2018

Design Principles Of Mammalian Transcriptional Regulation

Caroline R. Bartman
University of Pennsylvania, cbartman1@gmail.com

Design Principles Of Mammalian Transcriptional Regulation

Abstract
Transcriptional regulation occurs via changes to different biochemical steps of transcription, but it remains unclear which steps are subject to change upon biological perturbation. Single cell studies have revealed that transcription occurs in discontinuous bursts, suggesting that features of such bursts like burst fraction (what fraction of time a gene spends transcribing RNA) and burst intensity could be points of transcriptional regulation. Both how such features might be regulated and the prevalence of such modes of regulation are unclear. I first used a synthetic transcription factor to increase enhancer-promoter contact at the β -globin locus. Increasing promoter- enhancer contact specifically modulated the burst fraction of β -globin in both immortalized mouse and primary human erythroid cells. This finding raised the question of how generally important the phenomenon of burst fraction regulation might be, compared to other modes of regulation. For example, biochemical studies have suggested that stimuli predominantly affect the rate of RNA polymerase II (Pol II) binding and the rate of Pol II release from promoter-proximal pausing, but the prevalence of these modes of regulation compared to changes in bursting had not been examined. I combined Pol II ChIP-seq and single cell transcriptional measurements to reveal that an independently regulated burst initiation step is required before polymerase binding can occur, and that the change in burst fraction produced by increased enhancer-promoter contact was caused by an increased burst initiation rate. Using a number of global and targeted transcriptional regulatory perturbations, I showed that biological perturbations regulated both burst initiation and polymerase pause release rates, but seemed not to regulate polymerase binding rate. Our results suggest that transcriptional regulation primarily acts by changing the rates of burst initiation and polymerase pause release.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Immunology

First Advisor
Gerd A. Blobel

Second Advisor
Arjun Raj

Keywords
gene regulation, RNA FISH, single cell, transcription

Subject Categories
Allergy and Immunology | Genetics | Immunology and Infectious Disease | Medical Immunology

DESIGN PRINCIPLES OF MAMMALIAN TRANSCRIPTIONAL REGULATION

Caroline R. Bartman

A DISSERTATION

in

Immunology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

_____

Gerd A. Blobel, M.D., Ph.D.

Professor of Pediatrics

Co-Supervisor of Dissertation

_____

Arjun Raj, Ph.D.

Associate Professor of Bioengineering

Graduate Group Chairperson

_____

David Allman, Ph.D.

Professor of Pathology and Laboratory Medicine

Dissertation Committee

Craig H. Bassing, Ph.D., Associate Professor of Pathology and Laboratory Medicine

Marisa S. Bartolomei, Ph.D., Professor of Cell and Developmental Biology

Warren S. Pear, M.D., Ph.D., Professor of Pathology

DESIGN PRINCIPLES OF MAMMALIAN TRANSCRIPTIONAL REGULATION

ACKNOWLEDGMENT

Profound thanks to all the members of the Raj and Blobel labs for helpful discussion, technical assistance, and inspiration over the years of my Ph.D.

Thanks to Chris Hsiung for mentoring me during my rotation and afterward, Paul Ginart for helping me work through late nights in lab, Sarah Hsu as an inspiring example of careful experimentation and analysis, Vivek Behera and Ben Emert and Yogesh Goyal for scientific discussions, Connie Jiang and Sara Rouhanifard for contributing humor as well as looking at my second paper a number of times. Thanks also to the Immunology Graduate Group, especially Dave Allman and the excellent Mary Taylor for help and support, as well as my cohort, the entering class of 2013. Thanks to my supportive and patient thesis committee, Craig Bassing, Marisa Bartolomei and Warren Pear.

Thanks especially to my mentors Arjun and Gerd. Both have been extremely patient with me while I made my way through graduate school, and have given me all the freedom and support I could have asked for. As Francis Galton said of Charles Darwin's writings, 'They enlarged the horizon of my ideas. I drew from them the breath of a fuller scientific life.' My taste in science has become immeasurably better because of both my mentors.

Finally, thanks to my family, who have supported me throughout my Ph.D.

ABSTRACT

DESIGN PRINCIPLES OF MAMMALIAN TRANSCRIPTIONAL REGULATION
Caroline R. Bartman

Gerd A. Blobel

Arjun Raj

Transcriptional regulation occurs via changes to different biochemical steps of transcription, but it remains unclear which steps are subject to change upon biological perturbation. Single cell studies have revealed that transcription occurs in discontinuous bursts, suggesting that features of such bursts like burst fraction (what fraction of time a gene spends transcribing RNA) and burst intensity could be points of transcriptional regulation. Both how such features might be regulated and the prevalence of such modes of regulation are unclear. I first used a synthetic transcription factor to increase enhancer-promoter contact at the β -globin locus. Increasing promoter-enhancer contact specifically modulated the burst fraction of β -globin in both immortalized mouse and primary human erythroid cells. This finding raised the question of how generally important the phenomenon of burst fraction regulation might be, compared to other modes of regulation. For example, biochemical studies have suggested that stimuli predominantly affect the rate of RNA polymerase II (Pol II) binding and the rate of Pol II release from promoter-proximal pausing, but the prevalence of these modes of regulation compared to changes in bursting had not been examined. I combined Pol II ChIP-seq and single cell transcriptional measurements to reveal that an independently regulated burst initiation step is required before polymerase binding can occur, and that the change in burst fraction produced by increased enhancer-promoter contact was caused by an increased burst initiation rate. Using a number of global and targeted transcriptional regulatory perturbations, I showed that biological perturbations regulated both burst initiation and polymerase pause release rates, but seemed not to regulate polymerase binding rate. Our results suggest that transcriptional regulation primarily acts by changing the rates of burst initiation and polymerase pause release.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: Introduction

My graduate work focused on the broad question of how cells can change their state in response to stimuli. The cells of a eukaryotic organism all share the same genome; however, they differentiate from a single zygote into many different cell types that carry out different functions mediated by the expression of cell-type-specific suites of proteins. A major focus of biological science has been to understand how cells with the same genome can induce and maintain such divergent functional states. Relatedly, eukaryotic cells must be able to respond quickly to certain stimuli by changing protein expression: canonical examples of such stimuli include heat shock or inflammatory signals. Both cell-type identity and functional responses to signaling are chiefly governed at the level of DNA transcription into RNA, though other processes like protein post-translational modification and degradation also play important roles. Thus, understanding how eukaryotic cells induce and maintain the transcription of a set of genes is critical to understanding the function, health, and disease of eukaryotic organisms. Below, I will review the key aspects of transcriptional regulation of higher eukaryotes that have formed the basis for my graduate work. Figure 1.1 shows a schematized outline of this Introduction.

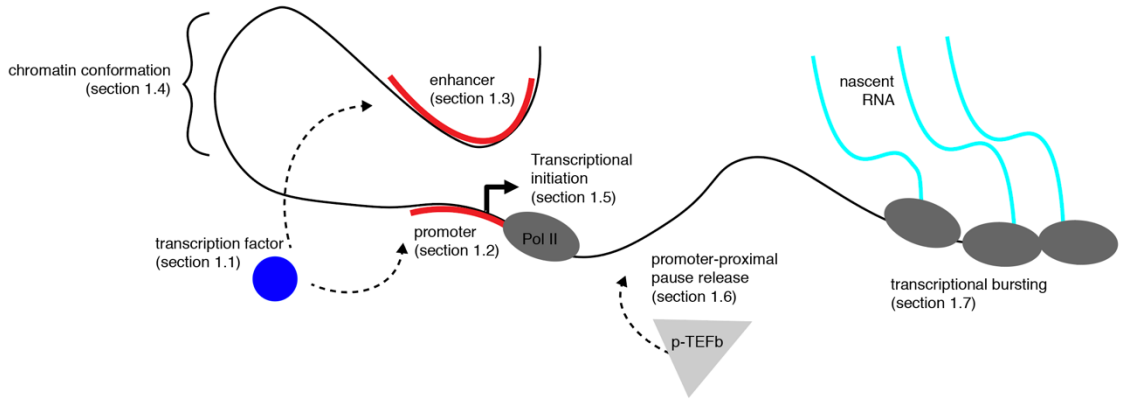


Figure 1. 1: Schematic of the events of RNA polymerase II transcription in mammalian cells, highlighting steps that will be reviewed in this Introduction.

1.1 Transcription Factors Orchestrate Cell Function

RNA transcription is a complex process which is governed at multiple levels, but arguably the prime movers in this process are transcription factors. Transcription factors are proteins that bind cognate target sequences in genomic DNA in order to induce or repress transcription. These factors can be sufficient to determine cellular function: for example, expressing certain sets of transcription factors can overwrite a cell's initial transcriptional program, and reprogram cells into a new cell type including dedifferentiating cells into fully pluripotent embryonic stem cells[1]. Conversely, transcription factors can also be necessary to establish cell type identity: certain transcription factors are absolutely required for differentiation of a given cell type, such as GATA-1 in erythroid cells[2], which will be further described at the end of this section.

Broadly, each transcription factor carries out its function using two modular domains: a DNA binding domain and a transcriptional transactivation domain[3]. The DNA-binding domain mediates sequence-specific binding to DNA. Many different families of conserved binding elements have been crystallized, and each structure confers a stereotypical type of DNA sequence recognition[4–6]. The quality and stringency of sequence specificity varies widely depending on the transcription factor, ranging from short motifs around 6-10 base pairs long with a wide range of degeneracy permitted to highly specific motifs up to 20 base pairs long in the case of CTCF[7–9] . Such transcription factor recognition sequences have been identified both using in vitro protein-DNA binding assays and by measuring binding in cells to plasmids or genomic DNA using methods such as chromatin immunoprecipitation (ChIP)[7]. Transcription factors bind to influence transcription of genes at two main types of locations in the genome: proximal to the transcriptional start site, in the region known as the gene promoter, or at distal regulatory regions known as enhancers (both discussed in detail below).

However, motif recognition is not sufficient to predict genome-wide transcription factor binding: many factors fail to bind the majority of their consensus motifs in the genome[8] . Beyond

direct sequence recognition, transcription factor binding is also influenced by the local sequence context surrounding its recognition motif. For example, clusters of binding sites serve to give a region a higher avidity for a transcription factor, and thus can raise the likelihood of transcription factor binding to a region[10]. The spacing between such binding motif sites may also influence the likelihood of transcription factor binding[11–13]. Binding of different transcription factors nearby can also promote transcription factor binding[14–17]. Such cooperative binding can be mediated by direct protein-protein interactions (for example in the case of GATA-1 and Tal1, described below), or by one factor helping to displace nucleosomes away from a second factor's binding site[7] .There is also evidence that factors can promote each other's binding indirectly via bending of DNA[18–20].

However, even sequence context is often not sufficient to predict genome-wide transcription factor binding patterns, due to other factors that influence binding. For example, genomic DNA in eukaryotic cells is wrapped around protein octamers known as nucleosomes, and nucleosome binding can block transcription factor binding[21–24]. However, certain transcription factors, known as 'pioneer factors', possess the ability to recognize partial cognate sites on nucleosome-bound DNA[23,25]. These factors can recruit nucleosome remodelers such as the Swi/SNF complex, which can remove nucleosomes to allow further non-pioneer factors to bind their cognate sites[23,25]. Such pioneer properties are thought to be critical in allowing regions of DNA to transition from the heterochromatinized state to a transcriptionally-accessible state, in which even transcription factors can access their cognate binding sites. Such a property is likely critical for cellular transdifferentiation, and also likely in normal cellular differentiation[26,27]. Given the complex interplay between transcription factor motif, nearby sequence context, and nucleosome occupancy, an open area of study addresses how to combine multiple types of data to predict transcription factor binding with better accuracy, to ultimately be able to reconstruct the mechanisms driving gene regulation in a given cell type[17,28].

Once a transcription factor successfully binds a cognate site, it must then carry out its function of transcriptional activation or suppression mediated by its transactivation domain. Transactivation domains are often characterized by an abundance of acidic residues[3,29]. These

domains can mediate their function by recruiting activating or repressive complexes, which can post-translationally modify surrounding histones with repressive or activating marks or can remodel nucleosomes to promote transcription factor binding and transcription itself[7,30]. For example, the interferon-beta enhancer supports the binding of multiple transcription factors, which ultimately results in the recruitment of acetyltransferases such as CBP/p300 to promote transcription of the interferon gene[20]. On the other hand, many zinc-finger transcription factors use their repressive KRAB domains to recruit a complex which imposes histone-3 lysine-9 trimethylation to silence transcription of a target region[31]. However, such black and white cases as repressive zinc fingers may be the exception rather than the rule: in many cases, the same transcription factor domain can recruit either activating or repressing complexes in a context dependent manner: GATA-1 described below is one such example[7,30]. Thus, neighboring factor binding can influence both the DNA binding and the transactivation properties of a given factor.

Transactivating domains have been more challenging to study structurally than DNA-binding domains of transcription factors, since many of these domains are disordered and thus elude crystallization[4]. Currently, there is a resurgence of interest in these disordered domains because of these same physical properties: some have hypothesized that such disordered protein domains may be able to drive transcription via ‘condensation’ of protein complexes using a phase separation mechanism[18,32–34]. Phase separation is the phenomenon that many weak cooperative interactions between multivalent molecules can bring together clusters of proteins that appear to behave similarly to liquid droplets, displaying properties like excluding outside proteins and flowing[18,33]. Such phase-separation interactions can be promoted by post-translational modifications of disordered protein domains, such as the phosphorylation of the C-terminal domain of Pol II itself and phosphorylation of other transcription factors[18,35,36]. This type of condensation-based interaction is proposed to drive the nucleation of the clusters of transcription factors and of Pol II that have recently been observed by high-resolution imaging in live cells[37,38]. Phase separation control of transcription will be a very exciting area of study, but it is challenging to disentangle the role of the physical properties conferred by such disordered

regions of transcription factors from more conventional mechanisms of interaction with and recruitment of protein partners.

1.1a: The GATA-1 transcription factor

In my work and in other studies from the Blobel lab, we have taken advantage of the requirement for the transcription factor GATA-1 in the red blood cell transcriptional program in order to differentiate red blood cells in an inducible manner. GATA-1 was the first-discovered member of the GATA transcription factor family, members of which bind to a (A/T)GATA(A/G) motif in the genome[39–41]. GATA-1 was first isolated due to its binding to the beta-globin locus in red blood cells[41,42]. Its induction in granulocyte-monocyte progenitors drives differentiation of precursors into red blood cells and megakaryocytes[41]. As a result of its necessity in these cellular programs, embryonic stem cells lacking GATA-1 can contribute to every tissue in mice except red blood cells and megakaryocytes[43,44], and mice lacking GATA-1 die at embryonic day 10.5 as a result of severe anemia[2]. GATA-1 is highly conserved from mice to humans, and humans with GATA-1 mutations display anemia and thrombocytopenia[45,46].

GATA-1 is an example of a transcription factor which both induces some genes and represses others in a context-dependent manner. Other transcription factors binding near GATA-1 at a specific binding site seem to govern whether activating or repressive complexes are recruited to a GATA-1 binding site[47]. For example, when Tal-1 and GATA-1 bind together, this combination tends to produce transcriptional activation due to direct protein-protein interaction between the factors; indeed, mutations of the surface of GATA-1 where it contacts Tal-1 reduces its ability to activate transcription [46]. However, at the majority of its sites GATA-1 mediates transcriptional repression, and thus it helps red blood cells to globally shut down transcription during their differentiation[41]. At such repressive binding sites, GATA-1 tends to bind alone in the absence of the Tal-1/SCL complex[47]. Such repression is mediated by the recruitment of the NuRD histone deacetylase complex by the GATA-1 binding partner FOG-1[47,48].

Taking advantage of GATA-1's necessity and sufficiency for erythroid differentiation, Weiss and colleagues knocked out this factor in a mouse erythroid precursor cell line thus blocking differentiation, and then re-inserted a GATA-1-estradiol receptor fusion transgene, creating the 'G1E-ER4' cell line[49–51]. Upon treatment with estradiol, GATA-1 protein is stabilized and induces a highly reproducible differentiation cascade similar to red blood cell differentiation *in vivo*[49–51]. G1E-ER4 cells exposed to estradiol undergo terminal differentiation similarly to red blood cells *in vivo*, inducing transcription of genes typical of the erythroid lineage including the *Slc4a1* (Band3) anion channel, the *Slc25a37* iron transport channel, and the alpha- and beta-globin subunits of hemoglobin[52–54]. This suite of proteins ultimately causes these cells to hemoglobinize and turn red, allowing visual tracking of their differentiation[55], and to suppress transcription of almost all genes, although cells do not fully progress to enucleation unlike their *in vivo* counterparts. These G1E-ER4 cells also repress stem-cell associated proteins that maintain pluripotency and the stem-cell transcriptional program, including the *Gata2* and *Myc* transcription factors, and the *Kit* cytokine receptor for stem cell factor cytokine[53,54]. This cell culture system thus represents a useful tool to study how a transcription factor can drive differentiation of a lineage and has been extensively previously characterized. I took advantage of these properties in my graduate work to compare this body of knowledge to the single-cell transcriptional properties produced by GATA-1 induction in these cells to better understand the regulation of transcriptional bursting.

1.2 Transcriptional regulatory regions: promoters

Transcription factors exert their effects at promoters, near the beginning of genes, and at enhancers, or distal regulatory regions. In bacteria, promoters consist of highly conserved sequence motifs that directly recruit polymerase in order to drive transcription of genes[56,57]. Therefore, promoters were the first studied regulatory regions in eukaryotes, and at first were assume to completely govern transcription as in bacteria[58]. However, in higher eukaryotes, the

sequences of promoters were found to be quite variable between organisms and between different genes in the same organism[56,59,60]. Higher eukaryotes display some conserved categories of promoter motifs shared by multiple promoters, such as the TATA box and the Initiator (Inr) element, but such motifs are short (less than 10 base pairs on average) and relatively degenerate[58,60]. Higher eukaryotic promoters fall into two broad categories based on their transcriptional activating patterns: CpG promoters are rich in C and G nucleotides and tend to initiate transcription in a broader region and are often associated with housekeeping genes, while 'narrow' promoters which often have a TATA box drive transcription starting from a more focused location in the genome[21,58,60]. Promoters also typically display the histone 3 lysine 4 trimethylation mark[56]. Promoters, especially 'narrow' promoters, also often include a well-placed nucleosomes at the transcription start site, which must be displaced in order for transcription to occur[21,56]. Due to the multiple categories of promoters with short and relatively degenerate motifs, higher eukaryotic promoters are difficult to identify using sequence alone: the gold standard techniques for promoter identification are experimental approaches that capture the 5' starting end of transcripts with high sensitivity, such as GRO-seq or CAGE[56,61]. Eukaryotic promoters serve as the assembly point for the polymerase initiation complex, a highly orchestrated process that will be discussed further below.

1.3 Transcriptional regulatory regions: enhancers

In higher eukaryotes, transcription factors binding at distal regulatory regions is critical to the establishment and maintenance of cell identity, perhaps as a result of the requirement for more complex spatial and temporal regulation associated with higher eukaryotic development [4]. Distal regulatory elements acting in mammalian cells were first identified in the SV40 virus[62]. In mammalian systems, enhancers are typically a few hundred base pairs in length (though this can be altered depending on the approach used to define enhancers, as discussed below) [8]. Enhancers can be located up to around a megabase from their target gene[63–65] or can be quite

close to their target gene, although it becomes difficult within around a kilobase to definitionally distinguish an 'enhancer' from an extended promoter region. Enhancers can also typically function independent of their orientation with respect to their target gene[66]. These two properties, transcriptional activation at a distance and orientation independence, are the canonical defining characteristics of enhancers[66].

Enhancers consist of multiple transcription factor binding sites, often including sites for multiple different factors. A given enhancer is often active only in certain developmental and cell-type contexts, in those contexts in which transcription factors are able to bind. In some situations, the assembly of this group of factors seems to occur in a specific order, and small sequence changes to the enhancer can abrogate all activating function[8,67]. However, in other cases, binding sites appear to drive transcription in a more additive manner, where different factors bind independent of each other to the same enhancer and each promotes enhancer activity[68,69]. A given gene can be controlled by different enhancers in different cellular contexts, such as in the case of the *Myc* gene[63,64]. The same gene can also be controlled by multiple distal regulatory regions, though whether these are considered to be separate enhancers or components of a massive 'superenhancer' can be somewhat semantic (discussed below)[70,71] .

Once transcription factors bind an enhancer, this complex must then activate transcription of the target gene by communicating to the promoter in some way. Enhancers are thought to promote transcription by transferring activating complexes to promoters or by increasing the local concentration of such activators. Transcriptional activators recruited by enhancers and brought to promoters include the CBP/p300 histone acetyltransferase complex and perhaps Pol II itself[4]. It has also been suggested that enhancers can transfer factors that help clear nucleosomes from promoters, or alternatively that enhancers can help promote promoter-proximal pause release of Pol II[8,72,73]. However, the mechanisms of how enhancers confer transcriptional activity to promoters are still somewhat murky, since it is difficult to specifically perturb recruitment of complexes to specific sites in a targeted manner.

Different models of enhancer-promoter communication have been proposed: for example, that activating factors tracked linearly along the genome from an active enhancer to its target gene, or alternatively that the enhancer transfers activating factors directly by looping to its target gene[66,74,75]. The latter looping-type mechanisms have been supported by findings from chromatin conformation capture techniques, which use covalent cross-linking methods to sequence distal regions of DNA that contact each other in the nucleus[76–78]. Moreover, the Blobel lab showed that enhancer-promoter looping can be sufficient to cause transcriptional activation at the beta-globin locus[79]. Conversely, in several systems, transcription factor mutations or deletions that abrogate enhancer-promoter contact also reduce transcription, although in such cases the causative role of looping is unclear[52,80,81]. At other gene loci however enhancers contact promoters prior to transcriptional activation, suggesting that a further event independent of enhancer-promoter looping may be necessary for transcription[82,83]. More complexity is conferred by high-resolution imaging studies that showed that in some systems, enhancers and promoters were actually farther away ('decondensed') during active transcription compared to a developmental stage with no transcription[84]. In general, it has been extremely difficult to experimentally uncouple the role of promoter-enhancer contact from other events that occur during transcription. Live-imaging approaches (described more extensively below) may be key to at least revealing the dynamics and sequence of events in transcription, if not their causative roles, though achieving the resolution required to visualize transcriptional events in intact loci is an ongoing challenge[38,85].

De novo identification of genomic regions that serve as enhancers is an ongoing challenge. The gold standard, more practical now given the development of CRISPR, is enhancer deletion or mutation in cells or organisms and measurement of resulting alteration of target gene transcription (for example,[70,86]). However, this approach is relatively cumbersome and low-throughput. The next best type of approach measures enhancer activity using a reporter system, a type of experiment that can be highly multiplexed[87,88]. However, there is some evidence that such assays may not always report faithfully on endogenous enhancer activity[89]. Approaches to

enhancer identification that don't directly measure transcriptional activity are the least costly and most efficient, but any individual criterion to identify putative enhancer regions has not been perfectly reliable[90]: such approaches include identifying evolutionarily conserved distal regions, measuring transcription factor binding of one or multiple factors using ChIP, mapping locations of histone-3 lysine-27 acetylation, of histone-3 lysine-4 monomethylation, of DNAse-I hypersensitivity distal to transcription start sites, or of Pol II binding distal to transcription start sites, or measuring transcription of distal regions themselves using a method like PRO-seq (described below)[66,91,92] . Hybrid approaches combine multiple of these markers and give more faithful information than any individual marker alone[93]; hopefully as data from genetic enhancer manipulation accumulates, such algorithms to identify enhancers will be optimized to help identify regulatory regions *de novo* in cell types and diseases of interest.

A great deal of enthusiasm in recent years has been devoted to the concept of 'superenhancers': large enhancer regions that function as more than the sum of their constituent regions[94–96]. These regions were proposed to not only be more powerful than the sum of their constituent regions, or than shorter 'typical' enhancers, but were also thought to be more sensitive to perturbation, and additionally to be more likely to direct transcription of important cell-identity genes like the transcription factors *Myc* (in several cell types) and *Sox2* (in ES cells)[96]. This categorization of enhancers implied that enhancer length and histone-modification density should be able to predict enhancer strength. However, careful studies have suggested that the existence of a discrete superenhancer category with distinct properties from 'typical' enhancers perhaps was overstated and not a general rule[97]: two elegant studies in particular suggested that regions of an enhancer might contribute additively to the activity of the whole enhancer[71,98].

Beyond the identification of enhancers which can drive transcription, the field has also struggled to identify rules for how an active enhancer selects which target gene to activate. Indeed, the initial observation of enhancer activity showed that the SV40 large T antigen enhancer could drive the mammalian beta-globin gene, immediately showing that enhancers could display some degree of promiscuity[62]. In some cases, enhancers are exquisitely specific,

skipping intervening genes to activate a further away specific target gene. For example, a study shows that in embryonic stem (ES) cells, only 2/3 of active promoters genome-wide contact the nearest enhancer, while the other genes appear to be controlled by non-neighboring enhancers[8,99]. However, in most transgenic experiments, an enhancer is able to drive a wide variety of promoters[87,100,101]. One simple solution is that in some cases specificity does not seem to be critical to biological function: in some cases, enhancer deletion influences multiple genes in a region[102,103]. In a series of elegant high-throughput studies in *Drosophila*, it was shown that though enhancers were not absolutely specific, they had preferences for promoters of the same 'type', i.e. housekeeping gene enhancers promoted the transcription of housekeeping gene promoters more strongly, while developmental gene enhancers promoted the transcription of developmental gene promoters most strongly[88,104]. To what extent this discovery of enhancer categories applies to mammals is as yet unclear. In some cases, genome organization seems to help promote enhancer target specificity: for example, the critical oncogene *Myc*, driven by different strong enhancers in different tissue types, sits in a relatively gene-poor region, perhaps to avoid inappropriate activation of the *Myc* gene, or of other genes by its enhancer[63,64]. Higher-order chromatin structure such as the delimitation of so-called 'topologically associating domains' (TADs) is also important for ensuring enhancer fidelity in some cases[105–108], as will be further discussed below. Heterochromatinization of stage-inappropriate genes may also prevent enhancer activation in some cases[26]. However, general mechanisms of enhancer specificity remain somewhat unclear: perhaps a combination of strong gene repression, and a robustness of cells to some inappropriate gene activation, is sufficient to explain how appropriate cell function is maintained without a high degree of intrinsic enhancer specificity.

1.3a: The beta-globin locus control region

One of the key model systems which has allowed a deeper understanding of transcriptional control by enhancers is the beta-globin locus in mammalian red blood cells. The mammalian beta-globin locus includes several developmentally regulated forms of beta-globin,

including embryonic, fetal, and adult forms, spread over around 100 kb[109,110]. All of these genes are regulated in their respective developmental stage by a large enhancer known as the beta-globin locus control region. This region consists of 5 DNase-I hypersensitive sites, each with binding sites for different transcription factors[111]. In particular, GATA-1 and KLF-1 must bind the LCR and mediate clustering of the hypersensitive sites with the promoter in order to promote transcription of the stage-appropriate beta-globin gene[52,81]. Indeed, the Blobel lab showed that redirecting this enhancer complex to a developmentally-inappropriate globin gene is sufficient to activate that gene: specifically, they were able to increase transcription of the fetal form of beta-globin in adult erythroid cells[112]. The LCR generally conforms to the additive model of enhancer function: deletion of each hypersensitive site reduces beta-globin transcription, and combined deletion of multiple HS sites has stronger effects than individual deletions[113]. The LCR may activate transcription of target genes via several mechanisms: deletion of parts of the LCR reduces PIC formation at the promoter, and also may impact promoter-proximal pause release[72,73]. The beta-globin locus has also been an interesting model of how an enhancer may drive transcription of multiple genes, and a number of studies have investigated whether the enhancer may drive transcription of multiple target genes simultaneously, or whether instead promoters compete for the activity of the LCR[114,115]. In my work, we took advantage of the previous thorough characterization of beta-globin locus regulation to investigate single-cell transcription and transcriptional bursting regulation in this system.

1.4 Chromatin structure: higher-order organization in transcription

One of the main issues raised by the existence of distal regulatory regions is how a specific enhancer is coupled to its target promoter, and how it avoids spurious activation of other nearby genes. One possible solution to this is specific promoter-enhancer looping, discussed above. Another hypothesis is that higher-order conformation of genomic DNA in the nucleus helps govern transcriptional fidelity. Historically, the study of DNA organization in the nucleus

began with imaging studies such as the first images of chromosomes performed by Boveri, leading to studies asking whether there were specific regions for each chromosome, or territories, in the nucleus[105,116,117]. Indeed, genomic regions are much more likely to contact regions within their own chromosome than between chromosomes, and there is evidence for some selective radial positioning of chromosomes, though chromosome locations are not identical in each cell nucleus and may change after a cell divides[118].

The finding of chromosome territories may explain the limitation of enhancer activity to its own chromosome; however, if genomic structure governs enhancer action, then finer layers of nuclear structure must be responsible to avoid spurious activation of genes nearby on the same chromosome. The local 'domain' structure of genomic DNA has been enthusiastically studied for the past ten years, especially after the advent of high-throughput chromosome conformation capture techniques[78]. Such methods employ formaldehyde crosslinking to capture physical contacts between linearly distal regions of DNA; high-throughput versions of these methods, such as Hi-C, enable more unbiased analysis of contacts without requiring selection of a single candidate region of interest[78,119] . Lieberman-Aiden et al. used genome-wide chromosome conformation capture to show that the nucleus is partitioned into active ('A') and repressed ('B') domains, and that genomic regions tend to clump like with like[78]. This 'compartment' structure of chromosomes in the nucleus may help avoid spurious activation of repressed domains, and may help ensure that even if enhancers act somewhat promiscuously on multiple genes, the other physically proximate regions are already being actively transcribed rather than suppressed. This compartmental clustering of active and of inactive regions has been validated directly by imaging[120]. Moreover, inactive domains are more likely to be excluded from the center of the nucleus and instead pushed to the lamina[121]. This laminar localization can directly disfavor transcription in some cases[122]. However, elegant single cell studies showed that localization to the nuclear lamina is non-deterministic, meaning that a lamina-associated region in one cell may shift away from the lamina when the cell undergoes division or vice versa[123] Thus, evidence suggests that compartment structure may help confine enhancer activity to active genes on the

same chromosome, but that such compartmentalization is not absolutely deterministic and can vary cell to cell.

A further level of organization thought to promote specificity of transcriptional regulation are TADs or topologically associating domains, first characterized in 2012 by several groups[105,106,124]. Satisfyingly, the existence of TADs was more recently confirmed by two orthogonal experimental approaches[120,125]. These regions average around 1 megabase, although different groups have used different algorithms to identify them and thus TAD sizes and locations vary between publications[126]. In general, however, TADs are computationally and functionally defined as genomic regions which interact more strongly within themselves than with regions outside their boundaries[105,106,124]. Many TADs have binding sites for CTCF, a protein first characterized as an 'enhancer insulator', at their boundaries, and these CTCF sites are typically convergently oriented[126,127]. CTCF has been shown to form homodimers, and thus such convergent binding sites are thought to mediate interactions between CTCF molecules sitting at the base of TAD loops [126]. It is hypothesized that DNA loop extrusion by the cohesin complex, proceeding from the center of TADs until cohesin is blocked by CTCF at the base of each TAD loop, establishes and maintains TADs as domains separated from each other[128–130]. Enhancers typically tend not to activate genes outside their TAD[108]. Moreover, TAD boundaries are relatively conserved compared to other intergenic regions, both between different cell types in the same organism, and across different mammalian species, and even appear to be somewhat consistent in single cells[129,131,132]. In some cases, mutation of TAD boundaries can lead to mis-activation of genes by enhancers from the neighboring TAD, as such a model would predict[107,133]. Moreover, global genome methylation in cancer cells led to enhancer-promoter miswiring in one cell system, which the authors suggested was due to the inability of CTCF to bind methylated chromatin[134]. However, in some cases, TAD boundary mutation has negligible effects on transcription[135]. Moreover, transient depletion of CTCF from cells, which dissolves the structure of most TADs, has moderate or negligible effects on transcription, depending on the system used[136,137]. (Note that permanent CTCF depletion is toxic for cells, so transient studies were critical to explore the

function of CTCF in transcription.) Given these equivocal results, it is likely that TADs may not be as critical for governing transcription as the earliest studies may have suggested. In future, imaging of TADs in single cells coupled with single cell transcriptional measurements may clarify to what extent TADs govern localization and transcription of genes within them.

1.5 The mechanics of transcription: initiation

Once a gene has the potential to transcribe due to transcription factor binding at its enhancer and promoter and recruitment of activating factors, the mechanics of transcription itself come into play. The polymerase initiation complex (PIC) must be assembled at the transcription start site, recruited to the degenerate promoter sequences described above[58,60] . The mammalian PIC consists of more than 40 proteins, including notably TFIIB, C, D, E, F, and H, which assemble in a stereotyped sequence and conformation[138–141]. For example, the TFIID complex directly recognizes the TATA box and other promoter motifs and binds DNA, and serves to nucleate the assembly of the other factors on DNA[139,142]. These other proteins act both to unwind the DNA at the promoter, a step which specifically requires the helicase activity of TFIIH[139,143], and to bring Pol II itself to the beginning of the gene. PIC assembly and unwinding of promoter DNA has been suggested to be a rate-limiting regulated step in eukaryotic transcription[139,144] .

Structural studies of the initiation complex have helped reveal the steps involved in PIC formation. The large size and numerous components of the PIC historically made it extremely challenging to crystallize using conventional methods[141,145]. However, the development of the cryo-EM technique, which applies a newly-developed high-resolution camera and computational alignment techniques to identify structural conformation of large multiprotein complexes to several angstrom resolution, has been valuable in clarifying the sequence of events involved in PIC assembly[141,145]. While conventional protein crystallization yields a single crystal structure, the cryo-EM technique averages together images of individual complexes: therefore, if the complex exists in several different conformational states, a single cryo-EM experiment can yield

information on all of these states, which has proved valuable in studying PIC assembly[141,145]. For example, the Nogales group found that TFIID, the factor which directly binds the TATA box of a promoter, mediates a sharp bend in promoter DNA when it binds, while the TFIIH complex which helps melt the promoter DNA to allow RNA transcription also changes conformation upon binding to the promoter[146–148]. Ultimately, such studies will continue to give us a better understanding of the dynamics and steps of transcriptional initiation.

1.6 The mechanics of transcription: promoter-proximal pause release

In bacteria, polymerase recruitment is the main checkpoint governing transcriptional activation of a gene, and initially this was assumed to be true of eukaryotic genes[4]. However, more recently a second major checkpoint controlling the rate of RNA production by Pol II was identified by the Lis lab and others: promoter-proximal pausing[61,149] . In this process, Pol II is stalled approximately 30 to 50 base pairs downstream of the transcriptional start site[149]. Elegant experiments blocking initiation and measuring the half-life of paused Pol II complexes have showed that such paused complexes can be quite stable, in the range of minutes to tens of minutes[150]. In order for polymerase to transcribe full-length RNA, the pTEF-b kinase complex, made up of cyclin T1 and CDK9 proteins, must phosphorylate the C-terminal domain of Pol II as well as negative elongation factors[149]. Indeed, direct tethering of pTEF-b in a *Drosophila* system is sufficient to drive transcription of heat shock genes in the absence of stimulus, demonstrating that this complex is key to the pause release process[151]. However, other proteins including BRD4 and the super-elongation complex also can collaborate in the pause release process[152,153]. Promoter-proximal pause release is globally required at all genes in higher eukaryotes, because pharmacological blockade of pTEF-b globally arrests transcription[150,154] . The degree of pausing at different genes is governed by the promoter elements of that gene[155]. It remains unclear exactly how the rate of pause release is regulated: there must be some gene-specific element to this regulation, because certain genes are maintained in the paused state without transcription

(for example heat shock genes in *Drosophila*), while other genes in the same cells undergo elongation. In future, live imaging approaches of RNA transcription and of proteins involved in pause release perhaps can be applied to help clarify the dynamics and regulation of this process.

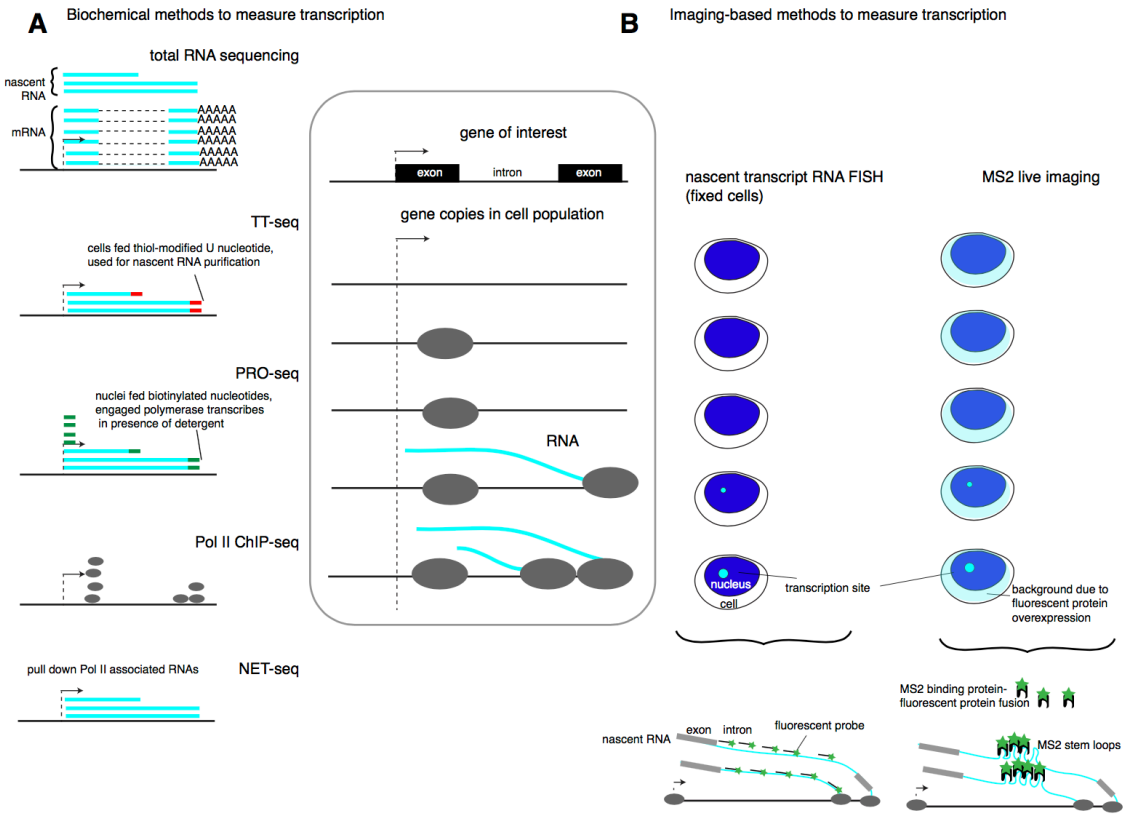


Figure 1.2: Experimental methods used to measure transcription. A) Biochemical population-averaging techniques discussed in section 1.6. B) Imaging-based techniques discussed in section 1.7.

Though often polymerase initiation and pause release are considered as independent and sequential processes, recent evidence demonstrates that they are interrelated. Specifically, the very size of the PIC means that if a polymerase complex is paused in the promoter-proximal region of a specific gene copy, no second polymerase molecule is able to initiate at the promoter until the first pause polymerase has moved into the gene body[156,157]. This cross-inhibition of initiation and pausing is a critical consideration when modeling the process of transcription in higher eukaryotes. Promoter-proximal pausing has also been suggested to be a mechanism by which genes can be quickly activated upon stimulus[149], however, since only 2-4 molecules of

RNA could be generated per cell by such an approach without further transcriptional initiation (depending on the cell cycle status of the cell in question), we suggest that it seems unlikely that this mechanism is as quantitatively essential as some have suggested.

Biochemical approaches have been used extensively to measure nascent transcription, and the choice of technique in large part should be dictated by the experimental question at issue (Figure 1.2). The most straightforward approach is total RNA sequencing (since poly-A selection will omit most nascent RNA species): the abundance of intronic sequences, which are degraded soon after transcription, can serve as a proxy for active transcription[158]. However this approach is needlessly costly since the majority of species sequenced are mature RNA. Thus, methods to enrich for nascent RNA can be more practical in measuring nascent transcription. Such enrichment approaches include TT-seq, which allows cells to transcribe briefly in the presence of 4-thiouridine: incorporation of this nucleotide allows biochemical enrichment and sequencing of recently transcribed RNAs[159]. GRO-seq and PRO-seq are other techniques that allow biochemical enrichment and sequencing of nascent RNAs[61,149,160]. PRO-seq and GRO-seq, like TT-seq, provide high-resolution information about nascent transcription, which can be critical to identify precise transcriptional start sites, and to provide directional information about transcription such as in cases of divergent transcription[149]. A downside of the PRO-seq/GRO-seq techniques is that they require transcription to take place in isolated nuclei after the addition of the detergent sarkosyl, which might bias results due to the nonphysiological conditions. A final type of approach uses chromatin immunoprecipitation to enrich for Pol II, either to measure Pol II binding as a proxy of transcription, or to purify Pol II associated RNAs and perform RNA sequencing (NET-seq)[161,162]. This approach is lower-resolution than PRO-seq/GRO-seq, not allowing nucleotide-level information of how much a region has been transcribed, but allows transcription to occur in situ in unmanipulated intact cells, and thus may provide the most physiological record of nascent transcription. Thus, choice of biochemical technique to measure transcription should be dictated by experimental question. Most of these approaches can help measure promoter-proximal pausing, with slight variations in the analysis method: the short nascent RNAs transcribed before

pausing are quite unstable and very short, so total RNA-sequencing is not a good approach for this sort of question. PRO-seq and GRO-seq reveal Pol II that is 'transcriptionally engaged' even if the RNA is unstable, so pausing will produce a strong promoter-proximal signal in these techniques[149]. Similarly, Pol II ChIP-seq will reveal a strong peak of Pol II if many cells contain paused Pol II near the promoter [149]. Note that the presence of these pause Pol II signals will be influenced by both the number of cells with paused Pol II, and by the average duration of time that polymerase remains paused without undergoing elongation at that gene. All such techniques average the transcriptional behavior of many cells, and this should be kept in mind while creating models based on such data.

1.7 Transcription in single cells: transcriptional bursting

The biochemical approaches to measuring nascent transcription described above have been extremely powerful; however, the advent of imaging studies of nascent transcription revealed that some of the complexity of transcription had been occluded by the population-based nature of such assays. First, it was observed that cells containing mRNA for highly expressed genes did not display active transcription of those genes at all times[163–165]. Moreover, the distribution of mRNAs for a given gene in a population of cells could not be fit by a Poisson distribution: this observation suggested that RNA transcription was likely to happen in 'bursts' of multiple RNAs, and there must be a step upstream of transcription that altered the likelihood of RNA production over time[166,167]. (This distributional argument could be compared to recording the number of cars on a stretch of rural highway- with no traffic control, cars are distributed in a roughly Poissonian manner- compared to recording the number of cars on a city street- in the latter case, clusters of cars are likely to appear, and thus one could infer the presence of a stoplight controlling traffic flow even if one couldn't see the stoplight itself. RNA patterns adhered to the latter pattern rather than the former.) Moreover, transcriptional bursting was found to be stochastic in the sense that different alleles in the same cell transcribed in an independent and

uncoupled manner[164,168,169]. These initial studies were performed with the single-molecule RNA fluorescence in-situ hybridization (FISH) technique (Figure 1.2)[163,170] ,which uses multiple fluorescently labeled DNA oligonucleotides to hybridize to RNAs of interest and allow direct quantitation of RNA counts in single cells. Since the initial observation of bursting in fixed cells described above, the RNA FISH technique has been applied to a wide variety of genes, organisms, and cell types, demonstrating that transcriptional bursting is an apparently universal property of transcription in higher eukaryotes, including in mammalian tissues[169,171,172].

Direct confirmation of this transcriptional bursting phenomenon was enabled by the development of the MS2 stem loop system to label nascent RNA in live cells [173,174]. This approach uses overexpression of a bacteriophage coat protein fused to a fluorescent protein such as GFP in cells or animals; this overexpressed protein binds the stem-loop RNA structure originally found in the bacteriophage, so that knocking 12 or 24 repeats of this stem-loop sequence into an RNA of interest creates an intense fluorescent signal wherever this RNA species is present in a cell. This MS2 approach revealed that transcription was discontinuous through time in *Drosophila* embryos and mammalian cells, as well as in primary mouse cells[172,175]. Moreover, by applying stimuli to cells and measuring nascent transcription with either live or fixed methods, it became clear that properties of these transcriptional bursts could be regulated[176,177].

Since the studies mentioned above directly measured RNA production but were unable to visualize protein regulators of this process, mathematical modeling has been applied extensively to such datasets to try to infer the regulation of transcriptional bursting. The most commonly used model, which fits the data well in many cases, is the so-called random telegraph model[163,178,179] : this model suggests that two rates regulate a gene's transcriptional output, the rate of 'burst initiation' which controls the frequency of transcriptional bursts, and the rate of 'RNA production', perhaps related to polymerase initiation, which controls the number of RNAs produced per burst, or the 'burst intensity'. Some live-imaging studies have found evidence for a third regulated rate, upstream of burst initiation, which can produce a so-called refractory period, meaning that there

tends to be a period of transcriptional silence between bursts[180,181]. Other studies have suggested yet more complex regulatory architectures, such as one which suggested there are many (more than 5) major regulated rates governing transcriptional bursting[182]. However, it is still unclear how these mathematically inferred regulated steps relate to actual molecular regulators of transcription, such as those described in previous sections.

Most promisingly, in recent years studies have started to apply super-resolution imaging approaches to visualize transcription in tandem with proteins that regulate transcription to clarify the regulation of the dynamics of transcription[37,183]. For example, the Cisse group has devised an approach to visualize Pol II molecules simultaneously with live imaging of a gene of interest. This approach revealed that Pol II clusters around a promoter just before a burst of transcription starts[38]. This clustering behavior has been used to support a model that Pol II is undergoing condensation or phase separation at the promoter to drive transcription[18,184] , however it is difficult to directly assay the physical properties of Pol II in such a system, or to examine the necessity of this type of condensation to drive transcription. In another multicolor live imaging approach, the Gregor group established a system to enable enhancer and promoter conformation imaging along with live imaging of transcription[85] (discussed further in the Discussion). In future, it will be valuable to visualize more protein players in such a system, perhaps including p-TEFb or the polymerase initiation complex components, to learn more about the dynamics of transcription. Perhaps such a system could be paired with drug-induced degradation approaches[136,137] to enable causal dissection of the events taking place during transcription in live cells. Moreover, a consistent difficulty in such studies has been the cumbersome genetic engineering required to allow both single protein molecules and nascent RNA to be visualized. CRISPR has helped ameliorate both these difficulties somewhat by making it easier to knock in stem-loops or tag sequences, however still only a limited number of genes and cell systems has been used for live imaging, potentially limiting the generalizability of the findings above. In future, it is likely that imaging techniques using Cas9 or other bacterial Cas proteins to bind RNA and DNA may be optimized, allowing wider application of such techniques[185,186].

1.8: Perspective on the field of transcriptional regulation

Transcriptional regulation is one of the key processes that allow cells to change or to maintain their functional state. A number of themes have emerged from the sixty or more years of study devoted to this topic. First, transcriptional regulation in mammals is complex: many molecular players and steps are required for transcription of any given gene. Moreover, simple and elegant rules do not always generalize easily between different genes and contexts, as demonstrated for example by the ongoing difficulties in computational prediction of enhancer function. In spite of this complexity, some general patterns have emerged, such as the requirement for PIC assembly and promoter-proximal pause release for RNA production to occur, as well as the generality of the transcriptional bursting phenomenon. Relatedly, the field has benefited from the development and application of multiple orthogonal experimental techniques to measure transcription and transcriptional regulation: for example, the development of RNA imaging techniques revealed transcriptional bursting behavior that was invisible by bulk biochemical techniques. High-resolution imaging allowing tracking of multiple protein, DNA and RNA components in live cells are becoming more accessible and hopefully will help deepen understanding of the dynamics and spatial characteristics of transcription in single cells [38,187,188]. Such studies should be coupled with targeted perturbations, to allow dissection of causative relationships in transcription. A general technical challenge in this field (as well as many other biological questions) is the candidate-based nature of protein measurements: for example, ChIP-based studies require antibodies, while imaging approaches require engineering of specific proteins. Perhaps in future, more sensitive unbiased approaches such as locus pulldown[189] single cell mass spectrometry[190], or antibody-free binding protein design[191] might allow unbiased identification and functional dissection of proteins involved in the different steps of transcriptional regulation. More conceptually, the field has now identified many of the players in transcriptional regulation: the next step, difficult so far, is to integrate these pieces of information to predict

transcriptional outcomes in a cell type or organism[28,93,192,193]. It is not clear how the newest concepts in the field which generate the most excitement, such as superenhancers or phase separation, will help us advance toward this goal of prediction. Rather, such topics of research seem to propose alternate mechanisms for transcriptional behaviors for which we already have explanations (incomplete though these explanations may be). It will be exciting to watch the field hopefully develop better predictive models for how the complexities of eukaryotic development are orchestrated in time and space.

CHAPTER 2: Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping

2.1 Introduction

Transcription of most genes in mammalian cells occurs in bursts interspersed with refractory periods of varying lengths, both in cell lines and in tissues [163,164,167,171,173,181,183,194,195]. Transcription of a given gene can be increased through modulation of several different burst parameters. For instance, the gene could increase in burst fraction (the number of alleles transcribing per cell, which is related to how frequently bursts occur), or burst size (the number of RNA molecules produced per burst). For example, serum-induced transcription activation of β-*actin* augments both burst frequency and burst size, while serum induction of *c-Fos* leads to an increase in burst frequency without changing burst size[176,177]. Recent evidence further suggests that maintenance of mRNA concentrations in response to changes in cell size are achieved through changes in burst size, while gene dosage compensation in the cell cycle is regulated by burst frequencies[196]. However, the molecular regulation of transcriptional bursts remains unclear.

Enhancers are distal genetic elements that regulate transcription of their target genes by diverse modes, which include altering chromatin structure and histone modifications, recruiting basal transcription factors, localizing genes to permissive nuclear compartments, or increasing transcriptional elongation[8,66]. However, most studies on enhancer function are population based, leaving open how enhancers affect transcription dynamics at individual alleles.

The β-globin locus control region (LCR) is a powerful distal enhancer that is required for high level transcription of all β-type globin genes[197–199]. The LCR engages the embryonic (ε-globin), fetal (γ-globin), and adult type (β-globin) globin genes through looped contacts in a developmentally appropriate manner[76,77,200]. Proposed mechanisms by which the LCR activates transcription include recruiting tissue-specific and general transcription factors as well as RNA polymerase II, moving the locus towards the center of the nucleus and outside the chromosome territory, as well as promoting transcription elongation[72,113,198]. How these mechanisms impact transcriptional burst size

versus fraction is unresolved. A related open question is how the LCR controls bursting in the context of multiple active or potentially active β-type globin genes within the locus.

A landmark study using RNA FISH demonstrated that at a fraction of alleles γ-globin and β-globin located on the same chromosome can be transcribed during the same developmental stage[114]. This study further provided evidence that the appearance of allelic co-expression of γ-globin and β-globin results from rapidly alternating interactions of these genes with the LCR. Subsequent studies that included kinetic experiments provided additional support for the idea that LCR-promoter contacts are so dynamic that the appearance of allelic co-transcription is actually the result of very rapid transcriptional switches[115,201,202]. A model in which only one promoter interacts with the LCR at any given time would also explain competition among the genes. However, none of these studies excluded the possibility that the LCR acts on both genes simultaneously. Besides simultaneous LCR-gene contacts, an additional scenario compatible with allelic co-expression is that one gene could be transcribed dependently and the other independently of the LCR. Indeed, the β-type globin genes are transcribed in the absence of the LCR albeit at much lower levels[197]. The application of quantitative RNA FISH to examine transcriptional bursting parameters might allow discrimination between these models.

Here, we used quantitative single-molecule RNA FISH[170,203] to measure transcriptional burst size and burst fraction of the β-globin gene during erythroid maturation. We observed increases in both transcriptional burst fraction and size during this process. The enhancing effects of the LCR on β-globin transcription are predominantly explained by augmenting burst fraction with a modest but significant contribution to burst size, as revealed by LCR deletion experiments. To study the relationship between LCR-promoter contact frequencies on bursting parameters we applied a forced chromatin looping system. Specifically, forced tethering of the nuclear factor Ldb1 via designer zinc fingers to a chosen β-type globin gene promoter leads to recruitment of the LCR and transcription activation[79,112]. We found that in murine erythroblasts, engagement of the LCR increased β-globin burst frequency but not burst size. Similarly, in primary adult human erythroid

cells, redirecting the LCR towards the γ-globin genes and away from β-globin genes increased γ-globin burst fraction while lowering β-globin burst fraction. A significant proportion of alleles co-transcribed β- and γ-globin, but allelic co-transcription of β- and γ-globin was statistically disfavored. Strikingly, the burst size of co-transcribed genes is not reduced compared to singly transcribing genes, further supporting the model that promoters compete for LCR activity and that LCR-promoter looping controls burst fraction. In concert, these results provide new insights into mechanisms of enhancer function and highlight the use of targeted alterations of chromatin architecture for functional studies of enhancers.

2.2 β-globin transcriptional burst fraction and burst size increase during erythroid maturation

An increase in transcription at a given allele can occur through elevated burst fraction, burst size or a combination of both (Figure 2.1A). We first measured β-globin transcriptional bursting parameters in the G1E-ER4 murine erythroblast cell line during the course of erythroid maturation, when β-globin transcription is strongly induced. Maturation of these cells is dependent on the hematopoietic transcription factor GATA1, which is expressed in these cells as an estrogen receptor fusion protein. Addition of estradiol triggers erythroid maturation and increases contacts between the LCR and the β-globin promoter, faithfully reproducing normal terminal erythroid maturation[49,50,52]. We measured transcription of the adult β-globin (*Hbb-b1*) gene at different time points during G1E-ER4 maturation using single-molecule RNA FISH[170,203]. Transcription sites were identified by nuclear colocalization of spectrally distinguishable signals from probes specific for β-globin introns and exons (Figure 2.1B)[168].

We verified that transcription sites faithfully reported recent transcriptional events by blocking transcription with actinomycin D and counting β-globin transcription sites in human erythroid cells (Figure A2.1A). This control showed that 90% of transcription sites were lost at 30 minutes post-drug treatment, and half disappeared by just 10 minutes, validating that our transcription site identification faithfully reported ongoing transcription.

Both burst fraction and average burst size increased during differentiation (Figure 2.1B-E, Figure A2.6), indicating that both parameters contribute to total β-globin production during cell maturation. Specifically, the increase in burst fraction from 4 to 24 hours of maturation was 3.4 fold, while the change in burst size from 4 to 24 hours was 1.8 fold (Figure 2.1D-E). The burst fraction increase corresponded to an increase in the proportion of cells transcribing 1,2, or 3 active β-globin alleles (Figure A2.1C). The overall increase in transcriptional output was consistent with the profile of RNA polymerase II (Pol II) binding at the β-globin locus, as determined by chromatin immunoprecipitation (ChIP) (Figure 2.1F). Moreover, we observed an increase in the fraction of Pol II in the gene body relative to the promoter (Figure 2.1G), reflecting an increase in elongating Pol II, consistent with the observed increases in burst fraction and size. To address possible variabilities in fluorescence intensities between slides and experiments we concurrently measured cyclin A2 mRNA in the experiments shown in 1B-1E (Figure A2.1B). We found that cyclin A2 mRNA intensities were consistent between experiments. The subtle downward trend in cyclin A2 signal during erythroid maturation is opposite to that of the β-globin transcription sites (Figure 2.1E), such that normalization of β -globin bursting size to cyclin A2 would result in slightly larger changes during maturation.

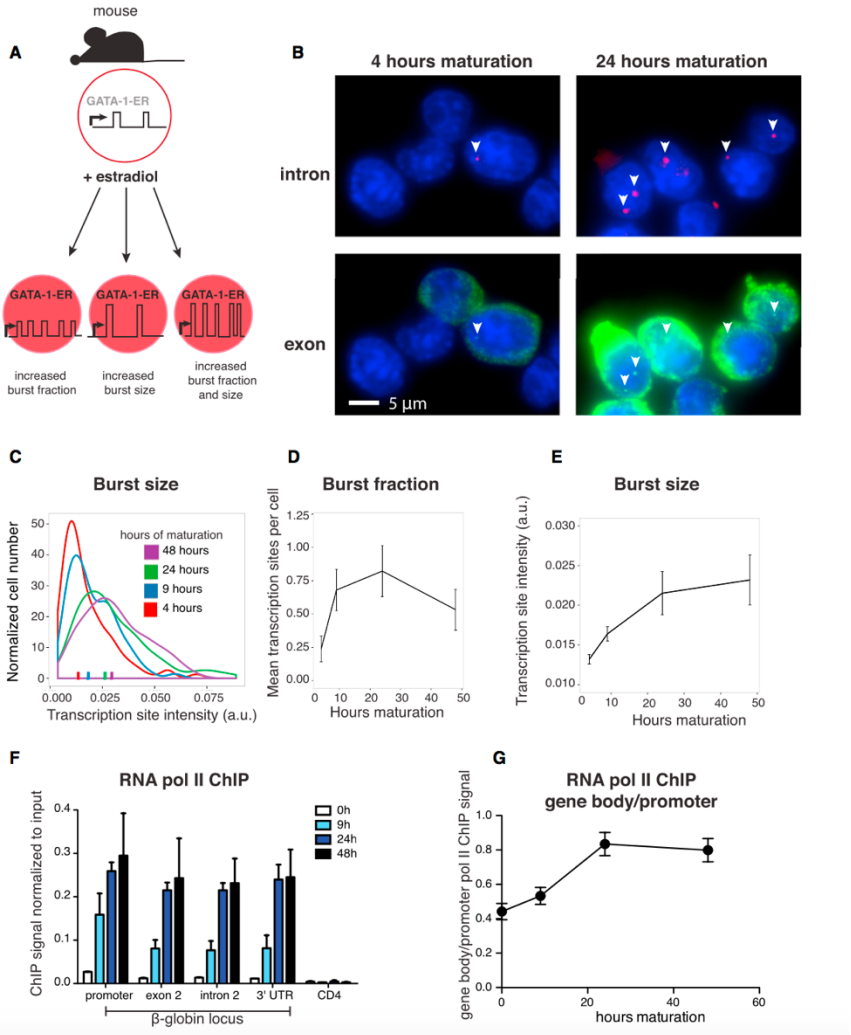


Figure 2. 1: β-globin transcriptional burst fraction and burst size increase during erythroid maturation. A) Simplified model of how bursts can be modulated during murine erythroid maturation to increase β-globin mRNA levels. B) Single-molecule RNA FISH using β-globin intron and exon probes to identify transcription sites in G1E-ER4 cells, 4 or 24 hours after estradiol addition (white arrows indicate transcription sites, intron and exon channels are set to same intensities across images). C) Representative experiment showing fluorescence intensities of β-globin transcription sites in G1E-ER4 cells at indicated time points after estradiol addition (tics on x axis=medians for this experiment). D) Mean number of alleles transcribing β-globin per cell at time points after estradiol addition. N=3 biological replicates. E) Median fluorescence intensities of β-globin transcription sites in G1E-ER4 cells following estradiol addition. N=3 biological replicates. F) Anti-RNA polymerase II ChIP in G1E-ER4 cells at indicated times after estradiol addition in the globin locus normalized to input. Primer pairs targeting distinct regions of the gene are listed on the x-axis. The silent CD4 locus served as a negative control. N=3 biological replicates. G) Ratio of RNA polymerase II ChIP signal at β-globin exon 2 to β-globin promoter. N=3 biological replicates. Error bars represent SEM. See also Figures A2.1 and A2.6.

One question raised by this analysis is whether the variation in burst size might be strongly regulated by global transcriptional differences between cells ('extrinsic noise')[204], as opposed to allele-level differences ('intrinsic noise'). If, for example, transcription factor concentrations between cells varied widely, and if this variation were to determine differences in burst size, then two alleles bursting in the same cell should tend to have similar burst sizes. To quantify the contribution of extrinsic noise to burst size, we measured whether the sizes of bursts were correlated in cells that displayed two transcriptional bursts at various stages of maturation. We found that the correlation was not strong (maximum R^2=0.293) and therefore differences between cells do not strongly determine individual burst sizes as measured by RNA FISH (Figure A2.1D-F). The correlation was slightly weaker at 9 hours of maturation compared to later time points, suggesting that cell-to-cell variability may play a slightly larger role in determining burst size variation later in differentiation. We note however that staggered initiation timing of bursts, as well as technical noise, can contribute to apparent intrinsic noise in this system.

To test whether global changes during erythroid maturation might nonspecifically alter the measurements of bursting parameters, we examined transcription of *Gata2*, a gene that is active in erythroid precursors but silenced during maturation[205]. We found that the burst size and burst fraction of *Gata2* are both reduced during maturation, suggesting that both parameters are regulated in a gene-specific manner (Figure A2.2).

In sum, β-globin burst fraction and burst size increase during erythroid maturation. This finding raises the question of how these bursting parameters are governed, and whether burst fraction and burst size are governed independently of each other.

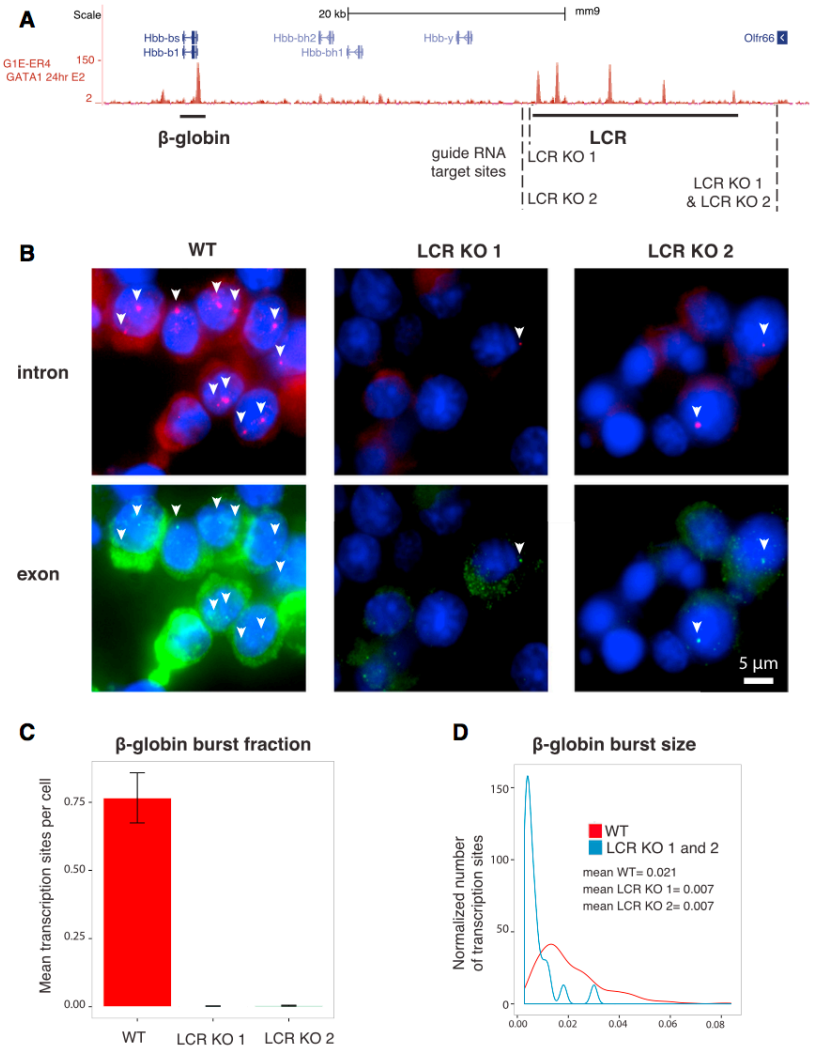


Figure 2. 2: Deletion of the LCR predominantly impairs β-globin burst fraction. A) GATA1 ChIP-Seq browser track view of the β-globin locus. Peaks demarcate the β-globin promoter and DNase I hypersensitive sites of the LCR. Guide RNA placements for Cas9-mediated LCR deletions are indicated. B) Single-molecule RNA FISH using β-globin intron and exon probes to identify transcription sites in wild-type and LCR-deleted G1E-ER4 cells 24 hours after estradiol addition (white arrows indicate transcription sites, intron channel is set to same intensities across images). C) Mean number of alleles transcribing β-globin per cell in G1E-ER4 wild-type and LCR-deleted cells, treated with estradiol for 24 hours. N=3 biological replicates. D) Fluorescence intensity of β-globin transcription sites in G1E-ER4 parental and LCR-deleted cells, exposed to estradiol for 24 hours. Wilcoxon non-parametric t-test, pooling the two LCR-edited clones, p=1.54e-10. Data from 3 biological replicates pooled. Error bars represent SEM. See also Figures A2.2 and A2.7.

2.3 Enhancer deletion by genome editing decreases β-globin burst fraction and size

Relatively little is known about the molecular control of bursting parameters. In particular, it is unknown how distal regulatory elements influence burst fraction and burst size. Previous studies that the LCR influences both burst fraction and size[113]. To quantify the contributions of the LCR to these parameters via single molecule RNA FISH, we generated two distinct G1E-ER4 cell lines with homozygous deletions of the LCR using different guide RNA pairs (Figure 2.2A, Figure A2.3A-G). Cells were induced with estradiol for 24 hours to examine the effect of the LCR in its most active state. The β-globin burst fraction LCR deleted cells is greatly reduced (311-fold and 194-fold in line 1 and line 2, respectively, Figure 2.2B-C, Figure A2.7) when compared to control cells. For measurements of burst size we had to examine a large number of cells (8216 cells total between all conditions and replicates) given the strong reduction in burst frequencies in cells lacking the LCR. Moreover, we pooled across biological replicates and both clones in order to gain the number of transcription sites necessary to make a statistical comparison of burst size. Burst size was reduced approximately 3-fold in cells lacking the LCR- compared to parental G1E-ER4 cells (3-fold for each clone, Figure 2.2B and 2.2D). We note that we measured an equal number of transcription sites from each clone, and the means of both populations were the similar, justifying the pooling the two clones in the final analysis. These results suggest that the LCR predominantly controls burst fraction with a lesser but significant impact on burst size.

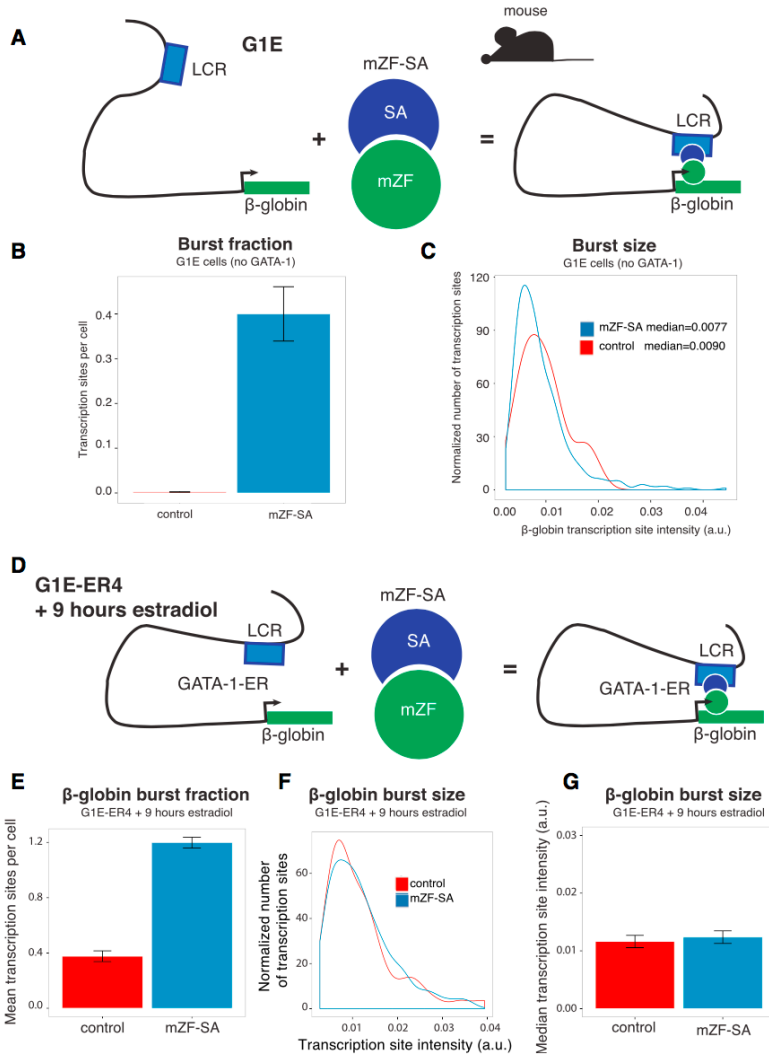


Figure 2. 3 Forcing LCR-promoter contacts increases β-globin burst fraction independent of burst size. A) Schematic of the forced looping strategy in murine G1E cells lacking GATA1. B) Mean number of alleles transcribing β-globin per cell infected with empty vector or vector expressing mZF-SA. N=3 biological replicates. C) Fluorescence intensity of β-globin transcription sites. Wilcoxon non-parametric test comparing control and mZF-SA intensities, p=0.42. Data were pooled from 3 biological replicates. D) Schematic of the forced looping strategy in murine G1E-ER4 cells induced for 9 hours with estradiol. E) Mean number of alleles transcribing β-globin per cell in G1E-ER4 cells infected either with control vector or mZF-SA, 9 hours after estradiol addition. N=3 biological replicates. F) Representative experiment showing fluorescence intensity of β-globin transcription sites in G1E-ER4 cells, infected either with control vector or vector expressing mZF-SA, 9 hours after estradiol addition. G) Median fluorescence intensity of β-globin transcription sites in G1E-ER4 cells infected either with control vector or mZF-SA, 9 hours after estradiol addition. N=3 biological replicates. Error bars represent SEM. See also Figures A2.3 and A2.8.

2.4 Forced enhancer-promoter contacts increase β-globin transcriptional burst fraction but not burst size

The LCR forms looped contacts with the promoters it regulates. Since the LCR influences both burst fraction and size, we set out to test whether these two effects are both controlled by LCR-promoter contact frequencies or whether they are mechanistically separable. We used a recently developed strategy to force LCR-promoter contacts that involves tethering of the self-association domain (SA) of nuclear factor Ldb1 by a designer zinc finger (ZF) to the β-globin gene promoter, which leads to recruitment of the LCR and transcription activation (Figure 2.3A)[79]. The unique advantage of this system is that it allows targeted formation of enhancer-promoter contacts without otherwise perturbing the locus or cell state. G1E cells lack GATA1 and as a result have few if any LCR-promoter contacts. G1E cells expressing a ZF-SA fusion protein specific for the murine β-globin promoter (mZF-SA) exhibit LCR-promoter contact frequencies approximating those of GATA1 replete cells. We infected G1E cells with a control retroviral vector or a derivative expressing mZF-SA as part of an IRES-GFP construct, followed by fluorescence-activated cell sorting (FACS) to enrich for infected cell populations. Increasing LCR-promoter contacts by mZF-SA significantly increased β-globin burst fraction compared to controls (Figure 2.3B, Figure A2.8). Surprisingly, burst size was unchanged by mZF-SA expression (Figure 2.3C).

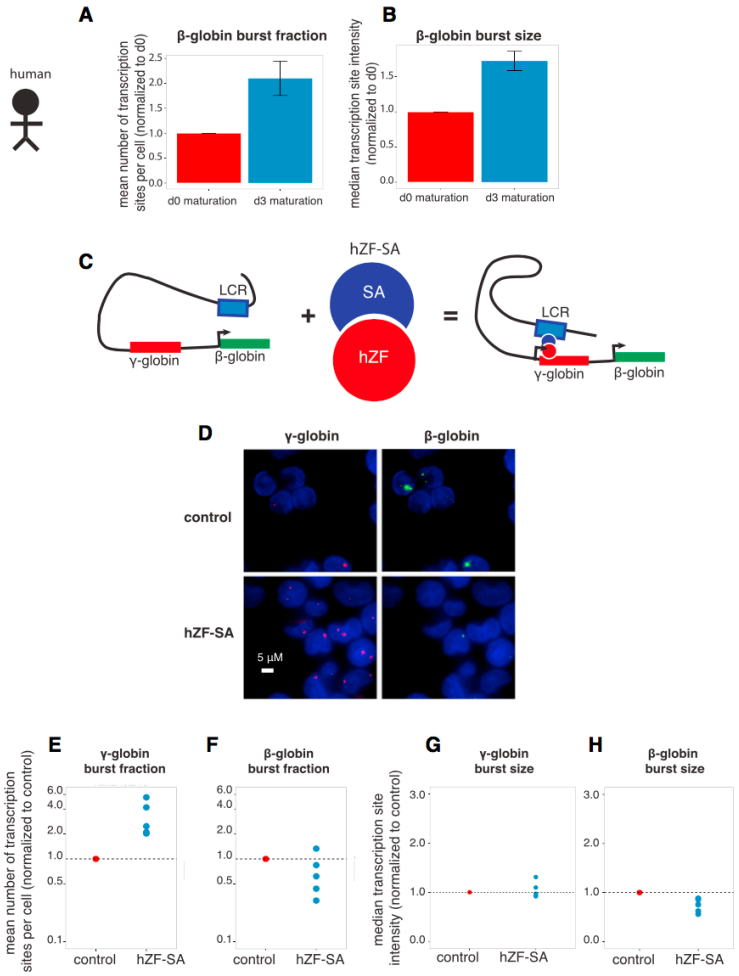


Figure 2. 4 Effects of maturation and forced LCR-promoter contacts on transcriptional bursting parameters in human erythroid cells. A) Mean number of alleles transcribing β-globin per cell, and B) median fluorescence intensity of β-globin transcription sites at 0 or 3 days of maturation induction. N=3 human donors. C) Schematic of forced LCR-γ-globin promoter loop strategy in primary human erythroid cells. D) Single-molecule RNA FISH of γ- and β-globin in human primary erythroid cells infected with hZF-SA expressing vector or control vector. E) Mean number of alleles transcribing γ-globin per cell in hZF-SA expressing cells or control cells, normalized to proportion transcribing in the control vector condition. Wilcoxon paired t-test, p=0.016. N=5 human donors. F) Mean number of alleles transcribing β-globin per cell normalized to control vector condition. Wilcoxon paired t-test, p=0.42, or p=0.2 excluding outlier. N=5 human donors. G) Median fluorescence intensities of -globin transcription sites cells infected with control vector or hZF-SA vector, normalized to proportion transcribing in the control vector condition. Wilcoxon paired t-test, p=0.84. N=5 human donors. H) Median fluorescence intensities of β-globin transcription sites in cells infected with control vector or hZF-SA vector, normalized to proportion transcribing in the control vector condition. Wilcoxon paired t-aest, p=0.09. N=5 human donors. Error bars represent SEM. See also Figures A2.4 and A2.9.

Since the LCR contains numerous GATA1 binding sites, its activity in G1E cells is compromised. Therefore, we carried out similar experiments under conditions in which GATA1 activity was partially restored. We retrovirally introduced mZF-SA and control constructs into G1E-ER4 cells and exposed them to estradiol for 9 hours, which results in partial differentiation and activation of β-globin transcription (Figure 2.3D). Similar to in the G1E system, mZF-SA increased β-globin burst fraction, but did not alter burst size (Figure 2.3E-G). β-globin transcription was confirmed to be increased by qPCR (Figure A2.4A). Together these results suggest that enhancer-promoter contacts augment the likelihood for a burst to occur, but that they do not determine burst size. The use of the mZF-SA reagent thus enabled the uncoupling of enhancer looping from other potential enhancer functions that might impact on the β-globin locus.

2.5 β-globin transcriptional burst fraction and size increase during human erythroid maturation

In order to investigate whether the regulation of β-globin transcriptional bursting during erythroid maturation is conserved between mice and humans, we examined primary human erythroblasts. Human CD34+ hematopoietic precursors from peripheral blood were expanded and differentiated towards the erythroid lineage using a previously described two-phase liquid culture system[206]. We measured adult type β-globin (*HBB*) burst size and fraction before and after 3 days of erythroid differentiation induction. During this process we observed an increase in both burst size and fraction of the β-globin gene, similar to our observations in maturing murine erythroid cells (Figure 2.4A-B). The increases in burst fraction (1.7 fold) and average burst size (1.6 fold) are comparable in magnitude and thus are both likely to contribute to increased β-globin production during erythroid maturation.

2.6 Forced switching of enhancer contacts between two promoters reciprocally alters burst fraction

To examine whether the human LCR regulates bursting in the same manner as the murine LCR, we employed a forced looping strategy (Figure 2.4C)[112]. The human β-globin locus spans several globin genes, including the adult type β-globin and a fetal form called γ-globin (*HBG*). We reported previously that expression of a ZF-SA construct targeting the γ-globin promoter (hZF-SA) in human primary erythroid cells increased LCR contacts with the γ-globin gene at the expense of the β-globin gene with corresponding changes in transcription, consistent with a mechanism in which these genes compete for LCR function. An advantageous feature of the human system is that perturbations of LCR looping leave unaltered endogenous nuclear factors, the LCR per se, and the maturation state of the cells.

We infected primary human cultures with lentiviral vectors expressing either GFP (control) or hZF-SA, purified infected cells by FACS, and quantified bursting parameters by RNA FISH. We observed that erythroblasts from different donors expressing control vector produced between 5 and 35% γ-globin and 65-95% β-globin. The variation between samples is presumably due to differences among donors or the way blood samples were obtained and processed. Forced enhancer-promoter looping between the LCR and γ-globin promoter increased the burst fraction of γ-globin transcription sites per cell, and decreased the burst fraction and size of β-globin sites slightly (Figures 2.4D-H, Figure A2.9), suggests the two genes compete for LCR activity. Furthermore, formation of γ-globin-LCR contacts increased γ-globin burst fraction without affecting γ-globin burst size, consistent with the results from the murine cell line (Figures 2.4E-H). Changes in burst fraction corresponded to changes in mRNA levels in the same samples: γ-globin mRNA increases with hZF-SA, while β-globin trends downward (Figure A2.4B-C). The human LCR thus governs burst fraction of target genes via enhancer-promoter contact, while governing burst size through another independent mechanism.

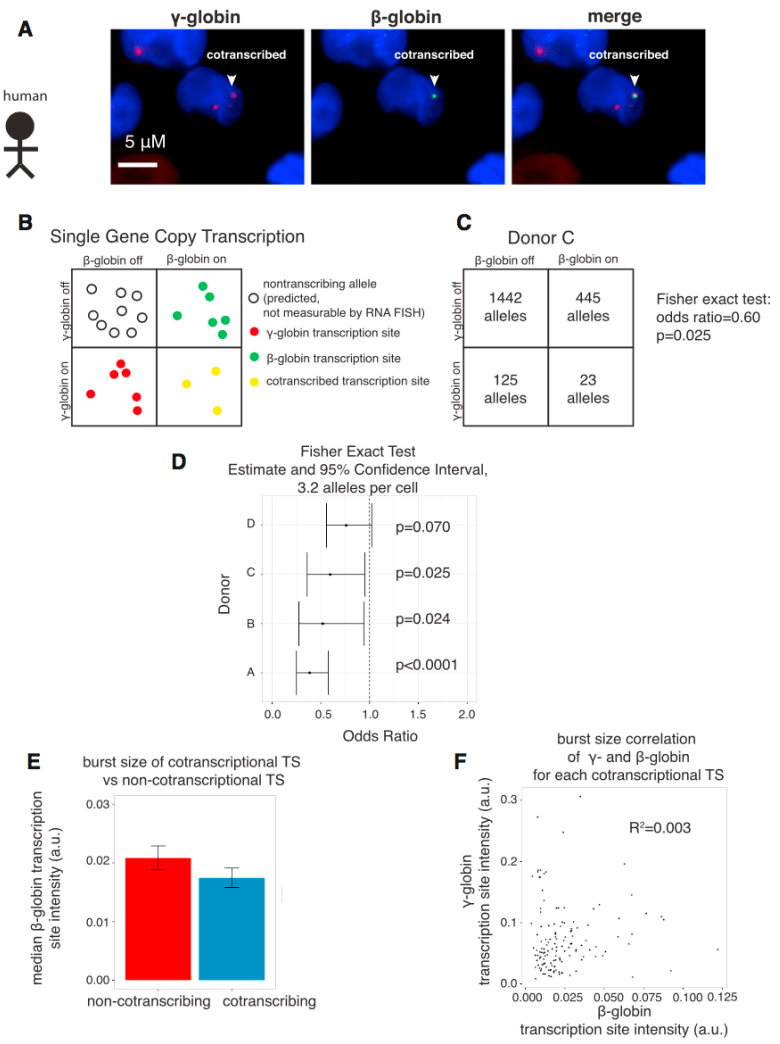


Figure 2. 5: β- and γ-globin promoters compete for enhancer activity when positioned in cis. A) Representative image of a primary human erythroid cell co-transcribing β- and γ-globin from the same allele. B) Model of how allelic β- and γ-globin transcription was quantitated using RNA FISH data. C) Example quantification of cellular β- and γ-globin transcription for one donor. Fisher exact test odds ratio=0.66, p-value=0.13. D) Fisher exact test odds ratio and 95% confidence interval for 4 human samples for cis-competition of globin alleles, estimating 3.2 alleles per cell. E) Fluorescence intensity of β-globin transcription sites in primary human cells, either non-colocalized transcription sites or those colocalized with γ-globin transcription sites. p=0.4, Wilcoxon paired rank-sum t-test, N=4 donors. F) Fluorescence intensity of β- and γ-globin transcription sites for each co-transcriptional transcription site, Pearson R^2=0.003, N=4 donors. See also Figures A2.5 and A2.10.

2.7 Allelic but not cellular co-transcription is disfavored in primary human erythroblasts

We observed co-transcription of β- and γ-globin on the same chromosome in a fraction of cells, as shown by colocalized γ-globin and β-globin intron RNA FISH signals, suggesting that competition between different globin genes might be not absolute (Figure 2.5A). Similar co-transcription of globin genes had been previously observed in murine systems[114,115,201,202]. Allelic co-transcription could mean (1) that the LCR switches rapidly between the β-globin and γ-globin genes to alternately trigger bursts, (2) that the LCR acts on both genes simultaneously, or that (3) occasional transcription of one or both genes occurs in an enhancer-independent manner. If the LCR truly switched rapidly (Model 1), co-transcriptional events should be less frequent than expected than if the two genes were transcribing independently.

To test whether co-transcription is indeed statistically disfavored, we first examined whether β- and γ-globin are transcribed independently in the same cell, regardless of whether the two genes are transcribed from the same chromosome or not. We quantified cells transcribing β-globin, γ-globin, both or neither, and then performed a Fisher exact test for independence of their transcription (Figure A2.5A-C). The null hypothesis is that a cell transcribing β-globin has the same likelihood of transcribing γ-globin as a cell NOT transcribing β-globin. This is represented by an odds ratio of one. Alternatively, an odds ratio less than one would show negative correlation between transcription of the two genes in a single cell, potentially suggesting some trans regulation mediating competition between the genes. We found that the odds ratio was not consistently greater or less than 1 in our analysis of the cells of 4 donors (Figure A2.5C). This suggests that on a cellular level, transcription of β- and γ-globin is independent.

Next, we investigated whether co-transcription on a single allele copy of γ- and β-globin was independent, treating all alleles as a pool irrespective of which cell they were in. This analysis required estimating the number of non-transcribing alleles in the population, which is not measurable by RNA FISH. We thus performed cell cycle analysis of cultured human cells using RNA FISH: cells with histone 4 mRNA are in S-phase, cells without histone 4 mRNA with a small

nucleus are in G1, while cells without histone RNA with a larger nucleus are in G2 (Padovan-Merhar et al., 2015). Using this method, we found that 40% of cells were in G1, 31% in S, and 29% in G2 (Figure A2.5D). Since the β-globin locus is early-replicating in erythroid cells[207] we conservatively assumed that each S and G2 cell had 4 copies, resulting in an average of 3.2 copies per cell (40% of 2 alleles + 60% of 4 alleles =3.2 alleles). However, we also carried out calculations with different possible numbers of alleles: 2.6 alleles (which assumes that globin is replicated at the end of S phase) and 3 alleles (if globin is replicated near the middle of S) (Figure A2.5E-F).

Using this estimate for allele number, we quantified allelic transcription in cells from four human donors. Cotranscription in cis represented 1.1% to 3.0% of total globin locus copies depending on the donor (example shown in Figure 2.5C). We then measured the relationship between transcription of β- and γ-globin in cis using the Fisher exact test (Figure 2.5B-D, Figure A2.10). The results suggest that allelic co-transcription of β- and γ-globin is mutually inhibitory, with odds ratios ranging from 0.39 to 0.76 depending on the donor (Figure 2.5B-D). Note that this was the most conservative measurement of the odds ratio of allelic β- and γ-globin co-transcription. If the globin locus were not replicated at the very beginning of S phase in all cells, then the average number of alleles per cell would be less than 3.2. Estimating either 3.0 or 2.6 globin alleles per cell on average reduces the odds ratios for each donor, and decreases all p-values below p=0.05 (Figure A2.5E-F). This analysis therefore supports the model that the LCR rapidly switches between promoters to allow co-transcriptional events to occur. In contrast, if the LCR could act on both genes at once or if one gene could transcribe without LCR contact (Models 2 and 3), we would expect to see cotranscription occurring at a higher rate.

Strikingly, we found that the average burst sizes of genes that are co-transcribing are not significantly different from those genes that are not, consistent with the model of dynamic enhancer switching (model 1) and the finding above that changing enhancer-promoter contacts does not alter burst size (Figure 2.5E).

If the LCR switches dynamically between genes (model 1) a further prediction would be that γ-globin burst sizes would not be correlated with β-globin burst sizes because bursts detected

by FISH appear largest in the middle of the bursting interval and dimmer during the beginning and end of the burst. In contrast, model 2 (simultaneous contacts of the LCR with both globin genes) would predict a positive correlation in the burst sizes since LCR activity on a target gene increases burst size. We observed a lack of correlation between allelic γ- and β-globin burst sizes, consistent with transcription of the two genes occurring in a staggered fashion and lending further support for model 1 (Figure 2.5F).

In sum, all our results agree with a model in which the LCR interactions with the □-globin gene disfavor those with the β-globin gene and vice versa, accounting for competition between the two genes. Definitive evidence for this model would require direct imaging of LCR-promoter contacts in live cells, which is not yet possible.

2.8 Discussion

We employed single-molecule RNA FISH to quantify transcriptional bursting parameters at the murine and human β-globin loci during cellular differentiation and following targeted changes to LCR-promoter contacts. During erythroid maturation both β-globin burst fraction and burst size increase. Deletion of the LCR dramatically reduces burst fraction with comparatively modest effects on burst size. In order to selectively interrogate the influence of LCR-promoter contacts on bursting parameters, looped contacts were forged via targeted tethering of the self association domain of Ldb1. Forced enhancer-promoter contacts increased burst fraction without affecting burst size. In concert, these findings suggest a connection between enhancer-promoter contact frequencies and burst fractions. The γ-globin and β-globin genes positioned in cis can be co-expressed but compete for LCR contacts and tend not to be initiated simultaneously, consistent with rapidly alternating and highly flexible looped enhancer promoter contacts.

The finding that the LCR influences both the increase in burst fraction and size is consistent with a previous report[113]. However, single molecule RNA FISH experiments in the present study enabled a more quantitative assessment of burst parameters, revealing that augmenting burst

fraction is the predominant mechanism by which the LCR enhances β-globin transcription. To bypass potential secondary effects of LCR deletion and to examine the consequences specifically of LCR-promoter contact frequency on bursting parameters, we employed a forced LCR-promoter looping approach. Notably, the looping inducing construct ZF-SA triggered increases burst frequency to the same extent as did GATA1-induced maturation but without modulating burst size. We saw essentially the same behavior in primary human cells under conditions in which the LCR associated protein complexes and cell maturation stage are presumed to be largely unperturbed. This confirms that the major regulatory function of the LCR is to increase the likelihood of the β-globin genes to be actively transcribed. These conclusions are in line with observations made with live cell imaging, which showed that increasing levels of transcription factor binding to a reporter construct predominantly increased burst fraction with little effect on burst size or duration[208].

These data beg the question of how exactly enhancer looping governs burst occurrence. Is enhancer-promoter contact required only to initiate a transcriptional burst or does it have to persist for the duration of the burst? Solving this issue definitively would require both live imaging of transcription as well as measuring the duration of enhancer-promoter contacts in live cells, the latter of which is technically difficult. The observation that cotranscriptional spots of γ-globin and β-globin only persist for a few minutes after cessation of transcription, suggests that enhancer-promoter contacts and the resulting transcriptional burst are indeed very dynamic on a scale of minutes in the case of this locus. We acknowledge that temporal inferences on LCR promoter-contact oscillations and burst lengths are limited by the half-life of the intron used to image transcription. Therefore, it remains possible that either the contacts durations and/or the burst lengths are even shorter than our measurements allow.

It remains unclear how regulatory elements such as the LCR contribute to burst size during cell maturation, although our data suggests that the mechanism is independent of enhancer-promoter contact. One possibility is that increases in transcription factor occupancy at the locus or progressive changes in histone modifications that accompany maturation promote RNA polymerase density or processivity[209–211]. LCR deletion may block such maturation-induced

changes: LCR deletion in G1E-ER4 cells led to reduced occupancy at the promoter by GATA1 (unpublished observation). However, loss of the LCR in whole animals left chromatin accessibility, histone acetylation and transcription factor occupancy at the promoters largely intact[52,80,198,199]. The LCR may also contribute to burst size by stimulating transcription elongation via physical contacts with the gene body[212–214]. Our forced looping approach predominantly facilitates enhancer contacts with the promoter, while increasing gene body contacts less than GATA1 induced maturation does[79]. An increase in elongation rate could allow more polymerase molecules to traverse the gene body in quick succession, thus increasing burst size. Therefore, decreased GATA1 occupancy, reduced LCR-gene-body contacts, or related consequences of LCR deletion likely account for the reduction in transcriptional burst size. However, we also acknowledge inherent technical limitations to the ability to distinguish burst size and fractions under conditions in which transcription is extremely rare, such as in the LCR-deficient cells.

Co-expression in cis of γ-and β-globin genes could result from simultaneous contacts of the LCR with both genes; or from rapidly alternating contacts of the LCR to the two genes such that one interaction disfavors the other. Our results clearly support the latter model since active transcription of one gene lowers the probability of transcription of the other gene. Additional support for alternating LCR-gene contacts derives from the lack of correlation in burst sizes between γ- and β-globin genes co-transcribed in cis. Prior experiments involving pharmacologic transcription elongation blockage followed by release into productive elongation showed that reappearance of co-transcriptional FISH signals for globin genes in cis re-appeared slower than signals from individually transcribing genes[115].Together with experiments showing that the LCR promotes phosphorylation of RNA polymerase II and elongation[72,80], the simplest model accommodating prior and current findings is that the LCR forms looped contacts with the promoters to increase the probability of paused RNA polymerase to convert into the actively elongating form.

If two genes in cis compete for LCR function in a mutually exclusive manner no co-transcription should be detected, yet co-transcription appears at approximately 1.1% to 3.0% of alleles. However, RNA FISH experiments are limited in their temporal resolution by the half life of

the introns under investigation. In our experiments, duration of RNA FISH signals was measured to average ~10 minutes following transcription inhibition. Thus, any fluctuations in LCR-gene contacts at a time scale below 10 min would appear as co-transcription even if LCR-promoter contacts are entirely mutually exclusive. However, we cannot not rule out rare simultaneous LCR contacts or occasional enhancer-independent firing that might lead to truly concurrent transcription. Addressing such possibilities might be feasible once live imaging techniques have reached the required spatial and temporal resolution.

Future studies will examine the generality of enhancer effects on transcriptional bursting. The LCR is a well-studied example, but distinct modalities of enhancer actions are a likely possibility. Moreover, different drug treatments or transcription factor perturbations may alter bursting in other ways that may help understand the molecular underpinnings of transcriptional control.

2.9 Experimental Procedures

Murine cell culture, infection and sorting

G1E cells and G1E-ER4 cells were cultured and differentiated as described (Weiss 1997).

Cells were infected with the MIGR-1 retrovirus expressing only GFP or mZF-SA followed by an IRES element and GFP[79]. Infections were performed as described[47]. Cells were expanded for two days and sorted using a BD FacsAria to purify GFP+ infected cells from control and mZF-SA samples. Finally, estradiol was added for 9 hours and transcription was measured by FISH or qPCR.

Human primary cell culture, infection and sorting

Human peripheral blood mononuclear cells were obtained from de-identified healthy blood donors after informed consent by the University of Pennsylvania Stem Cell Core. CD34+ hematopoietic precursor cells were isolated from peripheral blood mononuclear cells using CD34+ magnetic bead positive selection. Cells were expanded in SFEM media (StemCell) with 10^{-6}M hydrocortisone

(Sigma Aldrich), 100 ng/ml SCF, 5ng/ml IL-3, 100ng/ml erythropoietin (Peprotech), protocol adapted from[206,215,216] . For FISH studies measuring co-transcription, cells were examined after 9-12 days of expansion.

For hZF-SA experiments, cells were infected after 8-11 days of expansion. hZF-SA was expressed in a lentivirus driven by the ankyrin promoter, and coupled to IRES-GFP. The control vector expressed only GFP[112]. Infections were carried out as described[112]. The expansion phase was extended for 7 more days, and then cells were sorted and examined.

For human erythroid maturation experiments, cells were expanded for 11 days after isolation, then cells were differentiated for 3 days in SFEM with 3U/ml erythropoietin without other cytokines such that d0 and d3 (as in Figure 2.4A-B) correspond to day 11 and d14 of total culture respectively.

Actinomycin D treatment was performed by adding 1 microgram/ml of actinomycin D to human cells for the stated period of time, after 8-11 days of expansion of cells.

Generation of enhancer-deleted G1E-ER4 cells

Cas9 and guide RNA plasmids were transiently co-transfected into G1E-ER4 cells using an Amaxa Nucleofector 2b (Lonza, program G-016, reagent kit R). Transfected single cells were sorted into a 96 well plate using a FACS Aria II (BD Biosciences). Single cell clones were expanded and screened by PCR, followed by DNA sequencing.

Chromatin Immunoprecipitation

We performed ChIP as previously described[217], using the N-20 Pol II antibody (Santa Cruz sc899). ChIP-qPCR was performed with Power SYBR Green (Invitrogen).

RT-qPCR

We isolated RNA using TRIzol (Life Technologies). Reverse transcription was performed with iScript (Bio-Rad). qPCR was performed with Power SYBR Green (Invitrogen).

Single-molecule RNA FISH imaging

We performed single-molecule RNA FISH as described previously[203]. Briefly, we fixed cells in 1.85% formaldehyde for 10min at room temperature, and stored them in 70% ethanol at 4 degrees C until imaging. FISH probes consisted of oligonucleotides conjugated to fluorescent dyes. We hybridized pools of FISH probes to samples, followed by DAPI staining and wash steps performed in suspension. Samples were cytospun onto slides for imaging on a Nikon Ti-E inverted fluorescence microscope.

Image Analysis

We manually segmented boundaries of cells from bright field images and localized RNA spots using custom software written in MATLAB, with subsequent analyses performed in R. Transcription sites for mouse β-globin, *Gata2*, and human γ-globin are identified by co-localization of spots in the intron and exon channels for a given mRNA, and for human β-globin by bright nuclear intron spots. Alleles co-transcribing human γ-globin and β-globin are identified by colocalization of transcription sites. Fluorescence intensities of transcription sites were determined by 2D Gaussian fitting on processed image data.

Plotting and graphics

We used R packages dplyr and ggplot2 to produce nearly all figures, followed by cosmetic adjustments in Adobe Illustrator. Several figures were produced using Graphpad Prism.

CHAPTER 3: Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation

3.1 Introduction

Since I showed that increasing enhancer-promoter contact increased burst fraction at the β-globin locus in human and mouse cells (Chapter 2), I next asked how generally important this mode of transcriptional regulation might be. To address this question, I examined multiple genes and perturbations to see whether bursting regulation or other types of transcriptional regulation were more prevalent.

Studies using sequencing-based biochemical assays have suggested that transcriptional regulation primarily tunes two transcriptional steps: RNA polymerase II (Pol II) binding to a gene's promoter and the release of Pol II from promoter-proximal pausing [1,5,25]. Both polymerase binding and polymerase pause release are required to produce RNA, and studies have extensively characterized the proteins involved in facilitating both these steps.

However, given the data that bursting properties are regulated in some contexts (Chapter 2), I asked whether the fluctuations between transcriptional activity and inactivity, which I will call burst initiation and termination, might be regulated along with the steps of polymerase binding and polymerase pause release. The transitions between the transcriptionally-active state and the transcriptionally-inactive state are typically slower than the time required for polymerase binding and polymerase pause release (hours vs. tens of minutes)[37,38,150,166,180,218] . Therefore, changes to burst initiation and burst termination may form an independent layer of regulation superimposed on top of polymerase binding and polymerase pause release: each time a burst is initiated, multiple RNAs can be made by repeated cycles of polymerase binding and release. If burst initiation and termination are regulated, it is unclear how those steps relate to the steps of polymerase binding and pause release: either burst initiation could permit polymerases to bind to facilitate transcription of multiple RNAs quickly during a burst, or burst initiation could instead facilitate transcription by permitting bound polymerases to release from pausing. It is also unclear

whether regulation of bursting, polymerase binding and pause release are all equally important to mediate changes in gene expression.

To identify which of these steps of transcription are tuned, we first need to determine how polymerase binding and pause release fit together with bursting; specifically, whether polymerase binding occurs in bursts or if polymerase pause release occurs in bursts. It has proven difficult to distinguish these possibilities in part because single cell and population-averaging biochemical measures of transcription give different types of information: only single cell methods such as nascent transcript RNA FISH can be used to measure bursting properties such as burst fraction and intensity, while biochemical methods like Pol II ChIP-seq have been used to measure polymerase binding and pause release. We hypothesized that combining these methods could clarify how bursts relate to polymerase binding and pause release, and could thereby allow us to identify which transcriptional steps are regulated by perturbations. Nascent transcript RNA FISH uses probes specific to introns of a gene of interest to measure transcription of that gene in single cells, thus measuring average changes in burst frequency and amplitude[163,170,219]. In contrast, Pol II ChIP-seq (and similar approaches like NET-seq and PRO-seq) measures average changes in polymerase occupancy and pause release[61,150,154,155,220,221]. Using both these types of methods to simultaneously measure changes in bursting properties as well as polymerase occupancy and pause release should allow us to clarify whether either polymerase binding occurs in bursts or polymerase pause release occurs in bursts.

To determine which steps of transcription are changed by biological perturbations, we first constructed several possible models of transcription that fit together bursting with polymerase binding and pause release steps, including one framework where burst initiation permits polymerase to bind (‘polymerase binding occurs in bursts’) and another where burst initiation permits release of bound polymerases (‘polymerase pause release occurs in bursts’). Each framework made distinct predictions for how nascent transcript RNA FISH and Pol II ChIP-seq measurements should change in response to perturbations. We then performed both nascent transcript RNA FISH and Pol II ChIP-seq in the presence of a number of global and gene-specific

perturbations, and compared the results to the models' predictions. This approach revealed that polymerase binding can only occur after a burst is initiated, and excluded the possibility that burst initiation permits pause release of pre-bound polymerase. We then used our experimental data with this 'polymerase binding occurs in bursts' model of transcription to show that the alteration of burst initiation and polymerase pause release rates accounted for the majority of changes in the overall rate of transcription. In contrast to what was previously expected[58,138], most perturbations that we examined seemed not to greatly change the rate of polymerase binding in our system. Our study supports a model of transcription in which polymerase binding occurs in bursts, and it implicates transcriptional burst initiation and polymerase pause release as critical control points in transcriptional regulation.

3.2 Identifying the regulated steps of transcription requires selection of a model of transcription

In order to ultimately identify which steps of transcription are regulated by complex biological perturbations, we first needed to clarify how transcriptional bursting regulation fits together with polymerase binding and pause release regulation. We built quantitative models of transcription, each with a different configuration of the processes of burst initiation, polymerase binding, and pause release (Figure 3.1A). In each, we used simulations to determine what the experimental outcomes would be for changing the rates of specific processes, with the hypothesis that changing different rates would lead to a distinct set of experimental changes, ultimately allowing us to determine which rates were changing in complex biological perturbations. A key feature required of our models was that transcription occur in a bursting pattern (given that all genes we have examined show this behavior, Figure A3.1A). In one model, this bursting occurs because polymerase binding only occurs during transcriptionally active periods ('polymerase binding occurs in bursts'). In the other model, polymerase binding could happen any time, irrespective of whether the gene is active or not, and instead it is pause release that only occurs during transcriptionally-active periods ('polymerase pause release occurs in bursts').

To determine the experimental predictions of these two models of transcriptional regulation, we used Gillespie stochastic simulations to model a thousand gene copies that over time proceeded through the steps in a model framework: burst initiation, polymerase binding, polymerase pause release, and burst termination[222]. We then used the molecular events occurring over time in each thousand-gene-copy simulation to predict the resulting Pol II ChIP-seq and nascent transcript RNA FISH outcomes as follows. The mean polymerase binding along all gene copies was averaged to create a simulated Pol II ChIP-seq profile. We used this predicted Pol II ChIP-seq profile to calculate the Pol II traveling ratio: the ratio between gene body and promoter-proximal polymerase, which has been used to measure changes in pause release rate (although we note that experimentally, promoter-proximal Pol II includes both paused and early-transcribing polymerase[61,155,220,221]). Next, we used the number of polymerases that underwent pause release and produced nascent RNA over time for each gene copy to predict measurements from nascent transcript RNA FISH, a method that measures recent transcription in single cells using fluorescent probes specific to introns (Figure 3.1C, Figure A3.1B). Nascent transcript RNA FISH probes bind to nascent RNA to create a fluorescent spot near a gene's location in the nucleus (a 'transcription site') if a gene has transcribed recently (Figure 3.1C). We predicted the average number of active transcription sites per cell (the proportion of gene copies active at a time) and the mean transcription site fluorescence intensity (which measures the average number of RNAs recently transcribed by a single gene copy). We then repeated such simulations while changing the value of each rate parameter one at a time (Figure A3.3E). Thus our simulations generated predictions for how changing specific rates of transcriptional steps would change both Pol II ChIP-seq (population) and nascent transcript RNA FISH (single cell) measurements of transcription.

In this way, we made predictions for the effects of changing specific rates on the fold change in experimental measurements for both models of transcriptional regulation (Figure 3.1B, Figure A3.2A, Figure A3.6; see Figure A3.3 for other models examined). For each model, we found that changes to the rates of burst initiation, polymerase binding or pause release each

resulted in a distinct signature of experimental outcomes. Moreover, these signatures revealed a critical distinction between the two models: we predicted that altering the burst initiation rate would change the Pol II traveling ratio in the 'polymerase pause release occurs in bursts' model, but not in the 'polymerase binding occurs in bursts' model.

3.3 Combining nascent transcript RNA FISH and Pol II ChIP-seq reveals that polymerase binding occurs in bursts

Based on our simulations, by changing each step of transcription experimentally we should be able to first verify the shared experimental signatures of the two models, and then we could change the rate of burst initiation to select between the two models using the behavior of the Pol II traveling ratio (Figure 3.1B). We first changed the rate of polymerase pause release, with both the 'polymerase binding occurs in bursts' and 'polymerase pause release occurs in bursts' models predicting that all transcriptional measures should decrease: transcription sites per cell and transcription site intensity measured by nascent transcript RNA FISH, and Pol II traveling ratio measured by Pol II ChIP-seq (Figure 3.1A-B). Intuitively, if pause release occurred less frequently, a gene would be associated with nascent RNA for a lower fraction of the time (less transcription sites per cell), each time a gene is actively transcribing there would be fewer RNA molecules at a time (lower transcription site intensity), and Pol II would be released into the gene body less frequently (lower Pol II traveling ratio). We used flavopiridol to reduce the rate of pause release; flavopiridol inhibits the p-TEFb kinase complex[150,218,223]. As predicted by both models, we found decreased transcription sites per cell, decreased transcription site intensity, and decreased Pol II traveling ratio. Pol II traveling ratio was calculated using genes that maintained detectable Pol II occupancy after drug treatment (gene properties shown in Figure A3.1C), and was robust to choice of 3' Pol II binding region used as the numerator for the traveling ratio (Figure A3.1F). (Importantly, though some publications have applied high-dose flavopiridol to block new pause release and then have immediately measured transcription, in order to examine the properties of the receding wave of elongating Pol II, in our experiments we used lower drug doses and

measured transcription only after 60 minutes, so we could instead measure a new steady state of transcription with a reduced pause release rate[218].)

Note this experiment also eliminated the commonly used telegraph model model of transcription (Figure A3.3A). The telegraph model does not include an independently regulated pause release step. Thus, it has been sufficient to predict the distribution of mRNA levels in a cell population, but it cannot distinguish the effects of changing polymerase pause release versus polymerase binding rates[163,178,181,208]. The fact that changing the pause release rate changed the Pol II traveling ratio is inconsistent with the telegraph model (Figure A3.3A).

We next tested the predictions of our two models for the effect of changing the polymerase binding rate. Both the 'polymerase binding occurs in bursts' and the 'polymerase pause release occurs in bursts' models predicted that reducing polymerase binding would reduce transcription sites per cell and transcription site intensity but, unlike reducing the pause release rate, would leave Pol II traveling ratio unchanged. (The reason that transcription sites per cell are decreased is because of the period of time the gene is active, the percentage of that time during which there is a nascent transcript from an active polymerase would be less as the polymerase binding rate decreased.) We treated G1E-ER4 cells with triptolide to reduce the polymerase binding rate by inhibiting the helicase activity of TFIIH (Figure 3.1E)[150,218,224–226]. As predicted, triptolide treatment decreased both the transcription sites per cell and the transcription site intensity, while leaving the Pol II traveling ratio unchanged (Figure 3.1E, Figure A3.1D). This experiment supported the predictions of the two models and confirmed that we could experimentally distinguish a change in polymerase binding rate from a change in polymerase pause release rate. Together, these two experiments supported the predictions of both our models.

We then sought to change the burst initiation rate to determine whether transcriptional regulation was more consistent with the 'polymerase binding occurs in bursts' or the 'polymerase pause release occurs in bursts' model (Figure 3.1B). Both models predicted that changing burst initiation should change transcription sites per cell and should not change transcription site

intensity but they disagreed on the effect on Pol II traveling ratio. The 'polymerase binding occurs in bursts' model predicted no change in Pol II traveling ratio, while the 'polymerase pause release occurs in bursts' model predicted that increasing burst initiation should also increase the Pol II traveling ratio (Figure 3.1B). The molecular underpinnings of burst initiation are unclear, and there is no known pharmacological inhibitor of this transcriptional step. However, we previously showed that increasing enhancer-promoter contact for the *Hbb-b1* gene using a synthetic looping factor increased transcription sites per cell without changing transcription site intensity[227] . Both models predicted that such a result could only occur if forced enhancer-promoter looping changed the burst initiation rate. Therefore, we repeated this perturbation to see whether the Pol II traveling ratio changed or remained constant. We found that increasing promoter-enhancer contact did not alter the Pol II traveling ratio for *Hbb-b1*, suggesting that polymerase binding rather than pause release occurs in bursts for this gene. Thus our experimental data was most consistent with a framework in which a burst initiation step was required to permit polymerase binding to a gene (the 'polymerase binding occurs in bursts' model, Figure 3.1B). A caveat of the looping factor experiment was that we could only modulate burst initiation for one gene, but in Figure 3.2A we will show that erythroid differentiation changed burst initiation rate for the housekeeping gene *Prdx2* in addition to *Hbb-b1*, arguing for the generalizability of the 'polymerase binding occurs in bursts' model. In sum, changing three steps of transcription experimentally demonstrated that the 'polymerase binding occurs in bursts' model was most consistent with our data, providing a set of predictions we will subsequently use to identify transcriptional steps changed by complex perturbations.

We also evaluated several other possible model structures, but none of these models were more consistent with our experimental data than the 'polymerase binding occurs in bursts' model (Figure A3.3B-D). A model in which there was no independent burst initiation step and only polymerase binding and pause release rates could be regulated (Figure A3.3B) was inconsistent with the experimental data in Figure 3.1F because it could never predict a change in transcription sites per cell independent of changes in transcription site intensity and Pol II traveling ratio. We

also considered a modified ‘polymerase binding occurs in bursts’ model, in which some proportion of polymerases drop off from the pause site rather than undergoing elongation (Figure A3.3D)[150,228]. This modification did not change the model’s predictions for the experimental effects of changing burst initiation, polymerase binding, or polymerase pause release rates. We also tested a model in which a refractory period of transcriptional silence occurred after each burst (Figure A3.3C). (Some live imaging studies have found evidence for refractory periods but some have not[181,208] .) This model predicted that modulating the rate of escape from the refractory period escape would change transcription sites per cell but would not change transcription site intensity or Pol II traveling ratio, similar to the experimental result in Figure 3.1F. Our data is not inconsistent with this model, but our data also cannot show whether genes in our system transcribe with refractory periods. If indeed our genes transcribe with refractory periods, any change we herein ascribe to a change in burst initiation rate could also be achieved by a change in refractory period escape rate—either way, though, it is the effective rate at which the gene is reactivated that is subject to regulation. Thus, the possible regulation of refractory periods should be examined in future studies. This issue does not alter our ultimate conclusion that polymerase pause release and burst initiation/refractory period escape, and not polymerase binding, are the key regulated steps of transcription.

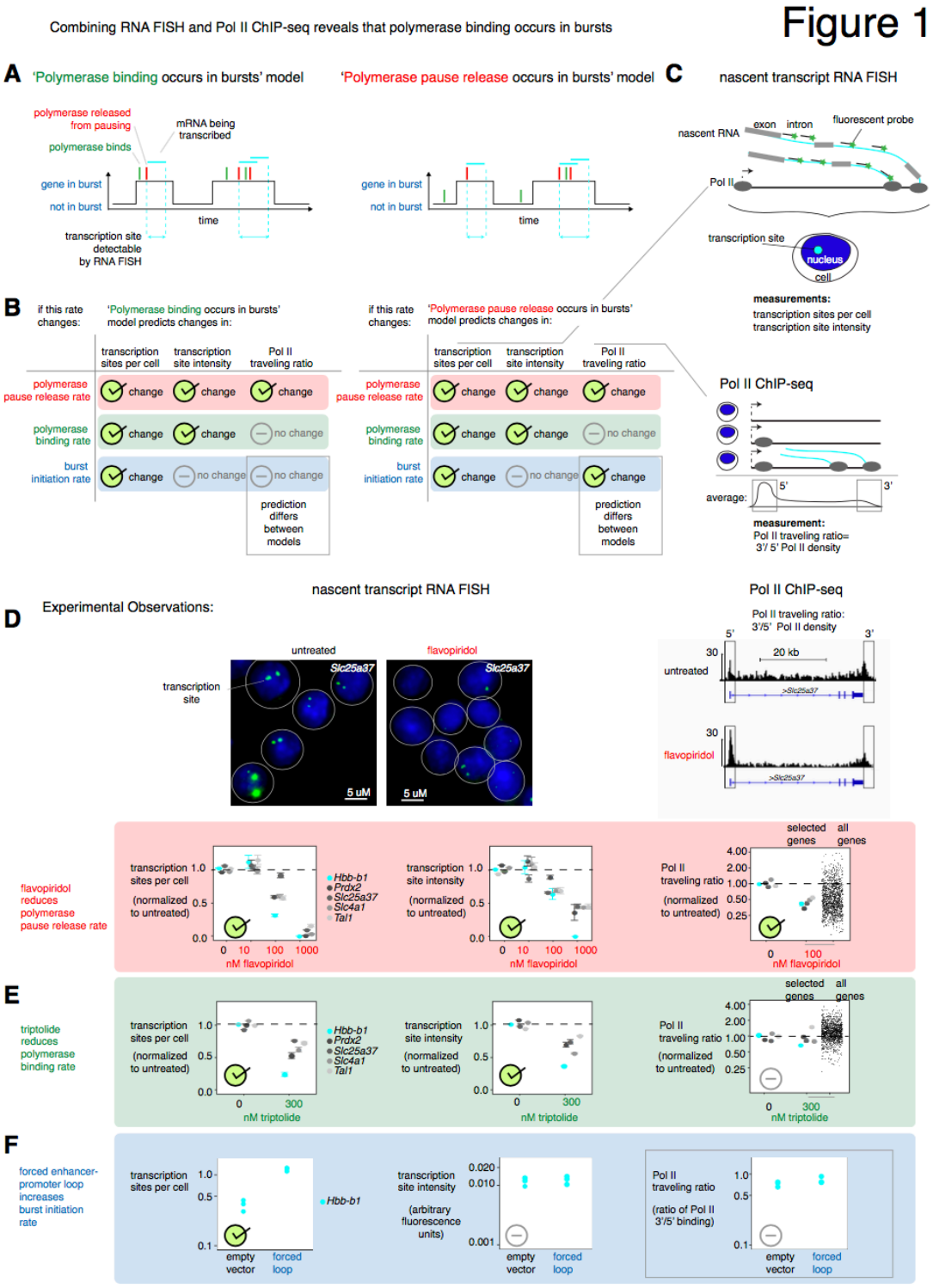


Figure 3. 1: Combining RNA FISH and Pol II ChIP-seq reveals that polymerase binding occurs in bursts A) Structure of the 'polymerase binding occurs in bursts' model and the 'polymerase

pause release occurs in bursts' model. B) Predictions of the models for how changing individual transcriptional rates should change nascent transcript RNA FISH and Pol II ChIP-seq measurements. See Figure A3.2A-B for the quantitative predictions that are summarized here. C) Schematics of nascent transcript RNA FISH and Pol II ChIP-seq. D) Top, representative images from nascent transcript RNA FISH and Pol II ChIP-seq after 60 minutes 100nM flavopiridol treatment of G1E-ER4 cells. Bottom, data summarizing nascent transcript RNA FISH and Pol II ChIP-seq after 0,10,100, or 1000 nM flavopiridol treatment for 60 minutes of G1E-ER4 cells that had been differentiated for 24 hours with 100nM estradiol to stabilize the GATA1-ER fusion protein expressed by the cells. n=3 biological replicates of nascent transcript RNA FISH (49-172 cells per gene per treatment per replicate, see Figure A3.1E, error bars display SEM) and n=3 biological replicates of Pol II ChIP-seq. For all figures, transcription sites per cell was measured by counting mean transcription sites, transcription site intensity was calculated as the mean fluorescence intensity of those sites calculated by 2D Gaussian fitting, and Pol II traveling ratio was calculated as (the Pol II density from the gene end to 1500bp past the gene end) divided by (the Pol II density 750 bp upstream to 750 bp downstream of the transcription start site), although effects were robust to choice of 3' region, see Figure A3.1F. For this figure, ChIP-seq data for 'selected genes' represents data for 5 genes corresponding to FISH experiments (*Hbb-b1, Prdx2, Slc25a37, Slc4a1,* and *Tal1*) and all genes represents 1416 genes still transcribed after drug treatment, see Figure A3.1C for characterization of genes. E) Nascent transcript RNA FISH and Pol II ChIP-seq after 0 or 300 nM triptolide was applied for 60 minutes to G1E-ER4 cells that had been differentiated for 24 hours with 100nM estradiol. n=3 biological replicates of nascent transcript RNA FISH (81-202 cells per gene per treatment per replicate, error bars display SEM) and n=3 biological replicates of Pol II ChIP-seq. For this figure, ChIP-seq data for 'selected genes' represents data for 5 genes corresponding to FISH experiments (*Hbb-b1, Prdx2, Slc25a37, Slc4a1,* and *Tal1*) and all genes represents 1416 genes still transcribed after drug treatment, see Figure A3.1C for characterization of genes. F) Nascent transcript RNA FISH and Pol II ChIP-qPCR after overexpression of looping factor or GFP-expressing empty vector control in G1E-ER4 cells that had been differentiated for 9 hours. n=3 biological replicates of nascent transcript RNA FISH (124-249 cells per treatment per replicate) and n=3 replicates of Pol II ChIP-qPCR. For this figure subsection, traveling ratio was calculated as (input-normalized ChIP-qPCR signal from the 3' gene region) divided by (input-normalized ChIP-qPCR signal from the transcription start site region).

3.4 Erythroid differentiation changes both burst initiation and polymerase pause release rates

Having established that the 'polymerase binding occurs in bursts' model could successfully predict experimental outcomes of changing transcriptional steps, we next sought to apply this model to identify which steps of transcription are regulated by more complex biological perturbations. To investigate the transcriptional effects of erythroid differentiation, we

differentiated G1E-ER4 erythroid cells[49] and performed both nascent transcript RNA FISH and Pol II ChIP-seq at several timepoints during differentiation. We expected that a complex biological process like erythroid differentiation would be associated with changes to multiple steps of transcription, but first we looked for the simplest type of scenario: genes where we could conclusively identify changes in individual steps of transcription.

Early in differentiation, after 4 hours, both the *Hbb-b1* gene, encoding an erythroid-specific subunit of hemoglobin, and the *Prdx2* gene, encoding a broadly-expressed antioxidant enzyme, had increased transcription sites per cell while their transcription site intensities and Pol II traveling ratios were unchanged (Figure 3.2A). According to the 'polymerase binding occurs in bursts' model, only a change in burst initiation could change transcription sites per cell without altering transcription site intensity or Pol II traveling ratio. Thus, we concluded that early erythroid differentiation specifically changes burst initiation for these genes. (Moreover, these data provided further support for the 'polymerase binding occurs in bursts' model and against the 'polymerase pause release occurs in bursts' model: in the latter framework, transcription sites per cell cannot be changed without also changing either Pol II traveling ratio or transcription site intensity, Figure 3.1B.)

Later in differentiation, after 13 hours, *Hbb-b1* and *Prdx2* had increased transcription sites per cell, transcription site intensity, and Pol II traveling ratio, which made it impossible to isolate just one specific rate as having changed at this timepoint for these genes (Figure 3.2B). Our model suggested that only a change in polymerase pause release rate could change Pol II traveling ratio, so erythroid differentiation must change at least polymerase pause release rate for these genes. (Other genes such as *Gata2* and *Myc* also displayed changes in Pol II traveling ratio and thus pause release rate as we will discuss below, Figure 3.3B). However, we could not exclude additional changes in other rates beside pause release rate: for example, if pause release rate and polymerase binding rate both increased after 13 hours of differentiation, the 'polymerase binding occurs in bursts' model would still predict increases in all three experimental measures: transcription sites per cell, transcription site intensity, and Pol II traveling ratio. So far

we concluded that erythroid differentiation changed both burst initiation (at 4 hours) and polymerase pause release rates (at 13 hours), and possibly other rates.

Erythroid differentiation changes both burst initiation and polymerase pause release rates **Figure 2**

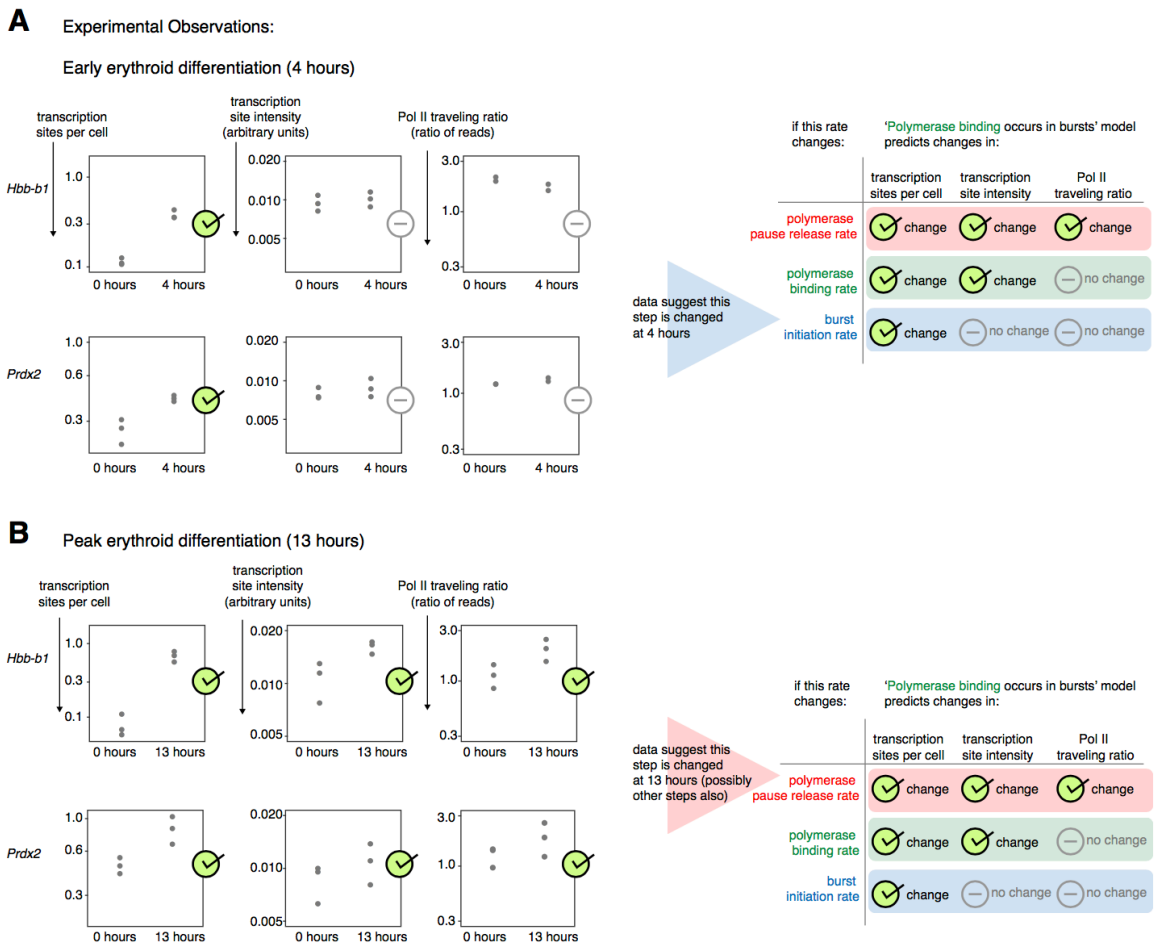


Figure 3. 2: Erythroid differentiation changes both burst initiation and polymerase pause release rates. A) Nascent transcript RNA FISH and Pol II ChIP-seq after 4 hours of differentiation with 100nM estradiol of G1E-ER4 cells for *Hbb-b1* and *Prdx2* genes. n=3 biological replicates of nascent transcript RNA FISH (274-534 cells per gene per treatment per replicate), n=2 biological replicates of Pol II ChIP-seq. B) Nascent transcript RNA FISH and Pol II ChIP-seq after 13 hours of erythroid differentiation with 100nM estradiol of G1E-ER4 cells for *Hbb-b1* and *Prdx2* genes, n=3 biological replicates of nascent transcript RNA FISH (84-665 cells per gene per treatment per replicate) and n=3 replicates of Pol II ChIP-seq.

3.5 Combining modeling and experimentation suggests that erythroid differentiation does not greatly change polymerase binding rate

Above, we identified which steps of transcription were changed in erythroid differentiation by measuring experimental changes one gene at a time, and comparing those data to qualitative model predictions for individual rate changes. However, we found that this logic fell short if multiple transcriptional rates changed simultaneously. The 'polymerase binding occurs in bursts' model, however, also makes predictions for how changing transcriptional rates affects the *quantitative relationship between changes* in nascent transcript RNA FISH and Pol II ChIP-seq. By measuring experimental changes across many genes, we sought to observe quantitative relationships between experimental changes to statistically identify which rates were most likely to be changing in the genes as a group.

Specifically, the quantitative relationship between transcription site intensity and Pol II traveling ratio changes allowed us to distinguish changes in multiple rates. To step through this logic, changes in different individual rates were predicted to result in distinct relationships between transcription site intensity and Pol II traveling ratio changes (Figure 3.3A, left). Each point in Figure 3.3A represents the increase or decrease in experimental measurements predicted to result from either an increase or decrease in the indicated rate, with all the points displayed demonstrating a 1000 fold range of rate values. If polymerase pause release rate changed, both Pol II traveling ratio and transcription site intensity were predicted to change in a correlated manner (Figure 3.3A, top left): if a perturbation like differentiation increased pause release rate, both Pol II traveling ratio and transcription site intensity should increase. A change in polymerase binding rate would instead change transcription site intensity but not change Pol II traveling ratio, resulting in a vertical line relationship (Figure 3.3A, middle left). (A similar outcome would result if burst termination rate was changed, Figure A3.4A). Finally, if burst initiation rate were changed, neither transcription site intensity nor Pol II traveling ratio were predicted to change greatly (Figure 3.3A, bottom left).

Combining modeling and experimentation suggests that erythroid differentiation does not greatly change polymerase binding rate

Figure 3

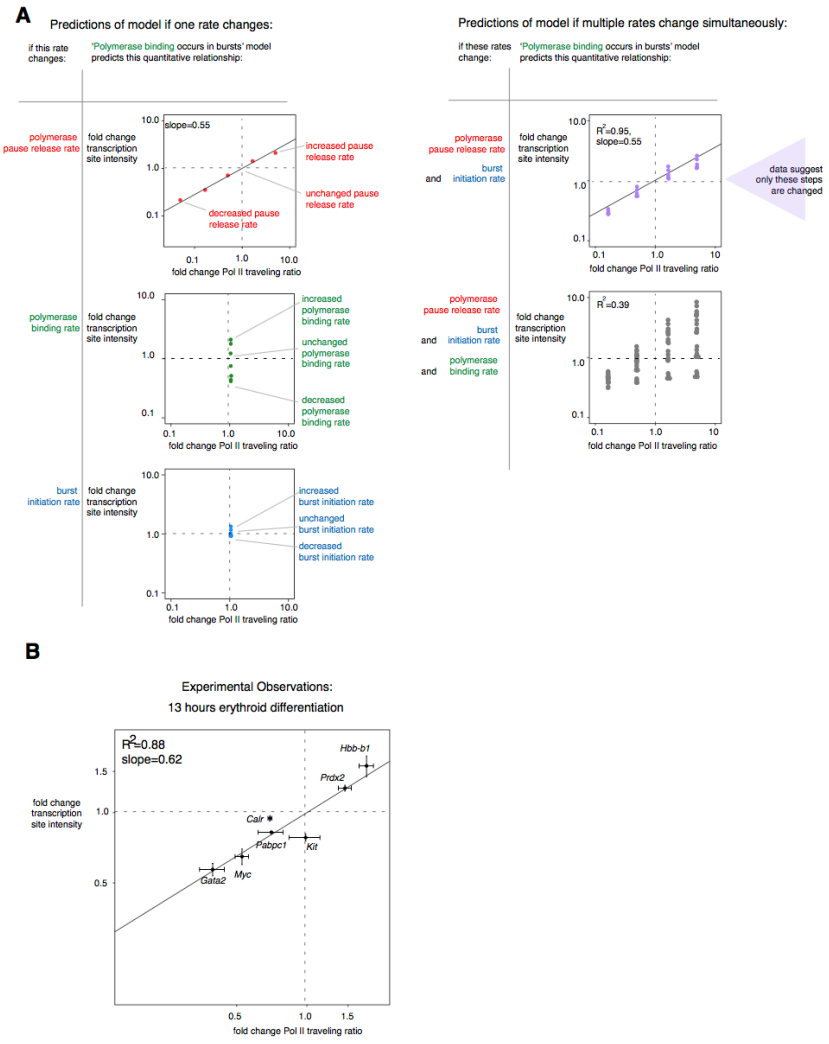


Figure 3. 3 Combining modeling and experimentation suggests that erythroid differentiation does not greatly change polymerase binding rate

a. Predictions of the 'polymerase binding occurs in bursts model' for how changing one or multiple rates of transcription affects changes in transcription site intensity and Pol II traveling ratio. Each figure represents changing the indicated rates from intermediates value up or down, with a total 1000-fold range of possible values represented for each rate (See Figure A3.3E for parameter values tested).

b. Fold change in transcription site intensity and Pol II traveling ratio for 7 genes after 13 hours of differentiation with 100nM estradiol of G1E-ER4 cells, n=3 biological replicates each of nascent transcript RNA FISH (84-665 cells per gene per treatment per replicate) and n=3 biological replicates of Pol II ChIP-seq. Error bars display SEM.

Similarly, changes in different combinations of rates were predicted to result in distinct quantitative relationships between transcription site intensity and Pol II traveling ratio changes. Examination of individual genes above (Figure 3.2) showed that erythroid differentiation could change both burst initiation and pause release rates; if these were the only rates altered, the model predicted that Pol II traveling ratio and transcription site intensity would change in a highly correlated manner (Figure 3.3A top right, R2=0.95), with a relationship of approximately log(fold change transcription site intensity)= 0.55*log(fold change traveling ratio), although the slope of relationship depends somewhat on model parameters (Figure A3.4C). This pattern of transcriptional changes is produced because changing the polymerase pause release rate produces a log y=0.55*log x relationship, while burst initiation rate has little effect on either x or y axis values (Figure 3.3A left). However, if polymerase binding rate were changed in addition to burst initiation and pause release rates, the model predicted that if we measured the effect of differentiation on number of genes, the changes in traveling ratio and transcription site intensity should NOT be strongly correlated (R2=0.39) (Figure 3.3A, bottom right). (Similarly if termination rate changed along with burst initiation and pause release rate, the changes in traveling ratio and transcription site intensity would also not be strongly correlated, Figure A3.4B.) Note that if two or more rates change, a range of behaviors can occur: a gene could display large changes in both rates, or a large change in the first and small change in the second, or vice versa, so multiple genes must be examined to characterize the whole range of experimental behaviors.

With this prediction in hand, we examined how 13 hours of erythroid differentiation affected the quantitative relationship between Pol II traveling ratio and transcription site intensity changes (Figure 3.3B). We found that the changes in transcription site intensity and Pol II traveling ratio were strongly correlated (R2=0.88): for example, both *Prdx2* and *Hbb-b1* genes had increased transcription site intensity and Pol II traveling ratio, while both *Myc* (encoding a broadly-expressed transcription factor controlling proliferation) and *Gata2* (encoding a hematopoietic stem cell transcription factor) genes had decreased transcription site intensity and Pol II traveling ratio. Moreover, this quantitative relationship fell near the log y= 0.55* log x

relationship predicted if only burst initiation and pause release were changed, with a slope of 0.62. Taken together, experimental data for these genes suggest that polymerase binding rate was not greatly changed in erythroid differentiation. (We can similarly exclude a large change in burst termination rate, by comparing to Figure A3.4B). Thus we conclude that burst initiation and polymerase pause release are the main rates changed by erythroid differentiation.

While our analysis suggested changes to mainly burst initiation and pause release in differentiation, the degree to which these rates changed is different for different genes. One hypothesis was that promoter accessibility could correlate to differential transcriptional changes, so we analyzed publicly available data to see whether the change in promoter DNAse sensitivity could predict the change in transcriptional measurements in response to differentiation (Figure A3.4D). We found no strong correlations. Further studies will be required to examine why different genes experience different transcriptional changes in response to erythroid differentiation.

3.6 BET inhibition changes both burst initiation and polymerase pause release rates

Given that previous literature had identified polymerase binding rate as a key target for transcriptional regulation, it was surprising that erythroid differentiation seemed not to greatly affect this rate. We thus examined several more complex biological perturbations to see whether they too chiefly affected burst initiation and polymerase pause release rates.

We next asked which rates of transcription were changed by BET inhibitor treatment. BET inhibitor treatment blocks the transcriptional activator proteins BRD2, BRD3 and BRD4 from binding to chromatin, and may inhibit multiple facets of gene regulation including polymerase pause release and enhancer activity[55,229,230]. We first looked for genes where we could identify individual transcriptional rate changes produced by BET inhibitor treatment. Acute BET inhibitor treatment reduced mean transcription sites per cell for the *Hbb-b1* gene without changing transcription site intensity or Pol II traveling ratio, showing that burst initiation was specifically reduced for this gene (Figure 3.4A). In contrast, for the genes *Tal1, Slc25a37,* and *Slc4a1* (which encode an erythroid transcription factor and two solute channels highly expressed in erythroid

cells), all three RNA FISH and Pol II ChIP-seq measures were reduced (Figure 3.4B). Thus for these genes, at least polymerase pause release rate was changed but potentially also other rates. So by using an approach of identifying single rate changes gene by gene, we found that BET inhibition can change both burst initiation rate and polymerase pause release rate, and might potentially change other rates.

3.7 Combining modeling and experimentation suggests that BET inhibition does not greatly change polymerase binding rate

We next sought to distinguish whether BET inhibition might alter polymerase binding rate, or if mainly burst initiation and pause release rates were changed. We again measured nascent transcript RNA FISH and Pol II ChIP-seq changes in multiple genes in response to BET inhibitor treatment (12 genes in this case), and examined the quantitative relationship between changes in transcription site intensity and Pol II traveling ratio.

If only burst initiation rate and polymerase pause release rate were altered by BET inhibition, the ‘polymerase binding occurs in bursts’ model predicted that Pol II traveling ratio and transcription site intensity would change in a highly correlated manner (R2=0.95), with a relationship of approximately log(transcription site intensity)= 0.55*log(traveling ratio) (Figure 3.3A, upper right); while if polymerase binding rate were also changed, changes in traveling ratio and transcription site intensity should not be strongly correlated (R2=0.39, Figure 3.3A lower right). Examining 12 genes, including both the important erythroid genes shown in Figure 3.4 as well as housekeeping genes such as *Pabpc1* and *Hnrnpl*, we found that BET inhibition produced a strongly correlated relationship between changes in Pol II traveling ratio and transcription site intensity (R2=0.77) falling near the log(transcription site intensity)= 0.55*log(traveling ratio) line (Figure 3.5A), with a slope of 0.41. We thus concluded that BET inhibitor treatment did not greatly change polymerase binding rate.

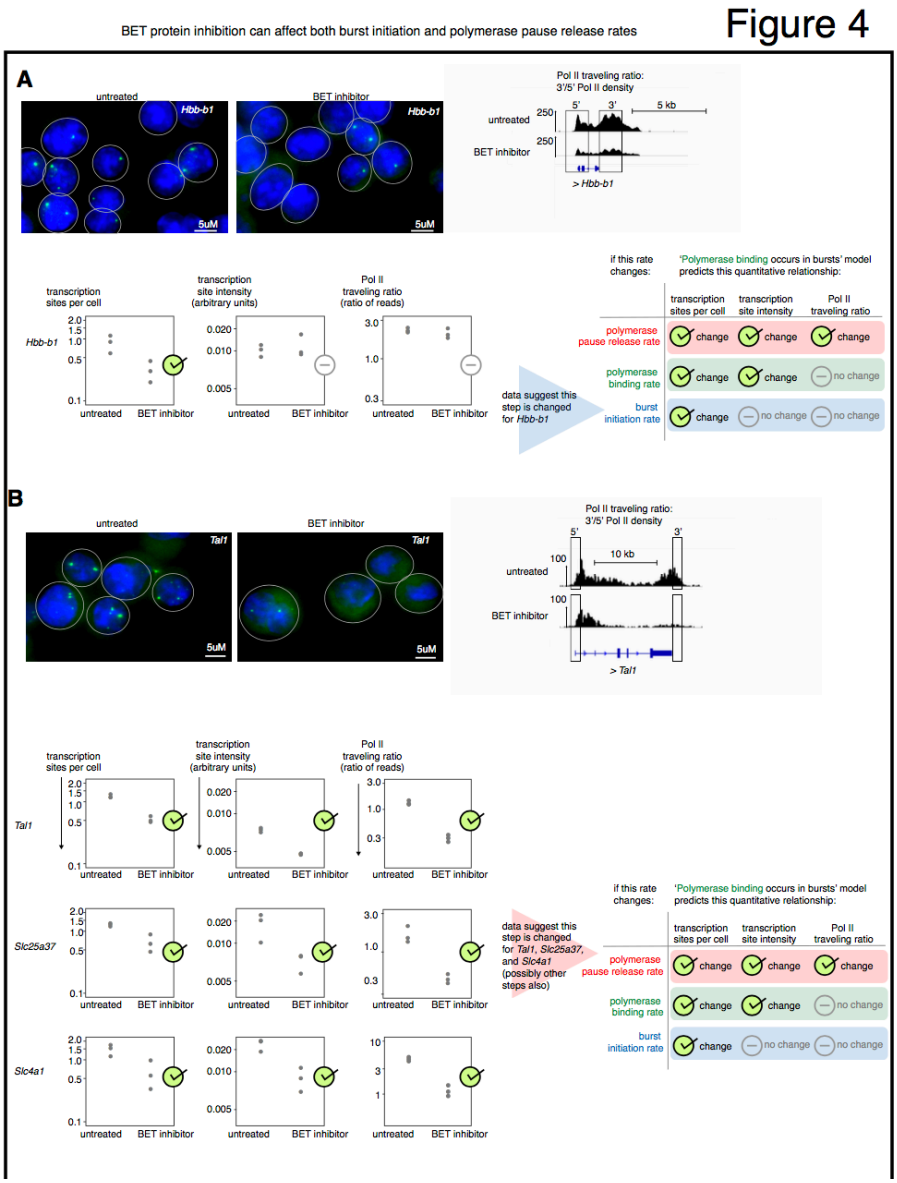


Figure 3. 4: BET inhibitor treatment changes both burst initiation and polymerase pause release rates. A) Representative and summary nascent transcript RNA FISH and Pol II ChIP-seq after 60 minutes of JQ1 BET inhibitor treatment of G1E-ER4 cells differentiated for 24 hours with 100nM estradiol for the *Hbb-b1* gene, n=3 biological replicates of nascent transcript RNA FISH (72-306 cells per replicate per experiment) and n=3 replicates of Pol II ChIP-seq. B) Nascent transcript RNA FISH and Pol II ChIP-seq after 60 minutes of JQ1 BET inhibitor treatment of G1E-ER4 cells differentiated for 24 hours with 100nM estradiol for *Slc25a37, Tal1,* and *Slc4a1* genes, n=3 biological replicates of nascent transcript RNA FISH (94-195 cells per replicate per experiment) and n=3 biological replicates of Pol II ChIP-seq. Top, representative RNA FISH and Pol II ChIP-seq data.

This analysis raised the question of why different genes experienced changes in rates of different transcriptional steps in response to BET inhibitor treatment. We analyzed publicly available data to see whether Histone 3 lysine 27 acetyl levels at gene promoters could affect the change in transcriptional measurements in response to BET inhibition, since BET proteins bind to acetylated histones. We found that promoter H3K27ac weakly correlated to change in transcription sites per cell in response to BET inhibition (Figure A3.5B), but not to changes in other transcriptional measurements. Further studies will be required to examine why different genes experience different transcriptional changes in response to this perturbation.

Some previous studies had suggested that BET inhibition might alter transcriptional elongation rate, a rate that was held constant in the 'polymerase binding occurs in bursts' model. To exclude this possibility, we blocked transcriptional pause release using DRB, and then allowed transcriptional elongation to resume in the presence or absence of BET inhibitor. By measuring the time it took to transcribe from an early to a late region of the nascent RNA using RNA FISH, we calculated the effect of BET inhibitor on elongation rate (Figure A3.5A). We found for the *Zfpm1* gene that BET inhibitor did not affect the elongation rate greatly, and that our calculated estimate for the elongation rate, 2.0 kilobases per minute, was similar to that from previous studies, (2-5 kilobases per minute[98,218,231]).

3.8 Combining modeling and experimentation suggests that *Slc25a37* enhancer mutation does not greatly change polymerase binding rate

We also used the 'polymerase binding occurs in bursts' model to examine what rates were changed by enhancer mutagenesis of the lineage-specific *Slc25a37* (mitoferrin-1) enhancer. The Xu group[70] previously created a series of G1E-ER4 cell lines with partial and full deletions of the large *Slc25a37* enhancer region which reduced the RNA output of the target gene (encoding an erythroid-specific mitochondrial iron transporter) to varying degrees (Figure A3.5C-D). We performed nascent transcript RNA FISH and Pol II ChIP for *Slc25a37* in the three partial-enhancer-deletion and one full-deletion cell lines to examine which rates are changed. The

enhancer-deletion cells had correlated reductions of Pol II traveling ratio and transcription site intensity (R2=0.89) falling near the log(transcription site intensity)= 0.55*log(traveling ratio) line, with a slope of 0.70 (Figure A3.5E). (Our modeling suggests that these enhancer mutations likely also reduce burst initiation rate, due to the strong reduction produced in Pol II promoter occupancy.) Thus, mutation of the *Slc25a37* enhancer reduced polymerase pause release rate, but appeared not to greatly change polymerase binding rate.

Combining modeling and experimentation suggests that BET inhibition does not greatly change polymerase binding rate

Figure 5

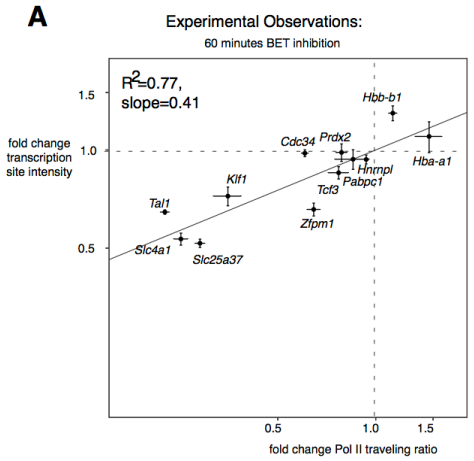


Figure 3. 5: Combining modeling and experimentation suggests that BET inhibitor treatment does not greatly change polymerase binding rate. A) Fold change in transcription site intensity and Pol II traveling ratio for 12 genes after 60 minutes of 250nM JQ1 BET inhibitor treatment of G1E-ER4 cells differentiated for 24 hours with 100nM estradiol, n=3 biological replicates of nascent transcript RNA FISH (72-306 cells per gene per replicate per experiment) and n=3 biological replicates of Pol II ChIP-seq. Error bars display SEM.

In sum, we developed a modeling and experimental approach to identify changes in transcriptional steps caused by complex perturbations. Surprisingly, we found that no perturbations aside from the targeted inhibitor triptolide seemed to strongly alter polymerase binding rate, while the main regulated steps of transcription were burst initiation and polymerase pause release.

3.9 Discussion

We sought to determine which steps of transcription are biologically regulated. To this end, we combined both single cell and bulk biochemical measures of transcription in order to measure changes to rates of bursting as well as rates of polymerase binding and pause release upon various regulatory stimuli. This approach revealed that polymerase binding, rather than polymerase pause release, is confined to active transcriptional periods, clarifying how bursting fits together with biochemical steps of transcription. We then applied a number of biological perturbations and found that burst initiation and polymerase pause release were the main regulated steps of transcription. Polymerase binding rate was not altered by the perturbations we examined except by triptolide, which is a known inhibitor of polymerase binding. We note however that such experiments should be also performed with other genes and in other systems to confirm the generality of these conclusions.

Importantly, this study showed that both Pol II ChIP-seq and nascent transcript RNA FISH were required to uniquely map changes in transcriptional steps to experimental changes. Nascent transcript RNA FISH distinguished between a change in polymerase binding rate and a change in burst initiation rate (using transcription site intensity), while Pol II ChIP-seq could not (Figure 3.1E-F). On the other hand, Pol II ChIP-seq but not RNA FISH could distinguish between changes in polymerase binding and polymerase pause release rates (Figures 3.1D-E). Therefore combining bulk and single-cell transcriptional measures is critical to determine which step of transcription was changed in a perturbation with unknown effects.

Our results point primarily to burst initiation rate and polymerase pause release rate as key regulated steps in response to the various biological perturbations of our experiments. Polymerase binding rate appeared not to be subject to regulation by the biological perturbations we examined, but this does not mean that polymerase binding is not rate limiting. Indeed, polymerase binding rate was in fact rate limiting in the context of triptolide treatment. The finding that polymerase binding rate is not typically altered by biological perturbations corresponds well

to recent single molecule imaging studies of Pol II: the authors found that even when a gene was lowly transcribed, many molecules of polymerase clustered near its promoter[37,38] . Since only one polymerase can bind a promoter at a time[156], it is possible that polymerase binding typically occurs with a very high rate in all conditions. It would be interesting in future to examine whether each gene has its own consistent polymerase binding rate, and whether this could be encoded by gene promoters.

Given that our study and others[85,177,232] have revealed the key role that control of transcriptional bursting plays in gene regulation, it will be essential in future to better characterize the biochemical underpinnings of transcriptional bursting. An intriguing hypothesis with some experimental support is that burst initiation is related to enhancer-promoter looping[85,227,232]. Here, we have shown that for *Hbb-b1*, increasing enhancer-promoter contact specifically increases burst initiation, consistent with other studies that have isolated looping from other potential enhancer activities. It is possible that each time an enhancer contacts a promoter, the gene undergoes burst initiation. This hypothesis is supported by a recent paper showing that enhancer-promoter looping occurs before every transcriptional burst[85]. Burst initiation could also be controlled by events independent from enhancer-promoter contact, such as transcription factor binding to the promoter or nucleosome remodeling. We believe that the ability to perform more biochemical assays in single cells will be required to understand the mechanistic basis of transcriptional bursts.

Our experiments with BET inhibitor treatment show that BET proteins can regulate both burst initiation rate and polymerase pause release rate. Many papers in the literature have suggested that a chief transcriptional effect of BET inhibition was to inhibit polymerase pause release, and we also found that this is a dominant effect[221,233–235] . However, we also observed significant effects of BET inhibitors on burst initiation rate: for example, BET inhibition of *Hbb-b1* only altered burst initiation rate and did not change polymerase pause release rate. Thus roles of BET proteins beyond just influencing polymerase pause release are likely to also be important for modulating gene expression[55,229,230]. Relatedly, an important future direction will be to explore

why different genes experienced changes in different rates of transcription in response to the same perturbation.

Some studies have raised the possibilities of other forms of transcriptional regulation not included in the 'polymerase binding occurs in bursts' model, such as refractory periods between bursts[181,236] or faster polymerase reinitiation after prior transcription[237] . We tested a model including refractory periods (Figure A3.3C), which corresponded to our data equally as well as the 'polymerase binding occurs in bursts' model. If there are refractory periods in our system, this modeling could suggest that the role we here attributed to burst initiation could be shared by the rate of escape from refractory period (i.e., early erythroid differentiation could regulate the rate of escape from a refractory period for *Hbb-b1* and *Prdx2*, see Figure 3.2A). The data shown here does not require a refractory period as part of the transcriptional framework to explain the data, and we could not observe a refractory period using these methods, but our data also cannot exclude a role for refractory period regulation. However, some live imaging studies seem to exclude the presence of a refractory period[173,183,208]. Therefore, it is possible that more rates may be regulated in which we were not able to detect changes (e.g. refractory period escape or independent polymerase re-binding rate).

Our study more generally demonstrates the utility of a model-based approach to identifying regulated transcriptional steps. In particular, previous studies of the burst initiation and termination phases have been largely phenomenological, characterized primarily by observables like burst frequency and transcription site intensity[177,180,238,239]. It has, however, proven difficult to discern any general rules or principles from these studies. Our study suggests that this may be due to the fact that these experimental observables can be convolved in counterintuitive ways; indeed, there is no reason *a priori* to believe that such observables map one-to-one to particular biological processes. For example, our study suggests that transcription site intensity was not independently regulated; rather, transcription sites per cell and transcription site intensity were both altered when polymerase binding or polymerase pause release are changed (Figure 3.1). By using a model-based approach informed by a combination of data types, we were instead able to

interpret these observables in terms of parameters of a simple model of transcription, thus revealing a more consistent underlying picture over a variety of perturbations.

3.10 Experimental Procedures

Murine cell culture, infection and sorting

G1E cells were previously derived through deletion of GATA1 in mouse embryonic stem cells, followed by in vitro differentiation[49]. We cultured a subline of G1E cells, G1E-ER4, in which GATA-1-ER was transduced by retrovirus(Weiss et al., 1997). G1E-ER4 cells were induced to mature by the addition of 100nM estradiol to culture media.

For differentiation experiments (Figures 3.2 and 3.3), cells were treated with estradiol for 4 or 13 hours as noted in the figure. For flavopiridol experiments (Figure 3.1D), the noted concentration of flavopiridol (either 10nM, 100nM, or 1uM) was added for 60 minutes to cells differentiated for 24 hours. For triptolide experiments (Figure 3.1E, Figure A3.1D), the noted concentration of flavopiridol (either 10nM, 100nM, 300nM, or 1uM) was added for 60 minutes to cells differentiated for 24 hours. For BET inhibition experiments (Figures 3.4 and 3.5), 250nM JQ1 was added for 60 minutes to cells differentiated for 24 hours. For forced looping experiments (Figure 3.1F), cells were infected with the MIGR-1 retrovirus expressing only GFP or expressing mZF-SA followed by an IRES element and GFP[79]. Cells were infected, expanded for two days and sorted using a BD FacsAria to purify GFP+ infected cells from control and mZF-SA samples, estradiol was added for 9 hours and transcription was measured by RNA FISH or ChIP-qPCR. For DRB transcriptional blocking to measure nascent transcript RNA FISH half-life (Figure A3.1B), 75uM DRB was added for 0, 10 or 60 minutes to cells differentiated for 24 hours.

For experiments examining the *Slc25a37* enhancer region (Figure A3.5), control and 4 enhancer deletion G1E-ER4 cell lines created previously were kindly provided by the Jian Xu lab [70]. Cells were differentiated for 24 hours using estradiol and transcription was measured by RNA FISH or ChIP-qPCR.

For experiments to measure the Pol II elongation rate (Figure A3.5A), 75uM DRB was added to cells differentiated 21 hours. After 3 hours (24 hours differentiation), DRB was removed by washing and cells were fixed at 10-minute intervals after DRB removal.

Chromatin Immunoprecipitation

We performed ChIP as previously described[217] , using the N-20 Pol II antibody (Santa Cruz sc899). For ChIP-sequencing, library construction was performed using Illumina's TruSeq ChIP sample preparation kit (Illumina, catalog no. IP-202-1012) according to manufacturer's specifications with the addition of a size selection using SPRIselect beads (Beckman Coulter, catalog no. B23318) prior to PCR amplification. Library size was determined (average ~340 bp) using the Agilent Bioanalyzer 2100, followed by quantitation using real-time PCR using the KAPA Library Quant Kit for Illumina (KAPA Biosystems catalog no. KK4835). Libraries were then pooled and sequenced on the Illumina NextSeq 500 using Illumina sequencing reagents according to manufacturer's instructions.

For perturbations in which only one gene was expected to change, forced enhancer-promoter looping and *Slc25a37* enhancer deletion, ChIP-qPCR was used to measure traveling ratio changes rather than ChIP-sequencing (Figure 3.1F, Figure A3.5C-E). Primers were tiled along *Hbb-b1* and *Slc25a37* respectively, and the ratio of TES-proximal to TSS-proximal promoter signal were used as traveling ratio, after normalizing to ChIP input.

ChIP-sequencing analysis

All Pol II ChIP-seq data was generated for this study, except the 0 and 13h G1E-ER4 differentiation (Figures 3.2 and 3.3), which was from[240]. We used bcl2fastq2 to convert and demultiplex the reads. We also applied fastQC to get read stats, Bowtie (0.12.8) to map reads to the mm9 genome (multiple and unique mappings), Samtools to convert SAM to BAM and get stats, MACS (1.3.7.1) to make wiggle files, and used wigToBigWig to convert wiggle files to bigWig files.

We then performed bigWigAverageOverBed to find Pol II binding signal at transcription start site (TSS) region (750 bp upstream of transcription start site to 750 bp downstream), and transcription end site (TES) region (transcription end site to 1500 bp downstream of transcription end site), and calculated the Pol II traveling ratio for TES/TSS Pol II signal. (In Figure A3.1F, we instead calculated the Pol II traveling ratio by (gene body)/TSS, where gene body included 750 bp downstream of TSS to 1500 bp downstream of TSS.) We displayed data from only the genes with detectable Pol II binding in all regions in all experimental conditions, so that we could avoid examining genes that aren't transcribed in one condition, by setting arbitrary cutoff values on the bigWigAverageOverBed results, and results were robust to changes in cutoff. To display Pol II ChIP-seq tracks in figures (e.g. Figure 3.1D), we displayed tracks normalized by number of aligned reads per sample.

For analysis of promoter DNAse hypersensitivity (Figure A3.4D), data was taken from (Hsiung et al. 2015). DNAse promoter sensitivity was determined using bigWigAverageOverBed in the TSS region (750 bp upstream to 750 bp downstream of TSS).

For analysis of promoter H3K27ac density (Figure A3.5B), data was taken from[240] .DNAse promoter sensitivity was determined using bigWigAverageOverBed in the TSS region (750 bp upstream to 750 bp downstream of TSS). This same dataset was used for the H3K27ac track displayed in Figure A3.5C.

Single-molecule RNA FISH imaging

We performed single-molecule RNA FISH as described previously[170,203] .. All probes used were complementary to introns of gene of interest (except the exon-targeted probe measures of mRNA shown in Figures A3.1A and A3.5C). All intron probes were approximately evenly spaced along length of RNA (except the transcriptional elongation rate measurements in Figure A3.5A, in which probes were placed at 5' or 3' end of *Zfpm1*).

Briefly, we fixed cells in 1.85% formaldehyde for 10min at room temperature, and stored them in 70% ethanol at 4 degrees C until imaging. We hybridized pools of FISH probes to samples, followed

by DAPI staining and wash steps performed in suspension. Samples were cytospun onto slides for imaging on a Nikon Ti-E inverted fluorescence microscope using a 100x Plan-Apo objective (numerical aperture of 1.43), a cooled CCD camera (Pixis 1024B from Princeton Instruments), and filter sets SP102v1 (Chroma), SP104v2 (Chroma), and 31000v2 (Chroma) for Cy3, Cy5, and DAPI, respectively. Custom filter (Omega) was used for Alexa594. We took 45 optical z-sections at intervals of 0.35 microns, spanning the vertical extent of cells, with 1s exposure time for Cy3, Cy5, and Alexa594, and 35ms for DAPI.

Image Analysis

We manually segmented boundaries of cells from bright field images and localized RNA spots using custom software written in MATLAB with subsequent analyses performed in R. Transcription sites were identified by bright nuclear intron spots. Fluorescence intensities of transcription sites were determined by 2D Gaussian fitting on processed image data.

Mathematical modeling

Mathematical models were constructed and simulations were performed in Matlab using Gillespie's stochastic simulation algorithm[222]. For the 'polymerase binding occurs in bursts' model, genes could be in three states (closed, open, polymerase bound). A gene in the closed (off, non-burst initiated) state transitions to the open (bursting) state at the burst initiation rate, while a gene in the open state can have polymerase bind the promoter with the polymerase binding rate. Once a polymerase is bound, that polymerase can be released to elongation with the rate of polymerase release from pausing, and the promoter thus returns to the open and unbound state. From either the open unbound or open polymerase bound state, the gene can transition to the off state with the rate of burst termination. We varied each of these rates through a 1000 fold range of values (as shown in Figure A3.3E). Other model variations ('polymerase pause release occurs in bursts', telegraph, 'bind-release', 'polymerase binding occurs in bursts with refractory period', and

'polymerase binding occurs in bursts with possibility of termination from pause release site') had different transition possibilities dictated by the frameworks displayed in Figures A3.2 and A3.3.

For every set of rates in a given model, we simulated 1000 gene copies, which were allowed to proceed through 2000 changes in state. We recorded the state of each gene copy at 1500 time intervals. The simulation equilibrated away from the initial condition (every gene copy started in the 'off' state) for every property within this time window, typically after ~100-200 time steps, and the value at which it converged was used for the following analyses. Each time a gene underwent the pause release (or RNA production) step, we allowed that polymerase to produce 1 RNA and elongate along the gene body for a short, fixed amount of time, and then for the polymerase to fall off the gene body and the nascent RNA to no longer be detected by RNA FISH. We then used the above information to calculate: transcription sites per cell (proportion of gene copies with at least 1 polymerase elongating in the gene body at a given time); transcription site intensity (average number of elongating polymerases on one gene for gene copies with at least 1 polymerase elongating in the gene body at a given time), polymerase binding signal at promoter (proportion of gene copies in the pol2 bound state), polymerase binding signal at gene body (average number of Pol molecules in the gene body). We calculated Pol II traveling ratio as gene body polymerase signal divided by promoter polymerase signal. (Note that none of these measures rely on identifying the beginning and end of a 'burst': if bursts are frequent enough, they will run together, leading to all gene copies being transcribed (maximum of 2 gene copies per cell) and leading to increased transcription site intensity).

Figure generation, plotting and graphics

We used R packages dplyr and ggplot2 to produce nearly all figures, followed by cosmetic adjustments in Adobe Illustrator. All code to perform modeling, data analysis and figure generation is available in Dropbox (https://www.dropbox.com/sh/ti5lfcn0e71p45i/AAD64o4EM6v3_1Gw3_88XuBza?dl=0).

CHAPTER 4: Discussion

4.1 The goal of my thesis work: using the lens of transcriptional bursting to improve our model of transcriptional regulation

In my graduate work, I asked how the phenomenon of transcriptional bursting might change the field's understanding of transcriptional regulation. Transcriptional bursting was first described relatively recently[163,203], when the field of transcriptional regulation was well-established, and this surprising observation provided the field new information about how transcription operates. However, there has been a lack of studies re-examining our previous long-standing and mechanistic understanding of transcription in light of these new developments.

Many of the first studies examining transcriptional bursting mainly addressed how the observed properties of bursting, such as burst frequency and intensity, were modulated in different cell types in response to different stimuli. Such studies clearly demonstrated that the properties of bursting could be regulated, but failed to address how bursting was regulated on a molecular level. Conversely, biochemical-based studies of transcription typically ignored the finding of transcriptional bursting when trying to advance our understanding of transcriptional regulation. Some studies seemed to recognize this gap between the observation of transcriptional bursting and our previous models of transcriptional regulation: studies sometimes alluded to the phenomenon of bursting, and suggest that it related to their findings[156,241], but in many cases this was a theoretical discussion, without actual experimental examination of bursting in their system.

This gap between the rich history of transcriptional regulation studies, and the young field of transcriptional bursting observations, was the gulf that my thesis work set out to address. We realized that given the extensive examination of transcriptional regulation in many systems, it was likely that molecular players involved in bursting regulation were previously identified in biochemical approaches. In effect, transcriptional bursting had always been present in biochemical studies of transcription with approaches like ChIP-seq and PRO-seq; such studies could not measure this phenomenon due to the averaging effects of combining millions of cells.

However, bursting was an inextricable part of transcription: if one could identify a way to reduce either the frequency or amplitude of a gene's bursts to zero, transcription would be blocked, even in assays that couldn't detect bursting at all. Thus, we hypothesized that by separating out which molecular processes might regulate bursting, imaging-based studies of transcriptional bursting might add information to disentangle how transcription is regulated.

4.2 Perspectives on enhancer regulation of transcriptional burst frequency

We reasoned that the best way to identify the regulation of transcriptional bursting was to measure transcriptional bursting in a system that was well-characterized from a biochemical perspective of transcription, thereby giving us the tools to identify how to connect transcriptional bursting to known regulation of transcription. I thus began my Ph.D. by examining the transcriptional bursting of the beta-globin locus during erythroid differentiation. Using the zinc-finger-based forced looping approach previously developed in the Blobel lab[79,112], I found that in both mouse and primary human erythroblasts, enhancer-promoter contact controlled the fraction of time the beta-globin gene transcribed, without changing the transcriptional intensity. This work identified a connection between a known biochemically-measurable process required for transcription, enhancer-promoter contact, and an isolated property of bursting (schematic in Figure 4.1).

Satisfyingly, an independent study published at the same time in an unrelated system confirmed that enhancer-promoter contact modulated burst frequency without changing burst size[232]. In this study, Fukaya and colleagues used the MS2 technology, which enables fluorescent labeling of a genetically-tagged nascent RNA in live cells, to measured bursting in transgenes in *Drosophila melanogaster*. They found that either increasing the distance between an enhancer and its target transgene, or separating the enhancer and transgene with an insulator sequence, decreased the burst frequency of the reporter gene without changing the burst

intensity. They confirmed this finding with several different and unrelated enhancer sequences. In effect, the main finding of both studies was in agreement.

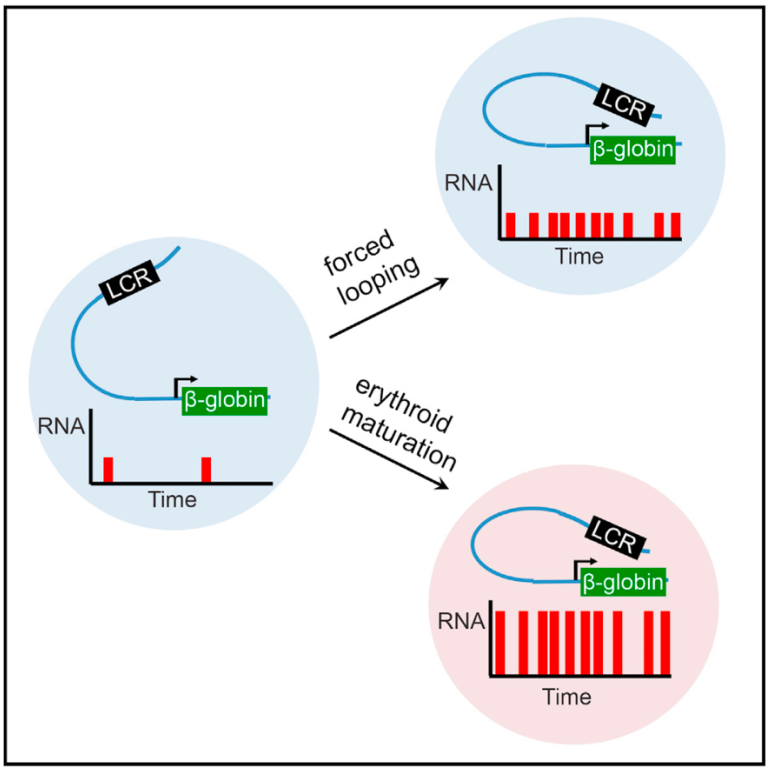


Figure 4. 1: Schematic of findings in Bartman et al. 2016. Forced enhancer-promoter looping increased burst fraction at the beta-globin locus, while erythroid maturation both increased burst fraction and burst intensity.

One interesting inconsistency between our study and Fukaya et al. related to the phenomenon of promoter competition for enhancer activity. A longstanding question in the field of transcriptional regulation has been whether an enhancer can only drive one gene at a time, and thus if driving multiple target genes by the same enhancer would effectively divide the power of that enhancer [4,18,114,115]. Given the prevalence of enhancers that contact multiple genes detected by chromosome conformation capture approaches, and of genes that contact multiple enhancers, understanding how such complex regulatory systems work may be important to understanding transcriptional regulation in cells[99,119,125]. Moreover, the phenomenon of bursting could be suggested to be a solution to the challenge of regulating multiple target genes with one enhancer:

rather than enhancer-promoter networks forming complex rosette-like structures in every cell[76,77], an alternate possibility was that enhancer-promoter contacts were relatively dynamic, and thus an enhancer could switch between driving different promoters on the minutes-to-hours timescale on which transcriptional bursts occur. In the human beta-globin system which we studied, the same enhancer, known as the globin locus control region (LCR), is required for the transcription of both adult and fetal beta-type globins, which in culture can be transcribed in the same cell population. We examined the transcription of fetal and adult globin in primary human red blood cell precursors, and found that these two genes could be transcribed simultaneously by the same gene copy in the same cell (i.e. one of the 2 copies in a G1 cell, 4 copies in a G2 cell). Statistical analysis showed that these gene copies were cotranscribed slightly less than would be predicted if their transcription was statistically independent. (Moreover, our measurements of transcriptional bursts using RNA FISH in fixed cells had by necessity a slight lag, since transcription sites were detectable for around 5-10 minutes after transcription was blocked (see Figure A2.1), and thus true cotranscription of the two genes was likely less than what we measured by RNA FISH.) We concluded that we could observe a slight effect of promoter competition for enhancers. In dramatic contrast, Fukaya and colleagues found that when they created a transgene where an enhancer was placed equidistant between two promoters, these promoters were actually highly likely to transcribe simultaneously. This same group went on to show that, in fact, an enhancer could drive simultaneous transcription between target promoters on different chromosomes in *Drosophila*, one promoter on the same chromosome as the enhancer and one on the homologous chromosome. This observation was very surprising, and different from our findings in mouse and human systems. Moreover, this observation has been cited extensively to support a phase-separation model of transcription in which active enhancers and promoters condense together (in effect, a fancy modern version of the rosette model), so it may be important to investigate this finding further.

Several possibilities for these different observations of enhancer-multiple-promoter transcription exist. The genes in the Fukaya et al. study were artificial transgenes in *Drosophila*

embryos measured in live cells, while ours were endogenous, non-genetically-manipulated loci in mammalian systems measured in fixed cells. *Drosophila* have been previously observed to have the ability for enhancers to drive transcription on neighboring chromosomes, known as transvection[242], while this is probably not common in mammalian systems. Moreover, transcription in *Drosophila* embryos may be qualitatively distinct from mammalian transcription in the sense that *Drosophila* genes are extremely highly transcribed, and rarely have pauses between bursts, while in mammalian cell culture most genes are transcribed less than 50% of the time[168]. Moreover, a different imaging-based study in *Drosophila* did observed promoter competition in a reporter system[187], suggesting that the finding of coordinated transcription of promoters may even be unusual in that species. Further studies will be required to see if this simultaneous bursting behavior of promoters driven by the same enhancers is specific to the transgene system of Fukaya et al., this developmental stage, or *Drosophila* in general, or alternatively whether other mammalian loci might exhibit such simultaneous bursting behavior and beta-globin is simply an exception.

The finding that enhancer-promoter looping controls burst frequency, the main conclusion of our study and the study of the Levine group, has been further investigated in the two years since these studies were published. One beautiful study in the *Drosophila* system supported this finding and in fact illustrated it more directly: they were able to use genetic engineering and multicolor live imaging to visualize the DNA regions of the enhancer and the promoter of a *Drosophila* transgene, as well as nascent RNA from the transgene[85]. Using this system, they found that enhancer-promoter contact controlled transcriptional burst frequency, and indeed they saw directly that transcription could only occur after the enhancer had contacted the promoter, and then transcription halted when the enhancer region moved away from the promoter in the nucleus. However, other studies have shown conflicting findings. Notably, one preprint that imaged the *Sox2* enhancer region, its target gene and the gene's nascent RNA in embryonic stem cells found that transcriptional bursts occurred without the enhancer approaching the promoter[243]. One possibility is that the regulation of this locus and the beta-globin locus are

fundamentally different in their regulation; a more interesting possibility is that enhancer-promoter looping could be correlated with, but not required for, a different event that is the proximal cause of burst initiation. Such a second event could be nucleosome remodeling, recruitment of Mediator or other activation proteins, or another event. Future work in our lab, described below, will hopefully help disentangle how exactly a transcriptional burst is begun.

4.3 Perspectives on transcriptional burst initiation as a key control point of transcriptional regulation and subsequent studies in the field

My first study identified a molecular regulator of an aspect of bursting, namely that promoter-enhancer contact at the beta-globin locus controlled burst frequency. But this finding raised as many questions as it answered. First, in differentiating erythroid cells, cellular differentiation increased enhancer-promoter contact at the beta-globin locus; in contrast, erythroid differentiation increased both burst fraction and burst intensity of beta-globin transcription. Thus, it was unclear which of the many global or locus-specific changes occurring during erythroid differentiation might be responsible for the increase in beta-globin burst intensity. Moreover, we wondered how such burst frequency modulation by enhancer contact might fit in with the numerous other modes of transcriptional regulation previously characterized[4,8,60]. Previous studies have suggested that enhancers might regulate multiple different aspects of transcription, including polymerase recruitment or polymerase pause release[4,8,72,73]. In effect, identifying that enhancer-promoter contact regulated transcriptional burst frequency still failed to clarify exactly what was happening at a target gene when a transcriptional burst occurred. Relatedly, some studies had theorized that when a transcriptional burst began, it then allowed polymerase to be recruited to a gene, while others had hypothesized that bursting could consist of ‘bursts’ of promoter-proximal pause release[149,156]. On the other hand, many studies that imaged transcriptional bursts ignored the step of pause release altogether, in spite of evidence that it was an important transcriptional checkpoint[177]. Finally, the majority of transcriptional bursting studies

focused on one or a few loci, which often were genetically modified to enable imaging of transcription, while we were well-placed to use RNA FISH to measure the transcription of a wide variety of intact endogenous loci. We set out to resolve how transcriptional burst regulation fit in with polymerase binding and pause release regulation.

We sought to reconcile biochemically-measured Pol II dynamics with bursting behaviors to create a more holistic model of transcription. To this end, we performed both types of measurements on the same cell population. We reasoned that since Pol II ChIP-seq and bursting both measured transcription with orthogonal techniques, they would be expected to correlate in predictable ways. Moreover, we realized that the patterns of correspondence between the two technologies actually provided information about how the underlying transcriptional behavior was regulated. In an ideal world, we would be able to perform live imaging of single Pol II molecules transcribing RNA on single gene copies, and thus directly measure how polymerase recruitment and pause release relate to pulsatile RNA transcription in endogenous, non-genetically-engineered loci. As a substitute, we used the technologies of Pol II ChIP-seq and nascent transcript RNA FISH to infer such dynamics.

We first considered how underlying transcription in a population of cells would be reflected in both Pol II ChIP-seq and RNA FISH measurements. For example, Pol II ChIP-seq is an average of polymerase binding averaged across a population of cells. On any given gene copy, theoretically, we might expect there to be no polymerase at all, or a polymerase paused near the promoter, or a polymerase or multiple molecules of polymerase elongating in the gene body. (Recent studies have shown that given the size of the Pol II initiation complex, only a single Pol II molecule can sit in the promoter-proximal region[156,157], so we built this requirement into our model.) Thus, if we simulated many copies of a gene in various different transcriptional states, averaging all of these together should yield a simulated Pol II ChIP-seq profile. Nascent transcript RNA FISH applied to the same population of gene copies would yield a different type of result: first, it would be able to measure the state of every gene copy in the

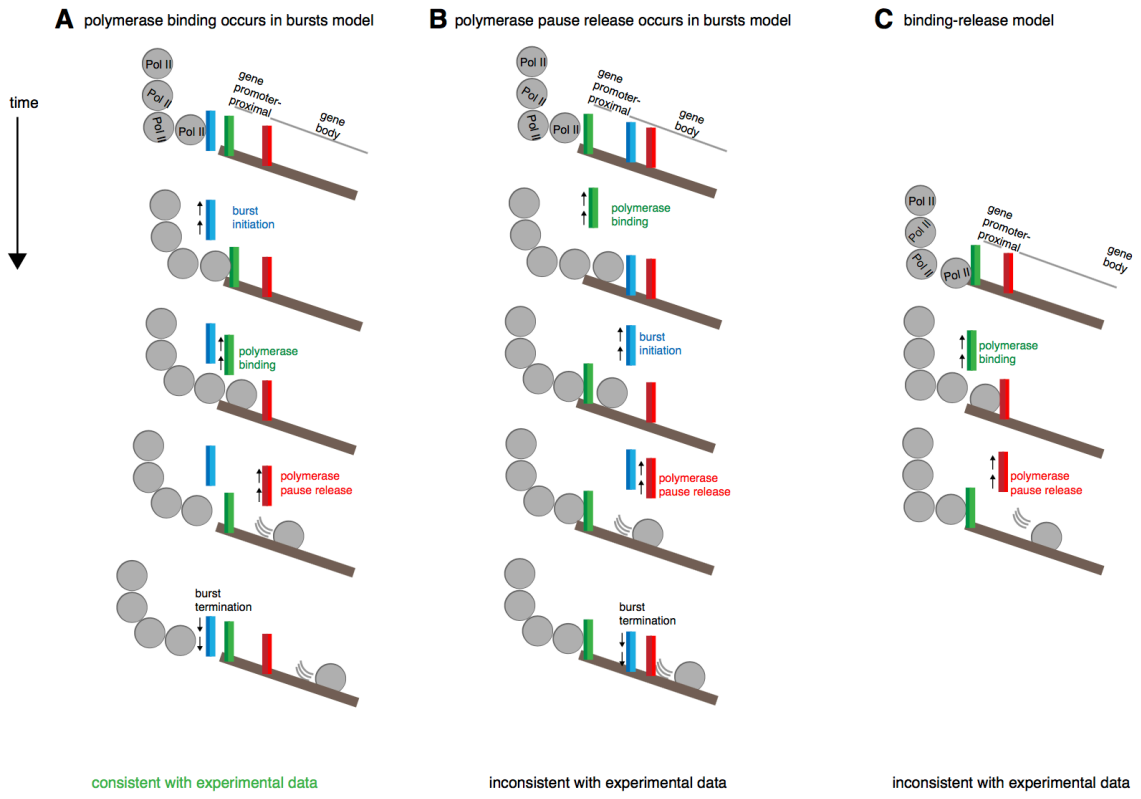


Figure 4. 2 Marble raceway schematic of transcriptional models investigated in Bartman et al. 2018. A) Polymerase binding occurs in bursts model: models are schematized as marble raceways, in which each step of transcription is a ‘gate’ that a polymerase has to pass by in order to reach the next step and ultimately to elongate through the gene body. Each time a polymerase (marble) passes a final gate and rolls to the end, 1 RNA is produced in the model. In this model, the rates of three gates opening can be regulated: the blue gate blocks access to the green gate (i.e. burst has to initiated before polymerase can bind). This model was consistent with our experimental data. B) Polymerase pause release occurs in burst model: in this model, the blue gate is after the green gate but before the red gate (i.e. polymerase can bind anytime, but polymerase can only undergo pause release after burst initiation. This model was not consistent with our experimental data. C) Binding-release model: polymerase has to bind, and then polymerase has to undergo pause release before it can elongate. This model was not consistent with our experimental data.

population, rather than yielding an average. However, both gene copies with no polymerase, or gene copies with polymerase paused at the promoter, would yield no signal in this assay. A

nascent transcript RNA FISH signal should appear at genes with at least one polymerase elongating and producing RNA at the time the cells were fixed, and the fluorescence intensity of the FISH signal should correlate to the number of polymerase molecules that had produced nascent RNAs in the past 5 minutes or so before fixation.

We then used these strictures to create multiple mathematical models of transcription (Figure A3.3, Figure 4.1). These models turned out to be critical for our conclusions: we had started out trying to make intuitive predictions for how transcriptional steps might have behaved, but this became untenable and confusing; in contrast, by setting up models, we had concrete predictions we could accept or reject using experimental data. For each model, we varied the rates of each included step of transcription, and predicted how each type of transcriptional change should be predicted to alter both Pol II ChIP-seq measurements and nascent transcript RNA FISH measurements of transcriptional bursting. We found that, depending on the way that transcription could be changed by varying different rates in the model, different experimental behaviors of these two methods were predicted to appear, suggesting that if we could experimentally measure the possible range of transcriptional measurements, we could see which mathematical model would line up to those possibilities, and thus we could infer which steps of transcription must be regulated in our system.

To complement this modeling approach, we performed a number of transcriptional perturbations in the mouse erythroid cell line used in the previous study, and measured the experimental changes in a number of different genes, in order to try to capture the full range of transcriptional behaviors possible in this system. We eventually examined 15 different genes, including of housekeeping genes, hematopoietic stem cell genes, and erythroid-specific induced genes, and examined their changes in response to a number of perturbations including different transcriptional inhibitors and genetic enhancer dissection. The ability to examine such a range of genes and perturbations set our study apart from previous studies.

We then compared our range of experimental observations to the different possible behaviors predicted by our computational models of transcription. Our previous finding, that burst

fraction could be controlled without any change to burst intensity, turned out to be key to this analysis, along with several other perturbation-gene pairings we found that resulted in independent change in burst fraction (Figure 3.2A, 3.3A). Specifically, previous literature could be taken to suggest that polymerase recruitment and polymerase pause release were the main regulated steps of transcription[4,58]. However, a computational model where only these two steps could be modulated predicted that burst fraction and burst intensity would always change in tandem, and thus such a model was inconsistent with our experimental findings. On the other hand, previous discussion had theorized that bursts might appear because polymerase binding occurred in a pulsatile manner, or alternatively because polymerase pause release occurred in a pulsatile manner[156,241]. We showed that our data were inconsistent with the latter model. Specifically, if polymerase pause release occurred in a pulsatile manner, then burst fraction changes would be predicted to move in tandem with the Pol II traveling ratio, or the amount of Pol II able to be released from the promoter proximal region, but in contrast, we never saw a change in the Pol II traveling ratio in the situations where burst fraction changed without burst intensity (see Figures 3.1F, 3.2A, 3.4A). Thus, applying a range of perturbations to a range of genes, and then measuring them with two orthogonal experimental approaches, was required for our conclusion that burst frequency could be controlled by an independent 'burst initiation' step, and that burst initiation controlled the ability of polymerase to bind the promoter rather than the ability of polymerase to be released from promoter-proximal pausing.

The combination of two experimental approaches to measure transcription helped us to draw conclusions about the underlying transcriptional behavior; analogously, it enabled us to propose explanations for how our conclusions might contrast to previous findings in the field. For example, most studies examining transcriptional bursting relied on the 'telegraph model' of transcription. In this model, the only two regulated steps are burst initiation, and then 'RNA production', which is an unspecified hybrid step combining polymerase recruitment, polymerase pause release, and any other step required to make RNA molecules. This model has been useful for the field because it can satisfactorily model the distribution of messenger RNA molecules in a

population of cells measured by RNA FISH or RNA live imaging[163,166,178]. However, we were able to show that to understand how stimuli regulate transcription, Pol II ChIP-seq is required also (or another similar technique); using such a method revealed that polymerase pause release is a dominant mode of transcriptional regulation, a more molecularly specific finding than regulation of 'RNA production'. We came to this conclusion by showing that most stimuli including erythroid differentiation and BET protein inhibition change Pol II traveling ratio, similar to an inhibitor of pause release and consistent with model-predicted effect of changing pause release, and inconsistent with experimental or model-predicted changes to polymerase recruitment.

Conversely, we showed that an independent burst initiation step upstream of polymerase binding rate can be a regulated step of transcription, for example when promoter-enhancer contact is modulated. Both modeling predictions and inhibitor experiments confirmed that the effect of promoter-enhancer contact was distinct from the effect of changing either polymerase recruitment or polymerase pause release. Our modeling showed that, from the standpoint of population averaging methods such as Pol II ChIP-seq, changes in burst initiation would be indistinguishable from changes in polymerase recruitment. We hypothesize that many studies that have pointed to changes in polymerase recruitment using bulk methods may have actually been measuring changes to burst initiation. Thus, for studies aiming to disentangle transcriptional mechanisms in future it will be important to combine single-cell approaches to distinguish effects on burst initiation from effects on polymerase recruitment.

4.4 Future Directions

Many unanswered questions related to the regulation of transcriptional bursting persist. One interesting future arena concerns the concept of phase separation. There is mounting evidence that condensation of proteins, mediated by disordered domains, might be critical to or at least correlated with active transcription (discussed in the Introduction)[18,34,38]. Indeed, the Fukaya et al. study showing tandem bursting of promoters controlled by the same enhancer has been

used as support for such a model[18]. Whether transcription in general, and the control of transcriptional bursting in particular, requires such a phase separation mechanism will be difficult to disentangle, but perhaps will shed light on new transcriptional mechanisms.

Another future direction directly proceeding from my studies is the molecular identity of the 'burst initiation' step that must occur upstream of polymerase binding. Our studies of the beta-globin locus suggested that this step was controlled by promoter enhancer looping, and this idea is strongly supported by the Gregor result that enhancer-promoter contact is necessary, independent from, and upstream of RNA transcription[85]. However, the *Sox2* study suggests that at the very least promoter-enhancer looping may not directly control burst initiation in all systems[243]. Future work in the Raj and Blobel labs will hope to address this issue. The Raj lab has developed a technique to sort cells expressing an RNA of interest from a population[244]. We will optimize this technique to allow specific enrichment of actively bursting cells from a population of unmanipulated cells, since for even highly expressed genes 50% or less of gene copies are actively transcribing at any given time (for example, see Figure 2.1D). This purification method will allow us to compare biochemical properties of a gene locus of interest in actively bursting and nonbursting cells. For example, we will perform chromatin conformation capture in transcribing and nontranscribing cells: if promoter-enhancer contact directly causes burst initiation, then promoter-enhancer contact should be strongly elevated in transcribing compared to non-transcribing cells in a population. Other candidates for burst initiation causation will be examined, including nucleosome remodeling, promoter binding by a transcription factor, and histone modification.

More broadly, high resolution live imaging is becoming more advanced to the point where it can be informative as to the events regulating transcription[38,187]. It will be exciting to see such techniques advance and be applied to clarify the sequence of events occurring during transcription.

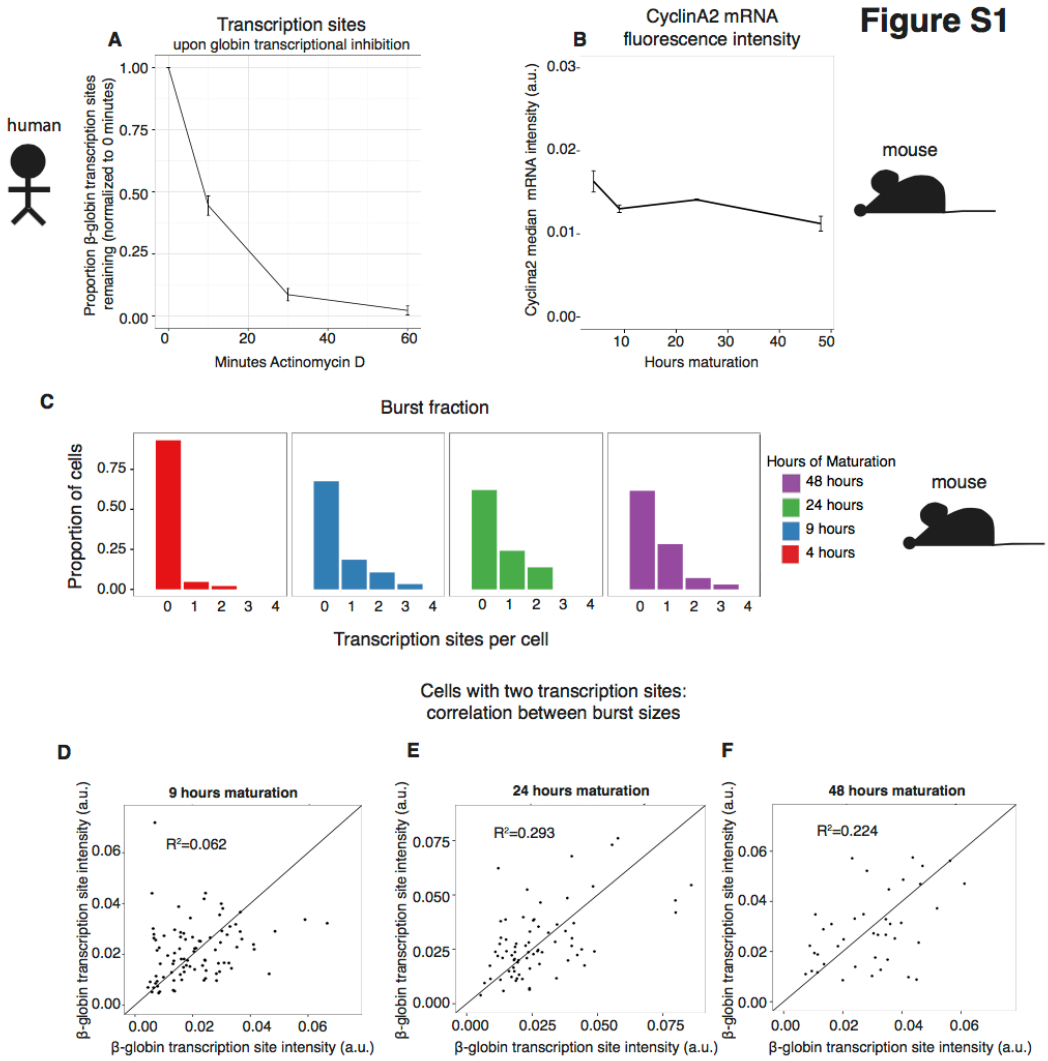


Figure S1, related to Figure 1. A) Effect of actinomycin D treatment on transcription site persistence, n=3 human donors, cell numbers for 0min, 10min, 30min and 60min samples were 105,59,64,120; 118,72,84,96; and 154,111,112,150 for each replicate respectively. B) Fluorescence intensity of cyclin A2 mRNA from the same experiments shown in 1D and 1E, n=3 biological replicates, number of cells for 4, 9,24 and 48h time points were 168, 180, 105, 98; 234, 151, 116,99; and 433,154,130 and 103 for the 3 replicates respectively. C) Proportion of cells with different numbers of alleles transcribing β-globin for one replicate from 1D and 1E. Cell numbers were 234, 151, 116,99 for 4hr, 9hr, 24hr and 48hr time points, representative of 3 biological replicates. D) Correlation of fluorescence intensities in cells with two transcription sites (pooled data from 3 experiments, 9h maturation time point as shown in Figure 1D-1E), number of cells with two transcription sites were 25, 16, and 47 for each replicate. E) Correlation of fluorescence intensities in cells with two transcription sites (pooled data from 3 experiments, 24h maturation time point as shown in Figure 1D-1E), number of cells with two transcription sites were 13, 16, and 43 for each replicate. F) Correlation of fluorescence intensities in cells with two transcription sites (pooled data from 3 experiments, 48h maturation time point as shown in Figure 1D-1E), number of cells with two transcription sites were 7, 7, and 26 for each replicate. All error bars represent SEM.

Figure A2. 1

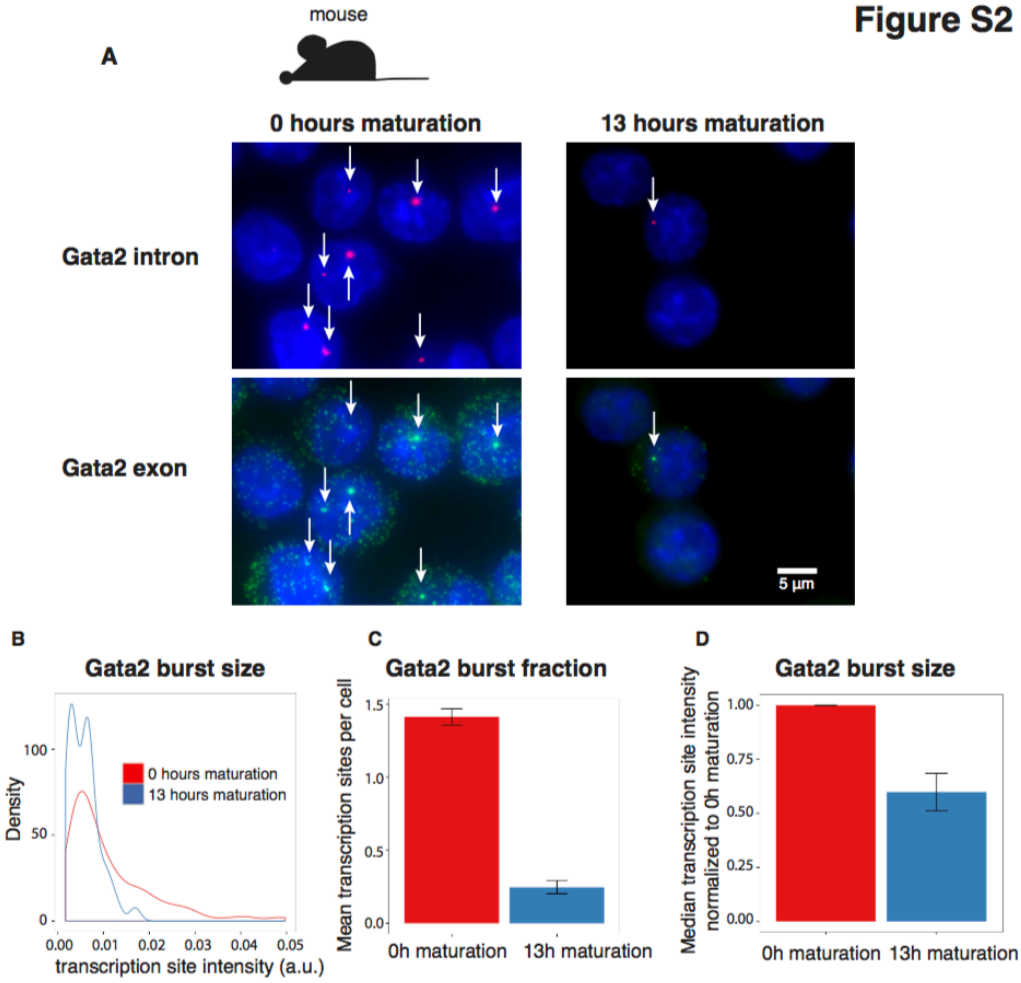


Figure S2, related to Figure 1. A) Representative images of Gata2 transcription sites in G1E-ER4 cells after 0 or 13 hours estradiol addition. Red: Gata2 intron probes; green: Gata2 exon probes. C) Fluorescence intensity of Gata2 transcription sites in G1E-ER4 cells at 0 or 13 after estradiol addition in a representative experiment (number of transcription sites per time point: 0hr 142, 13hr 43; number of cells per time point: 0hr 99, 13hr 131). D) Mean number of Gata2 transcription sites per cell at 0 or 13 hours after estradiol addition. N=3 biological replicates, number of cells in 0 hours condition was 99, 101 and 156, number of cells in 13 hours condition was 131, 120, 119. E) Fluorescence intensity of Gata2 transcription sites in G1E-ER4 cells at 0 or 13 hours after estradiol addition. N=3 biological replicates, number of cells in 0 hours condition was 99, 101 and 156, number of cells in 13 hours condition was 131, 120, 119; number of transcription sites in 0 hours condition was 142, 151, 204, number of transcription sites in 13 hours condition was 43, 21, 29. All error bars represent SEM.

Figure A2. 2

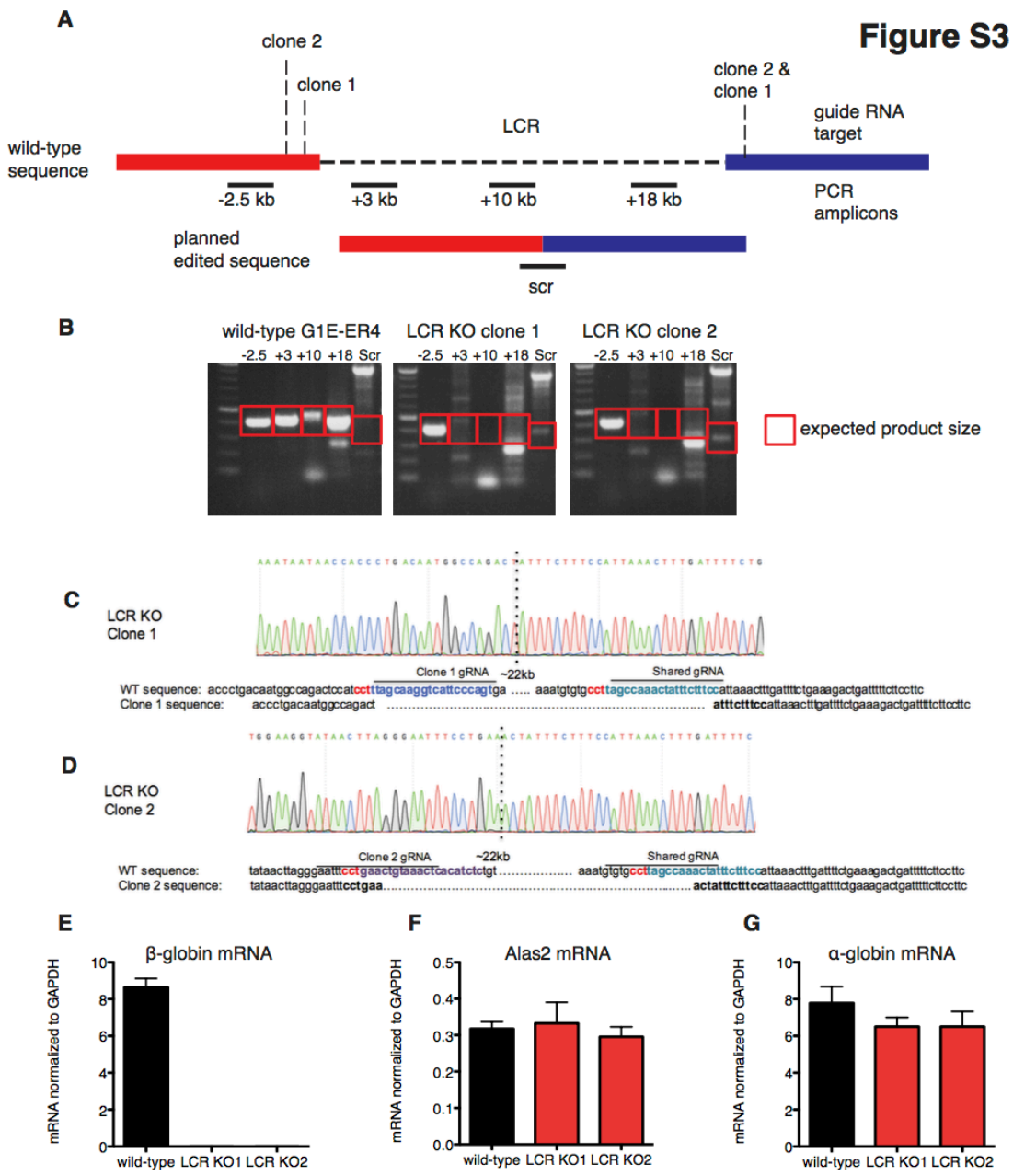


Figure S3, related to Figure 2. A) Schematic of Cas9 gRNA target sites for LCR deletion and PCR amplicon selection for screening of clones. B) PCR screening of WT G1E-ER4 cells and LCR KO clones 1 and 2, confirming deletion of the LCR and formation of a novel junction amplicon. C) Sanger sequencing of LCR region of LCR KO clone 1, showing sequence of deletion. D) Sanger sequencing of LCR region of LCR KO clone 1, showing sequence of deletion. E) qPCR of β-globin mRNA in wild-type and LCR KO clones matured for 24h with estradiol, N=3 biological replicates. F) qPCR of Alas2 mRNA in wild-type and LCR KO clones matured for 24h with estradiol, N=3 biological replicates. E) qPCR of α-globin mRNA in wild-type and LCR KO clones matured for 24h with estradiol, N=3 biological replicates.

Figure A2. 3

Figure S4

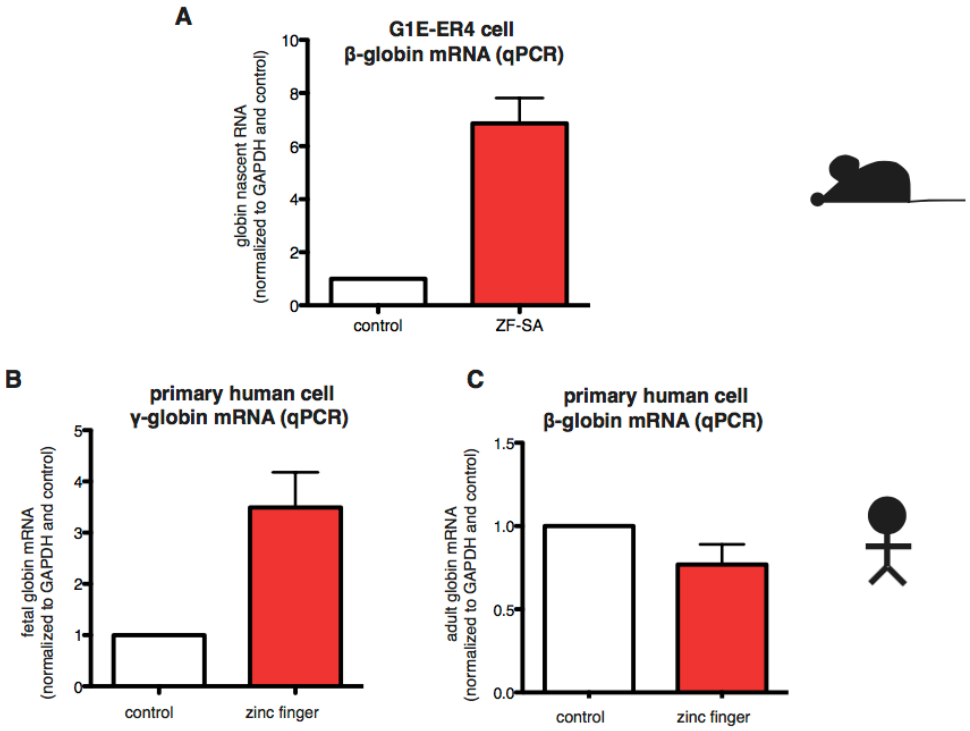


Figure S4, related to Figures 3 and 4. A) β-globin nascent RNA measured by RT-qPCR in G1E-ER4 cells induced for 9 hours with estradiol, infected with control vector or mZF-SA expressing vector, normalized to GAPDH mRNA and control (n=3). B) γ -globin mRNA measured by RT-qPCR in primary human erythroid cells infected with control vector or hZF-SA expressing vector, normalized to GAPDH mRNA and control (n=5 donors). C) β-globin mRNA measured by RT-qPCR in primary human erythroid cells infected with control vector or or hZF-SA expressing vector, normalized to GAPDH mRNA and control. Paired t-test, p=0.13. (n=5 donors). Error bars represent SEM.

Figure A2. 4

Figure S5

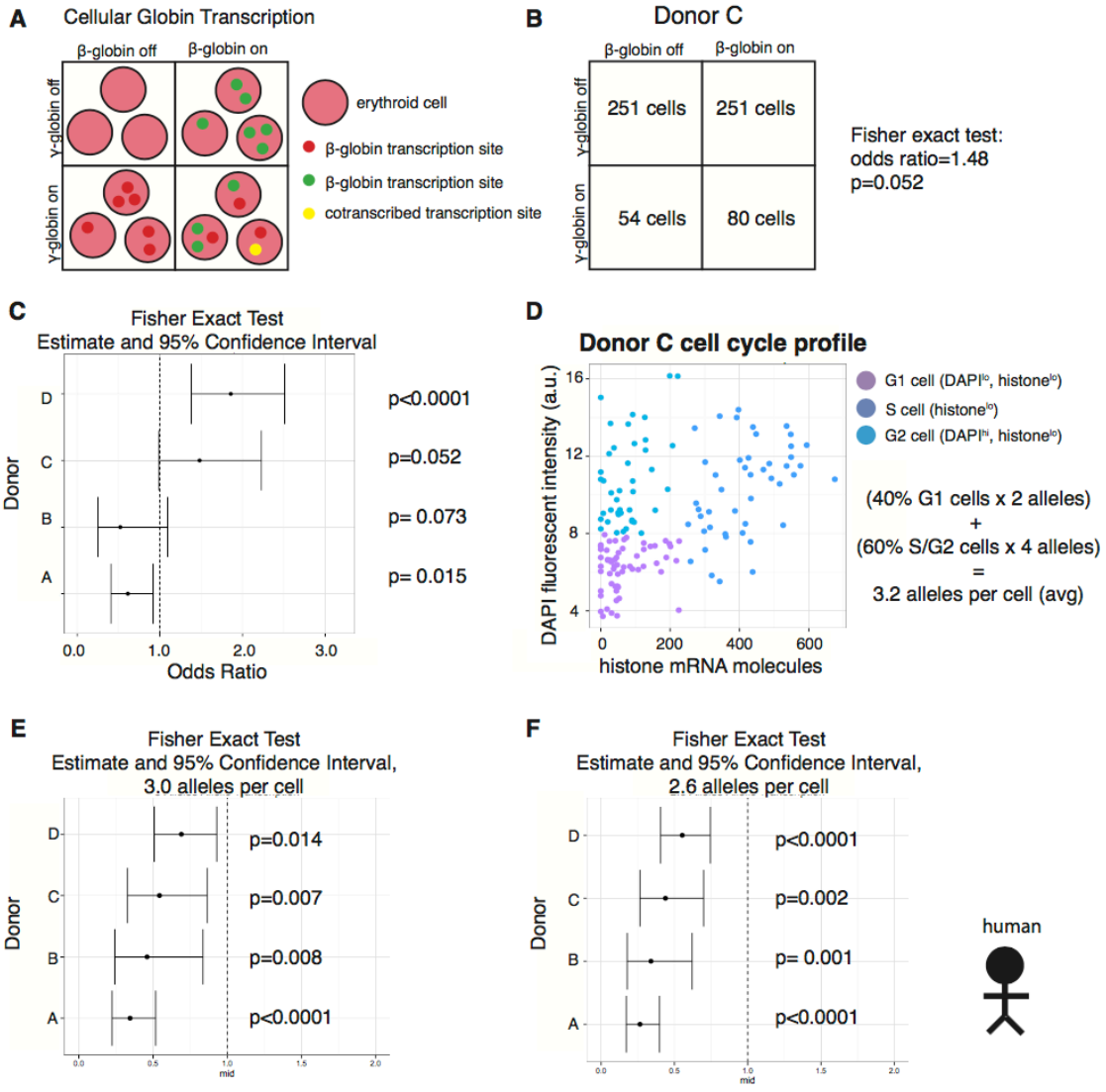


Figure S5, related to Figure 5. A) Model demonstrating how cellular β- and γ -globin transcription was quantitated in primary human cells using RNA FISH data. B) Example of cellular β- and γ -globin transcription from a single donor (636 total cells). Fisher exact test odds ratio=1.48, p-value=0.052. C) Fisher exact test odds ratio and 95% confidence interval for 4 human samples. (Cells per sample: A 423, B 179, C 636, D 1066). D) Cell cycle phase identification of human primary cells, and calculation of predicted average number of globin alleles per cell. (Prediction is required given that RNA FISH does not measure non-transcribing alleles, but this number is needed to calculate statistical independence of co-transcription). E) Fisher exact test odds ratio and 95% confidence interval for 4 human samples for competition between cis-competition of globin alleles, estimating 3.0 alleles per cell. (Cells per sample: A 423, B 179, C 636, D 1066). F) Fisher exact test odds ratio and 95% confidence interval for 4 human samples for cis-competition of γ- and β-globin alleles, estimating 2.6 alleles per cell. (Cells per sample: A 423, B 179, C 636, D 1066).

Figure A2. 5

Timepoint in maturation	Biological replicate	Number of cells	Number of transcription sites	Appears in 1C
4 hours	1	168	35	
9 hours	1	180	100	
24 hours	1	105	81	
48 hours	1	98	30	
4 hours	2	234	21	
9 hours	2	151	75	
24 hours	2	116	60	
48 hours	2	99	51	
4 hours	3	433	186	Yes
9 hours	3	154	152	Yes
24 hours	3	130	156	Yes
48 hours	3	103	83	Yes

Table S1, related to Figure 1.

Figure A2. 6

Cell line	Biological replicate	Number of cells	Number of transcription sites
WT G1E-ER4	1	167	101
LCR KO 1	1	1889	4
LCR KO 2	1	723	1
WT G1E-ER4	2	192	148
LCR KO 1	2	1184	2
LCR KO 2	2	969	7
WT G1E-ER4	3	143	132
LCR KO 1	3	1688	6
LCR KO 2	3	1261	4

Table S2, related to Figure 2.

Figure A2. 7

Cell line	Biological replicate	Number of cells	Number of transcription sites	Used in 3F
G1E+ ctrl vector	1	1311	4	
G1E+mZF-SA	1	407	201	
G1E+ ctrl vector	2	4041	7	
G1E+mZF-SA	2	283	81	
G1E+ ctrl vector	3	1008	2	
G1E+mZF-SA	3	150	63	
G1E-ER4+ ctrl vector, 9 hours maturation	1	171	52	Yes
G1E-ER4+mZF-SA, 9 hours maturation	1	141	171	yes
G1E-ER4+ ctrl vector, 9 hours maturation	2	160	62	
G1E-ER4+mZF-SA, 9 hours maturation	2	124	156	
G1E-ER4+ ctrl vector, 9 hours maturation	3	249	109	
G1E-ER4+mZF-SA, 9 hours maturation	3	137	154	

Table S3, related to Figure 3.

Figure A2. 8

Cells	Biological replicate (donor)	Number of cells	Number of β-globin transcription sites assessed	Number of γ-globin transcription sites assessed
D0 maturation	1	65	27	
D3 maturation	1	56	53	
D0 maturation	2	179	209	
D3 maturation	2	132	223	
D0 maturation	3	423	273	
D3 maturation	3	198	330	
Control vector	1	195	101	73
hZF-SA	1	238	76	219
Control vector	2	126	66	25
hZF-SA	2	129	28	56
Control vector	3	109	45	8
hZF-SA	3	209	115	85
Control vector	4	161	46	30
hZF-SA	4	156	14	61
Control vector	5	159	35	9
hZF-SA	5	135	25	32

Table S4, related to Figure 4.

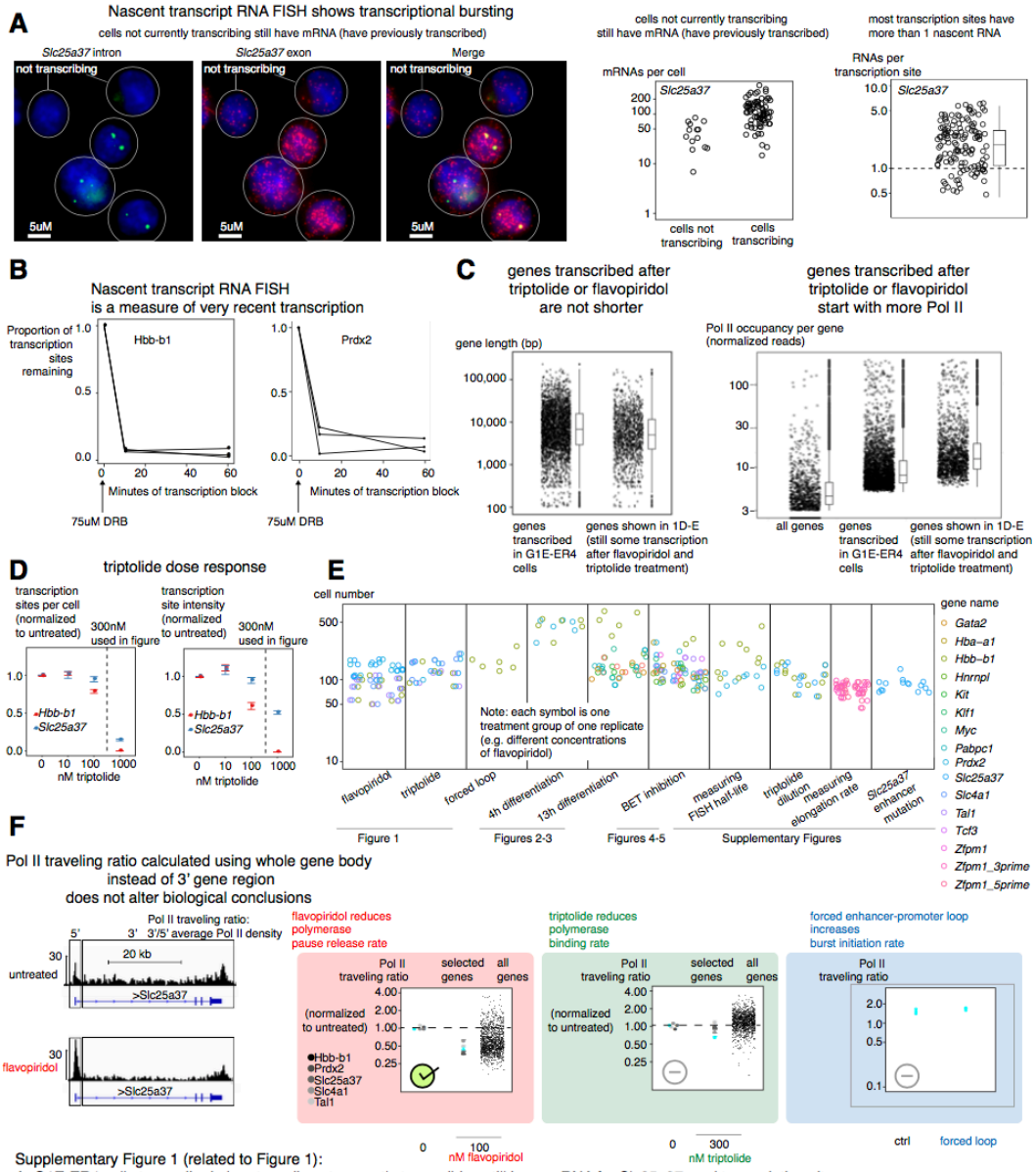
Figure A2. 9

Donor	Number of cells	Number of β-globin- only transcription sites	Number of γ-globin- only transcription sites	Number of cotranscribed alleles	Estimated nontranscribing alleles (estimating 3.2 per cell)
A	423	238	274	31	811
B	179	192	53	17	311
C	636	445	125	19	1442
D	1066	741	255	20	2354

Table S5, related to Figure 5

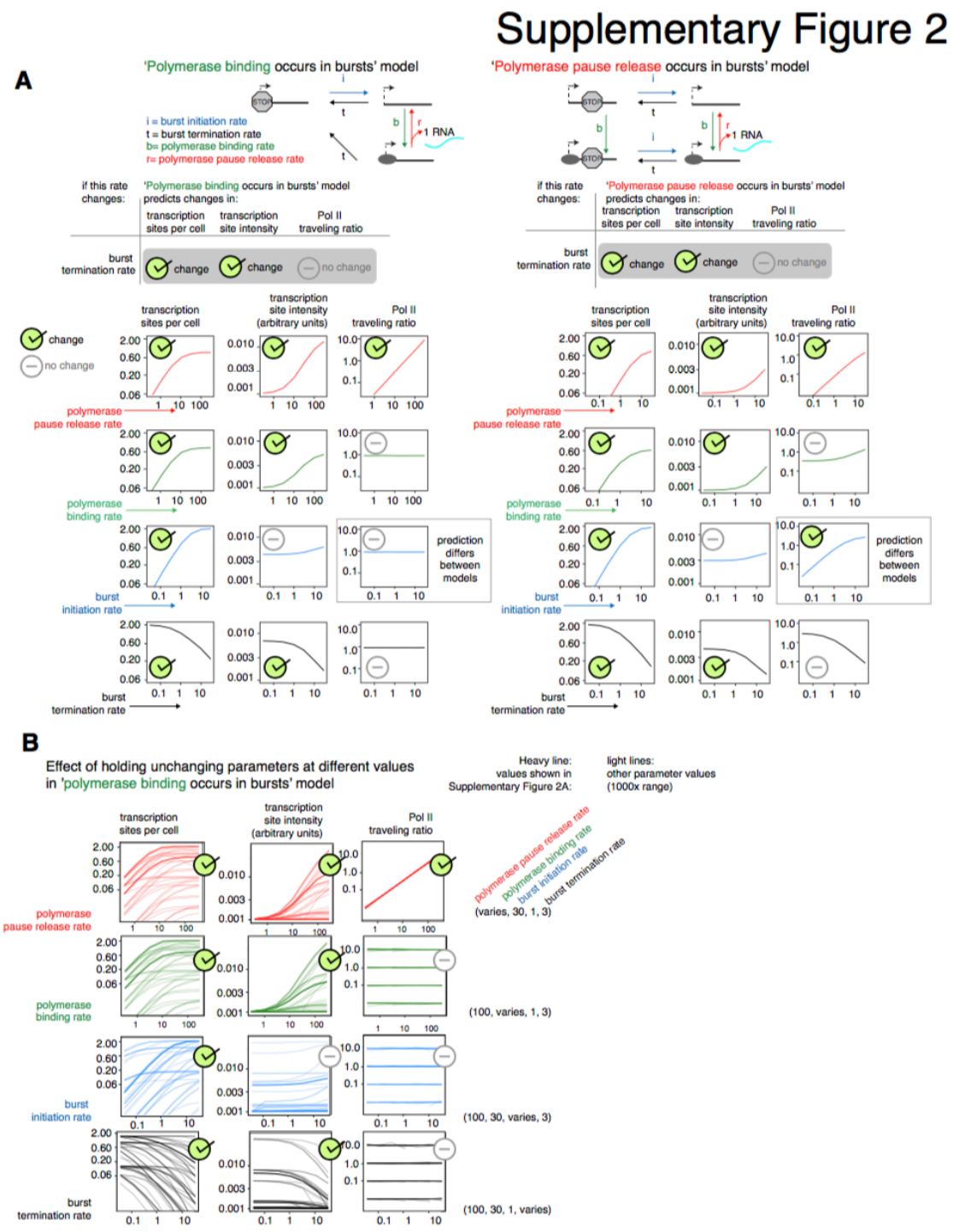
Figure A2. 10

Supplementary Figure 1



Supplementary Figure 1 (related to Figure 1):
A. G1E-ER4 cells transcribe in bursts: cells not currently transcribing still have mRNA for Slc25a37, and transcription sites on average have transcribed more than 1 RNA (as measured by normalizing fluorescence intensity of transcription sites to mRNA intensity.
B. Active-transcribing fraction in response to 75uM DRB, measuring half-life of transcription sites (n=3 biological replicates per gene).
C. Characteristics of genes displayed in traveling ratio analysis in Figures 1D-E: genes are not shorter than average, but do tend to have more Pol II in the untreated setting, so moderate drug dose does not fully block transcription (n=3 biological replicates).
D. Dose titration of triptolide using nascent transcript RNA FISH: 300nM reduced transcription sites per cell and transcription site intensity somewhat but not completely (n=3 biological replicates).
E. Numbers of cells counted per experimental condition per replicate for all FISH experiments in the paper (total= 327 gene-replicates performed, 43790 cells examined).
F. Using gene body instead of 3' end to calculate traveling ratio does not alter biological conclusions: traveling ratios from 1D-F right-side panels recalculated using whole gene body.

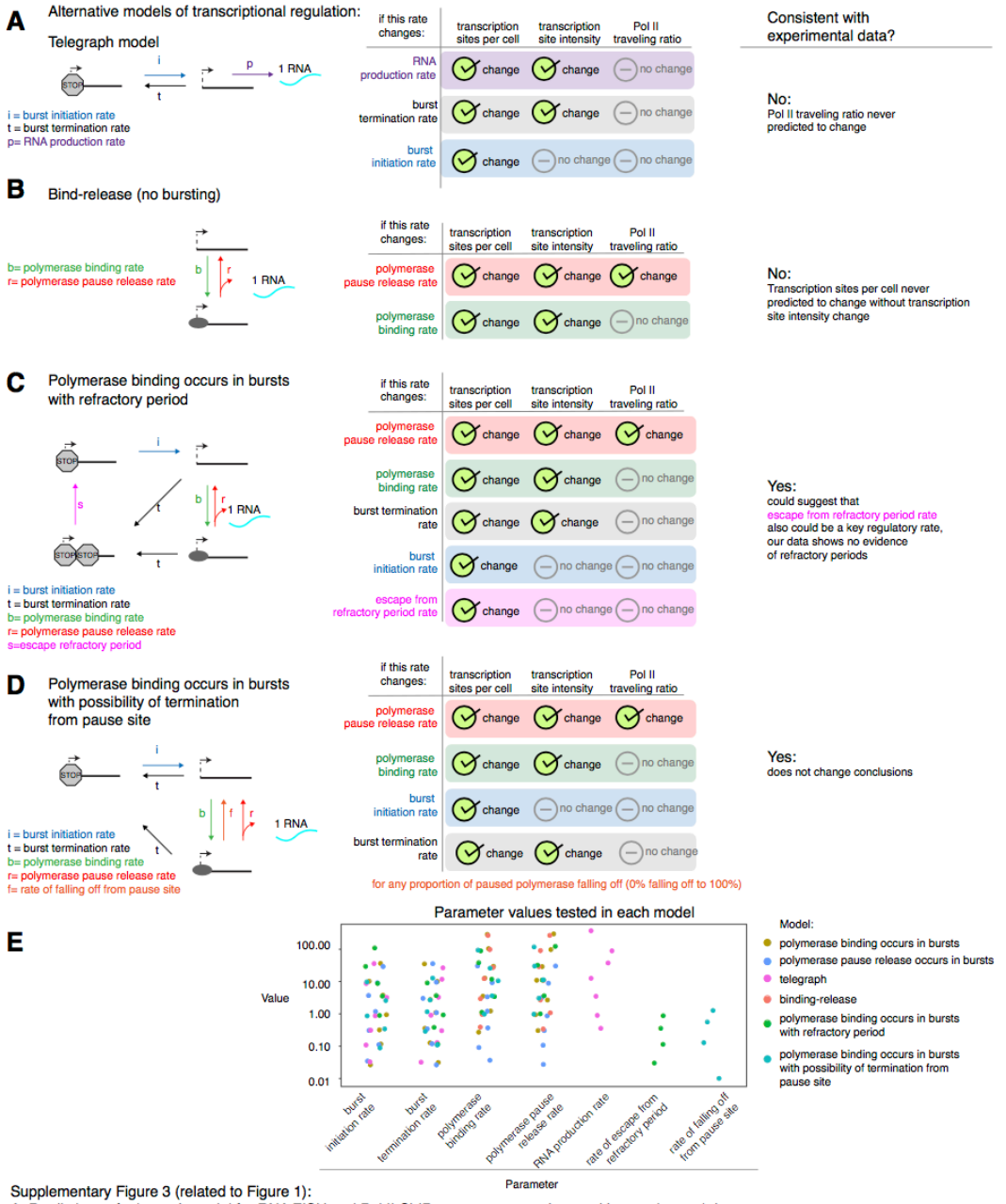
Figure A3. 1



Supplementary Figure 2 (related to Figure 1):
A.Structure and quantitative predictions of the 'polymerase binding occurs in bursts' and the 'polymerase pause release occurs in bursts' models. Rate parameters not labeled as changing in each graph held at the values of burst initiation rate=1, burst termination rate=3, polymerase binding rate = 100, polymerase pause release rate=30. These trends are representative of all parameter values.
B. Model predictions for all parameter values for 'polymerase binding occurs in bursts' model.

Figure A3. 2

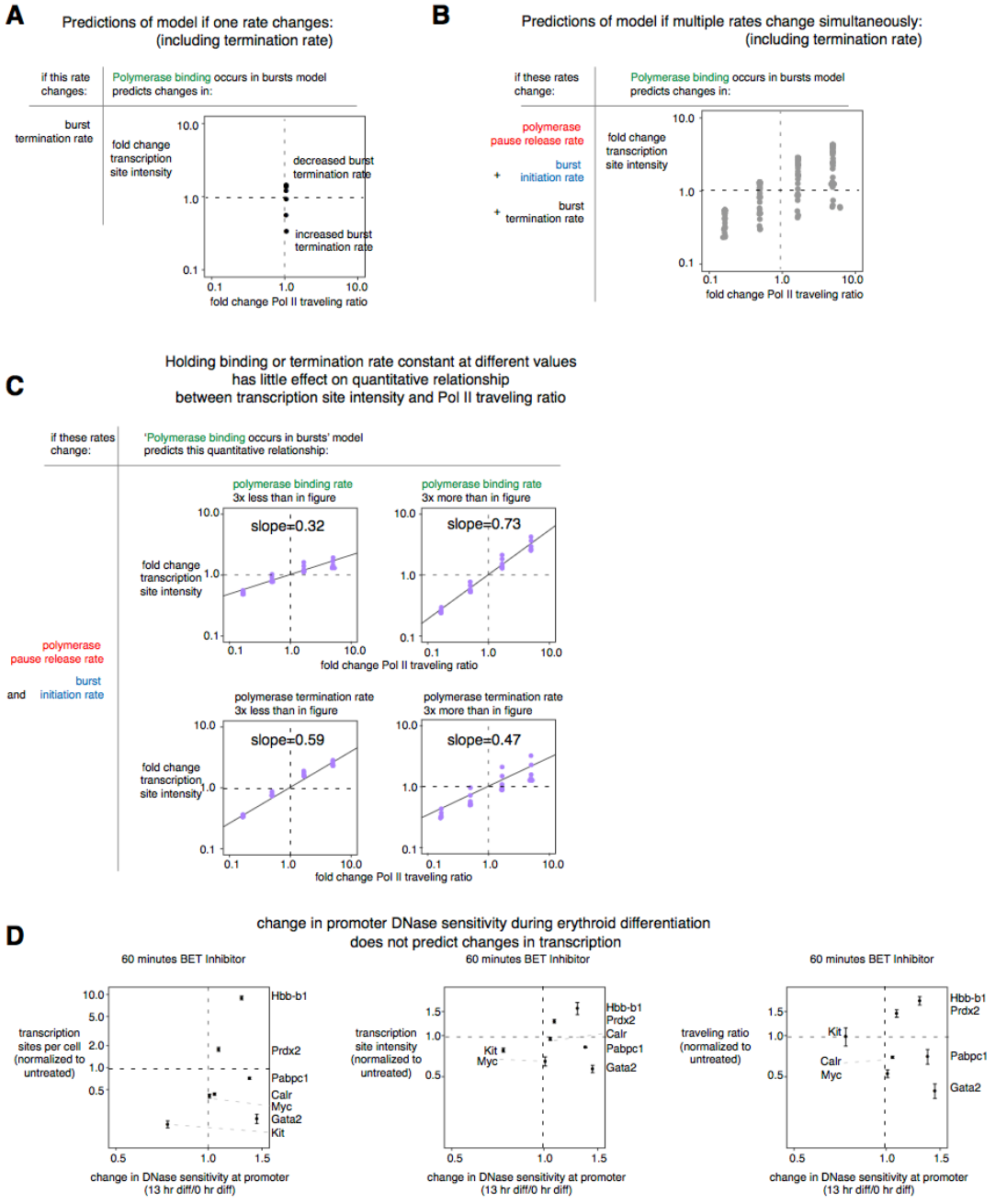
Supplementary Figure 3



Supplementary Figure 3 (related to Figure 1):
A. Predictions of telegraph model for RNA FISH and Pol II ChIP-seq are not consistent with experimental data.
B. Predictions of binding-release model without bursting for RNA FISH and Pol II ChIP-seq are not consistent with experimental data.
C. Predictions of bursty-binding model with refractory period are consistent with experimental data, and could suggest that rate of escape from refractory period is a regulated rate.
D. Predictions of bursty-binding model with possibility of termination from pause site are consistent with experimental data, but adding this possibility gives the same predictions as the bursty-binding model.
E. Parameter values tested for each rate in each model. Each combination of parameters was tested, for example binding-release model, 2 steps, 7 values each= 7^2=49 different simulations.

Figure A3. 3

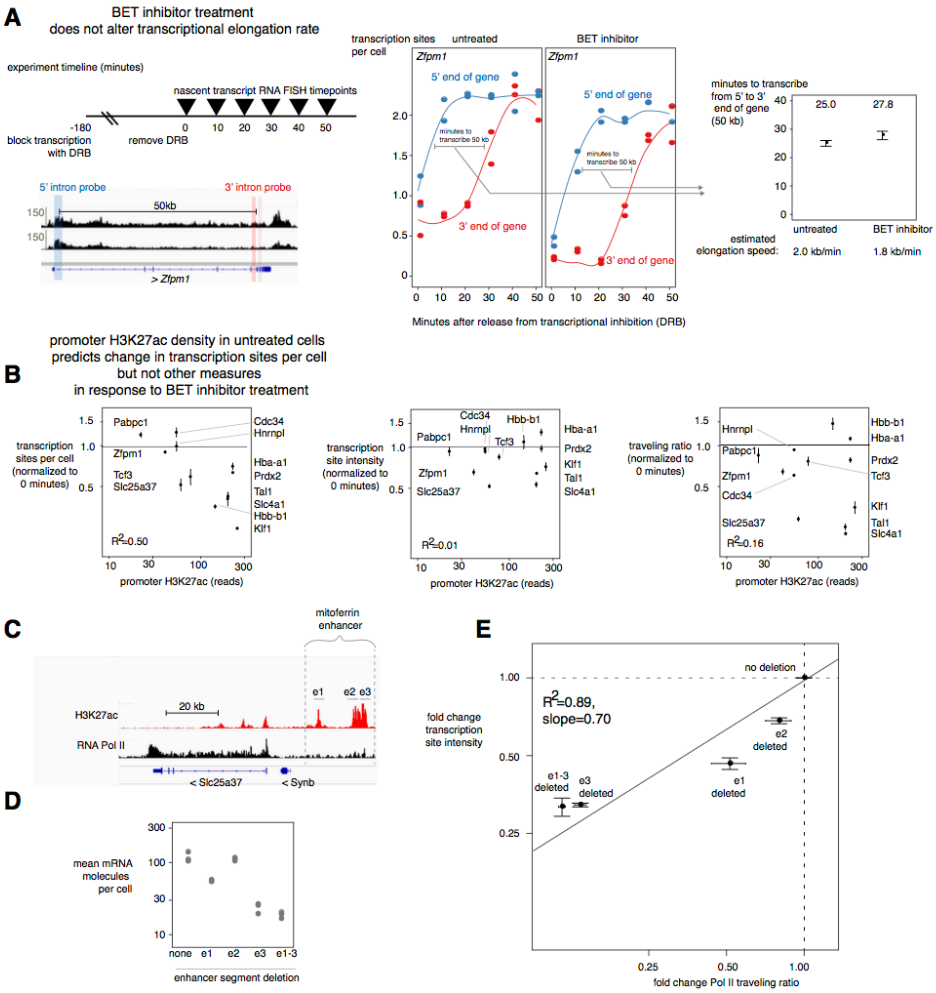
Supplementary Figure 4



Supplementary Figure 4 (related to Figure 3):
A. Effect of changing individual rates on the relationship between transcription site intensity and Pol II traveling ratio, including burst termination rate.
B. Effect of changing termination, initiation and pause release rates simultaneously on the relationship between transcription site intensity and Pol II traveling ratio.
C. Predicted effect of holding polymerase binding rate, or burst termination rate, constant at different levels on the relationship between nascent RNA intensity and traveling ratio when both polymerase pause release rate and burst initiation rate are changed.
D. Change in promoter DNase sensitivity during erythroid differentiation does not predict changes in transcription.

Figure A3. 4

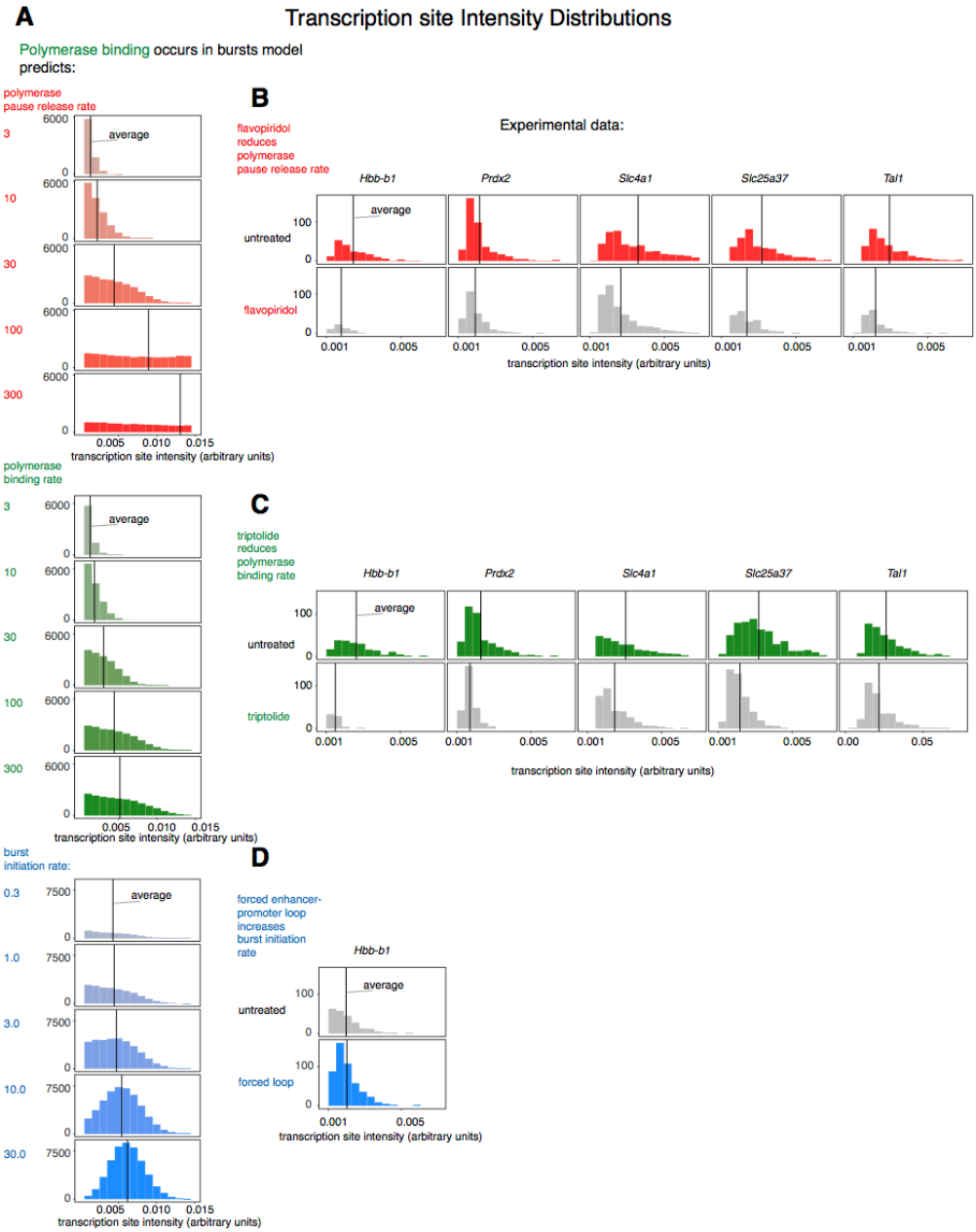
Supplementary Figure 5



Supplementary Figure 5 (related to Figures 4 and 5):
A. BET inhibitor treatment does not change elongation rate for *Zfpm1*, as measured by nascent transcript RNA FISH measurements at 5' and 3' ends of gene at timepoints after release of DRB transcriptional block.
B. Promoter H3K27ac density in untreated cells predicts change in transcription sites per cell but not other measures in response to BET inhibitor treatment.
C. Schematic of *Slc25a37* locus with labeled enhancer segments.
D. *Slc25a37* mRNA counts measured by RNA FISH in unmutated or enhancer-mutant G1E-ER4 cells differentiated for 24 hours (n=3 biological replicates of exon RNA FISH).
E. Effects of *Slc25a37* enhancer mutation on transcription site intensity and Pol II traveling ratio of *Slc25a37* (n=3 biological replicates each of nascent transcript RNA FISH and Pol II ChIP-qPCR).

Figure A3. 5

Supplementary Figure 6



Supplementary Figure 6 (related to Figure 1):
A. Distributions of transcription site intensities predicted by the 'polymerase binding occurs in bursts' model.
B. Distributions of transcription site intensities in cells treated with 100nM flavopiridol (n=3 biological replicates pooled).
C. Distributions of transcription site intensities in cells treated with 300nM triptolide (n=3 biological replicates pooled).
D. Distributions of transcription site intensities in cells expressing looping factor (n=3 biological replicates pooled).

Figure A3. 6

REFERENCES

1. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17,** 183–193 (2016).
2. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci.* **93,** 12355–12358 (1996).
3. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335,** 683–689 (1988).
4. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157,** 13–25 (2014).
5. Weirauch, M. T. & Hughes, T. R. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell. Biochem.* **52,** 25–73 (2011).
6. Damante, G. *et al.* Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. *Nucleic Acids Res.* **22,** 3075–3083 (1994).
7. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172,** 650–665 (2018).
8. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13,** 613–626 (2012).
9. Wallace, J. A. & Felsenfeld, G. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17,** 400–407 (2007).
10. Sandmann, T. *et al.* A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell* **10,** 797–807 (2006).

11. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106,** 159–164 (2015).
12. Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. & Levine, M. S. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 6508–6513 (2016).
13. Swanson, C. I., Evans, N. C. & Barolo, S. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18,** 359–370 (2010).
14. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **11,** 635–643 (2010).
15. Cao, Y. *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* **18,** 662–674 (2010).
16. Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527,** 384–388 (2015).
17. Behera, V. *et al.* Exploiting genetic variation to uncover rules of transcription factor binding and chromatin accessibility. *Nat. Commun.* **9,** 782 (2018).
18. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).
19. Falvo, J. V., Thanos, D. & Maniatis, T. Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83,** 1101–1111 (1995).
20. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129,** 1111–1123 (2007).
21. Lai, W. K. M. & Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* **18,** 548–562 (2017).

22. Archer, T. K. *et al.* Differential steroid hormone induction of transcription from the mouse mammary tumor virus promoter. *Mol. Endocrinol. Baltim. Md* **8,** 568–576 (1994).
23. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Dev. Camb. Engl.* **143,** 1833–1837 (2016).
24. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43,** 264–268 (2011).
25. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).
26. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet. TIG* **32,** 29–41 (2016).
27. Becker, J. S. *et al.* Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol. Cell* **68,** 1023-1037.e15 (2017).
28. Karimzadeh, M. & Hoffman, M. M. Virtual ChIP-seq: Predicting transcription factor binding by learning from the transcriptome. *bioRxiv* 168419 (2018). doi:10.1101/168419
29. Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346,** 1238–1242 (2014).
30. Frietze, S. & Farnham, P. J. Transcription factor effector domains. *Subcell. Biochem.* **52,** 261–277 (2011).
31. Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Dev. Camb. Engl.* **144,** 2719–2729 (2017).
32. Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155,** 1049–1060 (2013).

33. Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168,** 159-171.e14 (2017).
34. Portz, B. *et al.* Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* **8,** 15231 (2017).
35. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20,** 2922–2936 (2006).
36. Boehning, M. *et al.* RNA polymerase II clustering through CTD phase separation. *bioRxiv* 316372 (2018). doi:10.1101/316372
37. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341,** 664–667 (2013).
38. Cho, W.-K. *et al.* RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5,** (2016).
39. Evans, T. & Felsenfeld, G. The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell* **58,** 877–885 (1989).
40. Tsai, S. F. *et al.* Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339,** 446–451 (1989).
41. Ferreira, R., Ohneda, K., Yamamoto, M. & Philipsen, S. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell. Biol.* **25,** 1215–1227 (2005).
42. Wall, L., deBoer, E. & Grosveld, F. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes Dev.* **2,** 1089–1100 (1988).
43. Pevny, L. *et al.* Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349,** 257–260 (1991).
44. Pevny, L. *et al.* Development of hematopoietic cells lacking transcription factor GATA-1. *Dev. Camb. Engl.* **121,** 163–172 (1995).

45. Ciovacco, W. A., Raskind, W. H. & Kacena, M. A. Human phenotypes associated with GATA-1 mutations. *Gene* **427,** 1–6 (2008).
46. Campbell, A. E., Wilkinson-White, L., Mackay, J. P., Matthews, J. M. & Blobel, G. A. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood* **121,** 5218–5227 (2013).
47. Tripic, T. *et al.* SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113,** 2191–2201 (2009).
48. Hong, W. *et al.* FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J.* **24,** 2367–2378 (2005).
49. Weiss, M. J., Yu, C. & Orkin, S. H. Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **17,** 1642–1651 (1997).
50. Welch, J. J. *et al.* Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104,** 3136–3147 (2004).
51. Tsang, A. P. *et al.* FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90,** 109–119 (1997).
52. Vakoc, C. R. *et al.* Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17,** 453–462 (2005).
53. Jain, D. *et al.* Dynamics of GATA1 binding and expression response in a GATA1-induced erythroid differentiation system. *Genomics Data* **4,** 1–7 (2015).
54. Pilon, A. M. *et al.* Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* **118,** e139-148 (2011).

55. Stonestrom, A. J. *et al.* Functions of BET proteins in erythroid gene expression. *Blood* **125,** 2825–2834 (2015).

56. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13,** 233–245 (2012).

57. Ptashne, M. Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.* **30,** 275–279 (2005).

58. Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W. & Kadonaga, J. T. The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell Biol.* **20,** 253–259 (2008).

59. Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* **8,** 424–436 (2007).

60. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 40–51 (2012).

61. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).

62. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).

63. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27,** 2648–2662 (2013).

64. Yashiro-Ohtani, Y. *et al.* Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E4946-4953 (2014).

65. Amano, T. *et al.* Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16,** 47–57 (2009).

66. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144,** 327–339 (2011).

67. Johnson, A. D., Meyer, B. J. & Ptashne, M. Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5061–5065 (1979).

68. Giorgetti, L. *et al.* Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol. Cell* **37,** 418–428 (2010).

69. Arnosti, D. N. & Kulkarni, M. M. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* **94,** 890–898 (2005).

70. Huang, J. *et al.* Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev. Cell* **36,** 9–23 (2016).

71. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48,** 895–903 (2016).

72. Sawado, T., Halow, J., Bender, M. A. & Groudine, M. The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev.* **17,** 1009–1018 (2003).

73. Schübeler, D., Groudine, M. & Bender, M. A. The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 11432–11437 (2001).

74. Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281,** 60–63 (1998).

75. Griffith, J., Hochschild, A. & Ptashne, M. DNA loops induced by cooperative binding of lambda repressor. *Nature* **322,** 750–752 (1986).

76. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35,** 190–194 (2003).

77. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10,** 1453–1465 (2002).

78. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).

79. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149,** 1233–1244 (2012).

80. Song, S.-H., Hou, C. & Dean, A. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol. Cell* **28,** 810–822 (2007).

81. Drissen, R. *et al.* The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* **18,** 2485–2490 (2004).

82. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).

83. Spilianakis, C. G. & Flavell, R. A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5,** 1017–1027 (2004).

84. Williamson, I. *et al.* Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **28,** 2778–2791 (2014).

85. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0175-z

86. Gasperini, M. *et al.* CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am. J. Hum. Genet.* **101,** 192–205 (2017).
87. Ulirsch, J. C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **165,** 1530–1545 (2016).
88. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1074–1077 (2013).
89. Inoue, F. *et al.* A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* **27,** 38–52 (2017).
90. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49,** 825–837 (2013).
91. Berthelot, C., Villar, D., Horvath, J. E., Odom, D. T. & Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* **2,** 152–163 (2018).
92. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).
93. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354,** 769–773 (2016).
94. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
95. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159,** 374–387 (2014).

96. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).

97. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47,** 8–12 (2015).

98. Danko, C. G. *et al.* Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50,** 212–222 (2013).

99. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25,** 582–597 (2015).

100. Noordermeer, D. *et al.* Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region. *PLoS Genet.* **4,** e1000016 (2008).

101. Akhtar, W. *et al.* Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc.* **9,** 1255–1281 (2014).

102. Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E. & Liebhaber, S. A. Bystander gene activation by a locus control region. *EMBO J.* **23,** 3854–3863 (2004).

103. Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* **548,** 343–346 (2017).

104. Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518,** 556–559 (2015).

105. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

106. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

107. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161,** 1012–1025 (2015).

108. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24,** 390–400 (2014).

109. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12,** 283–293 (2011).

110. Sankaran, V. G. & Orkin, S. H. Genome-wide association studies of hematologic phenotypes: a window into human hematopoiesis. *Curr. Opin. Genet. Dev.* **23,** 339–344 (2013).

111. Mahajan, M. C. & Weissman, S. M. Multi-protein complexes at the beta-globin locus. *Brief. Funct. Genomic. Proteomic.* **5,** 62–65 (2006).

112. Deng, W. *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158,** 849–860 (2014).

113. Bender, M. A. *et al.* The hypersensitive sites of the murine β-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood* **119,** 3820–3827 (2012).

114. Wijgerde, M., Grosveld, F. & Fraser, P. Transcription complex stability and chromatin dynamics in vivo. *Nature* **377,** 209–213 (1995).

115. Gribnau, J. *et al.* Chromatin interaction mechanism of transcriptional control in vivo. *EMBO J.* **17,** 6020–6027 (1998).

116. Mahy, N. L., Perry, P. E. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159,** 753–763 (2002).

117. Bartman, C. R. & Blobel, G. A. Perturbing Chromatin Structure to Understand Mechanisms of Gene Expression. *Cold Spring Harb. Symp. Quant. Biol.* **80,** 207–212 (2015).

118. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2,** a003889 (2010).

119. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49,** 1522–1528 (2017).

120. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353,** 598–602 (2016).

121. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49,** 773–782 (2013).

122. Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452,** 243–247 (2008).

123. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163,** 134–147 (2015).

124. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–472 (2012).

125. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543,** 519–524 (2017).

126. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62,** 668–680 (2016).

127. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* **30,** 881–891 (2016).

128. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E6456-6465 (2015).

129. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82,** 45–55 (2017).

130. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360,** 102–105 (2018).

131. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10,** 1297–1309 (2015).

132. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59–64 (2013).

133. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30,** 2657–2662 (2016).

134. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529,** 110–114 (2016).

135. Rodríguez-Carballo, E. *et al.* The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31,** 2264–2281 (2017).

136. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930-944.e22 (2017).

137. Kubo, N. *et al.* Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. *bioRxiv* 118737 (2017). doi:10.1101/118737

138. Goodrich, J. A. & Tjian, R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat. Rev. Genet.* **11,** 549–558 (2010).

139. Roeder, R. G. Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. *Methods Enzymol.* **273,** 165–171 (1996).

140. Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **12,** 89–97 (2002).

141. Nogales, E., Louder, R. K. & He, Y. Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. *Curr. Opin. Struct. Biol.* **40,** 120–127 (2016).

142. Zawel, L. & Reinberg, D. Advances in RNA polymerase II transcription. *Curr. Opin. Cell Biol.* **4,** 488–495 (1992).

143. Zawel, L. & Reinberg, D. Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem.* **64,** 533–561 (1995).

144. Kouzine, F. *et al.* Global regulation of promoter melting in naive lymphocytes. *Cell* **153,** 988–999 (2013).

145. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 129–143 (2015).

146. Cianfrocco, M. A. *et al.* Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152,** 120–131 (2013).

147. Nogales, E., Louder, R. K. & He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu. Rev. Biophys.* **46,** 59–83 (2017).

148. Greber, B. J. *et al.* The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549,** 414–417 (2017).

149. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 167–177 (2015).

150. Henriques, T. *et al.* Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* **52,** 517–528 (2013).

151. Lis, J. T., Mason, P., Peng, J., Price, D. H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14,** 792–803 (2000).

152. Hargreaves, D. C., Horng, T. & Medzhitov, R. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138,** 129–145 (2009).

153. Luo, Z., Lin, C. & Shilatifard, A. The super elongation complex (SEC) family in transcriptional control. *Nat. Rev. Mol. Cell Biol.* **13,** 543–547 (2012).

154. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).

155. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339,** 950–953 (2013).

156. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49,** 1045–1051 (2017).

157. Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6,** (2017).

158. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18,** 1435–1440 (2011).

159. Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352,** 1225–1228 (2016).

160. Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11,** 1455–1476 (2016).

161. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161,** 526–540 (2015).

162. Wada, Y. *et al.* A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci.* **106,** 18357–18361 (2009).

163. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4,** e309 (2006).

164. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297,** 836–840 (2002).

165. Ko, M. S., Nakauchi, H. & Takahashi, N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.* **9,** 2835–2842 (1990).

166. Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14,** 572–584 (2013).

167. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15,** 1263–1271 (2008).

168. Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* **10,** 246–248 (2013).

169. Symmons, O. & Raj, A. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol. Cell* **62,** 788–802 (2016).

170. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280,** 585–590 (1998).

171. Bahar Halpern, K. *et al.* Bursty gene expression in the intact mammalian liver. *Mol. Cell* **58,** 147–156 (2015).

172. Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8,** 165–170 (2011).

173. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol. CB* **16,** 1018–1025 (2006).

174. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2,** 437–445 (1998).

175. Bothma, J. P. *et al.* Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 10598–10603 (2014).

176. Kalo, A. *et al.* Cellular Levels of Signaling Factors Are Sensed by β-actin Alleles to Modulate Transcriptional Pulse Intensity. *Cell Rep.* **13,** 1284–1285 (2015).

177. Senecal, A. *et al.* Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.* **8,** 75–83 (2014).

178. Peccoud, J. & Ycart, B. Markovian Modeling of Gene-Product Synthesis. *Theor. Popul. Biol.* **48,** 222–234 (1995).

179. Munsky, B., Neuert, G. & van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **336,** 183–187 (2012).

180. Tantale, K. *et al.* A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7,** 12248 (2016).

181. Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332,** 472–474 (2011).

182. Corrigan, A. M., Tunnacliffe, E., Cannon, D. & Chubb, J. R. A continuum model of transcriptional bursting. *eLife* **5,** (2016).

183. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123,** 1025–1036 (2005).

184. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361,** 412–415 (2018).

185. Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359,** 1050–1055 (2018).

186. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173,** 665-676.e14 (2018).

187. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50,** 1296–1303 (2018).

188. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529,** 418–422 (2016).

189. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170,** 1028-1043.e19 (2017).

190. Budnik, B., Levy, E., Harmange, G. & Slavov, N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *bioRxiv* 102681 (2018). doi:10.1101/102681

191. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550,** 74–79 (2017).

192. Agarwal, V. & Shendure, J. Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. *bioRxiv* 416685 (2018). doi:10.1101/416685

193. Fu, S. *et al.* Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers. *Nucleic Acids Res.* doi:10.1093/nar/gky753

194. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455,** 485–490 (2008).

195. Molina, N. *et al.* Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20563–20568 (2013).

196. Padovan-Merhar, O. *et al.* Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* **58,** 339–352 (2015).

197. Bender, M. A., Bulger, M., Close, J. & Groudine, M. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol. Cell* **5,** 387–393 (2000).

198. Epner, E. *et al.* The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol. Cell* **2,** 447–455 (1998).

199. Reik, A. *et al.* The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol. Cell. Biol.* **18,** 5992–6000 (1998).

200. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. & Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* **32,** 623–626 (2002).

201. Trimborn, T., Gribnau, J., Grosveld, F. & Fraser, P. Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. *Genes Dev.* **13,** 112–124 (1999).

202. Sankaran, V. G. *et al.* Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460,** 1093–1097 (2009).

203. Raj, A. & van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38,** 255–270 (2009).

204. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12795–12800 (2002).

205. Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8,** 1184–1197 (1994).

206. Sankaran, V. G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322,** 1839–1842 (2008).

207. Goren, A., Tabib, A., Hecht, M. & Cedar, H. DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. *Genes Dev.* **22,** 1319–1324 (2008).

208. Larson, D. R. *et al.* Direct observation of frequency modulated transcription in single cells using light activation. *eLife* **2,** e00750 (2013).

209. Forsberg, E. C. *et al.* Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 14494–14499 (2000).

210. Johnson, K. D. *et al.* Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. *Mol. Cell. Biol.* **23,** 6484–6493 (2003).

211. Letting, D. L., Rakowski, C., Weiss, M. J. & Blobel, G. A. Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Mol. Cell. Biol.* **23,** 1334–1340 (2003).

212. Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5,** 1 (2012).

213. Larkin, J. D., Cook, P. R. & Papantonis, A. Dynamic reconfiguration of long human genes during one transcription cycle. *Mol. Cell. Biol.* **32,** 2738–2747 (2012).

214. Lee, K., Hsiung, C. C.-S., Huang, P., Raj, A. & Blobel, G. A. Dynamic enhancer-gene body contacts during transcription elongation. *Genes Dev.* **29,** 1992–1997 (2015).

215. Guda, S. *et al.* miRNA-embedded shRNAs for Lineage-specific BCL11A Knockdown and Hemoglobin F Induction. *Mol. Ther. J. Am. Soc. Gene Ther.* **23,** 1465–1474 (2015).

216. Neildez-Nguyen, T. M. A. *et al.* Human erythroid cells produced ex vivo at large scale differentiate into red blood cells in vivo. *Nat. Biotechnol.* **20,** 467–472 (2002).

217. Letting, D. L., Chen, Y.-Y., Rakowski, C., Reedy, S. & Blobel, G. A. Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 476–481 (2004).

218. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3,** e02407 (2014).

219. Fremeau, R. T., Lundblad, J. R., Pritchett, D. B., Wilcox, J. N. & Roberts, J. L. Regulation of pro-opiomelanocortin gene transcription in individual cell nuclei. *Science* **234,** 1265–1269 (1986).

220. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469,** 368–373 (2011).

221. Winter, G. E. *et al.* BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Mol. Cell* **67,** 5-18.e19 (2017).

222. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22,** 403–434 (1976).

223. Zhou, Q., Li, T. & Price, D. H. RNA polymerase II elongation control. *Annu. Rev. Biochem.* **81,** 119–143 (2012).

224. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2,** 103–108 (2011).

225. Titov, D. V. *et al.* XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7,** 182–188 (2011).

226. Vispé, S. *et al.* Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. *Mol. Cancer Ther.* **8,** 2780–2790 (2009).

227. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* **62,** 237–247 (2016).

228. Wagschal, A. *et al.* Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* **150,** 1147–1157 (2012).

229. Shi, J. & Vakoc, C. R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol. Cell* **54,** 728–736 (2014).

230. Belkina, A. C. & Denis, G. V. BET domain co-regulators in obesity, inflammation and cancer. *Nat. Rev. Cancer* **12,** 465–477 (2012).

231. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* **16,** 1128–1133 (2009).

232. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).

233. Bartholomeeusen, K., Xiang, Y., Fujinaga, K. & Peterlin, B. M. Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient

release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein. *J. Biol. Chem.* **287,** 36609–36616 (2012).

234. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19,** 535–545 (2005).

235. Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19,** 523–534 (2005).

236. Harper, C. V. *et al.* Dynamic analysis of stochastic transcription cycles. *PLoS Biol.* **9,** e1000607 (2011).

237. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* **23,** 2610–2624 (2009).

238. Dar, R. D. *et al.* Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17454–17459 (2012).

239. Octavio, L. M., Gedeon, K. & Maheshri, N. Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet.* **5,** e1000673 (2009).

240. Hsiung, C. C.-S. *et al.* A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes Dev.* **30,** 1423–1439 (2016).

241. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13,** 720–731 (2012).

242. Bauer, C. R., Hartl, T. A. & Bosco, G. Condensin II promotes the formation of chromosome territories by inducing axial compaction of polyploid interphase chromosomes. *PLoS Genet.* **8,** e1002873 (2012).

243. Alexander, J. M., Guan, J., Huang, B., Lomvardas, S. & Weiner, O. D. Live-Cell Imaging Reveals Enhancer-dependent Sox2 Transcription in the Absence of Enhancer Proximity. *bioRxiv* 409672 (2018). doi:10.1101/409672

244. Rouhanifard, S. H. *et al.* Single-molecule fluorescent amplification of RNA using clampFISH probes. *bioRxiv* 222794 (2017). doi:10.1101/222794