Enhancer control of transcriptional activity *via* modulation of burst frequency

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Jana Tünnermann

Basel, 2024

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Erstbetreuer: Dr. Luca Giorgetti
Zweitbetreuer: Prof. Dr. Marc Bühler
Externe Expertin: Dr. Tineke Lenstra

Basel, den 19.11.2024

Prof. Dr. Marcel Mayor

Dekan

Table of Contents

Acknowledgments

I would not have been able to do this PhD without the support of others. I am grateful for the people who helped me develop professional and personal skills and stood by me while I tried to manage the level of frustration inherent in science. Luca and the FMI in general put together a great group of people who convert frustration into fun and motivation. Together with the people outside FMI, they enabled me to keep my mental well-being in place and at times, to also rediscover the passion for my project.

First, I would like to express my gratitude to Luca for the opportunity to pursue my PhD in his lab. I came to appreciate his guidance and the insight he provided, which kept me on track and moving forward. His dedication to science and commitment to fostering growth have greatly shaped my development as a researcher. Thanks to him, I am now 'there'.

I would also like to thank the members of my thesis committee, Marc Bühler and Tineke Lenstra. I greatly appreciate their time and valuable input. After our yearly meetings, I always left with motivation and inspiration from their scientific knowledge.

A very warm thank you to Gregory Roth. I am grateful I got to work together with you on this project. I learned many new things from you and your calm nature helped me in more than one way. This collaboration certainly is one of my personal highlights of my PhD journey.

A very big portion of my appreciation goes out to the former and current lab members of the Giorgetti group: Gregory Roth, Julie Cramard, Pia Mach, Pavel Kos, Ilya Flymer, Gergely Tihanyi, Julie Cordier, Nessim Louafi, Kristina Makasheva, Elena Testoni, Verena Mutzel, Ewa Piskadlo, Jessica Zuin, Alexis Cornec, Simon Gaudin, Yomna Gharib, Yinxiu Zhan, Josef Redolfi, Tania Distler, Marco Michalski and Mariya Kryzhanovska. Thank you, Julie the first, for saving my cells, all your time and sharing your knowledge on genome engineering. You are the true gem of the lab. Also thank you, Ilya and Pavel, for not telling me about CoPilot and seeing me as a dry lab scientist. Behind all that sarcasm, you guys are actually very supportive, who could have guessed. A very special thanks to all the TIRF users, especially Pia. I would not have tamed the TIRF without our discussions and your advice. I am grateful I got to share this with you. Also special thanks to Gregory, Ilya and Kristina for reading this thesis. To all lab members, thank you for all the scientific and not so scientific discussion. You are a true pleasure to work with.

I would also like to highlight and acknowledge the support of FAIM. Getting pretty pictures is easy, but making them meaningful would not have been possible without your guidance and help. Laurent, Laure and Sabine, you taught me a lot about microscopy and supported me whenever the TIRF did what it did best. Tim-Oliver and Jan, you sparked a passion for image analysis, shared your knowledge and inspired me to learn more.

Next, I would like to acknowledge Hubertus Köhler for cell sorting. Even when I had to repeat experiments, the good part of it was to work together with you again.

I would like to thank the PepNet community. Leonie Ringrose and Marc Rehmsmeier put together an amazing community who I was blessed to be a part of. Especially in times of lock downs and travel restriction, this community managed to make me excited about science and I am grateful for all the scientific exchanges and insights. A very special thanks to Leonie Ringrose for allowing me to work in her lab for a secondment.

My biggest emotional gratitude goes out to the people outside the work world: the Heidelberg climbing crew, the Basel crew, the broken knee people... There are many names and faces who supported me and I shared laughs and tears with. All the ski, climbing and hiking trips made these last years special. I especially would like to acknowledge Jens, Julia and Schniels. You celebrated the highs and made the lows bearable.

My by far biggest gratitude goes out to Christoph. Without you, I would not have eaten, my left foot would have frozen to death and I would still grieve about lost time. You helped me grow as a person and also as a scientist with all the discussions about image analysis. Whenever there was a difficult situation in the last years, you helped me tackle it. You are my family and my home. Naknak.

List of Abbreviations

3C chromosome conformation capture.

Bmal1 brain and muscle arnt-like protein-1.

bp base pairs.

BRD4 bromodomain containing 4.

CBP CREB-binding protein.

CDK7 cyclin-dependent kinase 7.

CDKN1a cyclin dependent kinase inhibitor 1A.

ChIP-seq chromatin immunoprecipitation-sequencing.

CREB cyclic adenosine monophosphate response element-binding protein.

CTCF CCCTC-binding factor.

CTD carboxy-terminal domain.

DNA deoxyribonucleic acid.

DNA-FISH DNA fluorescence in situ hybridization.

DPE downstream promoter elements.

EGFP enhanced green fluorescent protein.

EP enhancer-promoter.

ERK1/2 extracellular signal-regulated kinase 1/2.

eRNA enhancer RNA.

EWS Ewing sarcoma breakpoint region 1.

FACS fluorescence-activated cell sorting.

FLI1 friend leukemia integration 1.

FRAP fluorescence recovery after photobleaching.

FTz Fushi Tarazu.

Gal Galectin.

GCN4 general control nondepressible 4.

gRNA guide RNA.

GTF general transcription factor.

H3K27ac histone 3 lysine 27 acetylation.

H3K27me3 histone 3 lysine 27 trimethylation.

H3K4me1 histone 3 lysine 4 monomethylation.

H3K4me2 histone 3 lysine 4 dimethylation.

H3K4me3 histone 3 lysine 4 trimethylation.

Hi-C high-throughput chromosome conformation capture.

HIV human immunodeficiency viruses.

kb kilobase.

Klf4 Krüppel-like factor 4.

Mb megabase.

MCP MS2 coat protein.

mESC mouse embryonic stem cell.

mRNA messenger RNA.

MYC myelocytomatosis.

Oct4 octamer-binding transcription factor 4.

ORCA optical reconstruction of chromatin architecture.

P-TEFb positive transcription elongation factor b.

p300 tumor protein 300.

p38 tumor protein 38.

p53 tumor protein 53.

PCR Polymerase chain reaction.

PIC pre-initiation complex.

Pol II RNA polymerase II.

PTM post-translational modification.

RNA ribonucleic acid.

SCR Sox2 control region.

scRNA-seq single-cell RNA sequencing.

Ser2 Serine 2.

Shh Sonic hedgehog.

smRNA-FISH single-molecule RNA fluorescence in situ hybridization.

sna snail.

Sox2 SRY (sex-determining region Y)-box 2.

Sp1 specificity protein 1.

TAD topologically associating domain.

TAF15 TATA-box binding protein associated factor 15.

TAG transcription factor activity gradient.

TBP TATA-box binding protein.

TF transcription factor.

TFF1 trefoil factor 1.

TRIM24 tripartite motif-containing 24.

TSS transcription start site.

UTR untranslated region.

XPB xeroderma pigmentosum type B.

List of Figures

List of Tables

Abstract

In mammalian cells, transcriptional regulation strongly relies on *cis*-regulatory elements such as enhancers that can be located at large genomic distances from their target genes. Their action can be broadly defined as the ability to increase the amount of transcription generated at a cognate promoter. The Giorgetti laboratory and others have recently shown that this effect strongly depends on the genomic distance between the enhancer and promoter: enhancer action increases as the enhancer-promoter genomic distance decreases. However, how this is achieved at the level of single cells remains poorly understood. Within a single cell, transcription is a highly dynamic and stochastic process that occurs in bursts of promoter activity separated by periods of transcriptional inactivity. Whether different genomic distances from an enhancer translate in different promoter burst kinetics, and how this relates to the underlying mechanisms of enhancer-promoter communication remains unclear.

To answer these questions, we performed live-cell imaging of nascent transcription in mouse embryonic stem cell (mESC) lines harboring a bottom-up engineered genomic locus allowing to change enhancer-promoter distance without further regulatory confounding effects. By imaging transcription dynamics in multiple cell lines where an ectopic Sox2 promoter is located at different distances from the Sox2 control region (SCR) enhancer, we found that genomic distance from the enhancer controls the frequency, but not the duration, amplitude or size of bursts from the promoter. As a result, genomic distance from the enhancer also impacts the amount of cell-to-cell and temporal variability in transcriptional output from the promoter: a distal enhancer results in larger amounts of transcriptional noise than a proximal enhancer. Using mathematical modeling, we further show that the promoter operates as a multi-state system where the enhancer selectively increases the transition rate from a 'basal' lowly transcribing regime to more transient, transcriptionally active regime, in a way that depends on its distance from the promoter. Our results provide quantitative insight into how an enhancer acts on a promoter in single cells, and into how the large-scale architecture of a locus determine transcription levels as well as temporal and cell-to-cell variability.

Chapter 1

Introduction

1.1 Gene transcription and its stochastic nature

All roughly 200 cell types of the human body contain the same genetic code, yet it is translated into different outcomes. How a cell translates the genetic code into different functions is regulated at many layers, ranging from gene regulation to sensing the function of its environment. The focus of this thesis is gene regulation, which ensures that a gene is transcribed correctly in space, time and to the required quantitative level. At the cellular level, this precise regulation is achieved by modulating bursts of transcription, a random succession of transcriptional active and inactive states, enabling the cell to finely tune transcriptional output [1].

Already in the 1970s, irregular gaps between traveling transcription units on gene bodies have been observed [2]. First direct observations of transcription bursts were made 30 years later in prokaryotes [3] and recently a quasi-genome-wide study of transcription dynamics in human cells has shown that almost all genes stochastically burst [4].

While the molecular processes involved in transcription and their regulatory network have been extensively studied over the past decades, little is known about the kinetic rates at which they occur in living cells. Most likely, only a subset of all molecular processes will represent rate-limiting steps, ultimately shaping transcription bursts and output. To describe the impact of rate-limiting kinetic rates of single molecular processes on the kinetics of the whole transcriptional process, meaning transcription bursts, mathematical models are often applied. In one of the most simple models, transcription bursts can be described by three rates: the rate to switch on a gene, more precisely its promoter, the initiation rate of polymerase while the promoter is on, and the rate to switch off [5, 6] (Fig. 1.1). These promoter rates are then indirectly related to the kinetic rates of molecular processes.

Experimentally, promoter rates are typically inferred from measures informing on burst parameters such as burst duration, interburst duration, burst frequency, burst amplitude and burst size. Burst duration is the time during which transcription is actively occurring. Interburst duration is the time interval between consecutive tran-

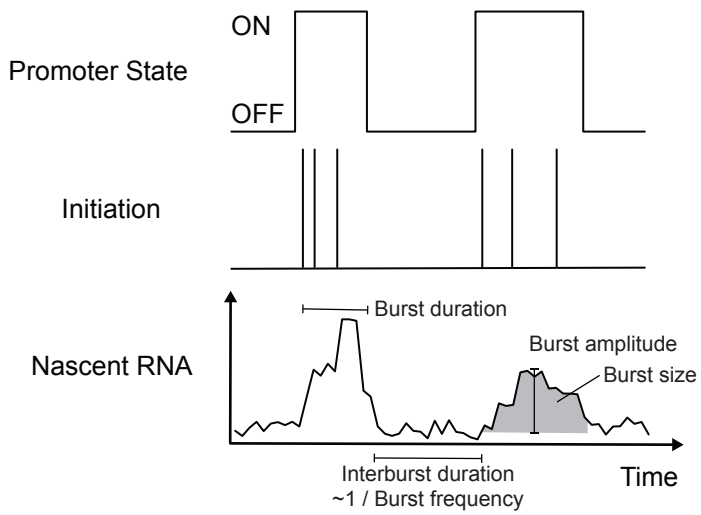

Figure 1.1: Schematic illustrating burst parameters. Top: The promoter can be in OFF-state (non-transmissive) or ON state (permissive) for transcription. Middle: During ON state, RNA polymerase II (Pol II) initiation can take place. Bottom: Nascent RNA accumulation reflects the promoter state and initiation rate. Burst parameters include: burst duration (active transcription time), interburst duration (time interval between transcription events), burst frequency (number of bursts per time window), burst amplitude (peak RNA amount per burst), and burst size (total RNA molecules per burst).

scription events. Burst frequency is inversely related to burst duration and interburst duration and describes the number of bursts within a given time window. Burst amplitude is the peak amount of ribonucleic acid (RNA) produced within a single burst and burst size is the total number of RNA molecules produced during one burst (Fig. 1.1).

Over the years, extensive research has been conducted on different aspects of transcriptional regulation attempting to dissect what steps of the transcriptional process are rate-limiting and how they shape promoter rates, transcription bursts and transcriptional output. Much of this work has concentrated on molecular players, such as transcription factors (TFs) or Pol II, while structural determinants like chromatin folding have received comparatively less attention. However, chromatin folding plays a role in enhancer-driven transcriptional regulation.

Enhancers are distal *cis*-regulatory elements, often located hundreds of kilobases (kbs) up to megabases (Mbs) away from their cognate promoter, and selectively increase transcriptional output [1, 7]. These vast genomic distances are likely bridged by chromatin folding, so that enhancers and promoters come into proximity in the three dimensional space. Moreover, chromatin folding is a dynamic process and mechanisms that regulate its dynamics, such as cohesin-mediated loop extrusion, are speculated to facilitate enhancer-promoter communication, thereby influencing

transcriptional regulation [8].

To investigate the role of chromatin folding and its dynamics in enhancer-promoter communication and transcriptional regulation, my PhD thesis focuses on one specific aspect: genomic distance. Previous work from our group and others has shown that the genomic distance between an enhancer and promoter correlates nonlinearly with population-averaged transcriptional output [9, 10, 11, 12]. Therefore, the main goal of my thesis is to relate these findings to single-cell-resolved burst dynamics.

In the first part of the introduction I will examine the molecular players involved in shaping transcription burst and promoter rates, with an emphasis on the biochemical aspects of enhancer-promoter communication. After explaining how the genome folds on the scale of enhancer and promoter, in the second part, I will explore the structural components of enhancer-promoter communication and how that may modulate transcription bursts.

1.2 Molecular determinants of transcriptional regulation and their dynamics

Transcription refers to the process of synthesizing RNA from a deoxyribonucleic acid (DNA) template. In eukaryotes, molecular steps necessary for this are the binding of TFs to specific DNA sequences, the formation of the pre-initiation complex (PIC) with the help of TFs and co-factors like mediator, recruitment of Pol II by PIC, the actual synthesis of RNA by elongating Pol II and ultimately termination and release of a newly synthesised RNA molecule [13]. Originally, it was thought that the transcription bursts come from the fact that most TFs are present in low copy numbers and their binding is inherently stochastic [14, 15].

However, a growing body of evidence shows that almost all steps of transcription can be modulated with distinct effects on burst parameters and ultimately transcriptional output. While this section focuses on the molecular aspects of transcriptional regulation that have been extensively studied, namely TFs, co-factors and PIC, Pol II pausing and enhancers, it should also be mentioned that other aspects are of importance, like DNA supercoiling [16, 17] or active repression via polycomb-group proteins [18].

1.2.1 Transcription factor residence time and concentration

Given the stochastic nature of TF binding, it is expected that the kinetics of their binding are closely linked to burst parameters. Indeed, factors such as TF residence time and concentration have been directly associated with these parameters. Since TF residence time and concentration can be independently modulated, they have distinct effects on burst parameters.

In yeast, TF residence time is directly linked to burst duration. For the *Gal3* gene, Gal4 binding sets the start and end of a burst with its residence time temporally correlated with burst duration [19]. For the *Gal10* gene, cooperative Gal4 binding is linked to longer burst duration, again linking residence time to burst duration [20]. However, the situation becomes more complex in mammals. For instance, residence time of p53 at the *CDKN1a* promoter and synthetic TFs at the *c-Fos* promoter have been associated with burst duration [21, 22], while the glucocorticoid receptor binding affects both burst duration and frequency [23]. For a synthetic promoter, residence time of the related TF only impacts burst frequency [24].

These seemingly contradictory results may be attributed to the different molecular roles of the studied TFs: the impact of residence time will vary depending on whether TF binding is required throughout the entire burst or to initiate it. If TF binding is required for the entire burst, residence time is directly linked to burst duration, as seen with Gal4 binding in yeast [19, 20]. However, TF binding is one of the first steps in a sequence of molecular events and downstream processes may be rate-limiting. Depending on the molecular function of the TF, its binding then can be required throughout the downstream process or only to initiate it, which ultimately affects burst duration, frequency, or both [24]. Taken together, studies focusing on endogenous loci [19, 20, 21, 22, 23] have shown that TF residence time primarily influences burst duration, suggesting it sets the time a promoter remains in an active state. Together with the example from the ectopic reporter [24], residence time can also be decoupled from the promoter active state, showing that the impact of TF residence time will depend on the TF's molecular function.

A similar rationale can be applied to interpreting the effect of TF concentration on burst parameters. Dose-response studies of the *c-Fos* gene with higher levels of phosphorylated ERK1/2 or p38 kinases, as well as the estrogen-responsive gene *TFF1*, show a distinct increase in burst frequency upon dose response [22, 25]. Similarly, p53 concentration has been associated with burst frequency during DNA damage response [26, 27]. For the above mentioned synthetic promoter [24], higher TF concentrations increase burst frequency without affecting burst duration. Unlike TF residence time, these studies suggest that TF concentration primarily regulates the interval between bursts rather than bursts themselves, influencing the transition of the promoter between active and inactive states. However, in case of the TF MYC, higher concentrations had genome-wide effects on burst frequency and duration [28], again highlighting that the impact of TF concentration depends on the specific molecular function of the TF.

1.2.2 Co-factors and pre-initiation complex

Unlike TFs, that bind specific sequences, co-factors, such as nucleosome remodelers, and components of the PIC lack sequence specificity. They are recruited to or stabilized on DNA with the help of TFs. Since the PIC represents a universal plat-

form for Pol II recruitment, studying its function poses inherent challenges due to its genome-wide influence. Nonetheless, specific features of the PIC have been linked to transcriptional regulation.

The PIC is a multi-protein complex that includes general transcription factor (GTF) such as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Pol II and Mediator. These factors assemble at the core promoter and facilitate the initiation of transcription. The Mediator complex plays a central role in controlling the formation and function of the PIC by organizing and positioning other proteins responsible for critical processes such as the opening of promoter DNA, facilitated by the helicase XPB, and the phosphorylation of Pol II's carboxy-terminal domain (CTD) by CDK7 [29].

Mediator, a multi-protein complex itself, carries out its distinct functions such as PIC assembly and transcription reinitiation depending on its composition. PIC assembly is a dynamic process, with rapid changes in the binding of factors as shown in studies using single-molecule imaging. Zhang *et al.* [30] revealed that the GTF TFIIB exhibits rapid binding dynamics during PIC assembly. Nguyen *et al.* [31] further demonstrated that PIC assembly in yeast is short-lived, emphasizing the transient nature of its components and it ability to rapidly change its function.

In case of transcription reinitiation, Mediator and a subset of the transcriptional machinery might reside at the promoter after the initial transcription event, forming a scaffold for the reassembly of a second transcription complex [29]. This allows for multiple rounds of transcription from a single initiation event, with the second round of transcription being initiated faster than the first one [32].

Due to the multi-protein composition of the PIC, individual components can be modulated with distinct effects on burst parameter and dynamics. Tantale *et al.* [33] observed that lower concentrations of Mediator correlate with reduced transcription initiation rates and shorter burst durations, highlighting mediator's role in transcription reinitiation and on the rate transcription can be initiated at a promoter. Furthermore, mutations in the binding site of the TATA-box binding protein (TBP), reducing both its residence time and binding affinity, impaired PIC assembly and subsequently altered burst frequency and transcription initiation dynamics [33, 34].

Further research is needed to fully understand the role of co-factors and PIC components in regulating burst parameters and dynamics. Due to co-factors' and PIC's ubiquitous roles in transcription, it will be particularly interesting to uncover fundamental principles governing how PIC dynamics can be modulated in the context of transcriptional regulation.

1.2.3 Chromatin-mediated regulation

In eukaryotes, DNA is organized into chromatin, where it is wrapped around histone proteins to form nucleosomes. These nucleosomes present significant obstacles to the transcriptional machinery, particularly in the context of TF binding and Pol II

elongation [35]. While I will discuss the impact of nucleosomes on Pol II elongation in Section 1.2.4, this section focuses on their role in regulating TF binding, co-factor binding and PIC assembly within chromatin.

In vitro studies have demonstrated that TF binding affinities can decrease by up to 1000-fold in the presence of nucleosomes compared to naked DNA [36, 37]. *In vivo* indications of altered TF binding affinities due to nucleosomes came from a study on human T lymphocyte-derived cell lines, in which the human immunodeficiency viruses (HIV) promoter was randomly integrated. Here, DNA accessibility, assessed by DNase I sensitivity assays, was highly correlated with burst frequency, with higher burst frequencies observed in open chromatin [38]. Later, direct evidence came from studies in yeast, in which a competition between either the TF Gal4 or PIC components with nucleosomes were found [19, 31, 34]. These findings highlight how the presence of a nucleosome can affect different steps of transcription.

The competition between nucleosomes and non-histone protein binding is further modulated by post-translational modifications (PTMs) of histones. For example, histone 3 lysine 27 acetylation (H3K27ac) and have typically been associates with active, open chromatin and enhanced transcription, whereas histone 3 lysine 27 trimethylation (H3K27me3) is found at closed repressed chromatin [39]. Although it remains a topic of debate whether the processes that set and interpret these PTMs are the cause or consequence of the resulting phenotype, these modifications have been studied in the context of transcription. A full review on PTMs would be out of the scope of this thesis and is reviewed elsewhere [39, 40, 41]. Here, I will briefly discuss H3K27ac in context of transcriptional regulation to showcase the complexity of this topic.

The transcription co-factors p300/CREB-binding protein (CBP) set and recognize H3K27ac with the ability to also acetylate proteins other than histones [42]. Presence of H3K27ac and p300/CBP was linked to active transcription, whereas H3K27ac alone does not exhibit the same effect [43, 44]. Hence, the impact of H3K27ac in transcriptional regulation has to be interpreted in the context of its functional role and was highlighted in several studies. Viñuelas *et al.* [45] demonstrated that histone acetylation, including H3K27ac, reduces nucleosome stability, leading to higher burst frequencies. Rodriguez *et al.* [25] studied the estrogen-responsive *TFF1* gene and observed that inhibiting the H3K27ac reader protein, TRIM24, resulted in smaller transcription bursts but with the same frequency. Nicolas *et al.* [46] explored the Bmal1 promoter, part of the circadian cycle, and found that H3K27ac levels covaried with transcription bursting, suggesting a direct link between acetylation and transcription burst frequency. Chen *et al.* [47] also showed that histone acetylation modulates both the size and frequency of transcription bursts in mouse neurons. Taken together, these studies highlight the impact of PTMs on transcription burst and promoter rates, while at the same time demonstrating the importance of context dependence interpretations. One of the main technical limitations in interpreting the effect of PTMs remains the dissection of their effect on nucleosome stability, nucleosome remodelers, direct or indirect protein recruitment and structural features. All of which are not necessarily exclusive.

1.2.4 Promoter-proximal pausing

Promoter-proximal pausing of Pol II represents another rate-limiting step in transcription and has been studied in the context of burst dynamics. This phenomenon is characterized by the accumulation of Pol II with the first nucleosome downstream of the transcription start site (TSS), typically within 20-100 base pairs (bp) [48, 49, 50]. The exit of this paused state into the actively transcribing elongation state is tightly linked to the phosphorylation of the CTD of Pol II on Serine 2 (Ser2) [50].

Genome-wide studies have revealed that most promoters exhibit significant Pol II peaks close to the promoter, indicating that promoter-proximal pausing is a common feature of transcriptional regulation [51, 52]. The duration of this pausing can vary dramatically, ranging from minutes to hours in mammals and *Drosophila*, implying different mechanism to control Pol II pausing [53, 54, 55, 56, 57].

Although the term "pausing" might suggest that Pol II is stably associated with the first nucleosome, recent studies have challenged this notion. Fluorescence recovery after photobleaching (FRAP) and single-molecule footprinting experiments have demonstrated that the stalled Pol II is highly dynamic, with continuous cycles of recruitment, initiation and early termination rather than stable stalling [57, 58].

In line with these findings, recent live-cell imaging studies have shown that not every Pol II molecule undergoes pausing; instead, only a fraction of Pol II molecules stochastically enter the paused state [59, 60]. This stochastic nature of pausing suggests a more nuanced role in transcription regulation than previously understood, whereas the different steps of Pol II pausing might be individually regulated. Indeed, the integration of single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) with bulk cell assays has revealed that the recruitment of Pol II and pause release can be individually be modulated [61]. So far, factors that have been linked to individual modulation are the promoter sequence and Pol II reinitiation guided through Mediator [59, 33]. In context of the promoter sequene, presense of an INR element in Drosophila, affected the interburst duration distribution, introducing an additional promoter off state that is linked to Poll II pausing [59].

Collectively, these studies highlight the complexity and dynamic nature of promoter-proximal pausing as a regulatory mechanism. Rather than serving merely as a passive checkpoint, pausing plays an active role in fine-tuning transcriptional output, making it a critical area of study for understanding transcriptional regulation.

1.2.5 Enhancers

Enhancers act as a major driver in development and are essential for activating cell-type specific transcription. These distal *cis*-regulatory elements are often placed hundreds of kbs up to Mbs away from their cognate promoter and can still selectively increase the transcriptional output [1, 7]. Since they are able to communicate over vast genomic distances, enhancer-promoter (EP) communication is an active field

of research, trying to elucidate both the nature of the information transferred and the mechanisms through which this communication occurs. Potential mechanism through which EP communication occurs will be discussed in Section 1.4. Here, I focus on the biochemical aspect of EP communication and the molecular players involved.

While enhancers were first described in the 1980s [62], even today it remains technically challenging to annotate them genome-wide. In 2012, the ENCODE consortium [63], defined a list of characteristics to identify putative enhancers based on next-generation sequencing methods: chromatin is accessible as measured e.g. by DNAse I hypersensitivity, sequences are bound by TFs as measured by chromatin immunoprecipitation-sequencing (ChIP-seq), histone marks such as H3K27ac, H3K4me1, H3K4me2 or potentially histone 3 lysine 4 trimethylation (H3K4me3) are present, DNA is not methylated as measured by bisulfite sequencing, the region is actively transcribed into enhancer RNA (eRNA) and they contact TSS as measured by chromosome conformation capture techniques. Using these characteristics, hundreds of thousands putative enhancers were defined. However, only a handful of these putative enhancers were functionally verified up until today, leaving the question open what their defining features are.

The challenge to define enhancers becomes even more apparent when they are compared to promoters: They have common TF binding sequences, share chromatin architecture and are both transcribed. Several studies have highlighted that promoters can act as enhancers on other promoters and enhancers can initiate transcription functioning as promoters (reviewed in [64]). At the same time, genome-wide eRNA transcription levels are lower than promoter-driven gene transcription levels, indicating that promoters might be optimized for transcription initiation [65, 66, 67]. A study in which one EP pair was isolated from each other by insertion of a insulator element, namely a CTCF-binding site, showed that Pol II levels were decreased at the promoter, but not the enhancer, in turn indicating an optimized function for the enhancer as Pol II loading side [68].

Taken together, these findings suggest the need for a revised definition of enhancers and promoters. While they may arise from a shared regulatory spectrum, they are optimized for distinct but complementary functions, ultimately defining them not as separate regulatory elements, but rather as components of the same regulatory spectrum.

1.2.5.1 Potential molecular functions of enhancers

Since enhancers and promoters share common TF binding sites, a central dogma in enhancer biology is that enhancers function by increasing the availability of proteins necessary for the transcriptional machinery at the promoter. This increase is not limited to TFs but extends to co-factors and/or Pol II recruitment, particularly involving p300/CBP (cf. Section 1.2.3) and Mediator (cf. Section 1.2.2) at active en-

hancers [7, 69]. An increase in the number of TFs, co-factors, and Pol II can potentially influence any step of the transcriptional process. Additionally, the effectiveness of an enhancer depends on how it communicates with the promoter (cf. Section 1.4). Although the molecular function of the enhancers is not fully understood, several consequences of their presence have been described.

For example, p300/CBP accumulates at active enhancers, leading to the acetylation of local histones, non-histone proteins, and promoter-associated proteins. This acetylation facilitates the recruitment of BRD4, which subsequently brings the elongation factor P-TEFb, ultimately influencing Pol II pause release [70, 71]. Acetylation is a dynamic process occurring over minutes, a timescale comparable to transcriptional bursting. Notably, the formation of p300 clusters has been linked to increased transcriptional initiation and burst duration [72]. Furthermore, histone acetylation, such as H3K27ac, is not only associated with BRD4 recruitment but also with nucleosome stability and the recruitment of other TFs and co-factors, adding additional layers of complexity to enhancer function (cf. Section 1.2.3).

Enhancers have also been linked to PIC assembly. *In vitro* studies have shown that GTFs, such as TFIIF and TFIIE, primarily accumulate and pre-assemble at enhancers before being recruited to the promoter [73]. Similarly, the Mediator complex stably accumulates at enhancers [29, 74]. Once these Mediator reservoirs interact with promoters [75], the Mediator complex undergoes compositional changes, making its binding more transient at promoters, potentially affecting promoter-proximal pause release [76, 77]. Since Mediator can serve various functions depending on its composition, this further underscores the multifaceted role of enhancers (cf. Section 1.2.2).

These are examples of the types of information that may be conveyed between an enhancer and a promoter. Further research is needed to comprehensively understand how enhancers influence the various molecular stages of the transcriptional process.

1.2.5.2 Enhancer-Promoter selectivity

Most promoters are regulated by multiple enhancers [78] and often enhancers do not regulate the nearest promoter based on genomic distance [79, 80]. Hence, to ensure correctly timed wiring of enhancers to promoters during development, EP communication must operate selectively. How EP selectivity is achieved is an active matter of debate, however it is likely created at two levels: first, by EP proximity in the three dimensional space, which will be discussed in Section 1.4, and second, through chemical compatibility. It is likely that these two mechanisms work together to ensure precise regulation.

For *Drosophila*, distinct classes of chemically compatible EP pairs have been identified, where compatibility is defined by preferential activation. While enhancers within these classes preferentially activate their corresponding promoters, they retain the ability to activate promoters outside their class to a lesser extent. These preferential EP compatibility classes are primarily distinguished by promoter motifs,

such as downstream promoter elements (DPE) or TATA-box motifs [81, 82]. It is hypothesized that TFs recruited by these motifs define chemical compatibility. For example, the TFs Caudal and FTz have been shown to specifically bind DPE-containing promoters, influencing their transcriptional output [83, 84].

More recently, in mammals, two studies utilizing massively parallel reporter assays reached seemingly contradictory conclusions regarding EP chemical compatibility. One study observed broad compatibility with minimal selectivity between 1,000 enhancer and 1,000 promoter fragments of approximately 260 bp in human K562 cells [85]. In contrast, the other study found that more than half of the 556 *cis*-regulatory 450 bp long elements, including enhancers and silencers, tested in mESC exhibited limited compatibility and strong selectivity [86]. The divergence in conclusions is likely attributable to differences in data analysis, as both studies reported that re-analysis of each other's datasets using their respective computational pipelines supported their own conclusions. Although the exact level of chemical compatibility remains unclear, both studies identified some degree of selectivity, leaving the question open of whether this compatibility is relevant at endogenous loci. Future investigations into chemical compatibility in a genomic context, particularly in the presence of chromatin, will be essential to determine whether these trends become more pronounced under native conditions.

1.3 Genome organisation on the scale of enhancer and promoter interactions

The genome is hierarchically organized at different scales, ranging from the formation of nucleosomes at approximately 100 bp to chromosome territories, in which individual chromosomes occupy distinct regions within the nucleus [87, 88]. Although the molecular processes underlying this organization differ across scales, a general principle is that contact probabilities scale with genomic distance according to a power-law decay. In this context, contact probabilities refer to the population-averaged frequency with which two DNA loci are found in close physical proximity in proximity-ligation and sequencing based methods, such as chromosome conformation capture (3C) techniques [89, 90].

At the scale of EP communication, specifically on the sub-megabase scale, the genome is organized into topologically associating domains (TADs). TADs are self-interacting genomic regions where genomic loci preferentially interact with one another compared to loci located outside the region [91, 92, 90]. Enhancers are often placed within the same TAD as their cognate promoters [93, 94] and genes that are co-regulated during development frequently reside within the same TAD [92, 95].

The proteins cohesin and CTCF are found at the borders of TADs [92, 95]. Together with the loop extrusion capability of cohesin [96, 97], they form a central component of a model proposing that TADs are formed through a process called loop

extrusion. In this model, cohesin is loaded onto chromatin and extrudes chromatin bidirectionally, forming a loop that continues to grow until the extruding cohesin complex encounters bound CTCF. At these sites, CTCF halts the extrusion process, potentially stabilizing the formed chromatin loop [96, 98, 99]. This model has been experimentally tested by degrading cohesin subunits and CTCF, demonstrating the loss of TADs following their depletion in 3C experiments (reviewed in [96]).

TADs should be understood as highly dynamic structures. Single-cell high-throughput chromosome conformation capture (Hi-C) experiments and DNA fluorescence in situ hybridization (DNA-FISH) methods have revealed significant variability in chromatin conformations between individual cells, indicating that TAD structures can only be resolved in population-averaged ensembles [100, 101]. Recent live-cell imaging studies have further demonstrated that CTCF-mediated loops are dynamic and short-lived, with average contact frequencies and duration lasting on the order of tens of minutes [102, 103].

Since EP pairs are often located within the same TAD, these domains are generally interpreted as insulating regions where enhancers and promoters preferentially interact with each other, rather than with other genomic loci. In addition, CTCF-bound sites are frequently found near enhancers and promoters [104], leading to the hypothesis that loop extrusion not only gives rise to TADs, but also facilitates EP contacts. Indeed, among many other examples, deletion of a CTCF site near the enhancer in the *Sonic hedgehog (Shh)* locus [105] or near the promoter of the *MYC* gene [106] results in reduced transcription levels.

However, the precise role of loop extrusion and TADs in gene regulation remains elusive. Global degradation studies of cohesin subunits, which reduce cohesin levels on chromatin, or CTCF have shown minimal changes in gene expression while fully disrupting TAD structures [107, 108, 109]. Higher-resolution 3C methods have revealed that although TADs are fully abolished following acute CTCF or cohesin loss, finer sub-kilobase structures, including EP contacts, are largely preserved [110, 111]. This suggests that while TADs may contribute to the robustness and precision of EP communication and transcriptional regulation, they are not strictly necessary for enabling these interactions.

At the same time, some EP interactions are clearly dependent on presence and functional activity of cohesin and CTCF [112, 113, 114, 115, 116]. For example, in macrophages, inducible gene expression relies on cohesin, linking it to myeloid differentiation [112]. Similarly, during neural maturation, a subclass of genes exhibits sensitivity to cohesin loss, with EP genomic distances in this subclass being notably longer than in non-sensitive genes. This suggests that the sensitivity to cohesin depletion may be influenced by the genomic distance between enhancers and promoters [116].

Despite advances in understanding TAD formation, loop extrusion and the roles of cohesin and CTCF, their precise impact on transcription remains unclear. The context dependent sensitivity to cohesin or CTCF loss in transcription suggests a nuanced

role, whereas further research is needed to understand the extend of their impact on EP communication and transcription.

1.4 Enhancer-promoter communication in space and time

In the genome, enhancers have the capability to skip their closest promoter and activate more distal ones [79, 80]. This selectivity potentially arises from chemical compatibility between EP pairs, silencing effects and the their physical proximity in the three dimensional space. Indeed, EP contact probabilities, as measured by 3C techniques, are a powerful predictor of enhancer-driven transcriptional activation [115, 117, 118] and changes in the genomic distance of an EP pair, and hence changes in their contact probabilities, dictate transcriptional output [9, 10, 11, 12].

However, an increase in contact probability does not differentiate between various types of physical proximity, as it is a population and time-averaged measure. It remains unclear whether a direct contact between enhancers and promoters, on the scale of a few nanometers, is necessary to transfer information, or if a certain level of proximity (hundreds of nanometers) is sufficient, e.g. by increasing the local concentration of certain molecules. The distance at which an enhancer potentially functions is often referred to as the 'radius of action'.

In addition to the radius of action, the temporal component of EP proximity may also play a significant role. The duration and frequency of their interactions could influence their ability to transfer information. Interestingly, two recent studies that measured chromatin fiber dynamics found that, for a cutoff distance of approximately 200 nm, CTCF-mediated contacts last for tens of minutes [102, 103], which is on the same timescale as transcriptional bursting.

Ultimately, how an enhancer communicates with its cognate promoter in both space and time is likely linked to the type of information being transferred. Some molecular processes may only function through direct contact, while others may work across larger distances. Determining which aspect is rate-limiting, ultimately dictating transcriptional bursting and overall transcriptional output, is still an active area of research. This leaves open the question of exactly how enhancers communicate with promoters.

1.4.1 Direct contacts

Contact models summarize the idea that the enhancer and promoter need to come into close physical proximity to transfer information, typically within a few nanometers, corresponding to the size of protein complexes such as the PIC (approximately 40 nm [119]). On a temporal scale, these contacts are likely transient [102, 103]. However, it should be noted that stable loop formation has been hypothesized, based on the idea that Mediator creates a stable binding platform between an EP pair [120].

Stabilized binding between an enhancer and promoter in *Drosophila* has indeed been shown to affect transcription initiation [121]. In mammals, forced EP contacts also lead to transcriptional activation [122, 123, 124], with direct evidence showing that EP contact impacts burst frequency [61]. While these studies demonstrate a correlation between transcriptional activation and close physical proximity, they do not account for the dynamic nature of EP contacts.

Keeping EP contacts on a physiological relevant time-scale failed to show a correlation between EP contacts and transcriptional bursting for the endogenous *Sox2* gene, as observed through live-cell imaging that visualized both EP position and transcriptional output simultaneously [125, 126]. Additionally, a study visualizing EP distance in fixed cells using optical reconstruction of chromatin architecture (ORCA) and relating it to transcription via smRNA-FISH also failed to measure a strong correlation between EP distance and transcription. While a weak correlation was observed, it was not more predictive than any other *cis*-contact [127]. However, these results do not provide a definitive argument against the contact models.

Due to technical limitations in microscopy-based approaches, typically resolution is around 200 nm for above mentioned three-color approaches [8] or 50 nm for high-resolution approaches like ORCA [127]. Since we do not fully understand what is communicated between an EP pair, we cannot determine the exact radius of action for direct contacts. Additionally, some EP contacts might be non-functional due to the enhancer not being bound by the molecules it is supposed to transfer. Lastly, if an EP contact is only required to initiate downstream processes, but continuous presence is not necessary, the time between contact and burst initiation will be influenced by additional kinetic parameters, resulting in a delay that follows a distribution rather than being fixed in time. Such models, referred to in the literature as 'hit-and-run' or 'kiss-and-kick,' predict low to no correlation between an EP contact and nascent transcription [8].

1.4.2 Communication at a distance

Two non-exclusive ideas have been proposed to explain how enhancers can function over larger distances beyond the size of a protein complex: either by forming clusters of TFs, co-factors, and Pol II, or by generating a gradient of action, referred to as transcription factor activity gradient (TAG) model.

The idea of clusters originates from imaging studies where clusters of TFs, co-factors and Pol II have been observed at promoters, in proximity to promoters, and/or co-localizing with enhancers [128, 129, 130, 131]. These clusters were shown to be functionally linked to transcription, e.g. by inhibition of cluster formation for the TFs Oct4 and GCN4 which led to reduced transcriptional activation [132].

For Pol II, Sox2, BRD4, and Mediator clusters, a direct link between cluster-promoter proximity and transcriptional initiation has been established [130, 133]. In this context, Du *et al.* [133] further suggested a temporal separation between cluster for-

mation at the enhancer and its interaction with the promoter. Pol II and Mediator clusters could diffuse between the enhancer and promoter, and burst frequency was linked to cluster distance from the promoter, with a cutoff distance of 1 µm defining two regimes of low and high transcriptional activity. Cluster size then influenced both burst size and frequency. It is important to note that clusters do not appear to function without an enhancer being present, suggesting they may serve as a communication platform.

Additionally, in studies where two promoters are co-activated by the same enhancer, spatial separation between the two promoters was observed. This space was bridged by the TF VP16 and Mediator clusters, leading to co-activation of both promoters [134, 135].

The formation of these clusters is a topic of active debate. For the low-complexity domains of TFs such as EWS/glsFLI1, TAF15 and Sp1, it has been shown that they can undergo liquid-liquid phase separation both *in vitro* and *in vivo* [136]. Similarly, liquid-like properties have been observed for BRD4 and Mediator clusters [129]. However, large-scale liquid-liquid phase separation can inhibit transcription, potentially by sequestering molecules away from non-colocalizing promoters [137, 138]. Additionally, differentiating liquid-liquid phase separation from other phase separation mechanisms, such as aggregation, is technically challenging, and it remains unclear whether these distinctions have functional significance in the context of enhancer-promoter communication [139].

An alternative model for enhancer action over larger distances is the TAG model. In this model, enhancer-bound co-activators such as the acetyltransferase p300/CBP serve as a source of acetylated proteins. As these acetylated proteins diffuse away from the enhancer, they are more likely to encounter deacetylases. These two opposing enzymatic activities establish a gradient of acetylated proteins around the enhancer. When the promoter comes into physical proximity with this gradient, it is more likely to encounter acetylated transcription factors and co-factors, thereby facilitating transcriptional activation. In this model enzymatic activity is not limited to acetylation, but can also affect other PTMs like phosphorylation [140].

1.5 Mathematical models to describe transcription bursts

Mathematical models of transcriptional bursting are powerful tools that connect measurable burst parameters to non-measurable, underlying molecular events and their kinetics. In a simplified view, mathematical models describe kinetic rates of rate-limiting steps in the transcriptional machinery, which may involve one or several molecular events. After calibration and validation with experimental data, a mathematical model provides a framework to make new testable predictions. Several models have been developed to explain transcriptional bursting, here I focus on those related to EP communication.

The simplest of these models is the one-state model, in which transcription can be initiated by a single stochastic rate (Fig. 1.2a). If this rate is low, it can result in gabs of single Pol II initiation events leading to fluctuations in transcription, measured as transcription bursts. While this model has been used to describe a few genes, it does not capture the burst behavior of most genes studied to date [141].

Hence, already in the 1990s, it was proposed that burst dynamics are better explained by a two-state model, also known as the random-telegraph model [5, 6]. In this model, the promoter alternates between an OFF-state, where no transcription can occur, and an ON state, from which transcription is initiated (Fig. 1.2b). This model has become widely accepted because it can explain many observed phenotypes and is the simplest model that accounts for complex burst dynamics.

For example, Larsson *et al.* [142] applied a two-state model to interpret changes in gene expression observed through single-cell RNA sequencing (scRNA-seq) data in mESC and differentiated fibroblasts. They found that enhancers primarily modulate the rate to switch ON, affecting burst frequency. Enhancer deletion also led to lower burst frequencies, reinforcing the role of enhancers in regulating this rate.

Similarly, Falo-Sanjuan *et al.* [143] used a two-state model to study Notch signaling in *Drosophila* development. By analyzing a reporter construct containing Notch-responsive enhancer and promoter elements, they demonstrated that Notch activity regulates burst size, as measured by live-cell imaging. This was interpreted as an effect on the rate at which the promoter switches from the ON to the OFF-state.

However, theoretical studies have shown that two-state models can fit static data, such as scRNA-seq or smRNA-FISH, even when the underlying system is more complex [144, 145]. As experimental techniques for measuring nascent transcription dynamics, such as MS2-tagging for live-cell imaging [146], have become more widespread, more complex multi-state models have been developed to better explain experimental results.

For example, Cheng *et al.* [131] used a three-state model to describe the dynamics of the *Klf4* gene in mESC. This model extends the classical two-state model by adding a deep OFF-state to represent cells that do not transcribe during the live-cell imaging window (Fig. 1.2c). They showed further that reducing the genomic separation between an EP pair resulted in higher burst frequency, indicating changes in promoter deep OFF-OFF and OFF-ON kinetics.

Rodriguez *et al.* [25] proposed an even more complex multi-state model for the estrogen-responsive *TFF1* gene, incorporating three promoter states and two RNA pre-processing steps (Fig. 1.2d). The *TFF1* gene is regulated by several enhancer and deletion of one led to lower burst frequency and burst size, affecting the transition between gene states, namely from OFF to ON.

Multi-state models have not only been applied to interpret EP communication but also to describe other systems, such as the HIV promoter [60], Polycomb-associated genes in mESC [18], core promoter motifs in *Drosophila* [59], and the developmental

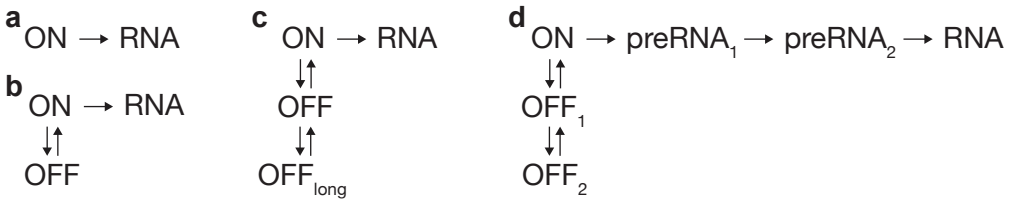

Figure 1.2: Mathematical models used to interpret transcriptional bursting. a. One-state model. **b.** Two-state model or random-telegraph model, e.g. used in [142, 143]. **c.** Three-state model model used in [131]. **d.** Multi-state model model used in [25].

gene *snail (sna)* in *Drosophila* [147]. These studies suggest that multi-state dynamics may be a more widespread phenomenon than previously thought.

1.6 Aim of this thesis

In recent years, significant progress has been made in understanding EP communication. However, the precise mechanisms governing these interactions remain incompletely understood. To advance this understanding, it is essential to generate qualitative and quantitative observations and employ a diverse set of perturbations to functionally test EP communication. While live-cell imaging has provided valuable qualitative insights, such as the presence of clusters and chromatin dynamics, and enhancer deletions or differentiation protocols have contributed to understanding functional aspects, these approaches have limitations.

Specifically, current perturbation methods, like enhancer deletions and differentiation protocols, primarily impact processes related to the molecular function of enhancers, like TF, co-factor and Pol II local concentrations and residence times. These approaches focus on the type of information exchanged between enhancers and promoters, but do not address how this communication occurs at a structural level. Although some live-cell imaging studies have offered insights into the spatial aspects of EP interactions, the underlying structural mechanisms remain unclear. This highlights the need for quantitative approaches to investigate the structural dynamics of EP communication.

Therefore, the aim of this thesis is to elucidate how changes in EP communication dynamics, without altering the molecular composition of the EP pair, influence burst dynamics and output. Specifically, I address the following key questions:

- How does EP genomic distance, and consequently their contact probability and population- and time-averaged proximity, affect transcription bursts?
- How do burst parameters relate to chromatin dynamics?
- Is EP genomic distance linked to cell-to-cell variability?

- Can we construct a mathematical model based on burst parameters that accurately describes our system, to makes testable predictions and gain functional insight?
- Which molecular processes of transcription might be influenced by enhancer distance?

To address these questions, I performed live-cell imaging of nascent transcription in mESC harboring a bottom-up engineered genomic locus that allows the modulation of EP genomic distance without additional regulatory confounding effects. By imaging burst dynamics in multiple cell lines where an ectopic Sox2 promoter is located at different distances from the SCR enhancer, I found that genomic distance from the enhancer controls the frequency, but not the duration, amplitude or size of bursts from the promoter.

Interestingly, while burst frequency is on the order of hours, the contact frequency for neutral genomic loci is typically on the order of minutes, challenging the presumed correlation between contact frequency and transcriptional bursting. Furthermore, genomic distance from the enhancer also influences the extent of cell-to-cell and temporal variability in transcriptional output: a distal enhancer results in higher levels of transcriptional noise compared to a proximal enhancer.

Using mathematical modeling, my colleagues and I demonstrate that the promoter operates as a multi-state system. In this model, the promoter transitions between 'basal' and 'enhanced' three-state regimes, in which it can occupy deep-OFF, OFF, or ON states. These regimes differ in the rate of switching from the OFF to the ON state (k_{on}) and the distance-dependent function of the enhancer affects the rate to switch from the 'basel' to the 'enhanced' regime. Since the rate that is modulated by enhancer genomic distance is followed by several rate-limiting steps, it suggests that early molecular processes of the transcriptional machinery are affected by enhancer proximity. This work is reported in Chapter 2.

Chapter 2

Enhancer control of transcriptional activity *via* modulation of burst frequency

Jana Tünnermann[1,2,*], Gregory Roth[1,*], Julie Cramard[1], Tim-Oliver Buchholz[1], Luca Giorgetti[1]

[1] Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
[2] University of Basel, Basel, Switzerland
* contributed equally to this work

Contribution

My contributions included the conception of this study, performance and interpretation of experiments and the writing of the manuscript. Namely, I conceptualized this study with contributions of Luca Giorgetti. I designed and performed all genome engineering, with the technical help of Julie Cramard. I carried out all other experiments including their analysis and interpretation. I conceptualized and implemented the live-cell imaging analysis pipeline. I implemented the smRNA-FISH image analysis pipeline with the technical help of Tim-Oliver Buchholz. Gregory Roth performed the mathematical modeling. Gregory Roth and I wrote the manuscript with support of Luca Giorgetti.

This chapter's results are unpublished.

2.1 Abstract

In mammalian cells, the cis-regulatory elements named enhancers can be positioned at considerable genomic distances from their target genes, fundamentally enhancing the transcriptional output at cognate promoters. Recent work has demonstrated that the effectiveness of enhancers is closely linked to the genomic distance between the enhancer and promoter: as this distance decreases, enhancer activity increases. However, the mechanisms behind this phenomenon at the single-cell level remain largely unclear. Within individual cells, transcription is a dynamic and stochastic process characterized by bursts of promoter activity interspersed with periods of inactivity. It is uncertain whether varying genomic distances from an enhancer result in different promoter burst kinetics and how these kinetics relate to enhancer-promoter communication.

To investigate these questions, we conducted live-cell imaging of nascent transcription in mouse embryonic stem cell (mESC) lines engineered with a bottom-up genomic locus that allows for the manipulation of enhancer-promoter genomic distances without additional regulatory influences. By analyzing transcription dynamics in multiple cell lines where an ectopic Sox2 promoter is situated at varying genomic distances from the Sox2 control region (SCR) enhancer, we discovered that genomic distance from the enhancer influences the frequency of transcriptional bursts, but does not affect their duration, amplitude, or size. Consequently, the distance from the enhancer also affects the cell-to-cell and temporal variability in transcriptional output: a more distal enhancer results in increased transcriptional noise compared to a proximal enhancer. Furthermore, we propose a mathematical model in which the promoter functions as a multi-state system, where the enhancer selectively enhances the transition rate from a low-transcription 'basal' regime to a more transient, transcriptionally active regime. Our findings offer quantitative insights into the role of enhancers in regulating promoter activity in single cells and highlight how the large-scale genomic architecture of a locus influences transcription levels, as well as temporal and cell-to-cell variability.

2.2 Introduction

In mammalian cells, control of gene expression critically relies on enhancers [4, 14] which determine the tissue specificity and developmental timing of most genes [1, 148]. Enhancers are *cis*-regulatory sequences located outside promoters, which can be positioned at a wide range of genomic distances from the gene that they control and dispersed into *cis*-regulatory landscapes that typically span several hundred kilobases. Despite their central role in gene regulation, the mechanisms that allow enhancers to choose and control their target genes from often large genomic distances remain elusive. Evidence provided by studies based on chromosome conformation capture (3C) methods suggests that enhancers communicate regulatory information to their promoters via direct physical proximity enabled by the three-dimensional (3D) folding of the chromatin fiber [9, 10, 115, 117]. However, the molecular processes that lead from physical proximity to promoter activation remain elusive.

Regulatory events at enhancers and promoters are inherently stochastic and dynamic due to the small numbers of transcription factors and co-factors that can bind each regulatory element inside single cells. This results in discontinuous bursts of promoter activity during which multiple RNA polymerase II (Pol II) are released into productive transcriptional elongation [7, 14, 15]. Yet how stochastic processes at promoters control the kinetics of transcription bursts is poorly understood, especially in mammalian cells where burst dynamics occurs on much slower timescales than transcription factor and co-factor binding. Recent studies have begun unraveling how transcription binding events at promoter regions determine burst initiation and duration in single living yeast cells [19, 20], and highlighted links between transcription factor binding kinetics and burst frequency and duration in *Drosophila* and mammalian cells [21, 22, 23, 25, 26, 27, 28]. However, how regulatory events at enhancers contribute to burst kinetics at promoters is only incompletely understood.

In addition to the stochastic molecular processes occurring at enhancer and promoter sequences, enhancer-promoter (EP) proximity itself likely contributes to determining burst kinetics as an additional stochastic variable. Physical distances between enhancers and promoters are thought to continuously fluctuate in time, so that molecular interactions at their interface can likely only occur within limited time windows during each cell cycle. In mammals, this is strongly supported by recent live-cell imaging studies of the looping dynamics of 'neutral' DNA sequences (that is, not enhancers or promoters) [102, 103] and was directly shown to be the case for an ectopic EP pair in *Drosophila* [12, 121]. Although it is highly likely that stochastic EP interactions contribute to determining burst kinetics, direct evidence for this notion in mammalian cells is lacking [125, 126]. This lack in correlation might be related to technical limitations in spatial resolution (~200 nm as compared to mammalian Mediator/pre-initiation complex size of approximately 42x37x24 nm [119]) and temporal resolution (~1 second [8]), but could also hint towards complex regulatory events at the promoter decoupling physical proximity events from transcription initiation.

It also remains unclear if all enhancers act similarly on target promoters by increasing one specific bursting parameter, or rather each enhancer acts in a different manner, possibly depending on cell type and locus contexts. While some studies have suggested that enhancers in mammals mainly control burst frequency [131, 142], others have shown that enhancers can also affect burst size [25, 124] at their target promoters. Even less is known of how other aspects of the organization of a gene's *cis*-regulatory landscape, such as the genomic distance separating it from its enhancer or the presence of other enhancers nearby, affect promoter bursting kinetics. Yet, understanding how enhancers control burst dynamics is fundamental for a better understanding of how both transcription levels and their cell-to-cell and temporal variability are encoded by genomic sequence.

We and others recently showed that an enhancer's effect decreases as its genomic distance from its target promoter increases [9, 10, 11, 12, 115, 149, 150]. This can be explained in terms of a nonlinear relationship between transcription levels and EP contact probabilities, which are inversely correlated with genomic distance [9, 10, 150, 151]. By interpreting single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) measurements with a simple two-state model of promoter operation [6], we attributed this effect to a decrease in burst frequency when EP proximity events become rarer [9].

Recent studies have questioned the validity of interpreting mammalian promoter burst kinetics in terms of simplistic two-state (ON/OFF) models of gene expression. Indeed, a growing body of research supports the idea of mammalian promoters functioning in the context of three-state models, where the promoter transitions between a repressive state ('deep' OFF state) and a permissive state (OFF state), which only can transition to an active state (ON state) [18, 25, 131]. Some of these studies then supported the hypothesis that enhancers play a role in modulating transitions between OFF states or between the permissive OFF and the ON state [25, 131]. Yet, whether all enhancers and promoters satisfy such 3-state behavior, or rather different models are necessary depending on enhancer and promoter remains unknown [33, 59, 60].

Here, we use a reductionist approach in mouse embryonic stem cell (mESC) allowing to study transcriptional dynamics of the Sox2 promoter while varying its genomic distance with its cognate enhancer in a controlled genomic region that minimizes regulatory and structural confounding effects. This allows us to isolate the effect of EP genomic distance - and hence contact probability, which inversely scales with genomic distance - on burst kinetics while keeping all other regulatory variables constant. Quantitative analysis of live-cell microscopy experiments reveals that burst dynamics at the Sox2 promoter cannot be explained by previously proposed two- or three-state models of promoter operation, irrespective of EP genomic distance. Interpretation of the data using mathematical modeling rather suggests that the promoter stochastically switches between two three-state regimes: a 'basal' regime with low burst frequency and an 'enhanced' regime with higher burst frequency. Changing genomic distance from the enhancer modulates the rate at which the promoter switches from the 'basal' to the 'enhanced' regime, with the effect of only modulat-

ing burst frequency, but not size and duration. Our data also reveal that larger EP distances also lead to higher cell-to-cell variability in the timing and number of transcription bursts, suggesting that the large-scale organization of a regulatory locus plays an important role in controlling transcriptional precision in space and time.

2.3 Results

2.3.1 Enhancer-promoter genomic distance modulates the probability of bursting

To study transcriptional dynamics as a function of enhancer-promoter (EP) genomic distance, we further developed an enhancer mobilization assay that we recently established [9] to enable the live-cell measurement of burst dynamics in the absence of confounding regulatory and structural confounding effects. This assay uses a piggyBac transposon cassette to mobilize an enhancer from a genomic location and reinsert it into many random loci around the initial position within a 'neutral' genomic region.

Specifically, we modified the founder mESC line from Zuin *et al.* [9] where the Sox2 promoter drives expression of a split *EGFP* transcript containing a piggyBac transposon cassette harboring the Sox2 control region (SCR) enhancer. This construct is integrated in a 'neutral' topologically associating domain (TAD) on chromosome 15 that is devoid of other active enhancers and promoters, active or repressive chromatin marks, or CTCF loops. To allow the visualization of RNA produced by this construct, we integrated 24 MS2 stem-loop repeats [146, 152] in the 3' untranslated region (UTR) of the split *EGFP* transgene (Fig. 2.1a). When transcribed, these loops can be visualized upon binding of the bacteriophage MS2 coat protein (MCP) fused to HaloTag, which was stably expressed following random integration of the corresponding expression construct. Upon expression of the PBase transposase, the enhancer transposon cassette is excised and randomly integrated into the genome (Fig. 2.1b). Reconstitution of a functional *EGFP* transcript followed by sorting of EGFP-positive cells enables the generation of clonal mESC lines, in each of which the SCR is located in a different position within the 'neutral' TAD. This approach ensures all clonal mESC lines to have the same MS2 coat protein (MCP)-HaloTag expression levels and the EP genomic distance to be the only varying factor, allowing subsequent live-cell imaging data and assessed burst parameters such as burst frequency, duration and size to be directly compared between clones (Fig. 2.1c).

Using this modified enhancer mobilization assay, we generated 15 clonal mESC lines with ranging genomic distances of 5-232 kb, including one cell line in which the SCR enhancer was reintegrated on a different chromosome from which it exerts no effect on the ectopic Sox2 promoter. To ensure that the modifications do not perturb the relation between transcriptional output and EP genomic distance that we previously measured, we assessed the EGFP protein level of all clonal mESC lines using flow

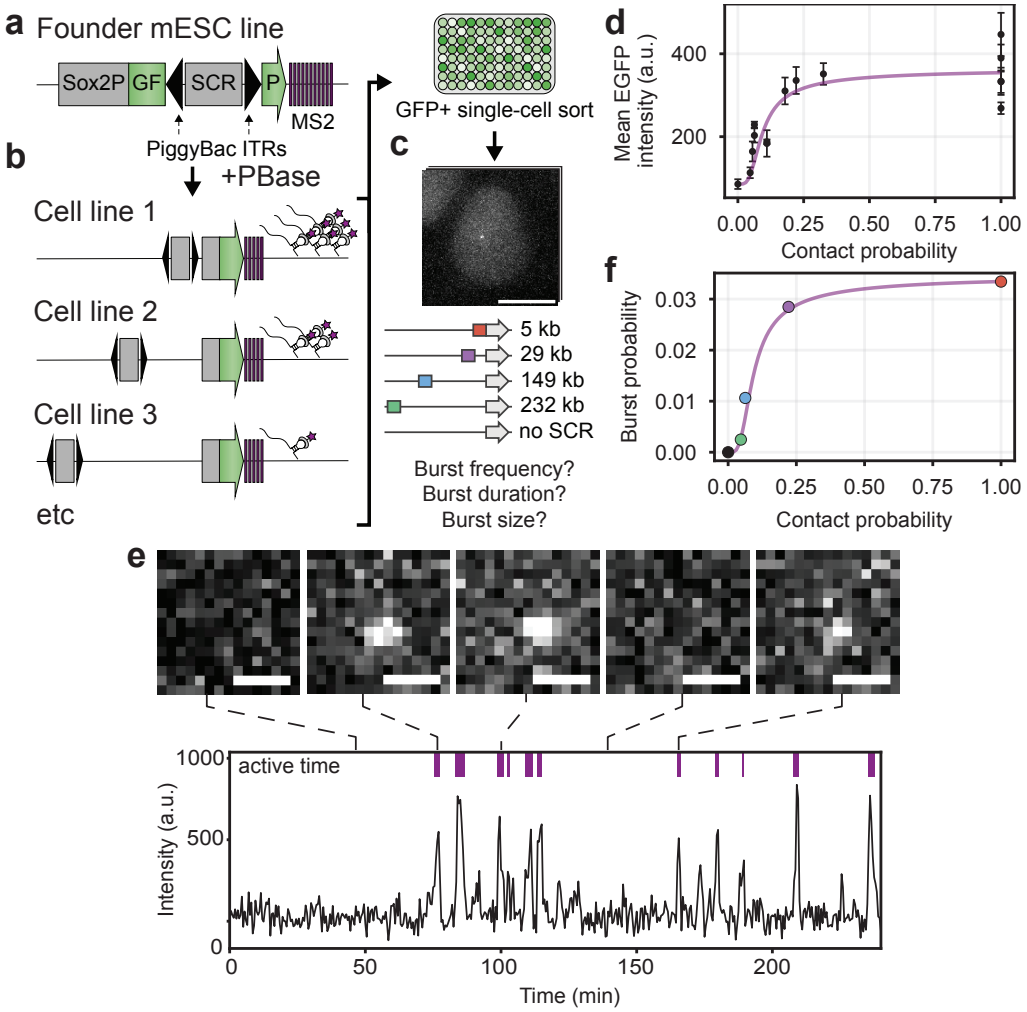

Figure 2.1: EP genomic distance dictates the probability of observing nascent transcription. **a.** Schematic of the transgene: the Sox2 promoter drives transcription of the *enhanced green fluorescent protein (EGFP)* gene split by the piggyBac-SCR enhancer cassette. The 3'-untranslated region (UTR) of the *EGFP* transgene is tagged with 24 MS2 stem loop repeats. ITR, inverted terminal repeats. **b.** Upon piggyBac transposon (PBase) expression, the enhancer cassette is excised and reintegrated elsewhere in the genome, reconstituting the *EGFP* transgene and enabling EGFP+ cells to be sorted into clonal cell lines. **c.** Live-cell imaging is used to measure burst parameters as a function of EP genomic distance for five cell lines with either no SCR present or varying EP genomic distances. Scale bar: 5 µm **d.** Mean EGFP intensities as assessed through flow cytometry in individual EGFP+ cell lines as a function of EP contact probability as measured by capture 3C. Purple trend line originates from Zuin *et al.* [9]. **e.** MS2 intensity trace of one cell from live-cell imaging with example images of nascent transcription spots. Scale bar: 1 µm. Transcriptionally active/inactive times were segmented based on the spot detection algorithm. Active time refers to a spot being detected. *continued on next page.*

Figure 2.1: *continued from previous page.* **f.** The probability of observing a burst in live-cell imaging as a function of contact probability. Burst probability describes the time-averaged probability of observing a burst at a given time point. Purple trend line originates from Zuin *et al.* [9].

cytometry. EGFP protein levels decreased with increasing EP genomic distance and scaled nonlinearly as a function of contact probability, as assessed through capture-3C, (Fig. 2.1d, Extended Data Fig. E.1a,b), making these clonal mESC lines ideally suited to study transcriptional burst dynamics as a function of contact probability.

We then selected a subset of five clonal mESC lines where the enhancer was reintegrated at genomic distances of 5, 30, 149, 232 kb and on a different chromosome for further analysis. In each of them, we acquired one EGFP 10 µm z-stack followed by movies of 10 µm z-stacks every 30 seconds for 5 hours using oblique illumination microscopy to reduce phototoxicity and photobleaching. Average EGFP protein levels as assessed by microscopy were highly correlated to those assessed by flow cytometry, illustrating that imaged cells were representative of the population (Extended Data Fig. E.1c,d). We then quantified the kinetics of nascent transcription by detecting transcription start site (TSS) as diffraction-limited spots in time-resolved image series. Imaging 3-7 technical replications per condition led to between 1000-2500 MS2 intensity traces per condition (example trace in Fig. 2.1e, Supplementary Table E.11). Interestingly, the time-averaged probability of seeing a burst at a given time point revealed the same nonlinear relationship as previously reported for transcriptional levels measured by mature messenger RNA (mRNA) or protein (Fig. 2.1f, [9]), suggesting that the SCR might predominantly control the probability of initiating a burst from the Sox2 promoter.

2.3.2 Enhancer-promoter genomic distance affects interburst duration

We next set out to understand how quantitative burst parameters change as a function of EP genomic distance, and hence contact probability. For this, we calculated the survival probabilities of burst size, burst amplitude, burst duration and the interburst duration in the five mESC lines with different SCR positions (Fig. 2.2).

Burst size (determined as the integrated intensity for the MS2 signal over an active transcription period) and burst amplitude (calculated as the maximum intensity of an active transcription period) showed no notable changes across different EP genomic distances, as confirmed by comparison with survival probabilities, which fell within the 95% confidence interval calculated from bootstrapping (10.000 repeats). We only noted a small trend for burst amplitudes to be smaller for larger EP genomic distances (Fig. 2.2a,b, Extended Data Fig. E.2a,b).

Given that we imaged the cells over finite windows of time, it is possible that some active or inactive periods initiated or finished outside the imaging window ('cen-

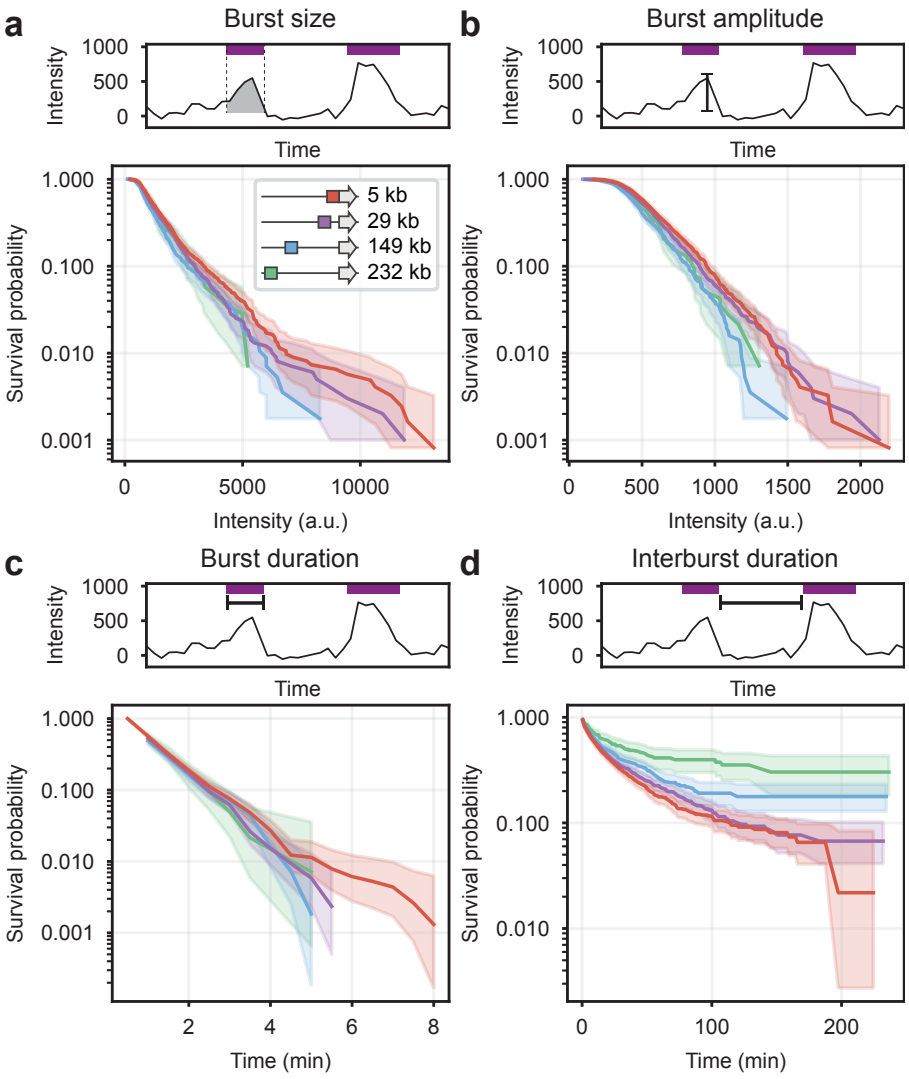

Figure 2.2: Burst parameter quantified from live-cell imaging. Upper panels describing how burst parameters were quantified. **a.** Survival probability of the burst size. 95% confidence intervals were estimated by bootstrapping (10.000 repeats). **b.** Survival probability of the burst amplitude. 95% confidence intervals were estimated by bootstrapping (10.000 repeats) **c.** Survival probability of burst duration as calculated with the Kaplan-Meier estimator. **d.** Survival probability of interburst duration as calculated with the Kaplan-Meier estimator.

sored' events). To account for this effect, we estimated burst duration and interburst interval distributions using the Kaplan-Meier estimator. Censoring-corrected burst durations appeared independent of EP genomic distance (Fig. 2.2c, Extended Data Fig. E.2c), with individual bursts being relatively short, exhibiting a median survival time of approximately 1.5 minutes. However, interburst durations substantially increased with increasing EP genomic distance (Fig. 2.2d, Extended Data Fig. E.2d), with median survival times ranging from 15 minutes for the two closest EP genomic distances (5 kb and 30 kb, respectively) up to 30 minutes for the longest EP genomic distance (232 kb), each displaying wide distributions from one minute to hundreds of minutes.

Irrespective of EP genomic distance, a substantial number of cells did not burst within the 4-hour imaging window. Since these events were censored, they were not included in the Kaplan-Meier calculations, but the fraction of non-bursting cells followed the same trend as the interburst durations, confirming the validity of these observations (Extended Data Fig. E.2e).

To ensure that these measures were not artifacts of the image analysis pipeline, we manually curated 10% of the live-cell imaging data from the two clones with EP genomic distances of 5 kb and 149 kb. Comparing the survival probabilities from the manually annotated data with the automatic image analysis pipeline revealed no differences in distribution or quantities, thus validating our data processing method (Extended Data Fig. E.2f-m).

2.3.3 Burst frequency and its variability scale nonlinear with contact probability

Since only interburst durations changed with EP genomic distance, we examined whether the related burst frequency recapitulated the previously observed nonlinear relationship. Burst frequency, calculated as the count of active transcription periods per hour, included data from a cell line where the enhancer was integrated into a different chromosome, referred to as the promoter-only cell line. In this cell line, no nascent transcription was observed over a 90 hour imaging period, resulting in a burst frequency of 0. Comparing burst frequencies between clones revealed that burst frequency fully recapitulated the previously observed nonlinear relationship between transcriptional output and contact probabilities (Fig. 2.3a).

Given the wide distribution of interburst durations, we further analyzed how average burst frequencies were manifested at the single-cell level. Specifically, for clones with EP genomic distances of 5 kb or 149 kb, the predicted average burst frequencies were approximately four and one burst per four hours, respectively. However, analysis of transcriptional traces spanning at least four hours revealed substantial cell-to-cell variability. Many cells exhibited more than four bursts in the 5 kb distance scenario, while others showed no bursts at the 149 kb distance (Fig. 2.3b,c).

Quantifying this cell-to-cell variability by the coefficient of variation of the burst fre-

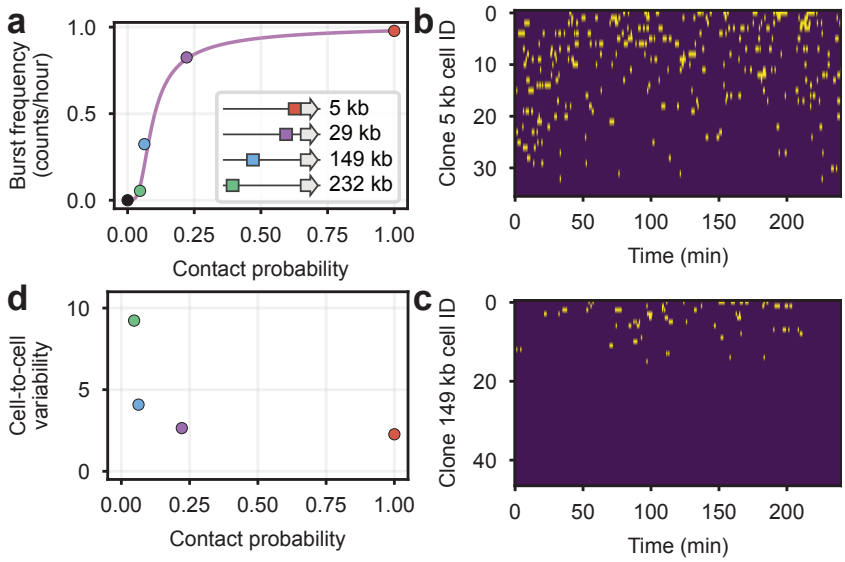

Figure 2.3: EP genomic distance modulates burst frequency and its variability. a. Burst frequency (number of bursts per hour) as a function of EP contact probability. Purple trendline originates from Zuin *et al.* [9]. **b-c.** Binarized intensity traces of at least 4 hours length for clonal cell lines of **(b)** 5 kb EP genomic distance and **(c)** 149 kb genomic distance. Every line corresponds to one cell. Traces are sorted by the number of bursts in the trace. **d.** Cell-to-cell variability of burst frequency as function of contact probability. Cell-to-cell variability is calculated as coefficient of variation.

quency demonstrates a nonlinear relationship, where the proximity of the enhancer to the promoter correlates with reduced cell-to-cell variability (Fig. 2.3d). This indicates a more consistent transcriptional behavior when the enhancer is closer to the promoter.

2.3.4 Silent promoter periods occur over three distinct timescales

Strikingly, the interburst durations do not follow a single exponential, but a weighted sum of at least three exponential distributions with lifetimes of approximately 3, 25, and >160 minutes (Fig. 2.4a-c and Extended Data Tab. E.1-E.4). Those multiple timescales over which no burst is observed suggest the existence of multiple OFF states (i.e. states in which transcription cannot initiate) in which the promoter would spend more or less time.

However, the shortest interburst duration (~3 minutes) could also potentially emerge from a low initiation rate in an ON state (i.e. state in which transcription can initiate), as previously proposed by Pimmet *et al.* [59]. In this case, we expect that bursts are terminated by chance because no Pol II was initiated during a time window longer than the elongation/release duration, creating a gap in the burst signal. To test this hypothesis, we examined the average TSS intensity profile after aligning intensity

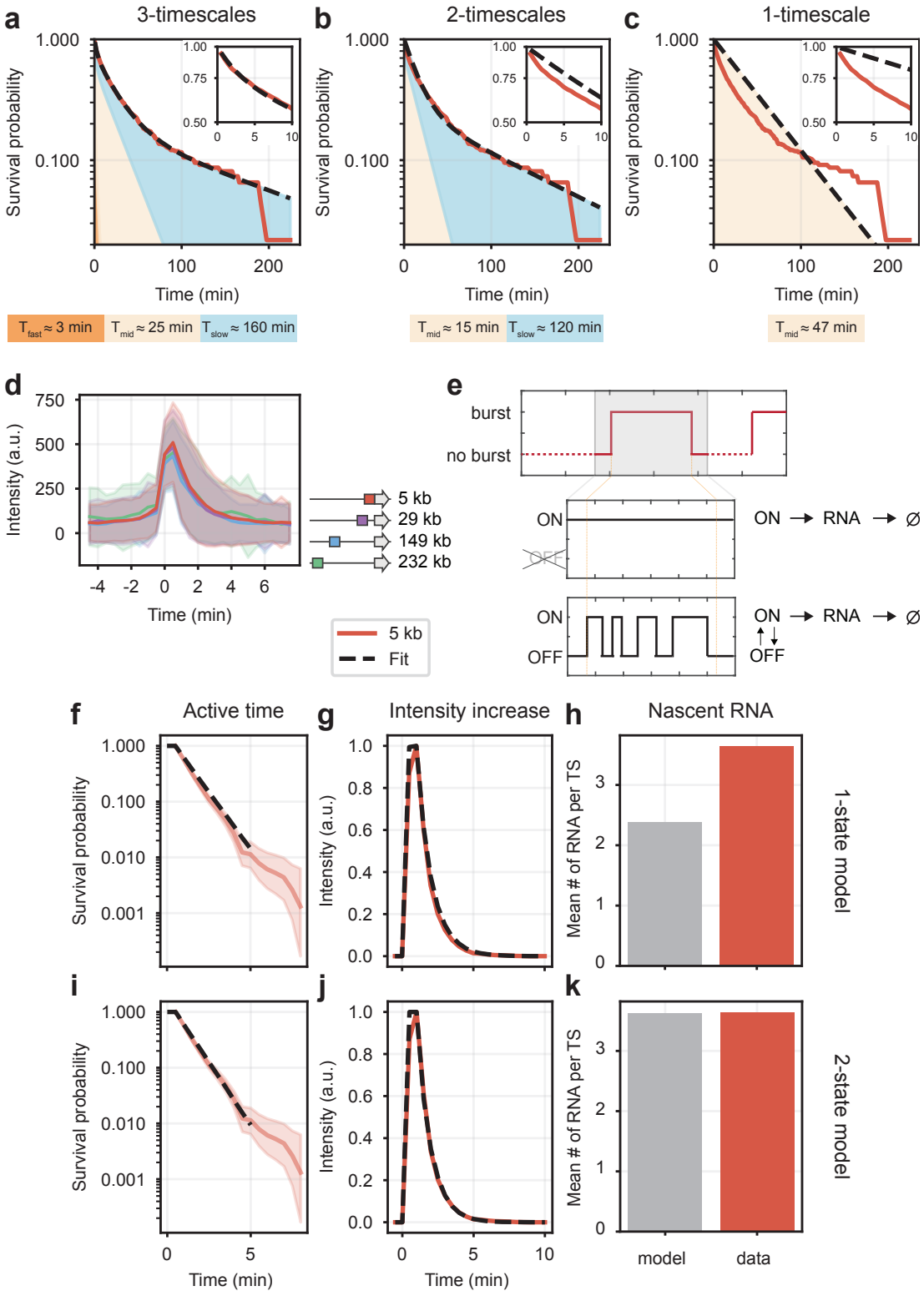

Figure 2.4: Interburst duration is best described by three time-scales. *continued on next page.*

Figure 2.4: *continued from previous page.* **a.** Best fit of the three-exponential model to the experimental survival probabilities of the interburst durations for the cell line where the SCR was located at 5 kb from the promoter. The best fit curve (black dotted line) is the weighted sum of three exponential curves represented in colored areas from the highest exponent (orange) to the lowest exponent (blue). The timescale of each exponential curve (1/exponent) is shown in the bottom. **b.** Similar to panel **(a)** but for the two-exponential model. **c.** Similar to panel **(a)** but for the single exponential model. **d.** Average intensity increases at the start of a burst. Per clone, transcriptionally active times were aligned to their start and intensity traces averaged. **e.** Illustration of a burst from the constitutive (one state) model and the two-state model. In the constitutive model, the burst is arrested because of a long time interval between two initiation events. In the two-state model, the burst is arrested because of the switch OFF. **f-h.** Best fit of the constitutive model to **(f)** the survival probability of burst duration, **(g)** the average TSS intensity profile and **(h)** to the mean number of nascent RNA. **i-k.** Similar to panels **(f-h)** but for the two-state model.

traces to the start of each burst. As expected from the largely overlapping burst size distributions, the average profiles from all clones overlapped, indicating similar initiation and termination rates (Fig. 2.4d, Extended Data Fig. E.2o). This is also in line with TSS intensity measurements by smRNA-FISH, which shows that the average number of nascent mRNAs is similar in all clones (around 3.5 mRNA/TSS on average, Extended Data Fig. E.3).

Next, we reasoned that if low initiation rate in an ON state was the main driver in burst termination, the average TSS intensity profile could be modeled by a simple Poisson initiation and termination (Fig. 2.4e). Interestingly, such a model cannot reproduce both the average intensity trace of a burst and the mean number of nascent mRNAs (Fig. 2.4f-h). This suggests that bursts are not stopped by random fluctuation in initiation but rather by switching the promoter to an OFF state. Indeed, a two-state model in which the promoter stochastically switches between an OFF and an ON state can reproduce both the average intensity trace of a burst and the mean number of nascent mRNAs (Fig. 2.4e,i-k).

2.3.5 Modeling the multi-state kinetics of promoter operation

Although the two-state model reproduces the observed initiation and termination of a burst, it cannot reproduce the three observed timescales of interburst durations. We previously proposed a mathematical model describing the simple hypothesis that, in single cells, the ON rate of the promoter is transiently increased after stochastic interactions with an enhancer via a small number of reversible regulatory steps (Fig. 2.5a). In this model, the promoter switches between a 'basal' two-state regime with a small on-rate and an 'enhanced' two-state regime with a higher on-rate. While it accurately reproduced the mean mRNA levels we previously measured in our enhancer mobi-

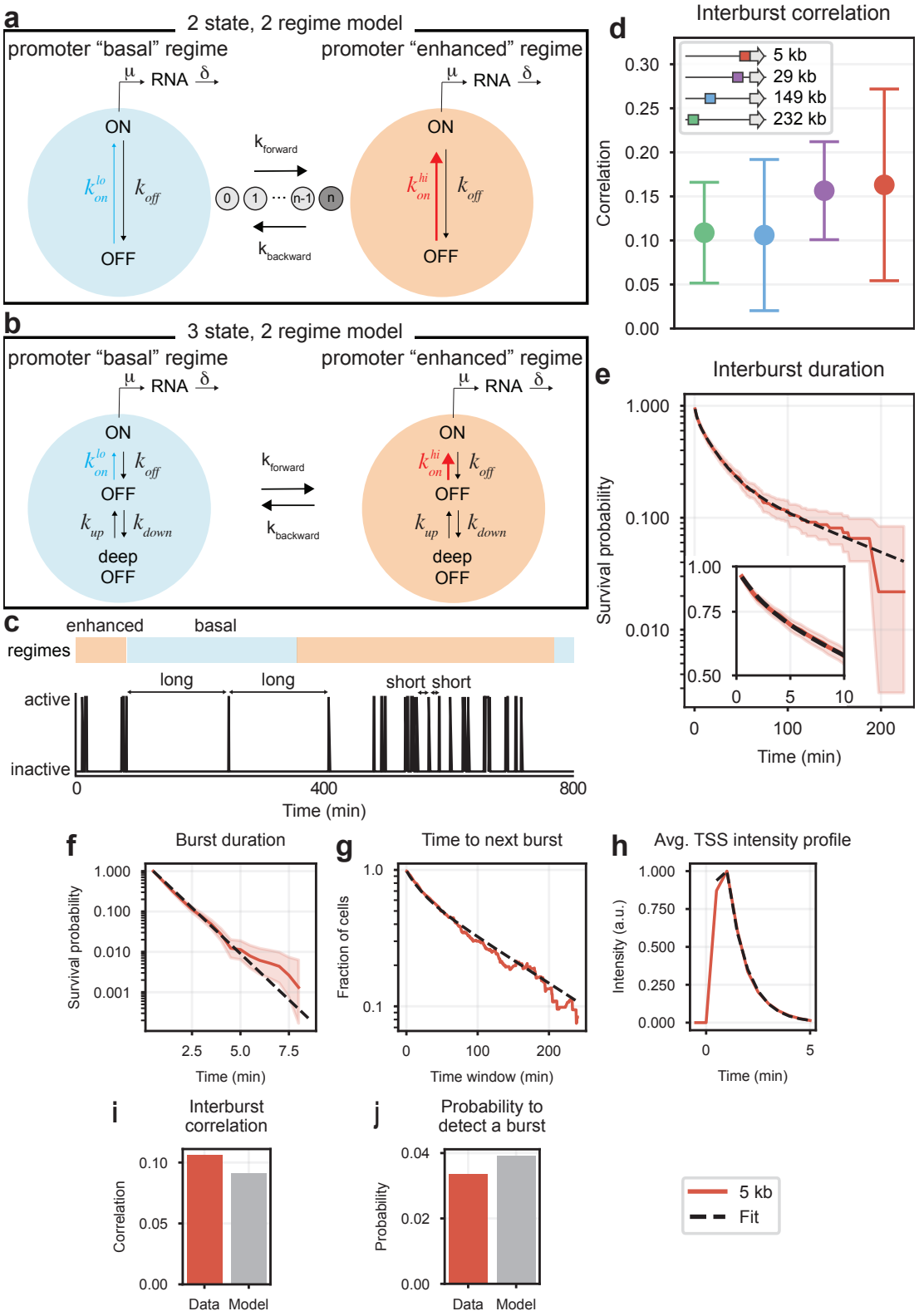

Figure 2.5: A multi-state model for the transcriptional dynamics of the ectopic Sox2 promoter. *continued on next page.*

Figure 2.5: *continued from previous page.* **a.** Diagram of a simplified version of the two-state, two-regime model in Zuin *et al.* [9]. The promoter operates in a 'basal' two-state regime with a small on-rate and can transiently enter an 'enhanced' two-state regime with a higher on-rate. In the original model, the promoter must complete n regulatory steps to successfully enter the 'enhanced' regime. **b.** Diagram of the three-state, two-regime model used to fit our live-cell imaging data. The promoter operates in a 'basal' three-state regime with a small on-rate and can transiently enter an 'enhanced' three-state regime with a higher on-rate. **c.** Example trajectory of the three-state, two-regime model. The active state corresponds to a burst (i.e. signal of at least one RNA) and the inactive state corresponds to no signal (i.e. zero RNA). The colored trace shows the underlying regime ('basal'/'enhanced') in which the promoter is. The interbust times are long in the 'basal' regime and short in the 'enhanced' regime, resulting in a positive correlation. **d.** Correlation between the durations of consecutive interburst events. The data are mean±s.d, based on 2.000 bootstrap repeats. **e-j.** Best fit of the three-state, two-regime model to the experimental **(e)** survival probability of the interburst duration, **(f)** survival probability of the burst duration, **(g)** survival probability of the time until the next burst, **(h)** average TSS intensity profile, **(i)** interburst correlation and **(j)** steady-state probability to detect a burst, for the cell line where the SCR was located at 5 kb from the promoter.

lization assay [9], this model predicts only two timescales of interburst durations and therefore fails at reproducing the actual promoter dynamics that we measured in live-cell imaging experiments. A natural extension of this model that accounts for three timescales in the interburst duration is the addition of a second OFF state (which we refer to as 'deep' OFF) in both regimes (Fig. 2.5b).

Similar to the double two-state promoter we previously proposed, this model predicts positive correlations between consecutive interburst events, since in the 'enhanced' regime, short interburst times tend to be followed by short interburst times, and vice versa in the 'basal' regime (Fig. 2.5c). Interestingly, our data validate this prediction (Fig. 2.5d) which disqualifies any alternative models that have a single ON state (e.g. typical multistate model used in the literature [18, 25, 131]), in which interburst events are independent from each other because the burst that separates two consecutive interburst events always resets the promoter to the same ON state.

To test whether this model can reproduce the statistics of the observed burst dynamics, we first fit the model to the data of a single clone (EP distance = 5 kb). To constrain the nine parameters of the model, we use a diverse set of observables. The model could simultaneously fit the survival probabilities of the interburst and burst durations, the survival probability of the time to the next burst, the average TSS intensity profile, the correlation between consecutive interburst event times, the steady-state probability to observe a burst and the mean number of nascent mRNA (Fig. 2.5d-j, Extended Data Tab. E.5). Best agreement occurred with an average burst duration of 52 seconds, an average interburst duration of 9 hours in the 'basal' regime, and 15 minutes in the 'enhanced' regime (Extended Data Tab. E.6). Every

2 hours on average, the promoter transitions from the 'basal' and the 'enhanced' regime, resulting in an average interburst duration of 30 minutes and 2.6 bursts per hour on average (Extended Data Tab. E.7).

We then asked if the data from the other clones can be explained by the model under the stringent hypothesis that EP distance modulates only the transition from the 'basal' to the 'enhanced' regimes, as in Zuin *et al.* [9]. For this we kept all other parameters fixed and asked the model to fit the survival probabilities of the interburst durations of all the other clones (i.e. EP distance = 29 kb, 149 kb, 232 kb) when only the forward rate is changed, and predict all other observables independently (Fig. 2.6a,b). Strikingly, the model was not only able to provide excellent fits to the OFF times, but it also predicted the survival probabilities of the burst durations, the correlation between consecutive interburst event times, the steady-state probability to observe a burst, and the survival probabilities of the time to the next burst of all the clones (Fig. 2.6c-e, Extended Data Tab. E.5). Best agreement occur with an average time spent in the 'basal' regime of 3 (EP distance = 29 kb), 11 (EP distance = 149 kb), and 41 hours (EP distance = 232 kb), resulting in a burst frequency of 2.1, 0.95, and 0.38 bursts per hour (Extended Data Tab. E.6).

2.4 Discussion

Our study provides quantitative measurements of promoter burst kinetics when varying its genomic distance over a range that is representative of those within mammalian *cis*-regulatory landscapes (from few to hundreds of kb). By only modifying the position of the enhancer within a 'neutral' genomic region with minimal structural and regulatory complexity [9, 102], we can isolate the effect of genomic distance from other confounding factors such as EP compatibility and influence from additional enhancers and promoters, which may confound interpretation of results when enhancer sequences are manipulated at endogenous gene loci.

In line with previous reports in which mammalian enhancers mainly controlled promoter burst frequency [124, 61, 131, 142, 25], we observe that EP genomic distance modulates the duration between consecutive bursts without any noticeable changes in burst size or duration. In addition to controlling population- and time-averaged transcription levels, this also results in different levels of heterogeneity across cells in the timing and number of bursts produced in a fixed time window - with enhancers located close to the promoter driving larger, but also more similar numbers of bursts per unit time compared to enhancers located at larger genomic distances. This suggests that the genomic position of an enhancer within a *cis*-regulatory landscape may have a substantial effect on the accuracy and robustness [1] of transcriptional programs.

Strikingly, in contrast to previous studies we observe three well-separated timescales (~3, ~25 and >160 minutes) in the silent periods between consecutive promoter

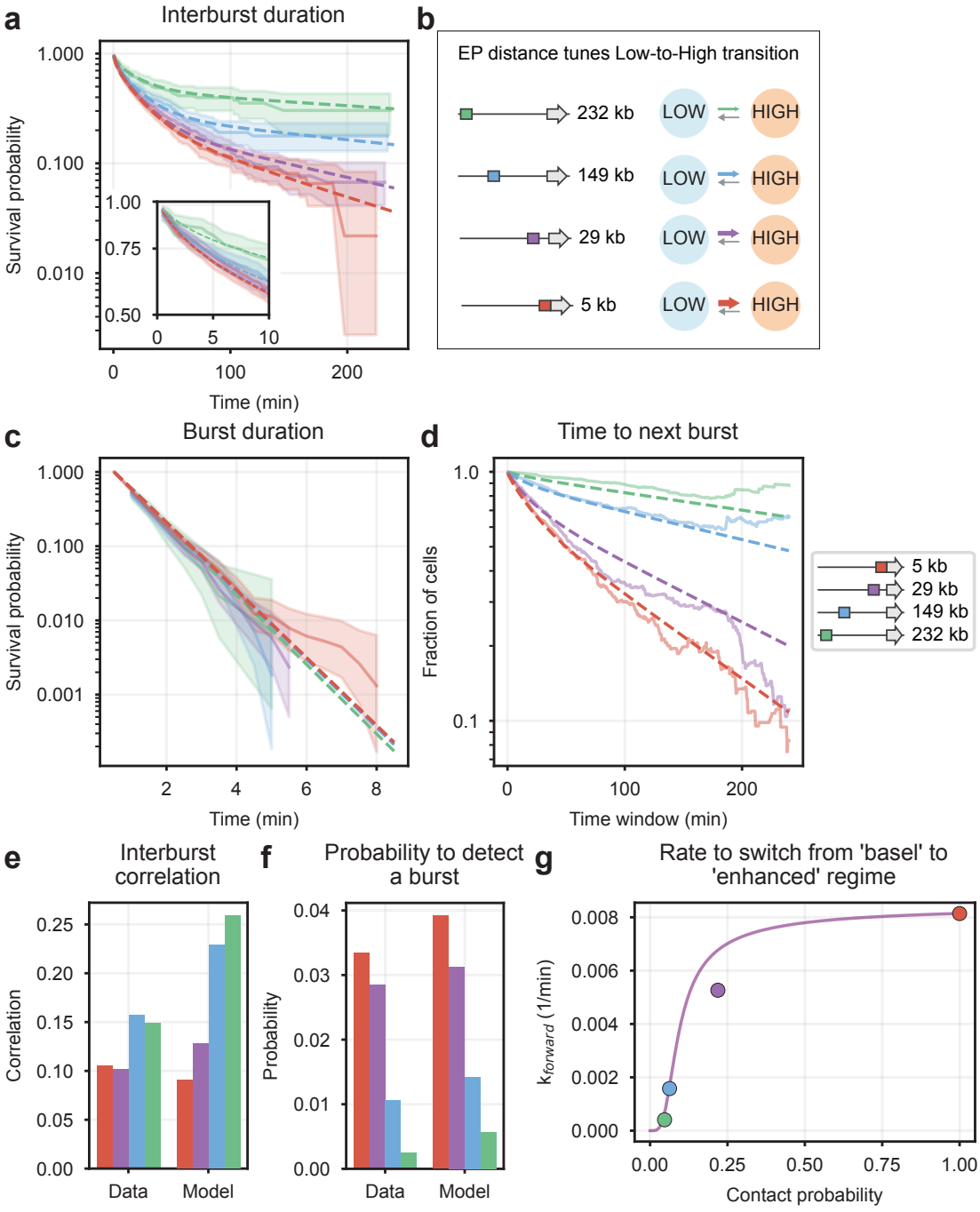

Figure 2.6: EP genomic distance affects the rate to switch from the 'basal' to the 'enhanced' transcription regime. *continued on next page.*

Figure 2.6: *continued from previous page.* **a-b. (a)** The survival probabilities of the interburst duration of the cell lines, where the SCR was located at 29 kb, 149 kb, and 232 kb from the promoter, can be predicted (purple, blue, and green dotted lines) from the model best fitted to the cell line where the SCR was located at 5 kb (red dotted line), **(b)** with a modified forward rate ($k_{forward}$). Solid lines are experimental data. **c-f.** Best fit of the model to the cell line, where the SCR was located at 5 kb (red), and model predictions of the other cell lines (purple, blue, green). The model was fitted on and predicted **(c)** the survival probabilities of the burst duration, **(d)** survival probabilities of the time until the next burst, **(e)** the correlation between the times of two consecutive interburst events and **(f)** steady-state probabilities to detect a burst. Dotted lines are model outputs and solid lines are experimental data. **g.** Best fit value of the forward rate ($k_{forward}$) plotted against contact probability. Sigmoidal transcriptional response observed in Zuin *et al.* [9] normalized to the larger $k_{forward}$ rate value is shown in purple.

bursts, irrespective of its genomic distance with the promoter. Interestingly, in contrast to previous studies in *Drosophila* [59], we show that the shortest of these timescales (~3 minutes) does not arise from fluctuations in transcriptional initiation, supporting the notion that the promoter has at least three OFF states and thus cannot be described by two- or three-state models. While it is possible that detection of these three very different timescales is enabled by our relatively high time resolution (30 seconds) and long imaging windows (5 hours), we cannot exclude that this mode of operation is specific to either the Sox2 promoter, or the specific ectopic location where it is inserted in our experiments.

Remarkably, we also observe that the duration of consecutive interburst silent periods are correlated. This is not compatible with any of the existing models of promoter operation, which all consider a single ON state. Indeed, a necessary requirement for positive correlation between consecutive interburst events is that there are at least two ON states and each one of them is directly connected to an OFF state with a different lifetime. Here, we propose a simple model that satisfies this requirement and has three timescales in the silent periods between bursts. In this model, the promoter switches between a 'basal' three-state regime with a small ON rate and an 'enhanced' three-state regime with a higher ON rate (Fig. 2.5b). This model extends the model we previously proposed to account for the population-averaged behavior we observed in the same experimental system [9]. There, in the absence of time-resolved measurements, we proposed the promoter to switch between 'basal' and 'enhanced' two-state regimes, which our work now compels to reconsider as two three-states regimes instead. In Zuin *et al.* [9], we also made the explicit hypothesis that the switch is triggered by stochastic EP interactions via one or more reversible regulatory steps that once completed, allow the promoter to enter the 'enhanced' regime. While the model proposed here does not explicitly describe EP interactions and their effect on the promoter, it still assumes implicitly that those processes affect the transition from the 'basal' to the 'enhanced' regime. Interestingly, the best fitted

values of this forward rate describe a sigmoidal relationship with the contact probability between the enhancer and the promoter, following a similar trend than the nonlinear transcriptional response observed in Zuin *et al.* [9] (Fig. 2.6g). Although alternative models having at least two ON states and three OFF states cannot be ruled out, our model provides a simple explanatory framework for the complex dynamics of enhancer-driven transcription in single cells. Future live-cell imaging experiments with higher time resolution and longer imaging windows will enable to test the existence of the 'basal' regime and to measure more accurately the transition between the regimes.

Although our experiments do not give access to the dynamics of EP distances, it is nevertheless possible to compare the burst frequency we measured at 149 kb distance to the frequency of physical proximity events that we previously measured in the same 'empty' TAD using Tet operator (TetO) and Lac operator (LacO) arrays located 150 kb apart [102]. On average, these neutral sequences came into physical proximity (defined as being in a range of ~150 nm) every six minutes, which is around 30 times more frequent than the average burst frequency of approximately one burst every three hours. If the dynamics of enhancers-promoter looping resemble those of non-regulatory DNA (i.e. LacO and TetO) at the same genomic distance, these figures suggest that the majority of such proximity events might either represent non-functional interactions, or consecutive proximity events leading to one productive regulatory event via transitions through intermediate rate-limiting regulatory steps, as we previously suggested based on model analysis [9].

Finally, we note that although bursts from the Sox2 promoter in this experimental system have a similar duration, they are much rarer than those that were previously detected at the endogenous Sox2 locus [125, 126], even when the enhancer is in much closer genomic proximity than the endogenous SCR. This might be due to the absence of the CTCF loops that connect the endogenous Sox2 promoter and SCR [153] or the lack of additional enhancers [149].

Taken together, our data uncover quantitative principles by which distal enhancers control transcription and its variability in single cells, and provide a quantitative framework for future studies to address mechanistically the interplay between enhancer-promoter interactions and transcription bursting.

2.5 Methods

2.5.1 Experimental methods

2.5.1.1 Cell culture of mESC lines

All cell lines are based on the E14Tg2a parental mESC line (karyotype 19, XY, 129/Ola isogenic background). Cells were cultured on gelatin-coated culture plates in Glas-

gow Minimum Essential Medium (Sigma-Aldrich, G5154) supplemented with 15% fetal calf serum (Eurobio Abcys), 1% L-Glutamine (Thermo Fisher Scientific, 25030024), 1% Sodium Pyruvate (Thermo Fisher Scientific, 11360039), 1% MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140035), 100 µM beta-mercaptoethanol (Thermo Fisher Scientific, 31350010), 20 U/ml leukemia inhibitory factor (Miltenyi Biotec, premium grade) in 8% CO_2 at 37°C. For enhancer mobilization and subsequent experiments, the culture medium was additionally supplemented with 2i inhibitors (1 µM MEK inhibitor PDO35901 (Axon, 1408) and 3 µM GSK3 inhibitor CHIR 99021 (Axon, 1386)). Cells were tested regularly for mycoplasma contamination and no contamination was detected.

2.5.1.2 Generation targeting vectors for MS2 cassette integration

The founder cell line carrying the piggyBac transgene and MS2 repeats is based on the piggyBac transgene founder cell line used in Fig. 1 of ref. [9]. This cell line allows for the piggyBac-mediated mobilization of the full-length SCR enhancer in the context of a mutant TAD where two internal CTCF sites were deleted.
We modified this cell line by inserting a 24x MS2 stem-loop cassette in the 3'UTR of the *EGFP* transcript. For knock-in of 24x MS2 repeats [152], three plasmids were used: a targeting vector, a gRNA vector to linearize the targeting vector and a gRNA vector targeting the genomic locus. The targeting vector contains 24x MS2 repeats flanked by homology arms and bacterial gRNA sequences for linearization of the insert inside a pUC57 backbone and was ordered from Genewiz. For linearization of the targeting vector, a modified pC2P plasmid including the bacterial gRNA sequences for plasmid linearization and the Cas9-P2A-puromycin cassette was gifted by Dirk Schübeler [154]. For targeting the genomic locus, namely the 3'UTR of the *EGFP* transcript and in front of SV40, the gRNA sequence was designed using an online tool (http://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) and purchased from Microsynth AG. This gRNA sequence was cloned into the PX459 plasmid (Addgene, cat# 48139) using the BsaI restriction site. Primers used are listed in Tab. E.8.

2.5.1.3 Generation of founder cell line carrying MS2 system

The founder line carrying the piggyBac transgene and MS2 repeats was generated by transfecting 1.5×10^6 cells with 1.5 µg MS2 targeting vector, 750 ng PX459 (Cas9/EGFP_gRNA) and 300 ng PX459 (Cas9/linearization_gRNA) using Lipofectamine-3000 (Thermo Fisher Scientific, L3000008) according to the manufacturer's guidelines. 24 h after transfection, 1 µg/ml of puromycin (InvivoGen, ant-pr-1) was added to the medium for 24 h. Cells were cultured in standard medium for an additional 4 days after which single cells were isolated in 96-well plates. Sorted cells were kept for 2 days in medium supplemented with 100 µg/µL primocin (InvivoGen, ant-pm-1) and 10 µM ROCK inhibitor (STEMCELL Technologies, Y-27632). 10 days after

sorting, plates were duplicated and genomic DNA extracted on-plates by lysing cells with lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 50 mM NaCl, 1 mg/mL proteinase K (Macherey-Nagel, 740506) and 0.05 mg/mL RNase (Thermo Fisher Scientific, EN0531)) and subsequent isopropanol precipitation. Individual cell lines were analyzed using genotyping PCR (genotyping primers listed in Tab. E.8) and correctly genotyped cell lines were expanded. Correct insertion was verified by Sanger sequencing of the genotyping PCR result covering the whole insert. Stable integration of stdMCP-stdHaloTag was achieved via lentiviral transduction (gifted by Jeffery Chao). 0.5×10^6 cells were transfected using Polybrene Infection/Transfection Reagent (Sigma-Aldrich, TR-1003) according to the manufacturer's guidelines and 10 days later sorted for HaloTag expression into single cells as described above. After 10 days, one-third of the cells were replated onto Cellvis Glass Bottom Microplates (96-well, Cellvis, P96-1.5H-N) and clonal lines were screened by microscopy for a good signal-to-noise ratio (SNR).

2.5.1.4 Mobilization of the piggyBac-enhancer cassette

Mobilization of the enhancer cassette was performed as previously described with minor adjustments [9]. In short, cells were cultured in medium additionally supplemented with 2i inhibitors. PiggyBac-enhancer cassette mobilization was achieved by 12 independent transfections using piggyBac transposase (PBase, gifted by Jesse Owens [155]). EGFP+ cell lines were isolated using fluorescence-activated cell sorting (FACS) in 96-well plate and multiplied for DNA extraction, flow cytometry, expansion and freezing.

2.5.1.5 Mapping piggyBac-enhancer insertion sites in individual cell lines

PiggyBac integration sites were mapped on-plate using Splinkerette PCR after DNA extraction as described previously [9]. Splinkerette primers are listed in Tab. E.8.

2.5.1.6 Flow cytometry

EGFP levels of individual cell lines were measured on-plate using the BD LSRII SORP flow cytometer equipped with BD High Throughput Sampler (HTS). For each clone, measurements were performed in triplicates and mean EGFP fluorescence intensities were calculated using FlowJo. All three replicates were averaged.
After expansion, before cell fixation and in parallel to live-cell imaging, EGFP levels of individual cell lines were measured in-tube using the BD LSRII SORP flow cytometer.

2.5.1.7 smRNA-FISH sample preparation

Cells were detached using accutase (Sigma-Aldrich, A6964) and spotted on poly-L-lysine (Sigma-Aldrich, P8920) pre-coated cover slips. 30 min after spotting, cover slips were washed once with PBS and cells fixed using 3% Paraformaldehyde (EMS, 15710) in PBS for 10 min at room temperature. Cover slips were washed with PBS and kept in 70% ethanol at -20°C for at least 24 h. Before probe hybridization, the cover slips were incubated for 10 min in freshly prepared wash buffer composed of 10% formamide (Millipore Sigma, S4117) in 2x SSC (Sigma-Aldrich, S6639). The cover slips were hybridized overnight (around 16 h) at 37°C in freshly prepared hybridization buffer (10% formamide, 10% dextran sulfate (Sigma-Aldrich, D6001) in 2x SSC) containing 125 nM of RNA-FISH probe sets against MS2 labeled with Atto-640 (gifted by Jeffery Chao [152]) and against EGFP labeled with Atto-561. For EGFP probes, 20 nucleotide long sequences were designed using Stellaris probe designer and the oligos were conjugated with Atto-561 as previously described [156]. EGFP targeting sequences are listed in Tab. E.9). After hybridization, the cover slips were washed with wash buffer pre-warmed to 37°C for 30 min at 37°C with shaking, followed by 5 min incubation with 500 ng/ml DAPI solution (Sigma-Aldrich, D9564) in PBS and a second round of washing. The cover slips were then mounted on slides with Prolong Gold medium (Thermo Fisher Scientific, P36934) and cured at room temperature for 24 h. After sealing, imaging was performed within 24 h.

2.5.1.8 smRNA-FISH image acquisition

Images were acquired on a Zeiss Axio Imager M2 microscope equipped with the Yokogawa CSU W1 with Dual T2 (pinhole size: 50 µm) spinning disk confocal scanning unit and sCMOS camera using 150 mW 405 nm, 150 mW 561 nm and 150 mW 639 nm Coherent Orbis diode lasers and a Plan-APOCHROMAT 100x/1.4 NA oil-immersion objective. DAPI signal was image with BP460/50 filter set (Chroma), 50% laser power and 50 ms exposure time, Atto-561 with BP609/54 filter set (Semrock), 100% laser power and 300 ms exposure time and Atto-640 with BP700/75 filter set (Chroma), 100% laser power and 700 ms exposure time. 51 z-stacks were acquired with a z-step size of 300 nm and a pixel size of 65x65 nm.

2.5.1.9 smRNA-FISH image analysis

Raw images were processed using custom Python code. All code is available upon request.
Cells were 3D segmented based on combined background intensities from the 561 nm and 640 nm channels using StarDist and the '2D_versatile_fluo' model [157, 158]. Cells wrongly segmented were manually excluded from the analysis. For spot detection, hmax and Laplacian of Gaussian (LoG) detection was used, whereas only spots detected in both approaches were kept. Spot size was estimated in xy by the Rayleigh

criterion and in z by the Abbe criterion. Afterwards, spots were assigned to cells and their maximum intensity read-out using an elliptical mask of the rounded spot size. Median cell background intensities were calculated using the cell segmentation and subtracted from the spot intensities.

2.5.1.10 Live-cell imaging

35-mm glass-bottom dishes (Mattek, P35G-1.5-14-C) were coated with 1–2 µg/mL laminin (Sigma-Aldrich, L2020) in PBS at 37°C overnight. The day before imaging, 0.8-1×10^6 cells were seeded in Fluorobrite DMEM (Gibco, A1896701) supplemented like culture medium containing 2i inhibitors (see above) and incubated at 37°C, 8% CO_2.
Before imaging, Halo-tagged MCP was labeled with Halo-Ligand-JaneliaFluor 549 (Tocris Bioscience, custom synthesis based on Cat. No. 6503). To this aim, medium was exchanged for medium containing 100 nM Halo-Ligand for 20 min. Subsequently, cells were washed three times with PBS and fresh Fluorobrite medium was added.
For imaging, a Nikon Eclipse Ti-E inverted widefield microscope equipped with a Total Internal Reflection Microscopy iLAS2 module (Roper Scientific), a Perfect Focus System (Nikon), a motorized Z-Piezo stage (ASI), Evolve 512 Delta EMCCD cameras with a pixel size of 16x16 µm^2 (Photometrics) and a CFI APO TIRF 100x/1.49 NA oil immersion objective (Nikon) was used. Following laser lines were used: for 488 nm, 200 mW Toptica iBEAM SMART laser and for 561 nm, 200 mW Coherent Sapphire laser. The microscope was operated in oblique illumination mode and using the Visiview software (Visitron). An enclosed microscope environmental control set-up (The BOX and The CUBE, Life Science Instruments) was used at 37°C and 8% CO_2.
Before acquisition of movies, one z-stack (21 planes, spacing of 300 nm) of EGFP levels was recorded using 10% 488 nm laser power, an exposure time of 150 ms and a gain of 50. To image MS2 signals, series of z-stacks (21 planes, spacing of 300 nm) were acquired every 30 s using 4% 561 nm laser power, an exposure time of 100 ms and a gain of 100.

2.5.1.11 Flatfield correction

Before the acquisition of movies, Mattek dishes filled with 20 nM Alexa Fluor™ 488 Hydrazide (ThermoFischer, A10436) or 100 nM Alexa Fluor™ 568 Hydrazide (ThermoFischer, A10437) were imaged as flatfield-images using the same acquisition settings (see above). 200 z-planes with 300 nm spacing were acquired and subsequently 21 planes selected close to the cover slip. Three z-stacks were averaged per day. Dark-images were acquired by setting laser power to 0%, 500 planes acquired and averaged. Flatfield correction was performed using following equation as step of

image processing (see below):

$$image_{corrected} = \frac{(image_{raw} - dark\text{-}image) * mean(flatfield\text{-}image - dark\text{-}image)}{(flatfield\text{-}image - dark\text{-}image)} \quad (2.1)$$

2.5.1.12 Image processing of live-cell imaging

Raw images were z-projected (mean projection for 488 nm channel, max projection for 561 nm channel) using Fiji and further processed using custom Python code. All code is available upon request.
Individual cells were segmented using StarDist and the '2D_versatile_fluo' model [157, 158] and tracked over time using linear assignment problem (LAP) tracker with no gaps allowed, a minimal cell size and a minimal track length of 10 frames.
Active transcription spots were detected using Trackpy (thresholding after a band pass filtering [159]) and subsequently filtered by size and intensity to remove spurious detections.
Spots were linked over time in individual cells and when no spot was visible between two bursts, the position of the promoter was linearly interpolated taking cell lateral movement and relative deformation into account. If two spots were detected in the same cell at the same time, the brighter spot was chosen for linking.
Spot intensity was read-out after projected images were flatfield corrected (see above) using a circular mask (7 pixel diameter). The local background intensity was read out using a ring-shaped mask (1 pixel thickness) around the spot mask separated by a 4-pixel gap. The global background intensity of the complete cell was read out using the cell mask.
To construct intensity traces, spot and local background intensities were bleach corrected by dividing the normalized global background intensity. Subsequently, the local background intensity was subtracted from the spot intensity.
To segment transcriptional active/inactive times, information on spot detection was used. Whenever a spot is detected it is defined as transcriptional active time. To account for changing levels of misdetection, the first hour of movies was excluded from analysis.

2.5.2 Quantification and statistical analysis

2.5.2.1 smRNA-FISH transcription site analysis

Background subtracted spot intensities were corrected for batch effects using z-score normalization. In detail, first log distributions were calculated. Then, z-scores for technical samples (cover slips) were calculated as follows:

$$z\text{-}score_{individual\ spot} = \frac{intensity_{individual\ spot} - mean(intensity_{sample})}{std(intensity_{sample})} \quad (2.2)$$

These z-scores of technical samples were normalized to the mean and standard deviation of complete dataset using:

$$z\text{-}score_{corrected} = z\text{-}score_{individual\ spot} * std(intensity_{dataset}) + mean(intensity_{dataset}) \quad (2.3)$$

In order to determine a threshold z-score between single mRNA spots and transcription sites, one image per clonal cell line was manually annotated. Resulting single mRNA z-scores were fitted using a Gaussian and the threshold defined as the mean + 3×standard deviation. In detail, bootstrapping was performed (n=1000) to assess the robustness of the threshold and the resulting mean of the thresholds was used.
For classified transcription sites, the number of mRNA at the site was estimated by calculating an enrichment score based on the background subtracted spot intensities.

$$enrichment\ score = \frac{transcription\ site\ intensity}{median\ spot\ intensity\ of\ that\ cell} \quad (2.4)$$

2.5.2.2 Time-averaged probability to detect a burst in live-cell imaging

To calculate the time-averaged probability of seeing a burst, binarized MS2 intensity traces (transcriptional active/inactive times) were used. Per biological sample (cell line), the probability to see a transcriptional active time per time point was calculated and subsequently averaged over time.

2.5.2.3 Survival probabilities calculated with Kaplan-Meier estimator

To account for transcriptional active/inactive times that start or end outside the imaging window ('censored' events), Kaplan-Meier estimator [160] was used to calculate survival probabilities of burst and interburst durations. This non-parametric method utilizes both uncensored (start and end within the imaging window) and right-censored data (start within the window).
From binarized MS2 intensity traces, interburst and burst durations are calculated from non-censored and right-censored events. The survival function *S(t)*, representing the probability that an event exceeds time *t*, is expressed as:

$$S(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})) \quad (2.5)$$

where t_i represents a time length of at least one event (e.g. one minute), d_i the number of events at time t_i (e.g. number of active times at one minute), and n_i the number of events surviving (e.g. number of active times longer than one minute, either non-censored or right-censored) up to t_i.
The python implementation 'lifelines' was used to calculate the Kaplam-Meier estimator [161].

2.5.2.4 Bootstrapped survival probabilities

Survival probabilities for burst size and burst amplitude were calculated from MS2 intensity traces. Only active transcriptional times starting and ending within the imaging window (uncensored) were considered. Burst size was determined as the integrated intensity for the MS2 signal over an active transcription time. Burst amplitude was calculated as the maximum intensity of an active transcription time. For resulting survival distributions, 95% confidence intervals were calculated from bootstrapping (n=10.000).

2.5.2.5 Intensity trace alignment and average intensity trace

To align the start of each transcriptional active time, time windows were defined around the onset of each active phase, with t=0 marking the start. From these windows, the mean and standard deviation of the intensity at the newly defined time points were calculated to generate average intensity traces.

2.5.3 Mathematical model and parameter fitting

The three state, two regime model (Fig. 2.5b) was fitted simultaneously to the survival probability of the burst and interburst duration, the steady-state probability to observe a burst, the correlation between the durations of two consecutive interburst events, the average TSS intensity profile, and the survival probability of the waiting time until the next burst, all measured in the cell line where the SCR was located at 5 kb from the promoter. All the observables were calculated analytically using the theory of continuous time Markov chain. The parameters of the model are the forward rate $k_{forward}$, the backward rate $k_{backward}$, the deep-OFF to OFF rate k_{up}, the OFF to deep-OFF rate k_{down}, the ON to OFF rate k_{off}, the basal ON to OFF rate ${k_{on}}^{b}$, the enhanced ON to OFF rate ${k_{on}}^{e}$, the initiation rate μ, and the elongation/release rate δ. All of these parameters were considered to be free in the fitting procedure. To test the hypothesis that EP distance modulates only the transition from the 'basal' to the 'enhanced' regimes, the model was fitted separately to the survival probabilities of the interburst duration, measured in each of the cell lines where the SCR was located at 29 kb, 149 kb, and 232 kb. In this case, only the parameter $k_{forward}$, was free parameter, whereas the other parameters were fixed to the best fit values obtained for the cell line where the SCR is located at 5 kb from the promoter.

2.6 Extended data

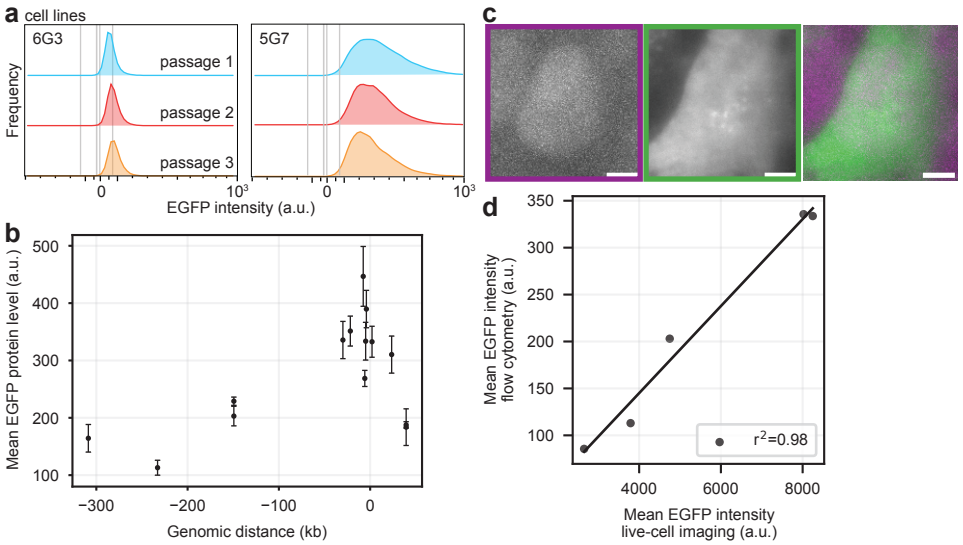

Figure E.1: EGFP levels correlate between flow cytometry and live-cell imaging experiments. a. On-plate EGFP levels of individual EGFP+ cell lines over cell passages as measured by flow cytometry. **b.** Mean on-plate EGFP intensity as assessed through flow cytometry in individual EGFP+ cell lines as a function of genomic distance. **c.** Representative images from live-cell imaging of the MCP-HaloTag channel (640 nm), EGFP channel (480 nm) and overlay (left to right). Scale bar: 5 µm. **d.** Correlation of mean EGFP levels measured on-plate using flow cytometry with mean EGFP levels measured in live-cell imaging. Pearson's coefficient of determination (r^2=0.98).

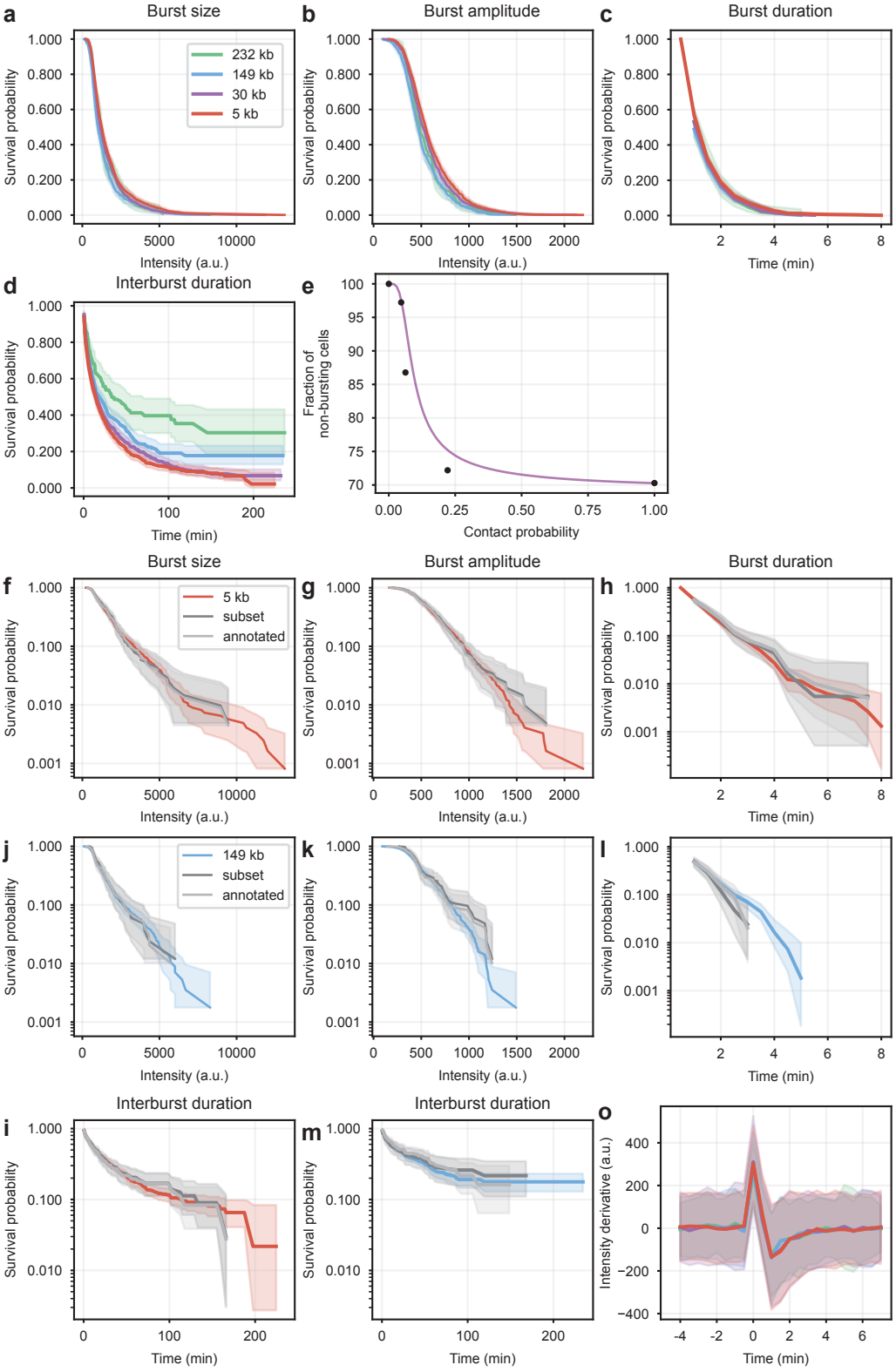

Figure E.2: Quantification of burst parameters is representative of underlying dynamics. *continued on next page.*

Figure E.2: *continued from previous page.* **a-d.** Survival probabilities of burst parameters in linear scale. **(a)** Survival probability of the burst size. 95% confidence intervals were estimated by bootstrapping (10.000 repeats). **(b)** Survival probability of the burst amplitude. 95% confidence intervals were estimated by bootstrapping (10.000 repeats). **(c)** Survival probability of burst duration as calculated with the Kaplan-Meier estimator. **(d)** Survival probability of interburst duration as calculated with the Kaplan-Meier estimator. **e.** Fraction of non-bursting cells as function of contact probability as measured by capture 3C. Purple trend line originates from Zuin *et al.* [9]. **f-m.** Comparison of burst parameters between the full dataset (colored line), a subset of the dataset (dark gray line) and manually annotated versions of the subset of the dataset (light gray line) for two clonal cell lines: **(f-i)** 5 kb EP genomic distance. **(j-m)** 149 kb EP genomic distance. **o.** Derivative of average intensity increases at the start of a burst.

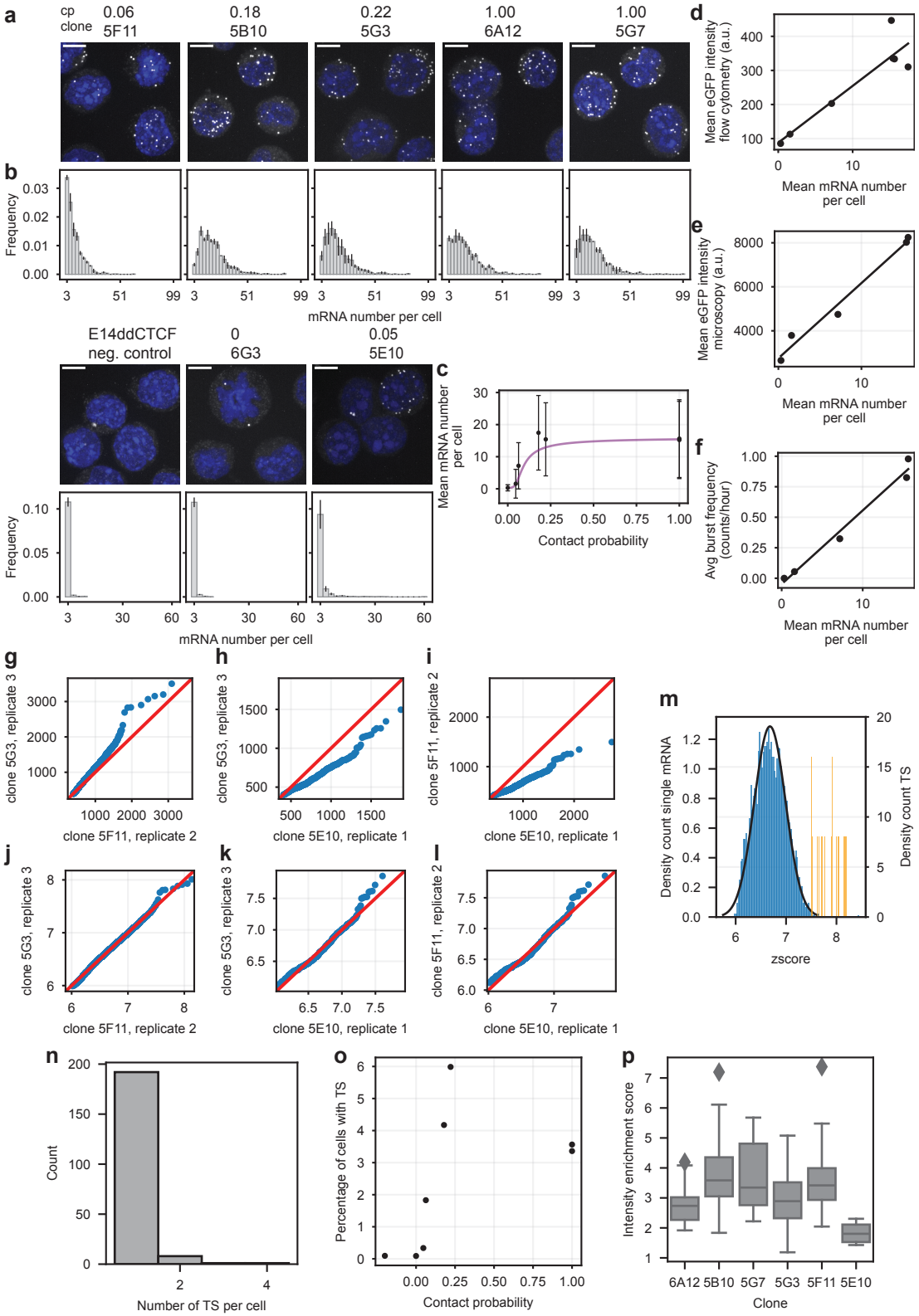

Figure E.3: smRNA-FISH of clonal cell lines with varying EP genomic distances, including TSS classification and quantification. *continued on next page.*

Figure E.3: *continued from previous page.* **a.** Representative smRNA-FISH images from E14ddCTCF lacking the transgene as negative control (lower left) and cell lines in which *EGFP* transcription is driven by the Sox2 promoter alone (6G3) or by the SCR located at different distances and contact probabilities. Scale bar: 5 µm. **b.** Distributions of mRNA numbers per cell measured in the cell lines shown in **(a)** (binning of histogram = 3). The error bars show the minimum and maximum frequency of three technical replicates. **c.** Mean number of mRNA per cell as function of contact probability as measured by capture 3C. The error bars show the standard deviation from the mean. Purple trend line originates from Zuin *et al.* [9]. **d.** Correlation of mean number of mRNA measured in smRNA-FISH with mean EGFP levels measured on-plate using flow cytometry. Pearson's coefficient of determination (r^2=0.84). **e.** Correlation of mean number of mRNA measured in smRNA-FISH with mean EGFP levels measured in live-cell imaging. Pearson's coefficient of determination (r^2=0.98). **f.** Correlation of mean number of mRNA measured in smRNA-FISH with burst frequency measured in live-cell imaging. Pearson's coefficient of determination (r^2=0.98). **g-i.** Quantile-quantile plots of smRNA-FISH spot intensity values from three example biological samples coming from three technical replications. Red line represents correlation=1. **j-l.** Quantile-quantile plot of z-score normalized spot intensities for samples shown in **(g-i)**. **m.** Density count plot of z-score normalized spot intensities from manually annotated smRNA-FISH data relating to single molecules (blue) or TSS (orange). Single molecule spot z-scores were fitted with a gaussian (black line). Line stops at three standard deviations from the mean and relates to the threshold at which a spot z-score is classified as TSS in subsequent analysis. **n.** Number of classified TSS per cell summarized from all clones. **o.** Percentage of cells showing at least one classified TSS as function of contact probability as measured by capture 3C. **p.** Intensity enrichment score of TSS classified per clone. Enrichment score relates to the number of single mRNA in TSS. The central line represents the median; the boxes indicate the lower (Q1) and upper (Q3) quartiles; whiskers extend to 1.5 times the interquartile range (IQR) below Q1 and above Q3; individual points represent outliers.

Table E.1: Parameter values for the best fitting one exponential model to the survival probabilities of the interburst durations. The inverse of the best fit exponent ($1/\lambda$) is shown in minutes. The residual of the least squares fit (res) and the Bayesian information criterion (bic) are included.

clone	EP genomic distance (bp)	$1/\lambda$	res	bic
5G7	-4942	47.3	38.84	-291.26
5G3	-29865	48.6	24.53	-398.99
5F11	-149334	50.2	9.81	-385.63
5E10	-232911	95.5	4.92	-211.16

Table E.2: Parameter values for the best fitting two exponential model to the survival probabilities of the interburst durations. The inverse of the best fit exponent ($1/\lambda$) is shown in minutes. Pslow is the weight associated with the slow exponential. The residual of the least squares fit (res) and the Bayesian information criterion (bic) are included.

clone	EP genomic distance (bp)	$1/\lambda_{\mathbf{slow}}$	$1/\lambda_{\mathbf{fast}}$	pslow	res	bic
5G7	-4942	120.9	15.4	0.26	0.73	-1027.42
5G3	-29865	90.8	11.6	0.41	0.55	-1129.49
5F11	-149334	100.1	9.2	0.51	0.36	-856.13
5E10	-232911	284.2	11.3	0.53	0.08	-527.77

Table E.3: Parameter values for the best fitting three exponential model to the survival probabilities of the interburst durations. The inverse of the best fit exponent ($1/\lambda$) is shown in minutes. P is the weight associated. The residual of the least squares fit (res) and the Bayesian information criterion (bic) are included.

clone	EP genomic distance (bp)	$1/\lambda_{fast}$	$1/\lambda_{middle}$	$1/\lambda_{slow}$	pslow	pmiddle	pfast	res	bic
5G7	-4942	163.9	23.5	2.6	0.19	0.58	0.23	0.21	-1248.6
5G3	-29865	inf	51.6	6.7	0.05	0.53	0.42	0.08	-1489.51
5F11	-149334	inf	39.8	3.9	0.15	0.57	0.29	0.12	-999.99
5E10	-232911	335	15.3	0.8	0.5	0.42	0.08	0.05	-549.5

Table E.4: Parameter values for the best fitting four exponential model to the survival probabilities of the interburst durations. The inverse of the best fit exponent ($1/\lambda$) is shown in minutes. P is the weight associated. The residual of the least squares fit (res) and the Bayesian information criterion (bic) are included.

clone	EP genomic distance (bp)	$1/\lambda_1$	$1/\lambda_2$	$1/\lambda_3$	$1/\lambda_4$	p1	p2	p3	p4	res	bic
5G7	-4942	163.9	163.9	23.5	23.5	0.19	0.23	0.09	0.49	0.21	-1238.12
5G3	-29865	inf	inf	51.6	42.9	0.05	0.42	0.53	0	0.08	-1478.96
5F11	-149334	inf	inf	39.8	39.8	0.15	0.29	0.28	0.29	0.12	-990.03
5E10	-232911	335	335	15.3	335	0.49	0.08	0.42	0.01	0.05	-540.78

Table E.5: Parameter values for the best fitting three state, two regime model. The unit of the rates is 1/minute. μ is the initiation rate, δ is the elongation/release rate.

clone	EP genomic distance (bp)	$k_{forward}$	$k_{backward}$	$k_{on}{}^{b}$	$k_{on}{}^{e}$	k_{up}	k_{down}	k_{off}	μ	δ
5G7	-4942	0.0081	0.008	0.007	0.25	0.12	0.22	1.14	24.59	6.95
5G3	-29865	0.0053	0.008	0.007	0.25	0.12	0.22	1.14	24.59	6.95
5F11	-149334	0.0016	0.008	0.007	0.25	0.12	0.22	1.14	24.59	6.95
5E10	-232911	0.0004	0.008	0.007	0.25	0.12	0.22	1.14	24.59	6.95

Table E.6: Statistics for the 'basal' and 'enhanced' regimes of the best fitting model. The times refer to averaged times in minutes.

Time ON basal	Time ON enhanced	Time OFF basal	Time OFF enhanced	ON freq. basal (#/min)	ON freq. enhanced (#/min)
0.87	0.87	553.7	15.3	0.002	0.084

Table E.7: Statistics for the best fitting three state, two regime model. The times refer to averaged times in minutes.

clone	EP genomic distance (bp)	Time basal	Time enhanced	Time ON	Time OFF	ON freq. (#/min)	ON freq. (#/h)
5G7	-4942	122.8	126.5	0.87	30.5	0.044	2.62
5G3	-29865	189.9	126.5	0.87	38.4	0.035	2.1
5F11	-149334	633.4	126.5	0.87	85.2	0.016	0.95
5E10	-232911	2446.4	126.5	0.87	212.3	0.006	0.38

Table E.8: List of oligonucleotide sequences used in this study, including gRNA sequences for MS2 insertion, primer sequences used for genotyping and Splinkerette primers.

gRNA targetting splitGFP/SV40		forward	caccgCTCAGTGTCCATAGCGACGC
		reverse	aaaacGCGTCGCTATGGACACTGAGc
Genotyping MS2 insert	complete insert	forward	CTTTATAGAAGAAATTTTGAG
		reverse	AGGGGTCCTTAATTAATAC
	5-prime	forward	AAATAAACCTCGATATACAG
		reverse	GATTTCTGTGTAATGTGTCT
	3-prime	forward	CCGACAACCCACAAACTTAC
		reverse	GGGACCCATCCCAAAACTT
Splinkerette	Adaptor	HMSpAa	CGAAGAGTAACCGTTGCTAGGAGAGACCGT GGCTGAATGAGACTGGTGTCGACACTAGTGG
		HMSpBb	GATCCCACTAGTGTCGACACCAGTCTCTAATT TTTTTTTTCAAAAAAA
	PCR1 5-prime	HMSp1	CGAAGAGTAACCGTTGCTAGGAGAGACC
		PB5-1	CAAAATCAGTGACACTTACCGCATTGACAA
	PCR2 5-prime	HMSp2	GTGGCTGAATGAGACTGGTGTCGAC
		PB5-2	CTTACCGCATTGACAAGCACGCCTCACGGG
	PCR1 3-prime	HMSp1	CGAAGAGTAACCGTTGCTAGGAGAGACC
		PB3-1	TAAATAAACCTCGATATACAGACCGATAAA
	PCR2 3-prime	HMSp2	GTGGCTGAATGAGACTGGTGTCGAC
		PB3-2	ATATACAGACCGATAAAACACATGCGTCAA

Table E.9: EGFP smRNA-FISH probe sequences.

cggtgaacagctcctcgc	ggcatggcggacttgaag	caggatgttgccgtcctc	cgatgggggtgttctgct
gaccaggatgggcaccac	ctcctggacgtagccttc	gttgtggctgttgtagtt	caggtagtggttgtcggg
gtttacgtcgccgtccag	gccgtcgtccttgaagaa	tgtcggccatgatatagа	ttaaggcggactgggtgc
acacgctgaacttgtggc	tcggcgcgggtcttgtag	ttgatgccgttcttctgc	cgttggggtctttgctta
acttcagggtcagcttgc	tgtcgccctcgaacttca	ggcggatcttgaagttca	accatgtgatcgcgcttc
ttgccggtggtgcagatg	ctcgatgcggttcaccag	ctgccgtcctcgatgttg	ggtcacgaactccagcag
gtagcggctgaagcactg	tgaagtcgatgcccttca	gtagtggtcggcgagctg	cttgtacagctcgtccat
agtcgtgctgcttcatgt			

Table E.10: Enhancer insertion sites of cell lines generated in this study.

clone	chromosome	strand	insertion site	E-P genomic distance (bp)	contact probability (cp)
5B06	chr15	+	11649416	2044	1
5B10	chr15	+	11670915	23543	0.1791502073
5D07	chr15	+	11338632	-308740	0.0555184076
5D11	chr15	+	11643227	-4145	1
5E02	chr15	-	11686806	39434	0.110271569
5E10	chr15	+	11414461	-232911	0.04666346
5F11	chr15	+	11498038	-149334	0.06280125
5G02	chr15	-	11686806	39434	0.110271569
5G03	chr15	-	11617507	-29865	0.2214260552
5G07	chr15	-	11642430	-4942	1
6A12	chr15	-	11639773	-7599	1
6C06	chr15	+	11498038	-149334	0.06280125
6D09	chr15	-	11625645	-21727	0.3250848675
6E02	chr15	+	11641496	-5876	1
6G03	chr10	-	82272353	70624981	0

Table E.11: Statistics of live-cell imaging data

clone	EP genomic distance (bp)	technical replicates	tracks	active transcrip. time	inactive transcrip. time
5G7	-4942	5	1299	1282	2525
5G3	-29865	3	1050	1021	2043
5F11	-149334	6	1679	584	2244
5E10	-232911	5	2490	143	2630
6G3	chr10	7	1534	0	1534

Table E.12: Statistics of smRNA-FISH data

clone	e-p genomic distance (bp)	technical replicates	number of cells
5B10	23543	3	1150
5E10	-232911	3	891
5F11	-149334	3	1203
5G3	-29865	3	969
5G7	-4942	3	952
6A12	-7599	3	1066
6G3	chr10	3	1119
E14 ddCTCF	neg. ctrl.	3	1064

Chapter 3

Discussion

In this thesis, I investigated the effect of EP genomic distance on burst dynamics at the promoter. I presented an experimental setup where only the position of the enhancer relative to its cognate promoter was altered in a 'neutral' genomic region. This setup allowed for the quantitative measurement of burst dynamics as a function of EP genomic distance. My findings revealed that only the duration between bursts (interburst duration) was modulated by EP genomic distance, while the bursts themselves remained unaffected.

In collaboration with Gregory Roth, we interpreted these findings using mathematical models to describe how promoter kinetics are impacted by EP genomic distance. We expanded on the previously reported mathematical model for the ectopic Sox2 promoter used in this study from Zuin *et al.* [9], which proposed that the promoter switches between 'basal' and 'enhanced' regimes. In the expanded model, in the 'basal' and 'enhanced' regimes, the promoter can occupy deep-OFF, OFF or ON states, with the key difference between regimes being the rate of switching from OFF to ON (k_{on}). As a result, EP genomic distance affects the rate of the promoter to switch from the 'basal' to the 'enhanced' regime.

In this chapter, I will place these findings in the context of the current state of knowledge in the field and explore their implications. I will also discuss the strengths and limitations of this study and provide an outlook on how this work can be follow up to gain further insight into open questions.

3.1 Multilayered enhancer function on promoter burst kinetics

As demonstrated in Section 1.2.5 and Section 1.4, enhancers function through an ensemble of many factors. Molecular processes and factors that modulate them such as TF and co-factor concentration, binding affinities or the chromatin environment have been linked to promoter kinetics and resulting burst dynamics. At the same time, the structural organization of the *cis*-regulatory landscape, such as chromatin structure,

genomic separation or chromatin dynamics, should influence burst dynamics and promoter kinetics. What remains unclear is if all factors of importance have been found and how these factors interact to create the unique burst dynamics observed for individual genes. To better understand this question, perturbations are a useful tool to gain valuable insights into each aspect independently. By examining how specific changes affect burst dynamics, we as a field can begin to identify patterns and rules governing how different combinations of these factors produce the final burst dynamics.

Before discussing the individual layers of enhancer function, it is useful to examine cell differentiation, as it represents a context in which all layers of enhancer function are modulated: global concentrations of TFs and co-factors change, along with the binding kinetics at the promoter and enhancer due to altered chromatin accessibility. The *cis*-regulatory landscape adapts as well, such as through changes in the contact probability between an enhancer and its promoter. Studies in the mouse beta-globin locus have demonstrated that erythroid maturation correlates with heightened EP interactions, leading to increased transcriptional burst frequencies and sizes [124]. Similarly, a genome-wide study in mouse fibroblast differentiation showed that for most genes, only burst frequency changed during differentiation [142]. Together, these findings show that differentiation modulates burst dynamics in a gene-specific manner with changes in burst frequency being a common factor across genes.

To better dissect the layers of enhancer function, several studies have investigated the effects of partial or full enhancer deletions. They can be interpreted as lowering the local enrichment of TFs and co-factors without changing their global concentrations or the surrounding *cis*-regulatory landscape, providing a powerful tool to investigate the information transferred between an enhancer and its cognate promoter. These experiments have highlighted the role of enhancers in modulating both transcriptional burst frequency and size [124, 25] or in some cases, burst frequency alone [142].

On the structural side, studies have examined forced or stabilized contacts between enhancers and promoters. The strength of these approaches lies in the fact that they do not alter the biochemical composition of the enhancer or promoter, meaning they do not change local TF or co-factor enrichment and their binding kinetics. Instead, they reduce the dynamic nature of EP contacts, eliminating the potentially rate-limiting step of these two elements coming into contact. In *Drosophila*, stabilized binding between an enhancer and its promoter has been shown to affect transcriptional activation [12, 121]. In mammals, forced EP contacts have also been associated with transcriptional activation [122, 123, 124], with direct evidence indicating that EP contact affects burst frequency [61]. However, as discussed in Section 1.4, it remains unclear how the dynamics of these contacts modulate burst dynamics.

To further untangle the impact of all layers of enhancer function, in my PhD work, I investigated an additional layer: EP genomic distance. In this study, I report on a system in which only the genomic distance between an EP was altered, and found that this change selectively affected burst frequency. The advantage of this approach

is that it does not alter the biochemical composition of the enhancer or promoter, nor does it change the dynamics of their contacts in ways that may be physiologically irrelevant. Instead, only the genomic distance, and thus the population-average contact probabilities, was varied, while preserving the underlying contact dynamics in a physiologically relevant context. One other study in mammals examined changes in genomic distance, alongside alterations in the *cis*-regulatory landscape by modifying TAD size, and found a selective affect on burst frequency [131]. Additionally, in *Drosophila*, changes in the genomic distance between an enhancer-promoter pair also resulted in a selective effect on burst frequency [12]. Both my results and those from these studies demonstrate that altering genomic distance has a selective impact on burst frequency, suggesting that burst frequency may be the predominant burst parameter modulated by changes in genomic distance.

As a field, we are only beginning to understand how enhancers modulate burst dynamics at a promoter. Modulation of burst frequency appears to be a central mechanism, as it was consistently observed in the studies mentioned above. However, the underlying rules and patterns governing this communication, such as the role of TF concentration or genomic distance, are only starting to be uncovered. In the future, it will be important to explore how these different layers of enhancer function interact. For example, how is genomic distance related to molecular processes at the enhancer and promoter? Does changing EP genomic distance influence the concentration of key proteins at the promoter or enhancer, for example of Mediator? The experimental framework presented in this work could serve as a foundation for further investigations into how multiple aspects of enhancer function combine to regulate transcriptional bursting.

3.2 Time-scales of transcription and enhancer-promoter contacts

A central question regarding EP communication is how enhancers and promoters transfer information (cf. Section 1.4). Do they need to come into close physical proximity, within just a few nanometers, or can they communicate effectively over larger distances, such as hundreds of nanometers? Additionally, how stable do these contacts need to be for effective communication? While this study does not provide a definitive answer to these questions, it has important implications for how experiments aimed at addressing these issues have been interpreted so far.

In *Drosophila*, three-color imaging studies, which visualize the positions of an enhancer, a promoter and nascent transcription, have revealed a clear correlation between EP proximity and the initiation of a transcriptional burst [12, 121]. However, similar studies in mammals failed to observe such a correlation [125, 126].

This lack of correlation in mammals may be partly due to technical limitations. Live-cell imaging is subject to spatial resolution limits, typically around 100 nm in two-

color co-localization experiments. Resolving the position of an enhancer and promoter using live-cell imaging would not be sufficient to distinguish between a direct contact model or a communication over a distance model for EP communication, meaning we cannot define the radius of action. In the above mentioned experiments in mammals [125, 126], many events classified as contacts may not represent true EP interactions, as the effective radius of action remains unknown. Additionally, polymer simulations suggest that EP contacts might occur on the order of seconds [8]. The studies in *Drosophila* [12, 121] primarily investigated stabilized EP contacts, which may increase this timescale, while the studies in mammalian cells [125, 126] examined more dynamic contacts. With time resolutions in the tens of seconds, the studies in mammalian cells [125, 126] may have missed transient contact events, thus failing to detect a correlation between contacts and bursting.

In addition to these technical limitations, we also do not understand if all EP contacts are functional. An argument for non-functional contacts can be drawn by comparing this study with chromatin dynamics studies [102, 103]. Mach *et al.* [102] imaged bacterial operator arrays at the same genomic locus as used in this study. Imaging with a 30-second time resolution revealed chromatin contact frequencies of about one contact every six minutes, with an average duration of five minutes for a genomic separation of 150 kb. One of the genomic separations investigated in this study was 149 kb, which exhibits an average burst frequency of one burst every three hours and an average burst duration of 1.5 minutes. Although the contact and burst durations are on the same order of magnitude, the frequencies differ by a factor of 30, suggesting that some contacts may be non-functional or that several consecutive contacts are required to form a single functional interaction. It is important to note that these inferred EP contact dynamics are based on operator array imaging in the absence of an enhancer and promoter.

The model we propose in Fig. 2.5b, may explain why a temporal correlation between EP contacts and bursting is not always observed, even if the resolution limits of microscopy are overcome or non-functional contacts are excluded. In this model, the promoter alternates between 'basal' and 'enhanced' regimes, where it can occupy deep-OFF, OFF, or ON states. The key difference between these regimes is the rate of switching from the OFF to ON state (k_{on}). EP genomic distance, and thus their contact probability and average physical proximity, affects the rate of switching from the 'basal' to the 'enhanced' regime. However, the promoter's ability to decouple contacts from burst initiation remains intact, as the probability of the promoter being in the ON state (k_{on}/k_{off}) and its initiation rate are not modulated by EP genomic distance. Testing the general applicability of this model will be an important future direction (cf. Section 3.3).

Despite the uncertainties surrounding the temporal coupling of contacts and transcription, as well as the definition of functional contacts and their radius of action, this work lays a foundation for addressing these questions in future research. While the exact radius of action or definition of a functional contact remains unclear, it would be valuable to investigate how contact dynamics, such as frequency and du-

ration, change as a function of genomic separation. By linking the changes in burst dynamics observed in this study to the chromatin contact dynamics as a function of genomic separation, we can establish connections without needing to strictly define what a functional contact is or its radius. This relationship could then guide our interpretation of future three-color imaging experiments, allowing us to either to classify specific contacts as functional or to identify cases where EP contacts and transcriptional bursts are temporally decoupled.

3.3 Modeling the ectopic Sox2 promoter

The model we propose in Fig. 2.5b differs from previously suggested multi-state models (cf. Section 1.5) and aims to describe our experimental findings more comprehensively. Our decision to investigate this model was driven by the identification of three distinct time regimes in the interburst duration (~3, ~25, and >160 minutes) and the observed correlation in these interburst durations. This model is a natural extension of what was previously proposed for the ectopic Sox2 promoter used in this study based on steady-state measurements [9]. In this extended model, the promoter transitions between 'basal' and 'enhanced' regimes, within which it can occupy deep-OFF, OFF, or ON states. The primary distinction between these regimes lies in the rate of switching from the OFF to the ON state (k_{on}). While the model's assumptions and parameters are inferred from our experiments, it is important to acknowledge that, like all models, it has limitations and may not necessarily apply universally and beyond the specific ectopic Sox2 promoter examined in this study.

The first assumption of the model arises from the presence of three distinct time regimes in the interburst durations. These multiple timescales suggest the existence of more OFF states than typically considered. For instance, a comparable number of time regimes was observed for a core promoter in *Drosophila* [59], where the shortest interburst duration was associated with a rapid initiation rate, which we excluded in our model. In mammalian systems, previous studies reported only two time regimes [25, 131], implying that the three regimes we observe may be unique to the ectopic Sox2 promoter.

Alternatively, the discrepancy in the number of time regimes could also stem from technical limitations inherent to different microscopy setups. For example, comparing to Rodriguez *et al.* [25], imaging over a 14-hour period with a frame rate of 100 seconds could not resolve our fast and intermediate time regimes (~3 and ~25 minutes). Similarly, Cheng *et al.* [131] used two OFF states to explain the presence of non-bursting cells within a 1-hour imaging window at a 30-second frame rate, which would fail to resolve the intermediate and long time regimes from this study (~25 and >160 minutes). Our experimental setup allowed imaging for 5 hours with a time resolution of 30 seconds, enabling the detection of these three regimes. However, because our analysis did not directly account for cells that were not bursting during the imaging period, we cannot rule out the possibility of additional, longer

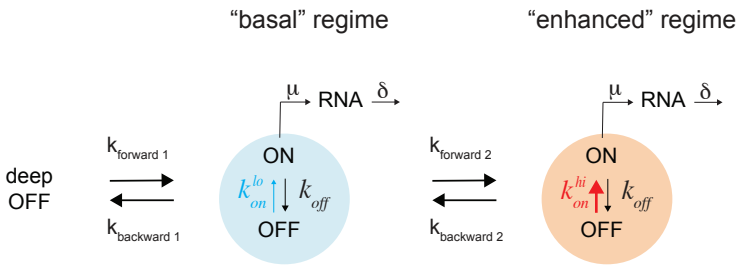

Figure 3.1: Alternative mathematical model to describe the ectopic Sox2 promoter.

time regimes beyond 160 minutes.

The next assumption of the model arises from the time correlation observed in interburst durations, indicating the presence of at least two different ON states. These ON states are entered by distinct k_{on} rates but share similar RNA initiation rates and rates for transitioning back to an OFF state. Such correlations have not been reported in previous studies. Our model addresses this aspect by allowing for two ON states with varying related k_{on} rates.

Although the proposed model meets the requirements to explain our data, it is not the only model capable of doing so. Ultimately, any mathematical model serves as a tool to represent a biological system and interpret its kinetic behavior. The original model from Zuin *et al.* [9], which we extended in Fig. 2.5b, was used to explain the nonlinear relationship between transcriptional output and contact probability potentially arising from multiple EP contacts. We chose it to explain a similar biological system. An alternative model that also meets our assumptions involves three distinct regimes: a deep-OFF regime (where no transcription occurs) and two two-state-like regimes ('basel' and 'enhanced'), each differing in k_{on} rates (Fig. 3.1). However, it is unclear what molecular processes these regimes would correspond to, how it would relate to multiple EP contacts and how EP genomic distance would influence the rates in this alternative model.

Moving forward, testing the proposed model against alternative models will be crucial. The choice of the most appropriate model will depend on its ability to fit the experimental data and predict measurable outcomes using orthogonal approaches such as smRNA-FISH or perturbation experiments. By comparing how well each model captures the dynamics we observe, we can refine our understanding of the molecular mechanisms underlying transcriptional bursting.

Bibliography

1. Lital Bentovim, Timothy T Harden, and Angela H DePace. Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development*, 144(21):3855–3866, November 2017.

2. Steven L McKnight and Oscar L Miller, Jr. Post-replicative nonribosomal transcription units in D. melanogaster embryos. *Cell*, 17(3):551–563, July 1979.

3. Ido Golding, Johan Paulsson, Scott M Zawilski, and Edward C Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, December 2005.

4. Yihan Wan, Dimitrios G Anastasakis, Joseph Rodriguez, Murali Palangat, Prabhakar Gudla, George Zaki, Mayank Tandon, Gianluca Pegoraro, Carson C Chow, Markus Hafner, and Daniel R Larson. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*, 184(11):2878–2895.e20, May 2021.

5. Minoru S Ko. A stochastic model for gene induction. *J. Theor. Biol.*, 153(2): 181–194, November 1991.

6. Jean Peccoud and Bernard Ycart. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.*, 48(2):222–234, October 1995.

7. Koji Kawasaki and Takashi Fukaya. Regulatory landscape of enhancer-mediated transcriptional activation. *Trends Cell Biol.*, 34(10):826–837, October 2024.

8. Jin H Yang and Anders S Hansen. Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation. *Nat. Rev. Mol. Cell Biol.*, 25(7):574–591, July 2024.

9. Jessica Zuin, Gregory Roth, Yinxiu Zhan, Julie Cramard, Josef Redolfi, Ewa Piskadlo, Pia Mach, Mariya Kryzhanovska, Gergely Tihanyi, Hubertus Kohler, Mathias Eder, Christ Leemans, Bas van Steensel, Peter Meister, Sebastien Smallwood, and Luca Giorgetti. Nonlinear control of transcription through enhancer–promoter interactions. *Nature*, 604(7906):571–577, April 2022.

10. Niels J Rinzema, Konstantinos Sofiadis, Sjoerd J D Tjalsma, Marjon J A M Verstegen, Yuva Oz, Christian Valdes-Quezada, Anna Karina Felder, Teodora Filipovska, Stefan van der Elst, Zaria de Andrade dos Ramos, Ruiqi Han, Peter H L Krijger, and Wouter de Laat. Building regulatory landscapes reveals that

an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat. Struct. Mol. Biol.*, 29(6):563–574, June 2022.

11. Henry Thomas, Songjie Feng, Marie Huber, Vincent Loubiere, Daria Vanina, Mattia Pitasi, Alexander Stark, and Christa Buecker. Enhancer cooperativity can compensate for loss of activity over large genomic distances. *bioRxiv*, (https://doi.org/10.1101/2023.12.06.570399), December 2023.

12. David B Brückner, Hongtao Chen, Lev Barinov, Benjamin Zoller, and Thomas Gregor. Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, June 2023.

13. Nicholas J Fuda, M Behfar Ardehali, and John T Lis. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, 461(7261):186–192, September 2009.

14. Joseph Rodriguez and Daniel R Larson. Transcription in living cells: Molecular mechanisms of bursting. *Annu. Rev. Biochem.*, 89(1):189–212, June 2020.

15. Joseph V W Meeussen and Tineke L Lenstra. Time will tell: comparing timescales to gain insight into transcriptional bursting. *Trends Genet.*, 40(2):160–174, February 2024.

16. Shasha Chong, Chongyi Chen, Hao Ge, and X Sunney Xie. Mechanism of transcriptional bursting in bacteria. *Cell*, 158(2):314–326, July 2014.

17. Heta P Patel, Stefano Coppola, Wim Pomp, Umberto Aiello, Ineke Brouwer, Domenico Libri, and Tineke L Lenstra. DNA supercoiling restricts the transcriptional bursting of neighboring eukaryotic genes. *Mol. Cell*, 83(10):1573–1587.e8, May 2023.

18. Aleksander T Szczurek, Emilia Dimitrova, Jessica R Kelley, Neil P Blackledge, and Robert J Klose. The polycomb system sustains promoters in a deep OFF state by limiting pre-initiation complex formation to counteract transcription. *Nat. Cell Biol.*, 26(10):1700–1711, October 2024.

19. Benjamin T Donovan, Anh Huynh, David A Ball, Heta P Patel, Michael G Poirier, Daniel R Larson, Matthew L Ferguson, and Tineke L Lenstra. Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *EMBO J.*, 38(12):e100809, June 2019.

20. Wim Pomp, Joseph V W Meeussen, and Tineke L Lenstra. Transcription factor exchange enables prolonged transcriptional bursts. *Mol. Cell*, 84(6):1036–1048.e9, March 2024.

21. Alessia Loffreda, Emanuela Jacchetti, Sofia Antunes, Paolo Rainone, Tiziana Daniele, Tatsuya Morisaki, Marco E Bianchi, Carlo Tacchetti, and Davide Mazza. Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nat. Commun.*, 8(1):313, August 2017.

22. Adrien Senecal, Brian Munsky, Florence Proux, Nathalie Ly, Floriane E Braye, Christophe Zimmer, Florian Mueller, and Xavier Darzacq. Transcription factors modulate c-fos transcriptional bursts. *Cell Rep.*, 8(1):75–83, July 2014.

23. Diana A Stavreva, David A Garcia, Gregory Fettweis, Prabhakar R Gudla, George F Zaki, Vikas Soni, Andrew McGowan, Geneva Williams, Anh Huynh, Murali Palangat, R Louis Schiltz, Thomas A Johnson, Diego M Presman, Matthew L Ferguson, Gianluca Pegoraro, Arpita Upadhyaya, and Gordon L Hager. Transcriptional bursting and co-bursting regulation by steroid hormone release pattern and transcription factor mobility. *Mol. Cell*, 75(6):1161–1177.e11, September 2019.

24. Achim P Popp, Johannes Hettich, and J Christof M Gebhardt. Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Res.*, 49(11):6249–6266, June 2021.

25. Joseph Rodriguez, Gang Ren, Christopher R Day, Keji Zhao, Carson C Chow, and Daniel R Larson. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell*, 176(1-2):213–226.e18, January 2019.

26. Dhana Friedrich, Laura Friedel, Ana Finzel, Andreas Herrmann, Stephan Preibisch, and Alexander Loewer. Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency. *Mol. Syst. Biol.*, 15 (12):e9068, December 2019.

27. Antonina Hafner, José Reyes, Jacob Stewart-Ornstein, Michael Tsabar, Ashwini Jambhekar, and Galit Lahav. Quantifying the central dogma in the p53 pathway in live single cells. *Cell Syst.*, 10(6):495–505.e4, June 2020.

28. Simona Patange, David A Ball, Yihan Wan, Tatiana S Karpova, Michelle Girvan, David Levens, and Daniel R Larson. MYC amplifies gene expression through global changes in transcription factor dynamics. *Cell Rep.*, 38(4):110292, January 2022.

29. William F Richter, Shraddha Nayak, Janet Iwasa, and Dylan J Taatjes. The mediator complex as a master regulator of transcription by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.*, 23(11):732–749, November 2022.

30. Zhengjian Zhang, Brian P English, Jonathan B Grimm, Stephanie A Kazane, Wenxin Hu, Albert Tsai, Carla Inouye, Changjiang You, Jacob Piehler, Peter G Schultz, Luke D Lavis, Andrey Revyakin, and Robert Tjian. Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.*, 30(18):2106–2118, September 2016.

31. Vu Q Nguyen, Anand Ranjan, Sheng Liu, Xiaona Tang, Yick Hin Ling, Jan Wisniewski, Gaku Mizuguchi, Kai Yu Li, Vivian Jou, Qinsi Zheng, Luke D Lavis,

Timothée Lionnet, and Carl Wu. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. *Mol. Cell*, 81(17):3560–3575.e6, September 2021.

32. Diane K Hawley and Robert G Roeder. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J. Biol. Chem.*, 262(8):3452–3461, March 1987.

33. Katjana Tantale, Florian Mueller, Alja Kozulic-Pirher, Annick Lesne, Jean Marc Victor, Marie Cecile Robert, Serena Capozi, Racha Chouaib, Volker Bäcker, Julio Mateos-Langerak, Xavier Darzacq, Christophe Zimmer, Eugenia Basyuk, and Edouard Bertrand. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.*, 7(1):12248, July 2016.

34. Ineke Brouwer, Emma Kerklingh, Fred van Leeuwen, and Tineke L Lenstra. Dynamic epistasis analysis reveals how chromatin remodeling regulates transcriptional bursting. *Nat. Struct. Mol. Biol.*, 30(5):692–702, May 2023.

35. Shun-Ichi Sekine, Haruhiko Ehara, Tomoya Kujirai, and Hitoshi Kurumizaka. Structural perspectives on transcription in chromatin. *Trends Cell Biol.*, 34(3): 211–224, March 2024.

36. Yi Luo, Justin A North, Sean D Rose, and Michael G Poirier. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.*, 42(5):3017–3027, March 2014.

37. Fangjie Zhu, Lucas Farnung, Eevi Kaasinen, Biswajyoti Sahu, Yimeng Yin, Bei Wei, Svetlana O Dodonova, Kazuhiro R Nitta, Ekaterina Morgunova, Minna Taipale, Patrick Cramer, and Jussi Taipale. The interaction landscape between transcription factors and the nucleosome. *Nature*, 562(7725):76–81, October 2018.

38. Siddharth S Dey, Jonathan E Foley, Prajit Limsirichai, David V Schaffer, and Adam P Arkin. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol. Syst. Biol.*, 11(5):806, May 2015.

39. Gonzalo Millán-Zambrano, Adam Burton, Andrew J Bannister, and Robert Schneider. Histone post-translational modifications - cause and consequence of genome function. *Nat. Rev. Genet.*, 23(9):563–580, September 2022.

40. Naoya Kitamura and James J Galligan. A global view of the human post-translational modification landscape. *Biochem. J.*, 480(16):1241–1265, August 2023.

41. Yanjun Zhang, Zhongxing Sun, Junqi Jia, Tianjiao Du, Nachuan Zhang, Yin Tang, Yuan Fang, and Dong Fang. Overview of histone modification. *Adv. Exp. Med. Biol.*, 1283:1–16, 2021.

42. Brian T Weinert, Takeo Narita, Shankha Satpathy, Balaji Srinivasan, Bogi K Hansen, Christian Schölz, William B Hamilton, Beth E Zucconi, Wesley W Wang, Wenshe R Liu, Joshua M Brickman, Edward A Kesicki, Albert Lai, Kenneth D Bromberg, Philip A Cole, and Chunaram Choudhary. Time-resolved analysis reveals rapid dynamics and broad scope of the CBP/p300 acetylome. *Cell*, 174(1):231–244.e12, June 2018.

43. Menno P Creyghton, Albert W Cheng, G Grant Welstead, Tristan Kooistra, Bryce W Carey, Eveline J Steine, Jacob Hanna, Michael A Lodato, Garrett M Frampton, Phillip A Sharp, Laurie A Boyer, Richard A Young, and Rudolf Jaenisch. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.*, 107(50):21931–21936, December 2010.

44. Alvaro Rada-Iglesias, Ruchi Bajpai, Tomek Swigut, Samantha A Brugmann, Ryan A Flynn, and Joanna Wysocka. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333):279–283, February 2011.

45. José Viñuelas, Gaël Kaneko, Antoine Coulon, Elodie Vallin, Valérie Morin, Camila Mejia-Pous, Jean-Jacques Kupiec, Guillaume Beslon, and Olivier Gandrillon. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol.*, 11(1):15, February 2013.

46. Damien Nicolas, Benjamin Zoller, David M Suter, and Felix Naef. Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U. S. A.*, 115(27):7153–7158, July 2018.

47. Liang-Fu Chen, Yen Ting Lin, David A Gallegos, Mariah F Hazlett, Mariana Gómez-Schiavon, Marty G Yang, Breanna Kalmeta, Allen S Zhou, Liad Holtzman, Charles A Gersbach, Jörg Grandl, Nicolas E Buchler, and Anne E West. Enhancer histone acetylation modulates transcriptional bursting dynamics of neuronal activity-inducible genes. *Cell Rep.*, 26(5):1174–1188.e5, January 2019.

48. David S Gilmour and John T Lis. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in drosophila melanogaster cells. *Mol. Cell. Biol.*, 6(11):3984–3989, November 1986.

49. Ann E Rougvie and John T Lis. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, 54(6):795–804, September 1988.

50. Leighton Core and Karen Adelman. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.*, 33(15-16):960–982, August 2019.

51. Ginger W Muse, Daniel A Gilchrist, Sergei Nechaev, Ruchir Shah, Joel S Parker, Sherry F Grissom, Julia Zeitlinger, and Karen Adelman. RNA polymerase is poised for activation across the genome. *Nat. Genet.*, 39(12):1507–1511, December 2007.

52. Julia Zeitlinger, Alexander Stark, Manolis Kellis, Joung-Woo Hong, Sergei Nechaev, Karen Adelman, Michael Levine, and Richard A Young. RNA polymerase stalling at developmental control genes in the drosophila melanogaster embryo. *Nat. Genet.*, 39(12):1512–1516, December 2007.

53. Telmo Henriques, Daniel A Gilchrist, Sergei Nechaev, Michael Bern, Ginger W Muse, Adam Burkholder, David C Fargo, and Karen Adelman. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell*, 52(4):517–528, November 2013.

54. Iris Jonkers, Hojoong Kwak, and John T Lis. Genome-wide dynamics of pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife*, 3(3):e02407, April 2014.

55. Martin S Buckley, Hojoong Kwak, Warren R Zipfel, and John T Lis. Kinetics of promoter pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes Dev.*, 28(1):14–19, January 2014.

56. Wanqing Shao and Julia Zeitlinger. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.*, 49(7):1045–1051, July 2017.

57. Arnaud R Krebs, Dilek Imanci, Leslie Hoerner, Dimos Gaidatzis, Lukas Burger, and Dirk Schübeler. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell*, 67(3):411–422.e4, August 2017.

58. Barbara Steurer, Roel C Janssens, Bart Geverts, Marit E Geijer, Franziska Wienholz, Arjan F Theil, Jiang Chang, Shannon Dealy, Joris Pothof, Wiggert A van Cappellen, Adriaan B Houtsmuller, and Jurgen A Marteijn. Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.*, 115(19):E4368–E4376, May 2018.

59. Virginia L Pimmett, Matthieu Dejean, Carola Fernandez, Antonio Trullo, Edouard Bertrand, Ovidiu Radulescu, and Mounia Lagha. Quantitative imaging of transcription in living drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. *Nat. Commun.*, 12(1):4504, July 2021.

60. Katjana Tantale, Encar Garcia-Oliver, Marie-Cécile Robert, Adèle L'Hostis, Yueyuxiao Yang, Nikolay Tsanov, Rachel Topno, Thierry Gostan, Alja Kozulic-Pirher, Meenakshi Basu-Shrivastava, Kamalika Mukherjee, Vera Slaninova, Jean-Christophe Andrau, Florian Mueller, Eugenia Basyuk, Ovidiu Radulescu, and

Edouard Bertrand. Stochastic pausing at latent HIV-1 promoters generates transcriptional bursting. *Nat. Commun.*, 12(1):4503, July 2021.

61. Caroline R Bartman, Nicole Hamagami, Cheryl A Keller, Belinda Giardine, Ross C Hardison, Gerd A Blobel, and Arjun Raj. Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol. Cell*, 73(3):519–532.e4, February 2019.

62. Julian Banerji, Sandro Rusconi, and Walter Schaffner. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1):299–308, December 1981.

63. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, September 2012.

64. Robin Andersson and Albin Sandelin. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.*, 21(2):71–87, February 2020.

65. Robin Andersson, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, Xiaobei Zhao, Christian Schmidl, Takahiro Suzuki, Evgenia Ntini, Erik Arner, Eivind Valen, Kang Li, Lucia Schwarzfischer, Dagmar Glatz, Johanna Raithel, Berit Lilje, Nicolas Rapin, Frederik Otzen Bagger, Mette Jørgensen, Peter Refsing Andersen, Nicolas Bertin, Owen Rackham, A Maxwell Burroughs, J Kenneth Baillie, Yuri Ishizu, Yuri Shimizu, Erina Furuhata, Shiori Maeda, Yutaka Negishi, Christopher J Mungall, Terrence F Meehan, Timo Lassmann, Masayoshi Itoh, Hideya Kawaji, Naoto Kondo, Jun Kawai, Andreas Lennartsson, Carsten O Daub, Peter Heutink, David A Hume, Torben Heick Jensen, Harukazu Suzuki, Yoshihide Hayashizaki, Ferenc Müller, Alistair R R Forrest, Piero Carninci, Michael Rehli, and Albin Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493): 455–461, March 2014.

66. Robin Andersson, Peter Refsing Andersen, Eivind Valen, Leighton J Core, Jette Bornholdt, Mette Boyd, Torben Heick Jensen, and Albin Sandelin. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.*, 5(1):5336, November 2014.

67. Leighton J Core, André L Martins, Charles G Danko, Colin T Waters, Adam Siepel, and John T Lis. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.*, 46(12): 1311–1320, December 2014.

68. Chunhui Hou, Hui Zhao, Keiji Tanimoto, and Ann Dean. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. U. S. A.*, 105(51):20398–20403, December 2008.

69. Andrew Field and Karen Adelman. Evaluating enhancer function and transcription. *Annu. Rev. Biochem.*, 89(1):213–234, June 2020.

70. Timothy J Stasevich, Yoko Hayashi-Takanaka, Yuko Sato, Kazumitsu Maehara, Yasuyuki Ohkawa, Kumiko Sakata-Sogawa, Makio Tokunaga, Takahiro Nagase, Naohito Nozaki, James G McNally, and Hiroshi Kimura. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*, 516 (7530):272–275, December 2014.

71. Takeo Narita, Shinsuke Ito, Yoshiki Higashijima, Wai Kit Chu, Katrin Neumann, Jonas Walter, Shankha Satpathy, Tim Liebner, William B Hamilton, Elina Maskey, Gabriela Prus, Marika Shibata, Vytautas Iesmantavicius, Joshua M Brickman, Konstantinos Anastassiadis, Haruhiko Koseki, and Chunaram Choudhary. Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. *Mol. Cell*, 81(10):2166–2182.e6, May 2021.

72. Liang Ma, Zeyue Gao, Jiegen Wu, Bijunyao Zhong, Yuchen Xie, Wen Huang, and Yihan Lin. Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. *Mol. Cell*, 81(8):1682–1697.e7, April 2021.

73. Inwha Baek, Larry J Friedman, Jeff Gelles, and Stephen Buratowski. Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. *Mol. Cell*, 81(17):3576–3588.e6, September 2021.

74. Julie Soutourina. Transcription regulation by the mediator complex. *Nat. Rev. Mol. Cell Biol.*, 19(4):262–274, April 2018.

75. Sebastian Grünberg, Steven Henikoff, Steven Hahn, and Gabriel E Zentner. Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. *EMBO J.*, 35(22):2435–2446, November 2016.

76. Célia Jeronimo and François Robert. Kin28 regulates the transient association of mediator with core promoters. *Nat. Struct. Mol. Biol.*, 21(5):449–455, May 2014.

77. Natalia Petrenko, Yi Jin, Koon Ho Wong, and Kevin Struhl. Mediator undergoes a compositional change during transcriptional activation. *Mol. Cell*, 64(3):443–454, November 2016.

78. Simon Fishilevich, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, Asher Kohn, Michal Twik, Marilyn Safran, Doron Lancet, and Dana Cohen. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*, 2017, January 2017.

79. Amartya Sanyal, Bryan R Lajoie, Gaurav Jain, and Job Dekker. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–113, September 2012.

80. Zhuoxin Chen, Valentina Snetkova, Grace Bower, Sandra Jacinto, Benjamin Clock, Atrin Dizehchi, Iros Barozzi, Brandon J Mannion, Ana Alcaina-Caro, Javier Lopez-Rios, Diane E Dickel, Axel Visel, Len A Pennacchio, and Evgeny Z Kvon. Increased enhancer-promoter interactions during developmental enhancer activation in mammals. *Nat. Genet.*, 56(4):675–685, April 2024.

81. Sumio Ohtsuki, Michael Levine, and Haini N Cai. Different core promoters possess distinct regulatory activities in the drosophila embryo. *Genes Dev.*, 12 (4):547–556, February 1998.

82. Jennifer E Butler and James T Kadonaga. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.*, 15(19):2515–2519, October 2001.

83. Tamar Juven-Gershon, Jer-Yuan Hsu, and James T Kadonaga. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev.*, 22 (20):2823–2830, October 2008.

84. Hila Shir-Shapira, Anna Sloutskin, Orit Adato, Avital Ovadia-Shochat, Diana Ideses, Yonathan Zehavi, George Kassavetis, James T Kadonaga, Ron Unger, and Tamar Juven-Gershon. Identification of evolutionarily conserved downstream core promoter elements required for the transcriptional regulation of fushi tarazu target genes. *PLoS One*, 14(4):e0215695, April 2019.

85. Drew T Bergman, Thouis R Jones, Vincent Liu, Judhajeet Ray, Evelyn Jagoda, Layla Siraj, Helen Y Kang, Joseph Nasser, Michael Kane, Antonio Rios, Tung H Nguyen, Sharon R Grossman, Charles P Fulco, Eric S Lander, and Jesse M Engreitz. Compatibility rules of human enhancer and promoter sequences. *Nature*, 607(7917):176–184, July 2022.

86. Miguel Martinez-Ara, Federico Comoglio, Joris van Arensbergen, and Bas van Steensel. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol. Cell*, 82(13):2519–2531.e6, July 2022.

87. Tom Misteli. The self-organizing genome: Principles of genome architecture and function. *Cell*, 183(1):28–45, October 2020.

88. Rieke Kempfer and Ana Pombo. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.*, 21(4):207–226, April 2020.

89. Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, October 2009.

90. Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458–472, February 2012.

91. Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, April 2012.

92. Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–385, April 2012.

93. Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and François Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.*, 24(3): 390–400, March 2014.

94. Malte Spielmann, Darío G Lupiáñez, and Stefan Mundlos. Structural variation in the 3D genome. *Nat. Rev. Genet.*, 19(7):453–467, July 2018.

95. Jesse R Dixon, David U Gorkin, and Bing Ren. Chromatin domains: The unit of chromosome organization. *Mol. Cell*, 62(5):668–680, June 2016.

96. Iain F Davidson, Benedikt Bauer, Daniela Goetz, Wen Tang, Gordana Wutz, and Jan-Michael Peters. DNA loop extrusion by human cohesin. *Science*, 366(6471): 1338–1345, December 2019.

97. Yoori Kim, Zhubing Shi, Hongshan Zhang, Ilya J Finkelstein, and Hongtao Yu. Human cohesin compacts DNA by loop extrusion. *Science*, 366(6471):1345–1349, December 2019.

98. Adrian L Sanborn, Suhas S P Rao, Su-Chen Huang, Neva C Durand, Miriam H Huntley, Andrew I Jewett, Ivan D Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K Stamenova, Eric S Lander, and Erez Lieberman Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.*, 112(47):E6456–65, November 2015.

99. Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of chromosomal domains by loop extrusion. *Cell Rep.*, 15(9):2038–2049, May 2016.

100. Ilya M Flyamer, Johanna Gassler, Maxim Imakaev, Hugo B Brandão, Sergey V Ulianov, Nezar Abdennur, Sergey V Razin, Leonid A Mirny, and Kikuë Tachibana-Konwalski. Single-nucleus hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, 544(7648):110–114, April 2017.

101. Bogdan Bintu, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362(6413):eaau1783, October 2018.

102. Pia Mach, Pavel I Kos, Yinxiu Zhan, Julie Cramard, Simon Gaudin, Jana Tünnermann, Edoardo Marchi, Jan Eglinger, Jessica Zuin, Mariya Kryzhanovska, Sebastien Smallwood, Laurent Gelman, Gregory Roth, Elphège P Nora, Guido Tiana, and Luca Giorgetti. Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.*, 54(12):1907–1918, December 2022.

103. Michele Gabriele, Hugo B Brandão, Simon Grosse-Holz, Asmita Jha, Gina M Dailey, Claudia Cattoglio, Tsung-Han S Hsieh, Leonid Mirny, Christoph Zechner, and Anders S Hansen. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, 376(6592):496–501, April 2022.

104. Naoki Kubo, Haruhiko Ishii, Xiong Xiong, Simona Bianco, Franz Meitinger, Rong Hu, James D Hocker, Mattia Conte, David Gorkin, Miao Yu, Bin Li, Jesse R Dixon, Ming Hu, Mario Nicodemi, Huimin Zhao, and Bing Ren. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.*, 28(2):152–161, February 2021.

105. Christina Paliou, Philine Guckelberger, Robert Schöpflin, Verena Heinrich, Andrea Esposito, Andrea M Chiariello, Simona Bianco, Carlo Annunziatella, Johannes Helmuth, Stefan Haas, Ivana Jerković, Norbert Brieske, Lars Wittler, Bernd Timmermann, Mario Nicodemi, Martin Vingron, Stefan Mundlos, and Guillaume Andrey. Preformed chromatin topology assists transcriptional robustness of shh during limb development. *Proc. Natl. Acad. Sci. U. S. A.*, 116(25): 12390–12399, June 2019.

106. Jurian Schuijers, John Colonnese Manteiga, Abraham Selby Weintraub, Daniel Sindt Day, Alicia Viridiana Zamudio, Denes Hnisz, Tong Ihn Lee, and Richard Allen Young. Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. *Cell Rep.*, 23(2):349–360, April 2018.

107. Elphège P Nora, Anton Goloborodko, Anne-Laure Valton, Johan H Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A Mirny, and Benoit G Bruneau. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5):930–944.e22, May 2017.

108. Suhas S P Rao, Su-Chen Huang, Brian Glenn St Hilaire, Jesse M Engreitz, Elizabeth M Perez, Kyong-Rim Kieffer-Kwon, Adrian L Sanborn, Sarah E Johnstone, Gavin D Bascom, Ivan D Bochkov, Xingfan Huang, Muhammad S Shamim, Jaeweon Shin, Douglass Turner, Ziyi Ye, Arina D Omer, James T Robinson, Tamar Schlick, Bradley E Bernstein, Rafael Casellas, Eric S Lander, and Erez Lieberman Aiden. Cohesin loss eliminates all loop domains. *Cell*, 171(2): 305–320.e24, October 2017.

109. Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A Fonseca, Wolfgang Huber, Christian H Haering, Leonid Mirny, and Francois Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551 (7678):51–56, November 2017.

110. Tsung-Han S Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S Hansen, Xavier Darzacq, and Robert Tjian. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.*, 54(12):1919–1932, December 2022.

111. Abrar Aljahani, Peng Hua, Magdalena A Karpinska, Kimberly Quililan, James O J Davies, and A Marieke Oudelaar. Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat. Commun.*, 13(1):2139, April 2022.

112. Sergi Cuartero, Felix D Weiss, Gopuraja Dharmalingam, Ya Guo, Elizabeth Ing-Simmons, Silvia Masella, Irene Robles-Rebollo, Xiaolin Xiao, Yi-Fang Wang, Iros Barozzi, Dounia Djeghloul, Mariane T Amano, Henri Niskanen, Enrico Petretto, Robin D Dowell, Kikuë Tachibana, Minna U Kaikkonen, Kim A Nasmyth, Boris Lenhard, Gioacchino Natoli, Amanda G Fisher, and Matthias Merkenschlager. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat. Immunol.*, 19(9):932–941, September 2018.

113. Maria Cristina Gambetta and Eileen E M Furlong. The insulator protein CTCF is required for correct hox gene expression, but not for embryonic development in drosophila. *Genetics*, 210(1):129–136, September 2018.

114. Anjali Kaushal, Giriram Mohana, Julien Dorier, Isa Özdemir, Arina Omer, Pascal Cousin, Anastasiia Semenova, Michael Taschner, Oleksandr Dergai, Flavia Marzetta, Christian Iseli, Yossi Eliaz, David Weisz, Muhammad Saad Shamim, Nicolas Guex, Erez Lieberman Aiden, and Maria Cristina Gambetta. CTCF loss has limited effects on global genome architecture in drosophila despite critical regulatory functions. *Nat. Commun.*, 12(1):1011, February 2021.

115. Lauren Kane, Iain Williamson, Ilya M Flyamer, Yatendra Kumar, Robert E Hill, Laura A Lettice, and Wendy A Bickmore. Cohesin is required for long-range enhancer action at the shh locus. *Nat. Struct. Mol. Biol.*, 29(9):891–897, September 2022.

116. Lesly Calderon, Felix D Weiss, Jonathan A Beagan, Marta S Oliveira, Radina Georgieva, Yi-Fang Wang, Thomas S Carroll, Gopuraja Dharmalingam, Wanfeng Gong, Kyoko Tossell, Vincenzo de Paola, Chad Whilding, Mark A Ungless, Amanda G Fisher, Jennifer E Phillips-Cremins, and Matthias Merkenschlager. Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *Elife*, 11, April 2022.

117. Hui Huang, Quan Zhu, Adam Jussila, Yuanyuan Han, Bogdan Bintu, Colin Kern, Mattia Conte, Yanxiao Zhang, Simona Bianco, Andrea M Chiariello, Miao Yu, Rong Hu, Melodi Tastemel, Ivan Juric, Ming Hu, Mario Nicodemi, Xiaowei Zhuang, and Bing Ren. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.*, 53(7):1064–1074, July 2021.

118. Charles P Fulco, Joseph Nasser, Thouis R Jones, Glen Munson, Drew T Bergman, Vidya Subramanian, Sharon R Grossman, Rockwell Anyoha, Benjamin R Doughty, Tejal A Patwardhan, Tung H Nguyen, Michael Kane, Elizabeth M Perez, Neva C Durand, Caleb A Lareau, Elena K Stamenova, Erez Lieberman Aiden, Eric S Lander, and Jesse M Engreitz. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.*, 51(12):1664–1669, December 2019.

119. Xizi Chen, Xiaotong Yin, Jiabei Li, Zihan Wu, Yilun Qi, Xinxin Wang, Weida Liu, and Yanhui Xu. Structures of the human mediator and mediator-bound preinitiation complex. *Science*, 372(6546):eabg0635, June 2021.

120. Olga Kyrchanova and Pavel Georgiev. Mechanisms of enhancer-promoter interactions in higher eukaryotes. *Int. J. Mol. Sci.*, 22(2):671, January 2021.

121. Hongtao Chen, Michal Levo, Lev Barinov, Miki Fujioka, James B Jaynes, and Thomas Gregor. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.*, 50(9):1296–1303, September 2018.

122. Wulan Deng, Jongjoo Lee, Hongxin Wang, Jeff Miller, Andreas Reik, Philip D Gregory, Ann Dean, and Gerd A Blobel. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, 149(6): 1233–1244, June 2012.

123. Wulan Deng, Jeremy W Rupon, Ivan Krivega, Laura Breda, Irene Motta, Kristen S Jahn, Andreas Reik, Philip D Gregory, Stefano Rivella, Ann Dean, and Gerd A Blobel. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, 158(4):849–860, August 2014.

124. Caroline R Bartman, Sarah C Hsu, Chris C-S Hsiung, Arjun Raj, and Gerd A Blobel. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell*, 62(2):237–247, April 2016.

125. Jeffrey M Alexander, Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D Weiner. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife*, 8, May 2019.

126. Angeliki Platania, Cathie Erb, Mariano Barbieri, Bastien Molcrette, Erwan Grandgirard, Marit Ac de Kort, Karen Meaburn, Tiegh Taylor, Virlana M Shchuka, Silvia Kocanova, Guilherme Monteiro Oliveira, Jennifer A Mitchell, Evi Soutoglou, Tineke L Lenstra, Nacho Molina, Argyris Papantonis, Kerstin Bystricky, and Tom Sexton. Competition between transcription and loop extrusion modulates promoter and enhancer dynamics. *bioRxiv*, (http://dx.doi.org/10.1101/2023.04.25.538222), April 2023.

127. Leslie J Mateo, Sedona E Murphy, Antonina Hafner, Isaac S Cinquini, Carly A Walker, and Alistair N Boettiger. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 568(7750):49–54, April 2019.

128. Won-Ki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I Cisse. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361(6400):412–415, July 2018.

129. Benjamin R Sabari, Alessandra Dall'Agnese, Ann Boija, Isaac A Klein, Eliot L Coffey, Krishna Shrinivas, Brian J Abraham, Nancy M Hannett, Alicia V Zamudio, John C Manteiga, Charles H Li, Yang E Guo, Daniel S Day, Jurian Schuijers, Eliza Vasile, Sohail Malik, Denes Hnisz, Tong Ihn Lee, Ibrahim I Cisse, Robert G Roeder, Phillip A Sharp, Arup K Chakraborty, and Richard A Young. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 361(6400), July 2018.

130. Jieru Li, Ankun Dong, Kamola Saydaminova, Hill Chang, Guanshi Wang, Hiroshi Ochiai, Takashi Yamamoto, and Alexandros Pertsinidis. Single-molecule nanoscopy elucidates RNA polymerase II transcription at single genes in live cells. *Cell*, 178(2):491–506.e28, July 2019.

131. Lingling Cheng, Chayan De, Jieru Li, and Alexandros Pertsinidis. Mechanisms of transcription control by distal enhancers from high-resolution single-gene imaging. *bioRxiv*, (https://doi.org/10.1101/2023.03.19.533190), March 2023.

132. Ann Boija, Isaac A Klein, Benjamin R Sabari, Alessandra Dall'Agnese, Eliot L Coffey, Alicia V Zamudio, Charles H Li, Krishna Shrinivas, John C Manteiga, Nancy M Hannett, Brian J Abraham, Lena K Afeyan, Yang E Guo, Jenna K Rimel, Charli B Fant, Jurian Schuijers, Tong Ihn Lee, Dylan J Taatjes, and Richard A Young. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell*, 175(7):1842–1855.e16, December 2018.

133. Manyu Du, Simon Hendrik Stitzinger, Jan-Hendrik Spille, Won-Ki Cho, Choongman Lee, Mohammed Hijaz, Andrea Quintana, and Ibrahim I Cissé. Direct observation of a condensate effect on super-enhancer controlled gene bursting. *Cell*, 187(2):331–344.e17, January 2024.

134. Tyler Heist, Takashi Fukaya, and Michael Levine. Large distances separate coregulated genes in living drosophila embryos. *Proc. Natl. Acad. Sci. U. S. A.*, 116(30):15062–15067, July 2019.

135. Koji Kawasaki and Takashi Fukaya. Functional coordination between transcription factor clustering and gene activity. *Mol. Cell*, 83(10):1605–1622.e9, May 2023.

136. Shasha Chong, Claire Dugast-Darzacq, Zhe Liu, Peng Dong, Gina M Dailey, Claudia Cattoglio, Alec Heckert, Sambashiva Banala, Luke Lavis, Xavier Darzacq, and Robert Tjian. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, 361(6400), July 2018.

137. Jorge Trojanowski, Lukas Frank, Anne Rademacher, Norbert Mücke, Pranas Grigaitis, and Karsten Rippe. Transcription activation is enhanced by multivalent interactions independent of phase separation. *Mol. Cell*, 82(10):1878–1893.e10, May 2022.

138. Shasha Chong, Thomas G W Graham, Claire Dugast-Darzacq, Gina M Dailey, Xavier Darzacq, and Robert Tjian. Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription. *Mol. Cell*, 82(11): 2084–2097.e5, June 2022.

139. Fabian Erdel and Karsten Rippe. Formation of chromatin subcompartments by phase separation. *Biophys. J.*, 114(10):2262–2270, May 2018.

140. Jonathan P Karr, John J Ferrie, Robert Tjian, and Xavier Darzacq. The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer–promoter communication. *Genes Dev.*, 36(1-2):7–16, January 2021.

141. Edward Tunnacliffe and Jonathan R Chubb. What is a transcriptional burst? *Trends Genet.*, 36(4):288–297, April 2020.

142. Anton J M Larsson, Per Johnsson, Michael Hagemann-Jensen, Leonard Hartmanis, Omid R Faridani, Björn Reinius, Åsa Segerstolpe, Chloe M Rivera, Bing Ren, and Rickard Sandberg. Genomic encoding of transcriptional burst kinetics. *Nature*, 565(7738):251–254, January 2019.

143. Julia Falo-Sanjuan, Nicholas C Lammers, Hernan G Garcia, and Sarah J Bray. Enhancer priming enables fast and sustained transcriptional responses to notch signaling. *Dev. Cell*, 50(4):411–425.e8, August 2019.

144. Jiajun Zhang and Tianshou Zhou. Promoter-mediated transcriptional dynamics. *Biophys. J.*, 106(2):479–488, January 2014.

145. Georg Rieckh and Gašper Tkačik. Noise and information transmission in promoters with multiple internal states. *Biophys. J.*, 106(5):1194–1204, March 2014.

146. Edouard Bertrand, Pascal Chartrand, Matthias Schaefer, Shailesh M Shenoy, Robert H Singer, and Roy M Long. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell*, 2(4):437–445, October 1998.

147. Virginia Pimmett, Maria Douaihy, Louise Maillard, Antonio Trullo, Jeremy Dufourt, Helene Lenden-Hasse, Ovidiu Radulescu, and Mounia Lagha. Dissecting the dynamics of coordinated active transcriptional repression in a multicellular organism. *bioRxiv*, (http://dx.doi.org/10.1101/2024.02.05.577724), July 2024.

148. Hannah K Long, Sara L Prescott, and Joanna Wysocka. Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell*, 167(5): 1170–1187, November 2016.

149. Karissa L Hansen, Annie S Adachi, Luca Braccioli, Smit Kadvani, Ryan M Boileau, Bozhena Pokorny, Rini Shah, Erika C Anderson, Moreno Martinovic, Kaite Zhang, Irié Carel, Kenya Bonitto, Robert Blelloch, Geoffrey Fudenberg, Elzo de Wit, and Elphège P Nora. Synergy between *cis*-regulatory elements can render cohesin dispensable for distal enhancer function. *bioRxiv*, (http://dx.doi.org/10.1101/2024.10.04.615095), October 2024.

150. Christina L Jensen, Liang-Fu Chen, Tomek Swigut, Olivia J Crocker, David Yao, Mike C Bassik, James E Ferrell, Alistair N Boettiger, and Joanna Wysocka. Long range regulation of transcription scales with genomic distance in a gene specific manner. *bioRxiv*, (http://dx.doi.org/10.1101/2024.07.19.604327), July 2024.

151. Jordan Yupeng Xiao, Antonina Hafner, and Alistair N Boettiger. How subtle changes in 3D structure can create large changes in transcription. *Elife*, 10, July 2021.

152. Franka Voigt, Hui Zhang, Xianying A Cui, Désirée Triebold, Ai Xin Liu, Jan Eglinger, Eliza S Lee, Jeffrey A Chao, and Alexander F Palazzo. Single-molecule quantification of translation-dependent association of mRNAs with the endoplasmic reticulum. *Cell Rep.*, 21(13):3740–3753, December 2017.

153. Mathias Eder, Christina J I Moene, Lise Dauban, Christ Leemans, and Bas van Steensel. Functional maps of a genomic locus reveal confinement of an enhancer by its target gene. *bioRxiv*, (http://dx.doi.org/10.1101/2024.08.26.609360), July 2024.

154. Ralph S Grand, Lukas Burger, Cathrin Gräwe, Alicia K Michael, Luke Isbel, Daniel Hess, Leslie Hoerner, Vytautas Iesmantavicius, Sevi Durdu, Marco Pregnolato, Arnaud R Krebs, Sébastien A Smallwood, Nicolas Thomä, Michiel Vermeulen, and Dirk Schübeler. BANP opens chromatin and activates CpG-island-regulated genes. *Nature*, 596(7870):133–137, August 2021.

155. Brian E Hew, Ryuei Sato, Damiano Mauro, Ilko Stoytchev, and Jesse B Owens. RNA-guided piggyBac transposition in human cells. *Synth. Biol.*, 4(1):ysz018, July 2019.

156. Imre Gaspar, Frank Wippich, and Anne Ephrussi. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA*, 23(10):1582–1591, October 2017.

157. Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Lecture notes in computer science, pages 265–273. Springer International Publishing, Cham, 2018.

158. Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. Star-convex polyhedra for 3D object detection and segmentation in microscopy. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3655–3662. IEEE, March 2020.

159. Daniel B Allan, Thomas Caswell, Nathan C Keim, Casper M van der Wel, and Ruben W Verweij. soft-matter/trackpy: Trackpy v0.5.0, 2021.

160. E L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, 53(282):457–481, June 1958.

161. Cameron Davidson-Pilon. lifelines: survival analysis in python. *J. Open Source Softw.*, 4(40):1317, August 2019.

Appendix

The appendix includes references to the publications I contributed to during my doctoral studies. The extended manuscript files can be assessed via the links provided. Their content is not discussed in this thesis.

Cohesin and CTCF control the dynamics of chromosome folding
Mach, P.,* Kos, P. I.*, Zhan, Y.*, Cramard, J., Gaudin, S., **Tünnermann, J.**, Marchi, E., Eglinger, J., Zuin, J., Kryzhanovska, M., Smallwood, S., Gelman, L., Roth, G., Nora, E. P., Tiana, G., & Giorgetti, L. Nature Genetics, 54(12), 1907–1918. (2022)
https://doi.org/10.1038/s41588-022-01232-7
*contributed equally

VisuStatR: visualizing motility and morphology statistics on images in R
Harmel, C., Ahmed, S. S., Koch, R., **Tünnermann, J.**, Distler, T., Imle, A., Giorgetti, L., Bahn, E., Fackler, O. T., Graw F. Bioinformatics, 38(10), 2970–2972. (2022)
https://doi.org/10.1093/bioinformatics/btac191