Published in final edited form as:

Nat Struct Mol Biol. 2019 September ; 26(9): 777–782. doi:10.1038/s41594-019-0288-9.

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

JT Lis¹

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

Abstract

Transcription regulation is critical to organismal development and homeostasis. The control and expression of our twenty thousand genes requires many hundreds of proteins acting through sophisticated multi-step mechanisms. This article highlights the progress that has been made in understanding eukaryotic transcriptional mechanisms using an array of disciplines and how this concerted effort has been driven by the development of new technologies and approaches.

The eukaryotic transcription field reached a landmark with the discovery of the machines that transcribe our genes: RNA polymerases I, II, and III (RNA Pols)¹. The technological power of biochemistry made possible the purification of RNA Pols from complex mixtures in nuclear extracts, and the history surrounding this landmark and the subsequent discovery of the accessory factors that guide the activity of the RNA Pols is clearly and authoritatively presented in the accompanying article by Bob Roeder [Nat.Struct.Mol. Biol]. We have come a long way in our understanding of eukaryotic transcription and in the tools and approaches used for these studies in past 50 years. For example, in these first decades of the transcription era, researchers routinely performed biochemical fractionations often requiring weeks/months of tortuous work done in the bone-numbing confines of the cold room; and now, researchers routinely analyze incredibly-massive DNA and RNA sequencing datasets often requiring weeks/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. Our quest will ultimately culminate in a full understanding of transcription and its regulation at every base of our 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^{2,3}. A full mechanistic understanding of just RNA polymerase II (Pol II) transcription of a single gene would involve 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 repertoire of particular 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, an enormous investment in effort and the development of an array of powerful approaches and methods (Table I) have moved the field at an exhilarating pace.

The early mechanistic studies of transcription depended heavily on classical biochemistry that identified and purified not only the RNA Pols but also General Transcription Factors (GTFs) that direct the RNA Pols to their transcription start sites (TSSs) and that regulated the frequency that the Pols transcribe genes⁴. These biochemical efforts were guided by codes of practice, which were eventually articulated as Ten Commandments by the late Arthur Kornberg⁵. Two commandments were particularly memorable: “IV. DO NOT WASTE CLEAN THINKING ON DIRTY ENZYMES”; V. “DO NOT WASTE CLEAN ENZYMES ON DIRTY SUBSTRATES”. Commandment IV was from the late Efraim Racker, an esteemed colleague in my department at Cornell. Therefore, it was 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 appropriate 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.

The recombinant DNA technology revolution⁶ 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, ‘observe’ transcription and transcription factor interactions in cells. With such probes, Dave Gilmour and I sought to develop methods for tracking Pol II and other factors on their native templates in cells. The idea was to crosslink proteins to DNA and fish out specific complexes. The use of an excellent antibody provided by Arno Greenleaf to RNA Pol II⁷ made it possible to test this idea in *Drosophila*. In heat shocked cells, the expectation was that Pol II would copurify with entire heat shock genes which could be readily quantified with pure gene probes. The resulting signal was indeed greater than from non-heat shock (control) cells, but even the control had a signal from the 5’ end of the genes that represented about one Pol II per gene⁸. Therefore, the Pol II, surprisingly, was already associated with the promoter before the gene was induced. Moreover, follow-up experiments showed the Pol II was transcriptionally engaged and paused 20-50 bases downstream of the TSS^{9,10}. 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 LTR of HIV virus was also acquired by the Peterlin Lab, as was support for the activation of productive Pol II elongation by Tat, the HIV regulatory protein¹¹. Transcriptionally-engaged Pol II on the extreme 5’-ends of Myc and Fos genes in mammalian cells was also discovered by the Eick and Groudine Laboratories^{12,13}, indicating paused Pol II and regulated early elongation was not simply restricted to heat shock genes and HIV promoters.

The effects of the purine nucleoside analog 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB), a kinase inhibitor, provided additional support for transcriptional regulation beyond Pol II recruitment and initiation. While DRB has no effect on in vitro transcription reconstituted from Pol II and GTFs, it is a potent inhibitor of mRNA production in cells¹⁴(reviewed in Yamaguchi et al¹⁵). 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 Lab¹⁶. P-TEFb stimulated transcription in vitro in a DRB sensitive manner. Other critical components were the two multiprotein complexes, DRB Sensitivity-Inducing Factor (DSIF)¹⁷ and negative elongation factor (NELF)¹⁸, which were purified by the Handa Lab. The Gilmour and Handa labs showed compellingly that DSIF and NELF when added to in vitro transcription systems, cause promoter-proximal pausing in vitro^{19,20}. The P-TEFb kinase can phosphorylate in vitro multiple sites on components of the paused complex (including Pol II, DSIF, NELF) to stimulate paused Pol II to enter into productive elongation²¹ (and references therein) (Figure 1). Interestingly the DSIF subunits, Spt4 and Spt5, were also uncovered in a genetic selection by Fred Winston²² – a selection 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 needs 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 (Figure 1).

With the revolution in genome-wide methods, the molecular features associated with particular types of regulation can be observed across all genes. The Adelman, Levine, and Young labs demonstrated that Pol II accumulated at many promoter regions initially using ChIP-chip assays where oligonucleotides of *Drosophila* and human genes on a microarray were used to probe the genomic DNA fragments associated with immuno-precipitated Pol II^{23,24}. Genome-wide GRO-seq assays demonstrated that a majority of promoters in human and *Drosophila* cells had the properties of promoter-proximal paused Pol II that had been associated with the extensively characterized *Drosophila Hsp70 gene*^{25,26}. 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 analyzed²⁷. 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 studies from many laboratories made it clear that the process of transcription and its regulation is a multistep odyssey for Pols, and especially for Pol II whose transcripts are extensively processed during transcription (Figure 1). The steps of transcription cycle begin with ensuring chromatin is open, a job executed by chromatin remodelers and pioneer factors that can bind specific DNA sequences on nucleosomes²⁸. This allows the recruitment to the promoter of Pol II (ushered by GTFs), initiation, and early elongation to the promoter-proximal pause. High resolution mapping of Pol II's location genome-wide using permanganate-ChIP-seq²⁰ and ChIP-exo²⁹ show 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 must, it must initiate efficiently and move to the pause site (although exceptions have been and will undoubtedly continue to be documented as in the activation of resting lymphocytes where limited cellular TFIID results in promoter melting

being a regulatory step³⁰). Thus, it appears that initiation *per se* formation of the first phosphodiester bond is not usually rate-limiting and unlikely to be a major point of transcription regulation. Paused Pol II must then respond to regulatory signals that are mediated by transcription factors (TFs) 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, and Pol II, leading to release of paused Pol II and productive elongation³¹. Once Pol II vacates the pause region, additional Pol II can be recruited^{32,33}. This increased recruitment can itself be modulated by the regulatory TFs, 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 TFs and changes in transcription program during stress and development³⁴⁻³⁶. Therefore, two major steps that are often rate-limiting and regulated appear to be 1) the recruitment of Pol II to promoters (long professed by Mark Ptashne³⁷) that rapidly leads to initiation and elongation to the promoter-proximal pause, and 2) 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 observed with high spatial and temporal resolution in cells. Genome-wide assays often work with 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 hypothesis with statistical rigor but provides a wealth of information that often stimulate the generation of new hypotheses.

A plethora of powerful genome-scale assays are available for probing transcription and its regulation. Multiple assays now allow the tracking of RNA Pol across the genome with base pair resolution, including PRO-seq³⁸ and NET-seq³⁹ and related methods [see Wissink et al. in press Nat Reviews]. TSSs can be efficiently mapped to the base pair genome-wide using nascent RNA in GRO-cap⁴⁰ and START-seq⁴¹ assays and variants thereof, while pause sites at promoters and enhancers can be mapped at base pair resolution with PRO-seq, START-seq, and NET-seq assays. Moreover, metabolic pulse-labeling of RNAs with various NTP analogues provide critical information genome-wide about RNA synthesis rates as seen by TT-seq and related methods⁴².

The locations of transcription factor (TF) binding can also be mapped genome-wide by ChIP-exo⁴³, CUT&RUN⁴⁴ and related methods at near base-pair resolution, if appropriate antibodies are available to the TF or if the TF has been tagged with an IP compatible polypeptide. Similarly, the position of nucleosomes and open chromatin can be determined with high resolution by MNase-seq⁴⁵ and ATAC-seq⁴⁶, respectively. Chemical probing strategies can be used to recognize specific features of DNA or chromatin and there is often a methodology that takes the analysis genome-wide and to base-pair resolution. For example, melted DNA associated with a transcription bubble can be mapped genome-wide with permanganate-ChIP-seq²⁰, and non-B form DNA can be mapped with

permanganate/S1 footprinting⁴⁷. Finally, the RNA structure in nascent RNA can be tracked with various RNA probing strategies⁴⁸. An impressive battery of tools already exist 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 with high resolution 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 a perturbation can ascribe a genome-wide change to a particular mechanistic process. Simple perturbations can be the activation or repression of sets of genes with specific inducing or repressing strategies as in response to heat shock or hormone addition. Then, molecular events and resulting transcriptional changes that are associated with these perturbations can be tracked at high temporal resolution. The location of factors and Pol II during a time course sets limits on any mechanistic model. Even more importantly, advances in RNAi knock-down⁴⁹, CRISPR-cas technologies⁵⁰, Degrons⁵¹, optogenetics⁵², small molecule inhibitors/drugs⁵³ and RNA aptamers⁵⁴ provide multiple powerful options to perturb the 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 quick and re-observation can be done in the minutes following a perturbation, then one can avoid the potential secondary effects associated with long term treatments like RNAi or mutant approaches.

A simple strategy of “Observe, Perturb and re-Observe” (OPreO [ap re o]) is now being used routinely to gain mechanistic information to transcription and its regulation in the natural context, in cells (Figure 2). The high specificity and resolution of these approaches can be used both to investigate processes de novo and to critically test molecular mechanisms of transcription and its regulation.

TFs regulate recruitment & pause release

The genome-wide OPreO approaches are revealing mechanistic features of transcription regulation. For example, it's clear that different steps in the transcription cycle can be rate-limiting and regulated. Many genes show no detectable association of Pol II in a particular cell line and growth condition^{23,24}. Many others show a pileup of Pol II at the promoter proximal pause sites²⁵. The simple interpretation of these genome-wide patterns is that recruitment and release of paused polymerase are rate-limited steps. Interestingly, these steps can be altered by induced changes in gene expression, and thus these steps appear to be points of transcription regulation^{35,36}.

TFs can act at either Pol II recruitment or pause release. The Bentley lab showed in transfection assays in the 90s that Sp1 seemed to act at “initiation”, which we would now refer to as recruitment, while Tat fused to a DNA binding domain can act at pause release⁵⁵. More recently, we have examined the genome-wide response to heat shock after perturbing the levels of GAGA-factor or the master heat shock regulatory factor, HSF1. The resulting patterns indicate GAGA-factor acts at recruitment and thereby establishing the paused Pol II³⁵, while HSF1 acts at the level of programing rapid release from the pause^{35,56}. These

observations imply that TFs can synergize to produce a highly-controlled and large dynamic range of transcription regulation, and that different cellular signals that activate different factors can have their effects integrated at promoters, with TFs and genes acting effectively as processors of different input signals.

Enhancers & chromatin architecture

Any mechanism of transcription regulation must account for the fact that many genes can be regulated by enhancer elements that reside variable distances, sometimes a mega-base or more, from the promoters they regulate^{57,58}. Enhancer and promoter elements must communicate over these distances, apparently by looping with architectural specificity, allowing only correct target promoters to be regulated. This has been intriguing ever since the discovery and pioneering work on enhancers by Schaffner and colleagues⁵⁸, and has recently become an area of intense investigation that is being pursued by both the 4D Nucleome and ENCODE Consortiums. High throughput targeted mutagenesis and CRISPRi targeting followed by read-outs of gene expression are rigorously identifying these enhancer-promoter regulatory connections^{59,60}. However, a full understanding of underlying mechanisms of enhancer stimulated transcription from promoters are lacking (see The Future)

A convergence of disciplines

Our 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 transcription. This was augmented by genetic approaches, both forward and reverse, that identify genes encoding critical transcription factors and transcriptional machines and test their function. And, as highlighted above, mechanisms are being tested genome-wide in their natural context with ‘in-cell biochemistry’ and use of targeted perturbation strategies. Finally, recent revolutionary advances in structural biology and high-resolution microscopy are taking our understanding of transcription mechanisms, respectively, to the atomic level and to single molecules in cells and in real time.

Structure of transcription intermediates

Determining the first x-ray crystal structures of RNA polymerases was a tour de force executed initially in Roger Kornberg’s lab⁶¹. These high resolution structures provided a first view for these magnificent machines and how they interacted with DNA templates and the RNAs as they are being made. Subsequent studies of Pol II complexes with GTFs provided views as to how these factors guide Pols to their target TSSs⁶². The interplay of transcription regulation with chromatin was likewise greatly accelerated by solving the crystal structure of the nucleosome⁶³

Now, the revolution in cryo-EM technologies, made possible with vastly improved detectors and computational approaches, has enabled high resolution imaging of even larger complexes, as in the Nogales Lab tracking of TFIID loading on a promoter⁶⁴, and the Cramer Lab tracking of reconstructed Pol II complexes that represent the changes in

structure accompanying the early steps of transcription. In the latter, a structure of the promoter-proximal paused Pol II with its pause stabilizing factors, DSIF and NELF was solved at 3.2Å⁶⁵. 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. The paper was accompanied by a second paper that showed Pol II in its elongation mode, having shed NELF and taken on elongation factors including the PAF1 complex and elongation factor Spt6²¹. The reconstructed states provided a physical high resolution model of what happens during this critical regulated step of pause release.

Microscopy applied to transcription

“Seeing is believing”: the verification of a model by visual observation can be very compelling. Early views of transcription and accompanying conclusions by examining ‘Miller spreads’ of chromatin by electron microscopy in the 70’s remain undisputed to this day⁶⁶. The amplification and higher resolution of polytene chromosomes provided a means of tracking the recruitment and dynamics of factors at specific genetic loci in the seconds following gene activation⁶⁷. Taking such studies to observing single TFs is now possible with Lattice light-sheet microscopy⁶⁸. Super resolution methods can map the location of proteins in cells at a resolution of tens of nanometers, and improvements in sensitivity with more intense fluorescent tags can even allow relatively high-resolution tracking of transcription factors in real time⁶⁹.

The Future: the next 50 years

Researchers have been examining transcription in cells for decades, in a sense applying the OPreO approach, and learning about features of transcription regulation. However, the stage has been set for the dissection of transcription mechanisms with unprecedented efficiency across genomes with the recent development of powerful of base-pair resolution methodologies to examine transcription, TFs, and chromatin at every position in the genome, including genes, promoters and enhancers. Moreover, the recent and continued development of sophisticated methods to perturb TFs and cofactors provide exquisite means of rigorously dissecting the function and mechanisms of TFs.

Both enhancers and promoters are transcribed and share a related architecture, where both are transcribed and often delimited by divergent core promoters and promoter-proximal paused Pol II^{36,70-72}. Yet, these regulatory elements function distinctly with only a small fraction of promoters serving as enhancers⁷³, and with promoters having evolved a more potent ability to efficiently transcribe stable pre-mRNA. 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 regulation⁷⁴.

Cryo-EM promises to continue to be effective at providing in some cases atomic resolution snapshots of Pol II and regulatory complexes, as well as at critical transitions between regulated steps. This will continue to be done in reconstructions from purified components,

but it will be important to examine by cryo-EM complexes assembled in their native environment – in cells. Effective methodologies need to be developed to rapidly and efficiently purify these large macromolecular complexes from the cells in forms homogeneous enough for cryo-EM.

The spectrum of stable and transient interactions of macromolecular complexes needs to be examined in context of intact cells and nuclei. Improved crosslinking and mass spectrometry technologies need to be developed that allow observation of even transient interactions. Pol II and TFs interact with many cofactors and many of these interactions are likely to be transient. Likewise, the interactions between enhancers and promoters, which may well be transient⁷⁵, 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 have evolved rapidly and are providing increasingly high spatial and temporal resolution of single molecules and their dynamics. The resolution and sensitivity to see individual molecules touching and follow their paths in real-time should test existing models and uncover new features of transcription regulation.

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

References

1. Roeder RG & Rutter WJ Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224, 234–7 (1969). [PubMed: 5344598]
2. Sawadogo M & Roeder RG Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc Natl Acad Sci U S A* 82, 4394–8 (1985). [PubMed: 3925456]
3. Matsui T, Segall J, Weil PA & Roeder RG Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem* 255, 11992–6 (1980). [PubMed: 7440580]
4. Burley SK & Roeder RG Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* 65, 769–99 (1996). [PubMed: 8811195]
5. Kornberg A Ten commandments: lessons from the enzymology of DNA replication. *J Bacteriol* 182, 3613–8 (2000). [PubMed: 10850972]
6. Ruddle FH A new era in mammalian gene mapping: somatic cell genetics and recombinant DNA methodologies. *Nature* 294, 115–20 (1981). [PubMed: 6272118]
7. Weeks JR, Coulter DE & Greenleaf AL Immunological studies of RNA polymerase II using antibodies to subunits of *Drosophila* and wheat germ enzyme. *J Biol Chem* 257, 5884–92 (1982). [PubMed: 6175634]
8. Gilmour DS & Lis JT RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol Cell Biol* 6, 3984–9 (1986). [PubMed: 3099167]
9. Rougvie AE & Lis JT The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* 54, 795–804 (1988). [PubMed: 3136931]
10. Rasmussen EB & Lis JT In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc Natl Acad Sci U S A* 90, 7923–7 (1993). [PubMed: 8367444]
11. Kao SY, Calman AF, Luciw PA & Peterlin BM Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* 330, 489–93 (1987). [PubMed: 2825027]

12. Eick D, Kohlhuber F, Wolf DA & Strobl LJ Activation of pausing RNA polymerases by nuclear run-on experiments. *Anal Biochem* 218, 347–51 (1994). [PubMed: 8074291]
13. 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, 2201–13 (1992). [PubMed: 1427080]
14. Chodosh LA, Fire A, Samuels M & Sharp PA 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J Biol Chem* 264, 2250–7 (1989). [PubMed: 2914905]
15. 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). [PubMed: 9581978]
16. Marshall NF & Price DH Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* 12, 2078–90 (1992). [PubMed: 1569941]
17. 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, 343–56 (1998). [PubMed: 9450929]
18. Yamaguchi Y et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97, 41–51 (1999). [PubMed: 10199401]
19. Wu CH et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev* 17, 1402–14 (2003). [PubMed: 12782658]
20. Li J et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol Cell* 50, 711–22 (2013). [PubMed: 23746353]
21. Vos SM et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* 560, 607–612 (2018). [PubMed: 30135578]
22. Winston F, Chaleff DT, Valent B & Fink GR Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* 107, 179–97 (1984). [PubMed: 6329902]
23. Muse GW et al. RNA polymerase is poised for activation across the genome. *Nat Genet* 39, 1507–11 (2007). [PubMed: 17994021]
24. Zeitlinger J et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39, 1512–6 (2007). [PubMed: 17994019]
25. Core LJ, Waterfall JJ & Lis JT Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–8 (2008). [PubMed: 19056941]
26. Core LJ et al. Defining the status of RNA polymerase at promoters. *Cell Rep* 2, 1025–35 (2012). [PubMed: 23062713]
27. Jonkers I, Kwak H & Lis JT Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 3, e02407 (2014). [PubMed: 24843027]
28. Mayran A & Drouin J Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* 293, 13795–13804 (2018). [PubMed: 29507097]
29. Pugh BF & Venter BJ Genomic Organization of Human Transcription Initiation Complexes. *PLoS One* 11, e0149339 (2016). [PubMed: 26866362]
30. Kouzine F et al. Global regulation of promoter melting in naive lymphocytes. *Cell* 153, 988–99 (2013). [PubMed: 23706737]
31. Peterlin BM & Price DH Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* 23, 297–305 (2006). [PubMed: 16885020]
32. Gressel S et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* 6(2017).
33. Shao W & Zeitlinger J Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* 49, 1045–1051 (2017). [PubMed: 28504701]
34. Boettiger AN & Levine M Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* 325, 471–3 (2009). [PubMed: 19628867]
35. Duarte FM et al. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev* 30, 1731–46 (2016). [PubMed: 27492368]
36. Vihervaara A et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat Commun* 8, 255 (2017). [PubMed: 28811569]

37. Ptashne M & Gann A Transcriptional activation by recruitment. *Nature* 386, 569–77 (1997). [PubMed: 9121580]
38. Kwak H, Fuda NJ, Core LJ & Lis JT Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950–3 (2013). [PubMed: 23430654]
39. Churchman LS & Weissman JS Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–73 (2011). [PubMed: 21248844]
40. Kruesi WS, Core LJ, Waters CT, Lis JT & Meyer BJ Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* 2, e00808 (2013). [PubMed: 23795297]
41. Nechaev S et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* 327, 335–8 (2010). [PubMed: 20007866]
42. Schwalb B et al. TT-seq maps the human transient transcriptome. *Science* 352, 1225–8 (2016). [PubMed: 27257258]
43. Rhee HS & Pugh BF Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–19 (2011). [PubMed: 22153082]
44. Skene PJ & Henikoff S An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6(2017).
45. Cui K & Zhao K Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol Biol* 833, 413–9 (2012). [PubMed: 22183607]
46. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213–8 (2013). [PubMed: 24097267]
47. 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). [PubMed: 28237796]
48. Mailler E, Paillart JC, Marquet R, Smyth RP & Vivet-Boudou V The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip Rev RNA* 10, e1518 (2019). [PubMed: 30485688]
49. Hannon GJ & Rossi JJ Unlocking the potential of the human genome with RNA interference. *Nature* 431, 371–8 (2004). [PubMed: 15372045]
50. Pickar-Oliver A & Gersbach CA The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol* (2019).
51. Natsume T & Kanemaki MT Conditional Degrons for Controlling Protein Expression at the Protein Level. *Annu Rev Genet* 51, 83–102 (2017). [PubMed: 29178817]
52. McDaniel SL et al. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Mol Cell* 74, 185–195 e4 (2019). [PubMed: 30797686]
53. Bensaude O Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2, 103–108 (2011). [PubMed: 21922053]
54. Shi H, Hoffman BE & Lis JT RNA aptamers as effective protein antagonists in a multicellular organism. *Proc Natl Acad Sci U S A* 96, 10033–8 (1999). [PubMed: 10468557]
55. Blau J et al. Three functional classes of transcriptional activation domain. *Mol Cell Biol* 16, 2044–55 (1996). [PubMed: 8628270]
56. Mahat DB, Salamanca HH, Duarte FM, Danko CG & Lis JT Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Mol Cell* 62, 63–78 (2016). [PubMed: 27052732]
57. Schoenfelder S & Fraser P Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* (2019).
58. Banerji J, Rusconi S & Schaffner W Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308 (1981). [PubMed: 6277502]
59. Gasperini M et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* 176, 377–390 e19 (2019). [PubMed: 30612741]
60. Fulco CP et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769–773 (2016). [PubMed: 27708057]

61. Cramer P, Bushnell DA & Kornberg RD Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–76 (2001). [PubMed: 11313498]
62. Kostrewa D et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323–30 (2009). [PubMed: 19820686]
63. Luger K, Mader AW, Richmond RK, Sargent DF & Richmond TJ Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–60 (1997). [PubMed: 9305837]
64. Patel AB et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* 362(2018).
65. Vos SM, Farnung L, Urlaub H & Cramer P Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* 560, 601–606 (2018). [PubMed: 30135580]
66. Laird CD & Chooi WY Morphology of transcription units in *Drosophila melanogaster*. *Chromosoma* 58, 193–218 (1976). [PubMed: 826377]
67. Zobeck KL, Buckley MS, Zipfel WR & Lis JT Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol Cell* 40, 965–75 (2010). [PubMed: 21172661]
68. Mir M et al. Single Molecule Imaging in Live Embryos Using Lattice Light-Sheet Microscopy. *Methods Mol Biol* 1814, 541–559 (2018). [PubMed: 29956254]
69. 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). [PubMed: 29440513]
70. Kim TK et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–7 (2010). [PubMed: 20393465]
71. Core LJ et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46, 1311–20 (2014). [PubMed: 25383968]
72. Henriques T et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* 32, 26–41 (2018). [PubMed: 29378787]
73. Arnold CD et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–7 (2013). [PubMed: 23328393]
74. Sabari BR et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361(2018).
75. Chen H et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* 50, 1296–1303 (2018). [PubMed: 30038397]
76. Roeder RG Multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in *Xenopus laevis*. Isolation and partial characterization. *J Biol Chem* 249, 241–8 (1974). [PubMed: 4855626]
77. Zawal L & Reinberg D Initiation of transcription by RNA polymerase II: a multi-step process. *Prog Nucleic Acid Res Mol Biol* 44, 67–108 (1993). [PubMed: 8434126]
78. Peterson CL & Tamkun JW The SWI-SNF complex: a chromatin remodeling machine? *Trends Biochem Sci* 20, 143–6 (1995). [PubMed: 7770913]
79. Grunstein M & Hogness DS Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci U S A* 72, 3961–5 (1975). [PubMed: 1105573]
80. Kadonaga JT, Carner KR, Masiarz FR & Tjian R Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 51, 1079–90 (1987). [PubMed: 3319186]
81. Lander ES Initial impact of the sequencing of the human genome. *Nature* 470, 187–97 (2011). [PubMed: 21307931]

Figure 1.

The transcription cycle consists of distinct steps that are targets of regulation. Promoters need to be accessible to Pol II and 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 (see text). These early events fall under the major regulated step of Pol II recruitment. The second major point of regulation is the P-TEFb kinase catalyzed release of paused Pol II into productive elongation. 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.

Figure 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.

Table I:

Major approaches & methods that have driven transcription regulation

Approach	Conceptual Advance/Breakthrough	Select Example(s)
Classical biochemistry	Allowed systematic identification and purification of key proteins starting with cell or nuclear extracts by biochemical fractionation and activity assays.	Purification of RNA polymerases I II & III ⁷⁶ Purification of Pol II initiation factors ⁷⁷ Purification of P-TEFb ¹⁶
Genetic forward screens	Allowed selection of genes that are part of a particular molecular transcriptional machine or affect a specific process	Independent screens produced swi and snf mutants led to discovery of the SWI/SNF chromatin remodeler ⁷⁸ Suppressors screens of Ty element insertions led to the identification of a collection of TFs, the Spts ²²
Recombinant DNA technologies	Allowed isolation of genes and regulatory elements for pure templates and probes of in vivo gene function and mechanisms.	Isolation of specific genes (⁷⁹) and to be used as templates and to express transcription factors ⁸⁰
Probing specific protein-DNA interactions 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 ⁸
Nucleic acid sequencing advances.	Allowed first view organizational principles of genomes of humans and model organisms And, a basis for organizing biological information across model organism.	Value of sequencing of the human genome assessed broadly on 10 th anniversary – most importantly it provided all the components need to study transcription ⁸¹
Technologies made possible by Next Generation Sequencing	Genome sequencing allowed adaptation of target gene probings to be executed genome-wide and often with base pair resolution. This meant that mechanistic features of transcription could be viewed in vivo.	ChIP-exo maps position of TF binding with base pair resolution ⁴³ . PRO-seq maps transcriptionally-compenent Pol II with base pair resolution ³⁸
Advances in structural biology	Transcription studies greatly benefited from crystal structures of Pols and the nucleosome. Now even larger complexes are being efficiently solved by Cryo-EM thanks to improvements in detectors and computational processing of images.	X-ray structure of RNA Pol II ⁶¹ . Cryo-EM structure of the RNA Pol II paused complex ⁶⁵
Advances in light microscopy	Super resolution microscopy & single molecule tracking in real time allows one to track the dynamics of transcription, TF binding, and enhancer-promoter interactions.	Tracking single TFs ⁶⁸ Watching enhance-promoter interactions in real time ⁷⁵