# Trends in Genetics

**What is a transcriptional burst?**

---Manuscript Draft---

<table>
<thead>
<tr>
<th><strong>Manuscript Number:</strong></th>
<th>TIGS-D-19-00210R1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Article Type:</strong></td>
<td>Review</td>
</tr>
<tr>
<td><strong>Keywords:</strong></td>
<td>transcriptional bursting; single cell gene expression; single molecule imaging; stochastic gene expression; MS2; smFISH</td>
</tr>
<tr>
<td><strong>Corresponding Author:</strong></td>
<td>Jonathan Chubb</td>
</tr>
<tr>
<td></td>
<td>London, UNITED KINGDOM</td>
</tr>
<tr>
<td><strong>First Author:</strong></td>
<td>Ed Tunncliffe</td>
</tr>
<tr>
<td><strong>Order of Authors:</strong></td>
<td>Ed Tunncliffe</td>
</tr>
<tr>
<td></td>
<td>Jonathan Chubb</td>
</tr>
<tr>
<td><strong>Abstract:</strong></td>
<td>The idea that gene activity can be discontinuous will not surprise many biologists – many genes are restricted in when and where they can be expressed. Yet during the past 15 years, a collection of observations compiled under the umbrella term ‘transcriptional bursting’ has received considerable interest. Direct visualization of the dynamics of discontinuous transcription has expanded our understanding of basic transcriptional mechanisms and their regulation, and provides a real-time readout of gene activity during the life of a cell. In this review, we try to reconcile the different views of the transcriptional process emerging from studies of bursting, and how this work contextualizes the relative importance of different regulatory inputs to normal dynamic ranges of gene activity.</td>
</tr>
</tbody>
</table>
Highlights

- Demystification of the term “transcriptional bursting”
- Models with one or two gene states are unable to accurately describe dynamic transcription for many genes
- Many alternative multi-state models have been proposed but these are likely to be highly context-specific
- Understanding the contributions of numerous different cellular features and processes to bursting is required to build more accurate and general models of transcription dynamics
- Emerging imaging technologies are beginning to facilitate the monitoring of these diverse sources of regulation
What is a transcriptional burst?

Ed Tunnacliffe¹ and Jonathan R Chubb²

¹MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DS. ²MRC Laboratory for Molecular Cell Biology, University College London, Gower Street, London, WC1E 6BT.

Correspondence to edward.tunnacliffe@ndcls.ox.ac.uk

Keywords: transcriptional bursting, single cell gene expression, stochastic gene expression, MS2, smFISH
The idea that gene activity can be discontinuous will not surprise many biologists – many genes are restricted in when and where they can be expressed. Yet during the past 15 years, a collection of observations compiled under the umbrella term ‘transcriptional bursting’ has received considerable interest. Direct visualization of the dynamics of discontinuous transcription has expanded our understanding of basic transcriptional mechanisms and their regulation, and provides a real-time readout of gene activity during the life of a cell. In this review, we try to reconcile the different views of the transcriptional process emerging from studies of bursting, and how this work contextualizes the relative importance of different regulatory inputs to normal dynamic ranges of gene activity.
An Introduction

If we accept that genes can be ‘on’ or ‘off’ and that cells are able to change their gene expression, then it requires no major leap of faith to accept the possibility that transcription can be discontinuous over time. Indeed, direct visual evidence of discontinuous transcription emerged as early as the 1970s. When viewed under an electron microscope, Miller chromatin spreads from the fruit fly embryo showed unequal distribution of nascent transcripts along gene sequences (Figure 1A) [1]. The gaps between groups of multiple transcripts were interpreted as interruptions in transcription initiation events.

Attempts to directly visualize transcription were far from the mainstream for the next 25 years, with the emphasis instead on defining transcriptional regulatory components and their interactions. These reductionist strategies were essential for determining the molecular players involved, but lacked certain features necessary for building a more complete view of the transcriptional process. Firstly, the measurements were static, merging transcription and RNA degradation into a single RNA quantity. Secondly, samples were ensemble, usually the average of millions of cells, blurring the dynamics of the activity of individual genes. Finally, the biochemical and genetic strategies used to define regulatory components, by their very nature, detach gene regulation from normal cell physiology, making it difficult to arrive at meaningful models of the transcriptional process.

Solving these issues of reductionism required the ability to see transcription of single genes, in single cells, in an appropriate physiological context. This needed improvements in fluorescence microscopy, to approach the speed and sensitivity required for single molecule imaging in living cells, combined with the development of appropriate RNA labeling strategies for transcript detection (Box 1). More specifically, the application of single molecule fluorescence in situ hybridization (smFISH) [2] on fixed cells and MS2 stem-loop-based detection (Figure 1B) [3] in living cells, corroborated the temporal transcriptional discontinuity inferred from Miller spreads [4-6]. These approaches also highlighted the dynamics of transcriptional events. The ‘bursts’ or ‘pulses’ of transcriptional activity were found to operate over timescales of a few minutes (Figure 1C), and were measurably responsive to features such as developmental time and local environment. The study of this phenomenon has since expanded to many different organisms, both prokaryotic and
eukaryotic, with an increasing number of mechanistic models used to explain the
different dynamic behaviors observed across a variety of genes. Here we collate and
review these models, focusing on recent studies where features of transcriptional
mechanism have been explored through the study of bursts, and how these features
can be linked to specific molecular regulatory events.

What is a Burst?
The term ‘transcriptional bursting’ has been employed to explain a range of
potentially different phenomena. There is nothing implicit in the word ‘burst’ that
implies a specific model, mechanism or dynamic behavior, beyond being
discontinuous over time. Although the term is vague, the descriptions of bursting
have often been highly quantitative and integrated with simple models of gene
activity. Typically, a model applied to a bursting phenotype will focus on the number
of ‘states’ or levels of activity at which a gene can be transcribed (Figure 2, Key
Figure). Although a simple ‘one-state’ model based on a fixed initiation rate can give
rise to fluctuations in transcriptional activity [7, 8] and such a model can fit well to
distributions of smFISH RNA counts for a few genes [9], the complexity offered by
these models is not sufficient to explain the dynamic behavior of most genes that
have been studied. A two-state or random telegraph model [10] has been more
widely adopted. Here, a gene is only transcribed when in an ‘active’ configuration.
Fluctuations between this ‘active’ state and an ‘inactive’ state, result in short spurts
of mRNA production interspersed with periods of no activity [6]. This model is now
widely used to explain how pulsatile mRNA synthesis is controlled, particularly when
inferring dynamics from fixed-cell smFISH transcript distributions. It has also been
used in genome-wide studies to show how transcriptional dynamics can explain
developmental gene expression heterogeneity [11] and to understand broad
mechanisms of sequence-encoded regulation [12].

As a recent example, using a two-state model [13] uncovered a common
regulatory mechanism governing transcription of gap genes in *Drosophila*.
Comparisons between mRNA count distributions showed almost identical statistical
relationships for all four genes studied. Modulation of promoter occupancy alone was
found to be sufficient to explain the common regulation, with tight coupling of ON and
OFF switching rates resulting in the emergence of a unified pattern of transcriptional
control across the gene set [13]. A similar coupling of switching rates was found in
live imaging experiments using *even-skipped*, which is regulated by the gap gene transcription factors (TFs) [14]. However, experiments on developmentally matched gene sets in *Dictyostelium* do not exhibit such statistical similarities in bursting activity [15] suggesting such unified control suits rigidly instructive forms of development, such as in the *Drosophila* embryo, rather than more responsive developmental systems.

Despite its widespread use, the assumptions of the standard two-state model – constant rates for initiation, degradation, and switching between active and inactive states – are unrealistic in many biological systems. Transcription changes in response to a multitude of signals, yet the model does not easily account for this. These assumptions rather marginalise bursting as a side issue of transcription, failing to accommodate extrinsic sources of variation (such as signaling to transcription), with bursting consigned to only those processes designated intrinsic (molecular noise). Ideally a model should be an informed attempt to explain the biology, rather than a device that inadvertently excludes much that is interesting. This case of the model ‘owning’ the bursting phenomenon is widespread, but rather unusual if one considers bursting as the dynamic manifestation of the complete transcriptional process. Beyond these issues, it is now clear the two-state model cannot accurately describe transcription kinetics for all genes, in all systems. The use of fixed-cell approaches can be limiting when exploring alternative models of regulation; theoretical work has shown how dynamic measurements, rather than transcript counting by smFISH, must be made in order to distinguish between certain promoter state conformations (such as two-state and some three-state models) [16, 17]. In keeping with this, a gene found by live imaging to show a spectrum of activity states would be well-described by a two-state model if assayed by smFISH [7].

**More Complexity, Less Consensus**

If a two-state model is largely unsatisfactory as a description of transcription, what alternatives are there? Models containing multiple promoter activity states have been employed theoretically to account for experimental data in numerous cell types [18-21], with imaging studies demonstrating more explicitly that an expansion of the two-state model architecture could be appropriate [22, 23]. In yeast, a four-state model with a single inactive state was identified as the best fit to smFISH data for a small number of stress-response genes [24]. Multiple timescales of transcriptional bursting
were also inferred from measurements of HIV-1 promoter activity using the MS2 system in mammalian cells [25]. Here, TATA-binding protein (TBP) and mediator were found to independently regulate gene activity on these alternate timescales, and a three-state model of transcription was proposed, with inactive and active states as well as an intermediate ‘permissive’ state.

Multi-state models containing a ‘refractory’ inactive state through which a gene promoter must pass before reactivation can occur have also emerged, from studies using destabilized reporter proteins (Box 1; Figure 2) [26-28]. Endogenous promoters were typically found to pass through 5-7 sequential inactive steps before reactivation, while synthetic or TATA-containing promoters had fewer inactive steps which resulted in noisier gene expression [29]. Bartman et al. [30] combined PolII-chromatin immunoprecipitation (ChIP) with smFISH and also identified refractory period-based models as consistent with their data, although their preferred model involved burst initiation and polymerase pause release as limiting steps in gene activation. Refractoriness is often associated with a ‘reset’ of molecular components in preparation for receiving a new stimulus. A less intuitive role for refractoriness in transcription may be to enable rapid and sensitive responses to stimuli [31]. Models of promoter progression, in which events at the promoter form an ordered sequence of recruitment of different parts of the transcriptional machinery, may be consistent with refractory behaviour [32]. An alternative view questions whether the refractory period is a transcriptional phenomenon, or merely an adaptation response in the upstream signaling, such as phosphatase activity or receptor down-regulation [33]. While refractoriness has now been described across several systems and genes, in terms of information transmission, this type of system may be less favourable than a simple two-state model of gene expression [16].

Rodriguez et al. [34] also found inefficient information transfer in multi-state transcription while studying TFF1 regulation in MCF7 cells. Here, a model containing three ‘gene states’ (two inactive, one permissive) and two ‘RNA steps’ (activity levels in the ON state) was the best fit to the data from an MS2 reporter cell line, with a highly inactive state occupied for extremely long periods of time. By measuring changes in chromatin contacts in response to an estradiol (E2) stimulus the authors showed that while cells can effectively sense multiple levels of E2 dose, the information transfer to transcriptional output is inefficient and slow. While it is unclear why such regulatory schemes have evolved in this way, it could represent a similar
process of robustness through sub-optimisation of the network, a concept also
applied to sub-optimal binding of TFs to developmental enhancers [35] and core
promoter sites [36]. Alternatively, these observations may imply that a coherent
transcriptional response is only likely in the presence of the full complement of
signals available in a normal tissue niche, with measurement of these additional
signals likely to provide more explanatory power [37].

Although the inclusion of additional activity states can improve the fit between
a model and experimental data, how far should one go with this? Is there an upper
limit to the descriptive benefits of increased model complexity? A continuum or
spectrum of activity states, rather than a discrete number, can provide the best fit to
dynamic expression data from genes in diverse systems (Figure 2) [7, 38].

Intuitively, this makes sense given the myriad of molecular inputs influencing gene
transcription. Whether a continuum actually represents many discrete activity states
which simply cannot be resolved is unclear and such distinctions may remain
elusive. In their paper describing a general multi-state mathematical framework for
transcriptional bursting, [39] show that it may be difficult to determine the precise
number of activity states, particularly if the time spent in each is very short. If the
number of regulatory inputs (and therefore perhaps the number of activity states) of
transcription is high, and the relative time spent in individual regulatory
conformations is low, it will be difficult to distinguish these states accurately. Along
these lines, a fast switching model emerged as the most appropriate scenario to
explain transcript output from the lysogeny maintenance promoter of lambda phage
[40].

Finally, it is not the case that simply adding more activity states to a
computational model provides a better fit to experimental data. Fritzsch *et al.* [41]
explored the E2-regulated *GREB1* gene in MCF7 cells and found that despite
sampling several multi-state models (with up to 10 discrete levels), a two-state model
gave the best fit to their data. Therefore, while use of a two-state model to describe
transcriptional bursting of a gene should not be the default position, equally, a multi-
state architecture of some form is not guaranteed to be more descriptive.

With so many different models describing gene regulation, is it possible to
derive general principles of transcriptional bursting? Which, if any, of the
conformations described above could be relevant more generally to describe
transcription? Should we even expect consensus, especially considering the
diversity in the genes and experimental systems, and the different methods that have been employed? Diversity in bursting is clear even in more closely-related contexts. Comparing separate detailed studies of bursting in oestrogen-inducible genes, where similar regulation might be expected, highlights different regulatory regimes. As previously mentioned, Rodriguez et al. [34] proposed a model containing five activity levels, including a deep repressive state defined by long periods of inactivity for TFF1, even at saturating E2 concentrations in MCF7 cells. On the other hand, in the same cells with similar saturating induction conditions, GREB1 showed near-constant activity in most cells and a simple two-state model was preferable to those with multiple states and circular architecture [41]. Despite their different cellular functions, these genes previously showed similarly strong induction by E2 stimulation in multiple cell types [42, 43]. While this comparison is somewhat limited in scope, it shows that even genes with superficially similar regulation can be subject to very different dynamic control. Therefore, the regulation of bursting may well be highly gene-specific and will depend, potentially to differing extents, on the multiple different inputs to gene regulation. Despite the apparent convergence of regulatory mechanisms in certain specific contexts [13], any substantial coherence between models of transcriptional bursting will require a more detailed understanding of the relative contributions of the processes affecting bursts.

Making Bursts

Cis-Regulation

As the scaffold for RNA polymerase loading onto a gene, the promoter represents an important integration zone for transcriptional control [44]. Sequence diversity permits enormous heterogeneity in transcriptional output [45] and individual promoter cis-regulatory elements have been shown to influence transcriptional bursting at the single-cell level [7, 46]. Even within a family of duplicated actin genes encoding exactly the same protein, considerable diversity in bursting patterns was identified [15]. The role of the upstream sequence was directly evaluated by a reciprocal switching experiment exchanging around 500bp of the proximal 5’ regulatory regions of genes with different bursting patterns. This treatment revealed bursting dynamics to be almost entirely instructed by the upstream regulatory sequence with only a minor role for features specific to genomic context.
At least superficially, this result goes against the grain of some earlier ideas on the origins of bursts, which suggested switching between ON and OFF states reflects chromatin remodeling [47]. Clearly, chromatin regulation is an important part of transcriptional control, and several studies have shown that disruption of the normal chromatin landscape can affect bursting [29, 48-51]. Recent live imaging studies directly showed an increase in H3K27ac levels immediately prior to the appearance of active forms of RNA pol II at transcriptionally active nuclear compartments in early zebrafish development [52, 53]. This is consistent with a prominent role for the chromatin environment in influencing transcriptional decisions, although it is not clear if the sensitivity of detecting the different chromatin and polymerase modifications is equivalent. Similarly, histone acetylation was also found to regulate burst frequency-mediated changes in circadian clock gene expression [54]. A role for chromatin modification and remodeling is evident – chromatin is the substrate, it is close to the action- it is almost expected that experimentally perturbing chromatin will affect transcription. But to what extent do chromatin changes drive bursting dynamics? Given the direct demonstration that actin gene bursts can be dominated by the promoter region [15] as well as other data showing similar bursting patterns at multiple genomic loci [29, 55], our current view is that although chromatin is crucial for the functional integrity of the bursting process, it does not instruct the dynamic behaviour.

Bursting is influenced by distal enhancers as well as proximal promoters, with these elements directly involved in regulating transcriptional bursts, predominantly by modulating the frequency of these events [31, 34, 41, 56-58]. Genome wide inferences from single cell RNAseq data suggest regulation of burst frequency is the most widespread method of modulating transcription during developmental progression [11], with enhancers likely to be a major control point for this regulation [12]. However, enhancer regulation by modulating burst frequency is not universal, with burst size regulation predominating in response to Notch signalling [59, 60]. Further complexity arises when considering the combined effects of multiple enhancers at different times and places during embryogenesis [61, 62].

The importance of enhancer-promoter proximity for bursts has recently been evaluated using dual labelling of both DNA and nascent RNA in live cells. Dynamic transcription was found to be both correlated [63] and uncorrelated [64] with enhancer-promoter proximity, suggesting a number of models are required to explain
enhancer activity. Indeed, such a dichotomy exists even at a single locus, as different tissue-specific enhancers of the Shh gene showed both increased [65] and decreased [66] enhancer-promoter proximity concomitant with gene activation. The rules governing enhancer-mediation of dynamic transcription in tissues are seemingly complex, and will likely depend on the specific transcription and structural factors bound there at any particular time, in addition to higher order features of the nuclear microenvironment [67, 68]. Current excitement for potential roles of liquid-liquid phase separation (LLPS) in forming compartments that enhance the efficiency of transcriptional reactions has been discussed elsewhere [69], although at the time of writing, there is a lack of convincing experimental evidence that compartments formed by LLPS bring any functional benefits [70].

Transcription Factors

The binding of transcription factors (TFs) to target motifs at both promoter and enhancer elements is key to activation of a gene yet, until recently, it has not been clear how TF binding events are dynamically related to transcriptional activity. Residence times of TF binding at target sites are typically on the order of seconds [71, 72], which contrasts the timescales of minutes usually associated with bursts. While single-molecule tracking (SMT) methods have enabled the study of individual TF molecule binding dynamics, it has remained challenging to assess the importance of these events to transcription of a specific gene of interest, given the many other potential binding sites for the TF within the genome. New imaging methods have made headway in solving this issue. One approach uses 3D orbital tracking (3DOT) to simultaneously monitor transcriptional dynamics from a PP7 reporter together with binding of individual Halo-tagged TF molecules. Unlike conventional confocal microscopy, 3DOT only collects intensity information from the site of transcription via orbital scanning of the sample, limiting the amount of photobleaching [73-75]. This method explicitly revealed the temporal coupling between TF binding and initiation of transcriptional bursts [76, 77]. In yeast, for example, an average TF (GAL4) binding time of 34 seconds initiates a mean burst duration of around 2.5 minutes. An analogous approach to computationally ‘fix’ the transcription site during imaging is target-locking 3D STED [78]. This live cell super-resolution technique was used for simultaneous molecular quantitation and spatial mapping of protein factors at the transcription site. A number of surprising features of
gene regulation were revealed for pluripotency markers. In particular, the gene encoding Oct4 (Pou5f1) appears to have around 20 molecules of the TF Sox2 clustered nearby when active, contrasting the textbook view of a single or dimeric TF binding and triggering a cascade of events. In addition, echoing the potential for transcription in the absence of enhancer-promoter communication [64], Sox2 TFs were spatially distinct from the active transcription site (Figure 1D). These approaches are a significant advance for the field, and will allow a more detailed understanding of the molecular interplay driving a transcriptional burst. In particular, a detailed dissection of the relative contributions of different proteins and complexes to multi-state models of dynamic transcription is now seemingly within reach.

Concluding Remarks

As bursting has finally entered mainstream thought, the challenge for the future is the same challenge faced by the entire study of transcription (see ‘Outstanding Questions’). How can we possibly formulate realistic models of dynamic transcriptional activity given the sheer number of factors influencing the process? The ability to directly visualize the interaction of different regulatory factors with transcriptional activity at loci of interest is a big step towards building such models. Limitations that need to be overcome include the restrictions on the number of different components that can be imaged in healthy living cells. Transcriptional regulation is often discussed in terms of complexes, but if one can only see a single component of a complex, then detailed mechanistic insight will remain elusive. Our impression is that the brute force approaches of drug treatment and genetics need to be superseded if we are to make more effective models. Optogenetics potentially provides a more subtle way of perturbing a system [79-81], although again, this takes the system outside its normal dynamic range, albeit in a potentially more sensitive manner. For all the reservations expressed here about the applicability of two state models, applying the new tools described above to genes which fit more simple regulatory regimes may yet provide the most straightforward route to a more complete understanding of bursting.
References


Figure 1. Approaches to Visualize Transcription.

A) Chromatin spreads from *Drosophila* embryos (image reproduced from [1]). The image shows a pair of sister chromatids aligned in parallel, with inferred initiation sites marked by α and β. Note the increasing size of the fibres (transcripts) extending from the central axis of each chromatid with increasing distance from the initiation sites (scale bar 1μm). Also note the fibre-free gaps (marked by arrows). B) Schematic of the MS2/MCP system for visualizing nascent transcripts. MS2 arrays are targeted into the gene of interest. The MS2 RNA forms stem loops, and can be detected at the site of transcription, as a fluorescent spot, by the MCP-GFP fusion protein. C) Transcription visualized using the MS2/MCP system, with stills from a movie sequence showing nascent RNA detected in bursts from the *act5* gene of *Dictyostelium* (scale bar 5μm). Normalised spot intensity values are shown in the plot below the film strip, with yellow dots corresponding to the images. D) Combining imaging of transcription, using MS2/MCP, with imaging of transcription regulators (image taken from Li et al. 2019). Images show nascent transcript foci from the mouse *Pou5f1* gene detected alongside different SNAP-tagged transcription factors (scale bars 300nm).

Figure 2. Key Figure. Models of Transcriptional Dynamics.

A selection of different model architectures used to describe transcriptional bursting dynamics.
Box 1. Popular Approaches to Measure Single Cell Transcription Dynamics

1. Live cell imaging of nascent RNA

These approaches use live cell RNA detection systems based upon stem-loop motifs from the genomes of RNA bacteriophages MS2 or PP7 [3, 82]. The distinct stem-loops structures have a high affinity interaction with the cognate coat proteins of the phages (MCP or PCP, respectively). By fusing fluorescent proteins to MCP or PCP, the stem loops recruit the fluorescent reporter, allowing live cell detection of the RNA. For imaging dynamic nascent transcript production, a sequence encoding an array of the stem loops (up to 128 repeats have been used) is targeted into the gene of interest. Upon transcription, the loops are incorporated into the nascent RNA and rapidly bind the fluorescent coat protein, allowing the nascent RNA to be visualized at the site of transcription as a fluorescent spot (see Figure 1B). The high specificity of these systems means MS2 and PP7 can be used together in the same cell to monitor activity of different genes, or to determine kinetic parameters of the transcriptional process, such as elongation rate, at a single gene.

2. Destabilised protein reporters

An alternative technique is to use protein reporters such as GFP or luciferase to observe activity of a particular gene. While these methods enable measurement of the output of a gene over time, and therefore provide dynamic information, using a protein rather than RNA reporter to model transcription requires the addition of several assumptions about intervening processes such as mRNA export and translation. Recent studies have tended to corroborate findings using other techniques, such as smFISH.

3. Single molecule RNA FISH (smFISH) and single-cell RNA sequencing (scRNaseq)

Fixed-cell measurements from methods such as smFISH produce distributions of both nascent and mature mRNA counts in single cells. From these data, parameters such as the frequency of burst initiation as well as the number of transcripts initiated (burst size) can be inferred. The approach of extracting dynamic behaviours from static measurement distributions has recently been extended to genome-wide approaches such as scRNaseq [11, 12]. In both cases, certain assumptions about
the regulation of the gene (i.e., whether it can be modelled as one-state, two-state or multi-state, see main text for further details) must be made which can limit the accuracy of such methods. More recent scRNAseq methods and analysis tools can give a coarse view of the changing gene expression of a cell [83, 84], adding an element of temporal detail onto otherwise static measurements.
Outstanding questions:

1. What are the relative contributions of the numerous regulatory inputs – involving tens, if not hundreds of cellular components – to transcriptional bursting?
2. How are the effects of these inputs integrated to generate the bursting patterns we observe?
3. What are the barriers to information transfer from cellular signalling to transcriptional apparatus? Is it really chromatin, or is the barrier function distributed through the regulatory network of the cell?
4. Does the apparently haphazard nature of transcription have any benefit for the organism, or is it simply a tolerable level of disorder?