Published in final edited form as:

Nat Rev Genet. 2019 December ; 20(12): 705–723. doi:10.1038/s41576-019-0159-6.

Nascent RNA Analyses: Tracking Transcription and Its Regulation

Erin M. Wissink[1,3], **Anniina Vihervaara**[1,3], **Nathaniel D. Tippens**[1,2], **John T. Lis**[1]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.

[2]Tri-Institutional Training Program in Computational Biology and Medicine, New York, New York 10065, USA.

[3]These authors contributed equally to this work.

Abstract

The genome encodes information to program an organism's development and maintenance, and its decoding begins with regulated transcription of genomic DNA into RNA. Transcription and its control can be tracked indirectly by measuring stable RNAs, or directly by measuring nascent RNAs. The latter reveals the immediate regulatory changes in response to developmental, environmental, disease, and metabolic signals. Multiple complementary methods have been developed to quantitatively track nascent transcription genome-wide at nucleotide-resolution, providing novel insights to mechanisms of gene regulation and transcription-coupled RNA processing. Here, we critically evaluate the array of strategies used for investigating nascent transcription and discuss recent conceptual advances they provide.

Introduction

Transcription is a process whereby RNA polymerases (Pol) synthesize RNA molecules from a DNA template. In eukaryotes, Pol I and III synthesize ribosomal RNAs and small structural RNAs, respectively; whereas Pol II produces protein-coding mRNAs, long non-coding RNAs, primary microRNAs, and enhancer RNAs (eRNAs) **[G]**. While every cell in an individual organism contains an identical genome, regulation of gene expression defines which RNAs and proteins are synthesized, and the level to which they are produced (reviewed in Ref. 1). Consequently, genome-wide coordination of transcription underlies cellular differentiation, responses to internal and extracellular signals, and organismal functions.

Transcription of genes is controlled *via* proximal and distal regulatory elements, termed promoters and enhancers, respectively (reviewed in Refs. 2 and 3). These regulatory elements contain binding sites for promoter - and enhancer-specific transcription factors (TFs) that define when a gene is active, and the frequency with which it is transcribed

* jtl10@cornell.edu.

Competing interests

Authors declare no competing interests.

(Figure 1). In mammals, promoters and enhancers contain two core initiation regions[4,5] **[G].** These regions are bound by general transcription factors (GTFs), that together with Pol II comprise the pre-initiation complex (PIC). Both promoters and enhancers have a very similar chromatin architecture, including the constellation of GTFs, TFs and Pol II complexes, and drive divergent transcription from their core initiation regions[4] (Figure 1). Enhancer transcripts, however, are generally short and unstable, whereas the coding strand of a gene produces transcripts (pre-mRNA) that are predominantly long and stable (Figure 1)[4,6,7]. The mechanisms of enhancer-promoter communication remain poorly understood, but recent reports suggest eRNA production correlates with its functional enhancer capacity[2,3] (reviewed in Ref. 8).

Beyond binding to genomic sequences, certain transcription regulators directly act at the transcription machinery. One of the major signaling platforms for transcription regulation is the C-terminal domain (CTD) of Pol II's Rbp1 subunit **[G]** (reviewed in Refs. 9 and 10). This CTD consists of multiple heptad repeats of a consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser sequence, and undergoes regulated post-translational modifications (PTMs) before and during transcription, leading to dramatic changes in the entourage of factors associated with Pol II as it progresses through the transcription cycle (reviewed in Ref. 10).

Detailed mechanistic studies at genes have revealed that transcription consists of multiple regulated steps (Figure 1). Transcription begins (step 1) when a pioneer factor binds closed chromatin and increases chromatin accessiblility by recruiting nucleosome remodelers and histone acetyltransferases (reviewed in Refs. 11 and 12). As chromatin opens, DNA elements become accessible to additional TFs and formation of PIC (step 2)[13] (reviewed in Ref. 14). Once the PIC is formed, its TFIIH subunit unwinds double-stranded DNA, allowing Pol II to initiate transcription (step 3)[15,16]. Engaged Pol II initially transcribes 20 to 60 nucleotides (nts), then undergoes promoter-proximal pausing, which in mammals and many other metazoans is a major rate-limiting step in gene expression, and a regulatory check-point for execution of transcription programs (step 4)[17–23] (reviewed in Ref. 24). During this early transcription, the nascent RNA is protected by 5′-capping[25]. Pausing of Pol II is stabilized by NELF and DSIF[26–28], whereas release from the pause (step 5) requires P-TEFb, whose CDK9 subunit phosphorylates NELF-E, DSIF, and Ser2 residue of the Pol II CTD (reviewed in Ref. 29). During productive elongation (step 6), multiple elongation factors enhance processivity of Pol II[30,31] and couple nascent transcription to processes such as splicing (reviewed in Ref. 32). Finally, at the 3′-end of the gene, the transcript is cleaved and the pre-mRNA polyadenylated[3] (reviewed in Ref. 33). This cleavage releases pre-mRNA from Pol II and leaves uncapped nascent RNA unshielded from XRN2-mediated degradation, which destabilizes Pol II and contributes to termination (step 7) (reviewed in Ref. 34). The terminating Pol II dissociates from the DNA and is recycled to a new round of transcription[35].

The current mechanistic view of transcription regulation (Figure 1) combines data from a variety of techniques that have mapped the chromatin composition and interrogated nascent RNA synthesis at each step. Over the past 50 years of tracking RNA synthesis[36–38], our ability to investigate transcription has evolved greatly, expanding single-locus analyses to probing transcriptional mechanisms across genomes. Moreover, advances in electron

microscopy has revealed structures of transcription factors and transcription complexes, including Pol II in its pause- and elongation complex[27,28,39,40]. Particularly, RNA sequencing methods have been crucial in quantifying RNA molecules in the cell, and more recently, allowed robust measures of genome-wide changes in nascent RNA production. Techniques that directly measure RNA synthesis differ in their biochemical approaches to capture transcripts from the total pool of RNA, and therefore, have different abilities to track distinct RNA species, such as pre-mRNAs, divergent transcripts **[G],** or eRNAs. In this Review, we compare and contrast some of these key strategies, discussing their strengths and limitations, and how their usage in combination can provide a holistical view of regulation throughout the transcription cycle. We begin with a description of the conceptual differences between nascent RNA sequencing methods, and outline complementary imaging techniques. We then discuss the technologies best suited to addressing particular biological and mechanistic questions while highlighting the discoveries made with these methods.

Methods to track nascent RNA synthesis

Methods that investigate RNA synthesis biochemically enrich for nascent or newly synthesized RNA species from the total pool of cellular RNA. In all of these techniques, isolated transcripts are reverse transcribed, ligated to adapters, deep sequenced, and after initial quality filtering and trimming, mapped against the reference genome. As a result, genome-wide sequencing of nascent or newly synthesized transcripts provides rich datasets, including density profiles of transcribing Pol II molecules across the genome, coordinates of transcription start sites or Pol II active sites at nucleotide resolution, and measures of gene and enhancer transcription. Albeit many of the library preparation and data analysis steps are shared between RNA-seq methodologies, distinct approaches differ remarkably in their ability to enrich for the nascent transcripts. Consequently, the method of choice defines which RNA species or steps of the transcription cycle can be analysed and affects the stringency and resolution of the generated data.

The various strategies to enrich for nascent RNA include isolation of chromatin-associated RNA (caRNA)[41–44] , Pol II–associated RNA[45,46], small capped RNA[47,48], recently synthesized RNA[49,50], or RNA from elongation-competent Pol II complexes[48,51,52] (Figure 2). Most of these methods can reliably discern changes in gene expression; however, they differ considerably in their sensitivity to detect different steps of transcription, spatiotemporal resolution, and abilities to identify distinct RNA species, such as eRNAs, divergent transcripts, un-spliced intermediates, and other unstable non-coding RNAs. While genome-wide nascent RNA sequencing methods can precisely map molecular-level regulation of Pol II at genes and enhancers[17,18,53–56], they do not yet distinguish cell-to-cell variation or provide spatial maps of trancription within the nucleus. By contrast, imaging-based methods have limited genomic resolution and ability to discern steps of transcripton, but they can quantify transcript production in real-time in the 3D space of the nucleus, tissue or organism[57]. Below, we summarize distinct sequencing and imaging methods for tracking RNA synthesis and emphasize key steps that determine the sensitivity, specificity, and resolution of each method (Figure 2).

Isolating chromatin-associated RNAs.

The simplest approach for isolating nascent RNA uses strong washes to separate RNA present on chromatin from other RNAs in the cells, thus relying on the stability of the association between nascent transcripts and polymerases during salt fractionation of chromatin (Figure 2a)[41–44,58]. Because cytoplasmic and nucleoplasmic RNAs are more abundant and stable than nascent RNAs, enriching for caRNAs significantly improves the dynamic range for detecting changes in gene and enhancer synthesis. Certain mature RNAs, however, are also stably associated with chromatin, and therefore are captured in caRNA-seq. These mature RNAs include long non-coding RNAs such as *Xist*, which coats and inactivates one of the two X chromosomes in female mammals, as well as small nuclear RNAs (snRNAs), including spliceosomal U1, U2, U4, U5, and U6. To further increase sensitivity and specificity for nascent RNAs associated with Pol II, Nascent Elongating Transcript Sequencing (NET-seq)[45] was developed in yeast and uses immunoprecipitation to capture Pol II-associated RNAs (Figure 2b). NET-seq shares many biochemical and data-analytical approaches with chromatin immunoprecipitation (ChIP-seq) **[G]** and RNA immunoprecipitation RIP-seq **[G]**[59], which has contributed to it being readily adopted by the transcription field. NET-seq was subsequently adapted for mammalian cells (mNET-seq); in this approach, Pol II immunoprecipitation was performed with antibodies against the post-translationally modified CTD of Pol II to identify regulatory status of the transcription complex at distinct regions of genes[46,60]. The originalNET-seq enriched for nascent RNAs by removing most non-nascent RNAs, as well as Pol I and Pol III associated transcripts. However, certain non-nascent RNAs that stably associated with Pol II remained, including the notable example of snRNAs[60]. Removal of non-nascent RNAs was improved by use of the detergent empigen, which dissociated splicing intermediates that bound the spliceosome, as well as miRNAs that associated with Microprocessor [61]. Comparing these two versions of the mNET-seq protocol provided valuable information about co-transcriptional splicing[60,61].

To specifically identify transcription start sites (TSSs), Start-seq enriches for chromatin-associated capped RNA-species[47]. The selection of capped RNA is conducted by enzymatic degradation of caRNAs that are not protected by the 5′-cap, such as rRNA or unshielded mRNA fragments generated during preparation of the sequencing library. Retaining capped caRNAs that are smaller than 80 nt in length, further selects for transcripts that are undergoing initiation, pausing, or early elongation. With paired-end sequencing **[G]**, these short, capped RNAs provide high-resolution information both on TSSs (5′-end of each read), and active sites of transcription (3′-end of each read) at genes and enhancers[47].

Isolating RNAs from transcriptionally-competent Pol II.

Unlike caRNA-seq and NET-seq, run-on assays specifically capture RNAs that are undergoing synthesis. The run-on experiment starts by placing the cells on ice, which stops Pol II from transcribing. Subsequently, the cells are permeabilized, or nuclei or chromatin isolated, to remove endogenous nucleotides and enable labelled NTPs to reach Pol II complexes at the chromatin. Traditional ‘run-on’ reactions used radio-labelled NTPs[62], which were detected at transcripts by hybridization to complementary DNA sequences[62,63]. With current high-throughput sequencing techniques, RNAs with incorporated labelled

nucleotides can be mapped across the genome[51], and the initiating base and active site identified at nucleotide resolution[4,48,52].

Like caRNA-seq, run-on reactions rely on the strong interaction between transcriptionally-engaged Pol II and DNA. During the run-on reaction, anionic detergent sarkosyl is used to remove proteins, including pausing factors, from chromatin, enabling transcription-competent Pol II to proceed along the genome *in vitro*. The addition of sarkosyl is particularly important to reactivate paused Pol II complexes, and therefore, to detect nascent RNAs in the pause complexes[64]. Since run-on techniques rely on incorporation of nucleotides, only transcription-competent polymerases are detected. Therefore, Pol II in initiation complex remains undetectable with nuclear run-on techniques[64]. Moreover, sarkosyl cannot rescue Pol II from a backtracked state (reviewed in Ref. 65), which occurs when Pol II moves back in the DNA template, for example as a result of misincorporation of a nucleotide, displacing the 3′-end of the nascent transcript. The first genome-wide adaptation of the run-on reaction was Global Run-On sequencing (GRO-seq), which labeled nascent transcripts with 5-bromouridine 5′ triphosphate (BrUTP), and immunopurified the transcripts using an antibody against BrUTP[51]. Precision Run-On sequencing (PRO-seq) refined this approach to generate nucleotide-resolution maps of active transcription. In PRO-seq, biotin-11-NTPs are incorporated into a competent Pol II's active site, and due to their bulkiness, only a single nucleotide is added to the transcript[48]. When using all four biotinylated nucleotides, which have similar rates of incorporation[48], all transcripts can be captured without biasing for presence for certain nucleotide sequences. After the run-on, biotinylated nascent transcripts are isolated using streptavidin-coated magnetic beads, providing nucleotide-resolution profile of the active sites of transcription (Figure 2c)[48]. The sensitivity of run-on methodologies is provided by the polymerase-catalyzed incorporation of a single biotinylated nucleotide to the active site of each nascent transcript, followed by and three biotin-affinity purifications to isolate nascent transcripts during library preparation. This results in very low background, and a nearly a million-fold dynamic range estimated from various data-sets[48,66]. While the *in vitro* aspect of run-on reactions may cause some concern that the read-out is distorted, run-on methods show agreement with permanganate IP-seq[48,67] (PIP-seq) **[G]** ChIP-seq[64], ChIP-exo[68], mNET-seq[46], and Start-seq[69] in that promoter-proximal pausing co-localizes genomewide[46] and occurs predominantly at the nucleotide preceding cytosine in run-on[52,70].

By using chromatin as a starting material for run-on reactions, a variant method, Chromatin Run-on sequencing (ChRO-seq), captures nascent RNAs from tissue samples that are not readily amenable to nuclei isolation protocols or have experienced RNA degradation during handling or storage[71]. Both GRO-seq and PRO-seq have been adapted for nucleotide-resolution identification of TSSs by enzymatic cap selection, similar to that used in Start-seq. These variants are known as GRO-cap and PRO-cap[4,48]. The PRO-cap protocol has recently been further modified to coordinate analysis of the initiating nucleotide, capping status, and length of RNA molecules (CoPRO)[52]. This simultaneous identification of the initiating base and active site of each nascent transcript allows systematic investigation of connections between transcriptional initiation and elongation, whereas capping status provides mechanistic insight into early co-transcriptional processes. We anticipate that

improvement of long-read sequencing techniques, which would generate sequencing reads that span multiple exons, will allow development of similar strategies to reach beyond the early coding sequences to map other co-transcriptional processes, such as termination and splicing.

Isolating metabolically labeled RNAs

In metabolic labeling, cells are grown in medium supplemented with modified cell-permeable nucleosides, such as 4-thiouridine[50,72–79] (4sU), bromouridine[80] (BrU), or 5-ethenyluridine[81,82] (EU), which are incorporated into nascent RNAs by active polymerases. Importantly, these nucleotides lack 5′ triphosphates, and are thus processed by the nucleoside salvage pathway, whereby U is converted to UTP in a series of enzymatic reactions[83], which explains the need for extended labeling times in cells (generally 10–60 minutes). In comparison, genome-wide run-on reactions start with isolated nuclei, permeabilized cells or chromatin, which allows NTPs to reach the active site of Pol II from the onset of the run-on reaction[48,51]. Metabolic labeling does not allow for tracking RNA synthesis at nucleotide resolution, but it can measure RNA synthesis in living cells, which allows following lifetime and turnover of RNAs (Figure 2d). Recently, advances have been made in shortening labeling with 4sU to 5 minutes in Transient Transcriptome sequencing (TT-seq)[50] and TimeLapse-seq[79], which has increased detection of newly synthesized RNAs. In TT-seq, metabolically labeled RNAs are enriched by using sonication to separate fragment RNA followed by immuno- or affinity-purification, then compared to unlabeled or total RNA[50]. Alternatively, RNAs can be chemically treated to convert incorporated 4sU to cytosine, creating single-nucleotide mutations that are quantified after deep sequencing[76–79]. Chemical conversion does not require isolation of labeled RNA, and compares nascent and non-nascent RNA in the same sample. This direct analysis of the levels of nascent or newly synthesized RNA (converted nucleotides) to unlabeled RNA allows estimates of average RNA production and decay rates genome-wide[76,78,79], as long as background rates of 4sU incorporation and sequencing errors are taken into account[77]. Besides allowing incorporation of labeled nucleotides in cultured cells, metabolic labeling of RNA has also been demonstrated in living animals[81,82].

Imaging-based methodologies.

Imaging-based methods have limited genomic resolution and capacity to identify regulatory steps, but can quantify transcript production in real-time in the 3D space of the nucleus, tissue or organism (reviewed in Ref. 84). Nascent transcripts can be imaged with techniques such as fluorescent *in situ* hybridization (FISH), in which labeled oligonucleotides are hybridized with transcripts (Fig 3A)[85–88]. Using these methods in mouse cells from parents of two different strains can provide information about the effects of genome variation in bursting within the same cell[89]. Alternatively, transcripts can be engineered to encode hairpin structures that are recognized *in vivo* by tagged cognate binding proteins, such as GFP-coupled bacterial coat protein MS2 (Fig 3B)[90,91]. In these techniques, the site of nascent RNA synthesis in the nucleus is identified as the brightest spot of transcript signal[90,92,93], by labeling the locus producing the transcript[94], or by targeting fluorophores specifically to introns that are present in short-lived pre-mRNA species[57,95,96]. To visualize single unspliced pre-mRNA molecules, several fluorophores must hybridize at each

molecule of a transcript's intron. Monitoring the synthesis of pre-mRNAs, one molecule at the time, can then be used as a proxy for elongating Pol II molecules at the gene. This transcript-by-transcript approach has revealed Pol II dynamics at single model loci. Together with analyses of factor dynamics at regulatory site[97,98] (reviewed in Ref. 99), transcription-driven mobility of chromatin[92,100], and enhancer-mediated gene activation[101–103], single molecule imaging of nascent RNA has elucidated kinetics of TF binding and chromatin regulation. Moreover, novel targeting and multiplexing strategies have begun to reveal chromatin features and RNA synthesis simultaneously at multiple endogenous loci[57,100], setting the stage for visualizing transcriptional programs and genome regulation within individual cells.

Different methodologies provide complementary information.

Each different approach that measures nascent RNA has unique strengths and limitations (Table 1). While sequencing mature mRNAs reliably measures steady-state levels, only nascent RNA-seq techniques enable robust quantification of changes in RNA synthesis and detect unstable RNA species. These distinct nascent techniques were designed to answer different questions about transcriptional regulation and co-transcriptional processses, and so comparing their findings provides a detailed look at many stages of gene regulation. Of the various nascent RNA-seq methods, PRO-seq[48] provides a direct measurement of positions of transcription complexes across the genome, enabling mechanistic studies of regulation of Pol II. On the other hand, metabolic labeling, such as TT-seq[50] and SLAM/TimeLapse-seq[76,79], measures the RNAs synthesized per unit of time. Generally, polymerase densities detected with mNET-seq and *in vitro* Run-On seq (RO-seq) techniques quantitatively agree across genes. However, mNET-seq can detect all chromatin-associated Pol II molecules, whereas RO-seq methods only detect active RNA in transcriptionally-competent complexes. The requirement in RO-seq for adding nucleotides *in vitro* excludes detection of Pol II in pre-initiation complexes[64], and is insensitive for stalled complexes, such as backtracked Pol II[19,104]. Reduced detection of initiation complexes is present also in caRNA and mNET-seq techniques due to mappability limit of short (<20 nt) reads. Although RNAs associated with newly initiated elongation complexes are efficiently captured biochemically in caRNA, NET-seq and RO-seq techniques, due to mappability limit of short (<20 nt) read, they are less likely to uniquely align to the reference genome. However, GRO-seq[51], and, length extended run-on sequencing (leChro-seq)[68] allow generation of longer transcripts during the *in vitro* transcription, which improves mappability of short transcripts and those that arise from repetitive genomic regions.

Metabolic labeling also tracks nascent or newly synthesized transcripts, but long labeling times used in traditional experiments causes poor resolution for measuring nascent transcription. Fragmentation of RNA and chemical conversion of labeled nucleotides have improved the positional resolution of the active site of RNA synthesis in metabolic labeling techniques by allowing selection of labeled fragments of transcripts after as little as five minutes of nucleotide incubation[50,79]. Moreover, metabolic labeling does not efficiently label nascent RNAs in stationary or slow moving Pol II, and thus is far less sensitive in detecting promoter-proximal pausing than PRO-seq and NET-seq (Figure 2). Isolation of RNA with antibodies to Pol II provides additional information about the status of Pol II. For

example, mNET-seq has provided maps of Pol II's post-translational modifications genome-wide[46]. While caRNA-seq is the least sensitive method for capturing nascent transcripts, it is straightforward to combine with methods that investigate co-transcriptional processes[42,105–107].

Nascent RNA sequencing methods have been adapted for genome-wide identification of enhancers and for analyses of promoter and enhancer architecture. Particularly, high-resolution and high-sensitivity maps of transcription initiation at putative enhancers have been obtained with GRO-cap, PRO-cap, and Start-seq, which identify far more TSSs than prior studies of capped RNA 5′ ends[4,18,68,108–110]. The distinct cap-selection methods use slightly different strategies to degrade un-capped and to enrich short capped RNAs for library preparation. Although these cap selection strategies highly correlate[48], the usage of different enzymes could generate differences in cap-selection datasets. Moreover, mis-annealing of reverse-transcript primer has been suggested to cause occasional mis-calling of pause-sites in any data that relies on reverse-transcription from the 3'-adaptor[108]. This mis-calling of inidividual pause sites is minor in high-quality datasets[111]. Using several nascent RNA methods together can therefore provide an integrated view of active transcription, co-transcriptional processes, and the half-lives of RNAs. Furthermore, integrating analyses of nascent transcription with mapping of chromatin states and TF binding reveals mechanistic regulation at different steps, such as: promoter or enhancer remodeling, assembly of the PIC, initiation of transcription, Pol II pausing and entry into productive elongation, and termination (Figure 1). Furthermore, biochemical and computational steps have been adjusted to analyze transcripts at early coding sequences, splice-junctions, and polyadenylation and cleavage sites to measure co-transcriptional processes, such as RNA capping, splicing, and cleavage.

Regulating the transcription cycle

Mechanistic regulation of Pol II at distinct steps of transcription underlies expression of individual genes, and coordinates transcription programs. In this section, we will describe different stages of the transcription cycle (Figure 1) and how nascent RNA methods have enhanced our understanding of the mechanistic regulation of the transcription steps and the co-occurring RNA processing (Table 2).

Gene activation is defined by initiation and promoter-proximal pausing.

Early experiments in bacteria and yeast suggested that the main rate-limiting step of transcription is recruitment of Pol II to promoters. In this model, gene activity is primarily controlled by chromatin state and PIC assembly. Promoter-proximal pausing in eukaryotes was characterized at *Drosophila* major heat shock genes[19,20,112], and similar analyses revealed human c-myc and c-fos genes appeared to have a similar promoter-proximal pausing[113–115]. This mode of regulation was initially considered by the field to be a regulatory step that was specific to a few highly-inducible genes; however, studies tracking nascent RNA synthesis, together with other methods, including genome-wide ChIP- and PIP-seq[67], demonstrated that promoter-proximal pausing is widespread throughout metazoan genomes[17,18,22,23,44,46,51,55,67,116–118] (reviewed in Refs. 24 and 65). Furthermore, many

transcriptional programs, such as tissue morphogenesis in *Drosophila* and heat shock response in mammals, are coordinated by pause-release[17,18,21,55,119–124]. However, some programs, including response to estrogen signaling, do primarily drive Pol II recruitment[53,118,125]. Perturbation of Pol II recruitment with triptolide and inhibition of promoter-proximal pause-release with flavopiridol confirmed that these two processes are the main rate-limiting steps for gene transcription. Triptolide interferes with transcription initiation by preventing TFIIH helicase from melting the DNA strands, which causes a genome-wide block in transcription initiation[126,127]. Flavopiridol, in turn, inhibits CDK9, preventing the release of Pol II from the promoter-proximal paused[128] (Figure 1). A GRO-seq time-course after flavopiridol treatment demonstrated that 95% of mouse genes require CDK9-dependent pause-release, even if pausing is not apparent in steady-state[129].

Polymerase recruitment and pausing are inextricably linked. Structural modeling of DNA-bound PIC, Mediator, and promoter-proximal paused Pol II indicated that pausing within 50 nucleotides of the TSS blocks new initiation owing to steric hindrance[70] (Figure 4A). A recent CoPRO study demonstrated the endogenous relevance of this work by measuring the distance of the pause site from the inititating base genome-wide, finding that pausing nearly always occurs within 50 nt of the TSS in human cells (Figure 4B)[52]. Combined usage of mNET-seq and TT-seq demonstrated that new initiation requires pause release of Pol II[70]. These results together strongly suggest that promoter-proximal pausing precludes a new round of initiation (Figure 4C).

Molecular events that trigger Pol II to switch from a paused state to an elongating complex have been biochemically characterized (reviewed in Ref. 29) and structurally resolved in context of the Pol II complex by cryo-EM **[G]**[27,28]. In the Pol II pause complex, NELF binding appears to facilitate the RNA-DNA hybrid adopting a tilted conformation that prevents NTPs accessing the Pol II active site[28]. The structure further suggests that during the switch to productive elongation, PAF binding displaces NELF, thus allowing NTPs to enter[27]. PRO-seq and mNET-seq measurements of endogenous Pol II pause sites have demonstrated Pol II pausing most likely occurs immediately prior to incorporation of a cytosine, which is the least abundant nucleotide, thus potentially slowing Pol II to permit NELF binding[52,70]. Interestingly, GC-rich sequences are prevalent at active promoter-proximal regions with pause sites located >36 nt from the start nucleotide. In comparison, Pol II pausing more proximal to the TSS (<35 nt from the TSS) is generally found at non-protein coding regulatory elements, such as enhancers, and contain less Pol II pausing signal[52].

Initiation and pausing are regulated at enhancers.

Like promoters, enhancers undergo widespread transcription, as first seen at the β-globin locus [130,131], and since demonstrated genome-wide in many cell types[4,18,53,54,109,110,125] (reviewed in Ref. 132). Nascent transcription assays have identified Pol II pausing also at enhancers, in fact the similar divergent pattern of transcription serves as a robust marker for *de novo* identification of transcribed enhancers genome-wide[4,18,68,69,109]. The mechanisms and importance of Pol II pausing at enhancers, however, remain largely uncharacterized. At enhancers, Pol II has been shown to exhibit short-lived pausing that is regulated by DSIF,

but unlike at promoters, this pausing is not sensitive to NELF knockdown, as seen by Start- and PRO-seq[69]. However, some enhancers have been shown to bind NELF upon gene activation, suggesting that eRNAs may 'sponge' NELF from promoters[133]. Similar steps of transcription regulation at genes and enhancers are further supported by the rapid transcriptional response to heat shock, which causes a genome-wide increase in Pol II density at promoter-proximal as well as enhancer-proximal pause sites[18,52]. The coordination of regulatory events at genes and their functionally connected enhancers remain to be characterized, however, release of paused Pol II from promoter-proximal regions can be stimulated by TF-binding to enhancers[134,135], which suggests that CDK9 may reach gene promoters *via* distal regulatory regions.

Bursting requires initiation and pause-release.

Imaging of nascent RNA production in individual cells has provided evidence that transcription occurs in discontinuous bursts. These bursts were first observed in EM-images of chromosomal spreads[136], then demonstrated *in vivo* in a variety of organisms from bacteria to human[21,83,137,138]. The exact mechanisms that control bursting remain elusive, and the occurrence and kinetics of bursting likely depend on the organism and the regulatory state of the gene[139]. Single molecule imaging of pre-mRNAs from yeast to humans has estimated that a burst is comprised of ~2–100 transcribing Pol II molecules[96,139,140]. The probability of a burst occurring (burst frequency) positively correlates with accessibility and priming of the promoter[101], and promoters with TATA and Initiator elements tend to have increased burst size[89]. Bursting frequency can be increased by forced enhancer connection[101] and inducing histone acetylation at enhancers *via* targeted recruitment of dCas9-p300[141]. TFs can influence burst frequency by modulating the chromatin state of promoters and enhancers, but they can also control burst duration, likely by directly activating Pol II machinery at gene promoters[101].

Many models of transcriptional bursting suggest that the rapid switching between on and off states is defined at a single regulatory step[85,96] (reviewed in Ref. 142). However, imaging-based methodologies rely on labeling introns in pre-mRNAs and therefore cannot distinguish between Pol II recruitment and initiation and the release of promoter-proximal paused Pol II to productive elongation (Figure 3). To send convoys of Pol II into productive elongation, the 'single step' of activation in metazoan species must constitute rapid rounds of coupled Pol II recruitment, initiation, and pause-release[70]. Recent computational modeling of bursting that used intron fluorescence in conjunction with Pol II ChIP-seq suggests that Pol II pause release, not recruitment, causes a burst to initiate[143].

Stability of promoter-proximal pausing.

The interplay between Pol II pausing, premature termination, and release into productive elongation has been a point of ongoing debate and investigation. Nascent RNA sequencing measurements[22,70,129,144] show highly varied pause duration, likely owing to distinct promoter sequences, chromatin accessibility, and rate of pause escape. For example, studiesthat inhibited transcription inititation with triptolide used a GRO-seq time-course to track Pol II as it escaped from promoter-proximal pausing, and found that the pause stability ranges genome-wide from 2.5 to 20 minutes (median = 6.9 minutes)[22,129,144]. Another

study, which modeled Pol II pausing, termination, and release into productive elongation by compaing mNET-seq and TT-seq, found similar pause durations[70]. These datasets agree well with an imaging and biochemical study that so far provides the only direct measure of the promoter-proximal Pol II half-life, quantifying Pol II dynamics at the *Hsp70* gene in polytene chromosomes of *Drosophila*. In its non-induced state, the *Hsp70* gene had a 5 minute Pol II half-life at the promoter, but upon heat-activation, the release of paused Pol II into productive transcription increased ~100 fold[145], corresponding to an estimated three second pause half-life in active state. Of note is that premature termination did not decrease upon heat-shock[145], demonstrating that the heat shock induced increase in transcription is regulated by an increase in the rate of released of paused Pol II into productive elongation rather than by an anti-termination mechanism.

Recently, other methodologies have estimated pause durations to be shorter than reported by nascent RNA sequencing studies. In one study, the kinetics of GFP-tagged Pol II were measured *en masse*. By photobleaching regions of nuclei and measuring the rates at which unbleached GFP-Pol II recovered the fluorescence, the Marteijn group deduced rates at which Pol II can be freed from chromatin[146]. This bulk visualization of Pol II, however, cannot discern whether the measured Pol II population came from promoters, gene bodies, enhancers, or even if it was bound to DNA. To overcome these limitations, the study used different drugs and computational modeling, arriving to a conclusion that 10% of Pol II molecules have a half-life of 2.4 seconds, 23% have a half-life of 42 seconds, and the rest have longer residence times in the genome, likely because they are engaged in pausing or gene body transcription[146]. Similarly, a single-molecule foot-printing experiment concluded that 68% of genes have pause half-lives up to 5 minutes[147], which is in rough agreement with triptolide inhibition and sequencing measurements[129]. This footprinting assay used methylation to track if DNA sequences were protected by proteins and used modeling to determine if they consituted the PIC or paused Pol II. However, methylation in this assay was performed with ~30-minute enzymatic treatments at physiological temperatures[147], during which time Pol II could escape pausing. Based on various sequencing and imaging studies conducted in distinct laboratories, we reiterate that the half-life of paused Pol II may range from seconds to many minutes, and that the kinetics of Pol II at the pause site are influenced by DNA and RNA sequence, chromatin environment and level of gene activity. In any case, the reported pause durations are long relative to the normal average elongation rate of Pol II (0.024 sec per nucleotide)[49,118,129].

Co-transcriptional processes

Releasing Pol II into elongation is only the beginning: co-transcriptional processes such as 5′ capping, intron splicing, and polyadenylation are essential for productive transcription, as well as for RNA stability and trafficking (Figure 1). Beyond being coupled to Pol II elongation, these co-transcriptional processes influence the efficiency of translation, and can create different transcript isoforms by alternative splicing or polyadenylation (reviewed in Refs. 148 and 149). Recent studies have begun to identify mechanisms of RNA processing, as well as elucidate roles for RNA epi-transcriptomic marks. RNA editing alters individual nucleotides, which can cause changes in splicing, stability, and the protein isoform production (reviewed in Ref. 150). Epi-transcriptomic marks, which include RNA

methylation, can alter RNA stability and translation efficiency (reviewed in Ref. 151). Given the importance of such modifications to proper RNA function, understanding where and when these processes occur relative to transcription can offer insights into whether these processes are regulated together, or as independent steps.

5′ capping of RNA occurs prior to or during promoter-proximal Pol II pause.

Shortly after transcription initiation, a 5′ inverted methyl-guanosine cap is added to the 5′-most nucleotide, which provides protection from exonucleases. Early run-on studies using chain-terminating biotin-dNTPs identified that capping occurs shortly after transcription initiation at two *Drosophila melanogaster* heat shock genes[20]. These run-on reactions were paired with de-capping assays that demonstrated that 5'-capping predominantly occurred at or before promoter-proximal pause sites[20]. Additional studies showed capping enzymes bind Ser5-phosphorylated Pol II CTD[152] and interact with DSIF[153,154]. Recently, CoPRO (Figure 2C) was used to investigate genome-wide connections between initiation, pausing, and capping in human cells[52]. In this study, the initiating base, active site, and capping status of individual transcripts as short as 18 nt were measured, demonstrating that capping begins when the 5'-end of the RNA emerges from Pol II exit channel, which corroborated earlier work on individual genes[20]. Capping of RNAs was further compared at pause sites that have either short (20–32 nt) or long (33–60 nt) distance to the TSS, demonstrating a connection between pausing and capping, and suggesting a regulatory role for sequences that influence the pause distance[52]. Promoter-proximal pausing acts as a quality control checkpoint[22] and as a mechanism for synchronous activation of genes during development or in response to signaling[17,18,21,24,55,124]. The distance between initiation and pausing may, therefore, add another dimension for coordinated gene-regulation, providing different opportunities for early RNA processing and Pol II CTD modifications, resulting in preferential recruitment, maturation, and release of Pol II, or folding and modification of the RNA.

Mechanisms of co-transcriptional splicing in yeasts and mammals.

During splicing, non-coding introns are removed from exons by the spliceosome, and these events largely occur co-transcriptionally (reviewed in Refs. 155 and 156). Splicing components have been visualized on nascent RNAs by electron microscopy[157] and ChIP studies have shown them to be associated with chromatin[158–160]. One of the most fundamental questions in the splicing field is how splice sites are recognized with a nucleotide precision from the vast variety of pre-mRNA sequences. It has been hypothesized that yeast, which have strong consensus splicing motifs in short introns, use intron definition **[G]** to demarcate splice sites (reviewed in Ref. 161). Mammals, however, have more degenerate motifs, and introns are generally multiple kilobases in length. These introns are therefore hypothesized to be recognized via exon definition **[G]** (reviewed in Ref. 161). Sequencing of nascent RNA has allowed more precise measurements of when and where splicing occurs in comparison to transcribing Pol II (reviewed in Ref. 162).

The most direct study of co-transcriptional splicing was performed in yeast using Single-Molecule Intron Tracking sequencing (SMIT-seq). *S. cerevisiae* yeast have only 250 intron-containing genes, most of which only have one intron. In SMIT-seq, total caRNA is isolated, then intron-containing genes are enriched by using primers specific to their first exons. By

using paired-end sequencing, the position of the polymerase is detected with the 3′ read, and the splicing status is measured with the 5′ read[163]. In *S. cerevisiae*, completed splicing was observed once Pol II had traveled 26 nt downstream of the 3′ splice site, which is the distance required for the spliceosome to access the site, providing strong evidence for intron definition[163]. SMIT-seq in *S. pombe*, which has more introns and more multi-intronic genes[164], demonstrated that transcripts that failed to splice co-transcriptionally were not cleaved and polyadenylated[165]. The coupled transcription, splicing and polyadenylation suggests connections between multiple co-transcriptional events.

SMIT-seq is not readily adapted to most metazoan introns because of their far greater length. Instead, metazoan co-transcriptional splicing has mostly been investigated with indirect methods such as metabolic labeling[75,166], caRNA-seq[42,105], and mNET-seq[46,60]. One form of evidence for exon definition would be if splicing was only detected once the subsequent exon had been fully transcribed. mNET-seq has provided some support for the exon definition model in mammals, because spliced transcripts were not detected, even in the reads where Pol II had transcribed into an exon[60]. However, long-read sequencing is needed to interrogate the Pol II position with respect to completed splicing in mammals. Results from NET-seq and GRO-seq demonstrate that Pol II accumulates at exons due to reduced elongation rate[44,129], suggesting that sequence and chromatin features of exons impact splicing. Interestingly, caRNA-seq demonstrated that transcripts largely remain chromatin-associated until splicing completes, even when those transcripts have already been cleaved and polyadenylated[42,105]. Although spliceosome components are present in non-chromatin regions of nuclei, these results suggest that proximity to chromatin promotes splicing efficiency[42,167]. Further, sequencing RNAs present on the chromatin and in the nucleoplasm demonstrated that constitutive introns were spliced almost entirely co-transcriptionally, whereas alternative introns were often spliced post-transcriptionally[105]. Exon definition for constitutive exons, which have stronger splice sites than alternative exons, therefore occurs efficiently during transcription, whereas additional, non-chromatin factors may be required for recognizing alternative exons, as weak splice sites may take more time to be recognized than strong ones.

RNA editing and methylation can occur co-transcriptionally.

RNA can be modified during and after synthesis, including “editing” of individual nucleotides (reviewed in Ref. 150) and modifications, the most common of which is adenosine methylation at the N-6 position (m6A) (reviewed in Ref. 151). Two groups combined nascent RNA sequencing techniques with immunoprecipitation of m6A and found that this methylation occurs rapidly during transcription[106,168]. Both groups also measured the effect on splicing in cells with reduced levels of the m6A methylase METTL3. While the Darnell group found that alternative splicing is not affected by m6A in caRNA from mouse embryonic cells[106], the Ørom group did detect an m6A-dependence on splicing rate as measured in HEK293 cells by metabolic labeling[168].

During RNA editing, one nucleotide is chemically converted into another (reviewed in Refs. 150 and 169). Both GRO-seq and PRO-seq show that this editing can occur very rapidly as differences between RNA sequence and the underlying DNA sequence can be detected

within 35 nt of the Pol II exit channel, likely occurring in R-loops **[G]**[170]. The most investigated form of editing, adenosine-to-inosine, has been studied with caRNA-seq and shown to affect splicing efficiency in both *Drosophila* and humans[107,171]. More work is needed to determine the timing of competing or synergizing co-transcriptional events.

RNA cleavage precedes transcription termination.

Directly measuring cleaved and polyadenylated RNAs with nascent RNA methodologies is challenging because the polyadenylated transcript is no longer associated with Pol II. However, increased Pol II density, caused by pausing or slowing down at cleavage and polyadenylation sites, is apparent in GRO-seq, PRO-seq, and NET-seq data[44,46,48,51,172]. Knockdown of the CPSF and CstF cleavage factors results in decreased accumulation of Pol II at the cleavage and polyadenylation site (CPS), which indicates a functional connection between Pol II and RNA-processing factors during co-transcriptional RNA cleavage (Figure 5)[46]. Importantly, knocking down cleavage factors and XRN2 also increases Pol II accumulation near the TSS[46], suggesting that many transcripts are cleaved at an early stage of transcription, likely as a part of premature termination. These prematurely cleaved transcripts often use cryptic cleavage sites in the first intron and are targeted for degradation by the nuclear exosome[173]. Alternative cleavage and polyadenylation is widespread in mammals (reviewed in Ref. 174); however, we do not yet have robust methods for studying the co-transcriptional dynamics of this process.

After cleavage and polyadenylation, Pol II continues transcribing, on average 8 kb, but it eventually dissociates from DNA and terminates transcription[34,46,50,51]. Two models for the mechanism of termination have been proposed: allosteric hindrance, in which Pol II interacts with other factors that destabilize it; and the torpedo model, in which an exonuclease ‘torpedo’ chases down Pol II to trigger termination (reviewed in Ref. 34). Termination is slowed upon degradation of the XRN2 exonuclease, which strongly supports the torpedo model (Figure 5)[172]. TT-seq data has been used to identify termination sites by capturing the uncapped RNAs that associate with Pol II after cleavage[50]. On average, genes have four transcription termination sites and the median termination window spans 3 kb. These termination sites are GC-rich, and they overlap with positions where Pol II pauses or slows, suggesting that pausing can aid termination[50]. Conversely, termination may be impaired during cellular stress, resulting in transcriptional readthrough[175,176].

Enhancer RNAs and microRNAs do not appear to use the same cleavage factors as genes. For example, a GRO-seq study showed that eRNAs are cleaved by Integrator[177], which also processes non-polyadenylated snoRNAs[178]. Primary microRNAs form a hairpin as they are transcribed, which is recognized and cleaved by the Microprocessor complex in the nucleus, leaving a characteristic 3’ overhang (reviewed in Ref. 179). Studies using caRNA-seq and mNET-seq found that this processing occurs co-transcriptionally[46,180].

Posttranslational modifications of Pol II and chromatin

Regulation of transcription complexes by the C-terminal domain of Pol II.

Pol II is a multisubunit protein that is regulated by GTFs, positive and negative elongation factors, chromatin environment, and sequence-specific TFs (reviewed in Refs. 9, 29, 181, and 182). ChIP-seq and mNET-seq data largely agree on the distribution of Pol II CTD phosphorylated Ser2, the density of which increases towards the 3′ ends of genes and is greatest at the regions where 3′-end processing and termination occur[46,183,184]. However, unlike ChIP-seq, mNET-seq does not detect significant accumulation of phosphorylated Ser5 CTDs at promoter-proximal regions of genes after normalizing CTD modification levels to the total amount of Pol II[46]. Moreover, mNET-seq reported high enrichment of Ser5 CTD phosphorylation at exons[46,60], a finding that has gained support from mass spectrometry analyses in yeast[185], but has not been detected by ChIP-seq (reviewed in Ref. 10). The upstream splicing RNA intermediates are likely closely tracking with phosphorylated Ser5 Pol II after the 5'splice site cleavage, thereby enriching mNET-seq libraries with RNAs at those positions[37,50]. The discrepancies between ChIP-seq and mNET-seq could be caused by a number of methodological differences. In essence, ChIP-seq detects DNA fragments that are occupied by Pol II after formaldehyde-mediated stabilization of protein-protein and protein-DNA interactions, whereas mNET-seq captures the RNAs that associate with Pol II-chromatin complex without cross-linking, and in the original protocol, without stringent washing[46]. Beyond enriching for distinct chromatin complexes, current NET-seq datasets under-report Pol II that occupies the first ~35 nt from the TSS, likely a consequence of size selection during library, reducing the efficiency of reporting paused Pol II complexes at genes and enhancers[46]. The differences in Pol II CTD phosphorylation obtained by ChIP-seq and mNET-seq need to be clarified to better understand Pol II regulation at initiation, pause-release and splicing.

Interplay of Pol II with local chromatin architecture.

Similar to the CTD code of Pol II, different chromatin modifications also are coupled to transcription. These modifications are suggested to both influence transcription, and to be affected themselves by the progression of Pol II and co-transcriptional processes (reviewed in Ref. 186). For example, histone H3 lysine 4 trimethylation (H3K4me3) is thought to increase transcriptional activity by maintaining an active chromatin state[187,188] and recruiting chromodomain helicase DNA binding protein 1 (CHD1). CHD1 is an ATP-dependent chromatin remodeler that, in turn, maintains trimethylation of H3K36 and H3K4 at transcribed genes, and promotes association of splicing factors and transcription elongation[189]. Intriguingly, co-transcriptional splicing of the first exon increases H3K4me3 at the promoter[190]. Furthermore, ChIP studies indicate that promoter levels of H3K4me3 inversely correlate with the distance to the first exon, and H3K4me3 at promoters decreases upon inhibition of splicing[190]. In a GRO-seq study, the rate of Pol II elongation and its degree of acceleration correlated with the length of the first intron[129]. Introns contain histone H3 lysine 79 dimethylation (H3K79me2)[191] and are less nucleosome-dense than exons[192], and long introns may therefore provide Pol II the opportunity to gain full speed[129] (reviewed in Ref. 181). H3K36me3 is deposited co-transcriptionally at genes and is enriched

in exons[193–196] (reviewed in Ref. 197), where ChIP studies show that it blocks cryptic transcriptional initiation in yeast[196] (reviewed in Ref. 181).

The characteristic patterns of histone modifications across the genome have been used to predict functions of specific regions[198]. In particular, histone modifications and DNA accessibility have been widely used to distinguish between active enhancers and promoters; the presence of histone H3 lysine 27 acetylation (H3K27ac) and the relative enrichment of mono-methylated H3K4 (H3K4me1) over H3K4me3 are commonly taken to be indicative of enhancer function[199]. However, recent studies have questioned whether histone modifications can be used to discern the classes of regulatory elements. A study from our lab identified transcription initiation sites using GRO-cap and categorized transcripts produced from these sites to stable RNAs (i.e., mRNAs) and unstable RNAs (i.e., uaRNA or eRNA) based on cap analysis gene expression (CAGE) **[G]** data[4]. The results showed that H3K4me1 and H3K4me3 are strongly correlated with polymerase occupancy at both promoters and distal enhancers. H3K4me1 was anti-correlated with polymerase density, whereas marks traditionally associated with promoters, such as acetylation of H3K9 (H3K9ac) and occupancy of GTFs, positively correlated with transcription at the locus, whether it was predicted to be an enhancer or promoter (Figure 6)[4]. Recent work from the Adelman lab extended these findings by showing that genomic sequences that have episomal enhancer activity in self-transcribing active regulatory region sequencing (STARR-seq)[200] **[G]** have prominent levels of H3K4me3 and transcription initiation *in vivo* as measured by ChIP-seq and Start-seq[69]. Similarly, using PRO-cap, the Furlong group found that bidirectional promoters were more likely to act as enhancers, and that certain enhancers could function as promoters[201]. Furthermore, blocking transcription using flavopiridol decreases H3K4me3 at enhancers[54]. These results demonstrate that H3K4me3 reflects transcription activity but may not distinguish promoters from enhancers. Taken together, the similarities in chromatin state and transcription factor constellation at promoters and enhancers suggest that regulatory elements may exist along a functional spectrum of promoter-ness and enhancer-ness where the histone modifications reflect transcription activity rather than functionally categorize regulatory elements[202,203] (reviewed in Ref. 132).

Conclusions and future perspectives

Control of RNA synthesis is critical in defining cell types, cellular responses, and organismal functions. Current methodologies can quantify transcribing Pol II complexes at nucleotide resolution with high sensitivity across the genome. The resulting genome-wide profiles of nascent RNA synthesis, coupled with high-resolution maps of factor binding, have uncovered mechanisms that drive gene transcription *via* regulatory steps, including Pol II recruitment and initiation, promoter-proximal pause-release, productive elongation, and termination (Figure 1). Despite recent advances, several seminal questions remain unanswered. Specific factors have been implicated in regulating certain steps of transcription, yet their detailed mechanistic roles, as well as their interplay with co-factors and the transcription machinery, need to be uncovered in detail. Moreover, RNA processing steps occur co-transcriptionally, but the timing and interconnections between splicing, RNA modifications, and Pol II elongation have not been disentangled. Further, understanding gene

regulation requires expanding the mechanistic studies of transcription from genes to transcribed distal regulatory elements, such as enhancers. How genes and enhancers produce fundamentally different transcripts from very similar chromatin architectures, as well as how a promoter establishes directionality to encode stable transcripts in only one direction, remain incompletely understood. It is also unclear how promoters and enhancers each contribute to different steps of transcription, such as Pol II recruitment, initiation, and release from the promoter-proximal pause. Consequently, how transcription is orchestrated in the networks of genes and their distal regulatory elements remain to be fully elucidated, and requires integrative techniques to track nascent RNA synthesis, chromatin connections and factor binding.

Addressing the above questions requires continued and improved use of approaches that monitor the mechanisms by which regulatory signals affect RNA synthesis, binding dynamics of regulatory factors, and chromatin architecture. Both existing and new tools are needed to rapidly perturb specific functions of targeted transcription factors and features of chromatin in living cells. These perturbation approaches include rapid degradation (degron, reviewed in Ref. 204 **[G]**) of TFs or their inhibition with high-affinity and high-specificity drugs or macromolecules, such as peptides or RNA aptamers (reviewed in Ref. 205) **[G]**. The perturbation-triggered changes in Pol II distribution can be instantly assessed by nascent RNA methods, providing insights to the mechanistic role of each factor before secondary effects confound the interpretation of results. To address coordinated execution of transcription and of co-transcriptional RNA processing, techniques that use long-read sequencing and require less starting material will open new avenues. Use of less starting material will expand analyses of transcription regulation to individual cells and cell types, including patient-derived samples, specific developmental stages, and comparison of cells from diseased *versus* healthy tissue. Particularly, the ability to interrogate nascent RNA synthesis in single cells is a prerequisite for understanding cellular heterogeneity, stochastic gene expression, and regulation of cells' responses in multicellular model systems, such as tissues and organisms. All of these *in vivo* analyses should be complemented by detailed atomic resolution structures of the large machines like Pol II complexes, chromatin remodelers and enhanceosomes that perform or regulate transcription. In this regard, cryo-EM is already providing precise structures of Pol II and other large regulatory machineries. In the future, applying cryo-EM to complexes in their native state will provide critical insights into interactions controlling transcription. Finally the dynamics of transcription need to be tracked in living cells, ideally with super-resolution **[G]** methods (reviewed in Ref. 205) to understand the interplay of multiple TFs and elongation factors at enhancers and promoters, and to clarify how transcription is orchestrated *in vivo*.

Acknowledgements

We thank members of the Lis laboratory for insightful discussions. We also thank our reviewers for their invaluable comments. This work was supported by the National Institute of General Medical Sciences grants R01-GM025232 (J.T.L.), F32-GM129904 (E.M.W.) and T32-HD057854 (N.D.T), and the Sigrid Jusélius Foundation (A.V.).

Glossary

CORE INITIATION REGION

Short (~ 60 nucleotide) region at promoter and enhancer TSSs that provides a binding platform for GTFs. At promoters, core promoter is used as a synonym for core initiation region

C-TERMINAL DOMAIN OF POL II RBP1
Contains multiple repeats of seven amino acids, and serves as a flexible binding scaffold for transcriptional regulators. The post-translational modifications of the heptad-repeat greatly influence the regulatory interactions, and therefore, transcriptional processes throughout the cycle

CRYO-EM
Electron microscopy technique that visualizes molecules in cryogenic (-200°C) temperatures. Allows near atomic resolution (<4Å) imaging of complex molecules and molecule complexes in their native conformation without crystallization or embedding of the sample

SUPER-RESOLUTION MICROSCOPY
A collective term for light microscopy techniques that provide higher resolution (< 200 nm) than imposed by the diffraction limit of visible light

ENHANCER RNAS
Short (50 – 2000 nt) non-coding RNAs that are produced by Pol II from enhancers. The production of eRNAs moderately correlate with the functional activity of the enhancer

DIVERGENT TRANSCRIPTION
In metazoan, genes and enhancers drive transcription from two core initiation regions from both strands in opposing directions. On genes, the mRNA-coding strand is termed sense, and the anti-sense strand produces the divergent transcript

ChIP-SEQ
Method in which protein-DNA interactions are stabilized, chromatin is sheared, and fragments with a protein of interest enriched using an antibody. Purified DNA from the enriched fragments is sequenced, providing genome-wide maps of protein localization

RIP-seq
A technique by which RNAs that interact with a given protein can be identified. Utilizes antibody-mediated enrichment of a protein, after which, its interacting RNAs can be isolated, reverse-transcribed, and sequenced

PAIRED-END SEQUENCING
High-throughput sequencing of DNA-molecules from both ends, which provides information from 3’- and 5’-ends of each DNA fragment, and allows more accurate mapping of the reads to the reference genome

PIP-SEQ
Permanganate footprinting detects unpaired thymine in the DNA, providing the exact locations of open transcription complexes (transcription bubble) across the genome

INTRON DEFINITION
Splicing model in which specific sequences that demarcate introns are sufficient for spliceosomes to recognize intron boundaries

EXON DEFINITION
A model by which proteins that bind to exons are required for the spliceosome to to recognize sequences that demarcate introns

R-LOOPS
Three-stranded DNA:RNA hybrid structures formed for example by template DNA, the complement nascent RNA and the non-template single-stranded DNA

CAGE
Cap analyses gene expression technique that measures RNA expression and maps TSS of gene promoters. Provides precise maps of TSSs of genes that produce long-lived transcripts

STARR-SEQ
Method that assays enhancer activity for millions of candidate sequences by cloning them downstream of a reporter gene and upstream of a CPS. Functionally active enhancers drive expression of RNA molecules that contains the candidate sequence

DEGRON
Tool in which genome editing adds a tag that is encodes a protein recognized by the E3 ubiquitin-ligase complex. After addition of a small molecule, the edited factor is inducibly degraded

RNA APTAMER
Structured RNA molecule selected for binding to a factor of interest to disrupt its functions or interactions

References

1. Liu Y, Beyer A & Aebersold R On the dependency of cellular protein levels on mRNA abundance. Cell 165, 535–550 (2016). [PubMed: 27104977]
2. Long HK, Prescott SL & Wysocka J Ever-changing landscapes: Transcriptional enhancers in development and evolution. Cell 167, 1170–1187 (2016). [PubMed: 27863239]
3. Zabidi MA & Stark A Regulatory enhancer-core-promoter communication via transcription factors and cofactors. Trends Genet. TIG 32, 801–814 (2016). [PubMed: 27816209]
4. Core LJ et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet 46, 1311–1320 (2014). [PubMed: 25383968] Identifies unified regulatory elements architecture at genes and enhancers, describing two core initiation regions to coordinate divergent pattern transcription. Maps enhancers across the human genome using their characteristic pattern of unstable, divergent transcription.
5. Scruggs BS et al. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol. Cell 58, 1101–12 (2015). [PubMed: 26028540]
6. Preker P et al. RNA Exosome depletion reveals transcription upstream of active human promoters. Science 322, 1851–1854 (2008). [PubMed: 19056938]
7. Ntini E et al. Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol 20, 923–928 (2013). [PubMed: 23851456]

8. Li W, Notani D & Rosenfeld MG Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet 17, 207–223 (2016). [PubMed: 26948815]
9. Chen FX, Smith ER & Shilatifard A Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol 19, 464–478 (2018). [PubMed: 29740129]
10. Zaborowska J, Egloff S & Murphy S The pol II CTD: new twists in the tail. Nat. Struct. Mol. Biol 23, 771–777 (2016). [PubMed: 27605205]
11. Zaret KS & Carroll JS Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25, 2227–2241 (2011). [PubMed: 22056668]
12. Mayran A & Drouin J Pioneer transcription factors shape the epigenetic landscape. J. Biol. Chem 293, 13795–13804 (2018). [PubMed: 29507097]
13. Schilbach S et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature 551, 204–209 (2017). [PubMed: 29088706]
14. Kadonaga JT Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip. Rev. Dev. Biol 1, 40–51 (2012). [PubMed: 23801666]
15. Conaway RC & Conaway JW An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. Proc. Natl. Acad. Sci 86, 7356–60 (1989). [PubMed: 2552440]
16. Tirode F, Busso D, Coin F & Egly JM Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol. Cell 3, 87–95 (1999). [PubMed: 10024882]
17. Mahat DB, Salamanca HH, Duarte FM, Danko CG & Lis JT Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. Mol. Cell 62, 63–78 (2016). [PubMed: 27052732]
18. Vihervaara A et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. Nat. Commun 8, 255 (2017). [PubMed: 28811569]
19. Rougvie AE & Lis JT The RNA polymerase II molecule at the 5’ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell 54, 795–804 (1988). [PubMed: 3136931]
20. Rasmussen EB & Lis JT In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. U. S. A 90, 7923–7927 (1993). [PubMed: 8367444]
21. Boettiger AN & Levine M Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science 325, 471–473 (2009). [PubMed: 19628867]
22. Henriques T et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell 52, 517–528 (2013). [PubMed: 24184211]
23. Williams LH et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol. Cell 58, 311–322 (2015). [PubMed: 25773599]
24. Vihervaara A, Duarte FM & Lis JT Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet 19, 385–397 (2018). [PubMed: 29556092]
25. Shimotohno K, Kodama Y, Hashimoto J & Miura KI Importance of 5’-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. Proc. Natl. Acad. Sci 74, 2734–8 (1977). [PubMed: 197518]
26. Wu C-H et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev 17, 1402–1414 (2003). [PubMed: 12782658]
27. Vos SM et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature 560, 607–612 (2018). [PubMed: 30135578] Provides atomic-resolution cryo-EM structure of elongation-competent Pol II complex. Describes composition and conformation changes in transcription complex as it proceeds from pause into elongation.
28. Vos SM, Farnung L, Urlaub H & Cramer P Structure of paused transcription complex Pol II–DSIF–NELF. Nature 560, 601 (2018). [PubMed: 30135580]
29. Fuda NJ, Ardehali MB & Lis JT Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192 (2009). [PubMed: 19741698]
30. Fitz J, Neumann T & Pavri R Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. EMBO J 37, e97965 (2018). [PubMed: 29514850]

31. Takeuchi A et al. Loss of Sfpq causes long-gene transcriptopathy in the brain. Cell Rep 23, 1326–1341 (2018). [PubMed: 29719248]
32. Hsin J-P & Manley JL The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev 26, 2119–2137 (2012). [PubMed: 23028141]
33. Tian B & Graber JH Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip. Rev. RNA 3, 385–396 (2012). [PubMed: 22012871]
34. Proudfoot NJ Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352, aad9926 (2016). [PubMed: 27284201]
35. Yudkovsky N, Ranish JA & Hahn S A transcription reinitiation intermediate that is stabilized by activator. Nature 408, 225–9 (2000) [PubMed: 11089979]
36. Pelling C Ribonukleinsäure-Synthese der Riesenchromosomen. Chromosoma 15, 71–122 (1964). [PubMed: 14171171]
37. Miller OL & Beatty BR Visualization of nucleolar genes. Science 164, 955–957 (1969). [PubMed: 5813982]
38. Miller OL, Hamkalo BA & Thomas CA Visualization of bacterial genes in action. Science 169, 392–395 (1970). [PubMed: 4915822]
39. Cramer P, Bushnell DA & Kornberg RD Structural basis of transcription: RNA Polymerase II at 2.8 Ångstrom resolution. Science 292, 1863–1876 (2001). [PubMed: 11313498]
40. Bernecky C, Herzog F, Baumeister W, Plitzko JM & Cramer P Structure of transcribing mammalian RNA polymerase II. Nature 529, 551–554 (2016). [PubMed: 26789250]
41. Henikoff S, Henikoff JG, Sakai A, Loeb GB & Ahmad K Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res 19, 460–469 (2009). [PubMed: 19088306]
42. Bhatt DM et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell 150, 279–290 (2012). [PubMed: 22817891]
43. Weber CM, Ramachandran S & Henikoff S Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol. Cell 53, 819–30 (2014). [PubMed: 24606920]
44. Mayer A et al. Native Elongating Transcript Sequencing reveals human transcriptional activity at nucleotide resolution. Cell 161, 541–554 (2015). [PubMed: 25910208]
45. Churchman LS & Weissman JS Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373 (2011). [PubMed: 21248844] This is the first demonstration of NET-seq, which was performed in S. cerevisiae by immunoprecipitating epitope-tagged Pol II.
46. Nojima T et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. Cell 161, 526–540 (2015). [PubMed: 25910207] Applies NET-seq to human cells using antibodies to Ser5-, Ser2-, and un-phosphorylated Pol II CTD, demonstrating the positions of different Pol II modifications genome-wide.
47. Nechaev S et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 327, 335–338 (2010). [PubMed: 20007866] The establishment of Start-seq, which identified initiating and promoter-proximal pausing positions across the D. melanogaster genome.
48. Kwak H, Fuda NJ, Core LJ & Lis JT Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339, 950–953 (2013). [PubMed: 23430654] Developed Precision Run-On sequencing (PRO-seq) that maps active Pol II positions at single-nucleotide resolution genome-wide.
49. Fuchs G et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol 15, R69 (2014). [PubMed: 24887486]
50. Schwalb B et al. TT-seq maps the human transient transcriptome. Science 352, 1225–1228 (2016). [PubMed: 27257258] Develops Transient Transcriptome sequencing (TT-seq) that uses short metabolic labeling times to capture newly transcribed RNAs in human cells.
51. Core LJ, Waterfall JJ & Lis JT Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848 (2008). [PubMed: 19056941] Developed the first global run-on technique (GRO-seq), and demonstrated widespread promoter-proximal pausing and bidirectional transcription across the human genome.

52. Tome JM, Tippens ND & Lis JT Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. Nat. Genet 50, 1533 (2018). [PubMed: 30349116]
53. Hah N et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell 145, 622–634 (2011). [PubMed: 21549415]
54. Kaikkonen MU et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell 51, 310–325 (2013). [PubMed: 23932714]
55. Duarte FM et al. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev 30, 1731–1746 (2016). [PubMed: 27492368]
56. Booth GT, Parua PK, Sansó M, Fisher RP & Lis JT Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. Nat. Commun 9, 543 (2018). [PubMed: 29416031]
57. Shah S et al. Dynamics and spatial genomics of the nascent transcriptome by Intron seqFISH. Cell 174, 363–376.e16 (2018). [PubMed: 29887381] This imaging technique provides visualization of the synthesis of >10,000 genes in single cells.
58. Wuarin J & Schibler U Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol 14, 7219–7225 (1994). [PubMed: 7523861]
59. Townley-Tilson WH, Pendergrass SA, Marzluff WF & Whitfield ML Genome-wide analysis of mRNAs bound to the histone stem–loop binding protein. RNA 12, 1853–67 (2006). [PubMed: 16931877]
60. Nojima T et al. RNA Polymerase II phosphorylated on CTD Serine 5 interacts with the spliceosome during co-transcriptional splicing. Mol. Cell 72, 369–379.e4 (2018). [PubMed: 30340024]
61. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M & Proudfoot NJ Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol. Cell 65, 25–38 (2017). [PubMed: 28017589]
62. Gariglio P, Buss J & Green MH Sarkosyl activation of RNA polymerase activity in mitotic mouse cells. FEBS Lett 44, 330–333 (1974). [PubMed: 4472294]
63. Smale ST Nuclear run-on assay. Cold Spring Harb. Protoc 2009, pdb.prot5329 (2009).
64. Core LJ et al. Defining the status of RNA polymerase at promoters. Cell Rep 2, 1025–1035 (2012). [PubMed: 23062713]
65. Adelman K & Lis JT Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet 13, 720–731 (2012). [PubMed: 22986266]
66. Min IM et al. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev 25, 742–754 (2011). [PubMed: 21460038]
67. Li J et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol. Cell 50, 711–722 (2013). [PubMed: 23746353]
68. Pugh BF & Venters BJ Genomic Organization of Human Transcription Initiation Complexes. PLOS One 11, e0149339 (2016). [PubMed: 26866362]
69. Henriques T et al. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev 32, 26–41 (2018). [PubMed: 29378787]
70. Gressel S et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. eLife 6, (2017).Demonstrated worth of combining multiple techniques to investigate transcription. Used TT-seq, mNET-seq, and structural modeling to examine the interplay of Pol II initiation and pausing.
71. Chu T et al. Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. Nat. Genet 50, 1553–1564 (2018). [PubMed: 30349114]
72. Kenzelmann M et al. Microarray analysis of newly synthesized RNA in cells and animals. Proc. Natl. Acad. Sci 104, 6164–9 (2007). [PubMed: 17405863]
73. Dölken L et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA 14, 1959–72 (2008). [PubMed: 18658122]

74. Rabani M et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol 29, 436–442 (2011). [PubMed: 21516085]
75. Rabani M et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell 159, 1698–1710 (2014). [PubMed: 25497548]
76. Herzog VA et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods 14, 1198–1204 (2017). [PubMed: 28945705]
77. Jürges C, Dölken L & Erhard F Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinforma. Oxf. Engl 34, i218–i226 (2018).
78. Muhar M et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science 18, 800–805 (2018).
79. Schofield JA, Duffy EE, Kiefer L, Sullivan MC & Simon MD TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods 15, 221–225 (2018). [PubMed: 29355846]
80. Paulsen MT et al. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. Proc. Natl. Acad. Sci. U. S. A 110, 2240–2245 (2013). [PubMed: 23345452]
81. Jao CY & Salic A Exploring RNA transcription and turnover in vivo by using click chemistry. Proc. Natl. Acad. Sci. U. S. A 105, 15779–15784 (2008). [PubMed: 18840688]
82. Akbalik G et al. Visualization of newly synthesized neuronal RNA in vitro and in vivo using click-chemistry. RNA Biol 14, 20–28 (2017). [PubMed: 27801616]
83. Nyhan WL Nucleotide Synthesis via Salvage Pathway. eLS (2001). doi: 10.1038/npg.els.0001399.
84. Lim B Imaging transcriptional dynamics. Curr. Opin. Biotechnol 52, 49–55 (2018). [PubMed: 29501816]
85. Gall JG & Pardue ML Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc. Natl. Acad. Sci. U. S. A 63, 378–383 (1969). [PubMed: 4895535]
86. Bauman JG, Wiegant J, Borst P & van Duijn P A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp. Cell Res 128, 485–490 (1980). [PubMed: 6157553]
87. Singer RH & Ward DC Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. Proc. Natl. Acad. Sci. U. S. A 79, 7331–7335 (1982). [PubMed: 6961411]
88. Vargas DY, Raj A, Marras SAE, Kramer FR & Tyagi S Mechanism of mRNA transport in the nucleus. Proc. Natl. Acad. Sci. U. S. A 102, 17008–17013 (2005). [PubMed: 16284251]
89. Larsson AJM et al. Genomic encoding of transcriptional burst kinetics. Nature 565, 251–254 (2019). [PubMed: 30602787]
90. Yunger S, Rosenfeld L, Garini Y & Shav-Tal Y Single-allele analysis of transcription kinetics in living mammalian cells. Nat. Methods 7, 631–633 (2010). [PubMed: 20639867]
91. Bertrand E et al. Localization of ASH1 mRNA particles in living yeast. Mol. Cell 2, 437–445 (1998). [PubMed: 9809065]
92. Larson DR, Zenklusen D, Wu B, Chao JA & Singer RH Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332, 475–478 (2011). [PubMed: 21512033]
93. Lionnet T et al. A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat. Methods 8, 165–170 (2011). [PubMed: 21240280]
94. Germier T et al. Real-time imaging of a single gene reveals transcription-initiated local confinement. Biophys. J 113, 1383–1394 (2017). [PubMed: 28978433]
95. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A & Tyagi S Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5, 877–879 (2008). [PubMed: 18806792]
96. Tantale K et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat. Commun 7, 12248 (2016). [PubMed: 27461529]

97. Gebhardt JCM et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. Nat. Methods 10, 421–426 (2013). [PubMed: 23524394]

98. Chen J et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 156, 1274–1285 (2014). [PubMed: 24630727]

99. Liu Z & Tjian R Visualizing transcription factor dynamics in living cells. J. Cell Biol 217, 1181–1191 (2018). [PubMed: 29378780]

100. Gu B et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science 359, 1050–1055 (2018). [PubMed: 29371426]

101. Bartman CR, Hsu SC, Hsiung CC-S, Raj A & Blobel GA Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. Mol. Cell 62, 237–247 (2016). [PubMed: 27067601]

102. Fukaya T, Lim B & Levine M Enhancer control of transcriptional bursting. Cell 166, 358–368 (2016). [PubMed: 27293191]

103. Mir M et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes Dev 31, 1784–1794 (2017). [PubMed: 28982761]

104. Adelman K et al. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol. Cell 17, 103–12 (2005). [PubMed: 15629721]

105. Pandya-Jones A et al. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. RNA 19, 811–827 (2013). [PubMed: 23616639]

106. Ke S et al. m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev 31, 990–1006 (2017). [PubMed: 28637692]

107. Hsiao Y-HE et al. RNA editing in nascent RNA affects pre-mRNA splicing. Genome Res 28, 812–823 (2018). [PubMed: 29724793]

108. Danko CG et al. Identification of active transcriptional regulatory elements from GRO-seq data. Nat. Methods 12, 433–438 (2015). [PubMed: 25799441] Identified enhancers and promoters from run-on data by identifying sites of bidirectional transcription.

109. Azofeifa JG et al. Enhancer RNA profiling predicts transcription factor activity. Genome Res 28: 334–344 (2018).Used computational tools to identify enhancers from run-on sequencing data and demonstrate that these data can identify active transcription factor motifs.

110. Kim HS et al. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. Nature 556, 510–514 (2018). [PubMed: 29670286]

111. Shivram H & Iyer VR Identification and removal of sequencing artifacts produced by mispriming during reverse transcription in multiple RNA-seq technologies. RNA 24, 1266–1274 (2018). [PubMed: 29950518]

112. Gilmour DS & Lis JT RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol. Cell. Biol 6, 3984–3989 (1986). [PubMed: 3099167]

113. Krumm A, Meulia T, Brunvand M & Groudine M The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes Dev 6, 2201–2213 (1992). [PubMed: 1427080]

114. Plet A, Eick D & Blanchard JM Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. Oncogene 10, 319–328 (1995). [PubMed: 7838531]

115. Meininghaus M & Eick D Requirement of the carboxy-terminal domain of RNA polymerase II for the transcriptional activation of chromosomal c-fos and hsp70A genes. FEBS Lett 446, 173–176 (1999). [PubMed: 10100637]

116. Muse GW et al. RNA polymerase is poised for activation across the genome. Nat. Genet 39, 1507–1511 (2007). [PubMed: 17994021]

117. Zeitlinger J et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet 39, 1512–1516 (2007). [PubMed: 17994019]

118. Danko CG et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol. Cell 50, 212–222 (2013). [PubMed: 23523369]

119. Adelman K et al. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. Proc. Natl. Acad. Sci 106, 18207–12 (2009). [PubMed: 19820169]
120. Donner AJ, Ebmeier CC, Taatjes DJ & Espinosa JM CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat. Struct. Mol. Biol 17, 194–201 (2010). [PubMed: 20098423]
121. Rahl PB c-Myc regulates transcriptional pause release. Cell 141, 432–45 (2010). [PubMed: 20434984]
122. Escoubet-Lozach L et al. Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. PLoS Genet 7, e1002401 (2011). [PubMed: 22174696]
123. Galbraith MD et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell 6, 1327–39 (2013).
124. Dukler N et al. Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. Genome Res 27, 1816–1829 (2017). [PubMed: 29025894]
125. Wang D et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474, 390–394 (2011). [PubMed: 21572438]
126. Vispé S et al. Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. Mol. Cancer Ther 8, 2780–2790 (2009). [PubMed: 19808979]
127. Titov DV et al. XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat. Chem. Biol 7, 182–188 (2011). [PubMed: 21278739]
128. Chao S-H & Price DH Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo. J. Biol. Chem 276, 31793–31799 (2001). [PubMed: 11431468]
129. Jonkers I, Kwak H & Lis JT Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife 3, e02407 (2014). [PubMed: 24843027]
130. Ashe HL, Monks J, Wijgerde M, Fraser P & Proudfoot NJ Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev 11, 2494–509 (1997). [PubMed: 9334315]
131. Routledge SJ & Proudfoot NJ Definition of transcriptional promoters in the human beta globin locus control region. J. Mol. Biol 323, 601–11 (2002). [PubMed: 12419253]
132. Tippens ND, Vihervaara A & Lis JT Enhancer transcription: what, where, when, and why? Genes Dev 32, 1–3 (2018). [PubMed: 29440223]
133. Schaukowitch K et al. Enhancer RNA facilitates NELF release from immediate early genes. Mol. Cell 56, 29–42 (2014). [PubMed: 25263592]
134. Chen FX et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. Science 357, 1294–1298 (2017). [PubMed: 28860207]
135. Galli GG et al. YAP drives growth by controlling transcriptional pause release from dynamic enhancers. Mol. Cell 60, 328–337 (2015). [PubMed: 26439301]
136. McKnight SL & Miller OL Post-replicative nonribosomal transcription units in D. melanogaster embryos. Cell 17, 551–563 (1979). [PubMed: 113103]
137. Golding I, Paulsson J, Zawilski SM & Cox EC Real-time kinetics of gene activity in individual bacteria. Cell 123, 1025–1036 (2005). [PubMed: 16360033]
138. Chubb JR, Trcek T, Shenoy SM & Singer RH Transcriptional pulsing of a developmental gene. Curr. Biol 16, 1018–1025 (2006). [PubMed: 16713960]
139. Lenstra TL, Coulon A, Chow CC & Larson DR Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. Mol. Cell 60, 597–610 (2015). [PubMed: 26549684]
140. Bothma JP et al. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc. Natl. Acad. Sci. U. S. A 111, 10598–10603 (2014). [PubMed: 24994903]
141. Chen LF et al. Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. Cell Rep 26, 1174–1188 (2019). [PubMed: 30699347]

142. Lenstra TL, Rodriguez J, Chen H & Larson DR Transcription Dynamics in Living Cells. Annu. Rev. Biophys 45, 25–47 (2016). [PubMed: 27145880]

143. Bartman CR et al. Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol. Cell 73, 519–532.e4 (2019). [PubMed: 30554946]

144. Chen F, Gao X & Shilatifard A Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. Genes Dev 29, 39–47 (2015). [PubMed: 25561494]

145. Buckley MS, Kwak H, Zipfel WR & Lis JT Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. Genes Dev 28, 14–19 (2014). [PubMed: 24395245]

146. Steurer B et al. Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. Proc. Natl. Acad. Sci. U. S. A 115, E4368–E4376 (2018). [PubMed: 29632207]

147. Krebs AR et al. Genome-wide single-molecule footprinting reveals high RNA Polymerase II turnover at paused promoters. Mol. Cell 67, 411–422.e4 (2017). [PubMed: 28735898]

148. Moore MJ From Birth to Death: The Complex Lives of Eukaryotic mRNAs. Science 309, 1514–1518 (2005). [PubMed: 16141059]

149. Singh G, Pratt G, Yeo GW & Moore MJ The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem 84, 325–354 (2015). [PubMed: 25784054]

150. Bass BL RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem 71, 817–846 (2002). [PubMed: 12045112]

151. Meyer KD & Jaffrey SR Rethinking m6A readers, writers, and erasers. Annu. Rev. Cell Dev. Biol 33, 319–342 (2017). [PubMed: 28759256]

152. McCracken S et al. 5’-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev 11, 3306–3318 (1997). [PubMed: 9407024]

153. Wen Y & Shatkin AJ Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev 13, 1774–1779 (1999). [PubMed: 10421630]

154. Mandal SS et al. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc. Natl. Acad. Sci. U. S. A 101, 7572–7577 (2004). [PubMed: 15136722]

155. Merkhofer EC, Hu P & Johnson TL Introduction to cotranscriptional RNA splicing. Methods Mol. Biol. Clifton NJ 1126, 83–96 (2014).

156. Naftelberg S, Schor IE, Ast G & Kornblihtt AR Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu. Rev. Biochem 84, 165–198 (2015). [PubMed: 26034889]

157. Osheim YN, Miller OL, J. & Beyer, A. L. RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell 43, 143–151 (1985). [PubMed: 3935315]

158. Görnemann J, Kotovic KM, Hujer K & Neugebauer KM Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell 19, 53–63 (2005). [PubMed: 15989964]

159. Lacadie SA & Rosbash M Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5’ss base pairing in yeast. Mol. Cell 19, 65–75 (2005). [PubMed: 15989965]

160. Listerman I, Sapra AK & Neugebauer KM Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat. Struct. Mol. Biol 13, 815–822 (2006). [PubMed: 16921380]

161. Keren H, Lev-Maor G & Ast G Alternative splicing and evolution: diversification, exon definition and function. Nat. Rev. Genet 11, 345–355 (2010). [PubMed: 20376054]

162. Herzel L, Ottoz DSM, Alpert T & Neugebauer KM Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol 18, 637–650 (2017). [PubMed: 28792005]

163. Carrillo Oesterreich F et al. Splicing of nascent RNA coincides with intron exit from RNA Polymerase II. Cell 165, 372–381 (2016). [PubMed: 27020755]

164. Prabhala G, Rosenberg GH & Käufer NF Architectural features of pre-mRNA introns in the fission yeast Schizosaccharomyces pombe. Yeast Chichester Engl 8, 171–182 (1992).
165. Herzel L, Straube K & Neugebauer KM Long-read sequencing of nascent RNA reveals coupling among RNA processing events. Genome Res 28, 1008–1019 (2018). [PubMed: 29903723]
166. Pai AA et al. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. eLife 6, (2017).
167. Khodor YL et al. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev 25, 2502–2512 (2011). [PubMed: 22156210]
168. Louloupi A, Ntini E, Conrad T & Ørom UAV Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. Cell Rep 23, 3429–3437 (2018). [PubMed: 29924987]
169. Eisenberg E & Levanon EY A-to-I RNA editing — immune protector and transcriptome diversifier. Nat. Rev. Genet 19, 473–490 (2018). [PubMed: 29692414]
170. Wang IX et al. RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. Cell Rep 6, 906–915 (2014). [PubMed: 24561252]
171. Rodriguez J, Menet JS & Rosbash M Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Mol. Cell 47, 27–37 (2012). [PubMed: 22658416]
172. Eaton JD et al. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev 32, 127–139 (2018). [PubMed: 29432121]
173. Chiu AC et al. Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. Mol. Cell 69, 648–663.e7 (2018). [PubMed: 29398447]
174. Tian B & Manley JL Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem. Sci 38, 312–320 (2013). [PubMed: 23632313]
175. Vilborg A et al. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci 114, E8362–E8371 (2017). [PubMed: 28928151]
176. Vilborg A, Passarelli MC, Yario TA, Tycowski KT & Steitz JA Widespread inducible transcription downstream of human genes. Mol. Cell 59, 449–461 (2015). [PubMed: 26190259]
177. Lai F, Gardini A, Zhang A & Shiekhattar R Integrator mediates the biogenesis of enhancer RNAs. Nature 525, 399–403 (2015). [PubMed: 26308897]
178. O'Reilly D et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. Nucleic Acids Res 42, 264–275 (2014). [PubMed: 24097444]
179. Ha M & Kim VN Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol 15, 509–524 (2014). [PubMed: 25027649]
180. Dhir A, Dhir S, Proudfoot NJ & Jopling CL Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. Nat. Struct. Mol. Biol 22, 319–327 (2015). [PubMed: 25730776]
181. Jonkers I & Lis JT Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol 16, 167–177 (2015). [PubMed: 25693130]
182. Venkatesh S & Workman JL Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol 16, 178–189 (2015). [PubMed: 25650798]
183. Hintermair C et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J 31, 2784–2797 (2012). [PubMed: 22549466]
184. Descostes N et al. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife 3, e02105 (2014). [PubMed: 24842994]
185. Harlen KM et al. Comprehensive RNA Polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. Cell Rep 15, 2147–2158 (2016). [PubMed: 27239037]
186. Gates LA, Foulds CE & O'Malley BW Histone marks in the 'driver's seat': Functional roles in steering the transcription cycle. Trends Biochem. Sci 42, 977–989 (2017). [PubMed: 29122461]
187. Vermeulen M et al. Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. Cell 131, 58–69 (2007). [PubMed: 17884155]

188. Klymenko T & Müller J The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep 5, 373–377 (2004). [PubMed: 15031712]

189. Sims RJ et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol. Cell 28, 665–676 (2007). [PubMed: 18042460]

190. Bieberstein NI, Carrillo Oesterreich F, Straube K & Neugebauer KM First exon length controls active chromatin signatures and transcription. Cell Rep 2, 62–68 (2012). [PubMed: 22840397]

191. Huff JT, Plocik AM, Guthrie C & Yamamoto KR Reciprocal intronic and exonic histone modification regions in humans. Nat. Struct. Mol. Biol 17, 1495–1499 (2010). [PubMed: 21057525]

192. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C & Komorowski J Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 19, 1732–1741 (2009). [PubMed: 19687145]

193. Keogh M-C et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123, 593–605 (2005). [PubMed: 16286008]

194. Carrozza MJ et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123, 581–592 (2005). [PubMed: 16286007]

195. Kolasinska-Zwierz P et al. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet 41, 376–381 (2009). [PubMed: 19182803]

196. Venkatesh S et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489, 452–455 (2012). [PubMed: 22914091]

197. McDaniel SL & Strahl BD Shaping the Cellular Landscape with Set2/SETD2 Methylation. Cell. Mol. Life Sci. CMLS 74, 3317–3334 (2017). [PubMed: 28386724]

198. Ernst J & Kellis M Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol 28, 817–825 (2010). [PubMed: 20657582]

199. Heintzman ND et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet 39, 311–318 (2007). [PubMed: 17277777]

200. Arnold CD et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339, 1074–7 (2013). [PubMed: 23328393]

201. Mikhaylichenko O et al. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev 32, 42–57 (2018). [PubMed: 29378788]

202. Andersson R, Sandelin A & Danko CG A unified architecture of transcriptional regulatory elements. Trends Genet. TIG 31, 426–433 (2015). [PubMed: 26073855]

203. Dao LTM et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat. Genet 49, 1073–1081 (2017). [PubMed: 28581502]

204. Natsume T & Kanemaki MT Conditional Degrons for Controlling Protein Expression at the Protein Level. Annu. Rev. Genet 51, 83–102 (2017). [PubMed: 29178817]

205. Ozer A, Pagano JM & Lis JT New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization. Mol. Ther. Nucleic Acids 3, e183 (2014). [PubMed: 25093707]

206. Schermelleh L et al. Super-resolution microscopy demystified. Nat. Cell Biol 21, 72 (2019). [PubMed: 30602772]

207. Petesch SJ & Lis JT Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. Cell 134, 74–84 (2008). [PubMed: 18614012]

208. Mueller B et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev 31, 451–462 (2017). [PubMed: 28356342]

209. Kristjánsdóttir K, Kwak Y, Tippens ND, Lis JT, Kang HM & Kwak H Population-scale study of eRNA transcription reveals bipartite functional enhancer architecture. bioRXiv (2018). doi: https://doi.org/10.110¼26908.

210. Michel M TT-seq captures enhancer landscapes immediately after T-cell stimulation. Mol. Syst. Biol 13, 920 (2017). [PubMed: 28270558]

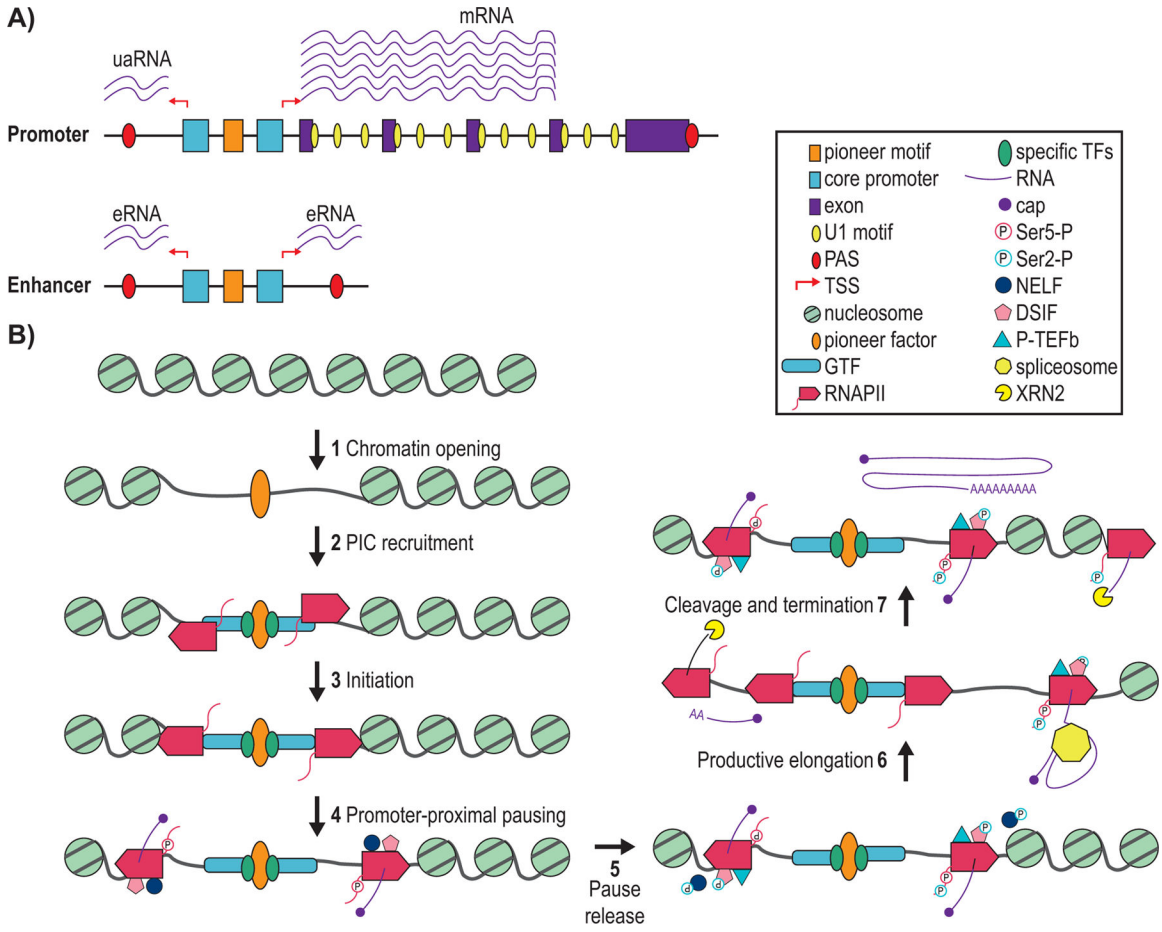

Figure 1 |. The transcription cycle.

a | Typical gene architecture depicting DNA elements that affect transcription and transcript stability. At promoters, binding sites for gene-specific transcription factors are found between two core promoters (i.e., core initiation regions) that drive divergent transcription. The coding strand (right) produces mRNA that is stabilized by the presence of splice sites, and the anti-sense strand (left) produces an unstable upstream antisense RNA. **b** | The transcription cycle consists of the following steps. 1) A pioneer transcription factor binds a specific sequence motif and increases chromatin accessibility. 2) Additional sequence-specific TFs bind near the pioneer factor. Core promoters recruit GTFs and Pol II to form the PIC. 3) The GTF TFIIH unwinds DNA, and Pol II initiates transcription. 4) Pol II undergoes promoter-proximal pausing after transcribing 20–60 nt, and the pause is stabilized by binding of DSIF and NELF. Before or during pausing, Pol II's CTD is phosphorylated at Ser5 and Ser7, and the RNA undergoes 5' capping. 5) Pol II escapes promoter-proximal pausing and enters productive elongation, largely due to CDK9 phosphorylating multiple targets: NELF (ending its interaction with Pol II), DSIF (converting it to an elongation factor), and Pol II Ser2 (which interacts with RNA processing factors). Additional elongation factors, such as PAF1, promote this escape. 6) During productive elongation, co-transcriptional processing, including splicing, RNA methylation, and RNA editing occur. Nucleosomes are removed in highly transcribed genes, while chromatin accessibility increases at moderately transcribed genes[207,208]. 7) The RNA is cleaved and polyadenylated.

After cleavage, Pol II continues elongating, but the nascent RNA lacks a cap and subject to XRN2-mediated degradation, which destabilizes Pol II and contributes to termination. After termination, Pol II can be recycled to new initiation, repeating the transcriptional cycle.

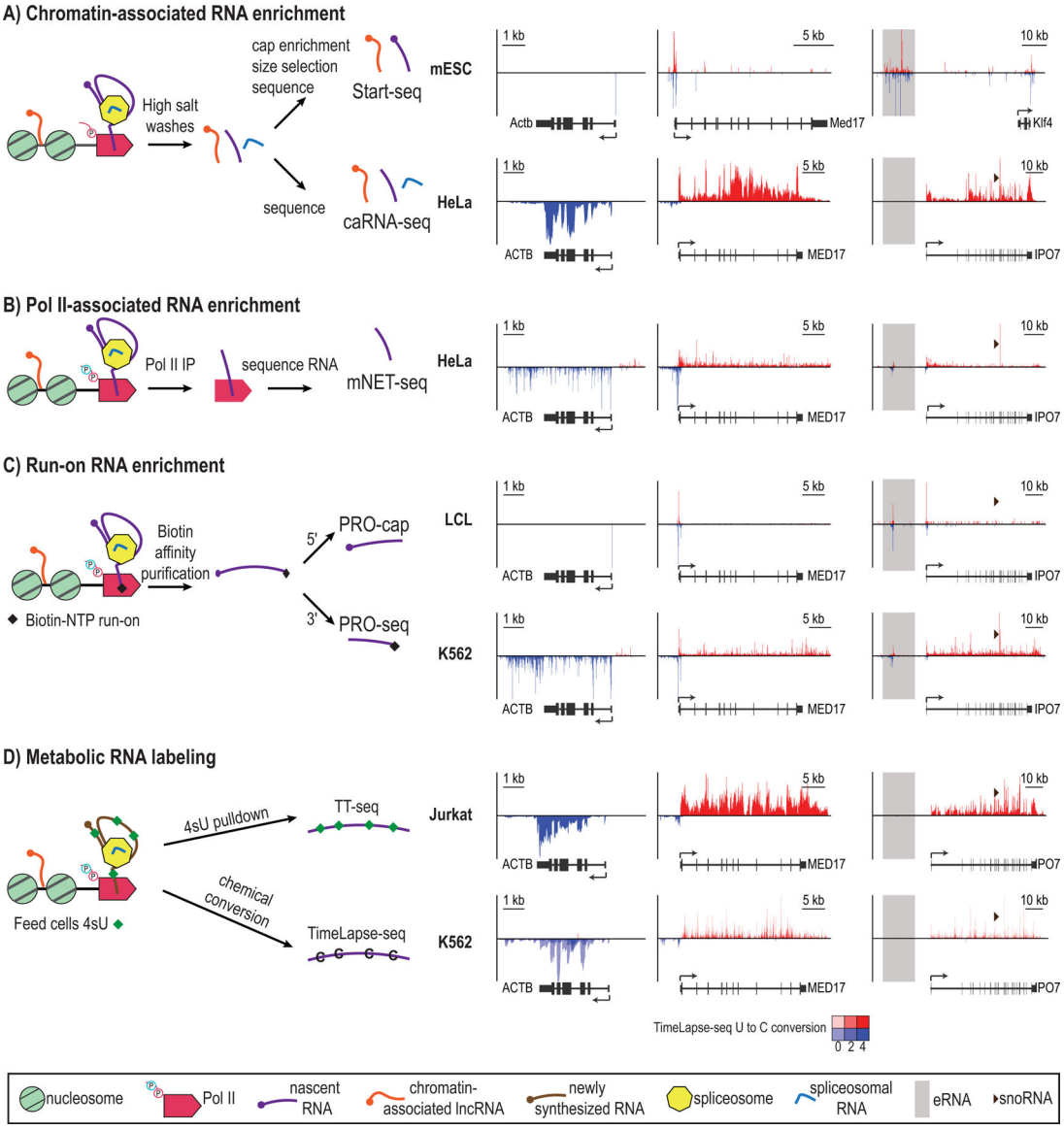

Figure 2 |. Comparison of nascent RNA enrichment and sequencing assays.

Transcription profiles of a highly expressed gene (*β-actin*), a highly paused gene (MED17), and a gene with a nearby eRNA (IPO7 for human cells, Klf4 for mouse cells), generated with distinct RNA-seq methodologies. **a** | In chromatin-associated RNA (caRNA) methods, high salt washes are used to isolate chromatin-bound RNAs. In traditional caRNA-seq (lower), that material is directly sequenced. Start-seq (upper) further enriches for capped caRNAs and uses size selection to capture initiation and pause sites of individual transcripts. Data for caRNA originates from Ref. 46 and Start-seq from Ref. 23. **b** | Immunoprecipitation of Pol II complexes enriches for RNAs that associate with Pol II in mNET-seq. Antibody-mediated isolation of Pol II removes most chromatin-bound RNAs. mNET-seq data that used an antibody targeting total Pol II was obtained from Ref. 61. . **c** | Run-on techniques mark nascent RNAs with labeled nucleotides. The use of the anionic detergent sarkosyl in the run-on reaction releases paused polymerases but not back-tracked or terminated Pol II. The

original genome-wide nuclear run-on assay, GRO-seq, has been adapted to provide single-nucleotide resolution of the position of engaged Pol II on nascent RNA (PRO-seq). Using cap selection and sequencing from the 5’ end of the labeled RNA reports the initiating base (PRO-cap, upper). Data for PRO-cap from Ref. 209 (GSE110638), for PRO-seq from Ref. 4. **d** | Metabolic RNA labeling methods feed living cells modified ribonucleotides that will be incorporated into nascent RNAs. After labeling, nascent RNA can be enriched from the total RNA pool with immuno-purification (TT-seq, upper). Additionally, 4-thiouridine allows U to C base conversion for mutation-based identification of nascent transcripts after sequencing (TT-TimeLapse-seq, lower). Data for TT-seq from Ref. 210, for TimeLapse-seq from Ref. 76.

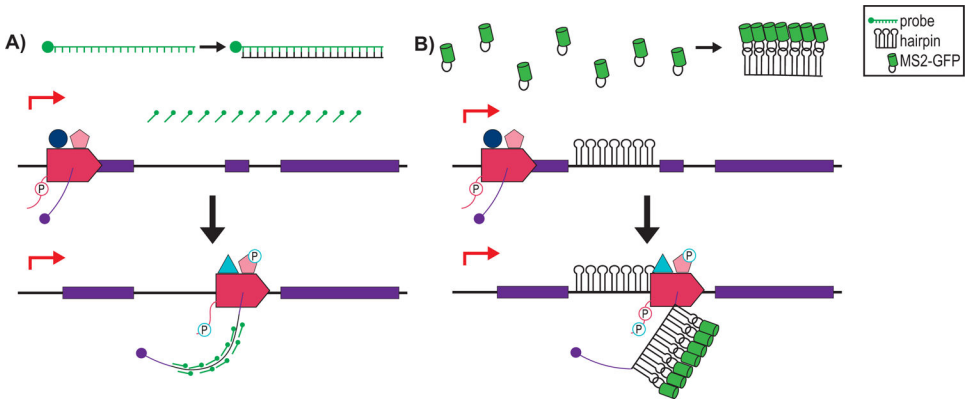

Figure 3 |. Imaging nascent RNA.

a | Nascent RNAs can be detected via fluorescence *in situ* hybridization. Dye-labeled probes that are complementary to RNA hybridize to intronic sequences, permitting detection of endogenous nascent transcripts. **b** | Hairpin-forming sequences that bind the GFP-tagged MS2 protein are engineered into an intron and can be imaged *in vivo*.

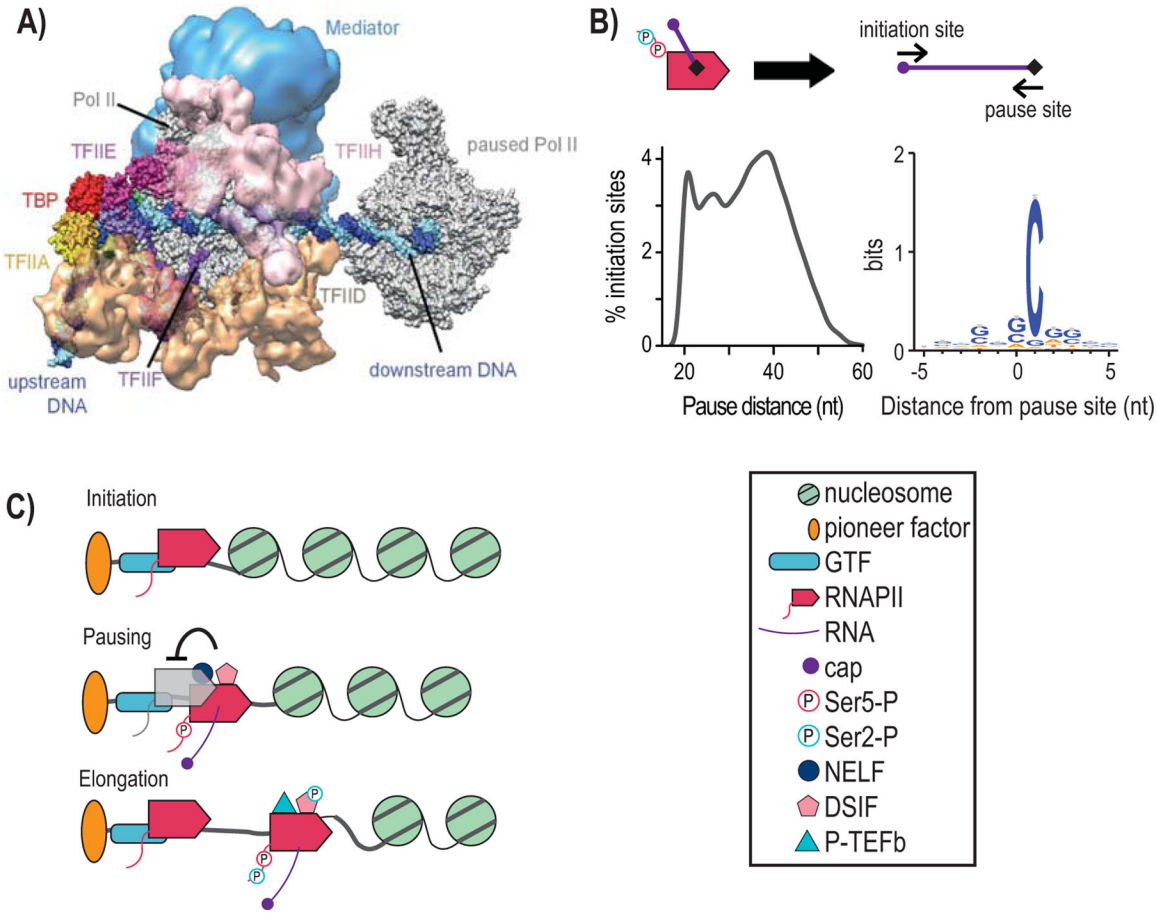

Figure 4 |. Promoter-proximal pausing interferes with transcription initiation.

a | Structural modeling shows that Pol II pausing within 50 nt of the initiation site precludes PIC binding due to steric hindrance. Reprinted from Ref. 70. **b** | Analysis of coPRO data demonstrates that promoter-proximal pausing occurs within 60 nt genome-wide. Pausing predominantly occurs one base upstream from a cytosine in a GC-rich region. Adapted from Ref. 52. **c** | Updated transcription cycle model demonstrating that paused Pol II blocks PIC formation.

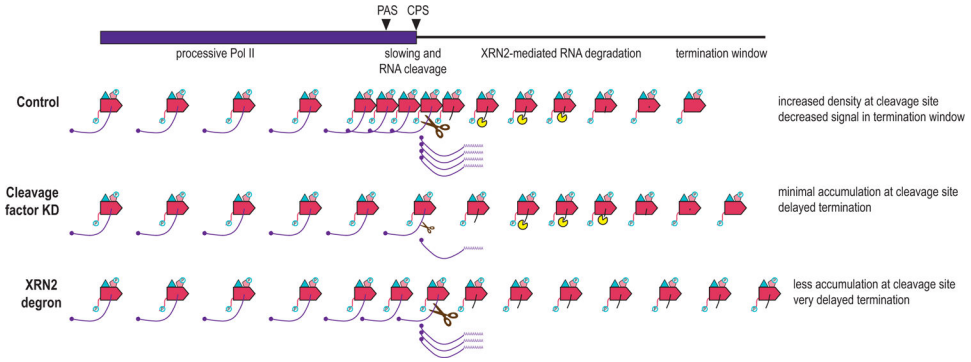

Figure 5 |. Observing cleavage and polyadenylation in nascent RNA datasets.

The 3’ end of a gene is demarcated by a polyadenylation signal (PAS) that is slightly upstream of the cleavage and polyadenylation site (CPS). In typical cells, Pol II accumulates at the CPS, which is indicative of a reduced elongation rate. After cleavage, Pol II continues transcribing, but its levels decrease as termination occurs in a termination window. During cleavage factor knock down, Pol II accumulates far less at the PAS and terminates later. Adapted from Ref. 46. When the XRN2 protein is rapidly degraded, Pol II has slightly less accumulation at the CPS but very delayed termination. Adapted from Ref. 172. These results support the torpedo model, whereby XRN2 chases down unshielded nascent RNA after the cleavage, destabilizing Pol II.

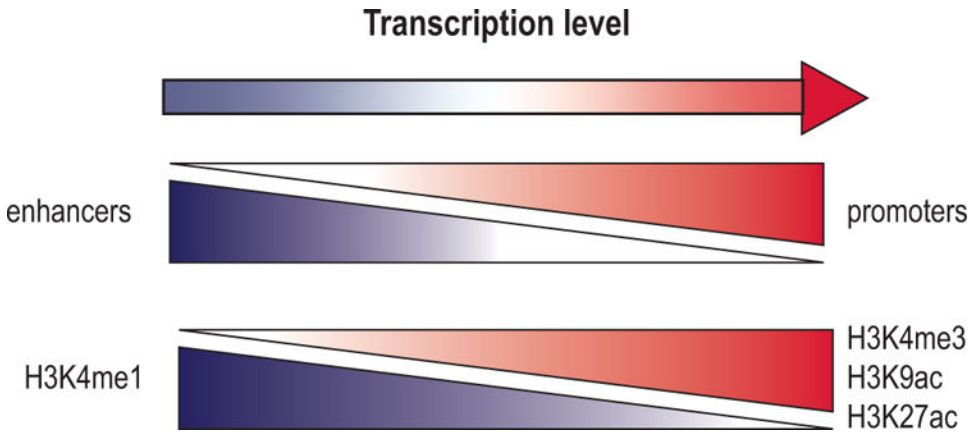

Figure 6 |. Histone modifications correlate with transcription level.

Previous work suggests that H3K4me1 demarcates enhancers and H3K4me3 demarcates promoters. Nascent RNA experiments demonstrate that the degree of H3K4 methylation instead correlates with transcription level. Enhancers are commonly enriched for H3K4me1 because they have lower transcription than promoters.

Table 1 |.

Strengths and limitations of nascent RNA methodologies

Method	Advantages	Considerations
Chromatin-associated RNA sequencing	• Can be used to isolate all chromatin-associated RNA species • Can be combined with methods that assay co-transcriptional processes, including RNA methylation and editing	• Sequences non-nascent RNAs that stably associate with chromatin also
Start-seq	• Simultaneously identifies initiation and pausing sites • Detects eRNAs	• Does not report transcription beyond the first ~100 nucleotides.
NET-seq	• Is Pol II-specific (antibody enrichment)	• Is limited to cells with epitope-tagged Pol II
mNET-seq	• Is Pol II-specific (antibody enrichment) • Can isolate Pol II with different modifications	• Includes RNAs that are stably associated with Pol II • Does not currently include RNA <30 nt in length
PRO-cap	• Identifies transcription initiation sites • Detects eRNAs	• Does not report transcription beyond the first ~100 nucleotides
PRO-seq	• Captures RNAs from transcriptionally-competent polymerases • Identifies positions of active transcription at nucleotide-resolution genome-wide • Detects eRNAs	• Does not measure polymerase backtracking • Captures RNAs being transcribed from Pol I and Pol III also
CoPRO	• Simultaneously identifies initiation and pausing sites • Measures RNA capping status	• Does not measure transcription beyond promoter-proximal pause site
SMIT-seq	• Measures splicing status during transcription	• Limited to species with short introns
TT-seq	• Captures RNAs from actively-transcribing polymerases • Identifies transcription termination sites	• Does not detect Pol II pausing
SLAM-seq and TimeLapse-seq	• Captures RNAs from actively-transcribing polymerases • Can be used to determine RNA stability	• Requires deep sequencing to measure chemical conversion rate • Long labeling times do not capture newly synthesized RNA
Intron seq-FISH	• Detects transcription of 1000s of genes in single cells • Contains positional information of transcribed genes in the 3D space of the nucleus.	• Does not report chromosomal positions of active Pol II complexes • Does not distinguish different steps of transcription • Requires a library of intron-targeting probes and series of hybridizations

Table 2 |.

Methods used to investigate different steps of transcription

Transcription step	Methods used
Transcription initiation site *	GRO-cap[4], PRO-cap[48], 5′ Start-seq[47], CoPRO[52]
RNA capping	CoPRO[52]
Promoter-proximal pausing	mNET-seq[46], GRO-seq[51], PRO-seq[48], CoPRO[52], Start-seq[47]
Co-transcriptional RNA processing	Chromatin-associated RNA-sequencing[42,105–107], mNET-seq[46,50,60], SMIT-seq[163], metabolic labelling[75,169], CoPRO[52]
Transcription termination	TT-seq[50], mNET-seq[46], PRO-seq[48]
Pol II CTD modification status	mNET-seq[46,60]
Transcription bursting	Intron seq-FISH[57]

* While the listed methods were designed specifically to detect TSSs, GRO-seq can infer TSSs, as can PRO-seq and NET-seq with long enough reads.