nature structural & molecular biology

FOCUS | PERSPECTIVE

https://doi.org/10.1038/s41594-019-0288-9

A 50 year history of technologies that drove discovery in eukaryotic transcription regulation

John T. Lis

Transcription regulation is critical to organism development and homeostasis. Control of expression of the 20,000 genes in human cells requires many hundreds of proteins acting through sophisticated multistep mechanisms. In this Historical Perspective, I highlight the progress that has been made in elucidating eukaryotic transcriptional mechanisms through an array of disciplines and approaches, and how this concerted effort has been driven by the development of new technologies.

The eukaryotic transcription field reached a landmark with the discovery of the machines that transcribe human genes: RNA polymerases I, II and III (Pol I, II and III)[1]. The technological power of biochemistry enabled the purification of RNA polymerases from complex mixtures in nuclear extracts, and the history surrounding this achievement and the subsequent identification of the accessory factors that guide the activity of the RNA polymerases at their cognate promoters is authoritatively presented in an accompanying Review by Robert (Bob) Roeder[2]. Understanding of eukaryotic transcription, as well as tools and approaches used for these studies, have come a long way in the past 50 years. For example, in the first decades of the transcription era, researchers routinely performed biochemical fractionations that often required weeks or months of tortuous work in the bone-numbing confines of a cold room; now, researchers routinely analyze incredibly massive DNA and RNA sequencing data sets that often require weeks or months of computational analyses done in the comfort of warm and cushy lounges. Transcription research may have lost some of the grit of its early days, but it is no less challenging and exciting. This quest will ultimately culminate in a full understanding of transcription and its regulation at every base of the human genome, and of any genome of interest.

'Observing' cells: a complement to classical biochemistry

Understanding transcription regulation is an enormous challenge. It was clear as early as the 1980s that a plethora of factors was needed simply to direct correct initiation on model eukaryotic genes[3,4]. A full mechanistic understanding of transcription of a single Pol II target gene would require the purification of many scores of factors and their proper assembly on appropriate chromatin templates. Given that Pol II was known to transcribe all mRNA-encoding genes, each with its particular repertoire of regulatory protein requirements, the field realized that a full understanding of transcription in molecular terms would require a huge effort as well as new technologies. Indeed, the enormous investment of effort and the development of an array of powerful approaches and methods (Table 1) have moved the field at an exhilarating pace.

The early mechanistic studies of transcription depended heavily on classical biochemistry approaches to identify and purify not only the RNA polymerases but also the general transcription factors (GTFs) that direct the RNA polymerases to their transcription start sites (TSSs) and that regulate the frequency at which the polymerases transcribe their cognate genes[5]. These biochemical efforts were guided by codes of practice that were eventually articulated as ten commandments by the late Arthur Kornberg[6]. Two commandments were particularly memorable: "IV. Do not waste clean thinking on dirty enzymes," and "V. Do not waste clean enzymes on dirty substrates." The original source of commandment IV was the late Efraim Racker, an esteemed colleague in my department at Cornell University. It was therefore with some trepidation that I decided, as an assistant professor, to study the mechanics of transcription and its regulation in the extremely 'dirty' milieu of whole cells. However, one could imagine that an approach for examining the process in living cells, where Pol II and its initiation and regulatory factors are present at physiological concentrations and properly assembled on chromatin templates, might, with measured investments, provide critical complementary information (signposts) that could guide the more Herculean biochemical efforts of reconstructing proper regulation from purified components.

The recombinant DNA technology revolution[7] provided a major breakthrough in transcription studies, making available pure gene substrates and expressed, recombinant proteins for use in biochemical reconstructions in vitro. It also provided highly specific probes to monitor, or in effect 'observe', transcription and transcription factor interactions within cells. With such probes, Dave Gilmour and I sought to develop methods for tracking Pol II and other factors on their native templates. The idea was to crosslink proteins to DNA and fish out specific complexes. The use of an excellent antibody to RNA Pol II[8] provided by the Greenleaf lab made it possible to test this idea in *Drosophila melanogaster*. In heat-shocked cells, the expectation was that Pol II would copurify with entire coding region heat-shock genes, which could be readily quantified with pure gene probes. Although the signal following heat-shock treatment was indeed greater than that from non-heat-shocked control cells, we detected a signal from the 5′ end of the genes that represented about one Pol II per gene independent of heat-shock treatment, indicating that Pol II was already associated with the promoter before the gene was induced[9]. Moreover, subsequent experiments showed the Pol II was transcriptionally engaged and paused 20–50 bases downstream of the TSS[10,11]. Thus, this promoter-proximal paused Pol II was a feature that needed to be accommodated in assessing mechanisms of transcription regulation, at least for heat-shock genes. Strong evidence for promoter-proximal, transcriptionally engaged Pol II on the uninduced long terminal repeat (LTR) of the human immunodeficiency virus (HIV) was also acquired by the Peterlin lab, as was support for the activation of productive Pol II elongation by Tat, the HIV regulatory protein[12]. Transcriptionally engaged Pol II on the extreme 5′ ends of *MYC* and *FOS* genes in mammalian cells was also discovered in Eick and Groudine laboratories[13,14], indicating paused Pol II and regulated early elongation was not restricted to heat-shock genes and HIV promoters.

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA. e-mail: johnlis@cornell.edu

Table 1 | Major approaches & methods that have driven transcription regulation

Approach	Conceptual advance/breatkthrough	Select example(s)
Classical biochemistry	Allowed systematic identification and purification of key proteins from cellular or nuclear extracts by biochemical fractionation and activity assays	Purification of RNA Pol I, II and III[78] Purification of Pol II initiation factors[79] Purification of P-TEFb[17]
Genetic forward screens	Allowed selection of genes that encode components of a particular molecular transcriptional machine or that affect a specific process	Independent screens generated *swi* and *snf* mutants that led to discovery of the SWI/SNF chromatin remodeler[80] Suppressors screens of Ty element insertions led to the identification of a collection of TFs, the Spts[23]
Recombinant DNA technologies	Allowed isolation of genes and regulatory elements for DNA templates and probes of in vivo gene function and mechanisms	Isolation of specific genes[81] to be used as templates for transcription and to express transcription factors[82]
Specific protein-DNA interaction probing in cells	Universal use of cloned DNAs as probes of transcription regulation in targeted gene studies	*Drosophila* Pol II association with promoters of uninduced genes mapped in cells[9]
Nucleic acid sequencing advances	Allowed first view of the organizational principles of genomes of humans and model organisms, and provided a basis for comparing biological information across organisms	Value of sequencing of the human genome assessed broadly on tenth anniversary; most importantly it provided all the components required to study transcription[83]
Technologies made possible by high-throughput sequencing	Genome sequencing allowed adaptation of target gene probing to be executed genome-wide, often with base-pair resolution, allowing mechanistic features of transcription to be viewed in vivo	ChIP-exo maps position of transcription factor binding with base-pair resolution[45] PRO-seq maps transcriptionally active polymerases with base-pair resolution[39]
Advances in structural biology	Transcription studies greatly benefited from crystal structures of polymerases and the nucleosome. Now structures of even larger complexes are being solved by cryo-EM owing to improvements in detectors and computational processing of images.	X-ray structure of RNA Pol II[63] Cryo-EM structure of the RNA Pol II paused complex[67]
Advances in light microscopy	Super-resolution microscopy and single-molecule tracking in real time allows one to track the dynamics of transcription, transcription factor binding and enhancer-promoter interactions	Tracking single transcription factors[70] Watching enhancer-promoter interactions in real time[77]

The effects of the purine nucleoside analog 5,6-dichloro-l-3-ribofuranosylbenzimidazole (DRB)

DRB, a kinase inhibitor, provided additional support for transcriptional regulation beyond Pol II recruitment and initiation. Although DRB has no effect on in vitro transcription reconstituted from purified Pol II and GTFs, it is a potent inhibitor of mRNA production in cells[15] (reviewed in Yamaguchi et al.[16]). This suggested an additional layer or step in regulation of transcription. One critical component in this layer of regulation, positive transcription elongation factor (P-TEFb), was purified by the Price lab[17]. P-TEFb stimulated transcription in vitro in a DRB-sensitive manner. Other critical components were the two multiprotein complexes, DRB sensitivity-inducing factor (DSIF)[18] and negative elongation factor (NELF)[19], which were purified by the Handa lab. Gilmour and Handa labs showed in a compelling manner that DSIF and NELF, when added to in vitro transcription systems, cause promoter-proximal pausing in vitro[20,21]. The P-TEFb kinase can phosphorylate multiple sites on components of the paused complex (including Pol II, DSIF and NELF) to stimulate paused Pol II to enter into productive elongation[22] (Fig. 1). Notably, the DSIF subunits, Spt4 and Spt5, were also uncovered in a genetic selection[23]—a selection in budding yeast that also identified TATA-binding protein, TBP (Spt15), the foundation factor for Pol II recruitment to promoters. Thus, it was clear that any mechanistic model of regulation needed to accommodate at least two distinct early steps in transcription: the recruitment of Pol II to promoters and the release of Pol II from its promoter-proximal pause.

With the subsequent revolution in genome-wide methods, the molecular features associated with particular types of regulation could be observed across all genes. Adelman, Levine and Young labs demonstrated that Pol II accumulated at many promoter regions, initially using chromatin immunoprecipitation followed by sequencing (ChIP-chip) assays in which oligonucleotides of *Drosophila* and human genes immobilized on a microarray were used to probe the genomic DNA fragments associated with immunoprecipitated Pol II[24,25]. Genome-wide global run-on sequencing (GRO-seq) assays demonstrated that a majority of promoters in human and *Drosophila* cells had the properties of promoter-proximal paused Pol II originally associated with the extensively characterized *Drosophila Hsp70* gene[26,27]. Additionally, virtually all mRNA coding genes were shown to go through this pausing step, even if the gene did not accumulate a paused Pol II under the conditions being analyzed[28]. Therefore, promoter-proximal pausing is a common step in Pol II transcription, and more often than not it is rate-limiting.

These and a variety of subsequent studies from many laboratories made it clear that the process of transcription and its regulation is a multistep odyssey for polymerases, and especially for Pol II, whose transcripts are extensively processed during transcription (Fig. 1). The steps of the transcription cycle begin with ensuring chromatin is open, a job executed by chromatin remodelers and pioneer factors that can bind specific DNA sequences on nucleosomes[29]. This allows the recruitment of Pol II to the promoter (ushered by GTFs), initiation and early elongation to the

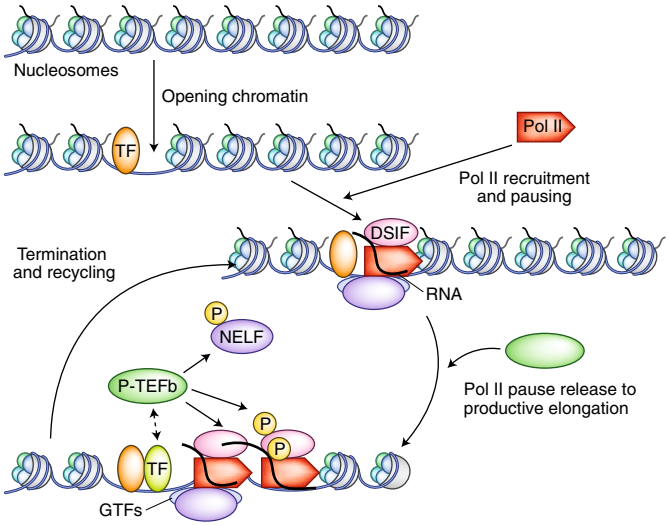

Fig. 1 | The transcription cycle consists of distinct steps that are targets of regulation. Promoters need to be accessible to Pol II and transcription factors (TFs), and in some cases they are assembled on nucleosomes or in higher-order chromatin structures. Opening of chromatin by specific TFs (orange ellipse) and chromatin remodelers (not shown) can be an important component in the recruitment of Pol II. Once Pol II enters, it rapidly initiates and elongates to the promoter-proximal pause sites. These early events fall under the major regulated step of Pol II recruitment. The second major point of regulation is the release of paused Pol II into productive elongation catalyzed by P-TEFb kinase. P-TEFb is recruited by specific TFs (green ellipse) leading to the phosphorylation of multiple components of the Pol II paused complex and Pol II release. Elongating Pol II eventually terminates and is available for another round of transcription at the same or other promoters. Adapted from ref. [84].

promoter-proximal pause. High-resolution mapping of Pol II's location genome-wide using permanganate ChIP-seq[21] and ChIP-exo[30] showed that the vast majority of promoter-associated Pol II is at the pause site and not the site of initiation. This indicates that once Pol II is recruited, it must initiate efficiently and move to the pause site (although exceptions will undoubtedly continue to be documented, as observed during activation of resting lymphocytes, where limited cellular TFIIH results in promoter melting being a key regulatory step[31]). Thus, it is becoming increasingly clear that initiation per se (defined by formation of the first phosphodiester bond) is not usually rate-limiting and is thus unlikely to be a major point of transcription regulation. Paused Pol II must then respond to regulatory signals that are mediated by transcription factors that bind DNA regulatory regions of genes and then recruit, directly or indirectly, P-TEFb kinase. P-TEFb kinase can phosphorylate pausing factors DSIF and NELF, as well as Pol II, leading to release of paused Pol II and productive elongation[32]. Once Pol II vacates the pause region, additional Pol II can be recruited[33,34]. This increased recruitment can itself be modulated by regulatory transcription factors, or it may simply reflect the realization of the intrinsic recruitment rate that is, under uninduced conditions, limited by paused Pol II occupancy of the promoter region. Many promoters are open and primed with paused Pol II, allowing instant access of transcription factors to direct changes in the transcription program during stress and development[35–37]. Thus, the two major steps that are often rate-limiting and regulated appear to be (i) recruitment of Pol II to promoters (long professed by Mark Ptashne[38]) that rapidly leads to initiation and elongation to the promoter-proximal pause, and (ii) P-TEFb-stimulated release of the paused Pol II to productive elongation.

Observe, perturb and re-observe

High-resolution in vivo observations, when coupled to highly specific perturbations, can be used to systematically dissect transcription mechanisms. Advances in biochemical and genomic methods allow molecular interactions and events occurring on the genome to be monitored with high spatial and temporal resolution in cells. Genome-wide assays often afford higher sensitivity and resolution than gene-targeted approaches, and one can acquire this information across entire genomes for a relatively modest cost. The resulting genome-wide information not only tests a hypothesis with statistical rigor, but also provides a wealth of information that often stimulates the generation of new hypotheses.

A plethora of powerful genome-scale assays are available for probing transcription and its regulation. Multiple approaches now allow the tracking of RNA polymerases across the genome with base-pair resolution, including PRO-seq[39] and NET-seq[40] and related methods[41]. TSSs can be efficiently mapped to the base pair genome-wide using nascent RNA in GRO-cap[42] and START-seq[43] assays and variants thereof, whereas pause sites at promoters and enhancers can be mapped at with PRO-seq, START-seq and NET-seq assays. Moreover, metabolic pulse-labeling of RNAs with various ribonucleotide analogs provide critical information genome-wide about RNA synthesis rates, as seen by TT-seq and related methods[44].

The locations of transcription factor binding can also be mapped genome-wide by ChIP-exo[45], CUT&RUN[46] and related methods at near-base-pair resolution if appropriate antibodies are available to the transcription factor or if the transcription factor has been tagged with a polypeptide that permits immunoprecipitation. Similarly, the position of nucleosomes and open chromatin can be determined with high resolution by MNase-seq[47] and ATAC-seq[48], respectively. Chemical probing strategies are now routinely used to recognize specific features of DNA or chromatin, and there is often a methodology that permits genome-wide analysis at base-pair resolution. For example, melted DNA associated with a transcription bubble can be mapped with permanganate-ChIP-seq[21], and non-B form DNA can be mapped by permanganate/S1 footprinting[49]. Finally, the secondary RNA structure of a nascent transcript can be tracked with various RNA probing strategies[50]. An impressive battery of tools already exists to probe RNA synthesis, chromatin structure and chromatin dynamics genome-wide at molecular resolution, and new tools are being generated at a furious pace, as evidenced by proliferation of the '-seq' technologies. Therefore, one can observe the architecture and mechanics of many nuclear structures and functions across genomes.

These methods of observation are becoming routinely coupled with approaches that perturb a process or factor with high specificity. Examining or 're-observing' immediately after perturbation can enable a genome-wide change to be ascribed to a specific cellular mechanism. Simple perturbations include expression modulation of sets of genes with specific inducing or repressing strategies, as in response to heat shock or hormone addition. The molecular events and resulting transcriptional changes associated with these perturbations can then be tracked with high temporal resolution. The location of transcriptional regulatory factors and Pol II during a time course sets limits on any mechanistic model. Moreover, advances in RNA interference (RNAi) knockdown[51], CRISPR-Cas technologies[52], Degrons[53], optogenetics[54], small-molecule inhibitors or drugs[55] and RNA aptamers[56] provide additional powerful options to perturb functions or specific interactions of factors with high specificity in cells at different time scales and can be used in conjunction with gene inducing or repressing stimuli. If the perturbation is rapid and re-observation can be done in the minutes following a perturbation, then one can avoid potential secondary effects associated with long-term treatments like RNAi or mutant approaches.

A simple strategy of 'observe, perturb and re-observe' (OPreO) is now routinely used to gain mechanistic insights into transcription and its regulation in a natural, cellular context (Fig. 2).

Transcription factors regulate Pol II recruitment and pause release

The genome-wide OPreO approaches have revealed mechanistic features of transcriptional regulation. For example, it is now clear that different steps in the transcription cycle can be rate-limiting and thus amenable to regulation. Many genes show no detectable association of Pol II in a particular cell type growth condition[24,25]. Many others show a pileup of Pol II at the promoter-proximal pause sites[26]. The simple interpretation of these genome-wide Pol II occupancy patterns is that recruitment and release of paused polymerase are rate-limited steps that are altered by induced changes in gene expression[36,37].

Transcription factors can act at either Pol II recruitment or pause-release steps. Transfection assays by the Bentley lab showed in the 1990s that Sp1 facilitates Pol II 'initiation', which we would now refer to as recruitment, and Tat fused to a DNA binding domain acts at pause release[57]. More recently, my colleagues and I have examined the genome-wide response to heat shock after perturbing the levels of either GAGA factor or the master heat-shock regulatory factor HSF1, and demonstrated that GAGA factor acts during recruitment and thereby establishes the paused Pol II[36], whereas HSF1 acts to program rapid release from the pause[36,58]. These observations imply that transcription factors can act in synergy to produce a highly controlled and dynamic range of transcription regulation, and that the individual inputs of different cellular signals are integrated at promoters through distinct factor subsets.

Enhancers and chromatin architecture

Any mechanistic model of transcription regulation must account for the fact that many genes are regulated by enhancer elements that reside at variable distances, sometimes as far as a megabase or more, from their promoter targets[59,60]. Enhancer and promoter elements must communicate over these distances, apparently by looping with architectural specificity, allowing only correct target promoters to be activated. This mode of regulation has been intriguing ever since the discovery and pioneering work on enhancers by Schaffner and colleagues[60], and has recently become an area of intense investigation that is currently being pursued by both the 4D Nucleome and ENCODE Consortiums. High-throughput, targeted mutagenesis and CRISPRi targeting followed by readouts of gene expression are being used to rigorously identify these enhancer-promoter regulatory connections[61,62]. However, the underlying mechanisms of enhancer-stimulated transcription from promoters remains to be determined.

A convergence of disciplines

The understanding of transcriptional mechanisms has benefited from attracting researchers from multiple disciplines. Classical biochemistry provided the first key strategies for identification, purification and study of the enzymes and factors that drive and regulate gene expression. These efforts were augmented by forward and reverse genetic approaches to identify genes encoding critical transcription factors and to test their function. As highlighted above, mechanisms are being tested genome-wide in their natural context by 'in-cell biochemistry' and use of targeted perturbation strategies. Finally, recent revolutionary advances in structural biology and high-resolution microscopy are taking the understanding of transcription mechanisms, respectively, to the atomic level and to single molecules, in cells and in real time.

Structures of transcription intermediates

Determining the first X-ray crystal structures of RNA polymerases was a tour de force executed initially in Roger Kornberg's lab[63]. These high-resolution structures provided a first view of these magnificent machines, and revealed how they interact with DNA templates and the RNAs as they are being made. Subsequent structural determination of Pol II complexes with GTFs provided views of how these factors guide polymerases to their target TSSs[64]. The interplay of transcription regulation with chromatin was likewise greatly accelerated by solving the crystal structure of the nucleosome[65].

Now, the revolution in cryogenic electron microscopy (cryo-EM) technologies, made possible with vastly improved detectors and computational approaches, has enabled high-resolution imaging of even larger complexes, exemplified by Nogales lab's tracking of TFIID loading on a promoter[66], and Cramer lab's tracking of reconstructed Pol II complexes that visualize the changes in structure that take place during early steps of transcription. In the latter case, a structure of promoter-proximal paused Pol II with its pause-stabilizing factors, DSIF and NELF, was solved at 3.2Å resolution[67]. This structure provided in near-atomic detail what many of us had been dreaming about for years, and it nicely fit all the predictions of classical biochemical studies and in-cell approaches. A subsequent cryo-EM structure of Pol II in elongation mode, having shed NELF and taken on elongation factors including the PAF1 complex and elongation factor Spt6 (ref. [22]) provided a physical high-resolution image of what happens during this critical regulated step of pause release.

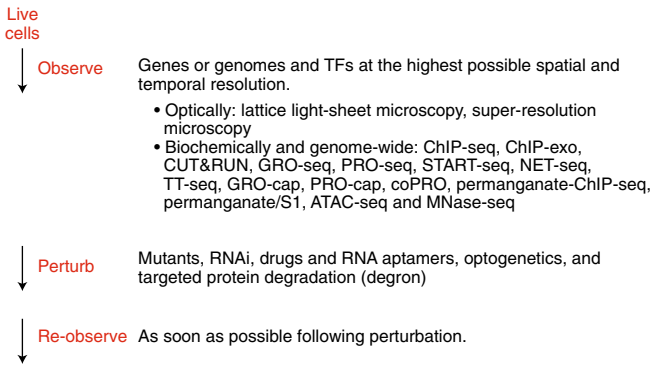

Fig. 2 | Dissecting eukaryotic transcriptional mechanisms in cells: the OPreO strategy. Three steps are observe, perturb and re-observe. See text for discussion of multiple high-resolution observation strategies for examining transcription and associated factor binding genome-wide. See text also for several perturbation strategies.

Microscopy applied to transcription

As 'seeing is believing', verification of a model by visual observation can be very compelling. Early views of transcription were provided by 'Miller spreads' of chromatin by electron microscopy in the 1970s, and conclusions drawn from those images remain undisputed to this day[68]. Amplification and increased resolution of polytene chromosomes provided a means of tracking recruitment and dynamics of factors at specific genetic loci in the seconds following gene activation[69]. It is now possible to take such studies to the level of single transcription factors with lattice light-sheet microscopy[70]. Super-resolution methods can map the location of proteins in cells at a resolution of tens of nanometers, and improvements in sensitivity, provided by more intense fluorescent tags, permit relatively high-resolution tracking of transcription factors in real time[71].

The future: the next 50 years

Researchers have been examining transcription in cells for decades, and have been applying the OPreO approach to discern features of transcription regulation. However, the stage is now set for the

dissection of transcription mechanisms with unprecedented efficiency with the recent development of powerful base-pair resolution methodologies to examine RNA polymerases, transcriptional regulatory factors and chromatin at every position in the genome, including genes, promoters and enhancers. Moreover, the progressive development of sophisticated methods to perturb transcription factors and cofactors provides exquisite means to rigorously dissect their functional mechanisms.

An understanding of enhancers and promoters and their dynamic interplay is essential in defining transcriptional regulation. Enhancers and promoters share a remarkably related architecture: both are often delimited by divergent core promoters and promoter-proximal paused Pol II[37,72–74]. However, these regulatory elements function distinctly. Promoters efficiently generate large pre-RNAs and rarely function to activate genes at a distance[75], whereas enhancers are effective at long-range transcription activation but generally produce only short unstable RNAs. A full understanding of the structural and mechanistic basis of these different functions remains to be resolved. Moreover, the specificity and dynamics governing how particular enhancers couple with particular promoters needs to be deciphered. Finally, it is intriguing to consider how liquid-like condensates of the transcriptional apparatus at enhancers mechanistically influence transcription regulation[76].

Cryo-EM promises to continue to provide near-atomic-resolution snapshots of Pol II and regulatory complexes, as well as critical transitions between regulated steps. This will continue to be accomplished by reconstructing active assemblies from purified components, but it will be equally important to examine complexes assembled in their native, cellular environment. Effective methodologies need to be developed to rapidly and efficiently purify such large macromolecular complexes from cells in forms that are sufficiently homogeneous for cryo-EM visualization.

The spectrum of stable and transient interactions of macromolecular complexes also needs to be examined in context of intact cells and nuclei. Improved crosslinking and mass spectrometry technologies are thus required to allow observation of even transient interactions among Pol II and its associated transcription factors and cofactors. Likewise, the interactions between enhancers and promoters, which may well be transient[77], require highly sensitive and sophisticated capture methodologies.

Finally, the power of real-time microscopy in evaluating dynamic mechanisms of transcription and its regulation cannot be overemphasized. New optical methods are providing increasingly high spatial and temporal resolution of single molecules and their dynamics. The resolution and sensitivity required to visualize individual molecules touching and follow their paths in real time will test existing models and uncover new features of transcription regulation.

Bob Roeder's initial discovery of eukaryotic polymerases initiated a field that has been vibrant for half a century. With the continued development of new ways to examine and dissect the mechanics of transcription and its regulation, the field promises to remain exciting for foreseeable future.

Received: 27 June 2019; Accepted: 26 July 2019;
Published online: 22 August 2019

References

1. Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
2. Roeder, R.G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol.* https://doi.org/10.1038/s41594-019-0287-x (2019).
3. Sawadogo, M. & Roeder, R. G. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci. USA* **82**, 4394–4398 (1985).
4. Matsui, T., Segall, J., Weil, P. A. & Roeder, R. G. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* **255**, 11992–11996 (1980).
5. Burley, S. K. & Roeder, R. G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**, 769–799 (1996).
6. Kornberg, A. Ten commandments: lessons from the enzymology of DNA replication. *J. Bacteriol.* **182**, 3613–3618 (2000).
7. Ruddle, F. H. A new era in mammalian gene mapping: somatic cell genetics and recombinant DNA methodologies. *Nature* **294**, 115–120 (1981).
8. Weeks, J. R., Coulter, D. E. & Greenleaf, A. L. Immunological studies of RNA polymerase II using antibodies to subunits of *Drosophila* and wheat germ enzyme. *J. Biol. Chem.* **257**, 5884–5892 (1982).
9. 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).
10. 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).
11. 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).
12. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**, 489–493 (1987).
13. Eick, D., Kohlhuber, F., Wolf, D. A. & Strobl, L. J. Activation of pausing RNA polymerases by nuclear run-on experiments. *Anal. Biochem.* **218**, 347–351 (1994).
14. 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).
15. Chodosh, L. A., Fire, A., Samuels, M. & Sharp, P. A. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J. Biol. Chem.* **264**, 2250–2257 (1989).
16. Yamaguchi, Y., Wada, T. & Handa, H. Interplay between positive and negative elongation factors: drawing a new view of DRB. *Genes Cells* **3**, 9–15 (1998).
17. Marshall, N. F. & Price, D. H. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol. Cell. Biol.* **12**, 2078–2090 (1992).
18. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
19. Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
20. Wu, C. H. et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414 (2003).
21. Li, J. et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol. Cell* **50**, 711–722 (2013).
22. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018).
23. Winston, F., Chaleff, D. T., Valent, B. & Fink, G. R. Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* **107**, 179–197 (1984).
24. Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
25. Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
26. 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).
27. Core, L. J. et al. Defining the status of RNA polymerase at promoters. *Cell Rep.* **2**, 1025–1035 (2012).
28. 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).
29. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
30. Pugh, B. F. & Venters, B. J. Genomic Organization of Human Transcription Initiation Complexes. *PLoS One* **11**, e0149339 (2016).
31. Kouzine, F. et al. Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988–999 (2013).
32. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
33. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, e29736 (2017).
34. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
35. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**, 471–473 (2009).
36. 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).

37. Vihervaara, A. et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat. Commun.* **8**, 255 (2017).
38. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
39. 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).
40. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
41. Wissink, E.M., Vihervaara, A., Tippens, N.D. & Lis, J.T. Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.* https://doi.org/10.1038/s41576-019-0159-6 (2019).
42. Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T. & Meyer, B. J. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife* **2**, e00808 (2013).
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).
44. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
45. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
46. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
47. Cui, K. & Zhao, K. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol. Biol.* **833**, 413–419 (2012).
48. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
49. Kouzine, F. et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst.* **4**, 344–356 e7 (2017).
50. Mailler, E., Paillart, J. C., Marquet, R., Smyth, R. P. & Vivet-Boudou, V. The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip. Rev. RNA* **10**, e1518 (2019).
51. Hannon, G. J. & Rossi, J. J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
52. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
53. Natsume, T. & Kanemaki, M. T. Conditional Degrons for Controlling Protein Expression at the Protein Level. *Annu. Rev. Genet.* **51**, 83–102 (2017).
54. McDaniel, S. L. et al. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Mol. Cell* **74**, 185–195.e4 (2019).
55. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
56. Shi, H., Hoffman, B. E. & Lis, J. T. RNA aptamers as effective protein antagonists in a multicellular organism. *Proc. Natl. Acad. Sci. USA* **96**, 10033–10038 (1999).
57. Blau, J. et al. Three functional classes of transcriptional activation domain. *Mol. Cell. Biol.* **16**, 2044–2055 (1996).
58. 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).
59. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
60. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
61. Gasperini, M. et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377–390.e19 (2019).
62. Fulco, C. P. et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
63. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
64. Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
65. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).
66. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
67. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601–606 (2018).
68. Laird, C. D. & Chooi, W. Y. Morphology of transcription units in *Drosophila melanogaster*. *Chromosoma* **58**, 193–218 (1976).
69. Zobeck, K. L., Buckley, M. S., Zipfel, W. R. & Lis, J. T. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* **40**, 965–975 (2010).
70. Mir, M. et al. Single Molecule Imaging in Live Embryos Using Lattice Light-Sheet Microscopy. *Methods Mol. Biol.* **1814**, 541–559 (2018).
71. Conic, S. et al. Imaging of native transcription factors and histone phosphorylation at high resolution in live cells. *J. Cell Biol.* **217**, 1537–1552 (2018).
72. Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
73. 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).
74. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
75. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
76. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
77. Chen, H. et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
78. Roeder, R. G. Multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in *Xenopus laevis*. Isolation and partial characterization. *J. Biol. Chem.* **249**, 241–248 (1974).
79. Zawel, L. & Reinberg, D. Initiation of transcription by RNA polymerase II: a multi-step process. *Prog. Nucleic Acid Res. Mol. Biol.* **44**, 67–108 (1993).
80. Peterson, C. L. & Tamkun, J. W. The SWI-SNF complex: a chromatin remodeling machine? *Trends Biochem. Sci.* **20**, 143–146 (1995).
81. Grunstein, M. & Hogness, D. S. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965 (1975).
82. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090 (1987).
83. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197 (2011).
84. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192 (2009).

Acknowledgements

The author thanks members of his lab for valuable comments. The work is supported by US Public Health Service Award R01 GM025232 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests

The author declares no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to J.T.L.

Peer review information: Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.