. A., Näär, A. M., Shilatifard, A., Cooch, N., & Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, *123*(2), 265–276. https://doi.org/10.1016/j.cell.2005.08.019

Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-Ferrat, N., & Peterlin, B. M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Molecular Cell*, *8*(2), 327–337. https://doi.org/10.1016/S1097-2765(01)00314-8

Barboric, M., Yik, J. H. N., Czudnochowski, N., Yang, Z., Chen, R., Contreras, X., Geyer, M., Peterlin, B. M., & Zhou, Q. (2007). Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Research*, *35*(6), 2003–2012. https://doi.org/10.1093/nar/gkm063

Barra, J., Gaidos, G. S., Blumenthal, E., Beckedorff, F., Tayar, M. M., Kirstein, N., Karakac, T. K., Jensen, T. H., Impens, F., Gevaert, K., Leucci, E., Shiekhattar, R., & Marine, J. C. (2020). Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA NEAT1. *Science Advances*, *6*(27). https://doi.org/10.1126/SCIADV.AAZ9072

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., & Greenleaf, A. L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes and Development*, *24*(20), 2303–2316. https://doi.org/10.1101/gad.1968210

Beckedorff, F., Blumenthal, E., daSilva, L. F., Aoi, Y., Cingaram, P. R., Yue, J., Zhang, A., Dokaneheifard, S., Valencia, M. G., Gaidosh, G., Shilatifard, A., & Shiekhattar, R. (2020).

The Human Integrator Complex Facilitates Transcriptional Elongation by Endonucleolytic Cleavage of Nascent Transcripts. *Cell Reports*, *32*(3). https://doi.org/10.1016/j.celrep.2020.107917

Belotserkovskaya, R., Saunders, A., Lis, J. T., & Reinberg, D. (2004). Transcription through chromatin: understanding a complex FACT. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1677*(1–3), 87–99. https://doi.org/10.1016/J.BBAEXP.2003.09.017

Bernecky, C., Plitzko, J. M., & Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural and Molecular Biology*, *24*(10), 809–815. https://doi.org/10.1038/nsmb.3465

Boeing, S., Rigault, C., Heidemann, M., Eick, D., & Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *Journal of Biological Chemistry*, *285*(1), 188–196. https://doi.org/10.1074/jbc.M109.046565

Boettiger, A., & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science*, *325*(5939), 471–473. https://doi.org/10.1126/science.1098641

Bondarenko, V. A., Steele, L. M., Újvári, A., Gaykalova, D. A., Kulaeva, O. I., Polikanov, Y. S., Luse, D. S., & Studitsky, V. M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Molecular Cell*, *24*(3), 469–479. https://doi.org/10.1016/j.molcel.2006.09.009

Bortvin, A., & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science (New York, N.Y.)*, *272*(5267), 1473–1476. https://doi.org/10.1126/SCIENCE.272.5267.1473

Bottardi, S., Mavoungou, L., Bourgoin, V., Mashtalir, N., Affar, E. B., & Milot, E. (2013). Direct protein interactions are responsible for Ikaros-GATA and Ikaros-Cdk9 cooperativeness in hematopoietic cells. *Molecular and Cellular Biology*, *33*(16), 3064–3076. https://doi.org/10.1128/mcb.00296-13

Bottardi, S., Zmiri, F. A., Bourgoin, V., Ross, J., Mavoungou, L., & Milot, E. (2011). Ikaros interacts with P-TEFb and cooperates with GATA-1 to enhance transcription elongation. *Nucleic Acids Research*, *39*(9), 3505–3519. https://doi.org/10.1093/nar/gkq1271

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., & Bentley, D. L. (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Molecular Cell*, *46*(3), 311–324. https://doi.org/10.1016/j.molcel.2012.03.006

Bunch, H., Lawney, B. P., Burkholder, A., Ma, D., Zheng, X., Motola, S., Fargo, D. C., Levine, S. S., Wang, Y. E., & Hu, G. (2016). RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. *Genomics*, *108*(2), 64–77. https://doi.org/10.1016/j.ygeno.2016.07.003

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Molecular Cell*, *36*(4), 541–546. https://doi.org/10.1016/J.MOLCEL.2009.10.019

Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., & Cramer, P. (2021). Efficient RNA polymerase II pause release requires U2 snRNP function. *Molecular Cell*, *81*(9), 1920-1934.e9. https://doi.org/10.1016/j.molcel.2021.02.016

Chanput, W., Mes, J. J., & Wichers, H. J. (2014). THP-1 cell line: an in vitro cell model for immune modulation approach. *International Immunopharmacology*, *23*(1), 37–45. https://doi.org/10.1016/J.INTIMP.2014.08.002

Chen, F., Gao, X., & Shilatifard, A. (2015). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes and Development*, *29*(1), 39–47. https://doi.org/10.1101/gad.246173.114

Chen, F. X., Smith, E. R., & Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology 2018 19:7*, *19*(7), 464–478. https://doi.org/10.1038/s41580-018-0010-5

Chen, S. R., Dai, Y., Zhao, J., Lin, L., Wang, Y., & Wang, Y. (2018). A mechanistic overview of triptolide and celastrol, natural products from Tripterygium wilfordii Hook F. *Frontiers in Pharmacology*, *9*(FEB), 331085. https://doi.org/10.3389/FPHAR.2018.00104/BIBTEX

Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., Varzavand, K., Cooper, J. J., Hu, X., Gnatt, A., Young, R. A., & Price, D. H. (2012). Functional association of gdown1 with RNA polymerase II poised on human genes. *Molecular Cell*, *45*(1), 38–50. https://doi.org/10.1016/j.molcel.2011.10.022

Cheung, A. C. M., & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, *471*(7337), 249–253. https://doi.org/10.1038/nature09785

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T. v., Graf, T., & Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. *ELife*, *10*, e65381. https://doi.org/10.7554/eLife.65381

Churchman, L. S., & Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), 368–373. https://doi.org/10.1038/nature09652

Conrad, T., & Ørom, U. A. (2017). Cellular fractionation and isolation of chromatin-associated RNA. *Methods in Molecular Biology (Clifton, N.J.)*, *1468*, 1–9. https://doi.org/10.1007/978-1-4939-4035-6_1

Core, L., & Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & Development*, *33*(15–16), 960–982. https://doi.org/10.1101/GAD.325142.119

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–1320. https://doi.org/10.1038/ng.3142

Core, L. J., Waterfall, J. J., Gilchrist, D. A., Fargo, D. C., Kwak, H., Adelman, K., & Lis, J. T. (2012). Defining the status of RNA polymerase at promoters. *Cell Reports*, *2*(4), 1025–1035. https://doi.org/10.1016/j.celrep.2012.08.034

Core, L., Waterfall, J., & Lis, J. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, *322*(5909), 1845–1848. https://doi.org/10.1261/rna.1246808

Cortazar, M. A., Erickson, B., Fong, N., Pradhan, S. J., Ntini, E., & Bentley, D. L. (2022). Xrn2 substrate mapping identifies torpedo loading sites and extensive premature termination of RNA pol II transcription. *Genes and Development*, *36*(19–20), 1062–1078. https://doi.org/10.1101/gad.350004.122

Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., & Bentley, D. L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Molecular Cell*, *76*(6), 896-908.e4. https://doi.org/10.1016/j.molcel.2019.09.031

Cramer, P. (2002). Multisubunit RNA polymerases. *Current Opinion in Structural Biology*, *12*(1), 89–97. https://doi.org/10.1016/S0959-440X(02)00294-4

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature 2019 573:7772*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

Czudnochowski, N., Bösken, C. A., & Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms1846

Dasilva, L. F., Blumenthal, E., Beckedorff, F., Cingaram, P. R., Santos, H. G. Dos, Edupuganti, R. R., Zhang, A., Dokaneheifard, S., Aoi, Y., Yue, J., Kirstein, N., Tayari, M. M., Shilatifard, A., & Shiekhattar, R. (2021). Integrator enforces the fidelity of transcriptional

termination at protein-coding genes. *Science Advances*, *7*(45). https://doi.org/10.1126/SCIADV.ABE3393

Davidson, L., Francis, L., Cordiner, R. A., Eaton, J. D., Estell, C., Macias, S., Cáceres, J. F., & West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. *Cell Reports*, *26*(10), 2779-2791.e5. https://doi.org/10.1016/j.celrep.2019.02.012

Day, D. S., Zhang, B., Stevens, S. M., Ferrari, F., Larschan, E. N., Park, P. J., & Pu, W. T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0984-2

Dienemann, C., Schwalb, B., Schilbach, S., & Cramer, P. (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Molecular Cell*, *73*(1), 97-106.e4. https://doi.org/10.1016/j.molcel.2018.10.014

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Duarte, F. M., Fuda, N. J., Mahat, D. B., Core, L. J., Guertin, M. J., & Lis, J. T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes and Development*, *30*(15), 1731–1746. https://doi.org/10.1101/gad.284430.116

Eaton, J. D., & West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics : TIG*, *36*(9), 664–675. https://doi.org/10.1016/J.TIG.2020.05.008

Ehrensberger, A. H., Kelly, G. P., & Svejstrup, J. Q. (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell*, *154*(4). https://doi.org/10.1016/J.CELL.2013.07.032

Eick, D., & Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical Reviews*, *113*(11), 8456–8490. https://doi.org/10.1021/CR400071F

Elagib, K. E., Mihaylov, I. S., Delehanty, L. L., Bullock, G. C., Ouma, K. D., Caronia, J. F., Gonias, S. L., & Goldfarb, A. N. (2008). Cross-talk of GATA-1 and p-TEFb in megakaryocyte differentiation. *Blood*, *112*(13), 4884–4894. https://doi.org/10.1182/blood-2008-03-145722

Elagib, K. E., Rubinstein, J. D., Delehanty, L. L., Ngoh, V. S., Greer, P. A., Li, S., Lee, J. K., Li, Z., Orkin, S. H., Mihaylov, I. S., & Goldfarb, A. N. (2013). Calpain 2 activation of P-TEFb drives megakaryocyte morphogenesis and is disrupted by leukemogenic GATA1

mutation. *Developmental Cell*, *27*(6), 607–620. https://doi.org/10.1016/j.devcel.2013.11.013

Elrod, N. D., Henriques, T., Huang, K. L., Tatomer, D. C., Wilusz, J. E., Wagner, E. J., & Adelman, K. (2019). The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Molecular Cell*, *76*(5), 738-752.e7. https://doi.org/10.1016/j.molcel.2019.10.034

Erhard, F., Baptista, M. A. P., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C. S., Theis, F. J., Saliba, A. E., & Dölken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature*, *571*(7765), 419–423. https://doi.org/10.1038/s41586-019-1369-y

Erickson, B., Sheridan, R. M., Cortazar, M., & Bentley, D. L. (2018). Dynamic turnover of paused pol II complexes at human promoters. *Genes and Development*, *32*(17–18), 1215–1225. https://doi.org/10.1101/gad.316810.118

Escoubet-Lozach, L., Benner, C., Kaikkonen, M. U., Lozach, J., Heinz, S., Spann, N. J., Crotti, A., Stender, J., Ghisletti, S., Reichart, D., Cheng, C. S., Luna, R., Ludka, C., Sasik, R., Garcia-Bassets, I., Hoffmann, A., Subramaniam, S., Hardiman, G., Rosenfeld, M. G., & Glass, C. K. (2011). Mechanisms establishing tlr4-responsive activation states of inflammatory response genes. *PLoS Genetics*, *7*(12). https://doi.org/10.1371/journal.pgen.1002401

Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., & Cramer, P. (2021). Structural basis of Integrator-mediated transcription regulation. *Science (New York, N.Y.)*, *374*(6569), 883–887. https://doi.org/10.1126/SCIENCE.ABK0154

Flajollet, S., Rachez, C., Ploton, M., Schulz, C., Gallais, R., Métivier, R., Pawlak, M., Leray, A., Issulahi, A. A., Héliot, L., Staels, B., Salbert, G., & Lefebvre, P. (2013). The elongation complex components BRD4 and MLLT3/AF9 are transcriptional coactivators of nuclear retinoid receptors. *PLoS ONE*, *8*(6), e64880. https://doi.org/10.1371/journal.pone.0064880

Flynn, R., Almada, A., Zamudio, J., & Sharp, P. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(26), 10460–10465. https://doi.org/10.1073/pnas.1106630108/-/DCSupplemental

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature 2009 461:7261*, *461*(7261), 186–192. https://doi.org/10.1038/nature08449

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., & Peterlin, B. M. (2004). Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and

Dissociates Negative Effectors from the Transactivation Response Element. *Molecular and Cellular Biology*, *24*(2), 787–795. https://doi.org/10.1128/mcb.24.2.787-795.2004

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A. S., Gaudenz, K., De Kumar, B., Krumlauf, R., & Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Reports*, *2*(6), 1670–1683. https://doi.org/10.1016/j.celrep.2012.11.024

Gaertner, B., & Zeitlinger, J. (2014). RNA polymerase II pausing during development. *Development (Cambridge)*, *141*(6), 1179–1183. https://doi.org/10.1242/dev.088492

Galbraith, M. D., Allen, M. A., Bensard, C. L., Wang, X., Schwinn, M. K., Qin, B., Long, H. W., Daniels, D. L., Hahn, W. C., Dowell, R. D., & Espinosa, J. M. (2013). XHIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell*, *153*(6), 1327. https://doi.org/10.1016/j.cell.2013.04.048

Gao, W. wei, Xiao, R. quan, Zhang, W. juan, Hu, Y. ren, Peng, B. ling, Li, W. juan, He, Y. hui, Shen, H. feng, Ding, J. cheng, Huang, Q. xuan, Ye, T. yi, Li, Y., Liu, Z. ying, Ding, R., Rosenfeld, M. G., & Liu, W. (2018). JMJD6 Licenses ERα-Dependent Enhancer and Coding Gene Activation by Modulating the Recruitment of the CARM1/MED12 Co-activator Complex. *Molecular Cell*, *70*(2), 340-357.e8. https://doi.org/10.1016/j.molcel.2018.03.006

Gao, X. D., Tu, L. C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S. A., Zhu, L. J., Wolfe, S. A., & Sontheimer, E. J. (2018). C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nature Methods*, *15*(6), 433–436. https://doi.org/10.1038/s41592-018-0006-2

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., … Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, *15*(3), 201–206. https://doi.org/10.1038/nmeth.4577

Gariglio, P., Bellard, M., & Chambon, P. (1981). Clustering of RNA polymerase B molecules in the 5’ moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Research*, *9*(11), 2589–2598. https://doi.org/10.1093/NAR/9.11.2589

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, *143*(4), 540–551. https://doi.org/10.1016/j.cell.2010.10.004

Gilchrist, D. A., Fromm, G., dos Santos, G., Pham, L. N., Mcdaniel, I. E., Burkholder, A., Fargo, D. C., & Adelman, K. (2012). Regulating the regulators: the pervasive effects of Pol II

pausing on stimulus-responsive gene networks. *Genes and Development*, *26*(9), 933–944. https://doi.org/10.1101/gad.187781.112

Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., Gilmour, D. S., & Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes and Development*, *22*(14), 1921–1933. https://doi.org/10.1101/gad.1643208

Gilmour, D. S., & Lis, J. T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Molecular and Cellular Biology*, *6*(11), 3984–3989. https://doi.org/10.1128/MCB.6.11.3984-3989.1986

Glover-Cutter, K., Kim, S., Espinosa, J., & Bentley, D. L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Structural and Molecular Biology*, *15*(1), 71–78. https://doi.org/10.1038/nsmb1352

Gorbovytska, V., Kim, S. K., Kuybu, F., Götze, M., Um, D., Kang, K., Pittroff, A., Brennecke, T., Schneider, L. M., Leitner, A., Kim, T. K., & Kuhn, C. D. (2022). Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-29934-w

Green, I. D., Pinello, N., Song, R., Lee, Q., Halstead, J. M., Kwok, C. T., Wong, A. C. H., Nair, S. S., Clark, S. J., Roediger, B., Schmitz, U., Larance, M., Hayashi, R., Rasko, J. E. J., & Wong, J. J. L. (2020). Macrophage development and activation involve coordinated intron retention in key inflammatory regulators. *Nucleic Acids Research*, *48*(12), 6513–6529. https://doi.org/10.1093/nar/gkaa435

Gressel, S., Schwalb, B., & Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11536-8

Gressel, S., Schwalb, B., Decker, T. M., Qin, W., Leonhardt, H., Eick, D., & Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *ELife*, *6*, e29736. https://doi.org/10.7554/eLife.29736.001

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell*, *130*(1), 77–88. https://doi.org/10.1016/j.cell.2007.05.042

Gupte, R., Muse, G. W., Chinenov, Y., Adelman, K., & Rogatsky, I. (2013). Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14616–14621. https://doi.org/10.1073/pnas.1309898110

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology 2018 19:10*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, *145*(4), 622–634. https://doi.org/10.1016/j.cell.2011.03.042

Hargreaves, D. C., Horng, T., & Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell*, *138*(1), 129–145. https://doi.org/10.1016/j.cell.2009.05.047

Hay, N., Skolnik-David, H., & Aloni, Y. (1982). Attenuation in the control of SV40 gene expression. *Cell*, *29*(1), 183–193. https://doi.org/10.1016/0092-8674(82)90102-7

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., & Zhou, Q. (2010). HIV-1 Tat and Host AFF4 Recruit Two Transcription Elongation Factors into a Bifunctional Complex for Coordinated Activation of HIV-1 Transcription. *Molecular Cell*, *38*(3), 428–438. https://doi.org/10.1016/j.molcel.2010.04.013

He, Q., Johnston, J., & Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nature Biotechnology*, *33*(4), 395–401. https://doi.org/10.1038/nbt.3121

He, Y., Fang, J., Taatjes, D. J., & Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature*, *495*(7442), 481–486. https://doi.org/10.1038/nature11991

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., & Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, *533*, 359–365. https://doi.org/10.1038/nature17970

Hendriks, G. J., Jung, L. A., Larsson, A. J. M., Lidschreiber, M., Andersson Forsman, O., Lidschreiber, K., Cramer, P., & Sandberg, R. (2019). NASC-seq monitors RNA synthesis in single cells. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11028-9

Henriques, T., Gilchrist, D. A., Nechaev, S., Bern, M., Muse, G. W., Burkholder, A., Fargo, D. C., & Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular Cell*, *52*(4), 517–528. https://doi.org/10.1016/j.molcel.2013.10.001

Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., Lavender, C. A., Fargo, D. C., & Adelman, K. (2018). Widespread transcriptional pausing

and elongation control at enhancers. *Genes and Development*, *32*(1), 26–41. https://doi.org/10.1101/gad.309351.117

Herzel, L., Ottoz, D. S. M., Alpert, T., & Neugebauer, K. M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology 2017 18:10*, *18*(10), 637–650. https://doi.org/10.1038/nrm.2017.63

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, K. L., Jee, D., Stein, C. B., Elrod, N. D., Henriques, T., Mascibroda, L. G., Baillat, D., Russell, W. K., Adelman, K., & Wagner, E. J. (2020). Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Molecular Cell*, *80*(2), 345-358.e9. https://doi.org/10.1016/j.molcel.2020.08.016

Italiani, P., Boraschi, D., & Ley, K. (2014). From monocytes to M1/M2 macrophages: phenotypical vs. functional differentiation. *Frontiers in Immunology*, *5*, 116283. https://doi.org/10.3389/FIMMU.2014.00514

Ivanov, D., Kwak, Y. T., Guo, J., & Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Molecular and Cellular Biology*, *20*(9), 2970–2983. https://doi.org/10.1128/MCB.20.9.2970-2983.2000

Izban, M. G., & Luse, D. S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes & Development*, *6*(7), 1342–1356. https://doi.org/10.1101/GAD.6.7.1342

Jaeger, M. G., Schwalb, B., Mackowiak, S. D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., Ferguson, F. M., Müller, A. C., Bergthaler, A., Gray, N. S., Bradner, J. E., Bock, C., Hnisz, D., Cramer, P., & Winter, G. E. (2020). Selective Mediator dependence of cell-type-specifying transcription. *Nature Genetics*, *52*(7), 719–727. https://doi.org/10.1038/s41588-020-0635-0

Jaeger, M. G., & Winter, G. E. (2021). Fast-acting chemical tools to delineate causality in transcriptional control. *Molecular Cell*, *81*(8), 1617–1630. https://doi.org/10.1016/J.MOLCEL.2021.02.015

Jeronimo, C., Angel, A., Nguyen, V. Q., Kim, J. M., Poitras, C., Lambert, E., Collin, P., Mellor, J., Wu, C., & Robert, F. (2021). FACT is recruited to the +1 nucleosome of transcribed genes and spreads in a Chd1-dependent manner. *Molecular Cell*, *81*(17), 3542-3559.e11. https://doi.org/10.1016/j.molcel.2021.07.010

Jeziorska, D. M., Tunnacliffe, E. A. J., Brown, J. M., Ayyub, H., Sloane-Stanley, J., Sharpe, J. A., Lagerholm, B. C., Babbs, C., Smith, A. J. H., Buckle, V. J., & Higgs, D. R. (2022). On-microscope staging of live cells reveals changes in the dynamics of transcriptional bursting during differentiation. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33977-4

Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics 2009 10:3*, *10*(3), 161–172. https://doi.org/10.1038/nrg2522

Jimeno-González, S., Ceballos-Chávez, M., & Reyes, J. C. (2015). A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Research*, *43*(6), 3068–3078. https://doi.org/10.1093/nar/gkv149

Johnson, D., Mortazavi, A., Myers, R., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, *316*(5830), 1497–1502. https://doi.org/10.1126/science.1142265

Jonkers, I., Kwak, H., & Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *ELife*, *2014*(3). https://doi.org/10.7554/eLife.02407

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology 2015 16:3*, *16*(3), 167–177. https://doi.org/10.1038/nrm3953

Juven-Gershon, T., Hsu, J. Y., Theisen, J. W., & Kadonaga, J. T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology*, *20*(3), 253–259. https://doi.org/10.1016/J.CEB.2008.03.003

Kamieniarz-Gdula, K., Gdula, M. R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J. R., Riepsaame, J., Brockdorff, N., Pauli, A., & Proudfoot, N. J. (2019). Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Molecular Cell*, *74*(1), 158-172.e9. https://doi.org/10.1016/j.molcel.2019.01.027

Kamieniarz-Gdula, K., & Proudfoot, N. J. (2019). Transcriptional control by premature termination: a forgotten mechanism. *Trends in Genetics : TIG*, *35*(8), 553–564. https://doi.org/10.1016/J.TIG.2019.05.005

Kao, S. Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, *330*(6147), 489–493. https://doi.org/10.1038/330489A0

Kebede, A. F., Schneider, R., & Daujat, S. (2015). Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *FEBS Journal*, *282*(9), 1658–1674. https://doi.org/10.1111/febs.13047

Kettenberger, H., Armache, K. J., & Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, *114*(3), 347–357. https://doi.org/10.1016/S0092-8674(03)00598-1

Kim, J. B., & Sharp, P. A. (2001). Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *Journal of Biological Chemistry*, *276*(15), 12317–12323. https://doi.org/10.1074/jbc.M010908200

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., & Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature*, *436*(7052), 876–880. https://doi.org/10.1038/nature03877

Kornberg, R. D. (1999). Eukaryotic transcriptional control. *Trends in Cell Biology*, *9*(12), M46–M49. https://doi.org/10.1016/S0962-8924(99)01679-7

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705. https://doi.org/10.1016/J.CELL.2007.02.005

Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., & Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell*, *67*(3), 411-422.e4. https://doi.org/10.1016/j.molcel.2017.06.027

Krueger, S., Weilert, M., & Zeitlinger, J. (2019). *ChIP-nexus (version 2019)*. http://www.nature.com/nbt/journal/v33/n4/full/nbt.3121.html

Krumm, A., Meulia, T., Brunvand, M., & Groudine, M. (1992). The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes & Development*, *6*(11), 2201–2213. https://doi.org/10.1101/GAD.6.11.2201

Kwak, H., Fuda, N., Core, L., & Lis, J. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, *339*(6122), 950–953. https://doi.org/10.5061/dryad.j5g7m

Kwak, H., & Lis, J. T. (2013). Control of transcriptional elongation. *Annual Review of Genetics*, *47*, 483–508. https://doi.org/10.1146/ANNUREV-GENET-110711-155440

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., … Kharchenko, P. V.

(2018). RNA velocity of single cells. *Nature*, *560*(7719), 494–498. https://doi.org/10.1038/s41586-018-0414-6

Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D. S., Zeitlinger, J., & Levine, M. S. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell*, *153*(5), 976. https://doi.org/10.1016/j.cell.2013.04.045

Lai, F., Gardini, A., Zhang, A., & Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature*, *525*(7569), 399–403. https://doi.org/10.1038/nature14906

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larke, M. S. C., Schwessinger, R., Nojima, T., Telenius, J., Beagrie, R. A., Downes, D. J., Oudelaar, A. M., Truch, J., Graham, B., Bender, M. A., Proudfoot, N. J., Higgs, D. R., & Hughes, J. R. (2021). Enhancers predominantly regulate gene expression during differentiation via transcription initiation. *Molecular Cell*, *81*(5), 983-997.e7. https://doi.org/10.1016/j.molcel.2021.01.002

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B., & Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature*, *565*(7738), 251–254. https://doi.org/10.1038/s41586-018-0836-1

Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics : TIG*, *32*(1), 42–56. https://doi.org/10.1016/J.TIG.2015.10.007

Lee, C. Z. W., Kozaki, T., & Ginhoux, F. (2018). Studying tissue macrophages in vitro: are iPSC-derived cells the answer? *Nature Reviews Immunology 2018 18:11*, *18*(11), 716–725. https://doi.org/10.1038/s41577-018-0054-y

Lee, D. K., Duan, H. O., & Chang, C. (2001). Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. *Journal of Biological Chemistry*, *276*(13), 9978–9984. https://doi.org/10.1074/jbc.M002285200

Lens, Z., Cantrelle, F. X., Peruzzini, R., Hanoulle, X., Dewitte, F., Ferreira, E., Baert, J. L., Monté, D., Aumercier, M., Villeret, V., Verger, A., & Landrieu, I. (2017). Solution Structure of the N-Terminal Domain of Mediator Subunit MED26 and Molecular Characterization of Its Interaction with EAF1 and TAF7. *Journal of Molecular Biology*, *429*(20), 3043–3055. https://doi.org/10.1016/j.jmb.2017.09.001

Li, J., & Liu, C. (2019). Coding or noncoding, the converging concepts of RNAs. *Frontiers in Genetics*, *10*(MAY), 456150. https://doi.org/10.3389/FGENE.2019.00496/BIBTEX

Li, Q., Price, J. P., Byers, S. A., Cheng, D., Peng, J., & Price, D. H. (2005). Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *Journal of Biological Chemistry*, *280*(31), 28819–28826. https://doi.org/10.1074/jbc.M502712200

Li, S., Song, L., Zhang, Y., Zhan, Z., Yang, Y., Yu, L., Zhu, H., Huang, W., Wang, W., Feng, H., & Li, Y. (2022). Optimizing the Method for Differentiation of Macrophages from Human Induced Pluripotent Stem Cells. *Stem Cells International*, *2022*. https://doi.org/10.1155/2022/6593403

Lin, C., Garrett, A. S., de Kumar, B., Smith, E. R., Gogol, M., Seidel, C., Krumlauf, R., & Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes and Development*, *25*(14), 1486–1498. https://doi.org/10.1101/gad.2059211

Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., & Shilatifard, A. (2010). AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. *Molecular Cell*, *37*(3), 429–437. https://doi.org/10.1016/j.molcel.2010.01.026

Lis, J. T., Mason, P., Peng, J., Price, D. H., & Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes & Development*, *14*(7), 792–803. https://doi.org/10.1101/GAD.14.7.792

Liu, B., Xu, Q., Wang, Q., Feng, S., Lai, F., Wang, P., Zheng, F., Xiang, Y., Wu, J., Nie, J., Qiu, C., Xia, W., Li, L., Yu, G., Lin, Z., Xu, K., Xiong, Z., Kong, F., Liu, L., … Xie, W. (2020). The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature*, *587*(7832), 139–144. https://doi.org/10.1038/s41586-020-2847-y

Liu, J., Wu, X., Zhang, H., Pfeifer, G. P., & Lu, Q. (2017). Dynamics of RNA polymerase II pausing and bivalent histone H3 methylation during neuronal differentiation in brain development. *Cell Reports*, *20*(6), 1307–1318. https://doi.org/10.1016/j.celrep.2017.07.046

Liu, L., Xu, Y., He, M., Zhang, M., Cui, F., Lu, L., Yao, M., Tian, W., Benda, C., Zhuang, Q., Huang, Z., Li, W., Li, X., Zhao, P., Fan, W., Luo, Z., Li, Y., Wu, Y., Hutchins, A. P., … Esteban, M. A. (2014). Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell*, *15*(5), 574–588. https://doi.org/10.1016/j.stem.2014.09.018

Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A. K., & Rosenfeld, M. G. (2013). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell*, *155*(7), 1581–1595. https://doi.org/10.1016/j.cell.2013.10.056

Liu, X., Guo, Z., Han, J., Peng, B., Zhang, B., Li, H., Hu, X., David, C. J., & Chen, M. (2022). The PAF1 complex promotes 3' processing of pervasive transcripts. *Cell Reports*, *38*(11). https://doi.org/10.1016/j.celrep.2022.110519

Liu, X., Kraus, W. L., & Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends in Biochemical Sciences*, *40*(9), 516–525. https://doi.org/10.1016/J.TIBS.2015.07.003

Lorch, Y., & Kornberg, R. D. (2017). Chromatin-remodeling for transcription. *Quarterly Reviews of Biophysics*, *50*. https://doi.org/10.1017/S003358351700004X

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature 1997 389:6648*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luo, Z., Lin, C., Guest, E., Garrett, A. S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M. P., & Shilatifard, A. (2012). The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. *Molecular and Cellular Biology*, *32*(13), 2608–2617. https://doi.org/10.1128/mcb.00182-12

Lykke-Andersen, S., Žumer, K., Molska, E. Š., Rouvière, J. O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., & Jensen, T. H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Molecular Cell*, *81*(3), 514-529.e6. https://doi.org/10.1016/j.molcel.2020.12.014

Maier, K. C., Gressel, S., Cramer, P., & Schwalb, B. (2020). Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Research*, *30*(9), 1332–1344. https://doi.org/10.1101/GR.257857.119

Mantovani, A., Allavena, P., Marchesi, F., & Garlanda, C. (2022). Macrophages as tools and targets in cancer therapy. *Nature Reviews Drug Discovery 2022 21:11*, *21*(11), 799–820. https://doi.org/10.1038/s41573-022-00520-5

Mantsoki, A., Devailly, G., & Joshi, A. (2018). Dynamics of promoter bivalency and RNAP II pausing in mouse stem and differentiated cells. *BMC Developmental Biology*, *18*(1). https://doi.org/10.1186/s12861-018-0163-7

Manzo, S. G., Zhou, Z. L., Wang, Y. Q., Marinello, J., He, J. X., Li, Y. C., Ding, J., Capranico, G., & Miao, Z. H. (2012). Natural product triptolide mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. *Cancer Research*, *72*(20), 5363–5373. https://doi.org/10.1158/0008-5472.CAN-12-1006

Marshall, N. F., Peng, J., Xie, Z., & Price, D. H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of Biological Chemistry*, *271*(43), 27176–27183. https://doi.org/10.1074/JBC.271.43.27176

Marshall, N. F., & Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of Biological Chemistry*, *270*(21), 12335–12338. https://doi.org/10.1074/JBC.270.21.12335

Martin, G., Gruber, A. R., Keller, W., & Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Reports*, *1*(6), 753–763. https://doi.org/10.1016/j.celrep.2012.05.003

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

Martinez, F. O., Gordon, S., Locati, M., & Mantovani, A. (2006). Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *The Journal of Immunology*, *177*(10), 7303–7311. https://doi.org/10.4049/jimmunol.177.10.7303

Martinez-Rucobo, F. W., Kohler, R., van de Waterbeemd, M., Heck, A. J. R., Hemann, M., Herzog, F., Stark, H., & Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular Cell*, *58*(6), 1079–1089. https://doi.org/10.1016/j.molcel.2015.04.004

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*, *7*, 29–59. https://doi.org/10.1146/ANNUREV.GENOM.7.080505.115623

Mayer, A., Di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A., & Churchman, L. S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*, *161*(3), 541–554. https://doi.org/10.1016/j.cell.2015.03.010

Mayor-Ruiz, C., & Winter, G. E. (2019). Identification and characterization of cancer vulnerabilities via targeted protein degradation. *Drug Discovery Today. Technologies*, *31*, 81–90. https://doi.org/10.1016/J.DDTEC.2018.12.003

McGhee, J. D., & Felsenfeld, G. (1980). Nucleosome structure. *Annual Review of Biochemistry*, *49*, 1115–1156. https://doi.org/10.1146/ANNUREV.BI.49.070180.005343

Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A., & Grosveld, F. (2006). Novel binding partners Ldb1 are required for haematopoietic development. *Development*, *133*(24), 4913–4924. https://doi.org/10.1242/dev.02656

Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y., & Taatjes, D. J. (2010). P53 activates transcription by directing structural shifts in Mediator. *Nature Structural and Molecular Biology*, *17*(6), 753–760. https://doi.org/10.1038/nsmb.1816

Min, I. M., Waterfall, J. J., Core, L. J., Munroe, R. J., Schimenti, J., & Lis, J. T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes and Development*, *25*(7), 742–754. https://doi.org/10.1101/gad.2005511

Mitra, P., Pereira, L. A., Drabsch, Y., Ramsay, R. G., & Gonda, T. J. (2012). Estrogen receptor-α recruits P-TEFb to overcome transcriptional pausing in intron 1 of the MYB gene. *Nucleic Acids Research*, *40*(13), 5988–6000. https://doi.org/10.1093/nar/gks286

Moon, K. J., Mochizuki, K., Zhou, M., Jeong, H. S., Brady, J. N., & Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell*, *19*(4), 523–534. https://doi.org/10.1016/j.molcel.2005.06.027

Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics 2014 15:6*, *15*(6), 423–437. https://doi.org/10.1038/nrg3722

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., & Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nature Genetics*, *39*(12), 1507–1511. https://doi.org/10.1038/ng.2007.21

Myers, S. A., Wright, J., Peckner, R., Kalish, B. T., Zhang, F., & Carr, S. A. (2018). Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nature Methods*, *15*(6), 437–439. https://doi.org/10.1038/s41592-018-0007-1

Nechaev, S., Fargo, D. C., Santos, G. Dos, Liu, L., Gao, Y., & Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science (New York, N.Y.)*, *327*(5963), 335–338. https://doi.org/10.1126/SCIENCE.1181421

Nguyen, V. T., Kiss, T., Michels, A. A., & Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, *414*(6861), 322–325. https://doi.org/10.1038/35104581

Nilson, K. A., Lawson, C. K., Mullen, N. J., Ball, C. B., Spector, B. M., Meier, J. L., & Price, D. H. (2017). Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Research*, *45*(19), 11088–11105. https://doi.org/10.1093/nar/gkx724

Nojima, T., Gomes, T., Carmo-Fonseca, M., & Proudfoot, N. J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nature Protocols*, *11*(3), 413–428. https://doi.org/10.1038/nprot.2016.012

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M., & Proudfoot, N. J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*, *161*(3), 526–540. https://doi.org/10.1016/j.cell.2015.03.027

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S. M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., & Proudfoot, N. J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Molecular Cell*, *72*(6), 970-984.e7. https://doi.org/10.1016/j.molcel.2018.10.011

Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes & Development*, *10*(21), 2657–2683. https://doi.org/10.1101/GAD.10.21.2657

Pagano, J. M., Kwak, H., Waters, C. T., Sprouse, R. O., White, B. S., Ozer, A., Szeto, K., Shalloway, D., Craighead, H. G., & Lis, J. T. (2014). Defining NELF-E RNA Binding in HIV-1 and Promoter-Proximal Pause Regions. *PLoS Genetics*, *10*(1). https://doi.org/10.1371/journal.pgen.1004090

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pei, Y., & Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *Journal of Biological Chemistry*, *277*(22), 19639–19648. https://doi.org/10.1074/jbc.M200015200

Petesch, S. J., & Lis, J. T. (2012). Overcoming the nucleosome barrier during transcript elongation. *Trends in Genetics : TIG*, *28*(6), 285–294. https://doi.org/10.1016/J.TIG.2012.02.005

Plet, A., Eick, D., & Blanchard, J. M. (1995). Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene*, *10*(2), 319–328. https://europepmc.org/article/med/7838531

Price, D. H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and Cellular Biology*, *20*(8), 2629–2634. https://doi.org/10.1128/MCB.20.8.2629-2634.2000

Quaresma, A. J. C., Bugai, A., & Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic Acids Research*, *44*(16), 7527–7539. https://doi.org/10.1093/NAR/GKW585

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., & Young, R. A. (2010). C-Myc regulates transcriptional pause release. *Cell*, *141*(3), 432–445. https://doi.org/10.1016/j.cell.2010.03.030

Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, *135*(2), 216–226. https://doi.org/10.1016/J.CELL.2008.09.050

Rapino, F., Robles, E. F., Richter-Larrea, J. A., Kallin, E. M., Martinez-Climent, J. A., & Graf, T. (2013). C/EBPα induces highly efficient macrophage transdifferentiation of B lymphoma and leukemia cell lines and impairs their tumorigenicity. *Cell Reports*, *3*(4), 1153–1163. https://doi.org/10.1016/j.celrep.2013.03.003

Reines, D. (1992). Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *Journal of Biological Chemistry*, *267*(6), 3795–3800. https://doi.org/10.1016/S0021-9258(19)50596-8

Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., & Cramer, P. (2021). Structure of the human Mediator–RNA polymerase II pre-initiation complex. *Nature*, *594*(7861), 129–133. https://doi.org/10.1038/s41586-021-03555-7

Reppas, N. B., Wade, J. T., Church, G. M. M., & Struhl, K. (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Molecular Cell*, *24*(5), 747–757. https://doi.org/10.1016/j.molcel.2006.10.030

Roeder, R. G., & Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, *224*(5216), 234–237. https://doi.org/10.1038/224234A0

Rougvie, A. E., & Lis, J. T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, *54*(6), 795–804. https://doi.org/10.1016/S0092-8674(88)91087-2

Rougvie, A. E., & Lis, J. T. (1990). Postinitiation transcriptional control in Drosophila melanogaster. *Molecular and Cellular Biology*, *10*(11), 6041–6045. https://doi.org/10.1128/MCB.10.11.6041-6045.1990

Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology 2015 16:3*, *16*(3), 129–143. https://doi.org/10.1038/nrm3952

Salzberg, S. L. (2018). Open questions: How many genes do we have? *BMC Biology*, *16*(1), 1–3. https://doi.org/10.1186/S12915-018-0564-X/TABLES/1

Sansó, M., & Fisher, R. P. (2013). Pause, play, repeat: CDKs push RNAP II's buttons. *Transcription*, *4*(4), 146–152. https://doi.org/10.4161/trns.25146

Sansó, M., Levin, R. S., Lipp, J. J., Wang, V. Y.-F., Greifenberg, A. K., Quezada, E. M., Ali, A., Ghosh, A., Larochelle, S., Rana, T. M., Geyer, M., Tong, L., Shokat, K. M., & Fisher, R. P. (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes and Development*, *30*(1), 117–131. https://doi.org/10.1101/gad.269589

Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T., & Ashe, H. L. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. *Genes and Development*, *27*(10), 1146–1158. https://doi.org/10.1101/gad.215459.113

Schaaf, C. A., Kwak, H., Koenig, A., Misulovin, Z., Gohara, D. W., Watson, A., Zhou, Y., Lis, J. T., & Dorsett, D. (2013). Genome-wide control of RNA polymerase II activity by cohesin. *PLoS Genetics*, *9*(3). https://doi.org/10.1371/journal.pgen.1003382

Schaukowitch, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., & Kim, T. K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, *56*(1), 29–42. https://doi.org/10.1016/j.molcel.2014.08.023

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. https://doi.org/10.1016/j.molcel.2016.11.029

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887–898. https://doi.org/10.1016/j.cell.2008.02.022

Schröder, S., Cho, S., Zeng, L., Zhang, Q., Kaehlcke, K., Mak, L., Lau, J., Bisgrove, D., Schnölzer, M., Verdin, E., Zhou, M. M., & Ott, M. (2012). Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. *Journal of Biological Chemistry*, *287*(2), 1090–1099. https://doi.org/10.1074/jbc.M111.282855

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., & Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, *352*(6290), 1225–1228. https://doi.org/10.1126/science.aad9335

Sentenac, A. (1985). Eukaryotic RNA polymerase. *Critical Reviews in Biochemistry and Molecular Biology*, *18*(1), 31–90. https://doi.org/10.3109/10409238509082539

Shao, W., & Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. *Nature Genetics*, *49*(7), 1045–1051. https://doi.org/10.1038/ng.3867

Sheridan, R. M., Fong, N., D'Alessandro, A., & Bentley, D. L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5′ Pause Release, Termination, and Transcription Elongation Rate. *Molecular Cell*, *73*(1), 107-118.e4. https://doi.org/10.1016/j.molcel.2018.10.031

Shkel, O., Kharkivska, Y., Kim, Y. K., & Lee, J. S. (2022). Proximity labeling techniques: a multi-omics toolbox. *Chemistry, an Asian Journal*, *17*(2). https://doi.org/10.1002/ASIA.202101240

Sigova, A. A., Mullen, A. C., Molinie, B., Gupta, S., Orlando, D. A., Guenther, M. G., Almada, A. E., Lin, C., Sharp, P. A., Giallourakis, C. C., & Young, R. A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(8), 2876–2881. https://doi.org/10.1073/pnas.1221904110

Smith, E., Lin, C., & Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. *Genes & Development*, *25*(7), 661–672. https://doi.org/10.1101/GAD.2015411

Soneson, C., Yao, Y., Bratus-Neuenschwander, A., Patrignani, A., Robinson, M. D., & Hussain, S. (2019). A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11272-z

Song, S. H., Kim, A. R., Ragoczy, T., Bender, M. A., Groudine, M., & Dean, A. (2010). Multiple functions of Ldb1 required for β-globin activation during erythroid differentiation. *Blood*, *116*(13), 2356–2364. https://doi.org/10.1182/blood-2010-03-272252

Stein, C. B., Field, A. R., Mimoso, C. A., Zhao, C., Huang, K.-L., Wagner, E. J., & Adelman, K. (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2022.10.004

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., Houtsmuller, A. B., & Marteijn, J. A. (2018).

Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4368–E4376. https://doi.org/10.1073/pnas.1717920115

Stik, G., Vidal, E., Barrero, M., Cuartero, S., Vila-Casadesús, M., Mendieta-Esteban, J., Tian, T. V., Choi, J., Berenguer, C., Abad, A., Borsari, B., le Dily, F., Cramer, P., Marti-Renom, M. A., Stadhouders, R., & Graf, T. (2020). CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nature Genetics*, *52*(7), 655–661. https://doi.org/10.1038/s41588-020-0643-0

Strobl, L. J., & Eick, D. (1992). Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *The EMBO Journal*, *11*(9), 3307–3314. https://doi.org/10.1002/j.1460-2075.1992.tb05409.x

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Takahashi, H., Parmely, T. J., Sato, S., Tomomori-Sato, C., Banks, C. A. S., Kong, S. E., Szutorisz, H., Swanson, S. K., Martin-Brown, S., Washburn, M. P., Florens, L., Seidel, C. W., Lin, C., Smith, E. R., Shilatifard, A., Conaway, R. C., & Conaway, J. W. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell*, *146*(1), 92–104. https://doi.org/10.1016/j.cell.2011.06.005

Tatomer, D. C., Elrod, N. D., Liang, D., Xiao, M. S., Jiang, J. Z., Jonathan, M., Huang, K. L., Wagner, E. J., Cherry, S., & Wilusz, J. E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes & Development*, *33*(21–22), 1525–1538. https://doi.org/10.1101/gad.330167.119

Tellier, M., Zaborowska, J., Caizzi, L., Mohammad, E., Velychko, T., Schwalb, B., Ferrer-Vicens, I., Blears, D., Nojima, T., Cramer, P., & Murphy, S. (2020). CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Research*, *48*(14), 7712–7727. https://doi.org/10.1093/nar/gkaa514

Thomas, M. C., & Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, *41*(3), 105–178. https://doi.org/10.1080/10409230600648736

Titov, D. V., Gilman, B., He, Q. L., Bhat, S., Low, W. K., Dang, Y., Smeaton, M., Demain, A. L., Miller, P. S., Kugel, J. F., Goodrich, J. A., & Liu, J. O. (2011). XPB, a subunit of TFIIH, is a

target of the natural product triptolide. *Nature Chemical Biology*, *7*(3), 182–188. https://doi.org/10.1038/nchembio.522

Tome, J. M., Tippens, N. D., & Lis, J. T. (2018). Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nature Genetics*, *50*(11), 1533–1541. https://doi.org/10.1038/s41588-018-0234-5

Tunnacliffe, E., & Chubb, J. R. (2020). What is a transcriptional burst? *Trends in Genetics : TIG*, *36*(4), 288–297. https://doi.org/10.1016/J.TIG.2020.01.003

Ummethum, H., & Hamperl, S. (2020). Proximity labeling techniques to study chromatin. *Frontiers in Genetics*, *11*, 536839. https://doi.org/10.3389/FGENE.2020.00450/BIBTEX

Van Oss, S. B., Shirra, M. K., Bataille, A. R., Wier, A. D., Yen, K., Vinayachandran, V., Byeon, I. J. L., Cucinotta, C. E., Héroux, A., Jeon, J., Kim, J., VanDemark, A. P., Pugh, B. F., & Arndt, K. M. (2016). The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6. *Molecular Cell*, *64*(4), 815–825. https://doi.org/10.1016/j.molcel.2016.10.008

Vispé, S., DeVries, L., Créancier, L., Besse, J., Bréand, S., Hobson, D. J., Svejstrup, J. Q., Annereau, J. P., Cussac, D., Dumontet, C., Guilbaud, N., Barret, J. M., & Bailly, C. (2009). Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. *Molecular Cancer Therapeutics*, *8*(10), 2780–2790. https://doi.org/10.1158/1535-7163.MCT-09-0549

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., & Cramer, P. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*, *560*(7720), 607–612. https://doi.org/10.1038/s41586-018-0440-4

Vos, S. M., Farnung, L., Urlaub, H., & Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*, *560*(7720), 601–606. https://doi.org/10.1038/s41586-018-0442-2

Vos, S. M., Pöllmann, D., Caizzi, L., Hofmann, K. B., Rombaut, P., Zimniak, T., Herzog, F., & Cramer, P. (2016). Architecture and RNA binding of the human negative elongation factor. *ELife*, *5*. https://doi.org/10.7554/eLife.14981.001

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., & Handa, H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*, *12*(3), 343–356. https://doi.org/10.1101/GAD.12.3.343

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., Boyer, F., Parrinello, H., Berkhout, B.,

Terzian, C., Benkirane, M., & Kiernan, R. (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell*, *150*(6), 1147–1157. https://doi.org/10.1016/j.cell.2012.08.004

Wang, Y., Lu, J. jian, He, L., & Yu, Q. (2011). Triptolide (TPL) inhibits global transcription by inducing proteasome-dependent degradation of RNA polymerase II (Pol II). *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0023993

Watanabe, S., Alexander, M., Misharin, A. V., & Budinger, G. R. S. (2019). The role of macrophages in the resolution of inflammation. *The Journal of Clinical Investigation*, *129*(7), 2619. https://doi.org/10.1172/JCI124615

Weber, C. M., Ramachandran, S., & Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase. *Molecular Cell*, *53*(5), 819–830. https://doi.org/10.1016/j.molcel.2014.02.014

Wen, Y., & Shatkin, A. J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & Development*, *13*(14), 1774. https://doi.org/10.1101/GAD.13.14.1774

Williams, L. H., Fromm, G., Gokey, N. G., Henriques, T., Muse, G. W., Burkholder, A., Fargo, D. C., Hu, G., & Adelman, K. (2015). Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Molecular Cell*, *58*(2), 311–322. https://doi.org/10.1016/j.molcel.2015.02.003

Wong, K. H., Jin, Y., & Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Molecular Cell*, *54*(4), 601–612. https://doi.org/10.1016/j.molcel.2014.03.024

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., Zuzarte, P. C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., de Jesus, J. G., Jones, K. L., Soulette, C. M., Snutch, T. P., Loman, N., Paten, B., Loose, M., … Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, *16*(12), 1297–1305. https://doi.org/10.1038/s41592-019-0617-2

Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., & Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes and Development*, *17*(11), 1402–1414. https://doi.org/10.1101/gad.1091403

Wynn, T. A., Chawla, A., & Pollard, J. W. (2013). Macrophage biology in development, homeostasis and disease. *Nature 2013 496:7446*, *496*(7446), 445–455. https://doi.org/10.1038/nature12034

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., & Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive

transcription elongation. *Molecular Cell*, *21*(2), 227–237. https://doi.org/10.1016/j.molcel.2005.11.024

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., & Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Molecular and Cellular Biology*, *22*(9), 2918–2927. https://doi.org/10.1128/mcb.22.9.2918-2927.2002

Yamaguchi, Y., Shibata, H., & Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochimica et Biophysica Acta*, *1829*(1), 98–104. https://doi.org/10.1016/J.BBAGRM.2012.11.007

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., & Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, *97*(1), 41–51. https://doi.org/10.1016/S0092-8674(00)80713-8

Yang, J., Zhang, L., Yu, C., Yang, X. F., & Wang, H. (2014). Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. *Biomarker Research*, *2*(1), 1. https://doi.org/10.1186/2050-7771-2-1

Yang, Z., Yik, J. H. N., Chen, R., He, N., Moon, K. J., Ozato, K., & Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*, *19*(4), 535–545. https://doi.org/10.1016/j.molcel.2005.06.029

Yang, Z., Zhu, Q., Luo, K., & Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature 2001 414:6861*, *414*(6861), 317–322. https://doi.org/10.1038/35104575

Yu, L., Zhang, B., Deochand, D., Sacta, M. A., Coppo, M., Shang, Y., Guo, Z., Zeng, X., Rollins, D. A., Tharmalingam, B., Li, R., Chinenov, Y., Rogatsky, I., & Hu, X. (2020). Negative elongation factor complex enables macrophage inflammatory responses by controlling anti-inflammatory gene expression. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16209-5

Zaborowska, J., Egloff, S., & Murphy, S. (2016). The pol II CTD: new twists in the tail. *Nature Structural & Molecular Biology 2016 23:9*, *23*(9), 771–777. https://doi.org/10.1038/nsmb.3285

Zeitlinger, J., Stark, A., Kellis, M., Hong, J. W., Nechaev, S., Adelman, K., Levine, M., & Young, R. A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Genetics*, *39*(12), 1512–1516. https://doi.org/10.1038/ng.2007.26

Zheng, B., Aoi, Y., Shah, A. P., Iwanaszko, M., Das, S., Rendleman, E. J., Zha, D., Khan, N., Smith, E. R., & Shilatifard, A. (2021). Acute perturbation strategies in interrogating RNA polymerase II elongation factor function in gene expression. *Genes and Development*, *35*(3), 273–285. https://doi.org/10.1101/GAD.346106.120

Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J., Wang, J., Wei, G., Liang, K., Chen, F. X., & Xu, Y. (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science*, *370*(6520). https://doi.org/10.1126/science.abb5872

Zhou, Q., Li, T., & Price, D. H. (2012). RNA polymerase II elongation control. *Annual Review of Biochemistry*, *81*, 119–143. https://doi.org/10.1146/annurev-biochem-052610-095910

Zimmer, J. T., Rosa-Mercado, N. A., Canzio, D., Steitz, J. A., & Simon, M. D. (2021). STL-seq reveals pause-release and termination kinetics for promoter-proximal paused RNA polymerase II transcripts. *Molecular Cell*, *81*(21), 4398-4412.e7. https://doi.org/10.1016/j.molcel.2021.08.019

VII. APPENDIX

1. LIST OF FIGURES

2. List of supplementary figures

3. List of supplementary tables

4. Abbreviations and acronyms

[a.u.]	arbitrary units
[bp]	base pairs
[h]	hour
[kbp]	kilobase pairs
[min]	minute
[nt]	nucleotide
aPD	apparent pause duration
ATP	adenosine triphosphate
BRD	bromodomain
Cas9	caspase 9
CDK	cyclin-dependent kinase
CF	cleavage factor
ChIP-on-chip	chromatin immunoprecipitation followed by DNA microarray
ChIP-seq	chromatin immunoprecipitation sequencing
cor	correlation
CPA	cleavage and polyadenylation
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulatory factor
CTD	carboxy-terminal domain
CycT1	cyclin T1
dCas9	dead mutant of caspase 9
DE	differential expression
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSG	disuccinimidyl glutarate
DSIF	DRB sensitivity-inducing factor
DTT	dithiothreitol

ER	estrogen receptor
eRNA	enhancer RNA
FACS	fluorescence-activated cell sorting
FACT	factor facilitating transcription through the nucleosomes
FBS	fetal bovine serum
FC	fold change
FRAP	fluorescence recovery after photobleaching
GAF	GAGA-associated factor
GenoSTAN	genomic state annotation
GFP	green fluorescent protein
GO	gene ontology
GRO-seq	global nuclear run-on sequencing
GTF	general transcription factor
hESCs	human embryonic stem cells
HIF1α	hypoxia-inducible factor 1α
HIV	human immunodeficiency virus
HRP	horseradish peroxidase
HS	heat shock
HSF	heat shock factor
IL	interleukin
iMac	induced macrophage
IP	immunoprecipitation
iPSCs	induced pluripotent stem cells
LFC	log fold change
lncRNA	long non-coding RNA
LPS	lipopolysaccharide
LTR	long terminal repeat
M-CSF	macrophage colony-stimulating factor
mESCs	mouse embryonic stem cells
MLL	mixed lineage leukemia
MNase	micrococcal nuclease

mNET-seq	mammalian native elongating transcript sequencing
mRNA	messenger RNA
MS	mass spectrometry
ncRNA	non-coding RNA
NDR	nucleosome-depleted region
NELF	negative elongation factor
NET-seq	native elongating transcript sequencing
NF-κB	nuclear factor κB
NTD	amino-terminal domain
ONT	Oxford Nanopore Technologies
P-TEFb	positive transcription elongation factor b
pA site	polyadenylation site
PacBio	Pacific Biosciences
PAF1	polymerase-associated factor 1
PIC	pre-initiation complex
pIF	productive initiation frequency
PNK	polynucleotide 5'-hydroxyl-kinase
Pol	polymerase
PP2A	protein phosphatase 2A
pre-B	precursor leukemia B-cell
pre-mRNA	precursor mRNA
PRO-seq	precision nuclear run-on sequencing
PVDF	polyvinylidene fluoride
PWC	pause window coverage
RA	retinoic acid
rep	replicate
RLE	run-length encoding
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RPK	reads per kilobase
rRNA	ribosomal RNA

RT	room temperature
RT-qPCR	reverse transcription followed by quantitative polymerase chain reaction
scRNA	short capped RNA
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	super elongation complex
sgRNA	single guide RNA
smFISH	single-molecule fluorescence in situ hybridization
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
Start-seq	short nascent RNA associated with early elongation complexes sequencing
STL-seq	Start-TimeLapse sequencing
TLR	toll-like receptor
TNFα	tumor necrosis factor α
TPM	transcripts per million
tRNA	transfer RNA
TRP	triptolide
TSS	transcription start site
tssRNA	TSS-associated RNA
TT-seq	transient transcriptome sequencing
TU	transcribed unit
uaRNA	upstream antisense RNA