10
Transcriptional Initiation: Frequency, Bursting, and Transcription Factories

Kieran Finan and Peter R. Cook

10.1
Transcription in Mammalian Nuclei

We know a great deal about the relative population-averaged rates of transcriptional initiation at many promoters, but little about absolute rates, and even less about the temporal distributions of initiations at single loci. Such data has now begun to accumulate; recently developed techniques such as RNA fluorescence *in situ* hybridization (FISH) and the MS2-GFP transcript-tagging system now allow single transcripts to be counted in single cells and in real time. Recent efforts have combined these methods with mathematical modeling to provide evidence that the transcriptional activity of a given gene can vary widely from cell to cell and from minute to minute; many so-called "active" genes seem to spend much of their time inactive, before switching to produce a brief "burst" of transcripts. We briefly review the basic mechanisms of transcription, before focusing on initiation rates. We discuss recent studies of initiation and relate findings to known mechanisms of regulation (concentrating on results obtained in mammalian systems).

10.1.1
General Introduction

The existence of RNA polymerases (RNAPs) in mammals was first demonstrated in 1959, when it was shown that isolated rat liver nuclei could incorporate [^{32}P] CTP into RNA in the presence of ATP, UTP, and GTP [1]. Shortly thereafter, three types of DNA-dependant RNA polymerizing enzymes (named I, II, III) were isolated from soluble cell lysates on the basis of differential binding to DEAE-cellulose columns, differential sensitivity to α-amanitin, and different structural properties [2]. As active RNA polymerases are insoluble and not easily extracted (see [3] and Section 10.4.4), one may conclude that these preparations likely contained a combination of unengaged enzyme and fragments of native polymerizing complexes that were detached by harsh buffers. Nevertheless, 50 years of

Genome Organization And Function In The Cell Nucleus, First Edition. Edited by Karsten Rippe.

Published 2011 by WILEY-VCH Verlag GmbH & Co. KGaA

exhaustive study has shown that these three enzymes perform the overwhelming majority of RNA synthesis in the nucleus [4].

RNAPs I, II, and III are of similar size (590, 550, and 690 kDa, respectively), contain 14, 12, and 17 subunits and are homologous to one another (e.g., five subunits are shared between all three) [5–7]. Various experimental approaches, including ChIP-seq, promoter analysis, and specific inhibition have shown that each RNAP specializes in transcribing a specific subset of genes, and is active at distinct sites in the nucleus [2] (see Section 4.4).

RNAP I is located in the nucleolus, where it transcribes rDNA transcription units (each encoding a copy of 18S, 28S, and 5.8S rRNA) [8]. It initiates at a very high frequency (once every 1.4 s at active rDNA promoters), and then elongates at ~100 bp/s [9]. RNAP III is located in the nucleoplasm and transcribes many small non-protein-coding RNAs (~300–500 transcription units detectable in mammals by ChIP-seq), including structural RNAs involved in translation (i.e., tRNA, 5S rRNA) and splicing (i.e., U6 RNA), as well as regulatory RNAs, including some miRNAs [10–11].

RNAP II transcribes most transcription units in mammalian genomes; ChIP-seq reveals at least 25 000 RNAP II promoters in HeLa cells (~2 orders of magnitude more than for all other RNAPs combined) [12]. Sequencing of mouse RNA reveals that ~60% of this genome is transcribed [13], and the overwhelming majority of this expression is probably performed by RNAP II.

The properties of purified and soluble RNA polymerases have been studied exhaustively [14]. While the core enzyme alone tends not to initiate specifically at promoters, the presence of additional proteins (i.e., transcription factors) can increase both the specificity and rate of RNAP II initiations *in vitro* [2, 15]. How this *in vitro* data acquired with soluble proteins relates to that obtained with the insoluble enzymes that are active *in vivo* is not fully understood.

RNAP II initially binds to promoters unstably and can often dissociate without producing a full-length transcript [16]. The transition from a transient "initiation complex" to an "elongation complex" (which is extremely stable and can transcribe up to several thousand base pairs) [17] is not well understood. After binding, RNAP II undergoes multiple rounds of "abortive initiation," transcribing the bases directly following the transcriptional start site multiple times before escaping into the rest of the transcription unit [15]. The transition to elongation also involves phosphorylation of the C-terminal domain (CTD) of the largest catalytic subunit of RNAPII, a long disordered tail containing (in humans) 52 tandem heptad repeats with the consensus sequence of YSPTSPS. The degree and type of CTD phosphorylation changes during the transcription cycle, and these probably underlie the recruitment of different RNA processing enzymes (i.e., the capping and splicing machineries) [15].

10.2
Transcription Is an Infrequent Event

In a HeLa cell, quantitative "western blotting" indicates ~65 000 molecules of RNAP II [3, 18] are active, while ChIP-seq reveals ≥25 000 RNAP II binding sites

per genome [12]. As this cell contains 3–4 chromosomal sets [19], one may conclude that there cannot be many active RNAPs on a typical transcription unit. This conclusion is at odds with what many of us were taught: that active genes are usually covered with elongating RNAPs, as in electron micrographs of "Miller spreads" which show ~100 active polymerases tightly packed on a gene (Figure 10.1a). However such micrographs typically show rRNA genes, which are the most active in the cell and which are transcribed by a special polymerase (i.e., RNAP I) in the nucleolus (Chapter 12). A systematic study of the other engaged RNAPs (i.e., mainly RNAP II) in such spreads reveals that ~66% have no neighbor within 12 kb (Figure 10.1b) [18].

Analysis of transcript numbers in mouse cells at different stages of development is consistent with these results (Figure 10.2). About half of all transcripts are present at less than one copy per cell regardless of developmental stage, and it is rare for a transcript to be present at >10 copies [21]. It is not the case that poorly expressed genes are simply "inactive"; studies in yeast show that many are essential [22].

Similarly low numbers of active polymerases/gene and transcripts/cell are found throughout the evolutionary tree. In *Saccharomyces cerevisiae*, microarray analyses reveal that ~80% mRNAs are present at 0.1–2.0 molecules/cell [23], with only 203 of 4942 ORFs being occupied by >0.5 RNAPs [24]. Although these studies

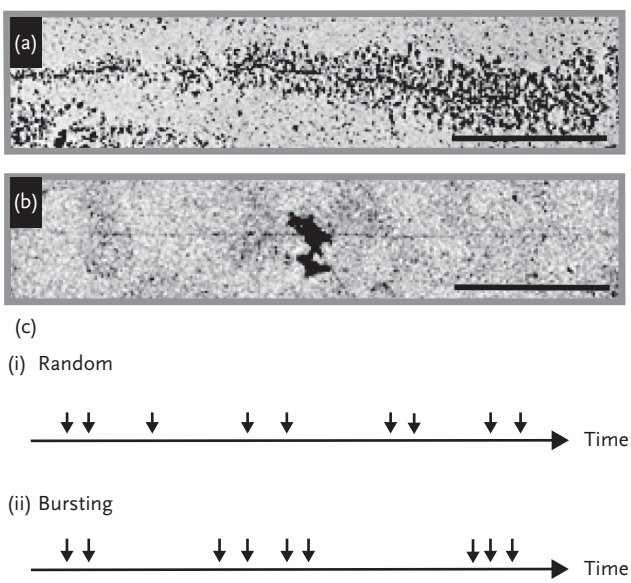

Figure 10.1 Electron micrographs of "genes in action" (from HeLa cells). Bars: 1 μm. (a) A "Miller" spread of an active ribosomal cistron; about 125 RNAP I complexes are engaged on the cistron. From [20] with permission of the Society of the European Journal of Endocrinology. (b) Analogous spread of a typical gene, probably being transcribed by one RNAP II complex. From [18]. (c) Two models: Initiations (arrows) occur randomly over time. This model gives the patterns in (a) and (b) if initiations occur frequently or rarely, respectively. Alternatively, initiations occur in "bursts"; the presence of a transcribing polymerase makes it more likely that another will initiate. This model could again give both patterns depending on the frequency of initiations within each burst.

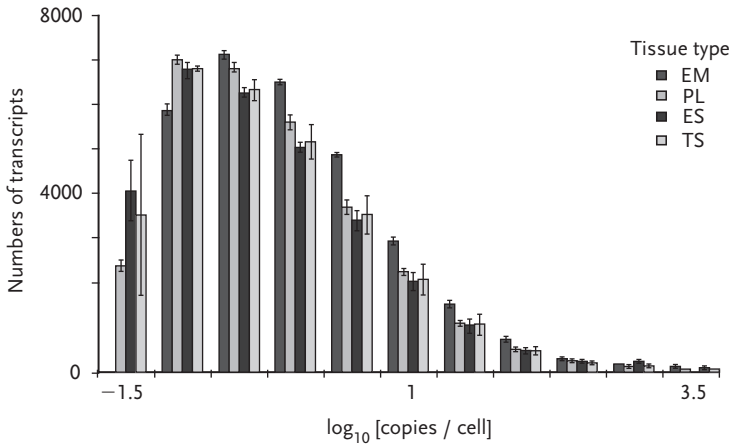

Figure 10.2 The number of different transcripts found at particular copy numbers in four different mouse cell types (EM = E12.5 embryo, PL = E12.5 placenta, ES = embryonic stem cells, TS = trophoblast stem cells). Most transcripts are present at ~1 copy per cell. Adapted from [21].

may have underestimated the number of transcripts and RNAPs/gene by a factor of 3–6 [25], including the appropriate correction does not significantly change the basic conclusion. Moreover, thorough measurements of a small number of yeast mRNAs using qRT-PCR shows a range of 0.001–100.0 per cell [22]. In *Escherichia coli*, where the original "Miller" spreads revealed 60–80 RNAPs closely packed on each ribosomal cistron, the overwhelming majority of RNAPs active on protein-coding genes were spaced 10–20 genes away from their nearest neighbor [26–27]. This view has been confirmed by independent analyses. As the numbers and degradation rates of transcripts are known, the rates of transcription (and so number of polymerases/gene) necessary to maintain transcript levels in the face of known degradation can be estimated; the conclusion is that nearly all active genes are transcribed at any moment by <1 RNAP, with many being transcribed less than once per cell cycle [24]. Similarly, when the results of RNAP ChIP on chip experiments [28] are normalized to the total number of RNAPs in the cell (in minimal media there are ~1500 RNAPs/genome [29, 30]), only ~80 genes have >1 RNAP, and only seven have >2 (excluding ribosomal cistrons; K. Finan, unpublished data). Although measurements of transcript or polymerase numbers can vary between studies, the general conclusion is always the same – most genes are not being transcribed most of the time.

10.3 Transcription Is Noisy

It is clear that, for an mRNA with an average copy number of less than one, different cells contain different numbers of transcripts. However even if the average

copy number is higher, it remains likely that a large variation in transcript copy number exists throughout a population. This is because the cell usually does not regulate the precise number of transcripts produced, but rather the average rate (or probability per unit time) at which transcriptional events occur. Transcription is typically *stochastic*: although it occurs with a defined probability per unit time, it also depends on chance. Thus, variability in molecular copy numbers (also referred to as noise) is an inevitable consequence of stochastic cellular processes. As we will see below, the type and degree of noise produced during transcription can reflect underlying regulatory mechanisms.

10.3.1
Extrinsic and Intrinsic Noise

Stochastic initiation is not the only source of noise contributing to variations in transcript copy numbers. Before discussing the various studies that have examined the stochastic nature of transcription, it is instructive to consider some of the origins of copy number variation. Let us model transcription as a simple chemical reaction, where n is the RNA copy number, and R is the transcription rate. We can then write:

$$n \xrightarrow{R(a,b,c,d)} n+1$$

where R depends on multiple factors, including concentrations of RNAPs and transcription factors, gene copy number, and so on (a,b,c,d...) [31]. R varies between cells – even in clonal populations – leading to variations in transcription rates and in the numbers of RNAs and proteins per cell. This type of variation (noise) is said to be extrinsic to transcription, because it does not originate in transcription itself, but is transmitted from upstream factors. [For the sake of simplification, we will generally neglect downstream factors (e.g., degradation) which also contribute to noise in protein/transcript copy numbers that is extrinsic to transcription].

However, even if the value of R is the same in every cell in a population, one would still expect some heterogeneity due to the stochastic nature of transcription. This noise is "intrinsic" to transcription, as it is not transmitted from upstream factors; rather, it originates from the probabilistic nature of the process. Whether noise in mRNA copy number is extrinsic or intrinsic to transcription can be easily distinguished experimentally. Let us consider a diploid organism, with one copy of a particular locus expressing CFP, and the other expressing YFP under the control of the same promoter (Figure 10.3) [32]. If noise were only extrinsic (i.e., transcription rates were different in every cell), then the numbers of CFP and YFP transcripts would be the same in every cell, and this number would vary between cells. This is because variations in upstream factors would affect both (identical) loci equally. However if noise were only intrinsic (i.e., every cell had the same rate constant and noise arose only from probabilistic initiation/degradation), then the CFP and YFP transcripts in the same cell would be no more correlated than the

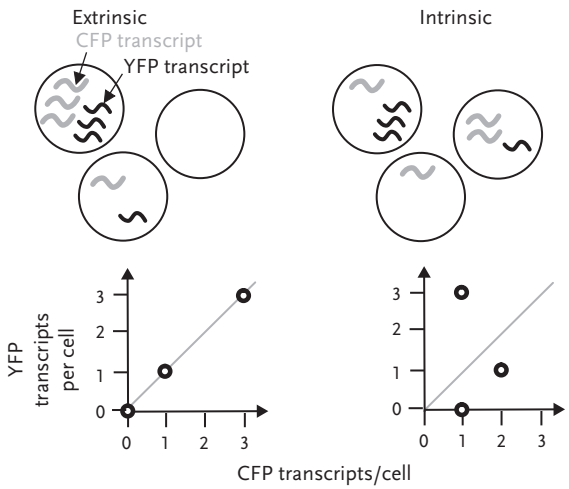

Figure 10.3 Distinguishing between extrinsic and intrinsic transcriptional noise using CFP and YFP expressed from identical promoters in a diploid cell. If noise is only extrinsic, the numbers of CFP and YFP transcripts in a cell should be perfectly correlated. In the cartoon, each cell has the same number of CFP and YFP transcripts, and in the plot all points lie along the diagonal. If noise is only intrinsic, the number of CFP transcripts in a cell varies independently of the numbers of YFP transcripts. In the cartoon, each cell can have the same or different numbers of YFP and CFP transcripts, and in the plot many points lie off the diagonal.

CFP transcripts in one cell and the YFP transcripts in another (i.e., completely uncorrelated).

The type and degree of noise intrinsic to transcription depends on the ways the process is regulated. In Figure 10.4, we compare two genes that express the same number of (rapidly degraded) transcripts per cell in a population of identical cells; in other words, all noise is intrinsic to transcription. Initiations on gene *a* occur independently, and a steady-state "Poisson" distribution of transcript numbers is seen across the cell population. Gene *b* is mostly inactive, but occasionally switches to produce a "burst" of transcripts (Figure 10.1c). In this case, we find three cell populations: one with no transcripts, one with close to the maximum, and a third that lies between the two (once the burst ceases, transcripts are degraded). Although both genes produce the same average number of transcripts, the populations differ. Quantitative models have been developed that use more subtle features to choose which model best fits a distribution (only extreme cases are presented here) [25, 33].

10.3.2 Studies of Noise Suggest Transcription Occurs in "Bursts"

RNA FISH is the most commonly used way of counting transcripts in a cell population; the number of transcripts per spot can be determined by comparing fluorescence intensities with the known intensity of a probe. Although it is not

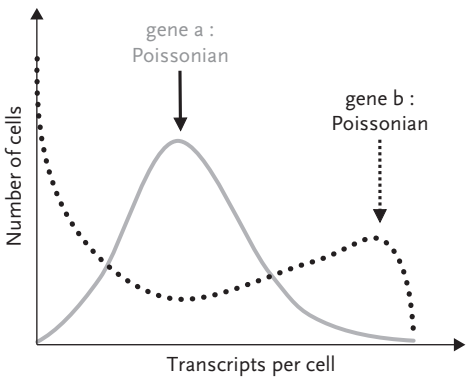

Figure 10.4 Different transcriptional modes yield different patterns of transcripts numbers in a population. See text for discussion.

immediately obvious that this method should allow the detection of most transcripts, or provide single-molecule sensitivity, extensive controls show that it does. RNA FISH gives the same average number of transcripts per cell as qRT-PCR [34], and – most impressively – probes targeting the 5' end of a nascent RNA can give signals that are once, twice, or three times higher than those targeting the 3' end (which in this case must reflect the distribution of 1, 2, or 3 polymerases engaged at different positions along the gene) [35].

RNA FISH provides evidence that mammalian transcripts are not produced independently of each other (i.e., that the process is non-Poissonian). In a seminal study, Raj and colleagues visualized transcripts copied from a genomically integrated transgene, and found the number per (CHO) cell varied dramatically [33]. Some 76% of cells contained no transcription site visible in the nucleus, with a mean of 74 transcripts/cell. The rest contained bright nuclear transcription sites (where many transcripts were concentrated) and a mean of 244 transcripts/cell (Figure 10.5). To show that this noise was intrinsic to transcription, they integrated two distinguishable (but otherwise identical) transgenes at different chromosomal positions. The expression patterns of the two genes were similar across the population, but poorly correlated with each other. This observation suggests that the observed noise did not result from variable upstream factors. As this distribution was not Poissonian (and so could not have arisen from a series of independent initiations), they proposed a model incorporating a transition from an inactive "OFF" state to a bursting "ON" state. This model fits the data well, yielding an average of $\sim$400 transcripts/burst. Moreover, reducing the initiation rate (by adding doxycycline to reduce binding of the tTA activator to the promoter) progressively decreased the number of transcripts per cell, whilst the transcriptional noise (defined as standard deviation/mean) was unaffected (if transcription were strictly Poissonian, noise would increase). Clearly, tTA binding affected transcription, but another regulator produced the noise. Modeling suggested that the reduction in binding of the transcription factor reduced the transcription rate

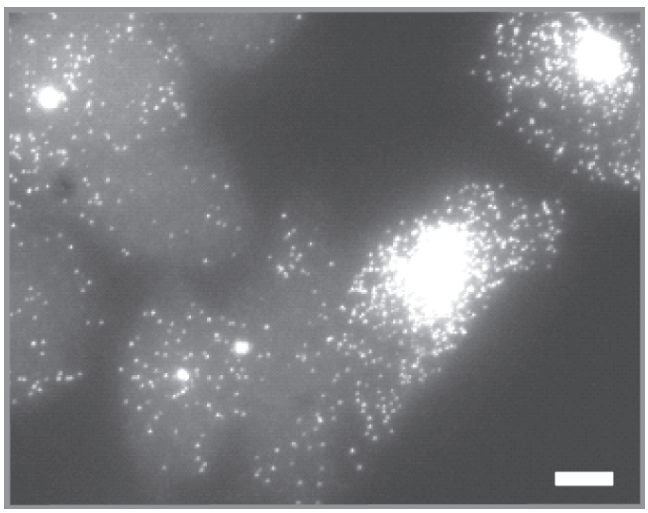

Figure 10.5 Heterogeneity in transcript copy number in CHO cells expressing YFP. YFP transcripts were detected by RNA FISH. Some cells express many transcripts, others only a few. Bar: 5 μm. Adapted from [33].

in the ON state, leaving transition between ON and OFF states unaffected. They now repeated the experiment, but this time integrated the two distinguishable transgenes into the same locus. Although both were noisy as before, this noise was now highly correlated; cells tended either to have high levels of both transcripts, or low levels of both transcripts. Thus, the source of the noise – the ON/OFF transition – acted simultaneously on adjacent genomic sites. It is unlikely that these dynamics are specific to the particular transgene used, as transcripts encoding the largest subunit of RNAP II show similar heterogeneity [33].

Other studies of cellular transcript copy numbers have uncovered non-Poissonian distributions, but it is unclear whether this noise is intrinsic to transcription. Bengtsson *et al.* used single-cell qRT-PCR to show that mRNA distributions in cells isolated from mouse pancreatic islets had log-normal (i.e., non-Poissonian) distributions [36]. However, expression patterns of two genes located on different chromosomes, *Ins1* and *Ins2*, were highly correlated, indicating that the noise was mostly extrinsic to transcription. This important issue is often ignored; for example, a heterogeneous distribution of transcript number in a *Drosophila* embryo was incorrectly used as evidence for bursting, without a demonstration that the noise was intrinsic [37]. Although it was argued (in the absence of mathematical modeling) that the poor correlation between nascent and matured transcripts meant that noise was intrinsic, this is insufficient; extrinsic factors could have varied over time (especially in developing cells), and the heterogeneity could easily have arisen from variability in rates of mRNA degradation.

Some yeast genes are also regulated by non-Poissonian mechanisms that seem to be similar to those found in mammals. For example, Raser and O'Shea isolated

noise intrinsic to transcription by engineering diploid yeast expressing CFP from one locus and YFP from the same locus on the homologous chromosome [38]. As they lacked knowledge of absolute transcript numbers, they modulated expression levels (using inducers) in order to query the source of the noise. For a Poissonian process, intrinsic noise strength (variance divided by mean) should be independent of expression level, as was the case for the *GAL1* and *PHO84* promoters; in other words, the relevant inducers probably regulated a Poissonian process (other types of processes can give such results) [39]. However, for the *PHO5* promoter, noise strength decreased strongly with expression level, suggesting an alternative mechanism was involved. The data could be explained if induction regulated an ON/OFF transition, without affecting the rate of transcription during the ON state.

Zenkleuksen *et al.* obtained similar results for some endogenous *S. cerevisiae* genes [25]. Using RNA FISH coupled to modeling to count and analyze nascent and completed transcripts, they argued that transcription in yeast was less intrinsically noisy than in mammals, and so more Poissonian. But there was one exception, *PDR5* – a gene selected because it was expected to be noisy. Transcript numbers for this gene were variable, and modeling showed the distribution to be consistent with bursting. However, this study, too, did not control for extrinsic variation, which other studies show may be large in yeast (it remains unclear whether such noisiness mainly results from position in the cell cycle [38, 40]).

In addition to the influence of extrinsic noise, there are other problems associated with using "snapshot" distributions of protein or transcript numbers to infer transcriptional dynamics. Even if a simple ON/OFF model is assumed, the data from transcript counting often fit a variety of scenarios [25]. Moreover, it is not clear that a simple ON/OFF model typically used does justice to biological complexity. Theoretical work shows that snapshots of steady-state distributions are insensitive to the shapes of distributions of both transcripts produced per burst and gestation times between bursts [39]. Moreover (as discussed below), there is good reason to expect that a binary ON/OFF switch does not describe the known behavior of many transcriptional regulators. Counting transcripts using RNA FISH does have the great advantage that it does not require genetic manipulation. Even so, the best way of studying transcription dynamics is to monitor transcription rates in single cells over time.

10.3.3
Visualizing Bursts of Transcription in Real Time

A powerful technique for localizing mRNAs in living cells combines expression of an MS2–GFP fusion protein and a message containing tens of hairpins encoding the sequence that binds MS2. One copy of the fusion protein binds to each hairpin soon after the message is made. Then, a fluorescent spot in the cell marks a nascent RNA molecule decorated with many fusion proteins, against a background of unbound (and rapidly diffusing) fusion proteins. When transcripts produced by the *lac/ara* promoter in *E. coli* are visualized, spots appear in bursts, one after another, followed by a silent period. These results do not fit a simple

model in which transcripts are randomly initiated according to a Poissonian process; rather, the gene seems to switch randomly from OFF to ON [41]. The *lac*/*ara* promoter stays ON for ~6 min (when it produces ~2 transcripts) before switching OFF for ~37 min (Figure 10.6). In those cases where it stayed ON for twice as long, twice as many transcripts tended to be made; this is consistent with transcripts being made consecutively during a burst (with one polymerase transcribing at any moment) rather than in parallel (with many engaged polymerases in a convoy). Similar results from the 1960s first described initiation at the tryptophan operon as periodic [42, 43], but the cause of such periodicity remains unknown.

Similar results were obtained using MS2-GFP tagging to study *dscA* mRNAs in the social amoeba, *Dictyostelium discoideum* [44]. As the population differentiates, *dscA* in some cells switches between ON/OFF states with similar lifetimes of 5–6 min, and global expression levels increase through an increase in the fraction of cells expressing the gene. The system was surprisingly rigid. There was no sign a gene that was ON could increase its activity either by increasing numbers of engaged polymerases (quantitative RNA FISH showed each gene was already maximally packed), or by decreasing the time between ON states. However, there was a transcriptional "memory"; a gene was more likely to be ON if it had been transcribed before, probably due to variability in upstream factors. Neighboring cells also tended to be ON together, perhaps because of they were synchronized through cAMP signaling.

Creative techniques using an indirect readout – detection of single fluorescent proteins in *E. coli* – have also been used to monitor the transcription that precedes protein production. Thus, Xie and colleagues used sensitive widefield microscopy to visualize, count, and then bleach single YFP-tagged membrane proteins as they

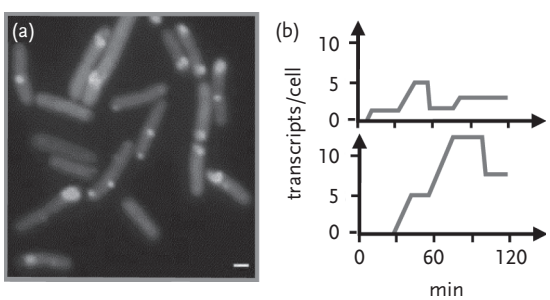

Figure 10.6 Visualizing mRNAs in living *E. coli* with MS2-GFP. An MS2-GFP fusion protein is used to tag transcripts encoding RFP linked to a tandem array of 96 MS2 binding sites. Adapted from [41] with permission. (a) In this black and white image, bright (GFP) foci mark one or more tagged mRNAs, against a diffuse background that fills the cell. The intensities of both foci and background vary from cell to cell. Bar: 1 μm. (b) Two time-courses indicating variations in the numbers of RFP transcripts in individual cells. Upper trace: a faint focus (marking one transcript) appears after ~10 s, and then becomes brighter (as an additional four transcripts appear) after ~40 s. Lower trace: five transcripts appear after ~30 s in the first burst, followed by an additional eight after 60 s.

were being produced under the control of a repressed *lac* promoter [45]. Proteins appeared in bursts, and modeling RNA copy numbers and degradation rates indicated each burst probably reflected translation of a single transcript. The distribution of the time between bursts fit an exponential decay, as expected of a Poissonian process. Later, the same group showed that short bursts probably resulted from partial dissociation of the tetrameric lac repressor, and longer ones from complete dissociation – which then led to a transcript being produced on the order of once per minute [46]. These techniques are only now being applied to higher eukaryotes, and we await results that should give us unprecedented insight into mechanisms of transcriptional regulation.

10.4
What Causes "Bursting"?

These studies on fixed and living cells make it possible to identify distinct modes of transcriptional regulation, and the authors sometimes link these modes to characterized regulatory mechanisms. However in most cases, the nature of the underlying mechanism remains unclear. We now review some known mechanisms and evaluate whether they might be involved. As these mechanisms act over different timescales, we first examine how long bursts last.

10.4.1
How Long Does a Burst Last?

In *D. discoideum* and *E. coli*, bursts last $\sim$5 min, but the number of transcripts produced per burst is either unknown or low. In mammals, modeling suggested that up to 400 transcripts are produced in one burst, and it is instructive to consider how long it would take this number of RNAPs to initiate on one gene. What is known about *in vivo* rates comes mostly from fluorescence recovery after photobleaching (FRAP) using arrays of induced genes that concentrate GFP-RNAP in one nuclear spot; after bleaching the spot, the residency time of the polymerases is inferred from the kinetics of fluorescence recovery. Several studies combine modeling and biological controls to conclude that polymerases exhibit different residency times at different stages in the transcription cycle. Darzacq *et al.* found that RNAPs bound to CMV promoters (in an array) for $\sim$6 s, initiated for $\sim$54 s, and then remained engaged for $\sim$10 min [16]. Initiation was inefficient; only $\sim$1% of RNAPs contacting a promoter proceeded to productive elongation. Such a rate would result in a successful initiation roughly once every 10 min, and a burst of 400 transcripts would take $\sim$3 days! However, Boireau *et al.* modeled an array of HIV-1 promoters, and found initiation to be too efficient to be observed as a distinct component [47]; another study on heat-shock puffs in *Drosophila* found initiation was also efficient [48]. Thus, initiation efficiencies may vary widely, especially as ChIP reveals wide variations in occupancy ratios between promoters and their genes [12]. However it seems unlikely that mammalian RNAP II initiates

faster than bacterial RNAPs on the ribosomal cistrons, which must initiate every 3.4 s to become tightly packed every ~85 bp (assuming an elongation rate of 25 nucleotides/s) [26, 41]. Moreover, we have seen that multiple polymerases are rarely seen together on spread DNA fibers (Figure 10.1b). If we consider an exceptional case with polymerases spaced 2 kbp apart, and assume an elongation rate at the maximum end of the range (i.e., 4 kbp/min) – the rate of initiation is still only once every 30 s, and a burst of 400 transcripts would take 3–4 h. This is consistent with the transgenes studied by Raj *et al.* being amongst the most active in the cell [33].

10.4.2
Typical Transcription Factors Do Not Bind Long Enough to Account for Bursting

The most obvious candidate for mediating ON/OFF transitions is the binding (or dissociation) of a transcriptional activator (or repressor). Although a transcription factor (TF) like the glucocorticoid receptor remains bound to its target site for hours *in vitro*, FRAP shows that it is bound only for seconds *in vivo* [49, 50]. A systematic study of nine other mammalian transcription factors also found that all exchanged fully within 2 min [51]. It could be that these TFs bind non-specifically to DNA, or to inactive promoters in an artificial array. However recent studies address this issue. *CUP1* in baker's yeast is present in a tandem array of 10 copies, and – in the presence of copper – Ace1 binds to the promoter to activate transcription of each gene in the array (demonstrated by MS2–GFP tagging of nascent RNA); GFP tagging showed that Ace1 bound to the array also exchanged fully within 2 min [52]. Clearly, most TFs dissociate within a minute or two, whereas bursts probably last much longer.

However, some TFs do have longer residency times. In heat-shocked *Drosophila*, the heat shock factor accumulates at specific puffs to remain stably bound with a half life of >6 min [53]. However, these stress-response genes are hyperactive, and this TF might be atypical. Another exception is the general TF, TBP, which is suspected of mediating ON/OFF transitions; it takes ~20 min to regain full intensity after photobleaching random spots in mammalian nuclei [54]. In an analogous experiment in yeast, all protein recovered in 15 s, but the authors noted their method would be unable to identify a small population that behaved as in mammals (e.g., a sub-population bound to promoters containing TATA boxes [55]). Moreover, global studies of noise in protein expression using a library of GFP-tagged yeast strains shows that the distribution of proteins expressed from promoters with TATA boxes is noisier than average – a finding consistent with TBP mediating a non-Poisson process [40]. Notably, transcription factors in primary tissues may exchange much more slowly. Thus, TFIIH in cell lines recovered (in a FRAP experiment) in less than 10 s [56], but in primary post-mitotic neurons ~90% remained bound for >1 h in a transcription-dependent manner [57]. Although it is possible that the ON/OFF transitions observed in *E. coli* were the result of TF binding/dissociation, little is known about the relevant kinetics. Xie and colleagues have shown that the lac repressor has long residency times on its

site *in vivo*, indicating that TFs could be responsible [58], but other mechanisms may underlie these bursts.

10.4.3 Effects of Chromatin Remodeling and Histone Modifications on Transcription Bursts

Chromatin remodeling proteins can reposition nucleosomes in an ATP-dependent manner to cover/uncover binding sites for the polymerase or its transcription factors [59], and non-Poissonian dynamics are often attributed to them. (More detail on remodeling proteins can be found in Chapter 5.) For example, knocking down remodelers increases noise strength (consistent with this process affecting ON/OFF rates [38]). Genome-wide measurements in *S. cerevisiae* also reveal that the expression of genes regulated by remodelers such as Swi/SNF and Isw2 is particularly noisy [40]. Other evidence that remodeling may act over the same long timescales as bursting comes from studies of periodic transcription. Transcription at the yeast *CUP1* locus oscillates with a $\sim$40-min period, and the oscillations are abolished by deleting *Rsc2* which encodes a chromatin remodeler [52]. However, it seems that nucleosome remodeling is a more dynamic and variable process with intermediary states that are not well described by a sharp ON/OFF model [52, 60]. Indeed, the only concrete finding of these studies is that nucleosome remodeling behaves in a non-Poissonian manner [39].

Changes in the covalent structure of nucleosomes (e.g., acetylation, phosphorylation, methylation) are described in detail in Chapter 4. The resulting effects on chromatin structure (e.g., to induce eu- or hetero-chromatin) may also underlie bursting. For example, genes with binding sites for enzymes that modify histones (such as SAGA) are noisier than average [40]. As these modifications and changed structures can spread down the fiber, they may be responsible for the ON/OFF switch identified by Raj *et al.*, as this switch acts simultaneously on adjacent loci [33].

10.4.4 Transcription Factories

The spatial restriction of transcription to "factories" is likely to have large effects on transcription dynamics. Below, we discuss the evidence for transcription factories and speculate on the relation between the spatial and temporal organization of transcript production. It is widely assumed that an RNAP becomes active by diffusing to a promoter, binding, and then tracking down the template as it makes its transcript. Accumulating evidence is consistent with an alternative: a promoter diffuses to a transcription "factory" where it binds to a transiently immobilized polymerase, which then reels in the template as the transcript is extruded [61]. We define a factory as a site containing at least two polymerases engaged on different templates. A typical nucleoplasmic factory in HeLa consists of a $\sim$90 nm core to which are tethered $\sim$16 loops – half through engaged polymerases and half

through transcription factors. The *raison d'être* of all factories is the same: to enhance production by concentrating relevant machines and raw materials in one place.

Support for this alternative view comes from various sources [61]:

1. After permeabilization, active polymerases resist detachment by nucleases, which places them at (or close to) points where loops are tied to the core.
2. High-resolution imaging shows nascent RNA is concentrated in a few sites: the factories (Figure 10.7a). As there are more active molecules of RNA polymerase

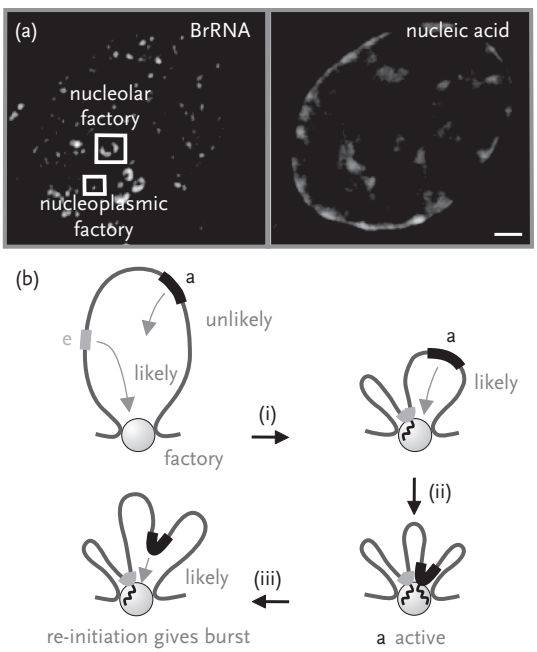

Figure 10.7 Transcription factories and bursting. (a) Factories in a HeLa cell. Cells were permeabilized, nascent RNA extended in Br-UTP, cells cryosectioned (100 nm), Br-RNA immunolabelled with FITC, nucleic acids counterstained with TOTO-3, and a fluorescence image collected on a confocal microscope; two views of one cell are shown. Newly-made Br-RNA is concentrated in factories in the cytoplasm (made by mitochondrial RNAPs), nucleoplasm (made by RNAPs II and III), and nucleoli (made by RNAP I). Stripping off and spreading one of the crescents in the nucleolar factory yields a "Christmas tree" like that seen in Figure 10.1a, while disrupting and spreading a typical nucleoplasmic factory yields ~8 structures like that seen in Figure 10.1b. Bar: 1 μm. Adapted from [62]. (b) A model for bursting. One loop is shown attached to a factory; transcription unit *e* is tethered closer to the factory than *a* – and so is more likely to initiate. (i) Transcription unit *e* now initiates by attaching to a polymerase in the factory. As a result *a* is brought closer to the factory, increasing the chances that it will initiate. In other words, *e* enhances the activity of *a* (and this is how we imagine many enhancers work). (ii) Gene *a* now initiates at a polymerase in the factory. (iii) Gene *a* now terminates and detaches from the polymerase/factory; once it has done this, it remains likely to reinitiate to create a burst (as it is still close to the factory).

II (in human, mouse, newt nuclei) than such factories, and as only one polymerase is typically engaged on a transcription unit, each factory must contain many different units.

3. Chromosome conformation capture (3C; for details, see Chapter 9) and FISH show sequences lying far apart on the genetic map can nevertheless lie close together in 3D space; significantly, contacting sequences are usually transcriptionally active [63, 64]. For example, after stimulating human cells with estrogen and mapping genome-wide contacts made by bound estrogen receptor, contacting partners are often associated with bound RNA polymerase II [65].

Active forms of the three nuclear polymerases are each concentrated in different dedicated factories. For example, it is well known that active polymerase I is nucleolar, and polymerases II and III are also concentrated in their own dedicated nucleoplasmic factories [66]. Results obtained using 3C and FISH suggest that polymerase II factories specialize even further to transcribe different gene subsets. For example, transcription units encoding factors involved in the globin pathway (e.g., *Hbb-b1*, its LCR, *Eraf*) on mouse chromosome 7 are often (but not always) together in factories when active [67–69], and the networks of genes attached to individual factories are now being examined [70, 71]. Moreover, two mini-chromosomes carrying essentially identical units are transcribed in the same factories, but inserting into one a different promoter now targets it to a different factory [72]. It is then easy to imagine that different factories contain factors required for the transcription of specific gene sub-sets.

While it is known that essentially all elongating RNAPs are found in factories, it is not known whether or not recruitment of promoters to factories is a rate-limiting step in initiation. Given that evolution seems to take every possible opportunity to regulate gene expression, it seems likely that it is (Figure 10.7b). A model where collision frequency with – and so spatial proximity to – a factory influences the activity of a gene predicts that dynamic reorganization of chromatin could have strong effects on gene activity. Modeling a chromatin loop attached to a factory shows that tethering a locus close to (or far from) a factory can activate (or repress) the locus [24]. Moreover, once a gene being transcribed terminates and dissociates, it would still be close to the factory, which contains the relevant transcription machinery; as a result, it would be likely to re-bind and re-initiate to give a burst. Such a model also predicts that tethering a transcription unit near a factory containing the "wrong" kind of transcription machinery would repress it. Consistent with this model, canonical "silencers" turn out to be transcription units [61], and tRNA genes (that sequester loci to RNAP III factories) are strong silencers of adjacent RNAP II genes [73]. Such attachment to factories could result in changes in activity that could last for minutes or hours. They could easily be responsible for some or all of the non-Poissonian transcriptional dynamics described above, producing "bursts" (e.g., as a gene re-initiates repeatedly in the same factory), or more gradual changes in activity (e.g., as the gene drifts away from, or towards, the relevant factory; Figure 10.7b).

We now know that the nucleus is highly compartmentalized ([74]; see also Chapters 12 and 17). For example, proximity to the nuclear lamina [75] or pore [76] (Chapter 8), or requirement for a remodeling complex like SAGA (which seems to be preferentially active at the periphery; Chapter 5) [69, 77, 78] can all have effects on transcription. In the absence of evidence concerning mechanism, and applying Ockham's razor, we suggest it is likely that all these effects work through the mechanisms described above (transcription factor binding, chromatin remodeling and modification, proximity to a factory).

10.5 Conclusion

A full understanding of how the cell regulates gene expression requires detailed and quantitative dissection of transcriptional dynamics in both cell populations and single cells. The pioneering studies we describe here have begun to provide such information, but the underlying mechanisms still remain to be uncovered. This undertaking will not be trivial, as we can already see that so many factors affecting activity are interconnected, including transcription factors, localization relative to factories, and global genomic structure. Still, a thorough exploitation of the methods described here should provide unprecedented understanding of how the cell controls transcription.

Acknowledgments

We thank Oscar Miller, Minoru Ko[illegible]n Raj, Ido Golding, and Ana Pombo for permission to reproduce their images. K.F. is supported by the E.P. Abraham Trust, a Clarendon award from the University of Oxford, and an Overseas Research Student award from the UK Government. Research in P.R.C.'s laboratory is supported by The Biotechnology and Biological Sciences Research Council, The Medical Research Council, and The Wellcome Trust.

References

1 Weiss, S.B. and Gladstone, L. (1959) A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J Am Chem Soc*, **81**, 4118–4119.
2 Chambon, P. (1975) Eukaryotic nuclear RNA polymerases. *Ann Rev Biochem*, **44**, 613–638.
3 Kimura, H., Tao, Y., Roeder, R.G., and Cook, P.R. (1999) Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Mol Cell Biol*, **19**, 5383–5392.
4 Koch, F., Jourquin, F., Ferrier, P., and Andrau, J.-C. (2008) Genome-wide RNA polymerase II: not genes only! *Trends Biochem Sci*, **33**, 265–273.
5 Jasiak, A.J., Armache, K.-J., Martens, B., Jansen, R.-P., and Cramer, P. (2006)

Structural biology of RNA polymerase III: subcomplex C17/25 x-ray structure and 11 subunit enzyme model. *Mol Cell*, **23**, 71–81.

6 Kuhn, C.-D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007) Functional architecture of RNA polymerase I. *Cell*, **131**, 1260–1272.

7 Cramer, P. (2004) RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev*, **14**, 218–226.

8 Miller, K.G. and Sollner-Webb, B. (1981) Transcription of mouse rRNA genes by RNA polymerase I: *in vitro* and *in vivo* initiation and processing sites. *Cell*, **27**, 165–174.

9 Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D., and Misteli, T. (2002) A kinetic framework for a mammalian RNA polymerase *in vivo*. *Science*, **298**, 1623–1626.

10 Oler, A.J., Alla, R.K., Roberts, D.N., Wong, A., Hollenhorst, P.C., Chandler, K.J., Cassiday, P.A., Nelson, C.A., Hagedorn, C.H., Graves, B.J., and Cairns, B.R. (2010) Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol*, **17**, 620–628.

11 Canella, D., Praz, V., Reina, J.H., Cousin, P., and Hernandez, N. (2010) Defining the RNA polymerase III transcriptome: genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Res*, **20**, 710–721.

12 Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol*, **27**, 66–75.

13 Consortium, T.F., Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.

14 Sawadogo, M. and Sentenac, A. (1990) RNA polymerase B (II) and general transcription factors. *Ann Rev Biochem*, **59**, 711–754.

15 Lee, T.I. and Young, R.A. (2000) Transcription of eukaryotic protein-coding genes. *Ann Rev Genet*, **34**, 77–137.

16 Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007) *In vivo* dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol*, **14**, 796–806.

17 Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A., Izumi, A., *et al.* (2009) A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci*, **106**, 18357–18361.

18 Jackson, D.A., Iborra, F.J., Manders, E. M.M., and Cook, P.R. (1998) Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell*, **9**, 1523–1536.

19 Bottomley, R.H., Trainer, A.L., and Griffin, M.J. (1969) Enzymatic and chromosomal characterization of HeLa variants. *J Cell Biol*, **41**, 806–815.

20 Miller, O.L. Jr and Bakken, A.H. (1972) Morphological studies of transcription. *Eur J Endocrinol*, **71**, S155–S177.

21 Carter, M., Sharov, A., VanBuren, V., Dudekula, D., Carmack, C., Nelson, C., and Ko, M. (2005) Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol*, **6**, R61.

22 Holland, M.J. (2002) Transcript abundance in yeast varies over six orders of magnitude. *J Biol Chem*, **277**, 14363–14366.

23 Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, **95**, 717–728.

24 Bon, M., McGowan, S.J., and Cook, P.R. (2006) Many expressed genes in bacteria

and yeast are transcribed only once per cell cycle. *FASEB J*, **20**, 1721–1723.
25 Zenklusen, D., Larson, D.R., and Singer, R.H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*, **15**, 1263–1271.
26 French, S.L. and Miller, O.L. Jr (1989) Transcription mapping of the Escherichia coli chromosome by electron microscopy. *J Bacteriol*, **171**, 4207–4216.
27 Miller, O.L. Jr, Hamkalo, B.A., and Thomas, C.A. Jr (1970) Visualization of bacterial genes in action. *Science*, **169**, 392–395.
28 Mooney, R.A., Davis, S.E., Peters, J.M., Rowland, J.L., Ansari, A.Z., and Landick, R. (2009) Regulator trafficking on bacterial transcription units *in vivo*. *Mol Cell*, **33**, 97–108.
29 Grigorova, I.L., Phleger, N.J., Mutalik, V. K., and Gross, C.A. (2006) Insights into transcriptional regulation and σ competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci*, **103**, 5332–5337.
30 Bremer, H. and Dennis, P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd edn (ed. F. Neidhardt), American Society for Microbiology, pp. 1553–1569.
31 Paulsson, J. (2004) Summing up the noise in gene networks. *Nature*, **427**, 415–418.
32 Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
33 Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*, **4**, e309.
34 Vargas, D.Y., Raj, A., Marras, S.A.E., Kramer, F.R., and Tyagi, S. (2005) Mechanism of mRNA transport in the nucleus. *Proc Natl Acad Sci USA*, **102**, 17008–17013.
35 Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998) Visualization of single RNA transcripts *in situ*. *Science*, **280**, 585–590.
36 Bengtsson, M., Ståhlberg, A., Rorsman, P., and Kubista, M. (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res*, **15**, 1388–1392.
37 Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009) Visualization of individual scr mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Curr Biol*, **19**, 2037–2042.
38 Raser, J.M. and O'Shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science*, **304**, 1811–1814.
39 Pedraza, J.M. and Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, **319**, 339–343.
40 Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, **441**, 840–846.
41 Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123**, 1025–1036.
42 Baker, R.F. and Yanofsky, C. (1968) Periodicity of RNA polymerase initiations – a new regulatory feature of transcription. *Proc Natl Acad Sci USA*, **60**, 313.
43 Imamoto, F. (1968) On the initiation of transcription of the tryptophan operon in *Escherichia coli*. *Proc Natl Acad Sci*, **60**, 305–312.
44 Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006) Transcriptional pulsing of a developmental gene. *Curr Biol*, **16**, 1018–1025.
45 Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006) Probing gene expression in live cells, one protein molecule at a time. *Science*, **311**, 1600–1603.
46 Choi, P.J., Cai, L., Frieda, K., and Xie, X. S. (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.
47 Boireau, S., Maiuri, P., Basyuk, E., Mata, M., Knezevich, A., Pradet-Balade, B.,

Bäcker, V., Kornblihtt, A., Marcello, A., and Bertrand, E. (2007) The transcriptional cycle of HIV-1 in real-time and live cells. *J Cell Biol*, **179**, 291–304.

48 Yao, J., Ardehali, M.B., Fecko, C.J., Webb, W.W., and Lis, J.T. (2007) Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol Cell*, **28**, 978–990.

49 Erlmann, T., Eriksson, P., and Wrange, O. (1990) Quantitative-analysis of the glucocorticoid receptor–DNA interaction at the mouse mammary-tumor virus glucocorticoid response element. *J Biol Chem*, **265**, 17222–17229.

50 McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*, **287**, 1262–1265.

51 Phair, R.D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M., and Misteli, T. (2004) Global nature of dynamic protein–chromatin interactions *in vivo*: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol Cell Biol*, **24**, 6393–6402.

52 Karpova, T.S., Kim, M.J., Spriet, C., Nalley, K., Stasevich, T.J., Kherrouche, Z., Heliot, L., and McNally, J.G. (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science*, **319**, 466–469.

53 Yao, J., Munson, K.M., Webb, W.W., and Lis, J.T. (2006) Dynamics of heat shock factor association with native gene loci in living cells. *Nature*, **442**, 1050–1053.

54 Chen, D., Hinkley, C.S., Henry, R.W., and Huang, S. (2002) TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes. *Mol Biol Cell*, **13**, 276–284.

55 Sprouse, R.O., Karpova, T.S., Mueller, F., Dasgupta, A., McNally, J.G., and Auble, D.T. (2008) Regulation of TATA-binding protein dynamics in living yeast cells. *Proc Natl Acad Sci*, **105**, 13304–13308.

56 Hoogstraten, D., Nigg, A.L., Heath, H., Mullenders, L.H.F., van Driel, R., Hoeijmakers, J.H.J., Vermeulen, W., and Houtsmuller, A.B. (2002) Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair *in vivo*. *Mol Cell*, **10**, 1163–1174.

57 Giglia-Mari, G., Theil, A.F., Mari, P.-O., Mourgues, S., Nonnekens, J., Andrieux, L.O., de Wit, J., Miquel, C., Wijgers, N., Maas, A., Fousteri, M., *et al.* (2009) Differentiation driven changes in the dynamic organization of basal transcription initiation. *PLoS Biology*, **7**, e1000220.

58 Elf, J., Li, G.-W., and Xie, X.S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.

59 Cairns, B.R. (2009) The logic of chromatin architecture and remodelling at promoters. *Nature*, **461**, 193–198.

60 Hager, G.L., McNally, J.G., and Misteli, T. (2009) Transcription dynamics. *Mol Cell*, **35**, 741–753.

61 Cook, P.R. (2010) A model for all genomes: the role of transcription factories. *J Mol Biol*, **395**, 1–10.

62 Cook, P.R. (1999) The organization of replication and transcription. *Science*, **284**, 1790–1795.

63 Osborne, C.S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, **36**, 1065–1071.

64 Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, **38**, 1348.

65 Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., Chew, E.G.Y., *et al.* (2009) An oestrogen-receptor-α-bound human chromatin interactome. *Nature*, **462**, 58–64.

66 Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G., and Cook, P.R. (1999) Regional specialization in

human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J*, **18**, 2241–2253.

67 Gondor, A. and Ohlsson, R. (2009) Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? *Nat Rev in Genetics*, **10**, 269–276.

68 Simonis, M. and de Laat, W. (2008) FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochim Biophys Acta*, **1783**, 2052–2060.

69 Sexton, T., Schober, H., Fraser, P., and Gasser, S.M. (2007) Gene regulation through nuclear organization. *Nat Struct Mol Biol*, **14**, 1049–1055.

70 Sandhu, K.S., Shi, C., Sjölinder, M., Zhao, Z., Göndör, A., Liu, L., Tiwari, V. K., Guibert, S., Emilsson, L., Imreh, M. P., and Ohlsson, R. (2009) Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev*, **23**, 2598–2603.

71 Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., Eskiw, C.H., *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*, **42**, 53–61.

72 Xu, M. and Cook, P.R. (2008) Similar active genes cluster in specialized transcription factories. *J Cell Biol*, **181**, 615–623.

73 Hull, M.W., Erickson, J., Johnston, M., and Engelke, D.R. (1994) tRNA genes as transcriptional repressor elements. *Mol Cell Biol*, **14**, 1266–1277.

74 Fraser, P. and Bickmore, W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413–417.

75 Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*, **394**, 592–595.

76 Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., Wilm, M., *et al.* (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell*, **21**, 811–823.

77 Cabal, G.G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., Feuerbach-Fournier, F., Olivo-Marin, J.-C., Hurt, E.C., and Nehrbass, U. (2006) SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, **441**, 770–773.

78 Luthra, R., Kerr, S.C., Harreman, M.T., Apponi, L.H., Fasken, M.B., Ramineni, S., Chaurasia, S., Valentini, S.R., and Corbett, A.H. (2007) Actively transcribed gal genes can be physically linked to the nuclear pore by the SAGA chromatin modifying complex. *J Biol Chem*, **282**, 3042–3049.

Keywords: green fluorescent protein, RNA FISH, RNA polymerase, single-molecule analysis, transcription