Nascent RNA analyses: tracking transcription and its regulation

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

Abstract | The programmes that direct an organism's development and maintenance are encoded in its genome. Decoding of this information begins with regulated transcription of genomic DNA into RNA. Although transcription and its control can be tracked indirectly by measuring stable RNAs, it is only by directly measuring nascent RNAs that the immediate regulatory changes in response to developmental, environmental, disease and metabolic signals are revealed. Multiple complementary methods have been developed to quantitatively track nascent transcription genome-wide at nucleotide resolution, all of which have contributed novel insights into the mechanisms of gene regulation and transcription-coupled RNA processing. Here we critically evaluate the array of strategies used for investigating nascent transcription and discuss the recent conceptual advances they have provided.

Enhancer RNAs
(eRNAs). Short (50–2000-nucleotide) non-coding RNAs that are produced by RNA polymerase II from enhancers. The production of enhancer RNAs moderately correlates with the functional activity of the enhancer.

Core initiation regions
Short (~60-nucleotide) regions at promoter and enhancer transcription start sites that provide a binding platform for general transcription factors. At promoters, 'core promoter' is used as a synonym for 'core initiation region'.

[1]*Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA.*
[2]*Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY, USA.*
[3]*These authors contributed equally: Erin M. Wissink, Anniina Vihervaara.*
**e-mail: jtl10@cornell.edu*

https://doi.org/10.1038/s41576-019-0159-6

Transcription is the process by which RNA polymerases synthesize RNA molecules from a DNA template. In eukaryotes, RNA polymerase I (Pol I) and Pol 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). Although 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. Here our focus is Pol II-dependent gene expression and assays used to measure its regulation.

Transcription of genes is controlled by proximal and distal regulatory elements, termed 'promoters' and 'enhancers', respectively (reviewed in REFS[2,3]). These regulatory elements contain binding sites for promoter-specific and enhancer-specific transcription factors (TFs) that define when a gene is active and the frequency with which it is transcribed (FIG. 1a). In mammals, promoters and enhancers predominantly contain two core initiation regions[4,5]. These regions are bound by general transcription factors (GTFs) that, together with Pol II, constitute the pre-initiation complex (PIC). Both promoters and enhancers have a very similar chromatin architecture, including the constellation of GTFs, TFs and paused Pol II complexes, and drive divergent transcription from their core initiation regions[4] (FIG. 1a). However, enhancer transcripts are generally short and unstable, whereas the coding strand of a gene produces predominantly long and stable transcripts (pre-mRNA)[4,6,7] (FIG. 1a). The mechanisms of enhancer–promoter communication remain poorly understood, but recent reports suggest that the amount of eRNA produced from a region correlates with its functional enhancer capacity (reviewed in REFS[2,3,8]). In addition to binding to genomic sequences, certain transcription regulators directly act at the transcription machinery. One of the major signalling platforms for transcription regulation is the C-terminal domain (CTD) of the RBP1 subunit of Pol II (reviewed in REFS[9,10]). This CTD consists of multiple repeats of a heptad consensus sequence (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) and it undergoes regulated post-translational modifications before and during transcription, which lead 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 (FIG. 1b). Binding of sequence-specific pioneer TFs that increase chromatin accessibility provides specificity for the genes that can be transcribed by allowing the recruitment of Pol II to a promoter. Engaged Pol II initially transcribes 20–60 nucleotides, then undergoes promoter-proximal pausing. In mammals and many other metazoans, the promoter-proximal pause is a major rate-limiting step in gene expression and a regulatory checkpoint for execution of transcription programmes[11–17] (reviewed in REF.[18]). Pausing of Pol II is stabilized by the negative elongation factor (NELF) and 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF)[19–21], whereas release

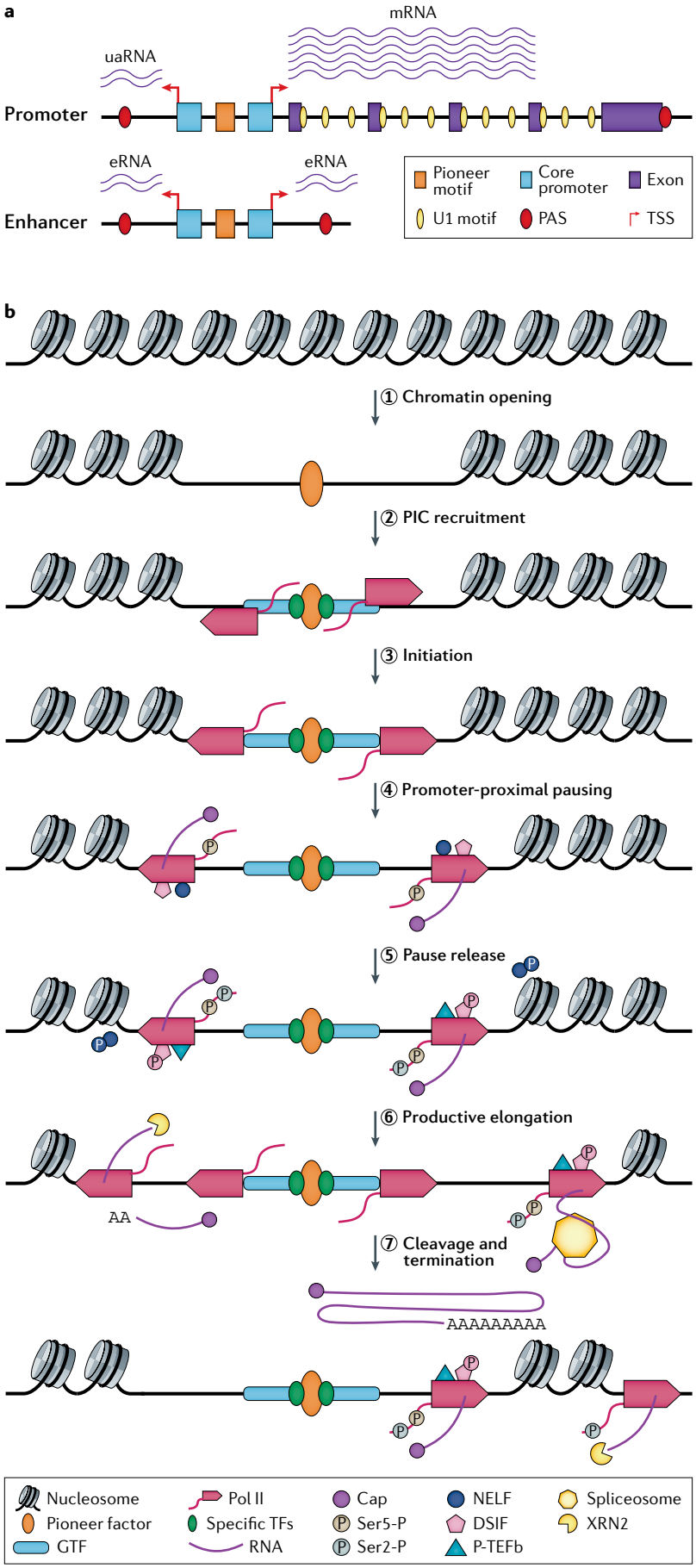


Fig. 1 | **The transcription cycle. a** | Shown is the architecture of a typical gene (upper panel) and enhancer (lower panel), depicting DNA elements that affect transcription and transcript stability. At both gene promoters and enhancers, binding sites for gene-specific transcription factors (TFs) are found between two core initiation regions that drive divergent transcription. At genes, transcription of the coding strand (shown on the right) is initiated at the transcription start site (TSS) and produces mRNA that is stabilized by the presence of 5′ splice sites (that is, U1 motifs)[203]. Transcription of the antisense strand (shown on the left) is initiated from a distinct TSS and produces an unstable upstream antisense RNA (uaRNA) that lacks U1 motifs[203]. Enhancer RNAs (eRNAs) are similar to uaRNAs in that they are short and unstable, and they have a polyadenylation signal (PAS) but not a U1 motif[4]. **b** | The transcription cycle consists of seven steps. In step 1, a pioneer TF binds to a specific sequence motif and increases chromatin accessibility[204,205]. In step 2, additional sequence-specific TFs bind near the pioneer factor[206]. Core promoters recruit general TFs (GTFs) and RNA polymerase II (Pol II) to form the pre-initiation complex (PIC)[207]. In step 3, The GTF TFIIH unwinds DNA, and Pol II initiates transcription[208,209]. In step 4, after transcribing 20–60 nucleotides, Pol II undergoes promoter-proximal pausing and the pause is stabilized by binding of 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) complex[19–21]. Before or during pausing, the C-terminal domain of Pol II is phosphorylated at Ser5 (REF.[210]), and the RNA undergoes 5′ capping[28]. In step 5, Pol II escapes promoter-proximal pausing and enters productive elongation, largely as a result of P-TEFb (which includes cyclin-dependent kinase 9 (CDK9)) phosphorylating multiple targets: these include NELF (ending its interaction with Pol II), DSIF (converting it into an elongation factor), and Pol II Ser2 (which interacts with RNA processing factors)[22]. Additional elongation factors not shown in the figure, such as polymerase-associated factor 1 (PAF1), promote this escape[22]. In step 6, during productive elongation, multiple elongation factors (not shown) enhance the processivity of Pol II[211,212]. Co-transcriptional processing, including splicing, RNA methylation and RNA editing, occurs[210]. Nucleosomes are removed in highly transcribed genes, and chromatin accessibility increases at moderately transcribed genes[213,214]. In step 7, the RNA is cleaved and polyadenylated[23]. After cleavage, Pol II continues elongating, but the nascent RNA lacks a cap and is subject to 5′–3′ exoribonuclease 2 (XRN2)-mediated degradation, which destabilizes Pol II and contributes to termination[24]. After termination, Pol II can be recycled to initiate a new round of the transcriptional cycle[25].

from the pause requires the P-TEFb complex, whose cyclin-dependent kinase 9 (CDK9) subunit phosphorylates NELF, DSIF and the Ser2 residue of the Pol II CTD (reviewed in REF.[22]). Finally, at the 3′ end of the gene, the transcript is cleaved and the pre-mRNA is polyadenylated (reviewed in REF.[23]), leaving the 5′ end of the nascent transcript unshielded, which destabilizes Pol II and contributes to transcription termination (reviewed in REF.[24]). The terminating Pol II dissociates from the DNA and is recycled to participate in a new round of transcription[25]. Several co-transcriptional processes alter the nascent RNA. These processes include 5′ capping of the RNA during initiation and promoter-proximal

pausing[26–29], removal of introns during productive elongation[30], and cleavage and polyadenylation at the 3′ end of the gene (reviewed in REF.[23]). Transcriptional regulation therefore determines which RNAs are expressed and contributes to the final isoform that is produced.

The current mechanistic view of transcription regulation (FIG. 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[31–33], our ability to investigate transcription has evolved greatly, expanding from single-locus analyses to probing transcriptional mechanisms across genomes. Moreover, advances in electron microscopy have revealed the structures of TFs and transcription complexes, including Pol II in its pause complex and elongation complex[19,20,34,35]. In particular, RNA sequencing methods have been crucial for quantifying RNA molecules in the cell and, more recently, for allowing robust measurements 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, and eRNAs.

In this Review, we compare and contrast some of these key strategies, discuss their strengths and limitations, and emphasize how using them in combination can provide a holistic view of regulation throughout the transcription cycle. We begin by describing the conceptual differences between nascent RNA sequencing methods and outlining 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

Sequencing-based methods that investigate RNA synthesis detect nascent or newly transcribed RNAs from the total pool of cellular RNA, either by biochemical enrichment or by chemically inducing point mutations. 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 data sets, including density profiles of transcribing Pol II molecules across the genome, coordinates of transcription start sites (TSSs) 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 sequencing methods, distinct approaches differ remarkably in their ability to enrich for or identify 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 data generated.

The various strategies to enrich for nascent RNA include isolation of chromatin-associated RNA (caRNA)[36–39], Pol II-associated RNA[40,41], small capped RNA[42,43], recently synthesized RNA[44–47], and RNA from elongation-competent Pol II complexes[42,48,49] (FIG. 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, unspliced intermediates and other unstable non-coding RNAs. Although genome-wide nascent RNA sequencing methods can precisely map molecular-level regulation of Pol II at genes and enhancers[13,16,50–53], they do not yet distinguish cell-to-cell variation or provide spatial maps of transcription within the nucleus. By contrast, imaging-based methods have limited genomic resolution and limited ability to discern steps of transcription, but they can quantify transcript production in real time in the 3D space of the nucleus, tissue or organism[54,55] (reviewed in REF.[56]). Next 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.

Isolating caRNAs. 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[36,37,39,57] (FIG. 2a). In one version of caRNA sequencing (caRNA-seq), the transcription inhibitor α-amanitin is added during chromatin isolation to help ensure RNA polymerase is mapped with base-pair precision[38]. Because cytoplasmic and nucleoplasmic RNAs are more abundant and stable than nascent RNAs, enriching for caRNAs significantly increases the dynamic range for detecting changes in RNA synthesis at genes and enhancers. Certain mature RNAs, however, are also stably associated with chromatin and are therefore 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, such as U1, U2, U4, U5 and U6 spliceosomal RNAs. Although caRNA-seq also captures non-nascent transcripts, it is straightforward and can be combined with methods that investigate co-transcriptional processes[36,58–60].

To further increase sensitivity and specificity for nascent RNAs associated with Pol II, native elongating transcript sequencing (NET-seq)[40] was developed in yeast to capture Pol II-associated RNAs using immunoprecipitation of epitope-tagged Pol II (FIG. 2b). Yeast NET-seq shares many biochemical and data-analytical approaches with chromatin immunoprecipitation followed by sequencing (ChIP–seq)[61] and RNA immunoprecipitation followed by sequencing (RIP–seq)[62], 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 is performed with antibodies against Pol II, including against specific CTD post-translational modifications, to identify the regulatory status of the transcription complex at distinct regions of genes[41,63,64]. The original mNET-seq protocols enrich for nascent RNAs by removing most non-nascent RNAs, as well as Pol I- and Pol III-associated

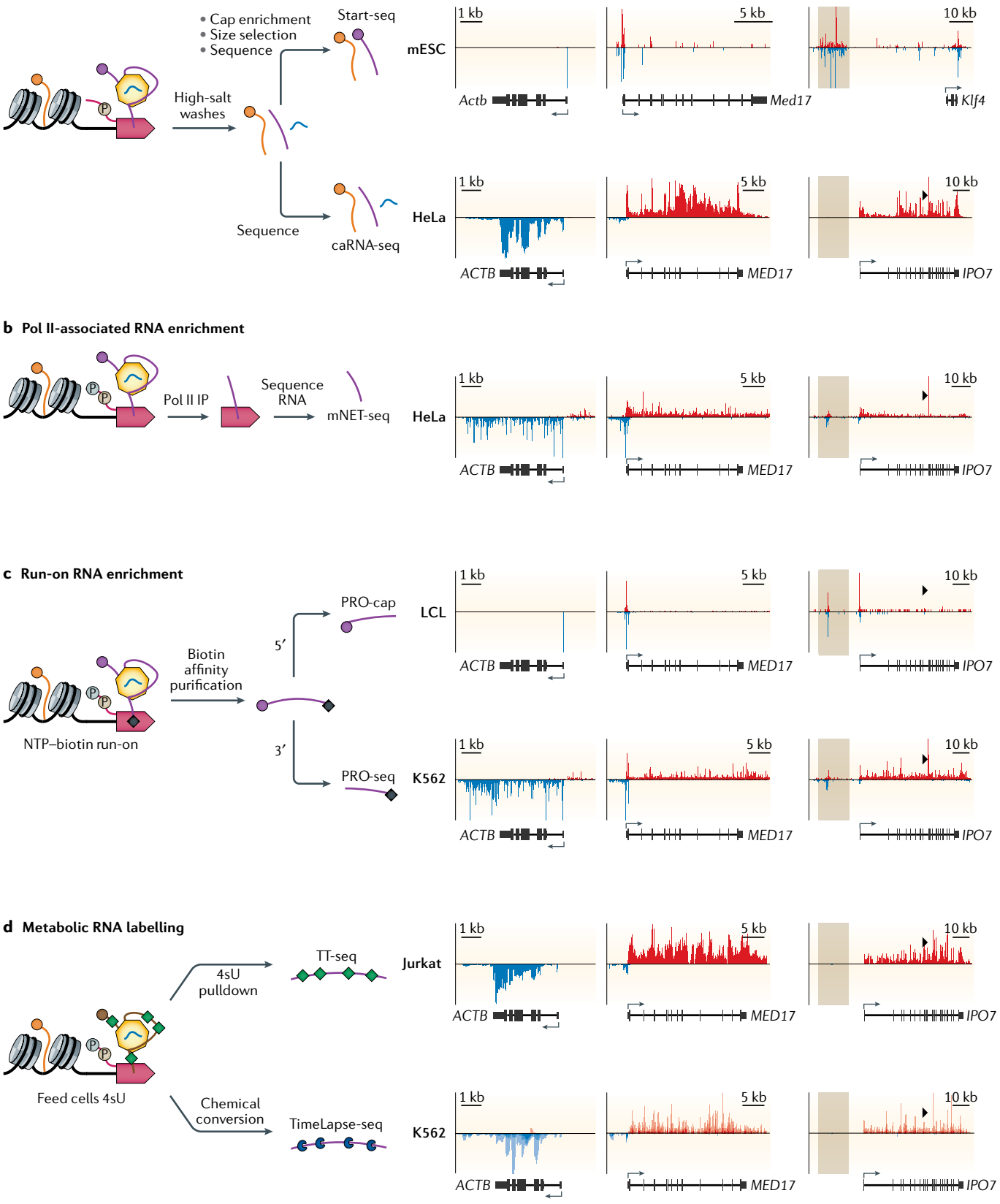
C-terminal domain (CTD) of the RBP1 subunit of Pol II
This domain contains multiple repeats of seven amino acids and serves as a flexible binding scaffold for transcriptional regulators. The post-translational modifications of the heptad repeat greatly influence the regulatory interactions and, therefore, transcriptional processes throughout the cycle.

Nascent RNA
RNA that is in the process of being synthesized via transcription.

Divergent transcripts
In metazoans, genes and enhancers drive transcription from two core initiation regions from both strands in opposing directions. In genes, the mRNA-coding strand is termed the 'sense strand', and the antisense strand produces the divergent transcript.

Chromatin immunoprecipitation followed by sequencing (ChIP–seq)
A method in which protein–DNA interactions are stabilized, chromatin is sheared and fragments with a protein of interest are enriched using an antibody. Purified DNA from the enriched fragments is sequenced, providing genome-wide maps of protein localization.

RNA immunoprecipitation followed by sequencing
A technique that is used to identify which RNAs interact with a given protein. It uses antibody-mediated enrichment of a protein, after which its interacting RNAs can be isolated, reverse-transcribed and sequenced.


a Chromatin-associated RNA enrichment
• Cap enrichment
• Size selection
• Sequence
Start-seq
High-salt washes
Sequence
caRNA-seq
mESC
HeLa
1 kb
5 kb
10 kb
Actb
Med17
Klf4
ACTB
MED17
IPO7
b Pol II-associated RNA enrichment
Pol II IP
Sequence RNA
mNET-seq
c Run-on RNA enrichment
Biotin affinity purification
NTP–biotin run-on
5′
3′
PRO-cap
PRO-seq
LCL
K562
d Metabolic RNA labelling
4sU pulldown
TT-seq
Feed cells 4sU
Chemical conversion
TimeLapse-seq
Jurkat
Nucleosome
Nascent RNA
Newly synthesized RNA
snoRNA
eRNA
4sU
NTP–biotin
Pol II
Ser5-P
Ser2-P
Chromatin-associated lncRNA
Spliceosomal RNA
Spliceosome
TimeLapse-seq U-to-C conversion
0 2 4

◀ Fig. 2 | **Comparison of nascent RNA enrichment and sequencing assays.** Shown are transcription profiles generated using distinct nascent RNA sequencing methods for a highly expressed gene (*ACTB*), a highly paused gene (*MED17*), and a gene with a nearby enhancer RNA (eRNA) (*Klf4* for mouse cells; *IPO7* for human cells, which also contains the small nucleolar RNA (snoRNA) *SNORA23* within an intron). **a** | In chromatin-associated RNA (caRNA) methods, high salt washes are used to isolate chromatin-bound RNAs. In traditional caRNA sequencing (caRNA-seq), the isolated material is directly sequenced. In Start-seq, the isolated material is further enriched for capped caRNAs and size selection is used to capture initiation and pause sites of individual transcripts. Data for caRNA-seq were obtained from REF.[41] and data for Start-seq were obtained from REF.[17]. **b** | In mammalian native elongating transcript sequencing (mNET-seq), immunoprecipitation (IP) of RNA polymerase II (Pol II) complexes enriches for RNAs that associate with Pol II. Antibody-mediated isolation of Pol II removes most chromatin-bound RNAs. Data for mNET-seq with an antibody targeting total Pol II were obtained from REF.[64]. **c** | Run-on techniques mark nascent RNAs with labelled nucleotides. The use of the anionic detergent sarkosyl in the run-on reaction releases paused polymerases but not backtracked or terminated Pol II. The original genome-wide nuclear run-on assay, global run-on sequencing (GRO-seq), has been adapted to provide single-nucleotide resolution of the position of engaged Pol II on nascent RNA (precision run-on sequencing (PRO-seq)). Use of cap selection and sequencing from the 5′ end of the labelled RNA reports the initiating base (precision run-on with cap selection (PRO-cap)). Data for PRO-cap were obtained from REF.[215] (GSE110638) and data for PRO-seq were obtained from REF.[4]. **d** | In metabolic RNA labelling methods, living cells are provided with modified ribonucleotides (such as 4-thiouridine (4sU)) that will be incorporated into nascent RNAs. After labelling, nascent RNA can be enriched from the total RNA pool with affinity purification (transient transcriptome sequencing (TT-seq)). Alternatively, 4sU allows U-to-C base conversion for mutation-based identification of nascent transcripts after sequencing (TimeLapse-seq). Data for TT-seq were obtained from REF.[216] and data for TimeLapse-seq were obtained from REF.[46]. The genomic coordinates depicted are as follows: *ACTB* (mm9: chr5:143,662,795-143,670,430; hg19: chr7:5,564,780-5,572,230; hg38: chr7:5,525,230-5,532,710), *MED17* (mm9: chr9:15,063,100-15,086,100; hg19: chr11:93,509,600-93,550,000; hg38: chr11:93,779,500-93,818,500), *IPO7* (hg19: chr11:9,378,500-9,479,500; hg38: chr11:9,349,000-9,454,500) and *Klf4* (mm9: chr4:55,479,000-55,567,000). The enhancer upstream of *IPO7* was selected because it is ubiquitously transcribed in human cells[217] and chromatin IP followed by sequencing (ChIP–seq) data indicate that similar levels of Pol II are present at this enhancer in the cell lines compared (GSE83777 and REF.[218]). LCL, lymphoblastoid cell line; lncRNA, long non-coding RNA; mESC, mouse embryonic stem cell; NTP, nucleoside 5′-triphosphate.

transcripts. However, certain non-nascent RNAs that stably associate with Pol II remain, including the notable example of small nuclear RNAs[63]. Newer versions of mNET-seq have addressed this issue by using the detergent Empigen BB to dissociate splicing intermediates that bind to the spliceosome and microRNAs that associate with the microprocessor complex[64]. Comparisons of data generated by these two versions of the mNET-seq protocol have provided valuable information about co-transcriptional splicing[63,64].

To specifically identify TSSs, Start-seq enriches for chromatin-associated capped RNA species[43]. The selection of capped RNA is achieved by enzymatic degradation of caRNAs that are not protected by the 5′ cap, such as ribosomal RNA and unshielded mRNA fragments generated during preparation of the sequencing library. Retaining capped caRNAs that are smaller than 80 nucleotides in length further selects for transcripts that are undergoing initiation, pausing or early elongation. With paired-end sequencing, these short, capped RNAs provide high-resolution information about TSSs (5′ end of each read) and active sites of transcription (3′ end of each read) at genes and enhancers[12,43,65].

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.

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 placement of the cells on ice, which stops Pol II from transcribing. Subsequently, the cells are permeabilized, or nuclei or chromatin is isolated, to remove endogenous nucleotides and to enable labelled nucleoside 5′-triphosphates (NTPs) to reach Pol II complexes at the chromatin. Traditional run-on reactions used radiolabelled NTPs[66], which were detected at transcripts by hybridization to complementary DNA sequences[66,67]. With current high-throughput sequencing techniques, RNAs with incorporated labelled nucleotides can be mapped across the genome[48], and the initiating base and active site can be identified at nucleotide resolution[42,48,49]. Thus, run-on sequencing (RO-seq)[42] provides a direct measurement of the positions of competent transcription complexes across the genome, allowing mechanistic studies of regulation of Pol II.

Like caRNA-seq, run-on reactions rely on the strong interaction between transcriptionally engaged Pol II and DNA. During the run-on reaction, the anionic detergent sarkosyl is used to remove proteins, including pausing factors, from chromatin, which enables 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[68]. As run-on techniques rely on the incorporation of nucleotides, only transcription-competent polymerases are detected. Therefore, Pol II in the initiation complex remains undetectable with nuclear run-on techniques[68]. Moreover, sarkosyl cannot rescue Pol II from a backtracked state (reviewed in REF.[69]), which occurs when Pol II moves back in the DNA template (for example as a result of misincorporation of a nucleotide) and displaces 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 labelled nascent transcripts with 5-bromouridine 5′-triphosphate (brUTP) and immunopurified them using an antibody against brUTP[48]. 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 the active site of competent Pol II complexes and, because of their bulkiness, only a single nucleotide is added to the transcript[42]. When all four biotinylated nucleotides, which have similar rates of incorporation[42], are used, all transcripts can be captured without biasing for the presence for certain nucleotide sequences. After the run-on, biotinylated nascent transcripts are isolated using streptavidin-coated magnetic beads, providing a nucleotide-resolution profile of the active sites of transcription[42] (FIG. 2c). The sensitivity of run-on methods is provided by the polymerase-catalysed incorporation of a single biotinylated nucleotide into the active site of each nascent transcript, followed by three biotin-affinity purifications to isolate nascent transcripts during library preparation. This results in very low background, and a nearly 10^6-fold dynamic range estimated from various data sets[42,70]. Although the in vitro aspect of run-on reactions may cause some concern that the conformation of Pol II would be distorted and prevent proper nucleotide incorporation, the transcription

profiles detected with run-on approaches highly correlate with profiles generated with other methods[16,68] (FIG. 2). Moreover, sites of promoter-proximal pausing that are detected by run-ons colocalize with those sites identified by other methods, including permanganate treatment and chromatin immunoprecipitation followed by sequencing (PIP–seq)[42,71], ChIP–seq[68], ChIP-seq with exonuclease treatment (ChIP-exo)[72], mNET-seq[41], and Start-seq[12], and occur predominantly at the nucleotide preceding a cytosine[49,73].

A number of variant methods based on GRO-seq and PRO-seq have been developed. Chromatin run-on and sequencing (ChRO–seq) uses chromatin as a starting material to capture nascent RNAs from tissue samples that are not readily amenable to nuclei isolation protocols or have experienced RNA degradation during handling or storage[74]. Global run-on and precision run-on with cap selection (GRO-cap and PRO-cap, respectively) incorporate enzymatic cap selection, similar to that used in Start-seq, to allow nucleotide-resolution identification of TSSs[4,42]. Coordinated precision run-on and sequencing (CoPRO) is a recent modification of the PRO-cap protocol that coordinates analysis of the initiating nucleotide, capping status and length of RNA molecules[49]. 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 labelled RNAs. Unlike run-on reactions, which use labelled NTPs, metabolic labelling experiments incubate cells in medium supplemented with modified cell-permeable nucleosides, such as 4-thiouridine[44–47,75–80], bromouridine[81] or 5-ethenyluridine[82,83]. Importantly, the cells need to be labelled for at least 5 min to allow the nucleoside salvage pathway to convert the modified uridine nucleosides into UTPs via a series of enzymatic reactions[84]; the NTPs are then incorporated into nascent RNAs by active polymerases. Metabolic labelling cannot track RNA synthesis at nucleotide resolution; however, because it measures RNA synthesis in living cells, it can be used to follow the lifetime and turnover of RNAs by measuring the RNAs synthesized per unit of time. Recent metabolic labelling variants include transient transcriptome sequencing (TT-seq)[47], TimeLapse-seq[46] and thiol (SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq)[45]. TT-seq and TimeLapse-seq have increased the detection of newly synthesized RNAs by restricting 4-thiouridine labelling times to 5 min so that only the 3′ end of nascent RNAs is labelled (FIG. 2d). In TT-seq, extracted total RNA is sonicated before affinity purification to enrich for the RNA sequences with labelled nucleotides, and the levels of labelled RNA are then compared with those of unlabelled or total RNA to identify the newly transcribed RNAs[47]. In SLAM-seq, which has to date used 60-min labelling duration, total RNA is treated with iodoacetamide to alkylate 4-thiouridine[45], and in TimeLapse-seq, which has so far used 5 or 60 min of labelling, total RNA is treated with 2,2,2-trifluoroethylamine and *meta*-chloroperoxybenzoic acid to oxidize the 4-thiouridine[46]. Both techniques rely on reverse transcription to recode the 4-thiouridine into cytosine, creating single-nucleotide T>C mutations that are quantified after deep sequencing[45,46,76,78]. Chemical conversion does not require enrichment of labelled RNA, and nascent RNA is compared with non-nascent RNA within the same sample, which requires less handling of the RNA and less starting material. This direct analysis of the levels of nascent or newly synthesized RNA (that is, RNA containing converted nucleotides) to unlabelled RNA allows average RNA production and decay rates to be estimated genome-wide[45,46,78], as long as the background rates of 4-thiouridine incorporation and sequencing errors are taken into account[76]. Importantly, metabolic labelling of RNA is the only nascent RNA method that has been demonstrated to work not only in cultured cells but also in living organisms, including mice[83], zebrafish[82], nematodes[85] and *Arabidopsis* plants[86].

Imaging-based methods. Current nascent RNA sequencing methods require many cells, and thereby report average transcription dynamics in a population. Imaging-based methods, instead, generally have poor genomic resolution, but can quantify transcription from single cells in real time in the 3D space of the nucleus, tissue or organism (reviewed in REF.[56]). Nascent transcripts can be imaged with techniques such as fluorescence in situ hybridization (FISH), in which labelled oligonucleotides are hybridized with transcripts[87–90] (FIG. 3a). 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 (REFS[91,92]) (FIG. 3b). In these imaging techniques, the site of nascent RNA synthesis in the nucleus is identified as the brightest spot of transcript signal[92–94], either by labelling the locus producing the transcript[95] or by targeting fluorophores specifically to introns that are present in short-lived pre-mRNA species[54,55,96,97]. To visualize single unspliced pre-mRNA molecules, several fluorophores must hybridize to one intron in each transcript. 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. These transcript-by-transcript approaches have revealed Pol II dynamics at model loci, uncovering regulatory principles that control the rate of releasing Pol II into transcription (reviewed in REF.[56]). When applied to cells from mice born to parents of two different strains, these methods have shown that genome variation in enhancers alters transcriptional bursting within the same cell, demonstrating that individual polymorphisms have different effects on the amplitude of transcription[98]. Together with analyses of factor dynamics at regulatory sites[99,100] (reviewed in REF.[101]), transcription-driven mobility of chromatin[102], and enhancer-mediated gene activation[103–105], single-molecule imaging of nascent

Permanganate treatment and chromatin immunoprecipitation followed by sequencing (PIP–seq)
A technique that uses permanganate, which oxidizes unpaired thymines in DNA, to detect the exact locations of open transcription complexes across the genome. RNA polymerase II is then immunoprecipitated, and DNA that has been oxidized by permanganate is cleaved using piperidine. These cleaved sites are identified by high-throughput sequencing.

Transcriptional bursting
Transcription occurs infrequently, and when a gene is turned on, many polymerases transcribe many copies of mRNA in a short time, which is known as transcriptional bursting. Bursting at a given gene is characterized by the duration, amplitude and frequency of transcription.

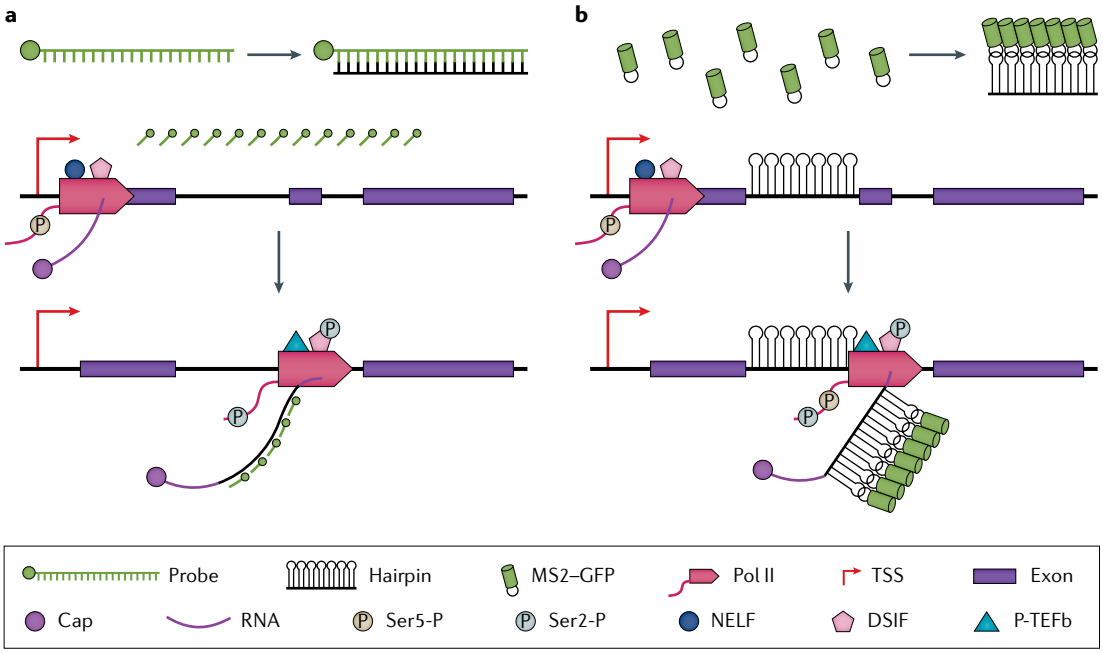


Fig. 3 | **Imaging nascent RNA. a** | Nascent RNAs can be detected by fluorescence in situ hybridization in fixed cells. Dye-labelled DNA probes that are complementary to RNA hybridize to transcribed intronic sequences and allow detection of endogenous nascent transcripts. DNA–RNA hybrids are more stable than DNA–DNA hybrids, so these experiments are conducted in conditions that denature interactions between the DNA probes and complementary DNA sequences. **b** | Hairpin-forming sequences that bind to the GFP-tagged protein MS2 are engineered into an intron and can be imaged in vivo. MS2 specifically binds to the structured RNA, not DNA, providing specificity for the nascent transcript. DSIF, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor; NELF, negative elongation factor; Pol II, RNA polymerase II; TSS, transcription start site.

RNA has elucidated the kinetics of TF binding and chromatin regulation. Moreover, novel targeting and multiplexing strategies, including intron chromosomal expression FISH (iceFISH)[54] and intron sequential FISH (intron seqFISH)[55], have begun to reveal chromatin features and RNA synthesis simultaneously at multiple endogenous loci, setting the stage for visualizing transcriptional programmes and genome regulation within individual cells.

Different methods provide complementary information. Sequencing mature mRNAs reliably measures their steady-state levels, but only nascent RNA sequencing techniques detect unstable RNA species and allow robust quantification of changes in RNA synthesis. Distinct techniques for measuring nascent RNA each have their own strengths and limitations (TABLE 1), arising from versatile strategies for selecting RNAs and the use of method-specific biochemical tools. Importantly, each approach was designed to answer different questions about transcriptional regulation and co-transcriptional processes, and therefore comparing findings from multiple methods provides a detailed look at different stages of gene regulation (FIG. 1), including initiation, pausing, elongation, termination and co-transcriptional processing (TABLE 2).

All of the sequencing methods that we have described can robustly quantify changes in gene expression, and many of them track Pol II progression through distinct steps of transcription. However, none of the current nascent RNA sequencing methods can detect Pol II in the PICs because no transcript has been produced by that step. The PICs can, instead, be captured with ChIP–seq and ChIP-exo[72], and differences in Pol II ChIP–eq and PRO-seq profiles have been used to quantify Pol II in PICs versus pausing complexes[68]. Once Pol II has started transcribing, it sometimes stalls due to backtracking, and RNAs associated with stalled Pol II are included in caRNA-seq, NET-seq[40] and Start-seq[43] libraries. By contrast, RO-seq methods selectively isolate RNAs in transcriptionally competent complexes, and metabolic labelling requires active transcription to label RNAs, which excludes RNAs from stalled Pol II complexes[15,106]. RNAs associated with newly initiated elongation complexes are efficiently captured by caRNA-seq, NET-seq and RO-seq, but RNAs that are shorter than 20 nucleotides are less likely to uniquely map to a reference genome. This challenge in mappability of short and repetitive sequences can be overcome in GRO-seq[48] and length extended chromatin run-on and sequencing (leChRO–seq)[74], which allow generation of longer transcripts during the in vitro transcription step. RO-seq, mNET-seq and Start-seq are all highly sensitive for detecting Pol II pausing, but metabolic labelling does not efficiently label nascent RNAs in stationary or slow-moving Pol II (FIG. 2). However, the metabolic labelling techniques do allow monitoring of transcripts beyond their release from Pol II and, unlike RO-seq and NET-seq, can measure the half-life of the RNA produced. To date, mNET-seq is the only nascent RNA-seq method that has isolated RNAs on the basis of the

Table 1 | **Strengths and limitations of nascent RNA methods**

Method	Advantages	Considerations
caRNA-seq	• 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	Also sequences non-nascent RNAs that stably associate with chromatin
Start-seq	• Simultaneously identifies initiation and pausing sites • Allows de novo calling of putative enhancers	Does not report transcription beyond the first ~100 nucleotides
Yeast NET-seq	• Is Pol II specific (antibody enrichment) • Identifies Pol II positions at nucleotide resolution genome-wide	Is limited to cells with epitope-tagged Pol II
mNET-seq	• Is Pol II specific (antibody enrichment) • Identifies Pol II positions at nucleotide resolution genome-wide • Can isolate Pol II with different post-translational modifications	• Includes RNAs that are stably associated with Pol II • Does not currently include RNA <30 nucleotides in length • Has detected eRNA transcription from previously called enhancers
PRO-cap	• Identifies transcription initiation sites • Allows de novo calling of putative enhancers	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 • Allows de novo calling of putative enhancers	• Does not measure polymerase backtracking • Also captures RNAs being transcribed from Pol I and Pol III
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 • Can be used to determine RNA stability • Identifies transcription termination sites	• Does not detect Pol II pausing • Has detected eRNA transcription from previously called enhancers
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 labelling times do not capture newly synthesized RNA
Intron sequential FISH	• Detects transcription of thousands 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

caRNA-seq, chromatin-associated RNA sequencing; CoPRO, coordinated precision run-on and sequencing; eRNA enhancer RNA; FISH, fluorescence in situ hybridization; mNET-seq, mammalian native elongating transcript sequencing; NET-seq, native elongating transcript sequencing; Pol, RNA polymerase; PRO-cap, precision run-on with cap selection; PRO-seq, precision run-on sequencing; SLAM-seq, thiol (SH)-linked alkylation for the metabolic sequencing of RNA; SMIT-seq, single-molecule intron tracking sequencing; TT-seq, transient transcriptome sequencing.

post-translational modification state of the Pol II[41], which changes through the transcription cycle (FIG. 1). Therefore, the steps of transcription being studied determine which method or combination of methods is appropriate.

The sensitivity of RO-seq and Start-seq in detecting unstable transcripts has allowed these methods to be used for identifying enhancers de novo. Beyond tracking the distinct patterns of transcription at genes and enhancers[16,74,107], the high-resolution and high-sensitivity maps of transcription initiation at putative enhancers obtained with GRO-cap, PRO-cap, and Start-seq have identified many more TSSs than prior studies of capped RNA 5′ ends[4,42,65,107,108]. The distinct cap-selection methods use slightly different strategies to degrade uncapped RNAs and to enrich short capped RNAs for library preparation. Although these cap-selection strategies highly correlate[42], the use of different enzymes could generate differences in cap-selection data sets. The misannealing of reverse-transcript primer has been suggested to cause occasional miscalling of pause sites in any data that rely on reverse transcription from the 3′ adaptor; however, this miscalling of individual pause sites is minor in high-quality data sets[109].

Use of several nascent RNA methods together can 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 remodelling, assembly of the PIC,

Table 2 | **Methods used to investigate different steps of transcription**

Method	Transcription step						
	TSS[a]	RNA capping	Promoter-proximal pausing	Co-transcriptional RNA processing	Transcription termination	Pol II CTD modification	Transcription bursting
Chromatin isolation-based methods							
caRNA-seq	No	No	No	Yes[42,105–107]	No	No	No
Start-seq	Yes[43]	No	Yes[43]	No	No	No	No
mNET-seq	No	No	Yes[41,73]	Yes[41,63,64]	Yes[41]	Yes[41,63]	No
SMIT-seq	No	No	No	Yes[159,160]	No	No	No
Run-on methods							
GRO-cap and PRO-cap	Yes[4,42]	No	No	No	No	No	No
GRO-seq, PRO-seq and ChRO–seq	No	No	Yes[42,48,74]	Yes[166]	Yes[42]	No	No
CoPRO	Yes[49]	Yes[49]	Yes[49]	No	No	No	No
Metabolic labelling methods							
TT-seq	No	No	No	No	Yes[47]	No	No
Imaging-based methods							
Intron sequential FISH	No	No	No	No	No	No	Yes[55]

caRNA-seq, chromatin-associated RNA sequencing; ChRO–seq, chromatin run-on and sequencing; CoPRO, coordinated precision run-on and sequencing; CTD, C-terminal domain; FISH, fluorescence in situ hybridization; GRO-cap, global run-on with cap selection; GRO-seq, global run-on sequencing; mNET-seq, mammalian native elongating transcript sequencing; Pol II, RNA polymerase II; PRO-cap, precision run-on with cap selection; PRO-seq, precision run-on sequencing; SMIT-seq, single-molecule intron tracking sequencing; TSS, transcription start site; TT-seq, transient transcriptome sequencing. [a]'Yes' indicates that the method was designed specifically to detect TSSs, but GRO-seq can infer TSSs, as can PRO-seq and mNET-seq with sufficiently long reads.

initiation of transcription, Pol II pausing and entry into productive elongation, and termination (FIG. 1). In addition, biochemical and computational steps have been adjusted to analyse 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

The regulation of Pol II at distinct steps of transcription underlies expression of individual genes and coordinates transcription programmes. In this section, we describe how nascent RNA methods have enhanced our understanding of the mechanistic regulation of the different steps of the transcription cycle (FIG. 1; 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 was first characterized at *Drosophila* major heat shock genes[14,15,110] and subsequently at human *MYC* and *FOS* genes[111–113], but was initially considered to be present only at a few highly inducible genes. However, studies tracking nascent RNA synthesis, together with experiments using methods such as genome-wide ChIP–seq[61] and PIP–seq[71], demonstrated that promoter-proximal pausing is widespread throughout metazoan genomes[12,13,16,17,38,41,48,51,71,114–116] (reviewed in REFS[18,69]). Furthermore, many transcriptional programmes, such as tissue morphogenesis in *Drosophila* and heat shock response in mammals, are coordinated by pause release[11,13,16,51,117–122]. Nevertheless, some programmes, including the response to oestrogen and androgen signalling, do primarily drive Pol II recruitment to target genes[52,114,123]. Perturbation of Pol II recruitment with triptolide (which causes a genome-wide block in transcription initiation by preventing TFIIH helicase from melting the DNA strands[124,125]) and inhibition of promoter-proximal pause release with flavopiridol (which inhibits CDK9, thereby preventing release of Pol II from the paused PIC[126]) confirmed that these two processes are the main rate-limiting steps for gene transcription[127] (FIG. 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 from steady-state tracking of transcribing Pol II[127].

Polymerase recruitment and pausing are inextricably linked. Structural modelling of DNA-bound PIC, the Mediator complex, and promoter-proximal paused Pol II indicated that pausing within 50 nucleotides of the TSS blocks new initiation owing to steric hindrance[73] (FIG. 4a). The endogenous relevance of this work was demonstrated by a recent CoPRO study in human cells that measured the distance between the pause site and the initiating base genome-wide and found that pausing nearly always occurs within 50 nucleotides of the TSS[49] (FIG. 4b). Furthermore, combined use of mNET-seq and TT-seq demonstrated that new initiation requires release of paused Pol II[73]. Taken together, these results strongly suggest that promoter-proximal pausing precludes a new round of initiation (FIG. 4c).

Molecular events that trigger Pol II to switch from a paused state to an elongating complex have been biochemically characterized (reviewed in REF.[22]) and structurally resolved in the context of the Pol II complex by cryogenic electron microscopy[19,20]. Binding of NELF to the Pol II pause complex seems to facilitate tilting of

Cryogenic electron microscopy
An electron microscopy technique that visualizes molecules at cryogenic (–200 °C) temperatures. It allows near atomic resolution (less than 4 Å) imaging of complex molecules and molecule complexes in their native conformation without crystallization or embedding of the sample.

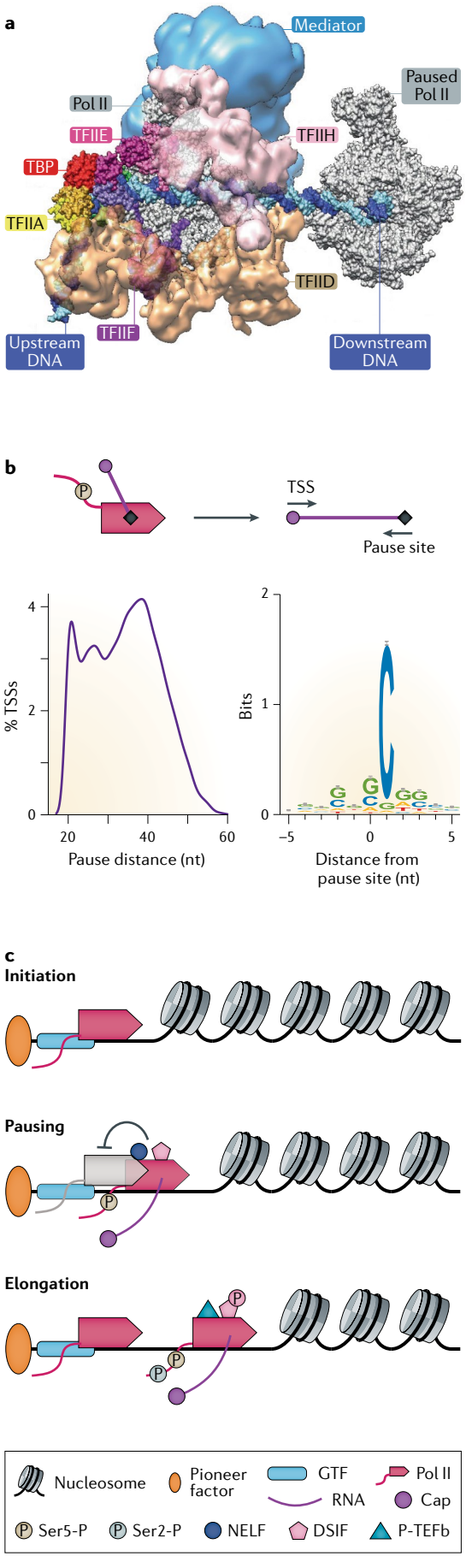


Fig. 4 | **Promoter-proximal pausing interferes with transcription initiation. a** | Structural modelling of paused RNA polymerase II (Pol II) (grey) 50 nucleotides (nt) downstream of the transcription start site (TSS). The pre-initiation complex (PIC) (modelled structure containing Pol II, Mediator, and general transcription factors) can bind to the TSS, but steric hindrance would preclude binding of the PIC if paused Pol II were closer to the TSS. **b** | Genome-wide analysis of coordinated precision run-on and sequencing (CoPRO) data demonstrates that promoter-proximal pausing occurs within 60 nt of the TSS. Pausing predominantly occurs one base upstream of a cytosine in a GC-rich region. **c** | Updated transcription cycle model demonstrating that paused Pol II blocks PIC formation. DSIF, 5,6-dichloro-1-β-D-ribofuranosyl-benzimidazole sensitivity-inducing factor; GTF, general transcription factor; NELF, negative elongation factor. Part **a** reproduced with permission from REF.[73], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.ast). Part **b** adapted from REF.[49], Springer Nature Limited.

the RNA–DNA hybrid into a conformation that prevents NTPs accessing the Pol II active site[20]. The structure further suggests that during the switch to productive elongation, binding of polymerase-associated factor (PAF) displaces NELF and restores a conformation that allows NTPs to enter the active site[19]. PRO-seq and mNET-seq measurements of endogenous Pol II pause sites have demonstrated that Pol II pausing most likely occurs immediately before incorporation of a cytosine; as cytosine is the least abundant nucleotide in mammals[128], its incorporation potentially slows the progress of Pol II sufficiently to permit NELF binding[49,73]. GC-rich sequences are prevalent at promoter-proximal regions of active genes in which pause sites are generally located more than 36 nucleotides from the start nucleotide. By contrast, the pause position at enhancers resides closer to the TSS (within 35 nucleotides), and the degree of pausing is less than that detected at active promoters[49].

Initiation and pausing are regulated at enhancers. Like promoters, enhancers undergo widespread transcription, as first seen at the human β-globin locus[129,130] and since demonstrated genome-wide in many mammalian cell types[4,16,52,53,107,123,131] (reviewed in REF.[132]). Nascent transcription assays have shown that Pol II pausing also occurs at enhancers[65]. As for promoters, the divergent pattern of transcription serves as a robust marker for de novo identification of transcribed enhancers genome-wide[4,65,107,108]. However, despite the similarities between Pol II regulation at proximal and distal regulatory elements, a full understanding of the mechanisms and importance of Pol II pausing at enhancers remains elusive. Short-lived Pol II pausing at enhancers is regulated by DSIF but, unlike at promoters, this pausing is not sensitive to NELF knockdown, as shown by analyses using Start-seq and PRO-seq[65]. However, some enhancers have been shown to bind to NELF on gene activation, which suggests that eRNAs may 'sponge' NELF from promoters[133]. The idea that transcription at genes and enhancers involves similar regulatory steps is further supported by the rapid transcriptional response to heat shock, which causes a genome-wide increase in Pol II density

at both promoter-proximal and enhancer-proximal pause sites[16,49]. Although the coordination of regulatory events at genes and their functionally connected enhancers remains to be characterized, release of paused Pol II from promoter-proximal regions can be stimulated by TF binding to enhancers[134,135], which suggests that CDK9 may act on 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 electron microscopy images of chromosomal spreads[136], and were then demonstrated in vivo in a variety of organisms from bacteria to humans[11,92,137,138] (reviewed in REF.[56]). The exact mechanisms that control bursting remain elusive, but 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 comprises ~2–100 transcribing Pol II molecules[97,139,140]. The probability of a burst occurring (that is, the burst frequency) positively correlates with accessibility and priming of the promoter[141], and promoters with TATA and initiator elements tend to release a larger number of Pol II molecules per burst (that is, they have a larger burst size) than promoters without these sequence elements[98]. Bursting frequency can be increased by forced enhancer connection[103] and inducing histone acetylation at enhancers via targeted recruitment of dead Cas9–p300 (REF.[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[103].

Many models of transcriptional bursting suggest that the rapid switching between on and off states is defined at a single regulatory step[97] (reviewed in REFS[56,139]). However, imaging-based methods rely on labelling introns in pre-mRNAs and therefore cannot distinguish Pol II recruitment and initiation from the release of promoter-proximal paused Pol II to productive elongation (FIG. 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[73]. Recent computational modelling of bursting that used data from intron fluorescence experiments and Pol II ChIP–seq experiments suggests that Pol II pause release, not recruitment, causes a burst to start[142].

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[12,73,127,143] show that pause duration varies widely across the genome, likely owing to distinct promoter sequences, chromatin accessibility and the rate of pause escape (reviewed in REF.[144]). For example, studies that inhibited transcription initiation with triptolide and used a GRO-seq time course to track Pol II as it escaped from promoter-proximal pausing found that the pause stability ranges genome-wide from 2.5 to 20 min (median 6.9 min)[12,127,143]. Another study, which modelled Pol II pausing, termination and release into productive elongation by comparing mNET-seq and TT-seq data, found similar pause durations, ranging from 0 to 10 min[73]. These data sets agree well with an imaging and biochemical study that quantified Pol II dynamics at the *Hsp70* gene in polytene chromosomes of *Drosophila*, which to date provides the only direct measure of the half-life of promoter-proximal Pol II pausing. In its non-induced state, the half-life of Pol II at the promoter of the *Hsp70* gene was 5 min, but, on heat activation, the release of paused Pol II into productive transcription increased ~100-fold[145], which corresponds to an estimated 3-s pause half-life in the active state. Notably, premature termination did not decrease on heat shock[145], which demonstrates that the heat shock-induced increase in transcription is regulated by an increase in the rate of release of paused Pol II into productive elongation rather than by an antitermination mechanism.

Recently, other methods 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: the rates at which Pol II can be freed from chromatin were deduced by photobleaching regions of nuclei and measuring how quickly these regions recovered fluorescence owing to movement of unbleached GFP-tagged Pol II[146]. However, this bulk visualization of Pol II cannot discern whether the measured Pol II population came from promoters, gene bodies or enhancers, or even if it was bound to DNA. To overcome these limitations, the study used Pol II-inhibiting drugs to block transcription at specific steps, and computationally modelled the kinetics of Pol II at the pause site in distinct conditions, arriving at a conclusion that 10% of Pol II molecules have a half-life of 2.4 s, 23% have a half-life of 42 s 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 min[147], which is in rough agreement with triptolide inhibition and sequencing measurements[127]. This footprinting assay used methylation to track if DNA sequences were protected by proteins and used computational modelling to determine if they constituted the PIC or paused Pol II. However, methylation in this assay was performed with ~30-min enzymatic treatments at physiological temperatures[147], during which time Pol II could escape pausing. On the basis of 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 the level of gene activity. In any case, the reported pause durations are long relative to the average elongation rate of Pol II (0.024 s per nucleotide)[44,114,127].

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

Intron definition
A 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 recognize sequences that demarcate introns.

(FIG. 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 and alternative polyadenylation (reviewed in REFS[148,149]). Moreover, RNA editing alters individual nucleotides, which can cause changes in splicing, stability and the protein isoform produced (reviewed in REF.[150]). In addition, epitranscriptomic marks, such as 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 offers insights into whether these processes are regulated together or as independent steps.

5′ capping of RNA occurs before or during promoter-proximal Pol II pause. Shortly after transcription initiation, a 5′ inverted methylguanosine cap is added to the 5′-most nucleotide of the nascent RNA, which provides protection from exonucleases. Early run-on studies using chain-terminating biotin-labelled dNTPs identified that capping occurs shortly after transcription initiation at two *Drosophila melanogaster* heat shock genes[14]. These run-on reactions were paired with decapping assays that demonstrated that 5′ capping predominantly occurred at or before promoter-proximal pause sites[14]. Additional studies showed that capping enzymes bind to Ser5-phosphorylated Pol II CTD[27] and interact with DSIF[26,29]. Recently, CoPRO was used to investigate genome-wide connections between initiation, pausing and capping in human cells[49]. In that study, the initiating base, active site and capping status of individual transcripts as short as 18 nucleotides were measured, which demonstrated that capping begins when the 5′ end of the RNA emerges from the Pol II exit channel and corroborated earlier work on individual genes[14]. Further comparison of RNA capping at pause sites either close to (20–32 nucleotides) or more distant (33–60 nucleotides) from the TSS demonstrated a connection between pausing and capping, and suggested a regulatory role for sequences that influence the pause distance[49]. Promoter-proximal pausing acts as a quality control checkpoint[12] and as a mechanism for synchronous activation of genes during development or in response to signalling[11,13,16,51,119]. The distance between initiation and pausing may, therefore, add another dimension for coordinated gene regulation by providing different opportunities for early RNA processing and Pol II CTD modification, which results 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[30,152]). Splicing components have been visualized on nascent RNAs by electron microscopy[153], and chromatin immunoprecipitation studies have shown them to be associated with chromatin[154–156]. One of the most fundamental questions in the splicing field is how splice sites are recognized with nucleotide precision from the vast variety of pre-mRNA sequences. It has been hypothesized that yeasts, which have strong consensus splicing motifs in short introns, use intron definition to demarcate splice sites (reviewed in REF.[157]). 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 (reviewed in REF.[157]). Sequencing of nascent RNA has allowed more precise measurements of when and where splicing occurs in comparison with transcribing Pol II (reviewed in REF.[158]).

The most direct study of co-transcriptional splicing was performed in yeasts using single-molecule intron tracking sequencing (SMIT-seq)[159,160]. *Saccharomyces cerevisiae* has only 250 intron-containing genes, most of which have only one intron. In SMIT-seq, total caRNA is isolated, and intron-containing genes are then enriched by PCR amplification with primers specific to their first exons. With paired-end sequencing, the position of the polymerase is detected with the 3′ read, and the splicing status is measured with the 5′ read[159]. In *S. cerevisiae*, completed splicing was observed once Pol II had travelled 26 nucleotides downstream of the 3′ splice site, which is the distance required for the spliceosome to access the site, thereby providing strong evidence for intron definition[159]. SMIT-seq in *Schizosaccharomyces pombe*, which has more introns and more multi-intronic genes[161], demonstrated that transcripts that failed to splice co-transcriptionally were not cleaved and polyadenylated[160]. This coupling of transcription, splicing and polyadenylation suggests that co-transcriptional events are co-regulated and can influence one another.

SMIT-seq is not readily adapted to most metazoan introns, which are much longer than yeast introns. Instead, metazoan co-transcriptional splicing has mostly been investigated with indirect methods such as metabolic labelling[80,162], caRNA-seq[36,60] and mNET-seq[41,63]. The exon definition model predicts that splicing would be detected only once the subsequent exon had been fully transcribed. Experiments using mNET-seq have provided some support for the exon definition model in mammals, because spliced transcripts were not detected, even for reads in which Pol II had transcribed into an exon[63]. However, long-read sequencing is needed to interrogate the Pol II position with respect to completed splicing in mammals. Both mNET-seq and GRO-seq experiments have demonstrated that Pol II slows down at exons, detected as higher Pol II density at exons than at introns[38,127]. The slower rate of elongation suggests that sequence and chromatin features of exons have an impact on Pol II progression, and likely contribute to the regulation of splicing. Interestingly, caRNA-seq demonstrated that transcripts largely remain associated with chromatin until splicing is completed, even when those transcripts have already been cleaved and polyadenylated[36,60]. Although spliceosome components are also present in non-chromatin regions of nuclei, including the nucleoplasm and nuclear speckles, these results suggest that proximity to chromatin promotes splicing efficiency[36,163]. Furthermore, sequencing RNAs present in the chromatin and in the nucleoplasm demonstrated that constitutive introns were spliced almost entirely

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.

co-transcriptionally, whereas alternative introns were often spliced post-transcriptionally[60]. 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 edited (reviewed in REF.[150]) and/or modified during and after synthesis. The most common modification of RNA is adenosine methylation at the N6 position (m6A) (reviewed in REF.[151]). Two studies combined nascent RNA sequencing techniques with immunoprecipitation of m6A and found that this methylation occurs rapidly during transcription[59,164]. Both studies also measured the effect on splicing in cells with reduced levels of the m6A methylase METTL3; one study found that alternative splicing is not affected by m6A in caRNA from mouse embryonic stem cells[59], whereas the other did detect an m6A dependence on splicing rate as measured in HEK293 cells by metabolic labelling[164]. These results suggest that m6A may have different effects on splicing in different cell types or that changing the rate of splicing does not have discernible impacts on alternative splicing.

During RNA editing, one nucleotide is chemically converted into another (reviewed in REFS[150,165]). Both GRO-seq and PRO-seq show that editing can occur very rapidly as differences between RNA sequence and the underlying DNA sequence can be detected within 35 nucleotides of the Pol II exit channel, likely occurring in R-loops[166]. The most investigated form of editing, adenosine to inosine, has been studied with caRNA-seq and metabolic labelling, and has been shown to affect splicing efficiency in both *Drosophila*[167] and humans[58]. 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 methods 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 mNET-seq data[38,41,42,48,168]. Knockdown of the CPSF and CstFS cleavage factors results in decreased accumulation of Pol II at the cleavage and polyadenylation site, which indicates a functional connection between Pol II and RNA-processing factors during co-transcriptional RNA cleavage[41] (FIG. 5). Importantly, knocking down cleavage factors and 5′–3′ exoribonuclease 2 (XRN2), which contributes to Pol II termination at the end of transcription (FIG. 1b), also increases Pol II accumulation near the TSS[41], suggesting that many transcripts are cleaved at an early stage of transcription, likely as 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[169]. Alternative cleavage and polyadenylation is widespread in mammals (reviewed in REF.[170]); 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 for, on average, 8 kb, but it eventually dissociates from DNA and terminates transcription[41,47,48] (reviewed in REF.[24]). 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.[24]). Termination is slowed on degradation of XRN2, which strongly supports the torpedo model[168] (FIG. 5). TT-seq data have been used to identify termination sites by capturing the transient RNA downstream of polyadenylation sites[47]. 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[47]. Conversely, termination may be impaired during cellular stress, resulting in transcriptional readthrough[171,172].

eRNAs and microRNAs do not seem to use the same cleavage factors as genes. For example, a GRO-seq study showed that eRNAs are cleaved by the Integrator complex[173], which also processes non-polyadenylated small nucleolar RNAs[174]. 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.[175]). Studies using caRNA-seq and mNET-seq found that this processing occurs co-transcriptionally[41,176].

Post-translational modifications

Regulation of transcription complexes by the CTD of Pol II. Pol II is a multisubunit protein that is regulated by GTFs, positive and negative elongation factors, the chromatin environment, and sequence-specific TFs (reviewed in REFS[9,22,177,178]). ChIP–seq and mNET-seq data largely agree on the distribution of Ser2-phosphorylated Pol II CTD, the density of which increases towards the 3′ ends of genes and is greatest at the regions where 3′-end processing and termination occur[41,179,180]. However, unlike ChIP–seq, mNET-seq does not detect substantial accumulation of Ser5-phosphorylated CTDs at promoter-proximal regions of genes after normalization of CTD modification levels to the total amount of Pol II[41]. Moreover, mNET-seq reported high enrichment of Ser5-phosphorylated CTD at exons[41,63], a finding that has gained support from mass spectrometry analyses in yeast[181], but has not been detected by ChIP–seq (reviewed in REF.[10]). It is likely that the upstream splicing RNA intermediates are closely tracking with Ser5-phosphorylated Pol II after the 5′ splice site cleavage, thereby enriching mNET-seq libraries with RNAs at those positions[41,63]. 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

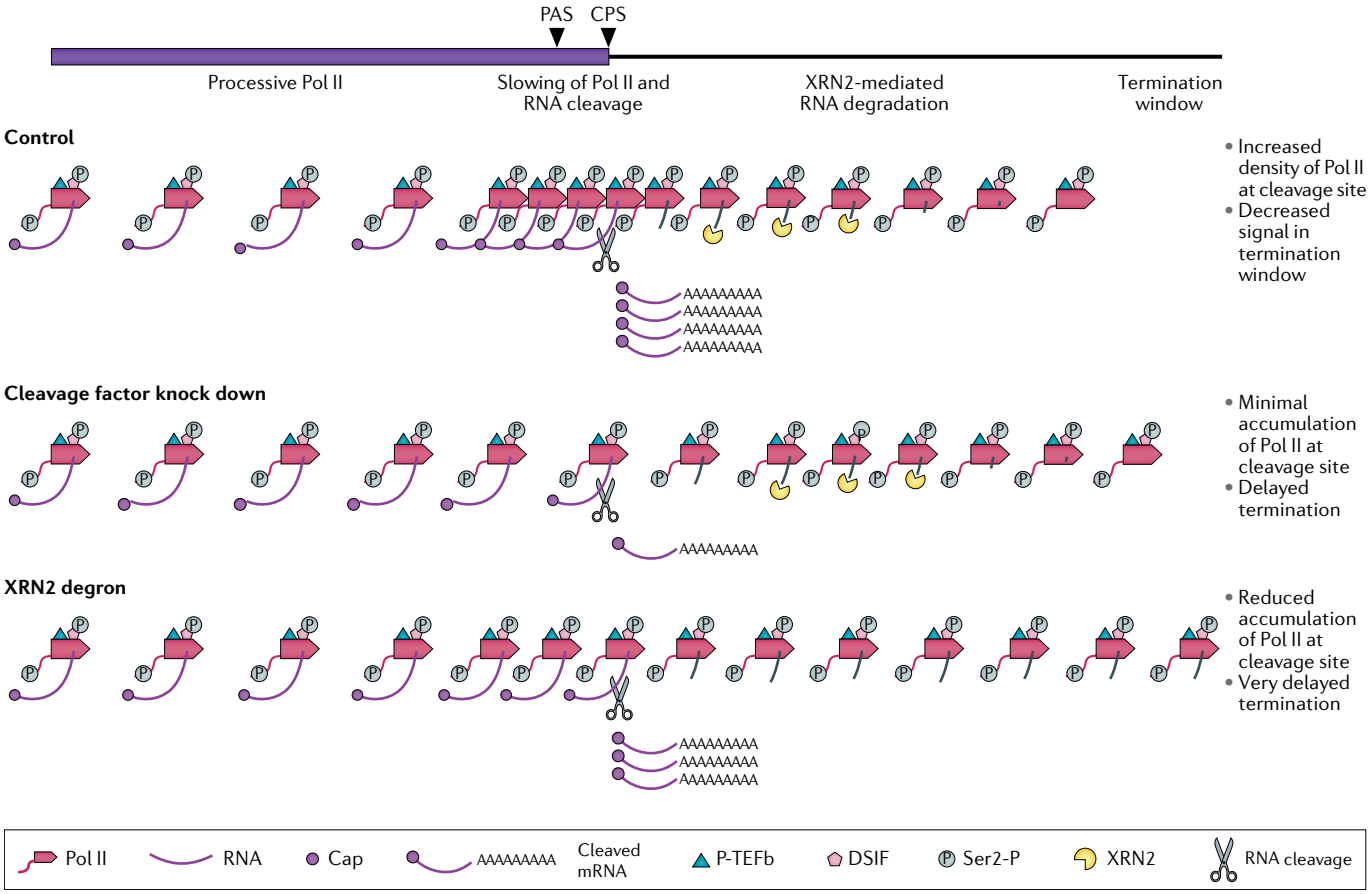


Fig. 5 | **Observing cleavage and polyadenylation in nascent RNA data sets.** The 3′ end of a gene is demarcated by a polyadenylation signal (PAS) that is slightly upstream of the cleavage and polyadenylation site (CPS). In typical cells, RNA polymerase II (Pol II) accumulates at the CPS, which is indicative of a reduced elongation rate. After cleavage, Pol II continues transcribing, but its levels decrease as termination occurs in a termination window. When cleavage factors are knocked down, much less Pol II accumulates at the PAS and transcription terminates later in comparison with control cells[41]. When 5′–3′ exoribonuclease 2 (XRN2) is rapidly degraded using the degron system, slightly less Pol II accumulates at the CPS and termination is substantially delayed in comparison with control cells[168]. Taken together, these results support the torpedo model of transcription termination, whereby XRN2 chases down unshielded nascent RNA after the cleavage, which destabilizes Pol II. DSIF, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity-inducing factor.

mNET-seq captures the RNAs that associate with the Pol II–chromatin complex without crosslinking and, in the original protocol, without stringent washing[41]. Current mNET-seq data sets under-report Pol II that occupies the first ~35 nucleotides downstream of the TSS, likely as a consequence of size selection during library preparation, and this reduces its efficiency for reporting paused Pol II complexes at genes and enhancers[41]. The differences in the distribution of phosphorylated Pol II CTD 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. Similarly to the CTD code of Pol II, different chromatin modifications are also coupled to transcription. These modifications are suggested to both influence transcription and be affected themselves by the progression of Pol II and co-transcriptional processes (reviewed in REF.[182]). For example, histone H3 Lys4 trimethylation (H3K4me3) is thought to increase transcriptional activity by maintaining an active chromatin state[183,184] and recruiting chromodomain helicase DNA-binding protein 1 (CHD1). CHD1 is an ATP-dependent chromatin remodeller that, in turn, maintains trimethylation of H3K36 and H3K4 at transcribed genes, and promotes association of splicing factors and transcription elongation with the Pol II CTD[185]. Intriguingly, co-transcriptional splicing of the first exon increases the level of H3K4me3 at the promoter[186]. Furthermore, chromatin immunoprecipitation studies indicate that the levels of H3K4me3 at promoters inversely correlate with the distance to the first exon and decrease on inhibition of splicing[186]. In a GRO-seq study, the rate of Pol II elongation and its degree of acceleration correlated with the length of the first intron[127]. Introns contain histone H3 Lys79 dimethylation (H3K79me2)[187] and are less nucleosome dense than exons[188], and long introns may therefore provide Pol II with the opportunity to gain full speed[127] (reviewed in REF.[177]). H3K36me3 is deposited co-transcriptionally at genes and is enriched in exons[189–192] (reviewed in REF.[193]), where chromatin

Cap analysis of gene expression
A technique that measures RNA expression and maps transcription start sites of gene promoters. It provides precise maps of transcription start sites of genes that produce long-lived transcripts.

Self-transcribing active regulatory region sequencing (STARR-seq)
A method that assays enhancer activity for millions ofcandidate sequences by cloning them downstream of a reporter gene and upstream of a cleavage and polyadenylation site Functionally active enhancers drive expression of RNA molecules that contains the candidate sequence.

Degron system
A tool for rapidly degrading a specific protein in a cell. Genome editing is used to tag the protein of interest with a protein domain that is recognized by the E3 ubiquitin ligase complex. On addition of a small molecule, the tagged factor is inducibly degraded by the proteasome.

RNA aptamers
Structured RNA molecules selected for binding to a factor of interest to disrupt its functions or interactions.

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.

immunoprecipitation studies show that it blocks cryptic transcriptional initiation in yeast[192] (reviewed in REF.[177]).

The characteristic patterns of histone modifications across the genome have been used to predict functions of specific regions[194]. In particular, histone modifications and DNA accessibility have been widely used to distinguish between active enhancers and promoters; the presence of histone H3 Lys27 acetylation and the relative enrichment of monomethylated H3K4 (H3K4me1) over H3K4me3 are commonly taken to be indicative of enhancer function[195]. However, recent studies have questioned whether histone modifications can be used to discern the classes of regulatory elements. A study from our laboratory identified TSSs using GRO-cap and categorized transcripts produced from these sites as either stable RNAs (that is, mRNAs) or unstable RNAs (that is, upstream antisense RNA or eRNA (FIG. 1a)) on the basis of cap analysis of gene expression data[4]. The results showed that H3K4me1 and H3K4me3 are strongly correlated with polymerase occupancy at both promoters and distal enhancers. H3K4me1 was anticorrelated with polymerase density, whereas marks traditionally associated with promoters, such as acetylation of histone H3 Lys9 and occupancy by GTFs, positively correlated with transcription at the locus, whether it was predicted to be an enhancer or promoter[4]. Recent work from the Adelman laboratory extended these findings by showing that genomic sequences that have episomal enhancer activity in self-transcribing active regulatory region sequencing (STARR-seq)[196] have prominent levels of H3K4me3 and transcription initiation in vivo as measured by ChIP–seq and Start-seq[65]. 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[197]. Furthermore, blocking transcription using flavopiridol reduces the levels of H3K4me3 at enhancers[53]. These results demonstrate that H3K4me3 reflects transcriptional activity but may not distinguish promoters from enhancers. Taken together, the similarities in chromatin state and TF constellation at promoters and enhancers suggest that regulatory elements may exist along a functional spectrum of promoter-ness and enhancer-ness whereby the histone modifications reflect transcriptional activity rather than functionally categorize regulatory elements[198,199] (reviewed in REF.[132]).

Conclusions and future perspectives

Control of RNA synthesis is critical for defining cell types, cellular responses, and organismal functions. Current methods 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 (FIG. 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 cofactors 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. Furthermore, a fuller understanding of gene regulation requires that mechanistic studies of transcription be expanded 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, remains 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 networks of genes and their distal regulatory elements remains to be fully elucidated.

Addressing the preceding 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 TFs and features of chromatin in living cells. Existing perturbation approaches include rapid degradation of TFs (for example, using the degron system, reviewed in REF.[200]) or their inhibition with high-affinity and high-specificity drugs or macromolecules, such as peptides or RNA aptamers (reviewed in REF.[201]). The perturbation-triggered changes in Pol II distribution can be instantly assessed by nascent RNA methods to provide insights into the mechanistic role of each factor before secondary effects confound the interpretation of results. Techniques that use long-read sequencing and require less starting material than existing methods will open new avenues for understanding the coordinated execution of transcription and of co-transcriptional RNA processing. In particular, reducing the amount of required starting material will allow transcription regulation to be studied and compared in individual cells and cell types, including patient-derived samples and specific developmental stages. Indeed, the ability to interrogate nascent RNA synthesis in single cells is a prerequisite for understanding cellular heterogeneity, stochastic gene expression and the regulation of cell 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 that perform or regulate transcription, including Pol II complexes, chromatin remodellers and enhanceosomes. In this regard, cryogenic electron microscopy is already providing precise structures of Pol II and other large regulatory machineries. In the future, application of cryogenic electron microscopy 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 microscopy methods (reviewed in REF.[202]), to understand the interplay of multiple TFs and

elongation factors at enhancers and promoters, and to clarify how transcription is orchestrated in vivo.

Over the past decade, the nascent RNA technologies have provided great insights to the mechanisms of transcription regulation and have delineated gene expression networks in many cell types. Moving forward, these methods — or new and/or improved versions of them — will bring us closer to answering the critical questions described above, and the resulting knowledge will inspire new research directions.

Published online 9 August 2019

1. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
2. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
3. Zabidi, M. A. & Stark, A. Regulatory enhancer-core-promoter communication via transcription factors and cofactors. *Trends Genet.* **32**, 801–814 (2016).
4. Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014).
This study identifies the unified architecture of regulatory elements at genes and enhancers, comprising two core initiation regions to coordinate divergent transcription. It also maps enhancers across the human genome using their characteristic pattern of unstable, divergent transcription.
5. Scruggs, B. S. et al. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* **58**, 1101–1112 (2015).
6. Ntini, E. et al. Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
7. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
8. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).
9. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).
10. Zaborowska, J., Egloff, S. & Murphy, S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777 (2016).
11. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science* **325**, 471–473 (2009).
12. 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).
13. Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G. & Lis, J. T. Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol. Cell* **62**, 63–78 (2016).
14. Rasmussen, E. B. & Lis, J. T. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc. Natl Acad. Sci. USA* **90**, 7923–7927 (1993).
15. Rougvie, A. E. & Lis, J. T. 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).
16. Vihervaara, A. et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat. Commun.* **8**, 255 (2017).
17. Williams, L. H. et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322 (2015).
18. Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* **19**, 385–397 (2018).
19. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018).
This article provides the atomic resolution cryogenic electron microscopy structure of the elongation-competent Pol II complex. It describes the changes in the composition and conformation of the transcription complex as it proceeds from pause into elongation.
20. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018).
21. Wu, C.-H. et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* **17**, 1402–1414 (2003).
22. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192 (2009).
23. Tian, B. & Graber, J. H. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* **3**, 385–396 (2012).
24. Proudfoot, N. J. Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).
25. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225 (2000).
26. Mandal, S. S. 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. USA* **101**, 7572–7577 (2004).
27. 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).
28. Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K. I. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl Acad. Sci. USA* **74**, 2734–2738 (1977).
29. Wen, Y. & Shatkin, A. J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13**, 1774–1779 (1999).
30. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
31. Miller, O. L. & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
32. Miller, O. L., Hamkalo, B. A. & Thomas, C. A. Visualization of bacterial genes in action. *Science* **169**, 392–395 (1970).
33. Pelling, C. Ribonukleinsäure-Synthese der Riesenchromosomen. *Chromosoma* **15**, 71–122 (1964).
34. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
35. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science* **292**, 1863–1876 (2001).
36. Bhatt, D. M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
37. Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19**, 460–469 (2009).
38. Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
39. Weber, C. M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).
40. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
This article introduces NET-seq, which was performed in *S. cerevisiae* by immunoprecipitating epitope-tagged Pol II.
41. Nojima, T. et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
This study applies NET-seq to human cells using antibodies to Ser5-phosphorylated, Ser2-phosphorylated and unphosphorylated Pol II CTD, demonstrating the positions of different Pol II modifications genome-wide.
42. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
This article describes PRO-seq, which maps active Pol II positions at single-nucleotide resolution genome-wide.
43. 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).
This study describes Start-seq, which was used to identify initiating and promoter-proximal pausing positions across the *D. melanogaster* genome.
44. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* **15**, R69 (2014).
45. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
46. Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C. & Simon, M. D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **15**, 221–225 (2018).
47. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
This article describes TT-seq, which uses short metabolic labeling times to capture newly transcribed RNAs in human cells.
48. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (**2008**).
This article describes GRO-seq and demonstrates widespread promoter-proximal pausing and bidirectional transcription across the human genome.
49. Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet.* **50**, 1533–1541 (2018).
50. Booth, G. T., Parua, P. K., Sansó, M., Fisher, R. P. & Lis, J. T. Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nat. Commun.* **9**, 543 (2018).
51. Duarte, F. M. et al. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev.* **30**, 1731–1746 (2016).
52. Hah, N. et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**, 622–634 (2011).
53. Kaikkonen, M. U. et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310–325 (2013).
54. Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* **10**, 246–248 (2013).
55. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376 (2018).
The imaging technique described in this article allows visualization of the synthesis of more than 10,000 genes in single cells.
56. Lim, B. Imaging transcriptional dynamics. *Curr. Opin. Biotechnol.* **52**, 49–55 (2018).
57. 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).
58. Hsiao, Y.-H. E. et al. RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res.* **28**, 812–823 (2018).

59. 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).
60. 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).
61. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
62. Townley-Tilson, W. H. D., Pendergrass, S. A., Marzluff, W. F. & Whitfield, M. L. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. *RNA* **12**, 1853–1867 (2006).
63. 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 (2018).
64. Schlackow, M. et al. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* **65**, 25–38 (2017).
65. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
66. Gariglio, P., Buss, J. & Green, M. H. Sarkosyl activation of RNA polymerase activity in mitotic mouse cells. *FEBS Lett* **44**, 330–333 (1974).
67. Smale, S. T. Nuclear run-on assay. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5329 (2009).
68. Core, L. J. et al. Defining the status of RNA polymerase at promoters. *Cell Rep* **2**, 1025–1035 (2012).
69. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
70. Min, I. M. et al. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).
71. Li, J. et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol. Cell* **50**, 711–722 (2013).
72. Pugh, B. F. & Venters, B. J. Genomic organization of human transcription initiation complexes. *PLOS ONE* **11**, e0149339 (2016).
73. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, e29736 (2017).
This article demonstrates the worth of combining multiple techniques to investigate transcription. The study uses TT-seq, mNET-seq, and structural modelling to examine the interplay of Pol II initiation and pausing.
74. Chu, T. et al. Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat. Genet.* **50**, 1553–1564 (2018).
75. Dölken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
76. Jürges, C., Dölken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218–i226 (2018).
77. Kenzelmann, M. et al. Microarray analysis of newly synthesized RNA in cells and animals. *Proc. Natl Acad. Sci. USA* **104**, 6164–6169 (2007).
78. Muhar, M. et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* **360**, 800–805 (2018).
79. 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).
80. Rabani, M. et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014).
81. Paulsen, M. T. et al. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl Acad. Sci. USA* **110**, 2240–2245 (2013).
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).
83. Jao, C. Y., & Salic, A. Exploring, R. N. A. transcription and turnover in vivo by using click chemistry. *Proc. Natl Acad. Sci. USA* **105**, 15779–15784 (2008).
84. Nyhan, W. L. Nucleotide synthesis via salvage pathway. *eLS* https://doi.org/10.1038/npg.els.0001399 (2001).
85. Stoeckius, M., Grün, D. & Rajewsky, N. Paternal RNA contributions in the Caenorhabditis elegans zygote. *EMBO J.* **33**, 1740–1750 (2014).
86. Sidaway-Lee, K., Costa, M. J., Rand, D. A., Finkenstadt, B. & Penfield, S. Direct measurement of transcription rates reveals multiple mechanisms for configuration of the Arabidopsis ambient temperature response. *Genome Biol.* **15**, R45 (2014).
87. Bauman, J. G., 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).
88. Gall, J. G. & Pardue, M. L. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl Acad. Sci. USA* **63**, 378–383 (1969).
89. Singer, R. H. & Ward, D. C. Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proc. Natl Acad. Sci. USA* **79**, 7331–7335 (1982).
90. Vargas, D. Y., Raj, A., Marras, S. A. E., Kramer, F. R. & Tyagi, S. Mechanism of mRNA transport in the nucleus. *Proc. Natl Acad. Sci. USA* **102**, 17008–17013 (2005).
91. Bertrand, E. et al. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
92. 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).
93. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–478 (2011).
94. Lionnet, T. et al. A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8**, 165–170 (2011).
95. Germier, T. et al. Real-time imaging of a single gene reveals transcription-initiated local confinement. *Biophys. J.* **113**, 1383–1394 (2017).
96. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
97. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
98. Larsson, A. J. M. et al. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
99. Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
100. Gebhardt, J. C. M. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421–426 (2013).
101. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* **217**, 1181–1191 (2018).
102. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050–1055 (2018).
103. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* **62**, 237–247 (2016).
104. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
105. Mir, M. et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev.* **31**, 1784–1794 (2017).
106. Adelman, K. et al. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* **17**, 103–112 (2005).
107. Azofeifa, J. G. et al. Enhancer RNA profiling predicts transcription factor activity. *Genome Res.* **28**, 334–344 (2018).
This study uses computational tools to identify enhancers from RO-seq data and demonstrates that these data can identify active transcription factor motifs.
108. Danko, C. G. et al. Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods* **12**, 433–438 (2015).
This study identifies enhancers and promoters from run-on data by identifying sites of bidirectional transcription.
109. Shivram, H. & Iyer, V. R. Identification and removal of sequencing artifacts produced by mispriming during reverse transcription in multiple RNA-seq technologies. *RNA* **24**, 1266–1274 (2018).
110. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Mol. Cell. Biol.* **6**, 3984–3989 (1986).
111. 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).
112. 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).
113. Plet, A., Eick, D. & Blanchard, J. M. Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene* **10**, 319–328 (1995).
114. Danko, C. G. et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50**, 212–222 (2013).
115. Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
116. Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
117. Adelman, K. et al. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proc. Natl Acad. Sci. USA* **106**, 18207–18212 (2009).
118. Donner, A. J., Ebmeier, C. C., Taatjes, D. J. & Espinosa, J. M. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nat. Struct. Mol. Biol.* **17**, 194–201 (2010).
119. 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).
120. Escoubet-Lozach, L. et al. Mechanisms establishing TLR4-responsive activation states of inflammatory response genes. *PLOS Genet.* **7**, e1002401 (2011).
121. Galbraith, M. D. et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* **153**, 1327–1339 (2013).
122. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
123. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
124. Titov, D. V. et al. XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–188 (2011).
125. 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).
126. Chao, S.-H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J. Biol. Chem.* **276**, 31793–31799 (2001).
127. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
128. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 (1994).
129. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* **11**, 2494–2509 (1997).
130. Routledge, S. J. E. & Proudfoot, N. J. Definition of transcriptional promoters in the human beta globin locus control region. *J. Mol. Biol.* **323**, 601–611 (2002).
131. Kim, H. S. et al. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature* **556**, 510–514 (2018).
132. Tippens, N. D., Vihervaara, A. & Lis, J. T. Enhancer transcription: what, where, when, and why? *Genes Dev.* **32**, 1–3 (2018).
133. Schaukowitch, K. et al. Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).
134. Chen, F. X. et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science* **357**, 1294–1298 (2017).
135. Galli, G. G. et al. YAP drives growth by controlling transcriptional pause release from dynamic enhancers. *Mol. Cell* **60**, 328–337 (2015).
136. McKnight, S. L. & Miller, O. L. Post-replicative nonribosomal transcription units in D. melanogaster embryos. *Cell* **17**, 551–563 (1979).

137. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
138. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
139. Lenstra, T. L., Coulon, A., Chow, C. C. & Larson, D. R. Single-molecule imaging reveals a switch between spurious and functional ncRNA transcription. *Mol. Cell* **60**, 597–610 (2015).
140. Bothma, J. P. et al. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc. Natl Acad. Sci. USA* **111**, 10598–10603 (2014).
141. Chen, L.-F. et al. Enhancer histone acetylation modulates transcriptional bursting dynamics of neuronal activity-inducible genes. *Cell Rep.* **26**, 1174–1188 (2019).
142. Bartman, C. R. et al. Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol. Cell* **73**, 519–532 (2019).
143. 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).
144. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev*. https://doi.org/10.1101/gad.325142.119 (2019).
145. Buckley, M. S., Kwak, H., Zipfel, W. R. & Lis, J. T. Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes Dev.* **28**, 14–19 (2014).
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. USA* **115**, E4368–E4376 (2018).
147. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422.e4 (2017).
148. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).
149. Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* **84**, 325–354 (2015).
150. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
151. Meyer, K. D. & Jaffrey, S. R. Rethinking m6A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* **33**, 319–342 (2017).
152. Merkhofer, E. C., Hu, P. & Johnson, T. L. Introduction to cotranscriptional RNA splicing. *Methods Mol. Biol. Clifton NJ* **1126**, 83–96 (2014).
153. Osheim, Y. N., O.L. Miller, Jr & Beyer, A. L. RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* **43**, 143–151 (1985).
154. Görnemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* **19**, 53–63 (2005).
155. Lacadie, S. A. & Rosbash, M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. *Mol. Cell* **19**, 65–75 (2005).
156. Listerman, I., Sapra, A. K. & Neugebauer, K. M. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* **13**, 815–822 (2006).
157. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
158. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
159. Carrillo Oesterreich, F. et al. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**, 372–381 (2016).
160. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).
161. Prabhala, G., Rosenberg, G. H. & Käufer, N. F. Architectural features of pre-mRNA introns in the fission yeast Schizosaccharomyces pombe. *Yeast* **8**, 171–182 (1992).
162. Pai, A. A. et al. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *eLife* **6**, e32537 (2017).
163. Khodor, Y. L. et al. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25**, 2502–2512 (2011).
164. Louloupi, A., Ntini, E., Conrad, T. & Ørom, U. A. V. Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. *Cell Rep* **23**, 3429–3437 (2018).
165. Eisenberg, E., & Levanon, E. Y. A-to-I, R. N. A. editing — immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
166. Wang, I. X. et al. RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep* **6**, 906–915 (2014).
167. Rodriguez, J., Menet, J. S. & Rosbash, M. Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. *Mol. Cell* **47**, 27–37 (2012).
168. Eaton, J. D. et al. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* **32**, 127–139 (2018).
169. Chiu, A. C. et al. Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. *Mol. Cell* **69**, 648–663 (2018).
170. Tian, B. & Manley, J. L. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.* **38**, 312–320 (2013).
171. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread inducible transcription downstream of human genes. *Mol. Cell* **59**, 449–461 (2015).
172. Vilborg, A. et al. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl Acad. Sci. USA* **114**, E8362–E8371 (2017).
173. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015).
174. O'Reilly, D. et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res* **42**, 264–275 (2014).
175. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
176. Dhir, A., Dhir, S., Proudfoot, N. J. & Jopling, C. L. Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat. Struct. Mol. Biol.* **22**, 319–327 (2015).
177. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
178. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
179. 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).
180. 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).
181. Harlen, K. M. et al. Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* **15**, 2147–2158 (2016).
182. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone marks in the 'driver's seat': Functional roles in steering the transcription cycle. *Trends Biochem. Sci.* **42**, 977–989 (2017).
183. Klymenko, T. & Müller, J. The histone methyltransferases trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).
184. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
185. Sims, R. J. 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).
186. Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K. & Neugebauer, K. M. First exon length controls active chromatin signatures and transcription. *Cell Rep* **2**, 62–68 (2012).
187. Huff, J. T., Plocik, A. M., Guthrie, C. & Yamamoto, K. R. Reciprocal intronic and exonic histone modification regions in humans. *Nat. Struct. Mol. Biol.* **17**, 1495–1499 (2010).
188. 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).
189. Carrozza, M. J. et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).
190. Keogh, M.-C. et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593–605 (2005).
191. Kolasinska-Zwierz, P. et al. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376–381 (2009).
192. Venkatesh, S. et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452–455 (2012).
193. McDaniel, S. L. & Strahl, B. D. Shaping the cellular landscape with Set2/SETD2 methylation. *Cell. Mol. Life Sci.* **74**, 3317–3334 (2017).
194. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
195. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
196. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
197. 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).
198. Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**, 426–433 (2015).
199. Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49**, 1073–1081 (2017).
200. Natsume, T. & Kanemaki, M. T. Conditional degrons for controlling protein expression at the protein level. *Annu. Rev. Genet.* **51**, 83–102 (2017).
201. Ozer, A., Pagano, J. M. & Lis, J. T. New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization. *Mol. Ther. Nucleic Acids* **3**, e183 (2014).
202. Schermelleh, L. et al. Super-resolution microscopy demystified. *Nat. Cell Biol.* **21**, 72 (2019).
203. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
204. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
205. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
206. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 40–51 (2012).
207. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204–209 (2017).
208. Conaway, R. C. & Conaway, J. W. 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. USA* **86**, 7356–7360 (1989).
209. Tirode, F., Busso, D., Coin, F. & Egly, J. M. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* **3**, 87–95 (1999).
210. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
211. 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).
212. Takeuchi, A. et al. Loss of Sfpq causes long-gene transcriptopathy in the brain. *Cell Rep* **23**, 1326–1341 (2018).

213. 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).
214. Petesch, S. J. & Lis, J. T. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134**, 74–84 (2008).
215. Kristjánsdóttir, K. et al. Population-scale study of eRNA transcription reveals bipartite functional enhancer architecture. Preprint at *bioRxiv* https://doi.org/10.1101/426908 (2018).
216. Michel, M. et al. TT-seq captures enhancer landscapes immediately after T-cell stimulation. *Mol. Syst. Biol.* **13**, 920 (2017).
217. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
218. ENCODE Project Consortium. A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLOS Biol.* **9**, e1001046 (2011).

Acknowledgements
The authors thank members of the Lis laboratory for insightful discussions. They also thank the 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.).

Author contributions
All authors contributed to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Genetics thanks N. Proudfoot and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.