Regulation of transcription by transcription factor residence time and multivalent interactions

Dissertation

submitted by Jorge Trojanowski

2021

Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Jorge Trojanowski, M.Sc.

born in: Augsburg, Germany
Oral examination: April 19 2021

ii

Regulation of transcription by transcription factor residence time and multivalent interactions

Referees: Prof. Dr. Karsten Rippe

Prof. Dr. Thomas Höfer

“The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same.”

Plutarch, *Theseus*

Summary

Transcription is a multistep process that is tightly regulated by transcription factors (TFs). TFs typically comprise two subdomains - a DNA-binding domain (DBD) and an activation domain (AD). Properties of TF target site binding are attributed to the DBD. The AD is thought to determine interactions with the transcriptional machinery, the acquisition of co-activating chromatin modifications and, potentially, the formation of phase-separated nuclear compartments via multivalent interactions. However, it remains unclear if DBD and AD really act independently in determining crucial parameters that govern the induction of transcription. In particular, it is unknown if TF binding kinetics and binding site residence times regulate transcription independent of equilibrium binding parameters and whether phase separation caused by multivalent TF interactions is functionally relevant for activation. In this thesis, experimental and analytical approaches were developed and applied that provide mechanistic insights from the highly informative analysis of TF binding and transcription kinetics. Techniques were introduced to measure TF binding kinetics and to follow transcription using light dependent TF recruitment. These approaches were automated and software packages for the analysis of the resulting data were developed to test a large number of conditions in single cells with high time resolution. The wide applicability of light-induced transcription time courses was demonstrated by two proof-of-concept applications: the detection of transcriptional memory and the discrimination of stochastic models using heterogeneous single-cell trajectories. The framework was then applied to reveal a functional link between binding properties of the DBD, multivalent interactions of the AD and the dynamics of transcriptional (co-)activation. Specifically, the following conclusions could be reached: (1) Reduced TF residence time decreased transcription, even for identical binding site occupancy. (2) Multivalent interactions of the AD stabilized chromatin binding of weakly bound TFs and led to the recruitment of an indirectly bound fraction of molecules. (3) ADs with strong multivalent interactions activated faster and more strongly. (4) Phase-separation into macroscopic droplets did not enhance transcription and could in some conditions even have a suppressive effect. (5) Acetylation of histone at lysine residue 27 (H3K27ac) and the binding of BRD4, which interacts with H3K27ac, were induced by indirectly and transiently bound activators under conditions that were not sufficient to induce RNA production. (6) H3K27ac and BRD4 were not strictly necessary for transcription, but had an enhancing effect. Based on these findings the thesis provides an integrated view of TF activity, in which multiple, interdependent properties of DBD and AD increase transcriptional output. These include long TF residence time, high binding site occupancy, complex stabilization by multivalent interactions and interactions with co-activators, but not phase-separation into macroscopic compartments. These findings provide insights into the different TF features that govern their ability to activate transcription and for the design of synthetic TFs.

Zusammenfassung

Transkription ist ein mehrstufiger Prozess, der von Transkriptionsfaktoren (TF) reguliert wird. TF bestehen typischerweise aus zwei Untereinheiten: der DNA-Bindungsdomäne (DBD) und der Aktivierungsdomäne (AD). Die Eigenschaften der Bindung von TF an ihre Zielbindestellen werden der DBD zugerechnet. Man denkt, dass die AD Interaktionen mit der Transkriptionsmaschinerie, die Anreicherung von koaktivierenden Chromatin-Modifikationen und möglicherweise die Ausbildung von nukleären Subkompartimenten durch Phasentrennung mittels multivalenter Interaktionen bestimmt. Es bleibt jedoch unklar, ob DBD und AD die wichtigen Parameter, die den Induktionsprozess bestimmen, wirklich unabhängig voneinander kontrollieren. Es ist insbesondere ungewiss, ob die TF-Bindungskinetik und die Bindezeit die Transkription unabhängig von den Gleichgewichtsparametern regulieren und ob Phasentrennung, die durch mutivalente TF-Interaktionen ausgelöst wird, funktionell wichtig für den Aktivierungsprozess ist. In der vorliegenden Arbeit wurden experimentelle und analytische Ansätze entwickelt und angewendet, die mechanistische Erkenntnisse aus der hochinformativen Analyse der TF-Bindungs- und Transkriptionskinetik gewinnen. Ein besonderer Schwerpunkt wurde darauf gelegt die TF-Bindungskinetik zu messen und den transkriptionellen Induktionsprozess mit hoher Zeitauflösung in einer großen Zahl von Einzelzellen mithilfe von lichtinduzierter TF Rekrutierung zu verfolgen. Die experimentellen Abläufe wurden automatisiert und es wurden Software-Pakete für die Datenanalyse entwickelt, um eine große Anzahl von Bedingungen in Einzelzellen mit hoher Zeitauflösung zu messen. Der breite Anwendungsbereich von lichtinduzierten Transkriptionszeitverläufen wurde durch zwei proof-of-concept Anwendungen demonstriert: Erstens für den Nachweis von transkriptionellen Erinnerungseffekten und zweitens für die Entscheidung zwischen konkurrierenden stochastischen Modellen unter Verwendung von heterogenen Einzelzell-Trajektorien. Die experimentellen und analytischen Ansätze wurden dann angewendet, um eine funktionelle Verbindung zwischen Bindungseigenschaften der DBD, multivalenten Interaktionen der AD und der Dynamik der transkriptionellen (Ko-)-Aktivierung nachzuweisen. Im Detail wurden die folgenden Schlussfolgerungen gezogen: (1) Eine verminderte Bindungszeit des TF verminderte die Transkription, sogar für gleichbleibende Besetzung der Bindestellen. (2) Multivalente Interaktionen der AD stabilisierten die Chromatinbindung von schwach gebundenen TF und führten zu einer

Fraktion von indirekt gebundenen Molekülen. (3) AD mit starken multivalenten Interaktionen aktivierten schneller und stärker. (4) Die Bildung von makroskopischen TF-Kompartimenten durch Phasentrennung führte nicht zu einer Verstärkung der Transkription und hatte unter gewissen Bedingungen sogar einen inhibitorischen Effekt. (5) Histon-Acetylierung an Lysinrest 27(H3K27ac) und die Bindung von BRD4, welches H3K27ac bindet, wurden durch indirekt und transient gebundene Aktivatoren unter Bedingungen induziert, die nicht ausreichend für die RNA-Produktion waren. (6) H3K27ac und BRD4 waren nicht strikt notwendig für die Transkription, aber hatten einen verstärkenden Effekt. Basierend auf diesen Erkenntnissen eröffnet die vorliegende Arbeit eine integrative Sichtweise auf die Aktivität von TF, in der mehrere voneinander abhängige Eigenschaften der DBD und AD die Transkription verstärken. Diese umfassen lange Bindungszeiten, hohen Besetzungsgrad der Bindestellen, Stabilisierung von TF-DNA-Komplexen durch multivalente Interationen und Interaktionen mit Ko-Aktivatoren, aber nicht die Bildung von makroskopischen Kompartimenten durch Phasentrennung. Diese Befunde geben einen Einblick in die TF-Eigenschaften, die für die Fähigkeit zur Transkriptionsaktivierung verantwortlich sind, und erlauben es starke synthetische TF zu designen.

List of publications

In the course of this thesis I contributed to the following publications:

1. Trojanowski, J., Rademacher, A., Erdel, F. & Rippe, K., Light-Induced Transcription Activation for Time-Lapse Microscopy Experiments in Living Cells. in *Imaging Gene Expression: Methods and Protocols* (ed Shav-Tal, Y.) 251–270 (Springer US, New York, NY, 2019). doi:*10.1007978-1-4939-9674-2_17*.
2. Rademacher, A., Erdel, F., Trojanowski, J. & Rippe, K., Tracing Reversible Light-Induced Chromatin Binding with Near-infrared Fluorescent Proteins. in *Photoswitching Proteins: Methods and Protocols* (ed Niopek, D.) 171–188 (Springer US, New York, NY, 2020). doi:*10.1007/978-1-0716-0755-8_12*.
3. Frank, L., Weinmann, R., Erdel, F., Trojanowski, J. & Rippe, K., Transcriptional activation of heterochromatin by recruitment of dCas9 activators. in *Enhancers and Promoters: Methods and Protocols* (eds Borggrefe, T. & Daniele Giaimo, B.)(Springer US, New York, NY, 2021), in press.
4. Trojanowski, J., Frank, L., Rademacher, A., Grigaitis, P. & Rippe, K., Transcription activation is enhanced by multivalent interactions independent of liquid-liquid phase separation. *bioRxiv* 2021.01.27.428421. doi:*10.1101/2021.01.27.428421*.

Abbreviations

(E)GFP	(enhanced) green fluorescent protein.
a.u.	arbitrary units.
AD	activation domain.
BLInCR	blue-light induced chromatin recruitment toolbox.
bp	base pairs.
BRD4	bromodomain containing protein 4.
CI	confidence interval.
CIBN	N-terminal region of CIB1.
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats.
DBD	DNA binding domain.
dCas9	catalytically inactive (dead) CRISPR associated protein 9.
ECDF	empirical cumulative distribution function.
FCS	fetal calf serum or fluorescence correlation spectroscopy.
GTFs	general transcription factors.
H3K27ac	histone H3 acetylation at lysine 27.
HAT	histone acetyl transferase.
HDAC	histone deacetylase.
IDR	intrinsically disordered region.
n.s.	not significant.
NLS	nuclear localization signal.

NSQFRAP	R-package Nuclear Spot Quantification for FRAP.
NSSQ	R-package Nuclear Segmentation and Spot Quantification.
PAM	protospacer adjacent motif.
PBS	phosphate buffered saline.
PFA	para-formaldehyde.
PHR	photolyase homology region (light-responsive domain of CRY2).
qRT-PCR	quantitative real-time polymerase chain reaction.
s.d.	standard deviation.
s.e.m.	standard error of the mean.
sgRNA	single guide RNA.
sm-RNA FISH	single molecule RNA fluorescence in situ hybridization.
SRRF	super-resolution radial fluctuations.
tdMCP	tandem MS2 coat protein.
tdTomato	tandem Tomato (red fluorescent protein).
TF	transcription factor.

List of Figures

Contents

Chapter 1

Introduction

1.1 Transcriptional induction is a dynamic multistep process

Transcription is induced by the binding of transcription factors (TFs). Thus, the temporal control of this process is of high interest. TF binding is governed by a large number of transient interactions with a high turnover [1]. It regulates different cellular gene expression programs [2–4] and can contribute to memory of past activation events [5, 6]. TFs exert a powerful control over cellular processes and are able to define cell types, as exemplified by the Yamanaka factors (Oct4, Sox2, Klf4, c-Myc) [7] that are sufficient to reprogram somatic cells into induced pluripotent stem (iPS) cells. They are also involved in malignant transformations into cancer cells [8]. Synthetic TFs hold great promises for therapeutic cellular reprograming [9] and epigenetic drugs that interfere with the transcriptional activation process are already applied clinically for cancer treatment [10]. Yet, our understanding of how distinct TF features quantitatively control the transcription process remains incomplete.

Binding of TFs to their target genes starts a multistep activation process. The core steps of RNA production can be divided into initiation, elongation and termination (reviewed in [11]). It is highly regulated on multiple levels that comprise the RNA polymerase and the local chromatin. Some of the relevant regulatory processes at the gene promoter are depicted in **Figure 1.1**. These partly sequential, partly parallel steps happen on different time scales.

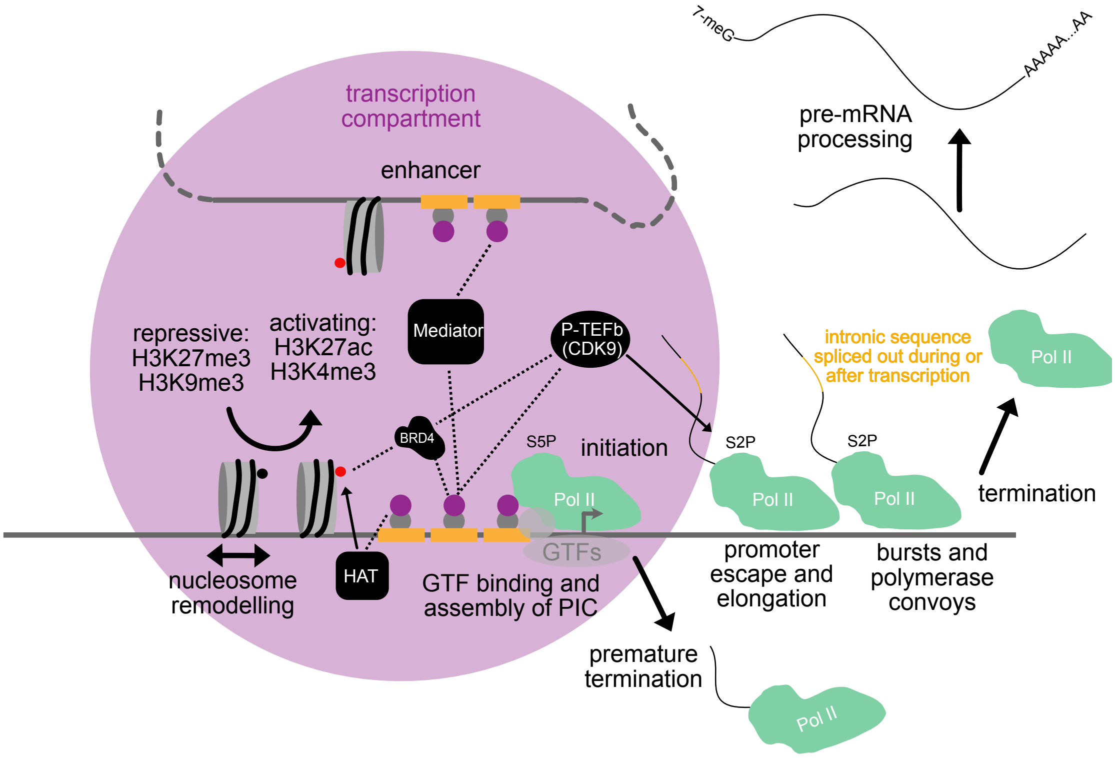


Figure 1.1: Overview of molecular events during transcriptional induction. The induction process is triggered by TFs composed of a DNA binding domain (DBD, grey) and an activation domain (AD) binding to their target sites (yellow). Recruitment of general transcription factors (GTFs, light grey) and RNA polymerase II (Pol II, green) leads to the formation of the pre-initiation complex (PIC). To start productive transcription PolII must be phosporylated first on Serine 5, then on Serine 2 of the heptameric repeats of its C-terminal domain (CTD). Without these modifications premature termination can occur. Once the polymerase escapes the promoter, elongation of the nascent transcript starts. Transcription can occur in bursts and polymerases initiated in fast succession can form polymerase convoys. Termination includes the dissociation of polymerase and nascent RNA from the DNA template. The pre-mRNA is processed by splicing, capping and polyadenylation to yield mRNA that can be exported from the nucleus. The activation process also leads to changes of the local chromatin: nucleosomes are shifted or evicted, repressive nucleosome modifications are replaced by activating marks, and DNA methylation may be removed. Co-activator molecules (black) help the activation process. Histone acetyl transferases (HATs) set acetylation marks, which can be read by bromodomain proteins, e.g. BRD4. (Continued on next page.)

Consequently, certain steps like stable chromatin modifications could provide long-term memory of activation and sustain the transcriptional response. Others have a high turnover, which could have a regulatory effect. The induction process itself is rapid leading to the observation of the first nascent RNA already after 2 min [12]. After successful initiation RNA polymerase traverses the gene body with a speed that has been estimated between 2 and 6 kb/min in human cells [13, 14]. For a gene with a length of 3 kb, like the reporter gene studied in this thesis, this would mean an elongation time of 0.5 to 1.5 min.

However, the time to full activation can be much longer on the range of an hour or more [12]. This indicates that the rate of gene induction is not determined by the core processes of RNA PolII engagement and RNA production, but by transitions of the promoter to a molecular state that allows RNA PolII initiation and release into elongation with a high rate. Termination releases the fully transcribed pre-mRNA from the DNA template and further processing including 5' capping and 3' poly-adenylation (reviewed in [15]). Splicing of the nascent transcript can occur already during transcription, but is fast requiring 5 to 10 min [13]. However, if the supply of splicing factors is low, splicing can become rate limiting for the release of nascent RNA [16].

The steps that directly lead to polymerases transcribing the gene body happen fast. These comprise the assembly of the pre-initiation complex (PIC), initiation of transcription and escape of the polymerase from the promoter. RNA PolII can stop its elongation after few bases, a process termed promoter proximal pausing that affects a significant proportion of genes [17]. High turnover of paused polymerases showed that there can be abortive transcription by premature termination [18]. In this way many of the activating process can already be ongoing without RNA production at an apparently silent promoter. Promoter escape depends on modifications of the C-terminal domain of RNA PolII, which consists of heptapeptide repeats with Serines at positions 2 and 5. Phosphorylation at Serine 5 allows initiation and at Serine 2 elongation (reviwed in [19]).

Serine 5 phosphorylation is set by the GTF TFIIH (containing CDK7) [20] after PIC assembly, whereas Serine 2 phosphorylation requires the Positive Transcription Elon-

---

Figure 1.1: (Continued) The Positive Transcription Elongation Factor b (P-TEFb) is recruited by interactions with BRD4 or directly with TFs and phosporylates PolII allowing promoter escape. Distal elements like enhancers can contact the promoter via the Mediator complex. Clusters of activator molecules, gene promoters and highly active enhancers have been observed to cluster together creating presumable transcription compartments.

gation Factor b (P-TEFb, containing CDK9) [21]. PIC (dis-)assembly is an example of a process that contains both steps with high turnover and slow steps that can allow memory to form. The GTFs TFIID and TFIIA have long residence times (several minutes), whereas TFIIB has a high turnover (1.5 s) until RNA PolII and TFIIF bind [1]. Only TFIID remains promoter bound after promoter escape of PolII [22]. One of its key components the TATA-box binding protein (TBP) remains stably bound to chromatin throughout mitosis creating a physical memory termed mitotic bookmarking [23].

Chromatin represents an additional layer for transcriptional regulation and can carry a physical memory of activation events in the form of remodelled (i.e. shifted or evicted) nucleosomes or of covalent post-translational modifications. Especially histones are extensively modified in order to epigenetically mark genomic loci for transcription or other biological processes. Important activating modifications are acetylation of Lysine 27 of histone H3 (H3K27ac) and (tri-)methylation of Lysine 4 (H3K4me3) [24], whereas methylation of Lysine 27 (H3K27me3) or Lysine 9 (H3K9me3) have a repressive effect (reviewed in [25]). Chromatin modifications have different life times ranging from short times for histone acetylation and H3K9me3 to several days and multiple cell generations for H3K27me3 and DNA methylation [26]. While these times were measured by artificially recruiting chromatin modifiers, the speed of the turnover can also be estimated using inhibitors. This revealed that acetylation is a relatively short-lived post-translational modification (PTM) with life times of few minutes [27]. The high turnover is due to continuous acetylation and deacetylation by histone acetyl transferases (HATs) and histone deacetylases (HDACs). The different persistence times of chromatin marks indicate that promoter states could persist for very different times depending on their modification profiles.

The setting of some chromatin marks also happens fast: The histone acetyl transferase p300 is recruited to the promoter early (within 3 minutes) and transiently and unbinds after auto-acetylating [28]. Histone acetylation enhances transcription directly by decreasing histone-DNA binding and loosening chromatin compaction [29], but the modified chromatin state at the gene promoter or enhancer also allows reader proteins with additional regulatory functions to bind. The co-activator protein BRD4 has two bromodomains which bind to acetylated histones [30]. It recruits P-TEFb to the promoter which initiates transcription [31] and it can also directly phosphorylate RNA PolII on Serine 2 [32]. Some TFs, like the HIV activator tat, can also take over BRD4's role and recruit P-TEFb directly [31].

From a quantitative point of view active genes are transcribed by between one and four polymerases based on electron microscopy of nuclear spreads [33] or around 17 polymerases based on calibrated super-resolution imaging [34]. Highly transcribed loci tend to co-localize in the nucleus giving rise to the concept of transcription factories [35]. Similarly, highly active enhancers with strong accumulation of master transcription factors and co-activators over long stretches of DNA form clusters and have been defined as super-enhancers [36]. They play an outstanding role in development and control cell identity, while their dis-regulation is observed in cancer [37]. Due to their reliance on co-activators like BRD4 they can be targeted by inhibitors and display an increased sensitivity to these perturbations [38]. The clustering of factors that regulate transcription has emerged as an important property of highly active genes. The most striking manifestation of this is the ability of transcription factors [39], co-factors [40] and even RNA PolII itself [41] to form phase-separated compartments in vitro and potentially also in vivo. It thereby could represent a mechanism by which nuclear subcompartments with high transcriptional activity are formed. This property will be introduced in more detail in section 1.2.1.

The promoter does not always react to stimulation by TF binding with the same response. The response of at least some genes depends on their previous activation history [5, 6], while in other cases transcription has been found to be without memory [42]. Moreover, there is a large heterogeneity in the transcriptional response of single cells [43] that has led to the discovery of transcriptional bursting, a phenomenon that was also observed in the transcriptional activity of a single gene over time [44]. Transcriptional bursting means that the promoter can reside in at least two states: an on-state from which RNA is produced and an off-state. The transitions between these states and their kinetics determine the properties of the bursting process, namely the bursting frequency, duration and the number of polymerases per burst. The average duration of the on-state was found to be around 5 min in the original study in *Dictyostelium* [44], whereas for mammalian genes on-times between 20 and 40 min were found [45]. Within these on-times shorter intervals of increased activity have been observed for the HIV promoter which have been attributed to RNA polymerase convoys, several polymerase molecules that have been initiated in fast succession transversing the gene body together [46]. Several observations have suggested that there must be promoter states in addition to a single on- and off-state. Among these is the detection of a promoter state that is refractory to re-stimulation [47, 48] and the modelling of promoter cycling that required at least five states [49]. The promoter states underlying the transcriptional bursting cycle have been attributed to different molecular mechanisms including binding and unbinding of a limited amount of TF [50, 51], episodic histone acetylation [52], binding of the Mediator co-activator [46] and transient binding of GTFs like the TATA box binding protein (TBP) [46]. In addition to the intrinsic noise produced by transcriptional bursting (reviewed in [53]), gene expression can also be affected by extrinsic noise sources like cell volume or gene copy number and the related compensatory global regulation mechanisms [54] . These multiple highly dynamic processes involved in all levels of transcriptional regulation ranging from chromatin to polymerase activity highlight the need for a quantitative and time-resolved analysis of gene induction at the single cell level. Moreover, they show that transcriptional induction should be considered as a complex reaction network that can be described by multiple distinct states of the gene promoter.

1.2 Transcription factors contain a DNA-binding domain and an activation domain

Transcription factors (TFs) control the first step of gene expression, mediate responses to signalling pathways and can determine cellular identity. They typically consist of a DNA binding domain (DBD), responsible for target gene binding, and an activation domain (AD) that mediates transcriptional induction (**Figure 1.2 A**). This is the typical blueprint for transcriptional activators, but there are two classes of TFs that modify this structure or function: Nuclear hormone receptors additionally contain a ligand binding domain that has a regulatory function [55]. Pioneer factors have a DBD that can bind to compact heterochromatin and their main role is to establish an accessible chromatin state, as is exemplified by FoxA whose winged helix domain can replace linker histones [56]. They often have a role in development and re-organize the chromatin state during cell differentiation (reviewed in [57]). TFs can also be repressors of transcription. In contrast to the disordered activation domains, repressors have an effector domain, which can have a ordered structure and be conserved like the KRAB domain for Krueppel-like TFs. KRAB interacts with an adaptor molecule and recruits repressive chromatin modifiers like HDACs and histone methyl transerases [58]. Prokaryotic repressors like LacI mediate repression by DNA looping. LacI binds to multiple binding site in the operator region and forms a tetramer that induces a repressive DNA loop [59]. Synthetic TFs can be constructed from modules that resemble the natural architecture of DBD and AD and hold great promises for cellular manipulations, for instance for cellular reprogramming [60]. While the binding properties of the DBD and, recently, the interactions and self-interactions of the AD have been characterized, their interdependence and combined effect on transcriptional activation still remain unclear.

1.2.1 DNA Binding Domain: structures and binding properties

The DNA binding domain of a TF folds into a defined structure and the conservation of their sequences has allowed to group TFs into families. In the human genome 2000-3000 genes are estimated to encode TFs of which 1391 genes have been compiled in a curated collection [61]. In this collection 80% of human DBDs belonged to one of three structural families: C2H2 zinc finger, homeodomain and helix-loop-helix [61]. The DBD structures have been reviewed and assigned to 54 families in ref. [62]: The C2H2 zinc finger domain contains two β-sheets and an α-helix and coordinates a zinc ion by two cysteine and two histidine residues. Homeodomain TFs and also bacterial repressors like LacI and TetR have a small helix-turn-helix motif [63]. Helix-loop-helix TFs have two long α-helices, one of which is responsible for dimerization and one for binding the DNA major groove.

TFs bind specifically to their target sequence which typically has a length of around 10 bp and a range from 5 to more than 30 bp in eukaryotic cells [64]. Target sites are typical in the promoter proximal regions or at enhancers. Only a fraction of theoretically amenable binding sites are usually bound. This is thought to be regulated by binding site accessibility and chromatin modifications [65].

Binding of a TF to DNA is mediated by interactions of its amino acid residues with the DNA strand. These interactions can be to a large extent non-specific: electrostatic interactions with the negatively charged phosphate backbone including release of counter ions and van-der-Waals interactions [66]. Specific interactions comprise van-der-Waals interactions and hydrogen bonds of amino acids with DNA bases [67].

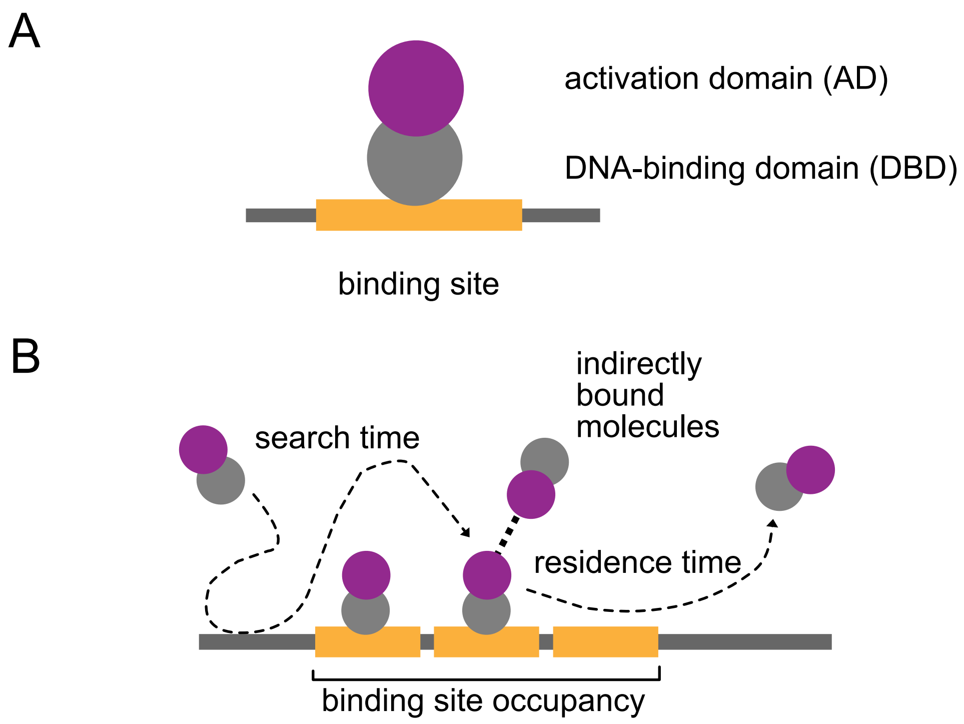


Figure 1.2: The TF binding properties. **(A)** A TF contains a DNA-binding domain (DBD) and an activation domain (AD) and binds to specific binding sites. **B** The binding of a TF is characterized by several parameters, comprising the search time (on-rate k_{on}), residence time (off-rate k_{off}) and the binding site occupancy (depending on TF concentration and dissociation constant K_D).

Specific binding may primarily not depend on stabilizing interactions (e.g. hydrogen bonds), which have a low binding energy, but on avoiding mismatch pairing of amino acid and DNA bases, which is energetically strongly disfavored [66]. The individual interactions contribute to the overall free binding energy ΔG via the binding reaction enthalpy H or via an increase of entropy S:

$$\Delta G = \Delta H - T \cdot \Delta S \tag{1.1}$$

The difference in binding energy for binding to target sites and non-specific background sites determines the specificity of TF recruitment and consequently gene activation. However, the binding to target sites cannot be indefinitely tight, because TFs need to dissociate on a time-scale that allows for other biological processes like DNA replication to occur [66]. Additionally, tolerance to mismatch binding has been suggested as a desirable trait that makes it easier for binding sites to evolve and rewire the regulation of TF target genes [68]. These constraints for the binding affinities indicate that additional

TF binding properties could be responsible for specific gene activation.

The different binding properties are visualized in **Figure 1.2 B**:

- The binding site occupancy that depends on the affinity described by the dissociation constant K_D.
- The association rate that depends on the search time the TF spends diffusing through the nucleus and screening non-specific sites before it finds its target site.
- The residence time that an individual TF molecule remains bound to its target site determined by the off-rate.
- The local concentration of TF molecules at the target site that is increased by indirectly bound TFs via protein-protein interactions.

The binding site occupancy θ be expressed as a saturation function of the TF concentration:

$$\theta = \frac{TF}{TF + K_D} \tag{1.2}$$

It can also be determined from the energies ϵ_i of the molecular micro-states (individual molecules bound at distinct sites) using partition functions Z that predict the probabilities for the different molecular micro-states.

$$Z = \sum_i e^{-\frac{\epsilon_i}{k_B T}} \tag{1.3}$$

The ratio of the partition function describing the desired micro-state and the sum of partition functions of all possible states yields the probability to find the system in this favorable state. This has been used to construct theoretical models to predict gene expression in prokaryotic cells from the DNA binding energies and concentrations of regulatory factors [69]. In this framework the probability to find a binding site occupied by a protein P is given by the number of non-specific binding sites N_{NS}, protein number P and the difference of binding energies at non-specific sites and the site of interest $\Delta\epsilon$ [69]:

$$p_{bound} = \frac{1}{1 + \frac{N_{NS}}{P} e^{\Delta\epsilon / k_B T}} \tag{1.4}$$

This approach is then extended to predict the probability of polymerase at the promoter site (as a measure of transcription rate) from the binding states of polymerase and an activator taking into account the binding energy for a direct interaction of activator and polymerase. The model can be further extended by incorporating binding energies between multiple activators leading to cooperative binding at clustered binding sites. Such thermodynamic models have also been used to predict transcription of genes in animal cells [70]. However, these models are based on the assumption that the equilibrium binding state alone regulates transcription ignoring the kinetic binding properties. Two kinetic binding parameters could mediate additional transcriptional regulation: the association rate constant (on-rate, k_{on}) and the dissociation rate constant (off-rate, k_{off}). Note that the thermodynamic and kinetic binding properties are linked by the dissociation constant: $K_D = \frac{k_{off}}{k_{on}}$, but that there can be at the same time high affinity binding and high turnover of individual bound molecules, if both k_{on} and k_{off} are high. The on-rate corresponds to the search time a TF spends in diffusing through the nucleus, binding to and unbinding from non-specific sites until it finally binds its target site (reviewed in [71]). The search time of an individual TF molecule has been estimated to be as long as 35 days for the bacterial repressor TetR [72], but at normal expression levels of the TF it will require only few minutes for the binding site occupancy to saturate. 1D-3D diffusion has been suggested as a mechanism that leads to fast target search. It comprises TF sliding along the 1D chromatin fiber interspersed with intervals of 3D diffusion [73]. Indeed, protein motifs like the nuclear localization signal have been shown to mediate protein sliding along DNA [74]. While TF search times have been invoked for the regulation of transcription burst firing rates [75], regulation of differential gene expression by different on-rates faces some challenges: The on-rate cannot be reduced indefinitely without slowing down the induction process, relies on the same search mechanism for all TF-target site pairs and depends on the (potentially heterogeneous) TF expression level.

The other rate constant that defines the binding kinetics is the off-rate k_{off}. Its inverse, $\frac{1}{k_off}$, gives a characteristic time that an individual TF molecules remains bound called residence time. The residence time will display marked differences for target site and non-specific binding. An important role for this quantity in transcriptional regulation is emerging, as it is regulated by signalling pathways and shows correlation with transcriptional activity. For the p53 TF it was shown that acetylation in response to DNA damage increases its residence time and leads to higher transcription of its target genes [76]. Modulation of residence times of synthetic repressor molecules (based on TALEs) revealed a correlation of repression with residence time [77]. Interestingly, the repression was mediated by a reduced transcriptional burst frequency indicating competition of repressors with activators. It is unclear if such a mechanism would also apply to the residence time of activator TFs at genes that are not mainly regulated by repressors.

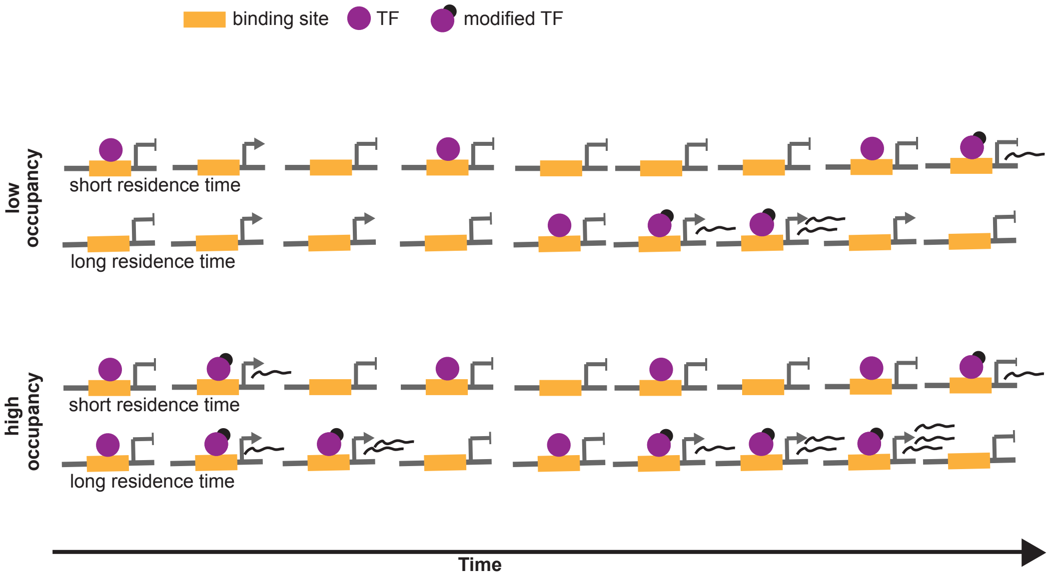


Figure 1.3: Different scenarios of TFs with high or low binding site occupancy and long or short residence time regulating a target gene. If the TF has to be covalently modified after binding its target gene and before inducing productive transcription, its residence time can regulate the transcriptional output independent of the binding site occupancy. At low TF concentration binding sites are mostly unbound for both short and long residence times (upper panel). At sufficiently high concentrations binding sites can be occupied for both cases (lower panel), but in the short residence time case individual TF molecules dissociate before being able to induce transcription.

Importantly, residence time and affinity are coupled, because the dissociation constant K_D will increase with an increasing off-rate if the on-rate remains the same. Studies of TF residence time, so far, did not distinguish if the residence time correlates with transcription due to its effect on binding affinity or if it regulates transcription on its own. While a genome-wide study found that binding turnover is better (anti-)correlated with transcription levels than binding site occupancy [78], the used competition ChIP method has neither the temporal resolution to measure second to minute time scale TF binding, nor is ChIP generally a good method to quantitatively measure binding site occupancy [79]. Nevertheless, it should be possible to study the effect of residence time independent of binding site occupancy. The binding rate depends on the TF concentration, so that it is possible to saturate binding site occupancy at high TF expression levels and still have a high turnover of bound molecules (**Figure 1.3, lower panel**). There could be mechanisms that regulate transcription by the residence time rather than the binding site occupancy. An intuitive example would be the modification of bound TFs that is required to start transcription. In such a scenario, there could be frequent short TF binding events for a short residence time TF leading to a high average binding site occupancy. However, the fast dissociation would prevent modified TFs from activating transcription, because they would be immediatly be replaced by un-modified TFs (**Figure 1.3**). A kinetic model implementing such a multistep activation mechanism will be presented in section 2.4.2. A recent theoretical study has demonstrated another residence time readout mechanism based on nucleosome remodelling in the presence of bound TF [80].

Finally, there can be multiple subpopulations of bound molecules with different binding kinetics. For instance, TF molecules bound indirectly via TF-TF interactions can have a higher binding turnover than directly DNA bound molecules. Some measurement methods, like single particle tracking, can resolve these subpopulations, whereas methods that measure the bulk behavior of the whole molecule population (like Fluorscence recovery after photobleaching) can only distinguish them if they exchange on distinct time scales.

1.2.2 Activation Domain: interactions of intrinsically disordered regions (IDRs)

In contrast to the DBD, the AD does not fold into a conserved structure that would allow ADs to be assigned to protein domain families. Rather, they contain intrinsically disordered regions (IDRs). These regions are much more prominent in ADs than in other classes of protein domains [81]. The AD sequence is usually not conserved and can even arise from a small fraction of random sequences. The activity of some TFs can be attributed to a short nine amino acid (9aa) motif which shows low conservation, divergent and convergent evolution and can randomly arise [82]. Despite the low sequence conservation, ADs can easily be transferred between species and are active in evolutionary distant organisms [83]. Instead of conserved domains they are characterized by general features. IDRs preferentially contain certain amino acids [84], among them glutamine, acidic amino acids (aspartic and glutamic acid) and proline, and ADs have been classified into three classes based on their content of these amino acids [83]. Interestingly, simple repeats of proline or glutamine are sufficient to activate transcription [85]. Acidic ADs contain repeats or related sequences that are responsible for their activation potential [86] and also glutamine rich ADs can be constructed from short segment repeats [87]. Screening of short artificial ADs revealed that hydrophobic amino acids are required for transcriptional activity [88]. These findings have been integrated into a design principle to explain necessary sequence features of acidic ADs. This design principle derived from a mutational screen requires acidic residues flanking a hydrophobic motif to ensure solubility [89, 90]. IDRs bind to partners with weak but specific interactions. Importantly, they can also specifically bind multiple distinct partners [84]. This weak binding with multiple specificity can be rationalized by the presence of multiple “mini-motifs” in the IDR: A machine learning approach found that the activation capacity of IDRs depends on multiple mini-motifs containing aromatic and acidic residues, but no positively charged residues and that these mini-motifs likely engange in multiple weak interactions.[91].

Different mechanisms could be responsible for the induction of transcription by IDRs in the AD:

While the AD can directly bind to RNA polymerase in prokaryotes [92], in eukaryotes direct interactions with GTFs like TBP and TFIIB and co-activators like SAGA have been shown for VP16 [93]. This binding could involve induced folding of the IDR upon target binding: an induced fit of the IDR upon binding to TBP was deduced from the biphasic binding kinetics of c-Myc, VP16 and GAL4 [94]. This was also seen for intrinsically disordered ADs interacting with CBP (CREB binding protein) co-activator [95]. ADs could also differ in their interaction repertoire and consequently only induce a subset of the steps necessary for productive transcription. ADs have been classified into three categories based on their ability to induce different phases of transcription: class I TFs, like SP1, induce only initiation, class IIa, like the HIV activator tat, only elongation and class IIb (e.g. VP16, p53) both initiation and elongation [96].

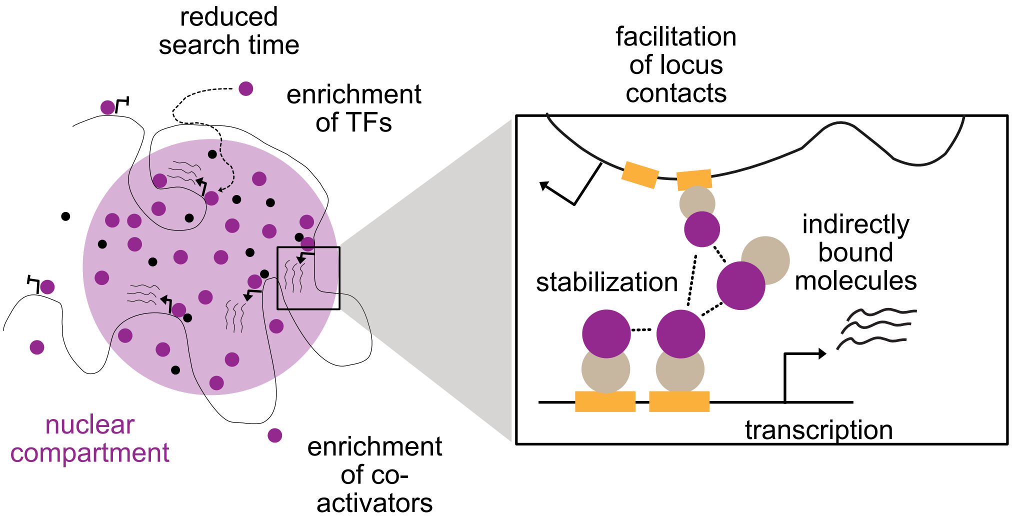


Figure 1.4: Potential effects of multivalent interactions of the AD on transcription by the formation of a phase-separated compartment or by local interactions: stabilization of chromatin binding, local enrichment of TFs and co-activators, facilitation of enhancer-promoter/promoter-promoter contacts.

In addition to direct interactions of ADs with GTFs and co-activators, other mechanism could induce or increase transcription. It has been shown that IDRs alone can be sufficient to localize a TF to its correct target promoters without the DBD, probably due to TFs already present there that interact with the IDR [97]. Many regulatory factors of transcription have the ability to undergo liquid-liquid phase separation [39–41] in vitro. This discovery makes a variety of activation mechanism appear possible that depend on phase-separated transcription compartments (or “transcriptional condensates” [39]).

Liquid-liquid phase separation (LLPS, reviewed in [98]) is characterized by the formation of two phases, one with a very high concentration and one depleted for the constituent molecule. Inside the high concentration compartment molecules retain a high mobility. There is a sharp transition in physico-chemical properties at the compartment boundary, e.g. of density, refractive index or molecular composition. Phase separation occurs abruptly if the concentration of separating molecules exceeds a critical concentration, that is specific for a certain temperature, salt concentration and pH. It may, however, be kinetically delayed if no nucleation sites for the newly arising phase are available. The homotypic interactions that drive phase separation also lead to surface tension that lead (in the abscence of additional constraints) to the formation of spherical droplets which can fuse and thereby reduce their surface-to-volume ratio.

LLPS of regulatory factors of transcription could have several functional consequences. An overview of putative mechanisms how phase-separated compartments could enhance transcription is shown in **Figure 1.4**. It has been reported to reduce target search time of TFs [99], and the cooperativity in factor binding was suggested to increase the sensitivity to perturbation [100]. Moreover, TF phase separation could constitute a positive feedback mechanism by enhanced nucleation at locations with high phosphorylated CTD concentration [101]. There can be enrichment and exclusion of molecules from droplets [102] and specific enrichment or exclusion between different phases [103]. This has been specifically shown for co-activators like mediator or BRD4 that phase separate in vitro and colocalize with other regulators and RNA Pol II [40, 104]. The formation of phase-separated compartments is most likely to occur at loci with clusters of binding sites and has therefore been prominently been described at super-enhancers which provide high binding valency [40]. The enrichment of regulatory factors could also involve RNA Pol II itself. Phase separation of its C-terminal domain (CTD) has been observed in dependence on its length [105]. Regulation by post-translation modifications (PTMs), allows to change the phase separation propensity or the enrichment in a given compartment. For instance, translocation of RNA Pol II from a compartment for transcription to one for RNA processing and splicing has been reported [106].

However, the physical nature of transcriptional condensates in vivo is disputed, as evidence for LLPS has often been found in vitro at expression levels that strongly exceed physiological levels and in vivo experiments were often insufficient to prove true LLPS [107]. LLPS can be distinguished from other aggregation states like gels, solids or polymer-polymer phase separation by its high internal mobility and its potentially unlimited increase of droplet volume with increasing concentration, as well as by droplet fusion and surface tension [98, 108]. A more detailed physico-chemical analysis of TF phase separation in vivo found that TFs assemble in hubs of a limited size rather than in liquid droplets [103]. These observations would also be compatible with TF assembly by surface condensation on DNA [109]. Nevertheless, it remains an interesting question, what effect LLPS would have on transcription if it occured naturally. Some mechanisms that could enhance transcription for liquid-like TF droplets could also be in force for TF hubs or surface condensates.

Understanding the mechanisms that drive condensate formation and potentially regulate transcription can have a direct impact for certain diseases. In some cancer entities fusion proteins that add the ability to undergo phase-separation or aggregation to endogenous TFs play a central role. For instance, oncogenic fusion TFs like EWS-FLI1 change the distribution of chromatin remodellers over the genome, a capacity that depends on their phase separating protein domain [8, 110]. Another protein domain that is frequently fused to oncogenic TFs is FUS [111]. It has the ability to phase separate and FUS mutations in amyotrophic lateral sclerosis (ALS) have been shown to increase the transition of a liquid to a solid phase [112].

1.2.3 Synthetic TFs: modules and applications

Due to the modular nature of DBD and AD synthetic TFs can be constructed by combining domains with the desired target sequence and activator strength [113]. Besides combining endogenous DBDs or ADs there are also synthetic modules that allow for stronger activation or nearly arbitrary sequence specificity. Currently there are three main strategies to construct synthetic DBDs to target a DNA sequence of interest: Zinc finger proteins, TALEs and CRISPR/dCas9. Construction of synthetic zinc finger TFs (reviewed in [114]) starts from the 30 amino acid zinc finger domain that occurs in endogenous TFs. This domain determines the binding specificity to three nucleotide motifs and this specificity can be adjusted by mutations, selection and design [115]. Fusions of several of these domains then allow to construct “multi-fingered” DBDs that specifically target longer DNA sequences [116]. TALEs (transcription activator-like effectors, reviewed in [117]) are factors secreted by plant-pathogenic bacteria to reprogram transcription of their host cells. They contain multiple tandem repeats of 34 amino acids. Each repeat has two variable residues that determine the binding specificity to a single DNA base [118]. This modular structure allows to construct synthetic DBDs for target sequences by fusing multiple repeats with the desired single base specificities. While the first two methods rely on the labor intensive design of synthetic proteins for each target site, the target sequence can be very easily adjusted in the CRISPR/dCas9 system. It is derived from the antiviral defense system of bacteria based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) [119] and CRISPR associated (Cas) proteins. The system consist of a nuclease (Cas9) that forms a complex with two RNA molecules and can be employed for targeted sequence specific cutting of DNA [120]. The two RNAs are the CRISPR RNA (crRNA) and the trans-activating crRNA (tracrRNA) which can be combined into a single guide RNA (sgRNA). The crRNA contains a 20 nt sequence that is complementary to its target site. A protospacer adjacent motif (PAM) is required next to the complementary sequence and has the form NGG for (d)Cas9. Mutation of Cas9 led to a catalytically dead variant, dCas9, which can be used as a DBD to target genomic loci without inducing DNA double strand breaks [121]. The (d)Cas9-sgRNA complex binds its target site rapidly within few second, however, the assembly of sgRNA and (d)Cas9 protein can be reduced if other RNAs are present that compete for (d)Cas9 binding [122]. Moreover, the CRISPR/dCas9 complex binds it target site very tightly (with low nanomolar affinity) and with high specificity [123, 124]. In the guide RNA stem loops (MS2, PP7, boxB) can be integrated and used for the assembly of multiple activators [125] or for multiplexing with different effectors or fluorescent proteins for specific loci [126].

There are also synthetic ADs that have been designed to increase activation capacity. In many applications the relatively strong AD of the viral TF VP16 is used and this domain has been used as a tandem tetramer to create the synthetic VP64 activator [116]. With the development of CRISPR/dCas9 for the activation of endogenous genes a new need for strong and robust activators has arisen and two multicoponent systems were designed that combine multiple ADs. VPR is composed of the ADs from VP64, p65 and Rta and was shown to activate more strongly than the individual components [127]. The synergistic activation mediator system (SAM) relies on the direct fusion of VP64 to dCas9 and recruits p65 and HSF1 via the MS2 coat protein (MCP) to MS2 loops of the dCas9 complex [125].

Introducing TFs into a cell allows to reprogram its transcriptome, as is most impressively demonstrated by the Yamanaka factors (Oct4, Sox2, Klf4, c-Myc) that are sufficient to create induced pluripotent stem cells (iPSCs) from somatic cells [7]. These reprogramming techniques can be employed therapeutically to trans-differentiate cells in-situ or for autologous cell transplantation and to replace damaged cell populations, e.g. in neurodegenerative diseases (reviewed in [128]). Examples of trans-differentiation comprise reprogramming of fibroblasts into muscle cells by the TF MyoD [129] or into neurons by a set of three TFs [130]. While these methods rely on the ectopic enforced expression of endogenous TFs, the possibility to construct synthetic TFs with desired sequence specificity and tunable activation strength opens up ways to more selectively activate subsets of target genes. Specifically, CRISPR/dCas9 based synthetic TFs could constitute an attractive class of easy-to-design biologicals [60].

1.3 Methods to study dynamic TF binding and activation

To study the dynamic aspects of transcriptional activation, time resolved measurements and inducible perturbations with good temporal control are needed. This regards the measurement of the time-variable promoter activity in reporter cell lines, controlled recruitment of effectors to the promoter by optogenetics and the measurement of TF binding kinetics by fluorescence recovery after photobleaching (FRAP) (**Figure 1.5**). Some part of the heterogeneous dynamics of single cells, e.g. due to bursting processes, would be lost in bulk measurements, so that single-cell measurements that can be provided by microscopy are advantageous. All measurements performed with these tools require quantitative measurements and modelling to understand the partly unobservable underlying mechanisms. Specifically, modelling of transcription time course data that could be driven by stochastic events, reaction-diffusion models for FRAP measurements and a reaction model to explain the binding kinetics of the optogenetic TFs.

1.3.1 Optogenetic TF recruitment with high temporal and spatial precision

The perturbation of cellular processes by light offers great opportunities to study dynamic living systems with high spatial and temporal precision and few collateral effects.

There are three main classes of optogenetic switches that are used for experimental manipulations: LOV domains, Cryptochromes and Phytochromes. Of these the CRY2 CIB1 system is the most widely used according to the optobase data base [131].

Light-oxygen-voltage (LOV) domains contain flavin mononucleotide (FMN) and confer light-sensitivity to multiple natural light responsive proteins in bacteria, plants and fungi [132]. Examples comprise EL222, a bacterial TF that dimerizes upon illumination and binds it target sites [133]. The inducible conformational changes have also been exploited for protein engineering e.g. to convert a anti-CRISPR protein to an inactive conformation by light and thereby release dCas9 to target its binding sites [134] or to induce nuclear translocation by exposing a nuclear localization signal (NLS) [135].

The Cryptochrome CRY2 controls the light-dependent initiation of flowering in *A.thaliana* [136]. It binds to CIB1 upon exposure to blue light (**Figure 1.5 A**). CRY2 contains FAD (Flavin adenosin dinucleotide) that is reduced to $FADH^{+}$ upon illumination which induces a conformational switch. Light-activated CRY2 has been determined by structural analysis to form tetramers [137]. The Photolyase Homology Region (PHR) of CRY2 and N-terminal region of CIB1 (CIBN) are sufficient for the light-dependent interaction and these smaller domains are easier to integrate in engineered protein constructs [138].

Recruitment of PHR to CIBN is fast and, reversible. Previous work in our group developed a toolbox called BLInCR (Blue light induced Chromatin Recruitment) for recruiting various effectors to chromatin targets [12]. The oligomerization of light-activated CRY2/PHR is responsible for its second light-induced activity, the formation of liquid condensates [139]. These “optodroplets” (**Figure 1.5 B**) have been determined to have liquid like properties and were used to induce protein phase separation in living cells by light exposure [140]. Optodroplets preferentially form at locations with low chromatin density [141] so that PHR-fusion proteins can be recruited either to chromatin bound CIBN-localizer proteins or droplets in the inter-chromatin region.

1.3.2 Recording transcriptional induction dynamics in a reporter cell line

Gene transcription can be imaged by single molecule RNA fluorescent in situ hybridization (smRNA FISH) in fixed cells, whereas in live cells reporter genes can be tagged, for instance by RNA stem loops that are bound by fluorescently labeled proteins, to visualize nascent transcripts (reviewed in [142]). Both techniques have single-cell resolution, but smRNA FISH only offers an end-point measurment, so that no connected single-cell time courses can be acquired. Tagging with MS2/PP7-loops can also be achieved for endogenous genes, as has for instance been demonstrated for a PP7 labeled target gene of the hormone regulated glucocorticoid receptor [143]. However, the endogenous TF binding is typically not observable and the dynamics of the transcriptional response is convolved with the (slower) dynamics of the signal transduction pathway, which makes it less useful for understanding the early processes of transcriptional induction.

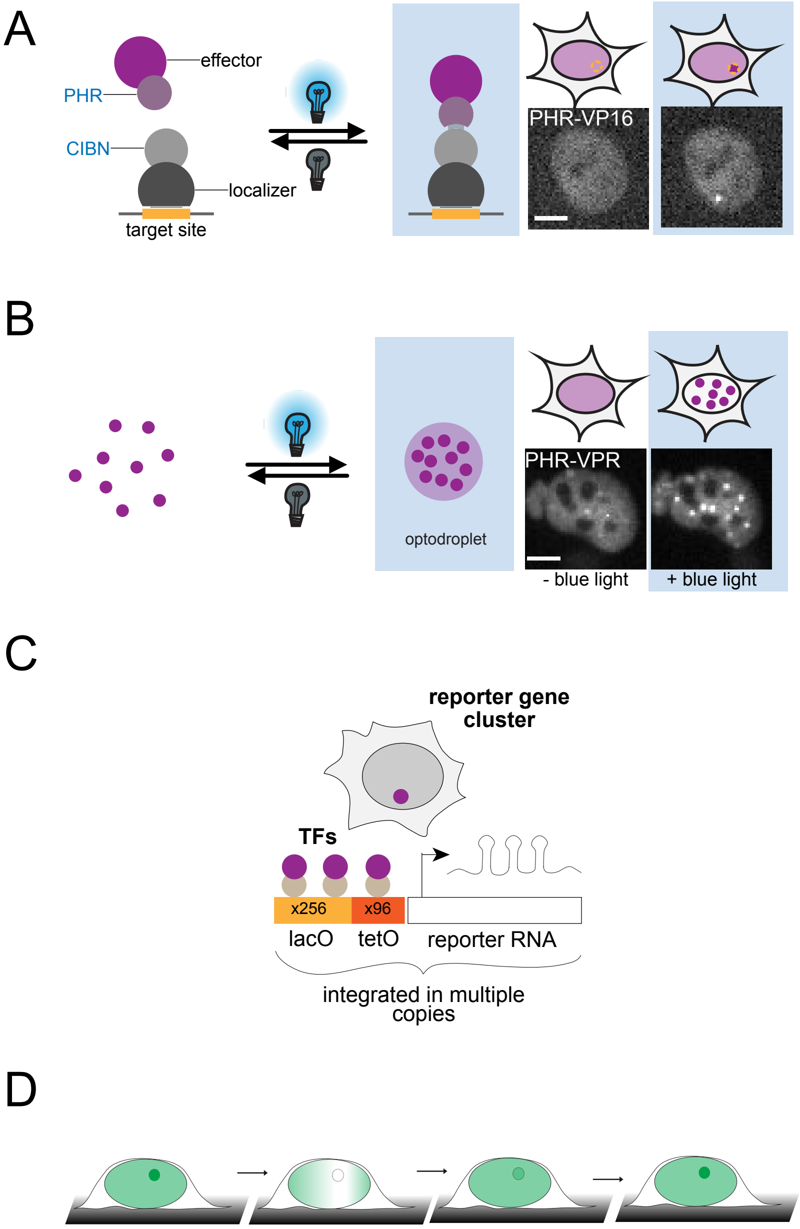


Figure 1.5: Tools to study TF binding, self-interactions, phase separation and transcription activation dynamics. (Continued on next page)

Using a custom super-resolution microscopy approach TF binding, RNA Pol II and co-factor recruitment and transcription of a single endogenous gene have been observed [34]. However, such experiments are technically challenging due to the instrumentation requirements, the low signal strength and difficulties to distinguish TF or RNA signal of the gene of interest from signal of near-by genes. Multi copy reporter gene arrays represent a more accessible system to study transcriptional regulation with conventional microscopy techniques. The multiple copies of TF binding sites and reporter gene bodies integrated at a single genomic locus provide sufficient signal amplification to see TF binding and RNA production on a fluorescence widefield microscope. A widely used reporter cell line U2OS 2-6-3 has been constructed by integrating a reporter construct in multiple copies in the subtelomeric region of chromosome 1 [144]. Each reporter gene unit contains a gene body encoding ECFP-SKL targeted to peroxisomes, 24 MS2 repeats and a globin intron-exon module. The minimal CMV promoter is under the control of 256 *lacO* and 96 *tetO* sites that allow to inducibly recruit effector proteins as LacI or (r)TetR fusions (**Figure 1.5 C**).

This reporter gene cluster has been shown to be in a silent heterochromatic state marked by H3K9me2/3 and heterochromatin protein 1 (HP1) before induction [144] and then to acquire activation marks (BRD4 and histone acetyl transferase GCN5) and to decondense, when transcription is induced [145]. Light-induced TF recruitment in this cell line by the BLInCR system has been used to acquire single-cell transcription induction time courses with high time resolution [12]. This revealed a fast on-set of RNA production, biphasic activation dynamics indicating a multistep or feedback mechanism and potentially sensitization of the genes to re-stimulation, but was limited to small cell numbers.

Figure 1.5: (Continued) **(A-B)** Two light-induced activities of the PHR-CIBN system. Images on the right show PHR-EGFP-AD fusions; scale bar 10 µm. **(A)** The PHR domain binds to CIBN upon blue light illumination. **(B)** PHR is also able to form homo-oligomers and phase separate into optodroplets, a property that depends on its fusion partner. **(C)** The reporter cell line contains multiple copies of a transcriptional reporter integrated as a tandem array. In each unit repeats of *lacO* and *tetO* binding sites control a minimal CMV promoter. The nascent RNA contains MS2 stem loops that can be bound by fluorescently labelled tdMCP molecules. **(D)** Fluorescence recovery after photobleaching (FRAP) allows to measure residence times at a cluster of TF binding sites. Bleaching of a cylindrical region of the nucleus is followed by the return of fluorescence intensity to the bleach region and gene cluster spot.

The multi-copy architecture of the reporter does not fully resemble the situation of isolated genes that will become activated by stimulus triggered TF binding. However, in many instances genes are organized in co-regulated clusters, for instance the major histocompatibility complex (MHC) cluster [146] or the histone gene cluster [147]. The activation of the reporter gene cluster could, thus, resemble the processes in a transcription factory [33] or at super-enhancers [36]. Since phase-separation of TFs and other activators has been proposed for super-enhancers [40], a large reporter gene cluster with multiple TF binding sites could represent a good model system. Moreover, there is a large overlap of chromatin marks that define enhancers and promoters [148, 149], so that cross-talk of activated reporter genes could also give general insights into cross-activation processes between distal genomic loci.

1.3.3 Measuring binding kinetics by fluorescence recovery after photobleaching (FRAP)

Fluorescence recovery after photobleaching (FRAP) is a technique that has been widely applied to determine diffusion and binding properties of biomolecules [150]. It is a fluorescence microscopy method and is based on the photobleaching of fluorescent molecules in a defined area by a high intensity laser beam. Recording the redistribution of fluorescence intensity, as bleached and unbleached molecules are exchanged, allows to understand the dynamics of the underlying molecular processes. Importantly, the bleaching only changes the fluorescent labeling, while the system remains in thermodynamic equilibrium. FRAP measures the steady state turnover of molecules that diffuse in and out of a region, bind to and unbind from their target sites. The intensity in the bleach region is measured over time and normalized to yield the recovery curve which can be fitted by a model to extract binding and diffusion parameters. The parameters comprise the (effective) diffusion coefficient D_{eff}, the unbinding rate constant k_{off} and the immobile fraction of molecules that do not dissociate from their binding sites during the observation period. The recovery process can be described by a reaction-diffusion model and there are limiting cases in which there are analytical solutions to simplified versions of this model [151]: If the molecule does not exhibit binding a diffusion-only model can be used. This model can also be used if the binding events are much shorter than the diffusion time, giving rise to an effective diffusion coefficient. In cases, where diffusion is much faster than binding, a binding-dominant model can be used. Fitting the data to a simplified model if the pre-conditions are not met can, however, result in the estimation of wrong parameter values. The recovery model must be adjusted if the binding sites are not homogenously distributed over the whole compartment but reside in a cluster at the bleach spot center. In this case the geometry of the bleach area, binding site spot and nucleus must be explicitly described in the model, typically in the form of concentric cylindrical volumes [152]. For the full reaction-diffusion model no analytical solutions exist and they must be numerically simulated.

The estimation of parameters from recovery curves is strongly dependent on the model used for data fitting. Ignoring either diffusion or binding, or using over-simplified initial conditions can lead to strong variations in the estimated parameters, but studies which found very different TF residence times could be brought into agreement when accounting for all of these factors [153]. Technical biases can be introduced if there is reversible photoswitching, or by bulk photobleaching. In reversible photoswitching the fluorophores transiently enter a dark state and will be detected as apparently bleached. The relatively fast return will be erroneously be detected as a fraction of molecules with fast diffusion or binding exchange. The impact of reversible photobleaching can be estimated and corrected by performing whole-cell bleaches or fusing the fluorophore to an immobile protein like a chromatin incorporated histone [154]. Bulk photobleaching refers to the reduction of the total amount of fluorescent molecules in the bleached compartment and can prevent that the original fluorescence intensity in the bleach region is reached at the end of the time course. This leads to an apparent immobile fraction. The effect, however, is mostly corrected by normalizing the bleach region intensity to the intensity of the whole compartment, e.g. the whole nucleus.

Overall, FRAP allows to measure informative parameters describing the diffusion and binding of molecules, but it requires careful analysis that accounts for all the processes occuring during recovery to extract meaningful fit parameters.

1.4 Scope of the thesis

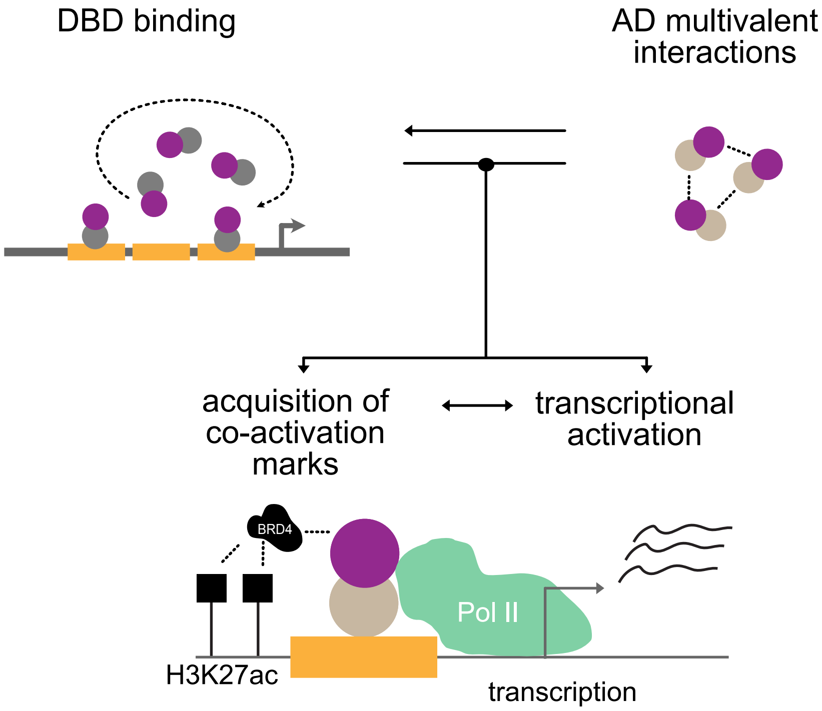


Figure 1.6: The TF properties studied in this thesis, their potential inter-dependencies and their combined regulation of transcription. Both binding site occupancy and binding site residence time can influence activation. The AD could affect not only activation but also affect binding. The multivalent interactions of the AD can provide additional stabilization of DNA bound TFs. In this manner multivalent interactions of TFs increase promoter occupancy and/or residence time or induce the formation of a phase-separated compartment.

TFs are key regulators of cell identity that convert biological signals to distinct patterns of gene expression. Their aberrant activity is a central part in the malignant transformation of cells into cancer cells. For instance, in Ewing's sarcoma a fusion TF (EWS-FLI1) drives the activation of specific enhancers and their oncogenic target genes [8]. One major task for TFs is to confer specificity to the control of their target genes in the presence of heterogeneous concentrations [155], low-affinity binding sites [156] and off-target binding [72]. Traditionally, binding site affinity has been considered a central TF property, but the possibility to compensate for low affinity by high TF concentrations poses a problem for specific regulation [157]. The impact of some TF features like TF binding site affinity and occupancy have been studied in detail and incorporated into quantitative models [69]. However, transcription is a dynamic process with a high turnover of regulatory factors [158] and multiple distinct states of the gene promoter [159]. Two dynamic properties of TFs have only recently gained more attention: the residence time and the ability to form phase-separated compartments mediated by multivalent interactions of the AD. Recent studies have implicated TF residence time [76, 80, 160] and phase-separation of the activation domain [39, 100] as important regulators of transcription. Residence time, i.e. the time a TF remains bound to its target site, is independent of TF concentration which makes it a good candidate for a TF property that controls specific gene activation. The multivalent interactions of ADs that underlie AD phase separation could cooperatively stabilize TF binding, especially at clusters of binding sites. Moreover, phase separation caused by AD-AD interactions has been controversially discussed as a mechanism for strong gene activation [100, 107]. This indicates that TF binding and activation could be interconnected. In **Figure 1.6** potential connections between the activities of DBD and AD are depicted. Nevertheless, major questions for this new dynamic model of gene regulation remain unanswered: Does the formation of phase-separated TF compartments have functional consequences for the gene induction process and, if so, which mechanism drives this enhanced activation? Is there a causal relationship between short residence times and low activation potential of TFs? Importantly, is there a true effect of the residence time or is reduced activation caused by low binding site occupancy? Are the dynamics of transcriptional induction, i.e. the speed with which a gene is activated, changed by TFs with different properties?

To answer these questions, methods are needed to measure the binding kinetics of transcription factors, the dynamics of transcriptional induction and to simultaneously observe the presence or absence of phase-separated compartments in heterogeneous cell populations. This thesis established tools for the analysis of the TF binding kinetics and light-induced transcription time courses. It systematically modulated TF binding kinetics and multivalent interactions and dissected their effect on the efficiency and dynamics of transcriptional induction and of the H3K27ac-BRD4 co-activation pathway.

The development and application of methods to the above described research questions proceeded via the following steps:

1. Light-activated transcription induction experiments in tens of single reporter cells were developed including automated widefield microscopy and image analysis (section 2.1). A software package for image analysis *NSSQ (Nuclear spot Seg-*

mentation and Quantification) is provided. Two proof-of-concept applications highlight the potentials of this method: In the first application example (section 2.1.6), promoter memory effects were studied, when the TF was recruited a second time after a dissociating during a dark phase. In the second application (section 2.1.7), stochastic models confronted alternative activation mechanisms with the heterogeneous single-cell data.

2. A fast and robust method for the measurement of residence times at a cluster of binding sites by widefield microscopy-based Fluorescence Recovery After Photobleaching (FRAP) is described, as well as a software package *NSQFRAP* for the subsequent automated image analysis and parameter estimation (section 2.2). The fast and simple acquisition of recovery curves allowed to probe the binding of multiple complexes in a large number of single cells.

3. TF binding and interaction properties were modulated using a set of synthetic TFs modularly constructed from bacterial repressor- or CRISPR/dCas9-DBDs and several viral or mammalian ADs (section 2.3). The ADs were investigated for their potential to phase-separate in PHR-AD fusions (section 2.3.2) and two ADs with respectively low and high phase separation propensity, VP16, a viral TF, and VPR, a synthetic multicomponent TF, were selected for a focused analysis of binding and (co-)activation properties.

4. Subjecting the panel of synthetic transcription factors to measurements of their binding kinetics by FRAP and to light-induced transcription induction experiments the following observations were made:
    - The ability of ADs to self-interact stabilized weakly bound chromatin complexes, thereby possibly enhancing their ability to successfully induce transcription (section 2.4.1).
    - Reducing TF residence times under controlled conditions strongly decreased transcription (section 2.4.3). This effect was independent of binding site occupancy. A kinetic proof-reading model of transcriptional induction offers a potential mechanism for such a residence time dependence (section 2.4.2).
    - TFs with strong multivalent interactions activated their target gene faster and more strongly (section 2.5.1), but the formation of phase-separated com-

partments did not enhance (section 2.5.2) and in certain configurations even suppressed transcription (section 2.5.3).

- The co-activation marks H3K27 acetylation and BRD4 recruitment were neither strictly necessary nor sufficient for the activation of the reporter gene studied here (section 2.6.1). Specifically, one TF configuration is described which induced co-activation marks but not transcription. The co-activation marks appear to have a boosting effect on transcription. Differences in the activation strengths of VP16 and VPR could be rationalized by differences in their ability to self-interact, directly recruit BRD4 and mediate H3K27 acetylation (sections 2.6.2 and 2.6.3).

The results of this thesis indicate that liquid-liquid phase separation of TFs is not required for strong transcriptional activation. Instead it reflects the underlying multivalent interactions of the AD which stabilize direct DNA binding and facilitate co-activator recruitment. The role for dynamic TF properties that was observed here raises new research questions. For instance, how gene activation is limited for TFs with short residence times. The observations of this thesis lead to a model of TF binding and activation in which contributions from both the DBD and the AD modulate transcription activation capacity. The design principles for strong synthetic TFs are discussed as well as the implications and potential interventions for oncogenic fusion TFs that have acquired a new intrinsically disordered domain with strong multivalent interactions.

Chapter 2

Results

2.1 Recording light-induced transcription time courses in single cells

The activation of a previously inactive gene is a dynamic process that comprises multiple molecular events like the recruitment of the transcriptional machinery, chromatin modifications and transitions of the local chromatin structure [11]. Methods to observe the activation in real-time promise to reveal regulatory mechanisms, clarify the temporal order of activation events and to identify essential activation steps. As an example, a chromatin modification that is found to occur after the start of productive transcription by such a time course experiment could be assumed to be non-essential for transcription initiation and elongation. While the activation process can be precipitated by natural stimuli, e.g. by cytokine stimulation [161], a mode of activation that does not confound the observed transcription dynamics with a potential delay caused by the signal transduction upstream of the TF recruitment is desirable. Such experimental systems that are decoupled from cellular signaling pathways are necessarily artificial, but offer tight control of stimulus timing. Light-induced TF recruitment has advantages over chemically induced TF recruitment in being faster and reversible without washing steps that would limit temporal resolution. The BLInCR toolbox [12] has provided the molecular tools to control the recruitment of a TF to a reporter gene by light using the optogenetic PHR-CIBN dimerization module. Here, an advancement of this method is presented that allows to automatically acquire and analyze transcription induction time courses in a large number of living cells. The technical details for the implementation of experiments and image analysis are described in **sections 4.2.2 and 4.2.3** of the **Materials and Methods** chapter. Single-cell transcription trajectories can be very heterogeneous and the ability to record tens to hundreds of time courses makes it possible to study this heterogeneity and to obtain robust estimates of activation speed and strength under different conditions.

2.1.1 The acquisition of light-induced transcription time courses was automated

The experimental strategy to induce a reporter gene by light-recruitment of an activation domain follows the approach of the original BLInCR toolbox [12], but introduces some modifications. As with the original BLInCR toolbox, the synthetic light-responsive TF is directed to the reporter gene cluster in the U2OS 2-6-3 cell line [144] (**Figure 2.1 A**). The cell line contains multiple copies of the reporter unit stably integrated in tandem at a single locus. Each unit contains 256 *lacO* and 96 *tetO* sites followed by a minimal CMV (Cytomegalus virus) promoter. The reporter gene body encodes peroxisome-targeted CFP-SKL and contains 24 repeats of the MS2 sequence which forms stem loops in the transcribed RNA and can be bound by the MS2 coat protein (MCP). Transcriptional activity of the reporter gene cluster can, thus, be monitored by microscopy of fluorescently tagged MCP that accumulates at the reporter cluster, as MS2 binding sites in the nascent RNA become available. Here, tdMCP-tdTomato was used, which has a higher affinity for MS2 stem loops because it does not require previous dimerization of MCP [162] and which does not accumulate in nucleoli (**Figure 2.1 B**) in contrast to other fluorescent protein (FP) fusions like tdMCP-EGFP. The *tetO* sites were targeted with CIBN-rTetR. The reverse tet repressor (rTetR, TetOn system) binds in the presence of tetracycline or doxycycline (dox) [163]. This has two adavantages over using TetR (TetOff), which binds only in the absence of the reagent. First, binding sites remain unbound until dox addition at the start of the experiment avoiding potential problems with DNA replication in cells with a tightly TetR bound repetitive array. Second, transfected cells will be sensitive to light only after dox addition reducing the chance of premature reporter gene activation by stray light. A similar approach has been developed in parallel by other research groups [164]. In transcription time course experiments dox is added 15 min before the start of image acquisition. This is sufficient, since rTetR binding at typical expression levels saturates after few minutes [72]. In the interval between

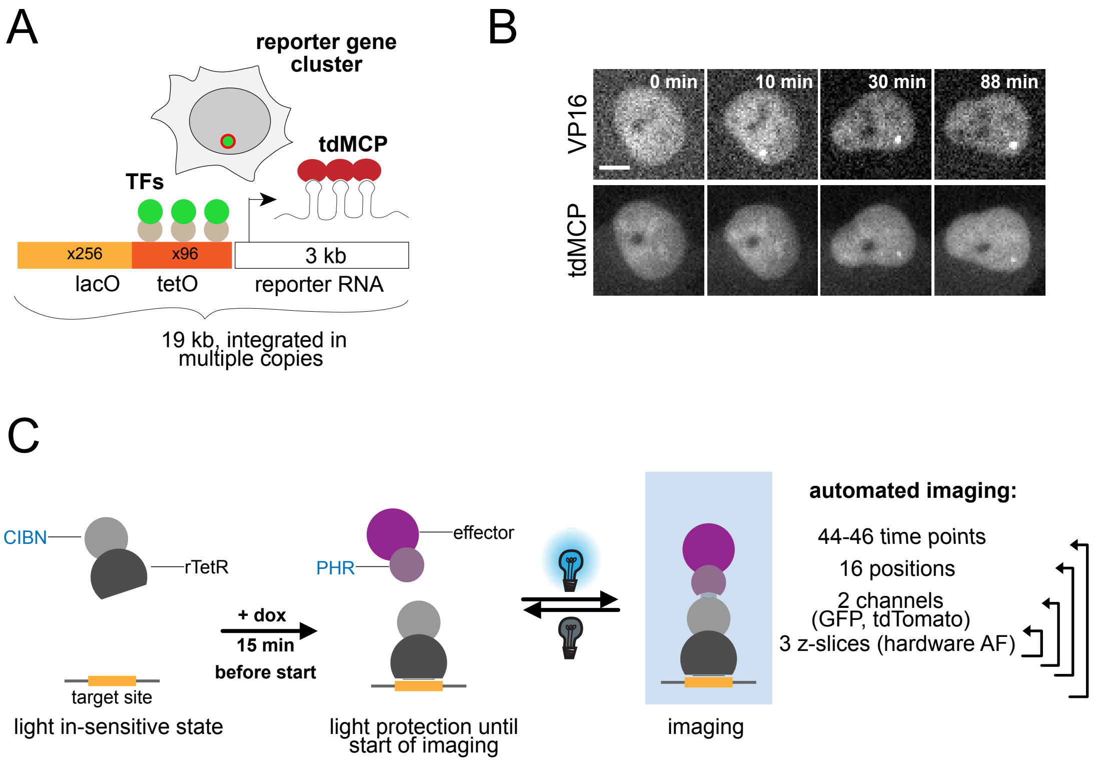


Figure 2.1: Light-induced transcription time course experiments. **(A)** Structure of reporter gene locus in the U2OS 2-6-3 cell line. Multiple copies of the array are integrated at a single locus in a tandem array. Each unit consist of repeated *lacO* and *tetO* binding sites followed by a minimal CMV promoter and the reporter gene body. The reporter RNA encodes peroxisome-targeted CFP and contains 24 MS2 sequences. The MS2 sequence forms stem loops that can be bound by fluorescently tagged tdMCP protein. **(B)** Exemplary image time courses for activation by the light-inducible TF PHR-EGFP-VP16 (upper row) inducing production of nascent RNA labelled by tdMCP-tdTomato (lower row). Scale bar: 10 µm. **(C)** Schematic of experimental procedure. CIBN-rTetR binds to the *tetO* sites, when doxycycline is added 15 min before the start of image acquisition. Imaging exposes cells to blue light leading to the binding of PHR-EGFP-effector to promoter bound CIBN-rTetR, so that transcription is induced and RNA production can be followed by imaging tdMCP-tdTomato. The TF dissociates, once imaging is interrupted removing the blue light stimulus. Typical settings for the automated multi-position, multi-channel time-course z-stack acquisition are shown on the right.

dox addition and the start of image acquisition cells need to be protected from any light exposure to avoid premature activation. The start of imaging exposes the cells to blue light which triggers a conformational change of the PHR domain in PHR-EGFP-VP16 and allows it to bind to the promoter bound CIBN-rTetR (**Figure 2.1 C, left**). From this time point on transcription is induced, as an activation domain is present at the promoter. Automated imaging is performed on a widefield microscope equipped with a 20x objective, a motorized stage and software modules for multiposition, multichannel z-stack acquisition over a defined time course (**Figure 2.1 C, right**). Such a microscope setup offers much lower resolution than a confocal microscope with a higher magnification objective, but is sufficient to measure spot intensities and is faster and more robust to movements of the sample in z-direction. Moreover, the large field of view allows to image more cells than would be possible with higher magnification. Using a hardware autofocus is a pre-requisite to acquire images with a stable focus over the duration of the time course. Imaging parameters were optimized to comprise three z-slices ($1.0\,\mu\mathrm{m}$ distance), two channels (tdTomato for nascent RNA, GFP for the TF and triggering recruitment), 16 positions (in a grid of 4x4) and 44-46 time points with 2 min intervals over 90 min duration. The complete protocol has been made available in reference [165]. These settings represent a good compromise between high throughput, low light exposure (to avoid fluorophore bleaching, cytotoxicity or overly strong optodroplet formation) and high time-resolution. The resulting image series allow to follow the recruitment of the TF to the reporter gene cluster and then the production of nascent RNA by the enrichment of fluorescently tagged tdMCP at the cluster (**Figure 2.1 B**). These image series can then be segmented, tracked and the intensity enrichment at the spot can be measured to extract quantitative RNA production trajectories. Typically, a gene cluster activation experiment results in a set of 4320 single images prohibiting manual segmentation, tracking and quantification of cluster spot intensities. Rather, this analysis needs to be at least partially automated.

2.1.2 The *NSSQ* software package provides an automated workflow for image time course analysis

The analysis of the multidimensional image series requires the selection of cells of interest, segmentation of the nucleus and of the gene cluster spot, tracking these regions between time points, measuring their intensities and a quality control step. Here, a semi-automated approach was chosen. Suitable cells that show recruitment of the TF to the spot are chosen manually and time courses are individually curated, but all processing steps inbetween are automated. This allows to measure large numbers of cells under different conditions, while the manual selection and curation steps are still feasible for several tens or hundreds of cells. The analysis pipeline was implemented in the R programming language [166] building on the *EBImage* package [167]. For ease of use and simple adaptivity to variations of the experimental design most tasks were implemented in functions and collected in the *NSSQ* package (Nuclear Spot Segmentation and Quantification) associated with reference [165] which also contains four example scripts demonstrating their application. Table 2.1 summarizes the functions and their usage.

Table 2.1: NSSQ functions

Function	Application
alwaysSegment	Segmentation function for nuclei that adjusts thresholding offset until there is an object at the image center.
correctBleach	Correct image intensity lost due to photobleaching
cropMovingObjects	Crop an image series around the coordinates of a moving object.
cropPosition	Crop the same region around this positions in all frames of the image series
fillTrackGaps	Fill up empty mask frames of a time series.
filterObjects	Filter objects according to a selected feature
filter_spots	Filter spots
followTrack	Follow a specified object along the pairwise mapping of consecutive frames
measureGradient	Create an image with gradient values for each pixel
NSSQ	Nuclear Spot Segmentation and Quantification: Segment and track nuclei and spots inside them and quantify their intensity above background
projectionTracking	Use the maximum projection of an image series to track the brightest object
readImageRegion	Load subregion of image series
readMask	Read mask images
selectInFocus	Select plane with best focus of z-stack
selectMainSpot	Tracking function for the brightest spot in the nucleus
selectPositions	Interactive selection of an arbitrary number of positions in each frame

spots_singleFrame	Segmentation of spots in the reference image inside the mask based on intensity quantiles
spot_segment	Segmentation of spots in the reference image inside the mask based on intensity quantiles
subtractBG	Local background subtraction
thresh_rad_border	Segmentation function for nuclei
tiltedPlane	Create image of background intensity gradient
track	Track objects between two consecutive mask frames
twoWayTrack	Track selected object forward and backward in time

The analysis workflow has six major steps: selection of cells, segmentation and tracking of the nucleus, z-plane selection, reporter gene spot segmentation and tracking, measurement of region areas and intensities and annotation of segmented image series (**Figure 2.2**). The first and the last step require interaction of the user and, therefore, the workflow is distributed on three analysis scripts of which the second contains all the fully automated steps. The analysis starts with the manual selection of cells with visible TF recruitment in the central z-slice at the third time point (4 min to 6 min). At this time PHR-EGFP-AD is already enriched at the reporter gene cluster due to the fast PHR-CIBN interaction kinetics. The manual control of this step ensures that cells without recruitment but potentially containing other bright regions are not used for the further analysis and prevents the analysis of cells with abnormal morphology, e.g. due to cell death. Next, each selected nucleus position is processed individually. To avoid exhausting the working memory only a defined image region around the selected nucleus position is loaded for all time points, channels and z-slices. The nuclear localization of the PHR-EGFP-AD construct allows to segment the nucleus by local thresholding and segmented nucleus objects are tracked between consecutive images by minimizing the euclidean distance of the object positions. Despite the hardware autofocus there can be a small drift of the nucleus z-position over the time course. To correct for this drift the z-plane with the best focus is determined individually for each time point using the sum of intensity gradient values inside the nucleus as a proxy for image sharpness. Then, the reporter cluster is segmented in the PHR-EGFP-AD channel based on a threshold determined from intensity quantiles in the nucleus region. The brightest object is tracked along the time course by a two-step approach: First the region of the spot is determined by creating and segmenting a maximum time-projection. Then the spot is tracked inside this refined region by distance minimization. This approach reduces the number of possible object mappings between time frames with a large number of segmented objects (e.g. in cases in which the nuclear GFP signal is very granular due to optodroplets). A mask of local background is created as a ring around the spot area. Segmentation and tracking of nuclei and gene clusters is challenging due to the large variability of construct expression levels and fluorescence intensity. To ensure robust processing the threshold for nuclear segmentation is adapted iteratively until there is successful segmentation and tracking over the whole time course. Once the nucleus, spot and background masks are created, their areas and mean fluorescence intensities in all channels are measured. Finally, the segmented image series has to be inspected in the nascent RNA (tdMCP-tdTomato) channel to exclude cells with lacking tdMCP-tdTomato expression, problematic morphology, failed nucleus or spot segmentation and failed tracking. Moreover, cells are annotated as responders if tdMCP-tdTomato intensity is visibly enriched in the spot region or as non-responders, otherwise.

The quantitative nascent RNA time course $R(t)$ is calculated from the measurements as the integrated fluorescence intensity in the spot area. First the local background $I_{bg}(t)$ is subtracted, then the mean spot intensity I_{spot} is multiplied by the spot area A_{spot}:

$$R(t) = (I_{spot} - I_{bg}) \cdot A_{spot} \tag{2.1}$$

The multi-position acquisition of time courses leads to a moderate time shift between cells imaged at the first and last position, since cells will be exposed to (stray) excitation light at all positions from the start of imaging. Consequently the exact time points need to be calculated from the position and imaging intervals. For averaging of single-cell time courses across positions the nascent RNA signal at common time points is approximated by linear interpolation.

If the strength of the transcriptional response is of interest the interpolated values of $R(t)$ can be averaged directly and the strength of the activator can be read from the time course plateau. For these comparisons experiments should ideally be performed in a single microscopy session to avoid differences in detection efficiency (e.g. due to aging of the excitation light source). If the dynamics of the transcriptional induction is of interest, the time courses can be normalized to their maximum before averaging.

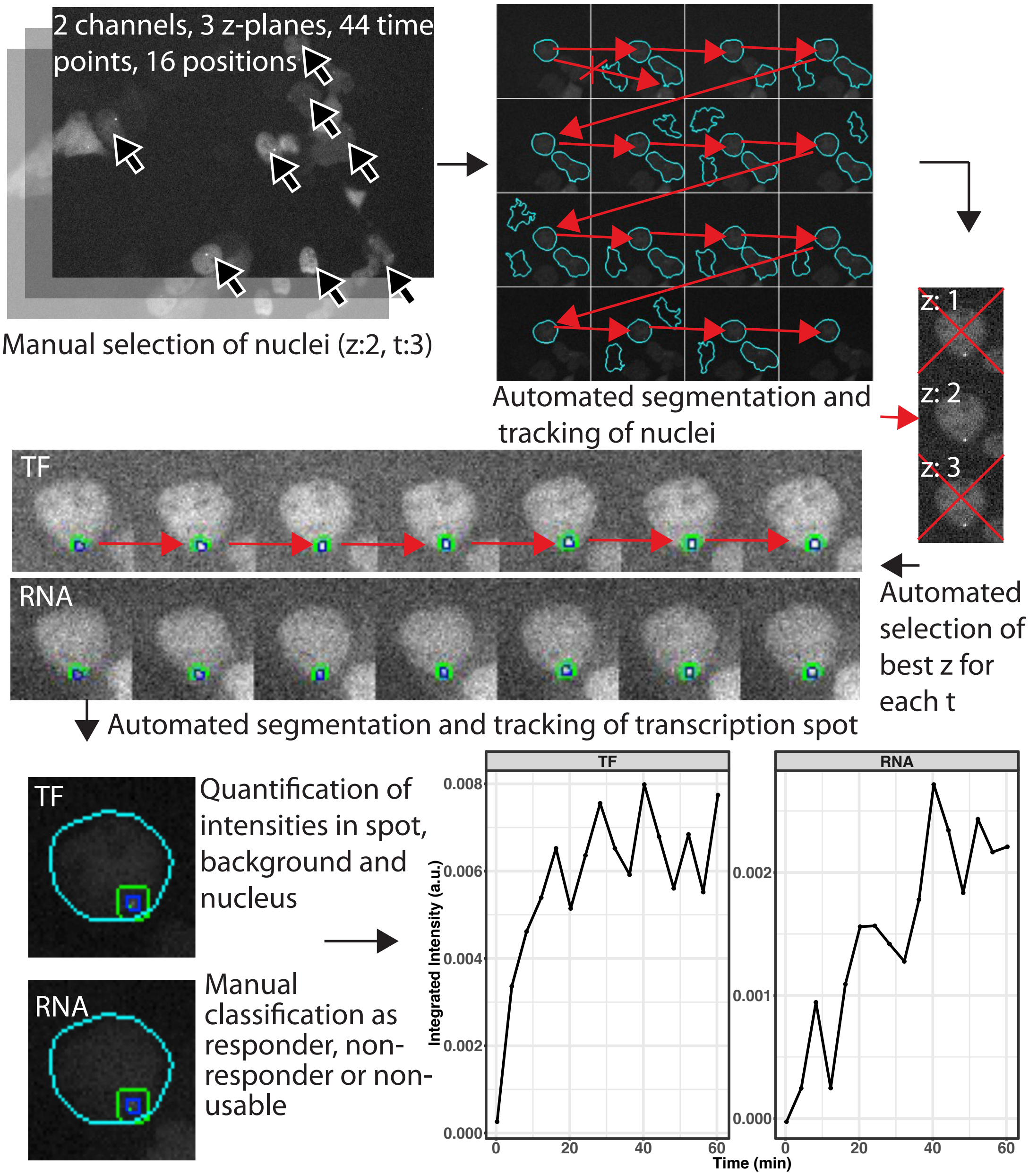


Figure 2.2: Automated image analysis workflow of transcription induction experiments with the R-package *NSSQ* (Nuclear Spot Segmentation and Quantification). The analysis uses the multidimensional image hyperstacks from the time course experiments. Positions of usable nuclei are selected manually in a single z-plane and time frame for each position. Nuclei are segmented and their position is tracked between consecutive time points. The best z-plane for each time point is selected based on the image intensity gradient. The transcription spot is segmented in the PHR-effector channel and the spot is tracked over time. Intensitities are then measured in the spot, a ring-shaped local background area around the spot and in the nuclear area. Finally, segmented cells are manually classified as responder or non-responder cells dependent on the visible accumulation of tdMCP-tdTomato intensity at the reporter spot. Cells with problems of the segmentation or tracking, or morphologies that indicate cell death or mitosis during the time course are sorted out. (Figure adapted from [165].)

The above described workflow comprising automated microscopy and image analysis can be easily adapted to study transcriptional induction under various conditions (**Figure 2.3**). In the base version of the protocol the dynamic activation behavior of a single activator is studied. The transcription dynamics and the heterogeneity in the response of single cells can give insight into the mechanism of transcriptional activation (see section 2.1.7). The activation domain in PHR-EGFP-AD can be easily exchanged, so that its influence on the activation speed can be investigated (see section 2.5).

If certain phases of the activation dynamics are suspected to depend on a molecular event, e.g. a chromatin modification, using inhibitors of this event allows to see the effect on the dynamics (see section 2.6.1). Finally, the reversible interaction of PHR and CIBN allows to remove the induction stimulus for a defined time before re-stimulation. Such experiments are informative about transcription persistence in the absence of stimulation and about promoter memory of past activation events (see section 2.1.6)

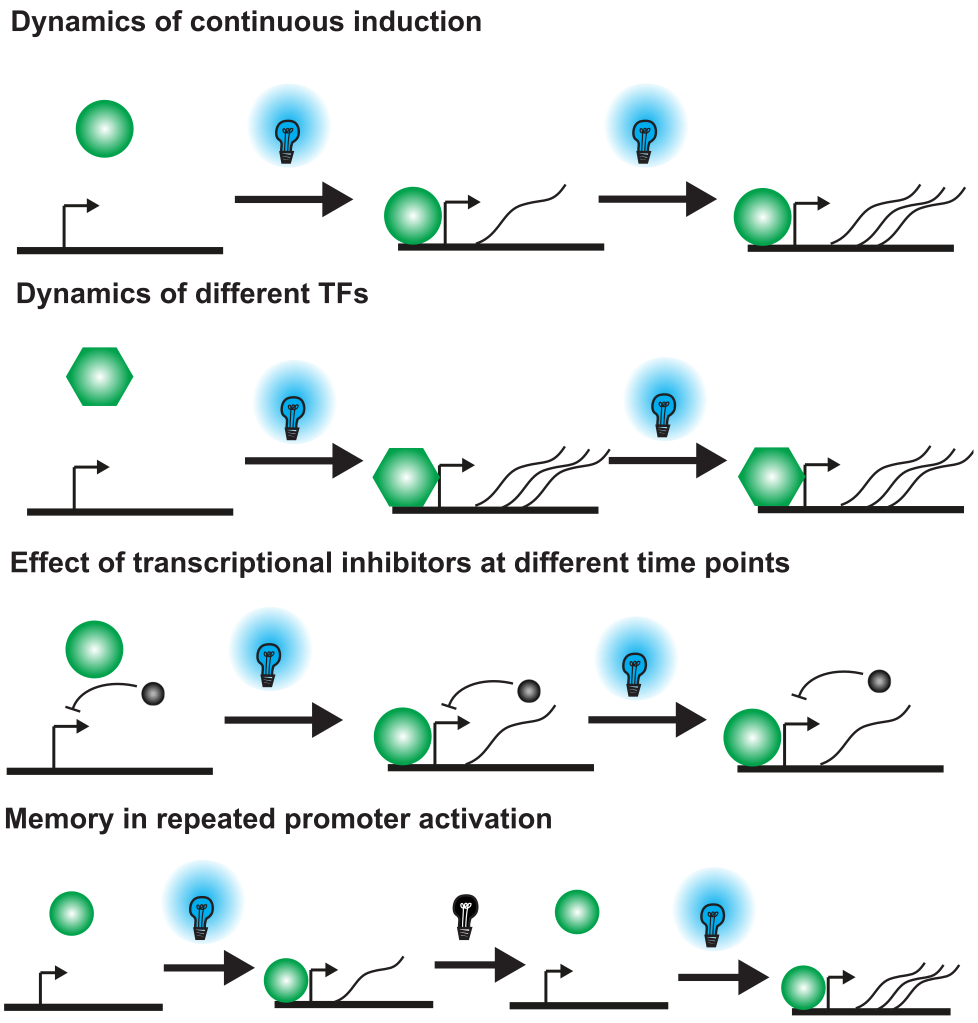


Figure 2.3: The protocol for light-induced transcription time courses can be modified to answer a wide range of biological questions. From top: In the standard approach an effector of interest is continuously recruited to the promoter at a defined time point and the dynamics of transcriptional induction are recorded. Using alternative activation domains allows to draw conclusions about differences in the timing and mechanisms by which TFs induce transcription. The use of inhibitors that target known steps of the transcriptional induction process can give insights, which steps of the induction process are responsible for which phases of the induction dynamics. Removing the light-stimulus and then re-stimulating allows to study promoter memory of past activation events and potential sensitization or refractoriness.

2.1.3 The workflow provides stable measurements of early induction time courses

The experimental and analytical workflow is subject to several sources of technical noise that could distort the quantitative intensity time courses. Among these are z-movements of the gene cluster out of the focus plane, imperfect segmentation of the reporter spot and bleaching of fluorophores. To asses, whether the setup is able to provide a stable readout of the biological signal, cells were transfected with EGFP-LacI which binds to the reporter array and should provide a constant signal over time. For most cells the spot intensity was stable over the 90 min acquisition time (**Figure 2.4 A**), even though there were few cases in which there was a drift of the signal. In one case this amounted to a change of 30% of intensity.

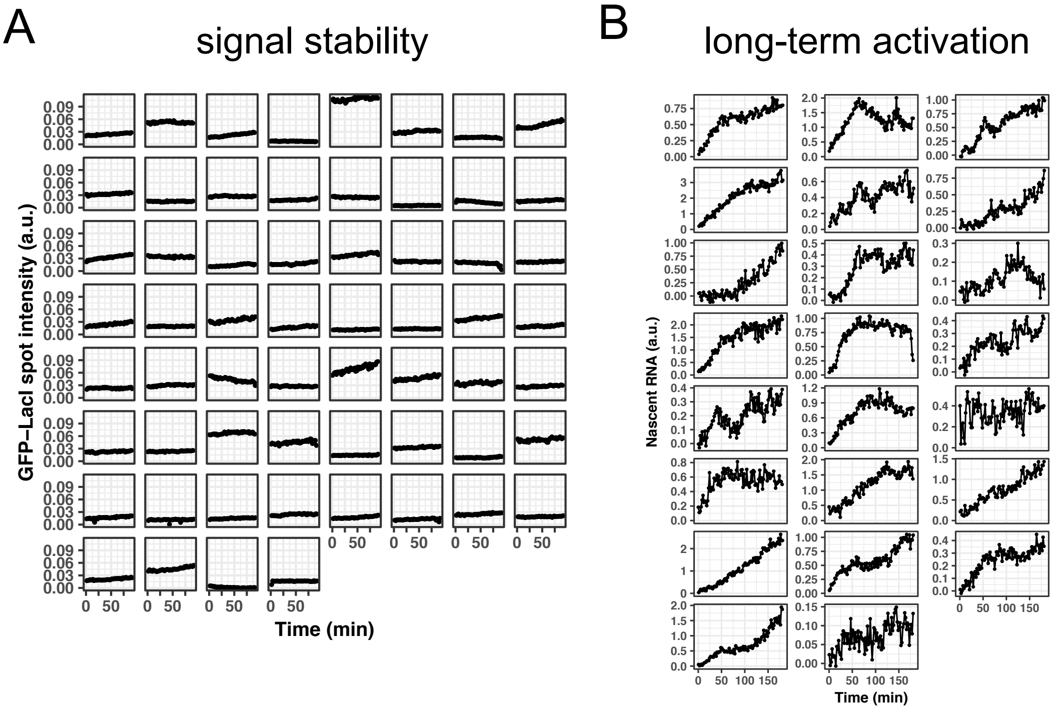


Figure 2.4: Robustness of experimental procedure in single-cell spot intensity time courses. **(A)** Signal stability of constitutively bound EGFP-LacI at the promoter cluster. The intensity remained constant for most cells over the whole time course. The variability of spot intensities between cells indicates heterogeneous reporter copy numbers. **(B)** Long term activation by PHR-EGFP-VP16 led to a saturation of the nascent RNA signal at 90 min in many cells, but there were several cells, for which the signal kept increasing over the whole 180 min time course.

Nevertheless, the experimental setup overall provides a robust readout of the accumulated intensity at the reporter gene cluster, but some trajectories could exhibit technical problems. These are expected to be smaller than the true signal increase due to induction of transcription.

Typical experiments have a duration of 90 min and RNA production appears to approach a plateau after this time. A long term activation of 3 h was performed to check if there is additional activation after this time point (**Figure 2.4 B**). While for many cells nascent RNA levels did not rise beyond the 90 min time point, for several cells the levels continued to rise or showed a second rise after an intermittent plateau phase. Consequently, the 90 min time courses can provide insight into the early activation dynamics and full activation can only be observed at later time points. This effect should be more pronounced for slow activators.

Normalization of single-cell time courses to their apparent maximum could lead to an over-estimation of activation speed for slow activators. In a comparison of fast and slow activators the difference in their dynamics could, thus, be underestimated, when comparing normalized time courses. Importantly, time course duration cannot be chosen to be arbitrarily long, since already during the first 90 min there is some cytotoxicity of the activated PHR-EGFP-AD construct leading to the loss of some cells. For longer time courses this could affect a large part of the transfected and activated cell population. The 90 min duration is, thus, a good compromise between capturing the early gene activation dynamics and imaging a large number of cells without strong cytotoxicity.

2.1.4 The copy number and RNA production capacity of the reporter gene were determined

A quantitative view of the reporter gene array and its activation is important to understand the order of magnitude for copy numbers of involved molecules and if stochastic effects are to be expected. This regards the copy number of reporter gene units and the number of reporter RNA molecules that are produced. In typical widefield experiments the cluster of reporter genes appears as a homogeneous spot, in which the labeling fluorescent proteins are enriched. In order to be able to distinguish and count single reporter gene units the chromatin of the gene cluster was decondensed by recruiting dCas9-EGFP-VPR to the *tetO* sites and *lacO* sites were labeled by SNAPtag-LacI. The activation domain VPR not only activates transcription, but also leads to strong chromatin decompaction in a subpopulation of cells [168, 169].

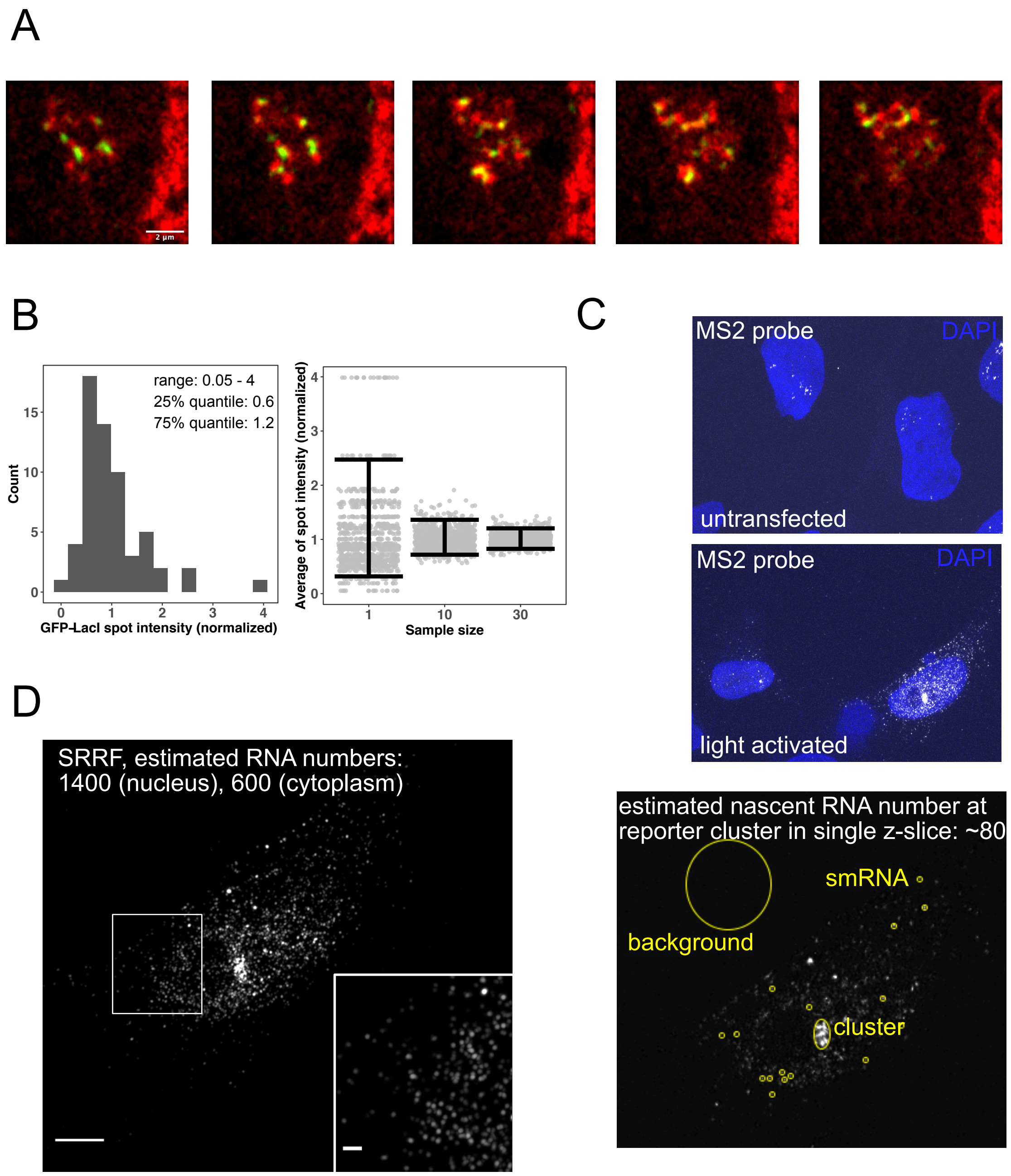


Figure 2.5: Structure and transcription quantification of reporter gene cluster **(A)** Structure of decondensed reporter locus bound by dCas9-EGFP-VPR (green) targeted to *tetO* and SNAPtag-LacI (red) imaged by SRRF. Super-resolution imaging by SRRF allows to distinguish presumably ∼ 20 single reporter gene units showing an alternating pattern of *tetO* and *lacO* sites. Scale bar: 2 µm. **(B)** Distribution of reporter gene copy number measured by EGFP-LacI spot intensity (left, normalized to mean value). Data corresponds to Fig. 2.4 A. (Continued on next page)

Imaging on a spinning-disc microscope with the super-resolution method SRRF (Super-resolution by radiality fluctuations) [170] then allowed to distinguish separate spots of dCas9-EGFP-VPR (green) and SNAPtag-LacI (red) (**Figure 2.5 A**). The alternating pattern of labeled *lacO* and *tetO* sites that resembles the tandem repeat architecture of the reporter gene locus suggests that these spots correspond to individual reporter gene units. Counting the GFP spots led to an estimate of 20 reporter gene units in the cluster. Repetitive sequences can undergo amplification and contraction by recombination events [171] and this is expected to be even more prominent in cancer cell lines, since one hallmark of cancer is genome instability [172]. Indeed, the copy number of reporter genes determined from the intensity of promoter bound EGFP-LacI showed a broad distribution (**Figure 2.5 B, left**). While the cell with the strongest spot had a 80-fold higher (integrated) spot intensity compared to the cell with the weakest spot, typical differences were around 2-fold (between the first and third quartile). The variability of reporter copy numbers could directly contribute to heterogeneous response amplitudes between single cells and needs to be considered, when interpreting the distribution of single-cell transcription responses.

The copy number heterogeneity should have a limited effect, when averaging over a sufficient number of cells as is demonstrated in **Figure 2.5 B (right)**, where the average of 1000 random samples drawn from the empirical EGFP-LacI distribution is shown. It is reasonable to assume that the amount of RNA produced at the gene cluster is directly related to the reporter gene copy number. Taking random samples from the

Figure 2.5: (Continued) The mean of samples taken from the empirical distribution (left) follows a distribution that becomes more narrow with increasing sample size (right). Dots: mean of 1000 random samples for each sample size; error bars: 5 and 95% quantiles. **(C)** Estimation of nascent RNA molecules number at reporter cluster by sm-RNA-FISH. The number of RNA molecules labeled by MS2-probe (white) was much lower for untransfected cells (top) compared to activated cells (center) transfected with CIBN-rTetR and PHR-EGFP-VP16 and light-exposed over night. Nuclei were stained with DAPI (blue). Comparison of the cluster intensity with single spot intensities (bottom) allowed to estimate the number of RNA molecules at the reporter gene cluster. **(D)** SRRF image (maximum projected z-stack) of activated cell shown in (C). RNA numbers in nucleus and cytoplasm were estimated by counting spots in defined regions and multiplying the point density with the nuclear or cytoplasmic area, respectively.

gene copy number distribution should thus allow to estimate the inter-cell variability introduced by this heterogeneous factor. For a sample size of 10 cells 95% of samples have a mean within ±35% of the full population average and for samples of 30 cells the deviation is reduced to ±20%. Nevertheless, heterogeneous gene copy numbers can confound the comparison of transcription time course amplitudes for single cells or small batches of cells. Ideally, copy number would be measured for each cell when acquiring transcription time courses, but light-induced time courses with three colors on a widefield microscope remain challenging.

The labeling of nascent RNAs by tdMCP-tdTomato allows to compare relative levels and acquire time courses in arbitrary units (a.u.). In order to also determine absolute numbers of RNA production single molecule RNA fluorescent in situ hybridization (sm-RNA FISH) was performed on cells that were either activated by light-induced TF recruitment (PHR-EGFP-VP16 + CIBN-rTetR, 24 h illumination) or left untransfected to determine basal levels. Basal levels amounted to few molecules that were mainly located in the nucleus, whereas light-induction led to strong activation in some cells (**Figure 2.5 C, top**). RNA could be detected in the cyotplasm, in the nucleus and as nascent RNA at the reporter gene cluster. To quantify the number of nascent RNA molecules at the gene cluster and mature RNA in the nucleus and cytoplams, a cell with high expression level was selected and imaged by spinning-disc microscopy and additionally by SRRF. The average integrated intensity of single molecule spots in the cell (**Figure 2.5 C, bottom**, yellow circles) was measured and the intensity at the gene cluster spot calibrated by the intensity of single spots (supposed to correspond to single molecules). From this calibration the number of nascent RNA molecules was estimated to be around 80 in a single z-slice for a strongly induced cell. The total number can be higher, when considering additional z-planes. Assuming a gene copy number of 20 each gene would, thus, be transcribed by four RNA-polymerases. This estimation has to be handled with caution due to the uncertainties of the copy number estimation, its broad distribution and the distribution of single molecule intensities in RNA FISH (up to a factor of 2). Nevertheless, it gives an order of magnitude for the transcriptional response in strongly responding cells. The total number of RNAs in the nucleus and in the cytoplasm can be determined by counting. For a better resolution that allows to distinguish single molecules even at high densities SRRF imaging was employed (**Figure 2.5 D**). Counting of single molecule spots in defined areas of the nucleus or cytoplasm and extrapolation to the full compartment areas showed that there were 1400 reporter RNA molecules in the nucleus and 600 in the cytoplasm. Even though this strongly responding cell is not necessarily representative for the whole population, it shows that basal expression levels of few molecules are clearly distinct from responding cells that produce several hundred reporter RNA molecules.

2.1.5 Light-induced binding of PHR to CIBN is a two-step process reversible on the minute time-scale

One major advantage of light-induced over chemically induced TF recruitment to study gene activation dynamics is its reversibility with relatively fast kinetics without the need of repeated washing steps. The kinetics of PHR dissociation from CIBN after removal of the light stimulus have been determined, previously [12]. It was shown that the PHR-CIBN complex has a life time of around 5 min. However, FRAP measurements revealed much shorter life times of the PHR-CIBN interaction of around 40 s (see below: section 2.4.1). Moreover, PHR dissociation curves after light removal can in some cases not be described by a single exponential decay, as would be expected for a one step dissociation mechanism. This is demonstrated by an exemplary single-cell dissociation time course of PHR-mCherry-VP16 from CIBN-LacI (**Figure 2.6 A**, data acquired by Lukas Frank) and has been described for PHR dissociating from CIBN-TetR, as well [173]. These curves have an initial short plateau before the TF intensity at the gene spot starts to drop. Potentially, this delayed dissociation could be due to optodroplets that stabilize the binding. However, PHR-mCherry-NLS/VP16 in combination with CIBN-TetR, as studied in [173] does typically not form optodroplets. In combination with CIBN-LacI the dissolution of PHR-EGFP-VP16-optodroplets was fast in those cells that formed them (**Figure 2.6 A, bottom right**). It is, thus, unlikely that the initial plateau in the dissociation time course is caused by a phase-separation process.

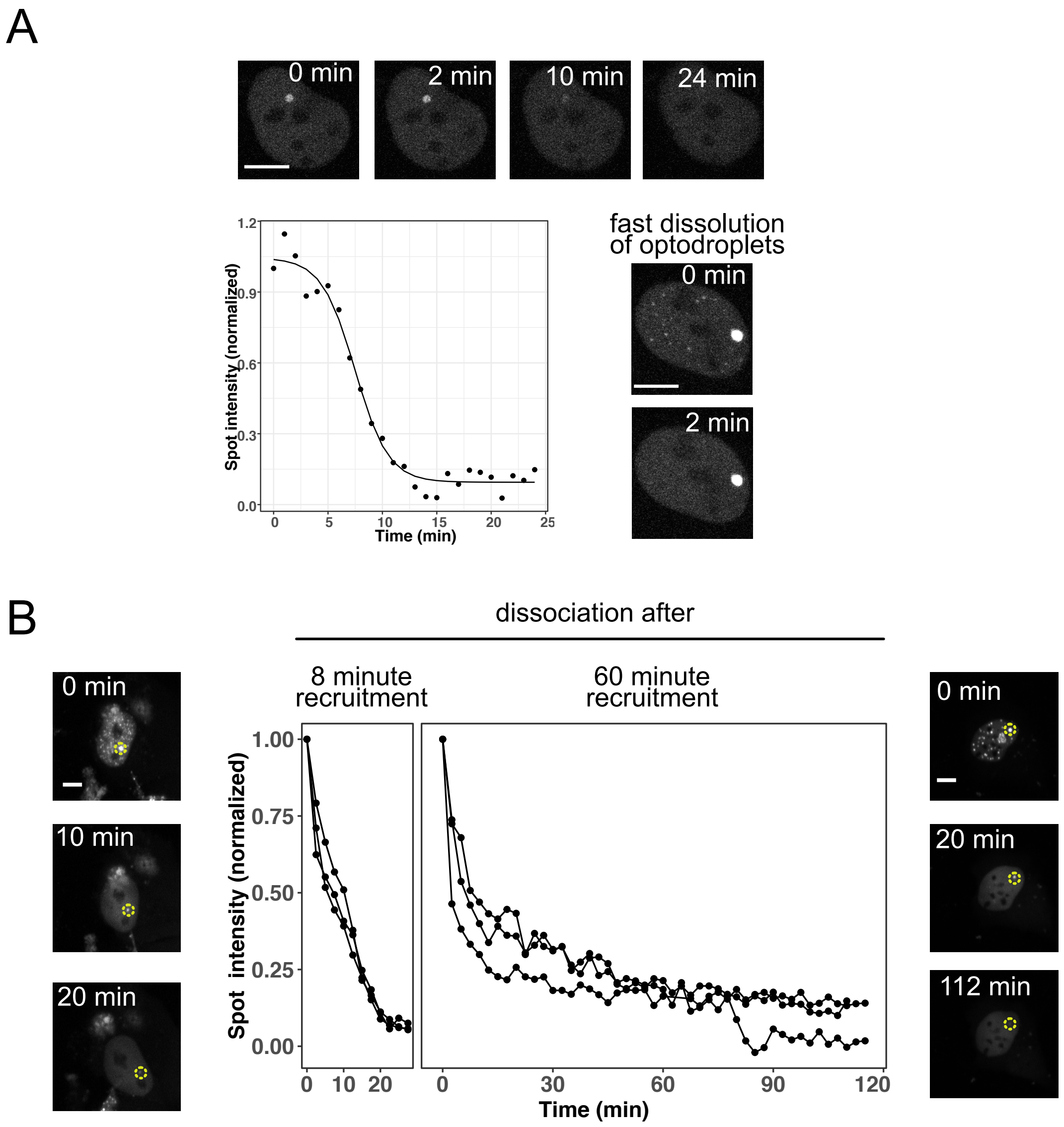


Figure 2.6: The dissociation kinetics of PHR from CIBN in some circumstances deviate from a simple one-step unbinding behavior. **(A)** Dissociation time course of PHR-mCherry-VP16 from CIBN-LacI fitted by a two-step model (see main text for details, data acquired by Lukas Frank). The cell quantified here, did not exhibit optodroplets (top row), but even for cells with optodroplets their dissolution was fast (right). (Continued on next page)

Instead, these observations can be explained by considering the mechanism of dissociation after light removal. PHR in its light activated conformation (PHR^*) can bind to and unbind from CIBN (offering free binding sites S) with the rates k_{on} and k_{off} leading to the formation or dissolution of the complex C. This binding turnover that takes place in both light and dark conditions is what is measured by FRAP. The short life times of the interaction indicate that this process is fast. When the light stimulus is removed PHR reverts from its light activated conformation (PHR^*) to its basal conformation (PHR). In the dark the conformational change occurs only from PHR^* to PHR and the following reactions take place:

$$PHR^* + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C \qquad PHR^* \xrightarrow{k_{revert}} \emptyset \tag{2.2}$$

Under the assumption that the concentration of $[PHR^*]$ is much higher than the concentration of free binding sites $[S]$ the evolution of molecule species concentrations over time can be described by a small set of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{d[PHR^*]}{dt} &= -k_{revert} \cdot [PHR^*] \\ \frac{d[C]}{dt} &= k_{on} \cdot [PHR^*] \cdot ([S_{total}] - [C]) - k_{off} \cdot [C] \end{aligned} \tag{2.3}$$

making use of the fact that the total concentration of binding sites is constant ($[S_{total}] = [S] + [C]$).

The comparison of light removal time courses with FRAP experiments (see below: section 2.4.1) suggests that the conformational reversion of PHR is slower than the binding and unbinding reactions. Under a quasi-equilibrium assumption the binding and unbinding reaction balance each other for each time point and the relation of bound (C) and unbound (PHR^*) PHR can be determined:

$$[C](t) = [S_{total}] \cdot \frac{[PHR^*](t)}{\frac{k_{off}}{k_{on}} + [PHR^*](t)} \tag{2.4}$$

Figure 2.6: (Continued) The dissociation time course (left) did not show an exponential decay, but could be fitted by a model that assumes light-dependent switching and target site binding and unbinding. **(B)** Long-term recruitment of PHR-iRFP-VP16 induced the formation of a second more stably bound fraction of molecules. Cells exposed to a short 8 min light stimulus (left) showed a fast and nearly complete dissociation of PHR-iRFP-VP16 from the reporter gene cluster in the dark, whereas there was a fraction of slowly dissociating molecules after long 60 min light exposure (right). Individual time courses for 3 cells each are shown. Note that cells with high spot intensity were chosen, which may not be representative for the whole cell population. Scale bars: 10 µm

This relation can be used to solve equation 2.3 and determine the kinetics of $[C]$ dissociation in the dark:

$$[C](t) = [S_{total}] \cdot \frac{[PHR^{*}](t=0) \cdot exp(-k_{revert} \cdot t)}{\frac{k_{off}}{k_{on}} + [PHR^{*}](t=0) \cdot exp(-k_{revert} \cdot t)} \tag{2.5}$$

This equation can be generalized by adding a basal level c and using it to fit the (normalized) dissociation kinetics of the gene spot intensity (**Figure 2.6 A**):

$$I(t) = I_0 \cdot \frac{exp(-k_{revert} \cdot t)}{b + exp(-k_{revert} \cdot t)} + c \tag{2.6}$$

The successful fit shows that the data is compatible with this two-step model of PHR conformational switching and (un-)binding.

Dissociation time courses are typically acquired after a short period of illumination that is sufficient for the saturation of binding sites. However, if the PHR-CIBN system is applied for reactivation experiments with prolonged phases of TF recruitment, dissociation and re-recruitment, the illumination period will be much longer. In order to test, whether this longer illumination affects the dissociation, cells were transfected with PHR-iRFP-VP16 and CIBN-rTetR. The near-infrared FP iRFP allows to image TF recruitment in the absence of blue light stimulation. It is also spectrally distinct from tdTomato and could be applied for the recording of TF and RNA accumulation in a blue light-independent manner. However, due to its low brightness, the need for sensitive detection and its strong propensity to form optodroplets in PHR fusions, its was only used for the dissociation experiments described here. Comparing the dissociation after a long (60 min) to a short (8 min) recruitment period revealed an additional fraction of molecules that remained bound for a longer time (**Figure 2.6 B**). The dissociation was almost complete after 25 min for the short stimulation, whereas for the long stimulation after 25 min there was still a sizeable fraction (around 25%) of molecules still bound that continued to dissociate over a time of 2 h.This was most likely not due to the formation of optodroplets, since these dissolved fast (**Figure 2.6 B, image series on the right**). Potentially, the prolonged binding could be caused by additional interactions of PHR-iRFP-VP16 with the chromatin at the promoter that are independent of the PHR-CIBN interactions. This may be coupled to transcriptional activation processes. For instance, VP16 could form long-lived interactions with components of the transcriptional machinery or modified histones could directly bind VP16 or stabilize VP16 via modfication readers. The prolonged binding of the light-controlled TF poses a technical limitation for the investigation of transcriptional persistence after stimulus removal. It may, however, also represent a biologically relevant mechanism, by which a promoter retains activators and remains in an activated state even after the stimulus has ended.

2.1.6 Proof of concept application I: Dynamics of transcriptional reactivation

One major advantage of light-controlled TF recruitment is the possibility to study the transcriptional response to discontinuous stimuli. Such situations can arise for TFs that exhibit an oscillatory translocation between cytoplasm and nucleus like p53 [4, 174, 175] and NF-κb [176]. Moreover, it can be used to determine if promoters retain memory of previous activation events. Such memory has been found in response to interferon stimulation [6] and in Drosophila development [5] and was retained over multiple cell division. Here, the experimental and analytical framework described above was applied to study transcriptional memory effects on a shorter time scale to prove that the approach is, in principle, able to detect memory effects. The experimental strategy is summarized in **Figure 2.7 A**. Cells expressing PHR-EGFP-VP16, CIBN-rTetR and tdTomato-tdMCP were exposed to blue light by imaging during two activation phases (first and second pulse) that were separated by a 60 min dark phase without imaging. In this setup the TF was recruited to the promoter during the first pulse and induced transcription (**Figure 2.7** B, upper row). In the following dark phase PHR reverted to it basal confirmation and the activator dissociated (at least partially, see section 2.1.5) from the promoter. The potential decrease of RNA production in this interval cannot be observed. During the second pulse the TF was again recruited and the recommencing RNA production was observed (**Figure 2.7** B, lower row).

Three different durations of the first pulse were chosen to determine if a potential memory effect depended on a sufficiently long first activation event: 10 min, 30 min and 60 min. Experiments were performed before the optimization of acquisition parameters in the experimental protocol was fully completed, so that there were small deviations from the procedure described above for continuous activation time courses. Foremost, the time resolution was lower (3.5 min instead of 2 min) and a larger number of positions (25 instead of 16) was acquired. This could prevent the detection of finer differences in the activation dynamics. Moreover, dox was added to all wells before the start of experiments and slides were not regularly exchanged with slides stored in the dark between experiments.

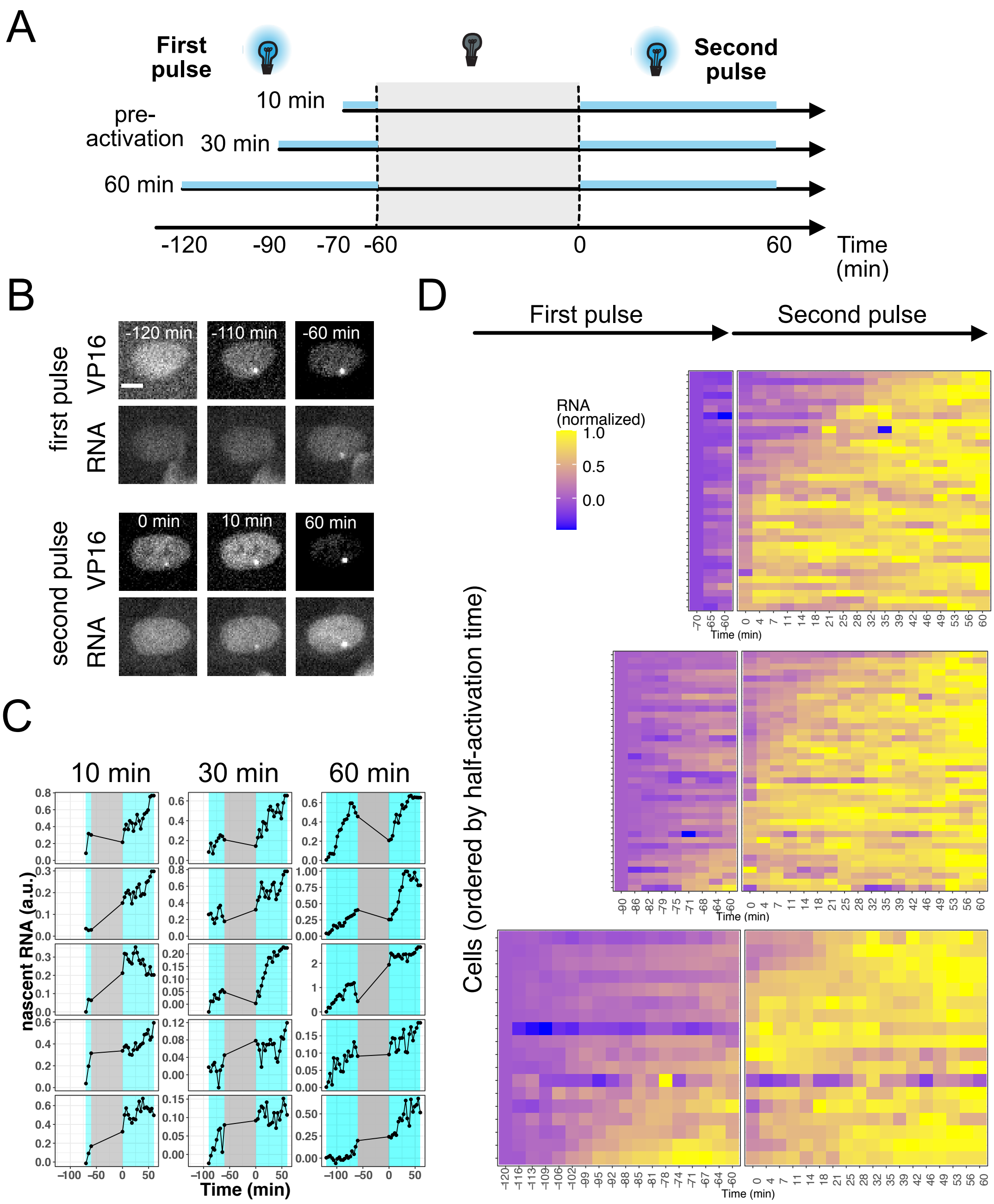


Figure 2.7: Transcriptional reactivation experiments in single cells. **(A)** Experimental strategy: Cells were exposed to blue light during imaging in a first pulse of variable duration (10 min to 60 min). This was followed by a 60 min dark phase without imaging and a second light pulse of 60 min, which was again recorded by imaging. **(B)** Exemplary image series of a 60 min pre-activation 60 min activation experiment of PHR-EGFP-VP16 (TF) and tdMCP-tdTomato (nascent RNA). Scale bar: 10 µm (Continued on next page)

]

It is noted that three replicates had to be removed from the experiment set because cells exhibited premature activation (indicated by high nascent RNA levels at the first time points or a very rapid onset of transcription) which was probably due to stray light exposure in the presence of dox. Nevertheless, the obtained results give first insights into effects of promoter reactivation and prove that this kind of experiments is technically feasible. They could easily be repeated with the optimized imaging settings and higher time resolution. Single-cell RNA production traces showed large variability in the induction dynamics for all three durations of the first pulse, as can be seen in representative examples (**Figure 2.7 C**). Yet, in all cases the RNA production reached higher levels in the second pulse. This was confirmed for normalized RNA tracks of the whole cell population (**Figure 2.7 D**). In nearly all cases maximum induction levels were only reached during the second activation pulse.

In many cases, maximum RNA production was reached early during the second pulse, while the rise of nascent RNA levels during the first pulse appeared more gradual. Interestingly, cells that reached (near-) maximal activation levels during the first pulse (**Figure 2.7 D**, lower rows for 60 min) took a longer time to again reach maximum activation during the second pulse, while a considerable fraction of cells that were not or weakly activated during the first pulse reached maximum activation at the beginning of the second pulse. This effect was not investigated in more detail, but highlights the value of single-cell time courses that allow to correlate events at different time points of the induction time course.

The overall response to the two activation pulses can be judged from averaged RNA and TF time courses. The PHR-EGFP-VP16 intensity at the gene cluster showed a fast increase during the first pulse and (for the 60 min pulse) reached a plateau (**Figure 2.8 A**). During the second pulse TF levels started from a higher base value and also

---

Figure 2.7: (Continued) **(C)** Exemplary single-cell nascent RNA time courses for the three different pre-activation durations. Illumination and imaging times are indicated in cyan. Nascent RNA signal corresponds to the integrated spot intensity in the tdMCP-tdTomato channel above local background. **(D)** Color-coded single-cell time traces of nascent RNA. Each time trace was normalized by subtracting the spot intensity at timepoint 0 min and dividing by the maximal value over the whole time course.

reached higher plateau values compared to the first pulse. The higher starting value can be attributed to incomplete dissociation of the TF, compatible with the observations in section 2.1.5, while the higher plateau might be caused by a slow increase of *tetO* binding site accessibility during the activation process, so that more CIBN-rTetR and consequently PHR-EGFP-VP16 can bind. This would also explain the relatively late saturation of TF recruitment in the first pulse that does not agree well with the association kinetics of PHR on CIBN [12, 173]. Increasing binding site accessibility is an activation mechanism mainly attributed to pioneer TFs [177], but could also happen as a consequence of promoter activation and chromatin modification by other TFs. For instance, histone acetylations loosens up chromatin by reducing histone-DNA interactions and could thereby allow additional TFs to bind to their binding site [178]. Nascent RNA levels rose more slowly compared to TF levels, but RNA production started shortly after TF recruitment (**Figure 2.8 B**). Transcriptional induction had a low increase in the beginning followed by a steeper rise in the second phase in agreement with the previously observed biphasic activation dynamics [12].

After the dark phase the starting RNA levels were approximately the same as the levels at the end of the first pulse indicating that the reporter gene activation persisted, even though it did not strongly increase. The nascent RNA levels rose faster at the start of the second pulse. Absolute values at the end of the complete time course were higher for longer durations of the first pulse, probably reflecting the longer cumulative activation time. Thus, the dark interval did not appear to reset the promoter to its state before activation.

To exclude that higher RNA levels during the second pulse were due to a global drift of intensities, the average nuclear tdMCP-tdTomato intensity was tracked over time (**Figure 2.8 C**). While there was a clear increase of nuclear intensities, this shift was on the range of ± 5 % and could be the cause of the strong differences of RNA levels observed between first and second pulse. The drift could potentially be caused by the moderate contraction of nuclei during imaging (data not shown), which concentrates the nuclear signal in a smaller area thereby increasing the signal detected by the (widefield) microscope.

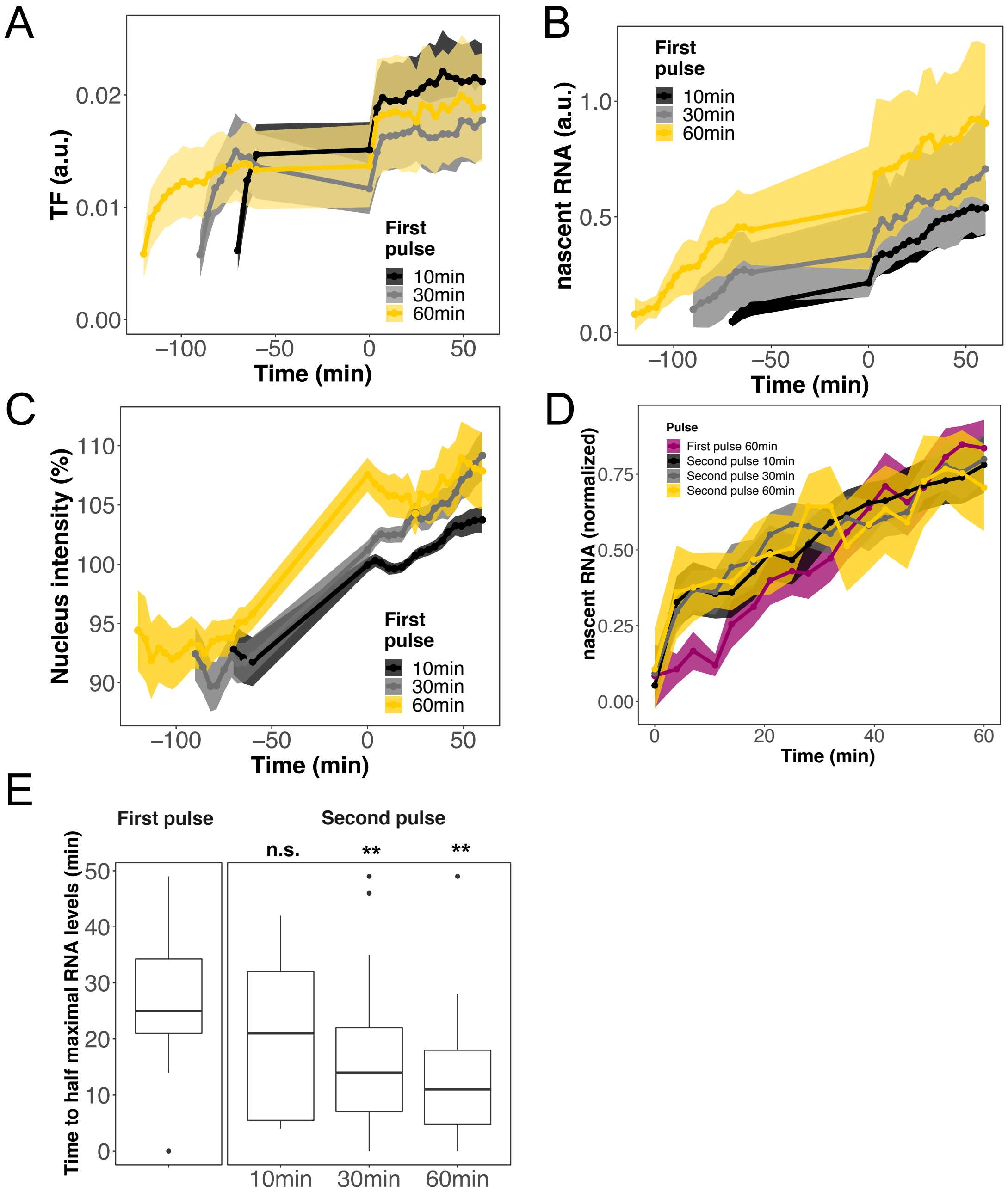


Figure 2.8: Promoters are sensitized to a second stimulus. **(A)** Average spot intensity of PHR-EGFP-VP16. Shaded area: 95% CI. **(B)** Average spot intensity of tdMCP-tdTomato. Shaded area: 95% CI. **(C)** Average nucleus intensity in tdMCP-tdTomato channel indicates moderate intensity drift. Shaded area: 95% CI. (Continued on next page)

The faster response to the second pulse could be detected by aligning time courses of the first (60 min) pulse and the second pulse and normalizing to the respective maximal values (**Figure 2.8 D**). There was a clearly delayed rise of nascent RNA levels of the first compared to the second pulse for all three conditions. Quantifying the dynamics by the time to half-maximal activation allowed to compare the dynamics on the single-cell level (**Figure 2.8 E**). While the effect was not so clear for 10 min pre-activation, pre-activation for 30 min and 60 min reduced this quantity from 26 min (first pulse) to 16 min and 13 min, respectively ($p < 0.01$, Welch two-sided t-test).

In summary, the first stimulus appears to sensitize the reporter gene promoters for re-activation. Relatively short activation events of 30 min appear to be sufficient, while the effect is less clear for 10 min activation. This is in line with observations that have been made for a small number of individual cells [12]. Even though the re-activation experiments were not performed with the optimized protocol described above, the new protocol allows to repeat them with better time-resolution and better protection from stray light activation (using multiple slides and adding dox only shortly before the experiment). The evolution of RNA levels during the dark phase could not be followed, because imaging would have exposed the cells to activating light. Nevertheless, the high RNA signal after the dark period together with the observed long-term association of a fraction of TF molecules suggests that the gene remains in an actively transcribing state, even though the gene induction process did not proceed to increase RNA production in this period. Potentially, some PHR-EGFP-VP16 molecules remain bound by interactions of VP16 with the transcriptional machinery and not by the binding of PHR to CIBN-rTetR at the promoter. This promoter state may retain some processes at a constant level like RNA Pol II initiation and elongation without a further progression of activation levels, but increase certain activation marks like binding site accessibility leading to increased TF recruitment in the second activation period. Further characterization of chromatin features at the primed promoter could give deeper insight into the mechanism of facilitated re-activation. Knowledge of such potential priming marks

---

Figure 2.8: (Continued) **(D)** Average nascent RNA levels rise faster during second light pulse. Time courses in first and second pulse were aligned by their start times. Spot intensity in tdMCP-tdTomato channel was normalized to maximal value after subtracting the value at pulse start (at -120 min or 0 min). Shaded area: 95% CI. **(E)** Distribution of times to half-maximal activation in single-cell time courses (time relative to start of respective light pulse). n.s.: not significant, **: $p < 0.01$, Welch two-sided t-test.

could then be exploited to artificially increase the response speed of target genes by selectively setting these modifications, e.g. with CRISPR/dCas9 fusion constructs, without activating the gene.

2.1.7 Proof of concept application II: Discrimination of stochastic models with single-cell transcription time course data

Transcription is a discontinuous process governed by transcriptional bursting events [44, 179] that add intrinsic noise to the RNA levels of individual genes to extrinsic sources of noise (like the cell cycle state) that affect transcription levels genome-wide [180]. The transcription dynamics described here were characterized by a large inter-cell heterogeneity both for continuous and re-activation experiments. And, while the fluctuations that characterize some single-cell trajectories can partly be attributed to technical factors, like loss-of-focus, there are also examples, in which the RNA signal shows fluctuations, even though the signal remains stable in the TF channel. This indicates that intrinsic noise sources can be detected in this setup. In contrast to typical studies of transcriptional bursting, which often study a single gene, the experimental system used here relies on a gene cluster. For genes that show independent burst events this source of noise should partly average out in the integrated cluster intensity depending on the gene copy number. If there is interaction between the promoters in the cluster, however, there could also be a collective bursting behavior. Indeed, for the system studied here a positive feedback mechanism that relies on promoter-promoter interactions has been proposed to explain the biphasic activation dynamics. Such a mechanism could explain the observed induction dynamics averaged over the whole cell population, but an alternative multi-step activation mechanism could not be ruled out [12]. Confronting the heterogeneous single-cell trajectories with mechanistic stochastic models could, thus, not only help to understand the heterogeneous response to a homogenous stimulus, but also discriminate between different activation mechanisms. Stochastic models of gene activation could in this way help to make fuller use of the information contained in the single-cell data compared to only considering averaged time courses. Here, a framework is presented to achieve these purposes. It comprises two competing mechanistic induction models, implementation of a stochastic simulation approach and a fitting procedure to confront the models with real or simulated data.

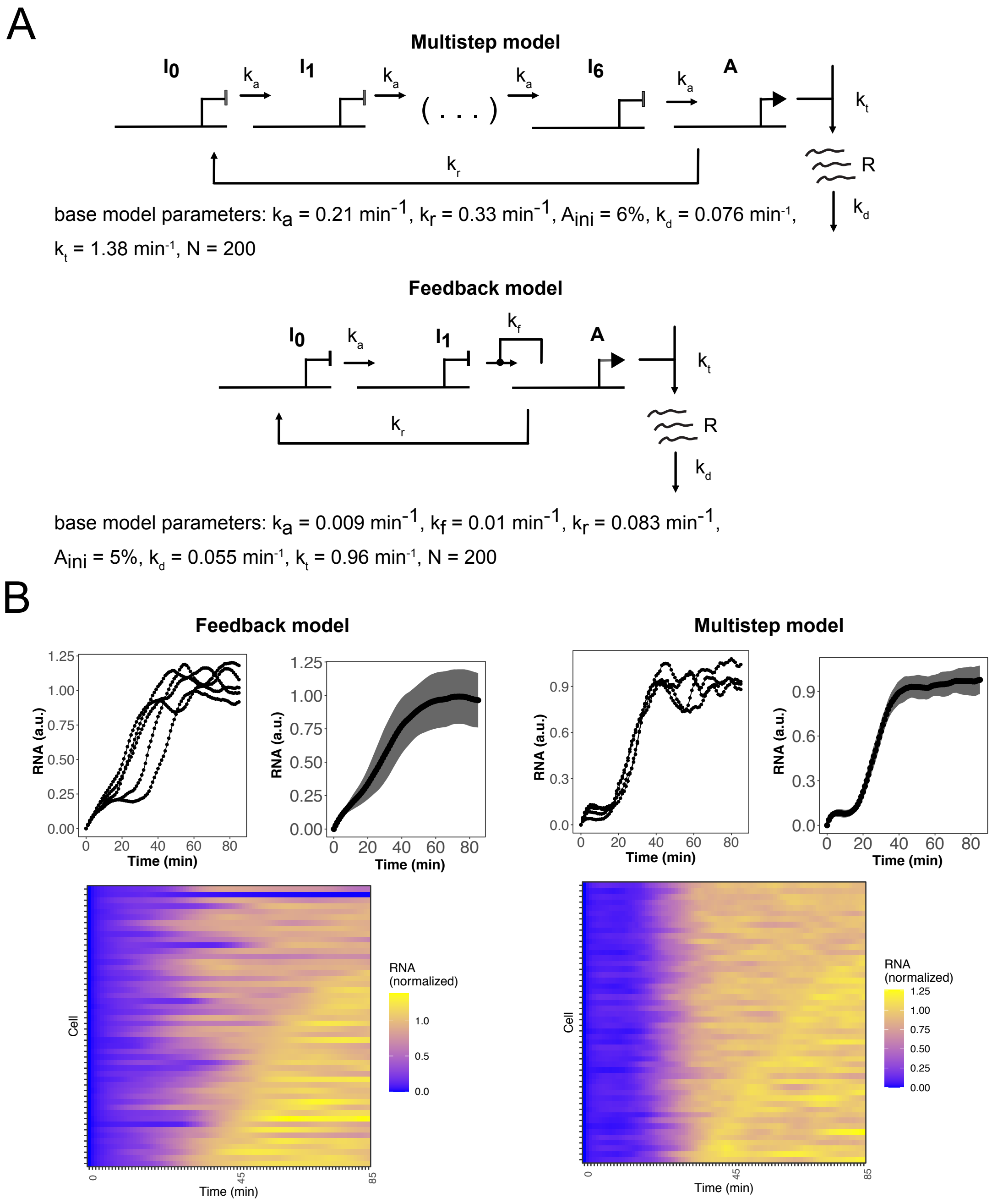


Figure 2.9: Stochastic models of promoter state transitions. **(A)** Two alternative models of promoter state transitions in transcriptional induction. Promoters can be in one of several inactive states *I* or in an active state *A* from which RNA *R* can be produced. In the multistep model promoters are activated by a sequence of seven inactive steps, whereas in the feedback model active promoters interact with inactive promoters in state I_1 to activate them in a positive feedback reaction. The respective base models were simulated with the indicated parameter sets. (Continued on next page)

While the current framework that is presented here neglects important factors like copy number variations and extrinsic and technical sources of noise and cannot be used to draw definite biological conclusions, it demonstrates that the combination of single-cell time courses and stochastic modeling can, in principle, distinguish between fundamentally different activation mechanisms. Models of transcriptional bursting assume that the promoter can be in different states that are either *on* or *off* meaning that they are capable to recruit, initiate and let RNA PolII escape from the promoter or not. Cyclic transitions between several inactive and one active state have been used to model bursting using three [48] or five [159] steps. Here, two cyclic multistep models were compared, one with seven states and one with three states and positive feedback (**Figure 2.9 A**). They are named Multistep and Feedback model. The Multistep model has a large number of activation steps to give it the flexibility to recapitulate the single-cell data that in some cases shows prolonged delays before RNA is produced. The purpose of the model comparison is to decide, if promoter-promoter interactions, that transfer activation marks in the Feedback model (and could thereby induce collective bursting of the gene cluster), are necessary to explain the heterogeneous time course data. The models start in a configuration in which the majority of the promoter population resides in the inactive state I_0 and a small fraction A_{ini} in the active state A. RNA is produced only from state A. TFs are assumed to be recruited instantaneously to the promoters at the start of simulated time courses thereby allowing all promoter state transitions to take place. For both models a parameter set was determined that leads to average time courses in rough agreement with the biphasic dynamics of the data (the base parameter set, **Figure 2.9 A, B**).

Gillespie simulations [181] for the two base models revealed that the Feedback model leads to a large variability in activation dynamics, as can be seen for individual trajectories and in the full set of simulated single-cell time courses represented by a color map (**Figure 2.9 B**).

Figure 2.9: (Continued) **(B)** Simulations of the models described in A. Representative single-cell simulations, mean trajectories ± one s.d. and complete sets of color coded single-cell simulations are shown. RNA values were normalized to the mean plateau values.

The Multistep model exhibited some fluctuations in the second plateau, but the transition from the first to the second plateau was well synchronized between simulated cells. The average time courses for both models showed small differences in the first plateau, but were overall similar. This illustrates the additional information contained in single cell as opposed to bulk data sets.

In order to investigate, whether the higher variability of time course dynamics of the Feedback model where specific for the chosen set of base parameter values or a general feature of this model the parameter values were successively changed (**Figure 2.10**). This also represents a crude but simple approach to check, if the model is sensitive to certain parameters. For each modified model simulation the average trajectory, the standard deviation and the autocorrelation matrix of time courses are shown. The autocorrelation matrices were obtained by calculating the Pearson correlation coefficient for RNA levels at each pair of time points across the population of simulated single-cell time courses. They can indicate, if stochastic events at a time point have a strong effect on the RNA levels at later time points. Importantly, they can be only calculated from single (live) cell trajectories. While there was strong long-range correlation of RNA levels in the Feedback model after around 15 min, in the Multistep model correlations were mainly between consecutive time points indicating that variations in RNA levels did not have long lasting effects for this mechanism. The Multistep model was not sensitive to the change of most parameters regarding the average, distribution width and correlation of RNA time courses. Only a strong increase of the initially active fraction of promoters could affect the dynamics producing an overshoot of RNA production early in the time course. In contrast, the induction dynamics showed strong changes for the Feedback model, when the the transition rate constants between promoter states, the fraction of initially active promoters or the total number of promoters was changed. The total number of promoters should (and did) not have any effect for the Multistep model except for the RNA trajectory amplitude, since there is no interaction of promoters.

Conversely, for the Feedback model the promoter-promoter interactions make this mechanism sensitive to this quantity. If the gene copy number and the RNA dynamics could be directly measured in the same experiment, a relation between these features would be an indication for a promoter-promoter activation mechanism. However, the difficulty to acquire three color time courses of RNA, TF and a promoter label like EGFP-

LacI has prevented this type of experiment.

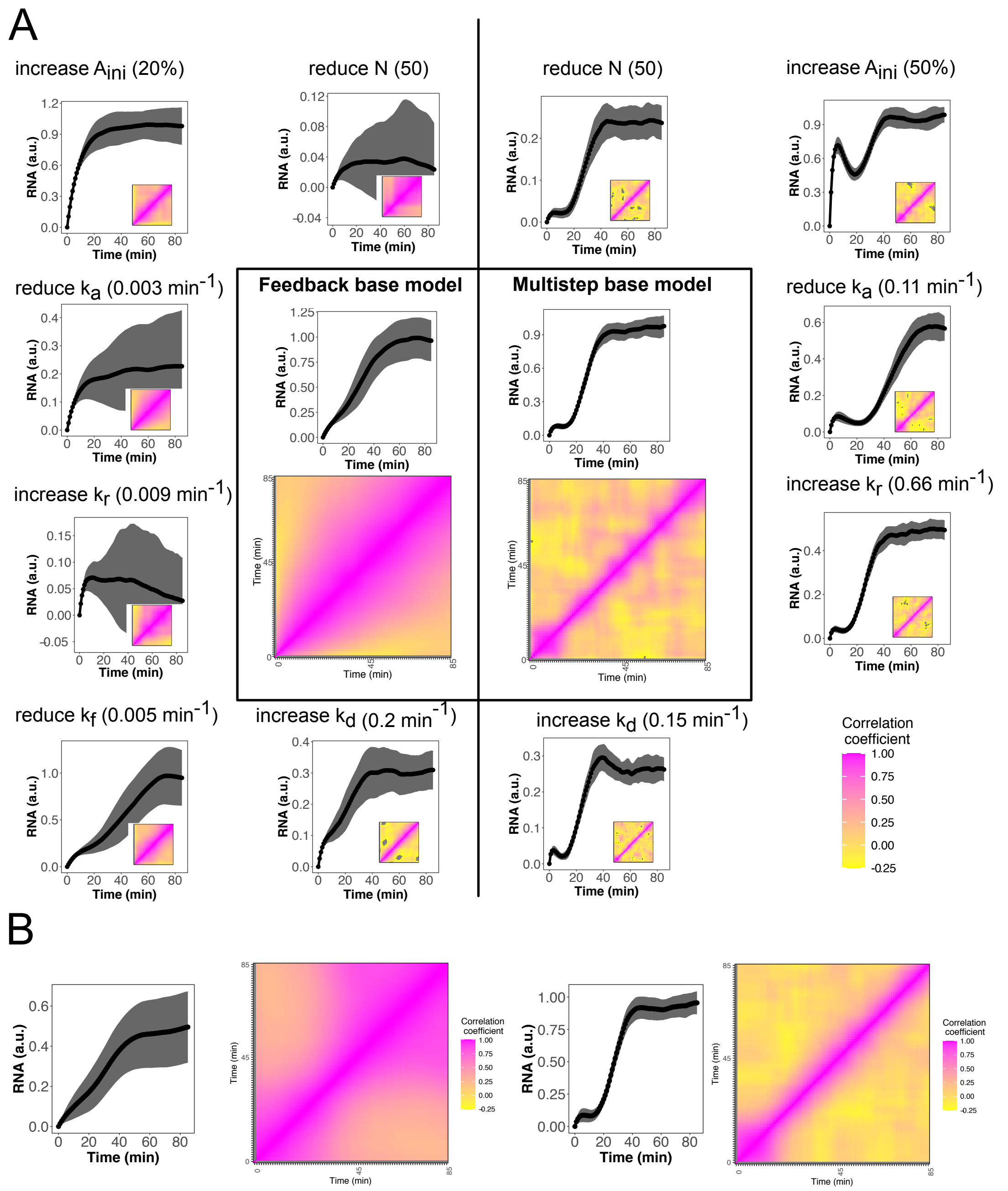


Figure 2.10: Effect of varying the parameter values of the Multistep and Feedback model. **(A)** Mean RNA trajectories ± one s.d. and correlation matrices of the Feedback (left) and Multistep model (right). Model simulation with the base parameter sets are shown at the center. Each condition represents 50 simulations. **(B)** Effect of heterogeneity in the fraction of initially active promoters for the Feedback (left) and Multistep model (right). The fraction of initially active promoters was drawn from a uniform distribution between 0 and 10% (multistep: 0 and 12%). 100 simulations for each condition.

For most parameter variations of the Feedback model the long-range correlation of RNA levels at later time points persisted, except for an increase of the termination rate k_d with which nascent RNA leaves the promoter cluster. This could mean that long dwell times of the nascent RNA are responsible for extended periods of high RNA concentration at the cluster leading to high correlation.

Biologically, it is reasonable to assume that the initially active fraction of promoters will be variable between cells. This fraction may be constituted of promoters that are localized in a activation prone chromatin environment [182], e.g. at the outer regions of the cluster, or of promoters that have more activating chromatin marks. For instance, the activating acetylation of histones shows a high turnover [183] and variable activities of histone acetyl transferases (HATs) and de-acetylases (HDACs) between cells could shift this equilibrium. Using a uniform distribution of initially active promoter fractions for single-cell simulations led to a broader distribution of time courses and additional correlations between RNA levels at the initial time points for the Feedback model. It did not have a strong effect on the Multistep model compared to using a fixed value of A_{ini} (**Figure 2.10 B**). Since a variable fraction of initially active promoters did not appear to give the Multistep model the ability to produce more heterogeneous or correlated time courses, for the further analysis a fixed value was used.

For the confrontation of models with experimental data, but also to explore the abilities of models to recapitulate simulated dynamics, an approach for model fitting is needed. This can be done by fitting the moments of the distribution of the observable at the different time points. This requires deriving expressions for these moments from the chemical master equation [184]. Here, a simple and more direct approach was chosen that is, however, computationally expensive, which will probably limit its use to the exploration of model properties. It consisted of confronting the statistics of single-cell simulations (average, cumulative distribution function, correlation matrix) directly with (simulated) data and minimizing the sum of squared residuals by varying model parameters according to a particle swarm algorithm. The time evolution of promoter states was simulated repeatedly using Gillespie simulations for each parameter set resulting in populations of around 50 time traces (for plotting, 20 for optimization), from which distributions and correlations could be calculated. The simulations and fitting could be done in this way because the number of reactions, states and parameters and, most importantly, the number of simulated molecules (promoters) was sufficiently small.

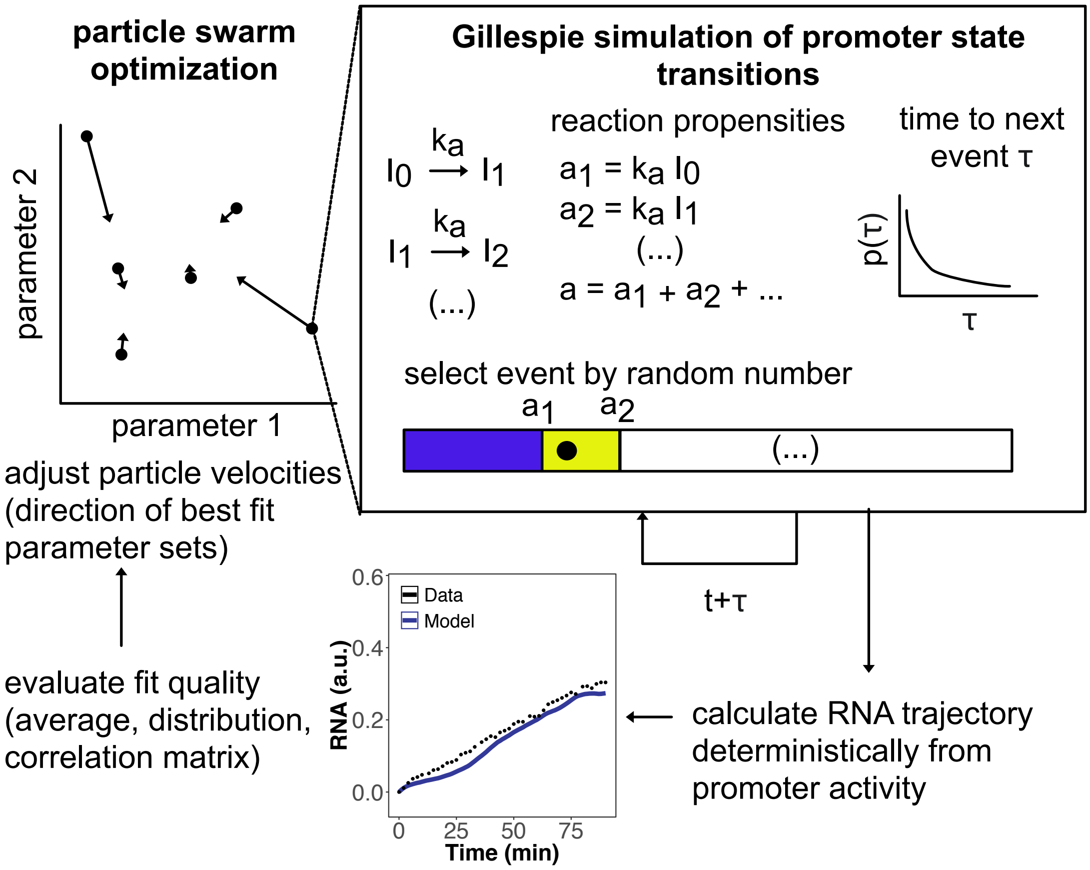
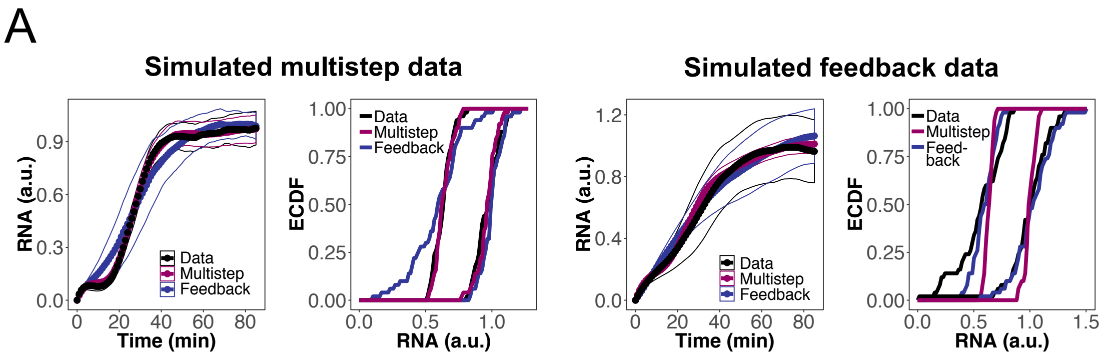


Figure 2.11: Fitting of stochastic models by particle swarm algorithm. The parameter sets ("particles") are iteratively changed and used for simulations. The velocity of each particle is adjusted in each optimization cycle so that it points towards the global and the particle's own best fit. Gillespie simulations for each parameter set comprise the calculation of reaction propensities from the rate constants and molecule numbers, the drawing of a random exponentially distributed waiting time τ and the random selection of a reaction according to the propensities. While promoter transitions are simulated stochastically the RNA trajectories are calculated deterministically. The goodness of fit is evaluated from the squared residuals of the RNA mean trajectory, distribution at each time point and correlation matrix.

To further reduce the computational cost, only the promoter state transitions were simulated stochastically and the resulting RNA time courses were calculated deterministically using the active promoter state trajectories and reaction rates for transcription and RNA dissociation (see Materials and Methods for details). Fitting and parameter estimation was achieved by the particle swarm optimization approach [185]. The implementation followed the approach described in [186]. The full fitting procedure is depicted in **Figure 2.11**. Parameter sets (particles) were used for Gillespie simulations of promoter state trajectories. The RNA trajectories were calculated from the active state trajectory optimizing the transcription rate constants to achieve a good fit to the (average) data time courses. Then the overall fit quality was assessed using the squared residuals of the average time course, of the empirical cumulative distribution function and of the autocorrelation matrix. This definition of the objective function was chosen in order to reflect both the bulk behavior of the population and the heterogeneity of single-cell trajectories. The particles were then moved in the direction of the best fit. A total of 30 particles was used in 30 optimization cycles. While this procedure was still feasible with calculation times of several hours, the method could become limited by computational power for larger models.

The fitting procedure was used to investigate, whether each of the two induction models had features that could not be recapitulated by the other model. Simulations of each model were used as data sets and fitted by both models. As expected, each data set could be fitted best by the model from which it was simulated (**Figure 2.12 A**). This could be mainly attributed to the broader distribution of RNA levels for the Feedback model. The Multistep model could not achieve this higher inter-cellular heterogeneity when fitting Feedback model data, whereas the Feedback model could not produce a more narrow distribution. The correlation matrices of the best fit simulations showed that the presence or absence of long-range correlation was an inherent property of the model that could not be adjusted to better fit the simulated data of the competing model (**Figure 2.12 B**). While the discrimination between Multistep and Feedback model is not yet possible without controlling for external sources of heterogeneity like gene copy number or readout noise, it can still be informative to fit both models to experimental data, to check, whether characteristic features of these models are present in the data. A data set of 64 cells activated by PHR-EGFP-VP16 recruited to CIBN-rTetR at the reporter gene promoter over 90 min was used to fit both models. The data set is discussed in detail in section 2.5. While both models could recapitulate the average RNA trajectory with some systematic deviations between 20 min and 40 min, only the Feedback model could achieve a comparable spread of the distribution (**Figure 2.13 A**). This was also the case for the long-range correlations (**Figure 2.13 B**). However, the differences of copy numbers between cells that affect time course amplitudes was not corrected for, so that the variability of signal strength between cells could be mainly responsible for the high correlation.



data parameters: $k_a = 0.21\ \text{min}^{-1}$, $k_r = 0.33\ \text{min}^{-1}$, $A_{ini} = 6\%$

fit parameters: $k_a = 0.23\ \text{min}^{-1}$, $k_r = 0.19\ \text{min}^{-1}$, $A_{ini} = 7\%$

data parameters: $k_a = 0.009\ \text{min}^{-1}$, $k_f = 0.01\ \text{min}^{-1}$, $k_r = 0.083\ \text{min}^{-1}$, $A_{ini} = 5\%$

fit parameters: $k_a = 0.004\ \text{min}^{-1}$, $k_f = 0.08\ \text{min}^{-1}$, $k_r = 0.039\ \text{min}^{-1}$, $A_{ini} = 4\%$

		data	
	objective function	**multistep**	**feedback**
model	**multistep**	203	206
	feedback	545	120

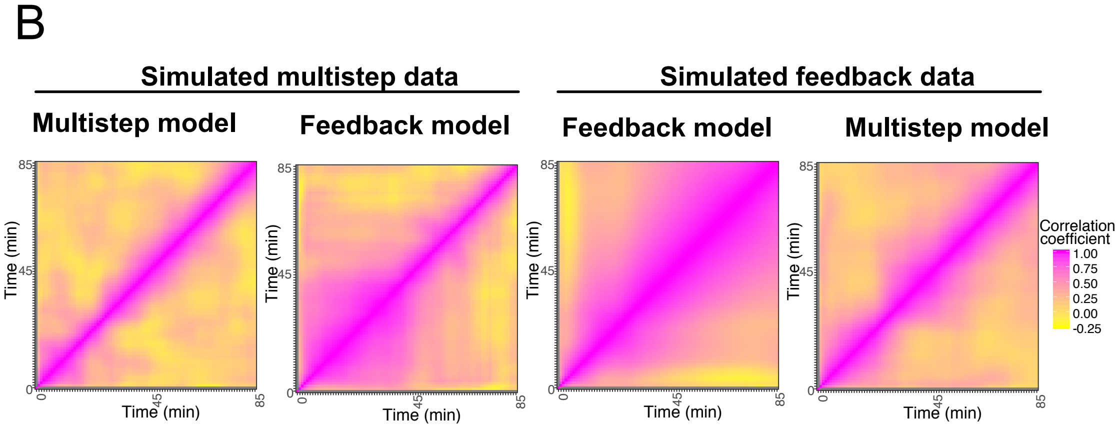
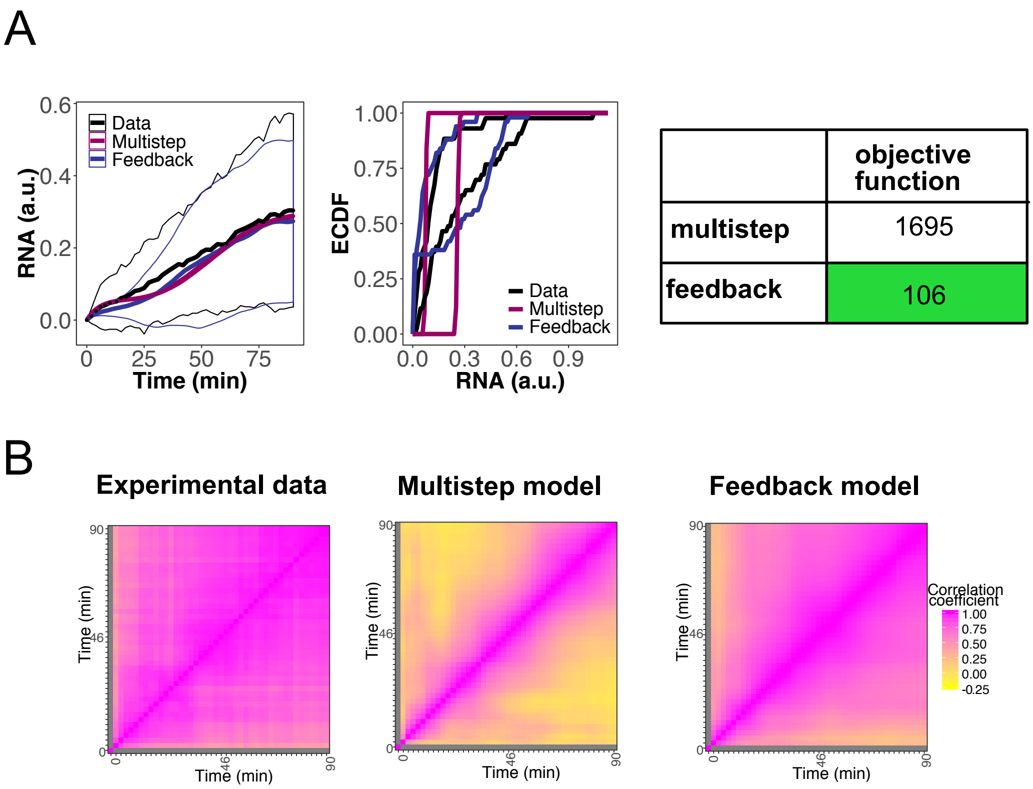


Figure 2.12: Feasibility of model discrimination demonstrated for simulated data. Sets of RNA trajectories simulated by the Feedback and the Multistep model were fitted by both models. **(A)** Mean RNA time courses and RNA distributions for time points 30 min and 75 min represented as empiricial cumulative distribution function (ECDF) for simulated data and the fits to it (left: simulated data from multistep model, right: feedback model). Parameters used for simulating the data and the estimated values by the best fit (Multistep for Multistep and Feedback for Feedback). (Continued on next page)

In conclusion, while no definite model discrimination could be achieved, the analysis showed that a sequential multistep activation mechanism and a promoter cross-talk mechanism lead to distinguishable features of the activation dynamics. These features comprise the distribution of nascent RNA levels and the (long-range) correlation of RNA levels between time points. There are two major challenges for a more reliable model discrimination: First, the variable copy number needs to be accounted for, ideally by measuring it together with the RNA time courses. Then the copy number could be integrated into the model. While the current experimental setup is limited to two colors, on a different microscope such three color experiments including a gene array marker should be technically feasible. Alternatively, the trajectories could be normalized to their maximum values. Second, for a fair comparison of models it must be ensured that the fitting procedure finds the best parameter set for each model.

While the approach of Gillespie simulations combined with particle swarm optimization has proven that it can successfully fit models to their own simulated data, it is limited due to the high computational cost. The simulation of only 20 trajectories per parameter set might be insufficient to identify a good parameter set reproducibly. Moreover, to distinguish between a mechanism with or without activation feedback, it would be useful to compare families of models to verify that the feedback mechanism and not a detail of model topology is responsible for the ability or failure to reproduce the data. For these purposes a different, less computational expensive approach like the calculation of distribution moments from the chemical master equation could be chosen. The results presented here indicate, that an advancement of the experimental and modeling approaches could be worthwhile and ultimately give insights into the mechanisms of transcriptional induction and into the sources of inter-cellular heterogeneity of activation dynamics.

Figure 2.12: (Continued) The sum of squared residuals of mean trajectory, ECDF and correlation matrix are given in the table. **(B)** Correlation matrices of the fitted models. Note that the multistep model seems to be unable to produce long-range correlations.



	objective function
multistep	1695
feedback	🟩 106

Figure 2.13: Confronting stochatic models with VP16 induction data. The data corresponds to the experiments described in detail in Fig. 2.32 for induction by PHR-EGFP-VP16. **(A)** Mean RNA time courses and RNA distributions for time points 30 min and 75 min represented as empiricial cumulative distribution function (ECDF) for data and fits. **(B)** Correlation matrices of data and the fitted models.

2.2 Measuring TF binding kinetics by widefield FRAP at a cluster of binding sites

The binding of TFs to their target sites is expected to quantitatively regulate transcription both by the equilibrium binding levels [69] and by the kinetics of the binding process [77, 80]. Moreover, multivalent interactions of the AD and the potentially resulting phase separation will affect the binding properties. Comparing binding kinetics could thereby help to detect multivalent interactions and to understand differences in the activation potential of TFs. Besides single molecule tracking (SMT) [103, 187], fluorescence recovery after photobleaching (FRAP) is a method to determine the binding kinetics of a fluorescently labeled TF. Typically, FRAP is measured on a laser scanning confocal microscope due to its ability to move the laser beam to the region of interest and bleach it. However, the confocal setup itself is not necessary for FRAP experiments. Any fluorescence microscope equipped with a laser scanner and a sufficiently strong laser source could be used. Here, a method is described to perform FRAP experiments on a widefield microscope equipped with an additional laser micromanipulation device. While having a lower axial resolution than a confocal microscope, this approach was expected to have several advantages:

- Lower sensitivity to movements of the bleached object in z-direction.
- Lower amount of photodamage.
- Fast acquisition of large image areas limited only by the minimum camera exposure and readout times. (Compared to potentially time consuming scanning of the image.)
- Retrofitting of a laser micromanipulator to a widefield microscope represents an economic alternative to acquiring a confocal setup.

As a result of the first two points recovery time courses can be acquired for longer durations without the need for a drift stabilization device. The possibility to acquire large image areas proved to be crucial for measuring large numbers of cells because suitable cells could be easily found without the time consuming moving of the scanning region and without switching between magnification factors. In summary, widefield FRAP though lower in resolution promises to be a simple, robust, fast and economic alternative to confocal FRAP. Even though a commercial solution for widefield FRAP is available [188], there are very few reports of using this technique and these are limited to studies of diffusion only [189, 190].

FRAP measurements of ligands binding at a cluster of *lacO* and *tetO* binding sites represented a special case for the modeling of the recovery process that required using numerical simulations of an adapted model of fluorescence recovery. Due to the large number of synthetic TFs and the need for sufficiently large numbers of cells measured for robust statistics the FRAP image analysis and fitting was partly automated. The necessary functions were collected in a package (*NSQFRAP*) in the R programming language. Technical details for FRAP experiments, image analysis and parameter estimation are also described in **sections 4.2.9, 4.2.10 and 4.2.11** of the **Materials and Methods** chapter.

2.2.1 Simple and robust widefield FRAP experiments were implemented

For measuring their binding kinetics the TFs were recruited to the *lacO* sites in the U2OS 2-6-3 cell line using LacI or dCas9 with the lacO-2xPP7 guide RNA. This was the preferred recruitment strategy due to the higher copy number of the operator sites and the resulting stronger fluorescence signal compared to *tetO*. Some complexes could only be assembled at *tetO* sites. They were based on the rTetR domain, or needed to have exactly the same composition as the complexes used for the transcriptional activation, which only effectively works from the more promoter proximal *tetO* sites. All measured constructs were labeled with EGFP. Light-dependent complexes were illuminated by imaging for at least 1 min with blue excitation light. Two types of FRAP measurements were performed: off-spot at a location outside of the binding site cluster to determine the effective diffusion coefficient D_{eff} and on-spot by placing the photobleaching laser beam at a central position of the gene cluster spot. Recovery in on-spot FRAP experiments depends on the diffusion of unbleached molecules into the bleach region and on the exchange of bleached bound molecules with unbleached molecules (**Figure 2.14**). Off-spot measurements had a duration of 60 s and 300 ms intervals, while on-spot measurements took 240 s in 1 s intervals. The FRAP experiments were conducted on a widefield microscope with a 70 mW, 473 nm laser micromanipulation device that was controlled independently of the image acquisition software. The laser beam position was calibrated to positions in the acquired images by aligning screen coordinates of the imaging software to coordinates in the software that controls the laser scanner using the accompanying routines. FRAP image time courses were acquired by selecting a bleach position, starting the image time series acquisition and manually triggering the laser pulse after few seconds of acquisition. As a consequence of this setup the number of pre-bleach frames can vary (typically between three and five) and the image acquisition time points are not synchronized with the start and end of the laser pulse. This leads to a variable shift between the end of bleaching and the first imaging time point that can maximally amount to the imaging interval (1 s for on-spot, 300 ms for off-spot). A low magnification (20x) objective was chosen for imaging resulting in a large field of view that allowed to quickly select the next cell for measurement, potentially in the same field of view. The relatively low excitation light intensity prevented strong acquisition bleaching of fluorophores and phototoxicity. Each single-cell measurement resulted in an image series with 200 (off-spot) or 240 (on-spot) time frames (**Figure 2.15, top**). The acquisition of FRAP series in hundreds of cells with this number of time frames made it impossible to analyze the data manually. Consequently the analysis needed to be automated.

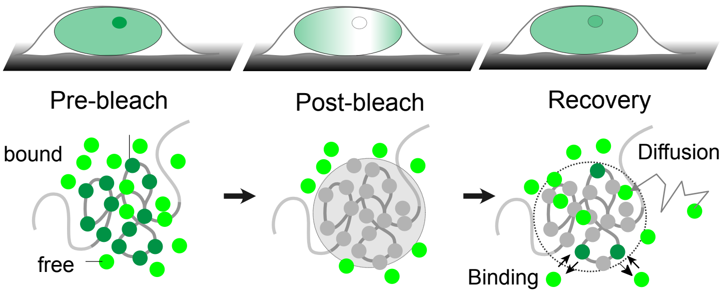


Figure 2.14: Principle of FRAP at a binding site cluster. Recovery happens by diffusion of unbleached molecules into the bleach region and by the exchange of bound bleached molecules with unbleached molecules.

2.2.2 The *NSQFRAP* software package provides an automated workflow for image analysis and data fitting

FRAP analysis consisted of three main steps: (1) segmentation and tracking of cell nuclei and bleach region, (2) intensity quantification and normalization and (3) parameter estimation from the resulting recovery curve (**Figure 2.15**). These steps were largely automated in the R programming language [166] and functions were collected in the software package *NSQFRAP* (*Nuclear Spot Quantification for FRAP*). It builds on the image processing capacities of the *EBImage* package [167] and on methods of numerical simulations for systems of partial differential equations of the *ReacTran* package [191]. **Table 2.2** summarizes the functions available in the package.

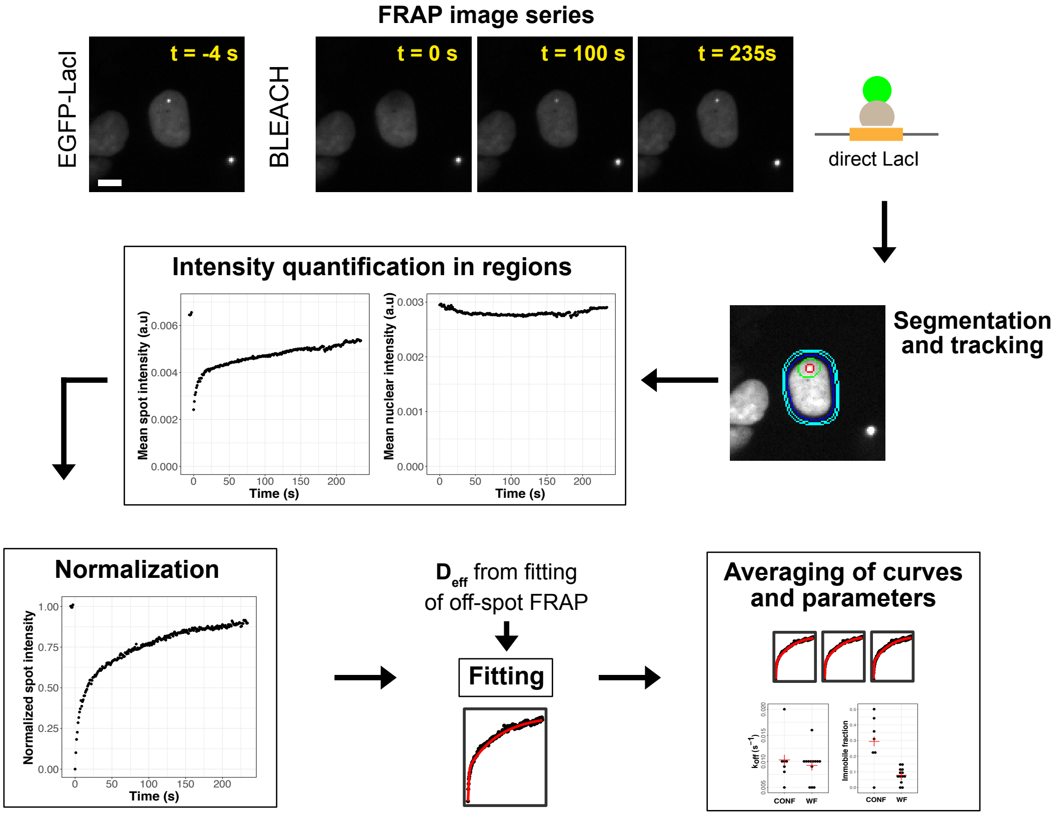


Figure 2.15: FRAP analysis workflow. Image series (here for EGFP-LacI) consisted of 240 time frames. Nuclei (blue), local nuclear background regions (cyan), gene cluster spots (red) and local spot background (green) were segmented and tracked by an automated, supervised image processing pipeline. Mean intensities were measured in all regions. The recovery curve was calculated by normalizing the spot intensity time course. Single-cell recovery curves were used for parameter estimation using a reaction-diffusion model. For on-spot fits the effective diffusion coefficient D_{eff} determined from off-spot measurements was used. For each synthetic TF construct single-cell trajectories, their fits and the estimated binding parameters were, finally, averaged.

The analysis pipeline is split into four parts. Exemplary scripts for each part are provided with the *NSQFRAP* package. Automated image analysis is followed by a manual curation step, in which tracking errors can be corrected. The resulting recovery curves are then used for automated parameter estimation either by a diffusion only model for off-spot FRAP, or by a reaction-diffusion model for on-spot FRAP. The first step of image processing requires no user interaction. The position of the measured cell and the time point of the bleach are determined from the very high intensity spot of the laser beam (which in contrast to confocal FRAP is visible in the acquired widefield images during the bleach). Images are cropped around the nucleus position and the nucleus, spot (for on-spot FRAP) and background regions are segmented. The bleach region is segmented in an image created from the difference of pre-bleach and first post-bleach images. The nucleus is tracked along the time course by a similiar approach described above for the *NSSQ* package. Since the bleaching leads to a transient disappearance of part of the nuclear intensity and of the spot additional measures need to be taken for object tracking. Apparent nuclear movements caused by a sudden decrease of nuclear area due to bleaching are excluded. The spot is not tracked directly, but the spot mask created from pre-bleach frames is moved according to the calculated nuclear movement. This does not always work perfectly, since the nucleus is not a rigid body, but can deform allowing an internal movement of the spot position. To account for this a manual curation step after segmentation and tracking allows to annotate cells as usable or not and to correct the spot position, if needed. Average intensities are then determined in the areas of interest as well as radial intensity profiles around the center of the bleach region. The profile in the first post-bleach frame represents the bleach profile that is used as a starting condition for model simulations.

Table 2.2: *NSQFRAP* functions

Function	Application
ApplyShift	Shift mask
CorrectTracking	Correct mask positions
dif2Dpolar	Simulate single diffusion step
difbind2Dpolar	Simulate single binding-diffusion step
fitSimDifBindSpotRatioImmo	Simulate binding and diffusion
fitSimMeanImmob	Simulate diffusion
getBleachCoordinates	Determine bleach position
getBleachFrames	Determine (pre-)bleach frames

gridFit	Parameter estimation on grid
IntInMask	Measure intensities
intProfile	Measure radial profile
makeBgMask	Create background mask
measureProfile	Measure radial profile
MoveParallelToNuc	Shift mask following nucleus movements
Multistart_nls	Fit using multiple start values
NSQFRAP	FRAP image analysis and fitting
paintOutlines	Visualize segmentation
ProfileToMean	Average intensity from radial profile
profile_func	Create bleach profile
segmentBleachRegion	Segment bleach region
segmentNucleus_series	Segment nucleus
segmentSpot	Segment spot

The raw mean intensity time courses were normalized to obtain recovery curves. Intensities were normalized by subtracting background values and dividing by the mean nuclear intensity to account for bulk photo-bleaching (the reduction of total nuclear intensity due to the bleach). Normalizing to pre-bleach levels ensured that a complete recovery corresponded to a plateau value of 1. For intensity profiles $I(r,t)$ the normalization led to normalized intensity profiles $I_{norm}(r,t)$:

$$\begin{aligned} I_1(r,t) &= I(r,t) - I_{nuc\_bg}(t) \\ I_{nuc\_norm}(t) &= I_{nuc}(t) - I_{nuc\_bg}(t) \\ I_2(r,t) &= \frac{I_1(r,t)}{I_{nuc\_norm}(t)} \\ I_3(r,t) &= I_2(r,t) - I_2(r = r_{center}, t = 0) \\ I_{norm}(r,t) &= \frac{I_3(r,t)}{mean(I_3(r,t<0))} \end{aligned} \tag{2.7}$$

The recovery curves for off-spot experiments were calculated by integrating the intensity profile from its center up to a radius of 3.5 µm. The innermost pixel was excluded due to its high noise levels. For on-spot recovery curves the average intensities in the spot area were normalized directly:

$$\begin{aligned}
I_{spot,1}(t) &= I_{spot}(t) - I_{nuc\_bg}(t) \\
I_{nuc\_norm}(t) &= I_{nuc}(t) - I_{nuc\_bg}(t) \\
I_{spot,2}(t) &= \frac{I_{spot,1}(r,t)}{I_{nuc\_norm}(t)} \\
I_{spot,3}(t) &= I_{spot,2}(t) - min(I_{spot,2}(t)) \\
I_{spot\_norm}(t) &= \frac{I_{spot,3}(t)}{mean(I_{spot,3}(t<0))}
\end{aligned} \tag{2.8}$$

The resulting recovery curves were used to first estimate diffusion coefficients from off-spot FRAP experiments for each construct and then binding parameters from on-spot experiments incorporating the estimated diffusion coefficients.

The estimation of binding and diffusion parameters from FRAP recovery curves is known to be highly dependent on the model fitted to the data [192]. While in certain cases the recovery can be described by a diffusion-dominant, an effective diffusion or a binding-dominant model [151], generally, the model should account for both diffusion and binding. In the case of the TF binding to a cluster of binding sites instead of evenly distributed sites the location of binding sites needs to be incorporated into the model. Otherwise, estimated parameters could deviate from their true values by an order of magnitude [152]. Analytical solutions to such a model have been derived [152], but the assumption of a bleach region of the same size as the bleached binding site cluster did not hold true for the FRAP experiments performed here. The bleach region was larger than the binding site spot and the (short, but variable) interval between bleaching and first post-bleach image acquisition led to a widening of the bleach profile. Consequently, numerical simulations were chosen as an alternative approach.

The model setup described here follows the description given by Sprague and colleagues [152]. It consist of a cluster of binding sites at the center of a cell nucleus. Both are assumed to be cylindrical with a spot radius r_s and a nuclear radius r_{nuc}. Binding sites are assumed to be evenly distributed in the cluster spot.

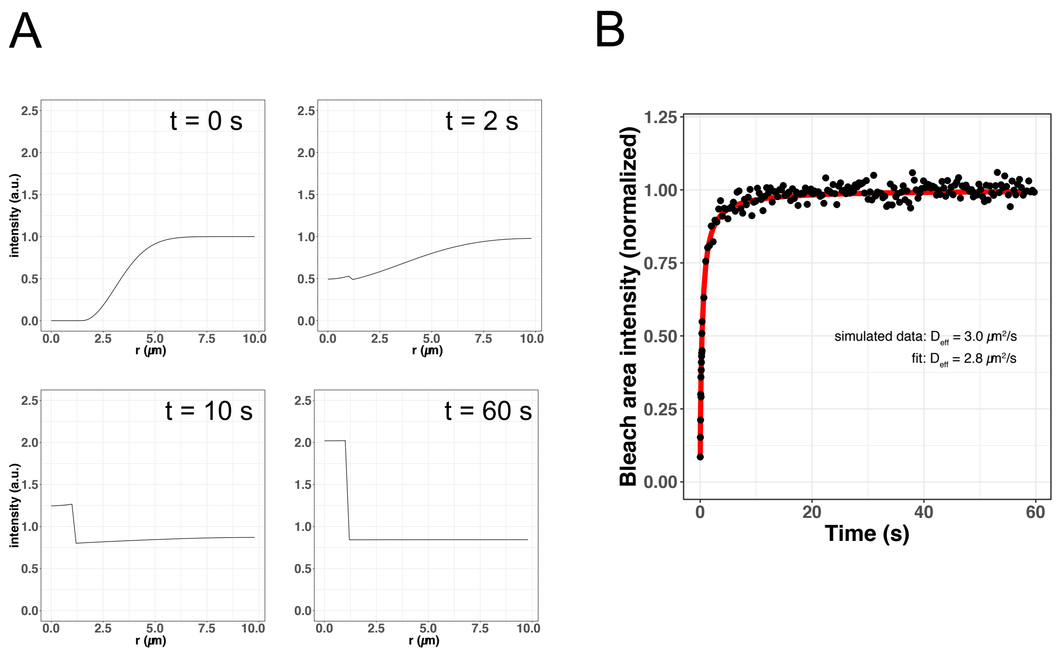


Figure 2.16: Simulation and parameter estimation in a reaction-diffusion model of FRAP at a cluster of binding sites **(A)** Simulated intensity profiles at different time points after onset of recovery. Parameter values in the simulation were: $r_{spot} = 1.0\,\mu m$ $D_{eff} = 2.0\,\frac{\mu m^2}{s}$, $k_{on} = 1.0\,s^{-1}$, $k_{off} = 0.07\,s^{-1}$. **(B)** Validation of simulation approach. Simulated data (black dots) of a diffusion-only recovery (off-spot) and a fit (red line) using the Bessel function as an analytical solution of the diffusion only model are shown. The estimated parameter D_{eff} showed good agreement with the value used for the simulation.

It has been shown that the simplifying assumption of binding sites extending along the whole z-axis does not strongly distort the estimated parameters [152]. In the model there are freely diffusing particle f, free binding sites b and complexes of binding sites occupied by a particle c. Particles diffusion is governed by an effective diffusion coefficient D_{eff} that also accounts for short non-specific binding events that could be for instance caused by non-specific interactions of TFs with chromatin [71]. Binding of free particles to free binding sites occurs with a rate constant k_{on} and dissociation with rate constant k_{off}. Since FRAP happens in equilibrium and only changes the visibility of particles, the concentration of free binding sites is constant and can be incorporated in the on-rate to yield a pseudo rate constant k^*_{on}. The spatio-temporal evolution of this reaction-diffusion model is described by the following set of partial differential equations in polar coordinates:

$$
\begin{aligned}
&\text{for} \quad r \leq r_s : \\
&\frac{\partial f(r,t)}{\partial t} = D_{eff} \cdot \nabla_r^2 f(r,t) - k_{on}^* \cdot f(r,t) + k_{off} \cdot c(r,t) \\
&\frac{\partial c(r,t)}{\partial t} = k_{on}^* \cdot f(r,t) - k_{off} \cdot c(r,t) \\
&\text{for} \quad r > r_s : \\
&\frac{\partial f(r,t)}{\partial t} = D_{eff} \cdot \nabla_r^2 f(r,t) \\
&c(r,t) = 0
\end{aligned}
\tag{2.9}
$$

For the off-spot case the PDE system simplifies, because no binding sites are available:

$$
\frac{\partial f(r,t)}{\partial t} = D_{eff} \cdot \nabla_r^2 f(r,t) \tag{2.10}
$$

The initial conditions are chosen as a gaussian profile with a central plateau with radius r_p:

$$
\begin{aligned}
f(r \leq r_p, t = 0) &= 0 \\
f(r > r_p, t = 0) &= A \cdot (1 - e^{\frac{-(r-r_p)^2}{\sigma}}) \\
c(r,t) &= 0
\end{aligned}
\tag{2.11}
$$

This profile accounts for recovery by diffusion during the bleach and during the interval between bleach and first post-bleach frame, while the (slower) binding turnover is expected to be negligible in these initial 1 s to 2 s.

The time evolution of this model was simulated numerically. The *NSQFRAP* package offers an easy-to-use interface for these simulations drawing on the finite difference method for numerical simulation of PDEs of the *ReacTran* package [191]. In this approach the continuous space in the nucleus is split into concentric grid cells assuming rotational symmetry. **Figure 2.16 A** shows simulations of the radial intensity profile as it recovers after the end of the bleach pulse at time point $t = 0\ s$. Simulated profile time courses are converted to recovery curves by integrating the profiles up to the radius of the spot (for on-spot simulations) or of the bleach area (for off-spot simulations) and applying the same normalization procedures as for the microscopy data. To confirm that the simulations produce realistic recovery time courses, a simulated recovery curve for the off-spot case was fitted by an analytical solution of a diffusion-only system [151]:

$$
\begin{aligned}
f(t) &= e^{\frac{-\tau_D}{2t}} \cdot (I_0(\frac{-\tau_d}{2t}) + I_1(\frac{-\tau_d}{2t})) \\
\tau_D &= \frac{r_{bleach}^2}{D}
\end{aligned}
\tag{2.12}
$$

where I_0 and I_1 are modified Bessel functions. The good agreement of the estimated diffusion coefficient ($D = 2.8\frac{\mu m^2}{s}$) with the value used for the simulation ($D = 3.0\frac{\mu m^2}{s}$) indicated that the simulation approach yields meaningful results (**Figure 2.16 B**).

Three parameters of the model were of major interest: the effective diffusion coefficient (D_{eff}), the dissociation rate constant k_{off} and the immobile fraction f_i. D_{eff} is mainly needed to account for diffusion in the full reaction-diffusion model, but the binding parameters are informative for the biological questions of residence time dependence of transcription and the role of multivalent interactions of the AD. Parameters were estimated by different fitting strategies for off- and on-spot fits. Off-spot recovery curves were fitted with the diffusion-only model using a local optimization method with multiple start values. The on-spot recovery curves were fitted with the full reaction-diffusion model using a two-step approach. In the first step, the immobile fraction was estimated by fitting a phenomenological function to a part of the curve, in which recovery by diffusion was mostly complete:

$$I(t|t > 30\ s) = A + B \cdot (1 - e^{-k \cdot t}) \tag{2.13}$$

In the second step the full model could be fitted, but the two parameters D_{eff} and f_i were fixed reducing the number of parameters that needed to be optimized. The equilibrium ratio of bound and unbound molecules is related to the rate constants and was used to express the value for k^*_{on}:

$$k^*_{on} = \frac{c}{f} \cdot k_{off} = spotRatio \cdot k_{off} \tag{2.14}$$

The parameter $spotRatio$ represents the ratio of bound and unbound molecules and an approximated value can be determined from the (equilibrium) fluorescence intensities in and outside of the spot before bleaching.

$$spotRatio = \frac{median(I_{spot}(t < 0) - I_{spot\_bg}(t < 0)}{median(I_{spot\_bg}(t < 0)} \tag{2.15}$$

This only represents a rough approximation and a start for further refinement by multiplying $spotRatio$ with a free fit parameter $bgRatio$. Consequently, there were two free fit parameters, k_{off} and $bgRatio$, but due to the long run times of model simulations a conventional fitting approach could not be applied. Instead, the model was evaluated on a grid of parameter values and the parameter set yielding the smallest sum of squared residuals was chosen as the best fit. For a more fine-grained determination of parameter values and better fit to the data a second round of optimization with a finer grid of parameter values centered around the best parameter set was performed. This approach had already been proposed by Sprague and colleagues [152].

2.2.3 Widefield FRAP yields a similar dissociation rate constant compared to confocal FRAP

The lower (especially axial) resolution of widefield microscopy compared to confocal microscopy is expected to affect the recovery curves and estimated parameters from widefield FRAP. Moreover, confocal microscopy has been the standard approach for FRAP experiments. A side-by-side comparison of FRAP by both imaging modes could, thus, help to confirm the value of widefield FRAP as an alternative approach to measure binding properties, when large number of cells and conditions need to be measured. EGFP-LacI was recruited to the *lacO* sites and FRAP measurements were taken by both approaches. Widefield FRAP was measured by the workflow described above (**Figure 2.17 A**). Confocal time courses had to be shorter due to movement of the bleached spot out of focus, but were acquired with a higher temporal and spatial resolution using a 63x objective (**Figure 2.17 B, C**).

Single-cell recovery curves showed a larger variability for the confocal imaging mode, but for both modes the reaction-diffusion model produced good fits (**Figure 2.17 A, C**). Single data points displayed a wider scattering around the fit for confocal mode. While there was a more complete recovery for two cells in confocal mode, most cells showed incomplete recovery during the acquisition time. This was also reflected in the average curves and fits (**Figure 2.17 D**). In widefield mode there was nearly complete recovery. This can be partly attributed to the longer observation period. However, also for early time points there was stronger and faster recovery. The initial steep rise of widefield curves could be attributed to a larger fraction of observed freely diffusing particles. This is consistent with the fact that in widefield mode additional fluorescence intensity of unbound molecules above and below the binding site cluster are detected. This freely diffusive fraction is also present, but smaller in confocal FRAP. Since the fit of the full model can accomodate variable diffusive and bound fractions, this should not pose a problem for the estimation of the dissociation rate constant. Indeed, off-rates estimated from confocal and widefield FRAP curves were nearly identical around $0.01s^{-1}$ (**Figure 2.17 E**).

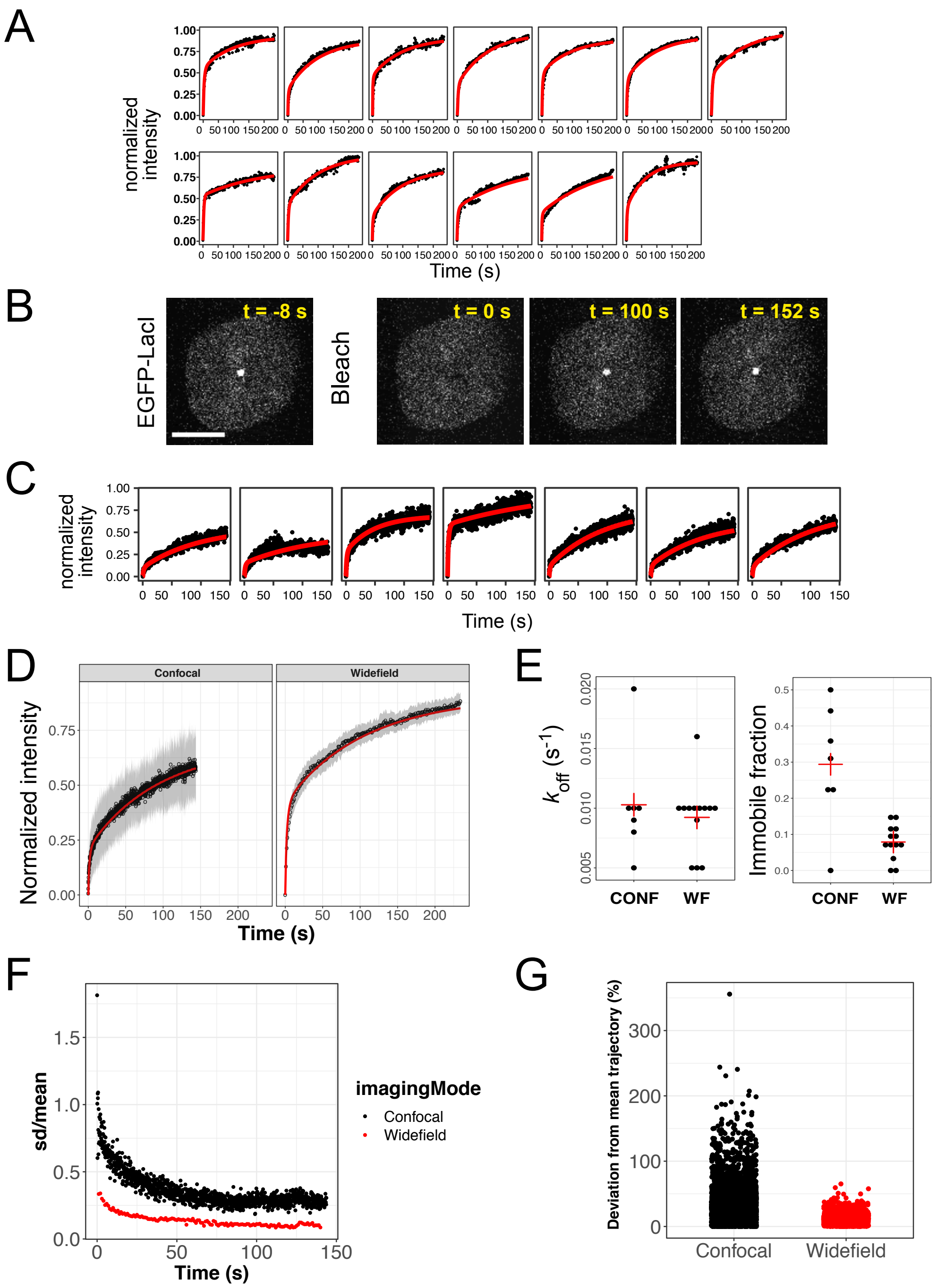


Figure 2.17: Comparison of widefield FRAP of EGFP-LacI with confocal FRAP. (The confocal FRAP data was acquired by Anne Rademacher.) **(A)** On-spot recovery curves of single-cells acquired with widefield FRAP. **(B)** Image series of confocal FRAP. Scale bar: 10 µm **(C)** On-spot recovery curves of single-cells acquired with confocal FRAP. **(D)** Mean recovery curves of both imaging modes. Red line: fit, shaded area: 95% CI, n = 7 (confocal), 13 (widefield). Note that confocal FRAP was acquired for a shorter time, but with higher time resolution. (Continued on next page)

However, there was a clear difference between the estimated immobile fractions (around 8 % vs. 30%). This finding could be partly caused by the combination of a shorter observation time in confocal FRAP and the separate fitting of the immobile fraction as a first step of the parameter estimation. If the shape of the confocal recovery curve for later (unobserved) time points deviated from the phenomenological fit function (eq. 2.13), it could mistakenly assume that the curve had reached its plateau value at the end of the observation interval.

Independent of potential fitting problems, a deviation of the immobile fraction is also expected, if there is a larger diffusive fraction in widefield FRAP, that will reduce the relative weight of the other molecular fractions. It is, thus, reasonable to assume that widefield FRAP will typically underestimate the immobile fraction, but give accurate estimates of the dissociation rate. It is noted, that the relative weights of bound and diffusive fraction (in contrast to the immobile fraction) do not contain biological information, because they depend on expression levels, binding site copy number, condensation state of the reporter gene cluster and the observed focus volume.

The wider scattering of data points around the average recovery curve could also be confirmed for the distribution of single-cell values for each time point. The relative standard deviation (coefficient of variation) was larger for confocal recovery curves at every time point (**Figure 2.17 F**). The same could be observed, when comparing the relative deviation of single-cell data points from the mean value of the respective time point (**Figure 2.17 G**). This confirmed the expectation that widefield imaging with a lower axial resolution will be more robust for observing the intensity of a bleached object that can move out of focus. In summary, widefield FRAP represents a robust technique to measure dissociation rates in large numbers of cells, but its tendency to underestimate the immobile fraction must be kept in mind, when interpreting recovery curves.

Figure 2.17: (Continued) **(E)** Binding parameters estimated from recovery curves of both imaging modes. Red crosses indicate the median value. **(F)** Comparison of acquisition noise. Shown is the standard deviation of the normalized spot intensity divided by the mean over time across single cells. **(G)** Distribution of the relative deviation from the population mean for all time points and cells.

2.3 Characterizing phase-separation and activation by synthetic TFs

The modular nature of TF protein domains allows to construct synthetic TFs in a combinatorial manner and, hence, to screen for interdependencies of the properties of the DNA binding (DBD) and activation domain (AD). Here, a collection of transcriptional activation complexes was assembled and tested for their binding, phase separation and transcriptional activation properties. This collection covered single- and multi-component complexes, CRISPR/dCas9- and bacterial repressor-mediated binding, endogenous, viral and synthetic ADs and constitutive and light-dependent activation. The characterization of this set of synthetic TFs regarding their binding, phase separation and activation properties is part of a collaborative project with Lukas Frank and has resulted in a separate manuscript [193] that is currently in the submission process. It is presented completely, here and in the following sections (sections 2.3-2.6), in order to explain all the resulting conclusions. In cases, where experiments or analyses were not performed by the author of this thesis, but needed to be presented to understand the context of the study, this is clearly indicated. This regards mainly the development of CRISPR/dCas9 based TF architectures, qRT-PCR experiments and acquisition of representative images. Cases, where data acquisition was shared by both project participants (e.g. FRAP image series or images for radial profiles), are not indicated separately.

2.3.1 Constitutive and light-responsive synthetic TFs successfully target reporter binding sites

Different DBDs and ADs were combined in a modular manner to induce gene expression of the gene reporter construct described above (**Figure 2.1**): (1) DBDs comprised LacI for direct binding to lacO repeats, rTetR for binding to tetO repeats and dCas9 with suitable guide RNAs. (2) As ADs the activation domains of VP16, VPR, p65, Rta and STAT2 were tested and as a mock effector EGFP. (3) The linkage of the DBDs with the ADs involved direct fusions to a given DBD, constitutive recruitment of tdPCP-AD to its cognate PP7-loops in the guide RNA, light-induced recruitment of PHR-AD to CIBN-dCas9-CIBN or light-induced recruitment of PHR-AD to tdPCP-CIBN bound to PP7-guide RNA.

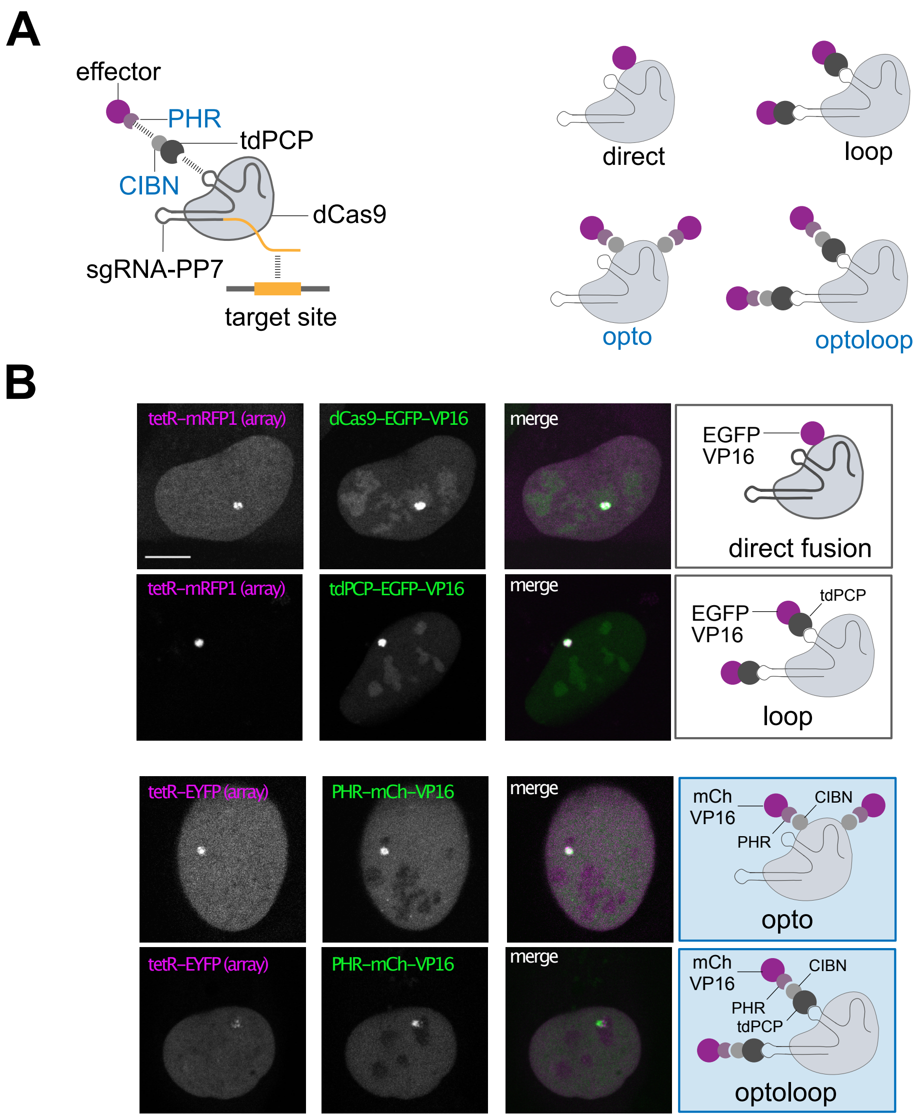


Figure 2.18: Targeting the reporter gene promoter with a set of constitutive and light-responsive synthetic TF architectures. **(A)** Explosion graph of the ”optoloop” complex shows molecular interactions of the components (left). Four different TF architectures allow to recruit the effector to the dCas9 complex (right). Direct fusion, recruitment via the gRNA PP7 loops by tdPCP or light-triggered interaction of PHR and CIBN. CIBN can either be directly fused to the dCas9 molecule or recruited to the PP7 loops in the tdPCP-CIBN fusion. Visualizations are shown with kind permission of Lukas Frank. **(B)** VP16 fusions of the dCas9 based synthetic TFs can be detected at the TetR-mRFP (or TetR-YFP) labeled promoters (magenta), when recruited to the *lacO* sites. Effector constructs were labeled with EGFP or mCherry (green). (Images were acquired by Lukas Frank and are shown for completeness.) Scale bar: 10 µm.

The recruitment modes were named direct, loop, opto and optoloop recruitment, respectively (**Figure 2.18 A**). The fusion protein containing the AD was always labeled with EGFP. CRISPR/dCas9-based constructs were developed by Lukas Frank. The opto complex has been developed before under the name LACE (light-activated CRSIPR-Cas9 effector) and was employed for the light-dependent activation of endogenous genes [194]. Some of the activation complex architectures differed from endogenous TFs, which consist of a single component and are typically smaller than the dCas9 protein. The large size of DNA binding component or the multi-subunit architecture could negatively affect transcription. However, these variable architectures have the potential to lead to a large range of AD residence times at the promoter, which is a main focus of this study. Flexible multi-component complexes can therefore be a model system for TFs with a broad range of activator residence times and phase-separation propensities.

Recruitment of the activators to their target site and induction of transcription was analyzed in the reporter cell line U2OS 2-6-3. Binding of fusion proteins to the operator repeats can be quantified by the accumulation of fluorescence intensity at the tandem array of reporter gene units in parallel with the production of nascent RNA by fluorescently labeled tdMCP. This approach allowed to evaluate AD turnover rates and the dynamic induction of transcription in the same cellular system.

Successful binding was observed for all complex architectures with the VP16 AD (**Figure 2.18 B**) and the VPR AD (data not shown). Binding was absent in the dark and reversibly induced by blue light illumination for PHR-mCherry-VP16 targeted to CIBN-LacI, CIBN-dCas9-CIBN (opto complex) and dCas9 + tdPCP-CIBN (optoloop complex) with binding kinetics comparable to PHR recruitment to CIBN-TetR [12] (data not shown). Thus, all activation complex architectures could successfully recruit the AD either constitutively or upon light stimulation.

2.3.2 The propensity to form optodroplets distinguishes ADs with weak and strong multivalent interactions

Next, the propensity of the ADs to self-interact was characterized by comparing their ability to induce the formation of PHR-AD optodroplets. The PHR domain has been shown to undergo phase separation leading to the formation of spherical liquid-like condensates (optodroplets) [140] and this ability is increased by homotypic interactions of its fusion partner [101]. Microscopy images of transfected cells expressing heterogeneous levels of PHR-EGFP-AD and CIBN-rTetR in the presence of doxycycline were acquired.

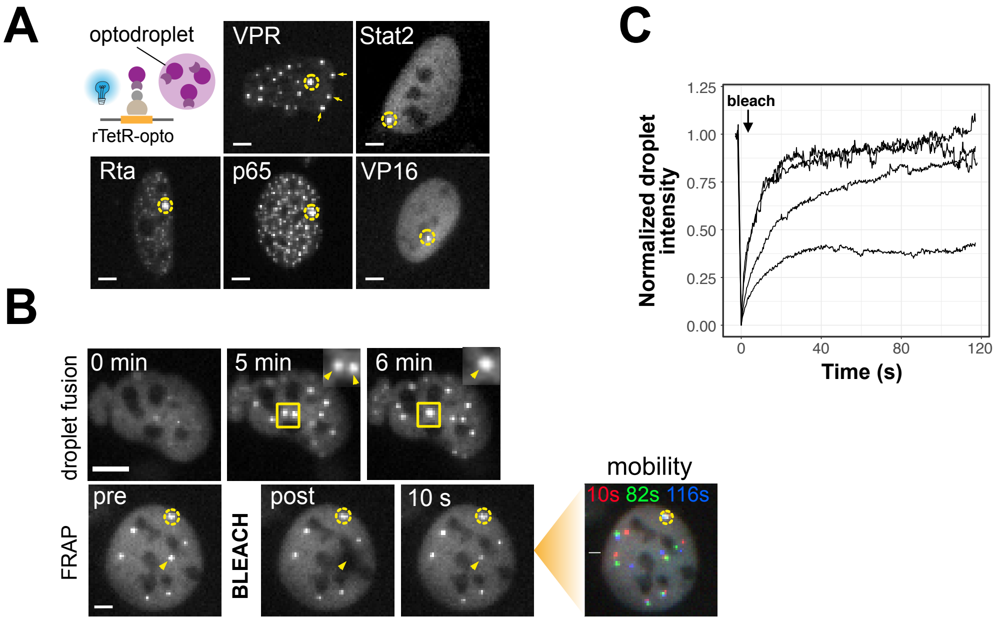


Figure 2.19: PHR-EGFP-AD fusions upon illumination form optodroplets which exhibit hallmarks of a liquid phase. **(A)** The formation of PHR optodroplets depends on its fusion partner. Cells transfected with CIBN-rTetR and PHR-AD showed droplets for VPR, p65 and Rta, but not VP16 and STAT2. Yellow ring marks the reporter cluster. Scale bar 5 µm. **(B - C)** Hallmarks of liquid droplets for PHR-EGFP-VPR. **(B)** Image time course of droplet formation after illumination. The yellow box highlights the fusion of two optodroplets (top). FRAP image series of optodroplet (bottom, left). Promoter array marked by yellow circle, bleach region indicated by yellow arrow. Mobility of gene cluster and optodroplets visualized by an overlay of three color-coded time points (bottom, right). Scale bar: 10 µm. **(C)** Recovery curves of bleached optodroplets in four cells.

After several cycles of repeated imaging the blue light exposure led to the formation of spherical structures (optodroplets) in addition to the spot at the cluster of *tetO* binding sites (marked by dashed yellow circle, **Figure 2.19 A**) for the VPR, p65 and Rta fusions, whereas such structures were seen less often or not at all for VP16 and STAT2. The propensity for optodroplet formation was reflected by the fraction of cells with visible droplets ranging from 0 % for STAT2 and 27 % for VP16 to 41-89 % for VPR, p65 and Rta (**Table 2.3**). Droplets showed hallmarks of a liquid phase like droplet fusion and fast recovery after bleaching in FRAP (**Figure 2.19 B, C**). In some cells, however, recovery was reduced (**Figure 2.19 C**) indicating that there might be a maturation to a more solid or gel-like state. Droplets appeared later than the spot at the *tetO* site cluster and were more mobile (**Figure 2.19 B**) suggesting that they are not chromatin bound. Since droplet formation only occurs, once the concentration of PHR-AD molecules in the light-induced conformation exceeds a critical value, light-switched PHR-AD molecules under sub-critical conditions will first bind to CIBN-rTetR at the *tetO* cluster and only later form droplets.

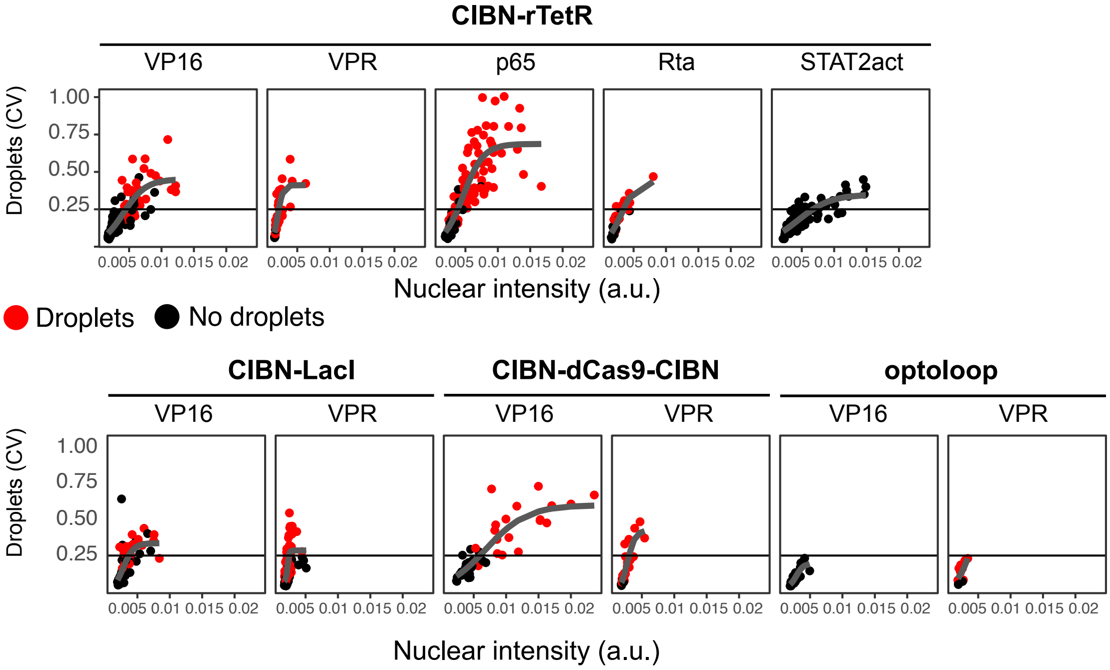


Figure 2.20: Quantification of critical values for phase separation of the synthetic TF complexes. The binding targets are indicated in bold above the plots and the AD in the PHR-EGFP-AD fusion directly above the plots. The abundance of optodroplets (measured by the CV of the nuclear intensity) increased with expression level. The relation of nuclear intensity and droplet abundance was fitted by a logistic function (grey line). Manual classification of cells as droplet containing (red) or not (black) is color coded. The CV-threshold defining the critical value was set to 0.25 (black horizontal line).

In order to quantify the propensity to induce droplets, the critical concentration for droplet formation of the different AD fusions was determined. The droplet abundance across the imaged cell population, measured by the coefficient of variation (CV) of the nuclear PHR-EGFP-AD intensity, was fitted as a function of the expression level (nuclear intensity) and the intensity at which it crossed a threshold determined (**Figure 2.20**). This threshold was chosen so that it separated cells with and without visible droplets according to the manual annotation. The critical values ranged from 0.002 a.u. for VPR with a high self-interaction propensity to 0.007 a.u. for STAT2 with a low self-interaction propensity (**Table 2.3**).

Table 2.3: Optodroplet Propensities

AD	Target	n	Fraction(%)	c_{crit}
VP16	CIBN-rTetR	122	27	0.0045
VP16	CIBN-LacI	73	32	0.0038
VP16	optoloop	41	0	outside range
VP16	opto	49	39	0.0064
VPR	CIBN-rTetR	35	89	0.0022
VPR	CIBN-LacI	101	63	0.0025
VPR	optoloop	15	80	outside range
VPR	opto	35	60	0.003
p65	CIBN-rTetR	123	71	0.0038
STAT2	CIBN-rTetR	97	0	0.007
Rta	CIBN-rTetR	34	41	0.0036

There was also an effect of the binding partner. While VP16 had a low self-interaction propensity when recruited to CIBN-rTetR (0.005 a.u.), it increased when recruited to CIBN-LacI (0.004 a.u.). LacI forms dimers and tetramers [59] raising the possibility that it could work as a multivalent bridging factor between PHR-AD molecules favouring droplet formation. For the opto and optoloop complexes the optodroplet propensity was reduced in comparison to recruitment to CIBN-rTetR, but showed the same trend for the ADs: low propensity for VP16 and high for VPR.

2.3.3 Architecture and AD type determine TF activation capacity

The ability of the synthetic TF architectures to induce transcriptional activation and co-activation marks was checked. The bulk production of reporter RNA was checked by qRT-PCR, while nascent RNA levels were measured in single cells by microscopy. Importantly, this readout does not necessarily agree quantitatively with single-cell readouts of nascent RNA at the promoter. Bulk RNA levels represent the accumulated products of promoter activity. Biologically, they also depend on the RNA degradation rate, while the levels of nascent RNA besides promoter activity also depend on the rate with which fully transcribed RNA molecules leave the transcription site. Technically, differences between single-cell microscopy and bulk qRT-PCR detection can arise, because cells are selected for quantification in microscopy, whereas the average across all cells is measured by qRT-PCR. Consequently, qRT-PCR results can depend on the number of actively transcribing cells, which is influenced by transfection efficiency and cytotoxicity of the transfected constructs. Microscopy analyses can be biased by the unintended selection of cell groups, e.g. by segmentation difficulties in cell sub-populations. Measuring transcriptional activity by the two complementary methods gives a biologically more complete picture and allows to identify cases, in which technical biases could be present.

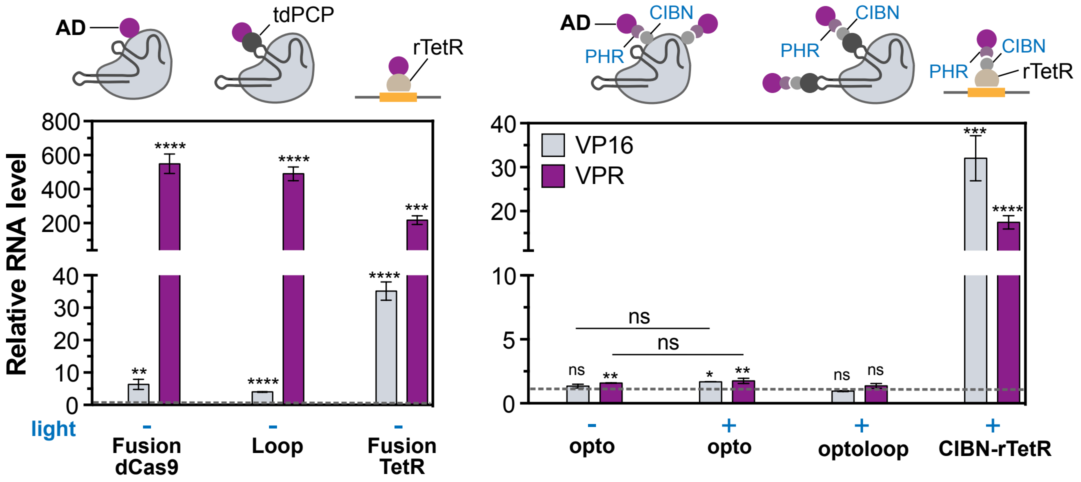


Figure 2.21: Activation capacity of synthetic TFs at reporter gene cluster measured as bulk RNA by qRT-PCR. Complex architectures for constitutive (left) and light-responsive synthetic TFs are indicated by cartoons above. Bars represent mean RNA levels relative to mock transfected cells. (qRT-PCR experiments were performed and analyzed by Lukas Frank and are shown for completeness.) Dotted line: mock transfection (normalization reference). Error bars: $\pm 1 s.d.$, n.s.: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$, two-sided unpaired Student's t-test.

Bulk measurements by qRT-PCR (performed and analyzed by Lukas Frank) 24 h after transfection with the synthetic TF constructs confirmed the ability of most TF architectures to induce RNA production. The two light-dependent dCas9 complexes, opto and optoloop, constitute an exception. The optoloop complex did not induce bulk RNA levels significantly above mock transfection levels for neither VP16 nor VPR (**Figure 2.21**). While there was weak induction by the opto complex (VP16: 1.7-fold, VPR 1.8-fold) the induction levels were not greater under illumination than in the dark. It is concluded, that the CRISPR/dCas9 activation complexes are not able to induce transcription of the reporter gene studied here in a light-controlled manner. When comparing the strength of ADs, VPR was a much stronger activator than VP16 in most conditions.

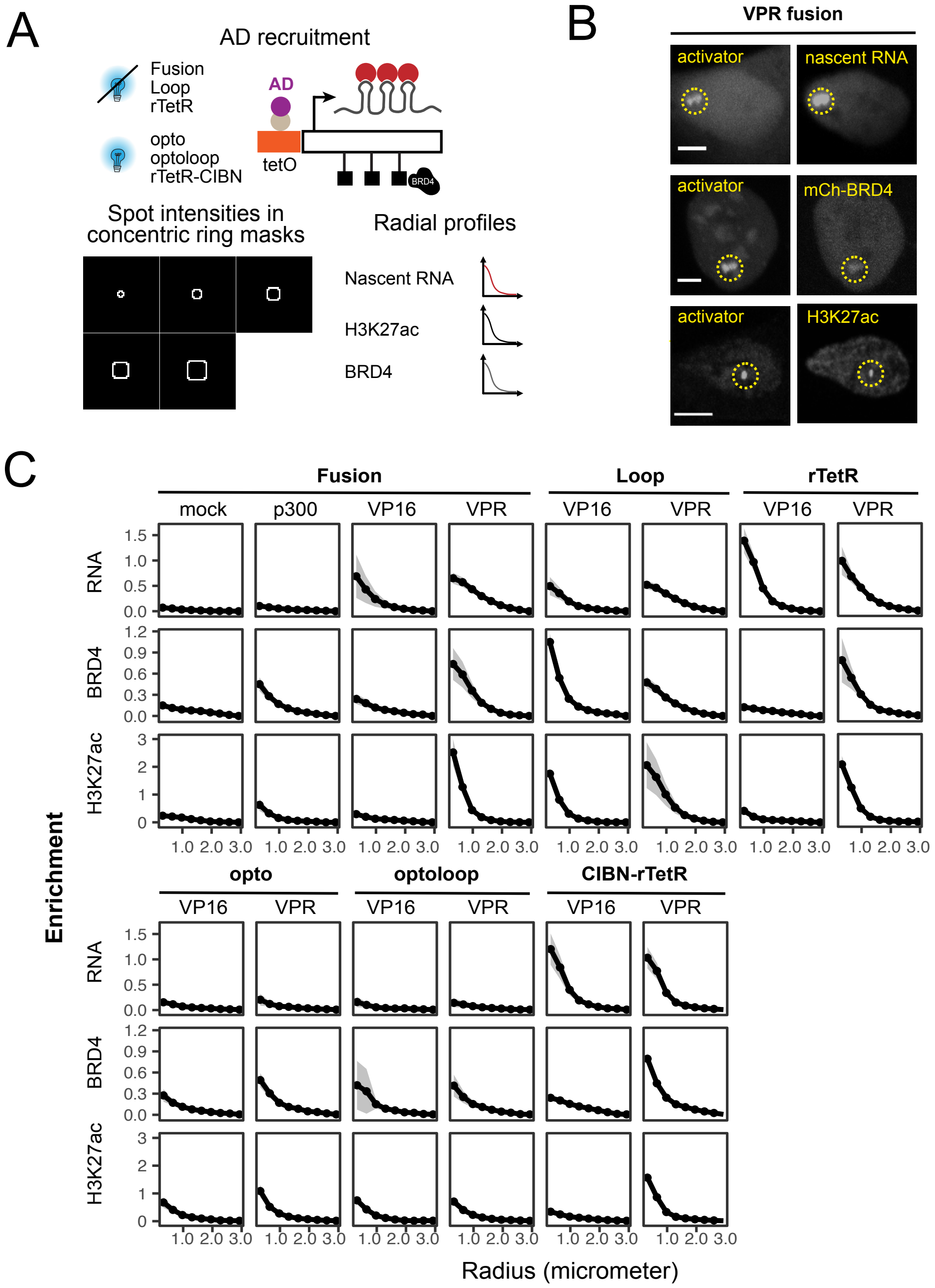


Figure 2.22: Promoter activity induced by synthetic TFs visualized by intensity profiles. **(A)** Readouts of transcriptional (co-)activation included nascent RNA (labeled by tdMCP-tdTomato), mCherry-BRD4 and immunofluorescence staining for H3K27ac (top). Intensity profiles of the three readouts (bottom right) were obtained from microscopy images measuring the median intensity in concentric ring shaped masks. (For details see section Material and Methods). **(B)** Exemplary images of nascent RNA, H3K27ac and BRD4 induced by dCas9-EGFP-VPR. (Continued on next page.)

For the single-cell readout activation and co-activation marks were measured by recruiting the different complexes containing VP16 or VPR to the *tetO* sites and recording (1) tdMCP-tdTomato binding to MS2 loops in the nascent reporter RNA, (2) BRD4-mCherry co-recruited to the cluster of promoters in live cells and (3) histone H3 actetylation of lysine 27 (H3K27ac) by immunofluorescence staining. Cells were illuminated over 24 h and 81 positions for each condition were imaged on a spinning-disc microscope. For a semi-quantitative comparison enrichment scores of fluorescence intensity above the nuclear background intensity at the reporter gene cluster were calculated and averaged for concentric rings around the spot center to obtain radial profiles (**Figure 2.22 A**).

Note that the intensity enrichment profiles are normalized in a way that prevents a quantitative comparison (for details see Materials and Methods). For instance, a broadening of the profile that could be caused by VPR decondensing the gene array will lead to lower enrichment score values at the profile center. Instead they offer a way to judge the presence or absence of intensity enrichment.

The strongest activator dCas9-EGFP-VPR showed spot enrichment of the signal for all three readouts (**Figure 2.22 B**) and the nascent RNA profile was clearly above the negative control, the mock effector dCas9-EGFP (**Figure 2.22 C**). Production of nascent RNA was induced for the directly binding complexes based on dCas9 and rTetR, as well as for the loop complexes and the opto-rTetR complexes. Consistent with the qRT-PCR readout there was no transcriptional induction by the opto or optoloop complexes. Further characterization of activation deficient complexes is presented in section 2.6.1.

Figure 2.22: (Continued) **(C)** Mean radial intensity enrichment profiles of the (co-)activation readouts for the set of studied synthetic TFs. The enrichment score is affected by intensity enrichment at the gene cluster, but also by additional quantities like spot size and background levels and by the variability between single-cell profiles. It should not be interpreted quantitatively, but rather qualitatively to judge the presence or absence of readout intensity enrichment in the spot. Shaded area: 95 % CI, n = 16 - 520 cells per condition.

Interestingly, there was strong correlation of the two measured co-activation marks BRD4 and H3K27ac. For each TF construct both marks were present or absent together. The core domain of a histone acetyl transferase (HAT), p300, was included as an effector domain. dCas9-EGFP-p300 did not induce production of RNA, but histone acetylation and BRD4 recruitment. Also all complexes containing VPR induced these co-activation marks. Strikingly, this also comprised the transcription-deficient opto and optoloop complex. Since the VPR opto and optoloop complexes successfully induced histone acetylation and BRD4 recruitment, it is hypothesized, that these co-activation marks could be sufficient for the induction of some genes, but not for the reporter cluster studied here, which has been reported to be in a heterochromatic state [144]. The VP16 complexes, generally, did not induce strong acetylation or BRD4 recruitment, with the exception of loop VP16.

Overall, the activation and co-activation assay demonstrated that the architecture of the synthetic TFs determines, whether the complex is able to activate transcription or not, whereas the recruitment of co-activators appears to depend on the AD and to be mostly independent of the TF architecture that recruits the AD to the promoter.

2.4 TF binding and activation: stabilization by multivalent interactions and the role of residence time

The toolbox of synthetic TFs described in section 2.3 with its diverse DBD and AD modules opens a way to systematically and independently vary the DNA binding strength and the strength of multivalent interactions. Here, these activation complexes were subjected to the FRAP workflow introduced in section 2.2 to dissect the contribution of the binding properties of the two TF domains to transcriptional activation and co-activation. Moreover, the contribution of multivalent interactions to stable target site association of TFs was investigated.

The dissociation rate constant k_{off} and its inverse, the residence time, are the main binding parameters that can be determined by FRAP and constitute interesting candidates for regulatory properties of TFs. Binding is characterized by the association rate with rate constant k_{on}, the dissociation rate constant k_{off} and the dissociation constant K_D which corresponds to the ratio $\frac{k_{off}}{k_{on}}$ and determines the binding site occupancy in equilibrium. The inverse of the rate constants are the TF search time and the residence time. Since binding is a bimolecular reaction (TF and binding site) the association rate depends on the TF expression level. In contrast, the dissociation is independent of expression levels making k_{off} or the residence times quantities that can be compared between cells with variable expression levels. While TF search times have been proposed to regulate transcription [99], it seems unlikely that it does so beyond its influence on binding site occupancy. With a binding rate k_{on} of $9 \cdot 10^4\ M^{-1}s^{-1}$ TF binding to its target sites happens in $10 - 100\ s$ at typical expression levels of $100\ nM - 1\ \mu M$ [72]. Once binding equilibrium is achieved it is difficult to imagine a scenario, in which k_{on} would regulate transcription independent of binding site affinity and occupancy. Contrarily, in section 2.4.2 a model of transcriptional activation is proposed that depends on the residence time independent of its influence on binding site occupancy. The dissociation rate k_{off} can be determined in FRAP measurements. For these reasons it was chosen as the quantity of interest to compare the binding of the synthetic TFs studied here.

2.4.1 FRAP of activation complexes reveals stabilizing effect of multivalent interactions of the AD

To determine the off-rates, the apparent turnover of the activation complexes bound to the *lacO* sites of the reporter gene (on-spot) was measured, whereas the effective diffusion coefficient was measured at a nuclear position outside of the gene cluster (off-spot). For each complex, the moiety carrying the activation domain was tagged with EGFP. Fusions to EGFP alone were used as mock effectors. FRAP measurements were performed as described in section 2.2 for the four dCas9-based complex architectures described above, direct fusions to rTetR and PHR recruited to CIBN-rTetR and with or without fusions of VP16 and VPR . FRAP was measured for on-spot and off-spot positions and the normalized recovery curves were fitted by a reaction-diffusion model for clustered binding sites comprising an immobile fraction and using the effective diffusion coefficient determined from the off-spot FRAP measurements. Recovery curves for single-cells were averaged after fitting.

Recovery by diffusion was rapid on the time scale of several seconds for all constructs (**Figure 2.23 A**). The estimated effective diffusion coefficients ranged between 1.0 and 4.0 $\mu m^2 s^{-1}$ (**Figure 2.23 B, Table 6**). Direct fusion constructs containing dCas9 or LacI as a DBD and PHR constructs with a high propensity to form optodroplets

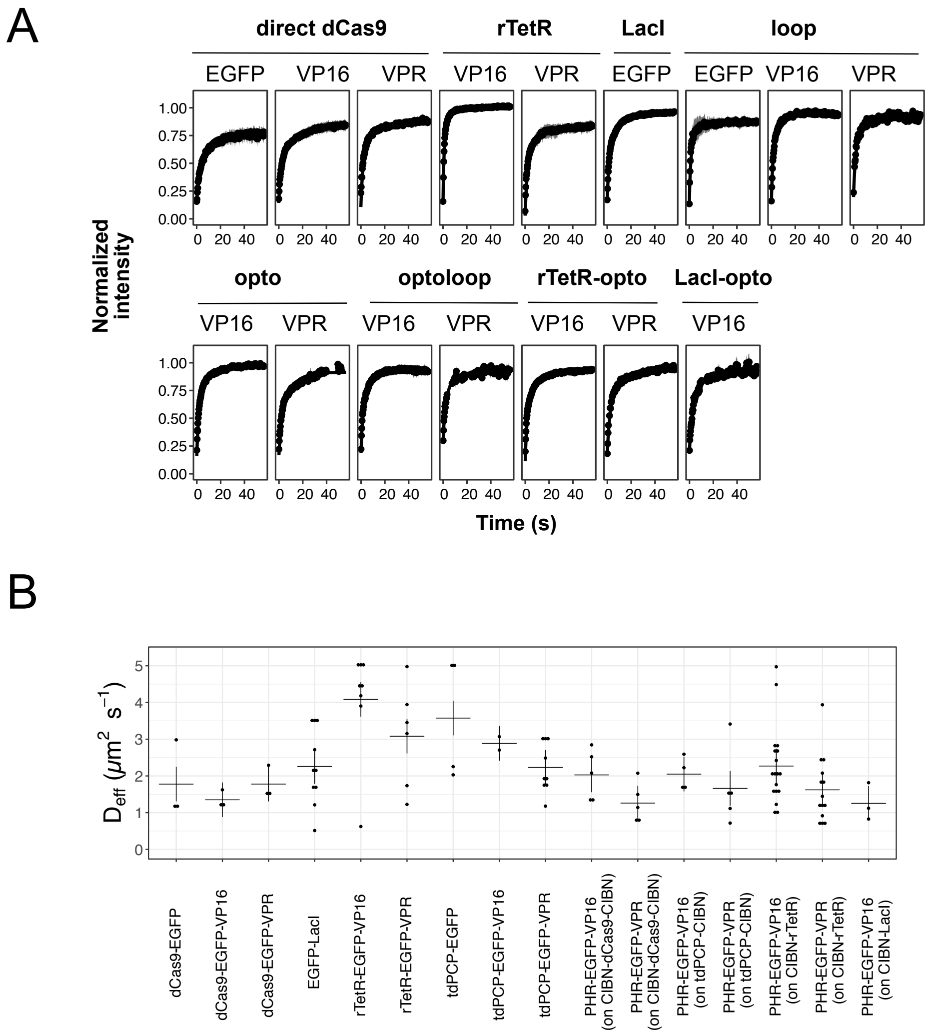


Figure 2.23: Diffusion of synthetic TF complexes measured by off-spot FRAP. **(A)** Average recovery curves with fit (solid line). Shaded area: 95 % CI, n = 2 - 18 cells per condition. **(B)** Effective diffusion coefficent estimated from single off-spot recovery curves. Crosses indicate median values.

(PHR-VPR constructs and PHR-VP16 in combination with CIBN-LacI) had lower values of D_{eff}. This could be the result of non-specific DNA binding by the DBDs and of the formation of large and slowly diffusion PHR-AD oligomers or optodroplets. For LacI 90% of molecules have been shown to be bound non-specifically to DNA with short residence times [195] supporting this possibility.

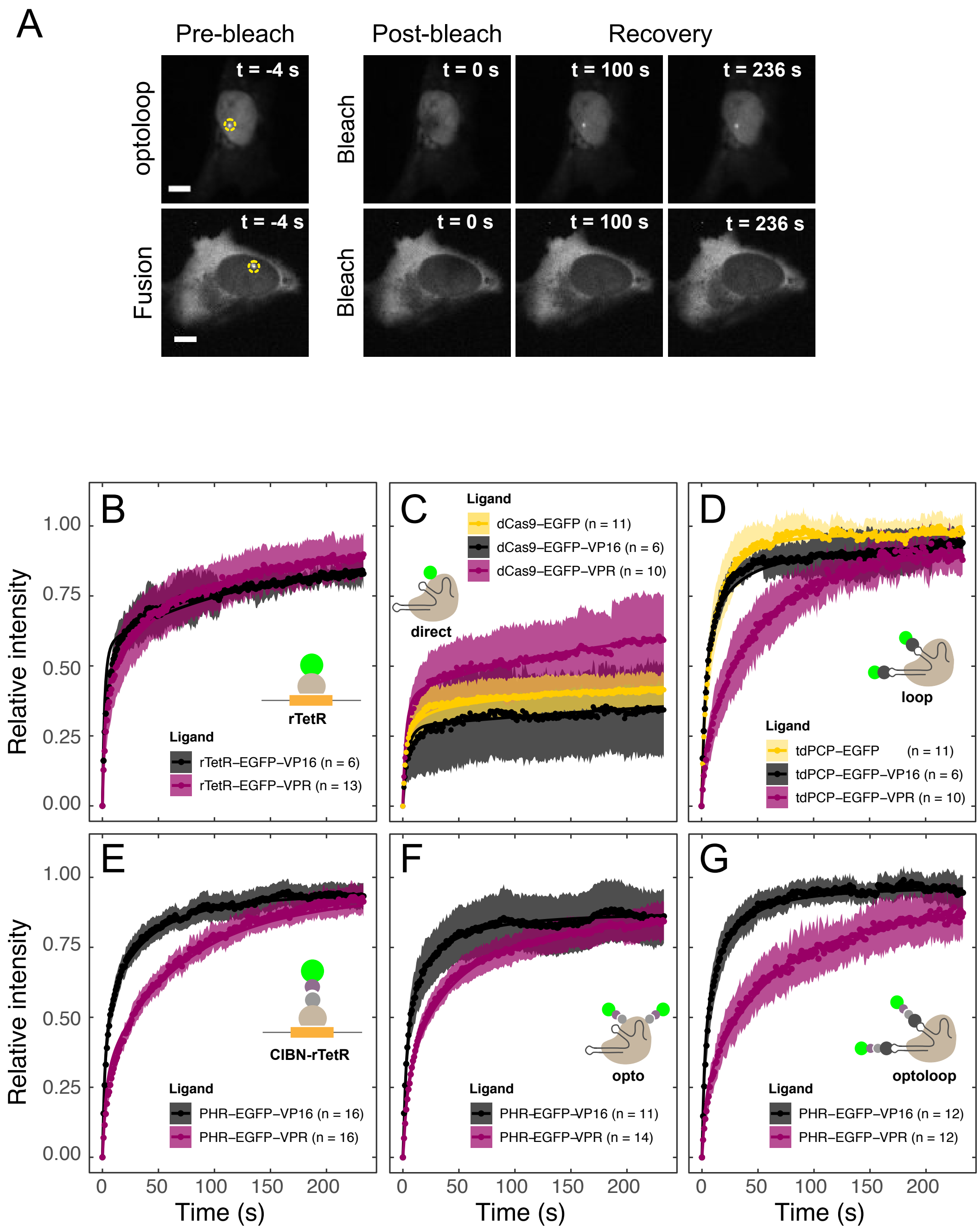


Figure 2.24: Binding kinetics of synthetic TF complexes recruited to *lacO* sites (or *tetO* sites, if indicated) measured by on-spot FRAP. **(A)** Exemplary image series of complexes with very slow and fast binding turnover: dCas9-EGFP (on DNA, "Fusion") and tdPCP-EGFP-VP16 (on PP7 loops of dCas9 complex, "optoloop"). Yellow circle: gene cluster spot. Scale bar: 10 µm. **(B-G)** Average recovery curves (points), fit (solid line), 95 % CI (shaded area). **(B)** Direct fusion of EGFP-AD to rTetR. Recruited to *tetO* sites. **(C)** Direct fusion of EGFP-AD to dCas9. (Continued on next page.)

In on-spot FRAP experiments strong differences could be observed between complexes. While there was only recovery by diffusion, but not by binding turnover, e.g. for dCas9-EGFP, other constructs like the PHR-EGFP-VP16 in the optoloop complex exhibited a relatively fast return of fluorescence intensity also by exchange of bound molecules at the spot (**Figure 2.24 A**). Direct fusions to dCas9 displayed the slowest recovery (after the fast diffusive recovery of unbound molecules), whereas tdPCP-EGFP on the PP7-loops of the dCas9 complex recovered to its initial intensity within few seconds (**Figure 2.24 B-G**). The light-induced interaction of PHR and CIBN led to recovery times comparable to tdPCP on PP7 for the respective VP16 or VPR fusions. For the loop, opto and optoloop complex as well as for PHR-fusions binding to CIBN-rTetR recovery was slowed down by VPR compared to VP16, whereas dCas9-EGFP-VPR showed a recovery to higher levels compared to dCas9-EGFP-VP16. In the light of VPR's ability to self-interact these observations suggest that VPR can stabilize the binding of transiently bound complexes, whereas for strongly bound complexes the additionally recruited, indirectly bound fraction of molecules can be observed as an additional recovery.

The estimated effector off-rates reflected the observations made for the recovery curves (**Figure 2.25, top**). Estimated FRAP parameters are summarized in **Table 6**. The residence times were calculated as the inverse of k_{off}. They ranged from 12 s for tdPCP-EGFP on PP7 to more than the observation time of 240 s for dCas9-EGFP-VP16 on the lacO sites. Residence times for the loop, opto, opto-rTetR and optoloop complexes were between 28 and 42 s for VP16 and between 47 and 71 s for VPR, with the residence times of VPR for each complex architecture consistently longer by at least 14 s. For the fusion constructs that directly bound to DNA, rTetR and dCas9, the VPR fusion had shorter residence times (56 s and 204 s) than VP16 (132 s and more than 240 s) supporting the idea that its indirectly bound fraction exchanges faster than the direct interaction of the complex with its binding site. The immobile fraction that did not

---

Figure 2.24: (Continued) **(D)** tdPCP-EGFP-AD bound to PP7 loops of the dCas9 complex. **(E-G)** PHR-EGFP-AD binding to CIBN-fusions after at least 1 min blue light exposure. **(E)** CIBN-rTetR recruited to *tetO* sites. **(F)** CIBN-dCas9-CIBN. **(G)** tdPCP-CIBN bound to PP7 loops of dCas9 complex.

show recovery during the observation period was low for the loop complexes (between 2.5 % for tdPCP-EGFP and 6.6 % for tdPCP-EGFP-VPR), low to intermediate for the light inducible complexes (2.0 % for optoloop VP16, 14 % for opto VPR) and consistently higher for the VPR fusions compared to VP16 (**Figure 2.25, bottom**). For directly binding complexes immobile fractions were comparable for rTetR (VP16: 5.3 %, VPR: 6.9 %), but for the dCas9 fusions VPR had a strongly reduced immobile fractions (19% compared to 37 % for dCas9-EGFP-VP16), probably due to the additional indirectly bound fraction of molecules with a faster exchange.

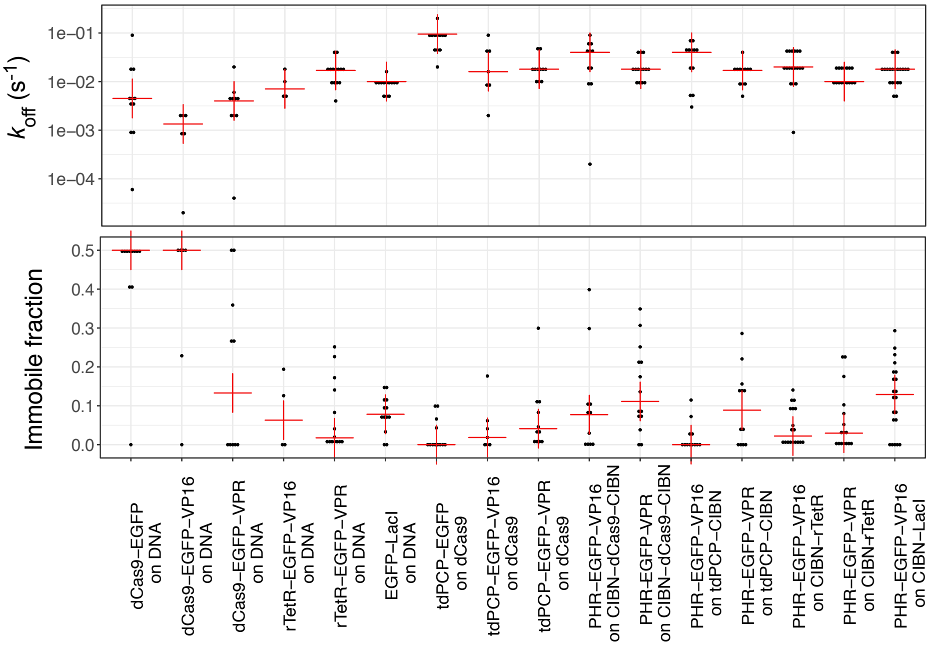


Figure 2.25: Binding parameters of synthetic TF complexes estimated from on-spot FRAP recovery curves in **FIgure 2.24**. Red crosses indicate median values.

To check, whether the additional fraction of indirectly recruited molecules could also be detected as an increased local concentration of activation complex at the reporter gene cluster, the loop complexes with VP16 or VPR recruited to the *tetO* sites were imaged with identical settings and their intensities relative to a marker of the reporter gene cluster, tagBFP-LacI, were quantified (**Figure 2.26**). There was a 1.9-fold higher spot intensity for tdPCP-EGFP-VPR compared to tdPCP-EGFP-VP16 ($p = 0.0006$, Welch two-sample t-test), confirming additional recruitment (**Table 7**).

In summary, the FRAP analysis revealed large differences in the residence times and immobile fractions of the synthetic TF constructs. Binding kinetics did not only depend on the DBD, but VPR, an AD with strong multivalent interactions, stabilized weakly bound complexes and led to the recruitment of an indirectly bound fraction of TF molecules resulting, overall, in a larger local concentration of activator at the promoter.

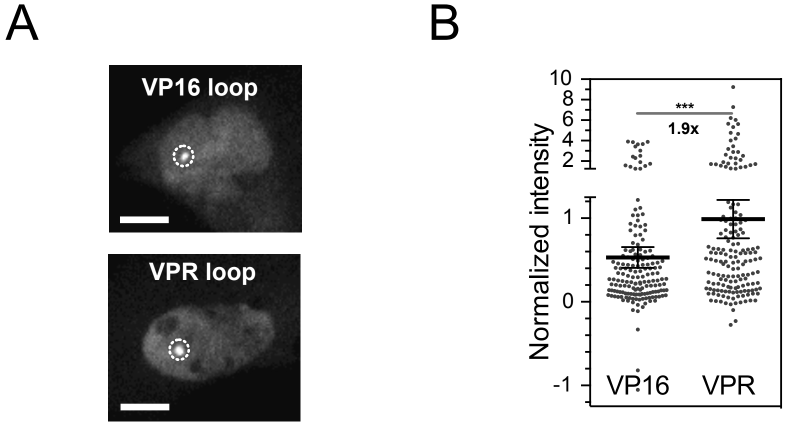


Figure 2.26: Effect of binding stabilization by multivalent interactions on binding site occupancy for the loop complexes. **(A)** Representative images of tdPCP-EGFP-VP16 and -VPR targeted to the PP7-loops of the dCas9 complex at *tetO* sites. **(B)** Distribution of local relative intensity of loop VPR complex compared to loop VP16 complex at *tetO* sites. Intensities were normalized to the tagBFP-LacI spot intensity. Dots: single cell values, bars: mean, error bars: 95% CI, ***: $p < 0.001$, two-sided unpaired Student's t-test. Local concentration was 1.9-fold higher for VPR.

2.4.2 A kinetic proof-reading model suggests residence time dependence of transcription

TF binding can modulate transcription via two parameters: binding site occupancy and TF residence time on the binding site. The former is obvious, since a higher occupancy means that there is more TF at the promoter or it is there for a larger fraction of time. The latter has been suggested as a parameter that regulates transcription independently of the binding site occupancy, e.g. by kinetic proof-reading [196] involving the removal of a promoter-occluding nucleosome [80]. We propose an additional model of residence time dependent transcription that requires the promoter bound TF to be modified before it can induce transcription (**Figure 2.27 A**). Several post-translational modifications of

TFs are known, comprising acetylation [175, 197, 198] and ubiquitinylation [199, 200] and have been shown to increase transcription. In this model the reduction of the binding sites occupancy by a high off-rate can be compensated by an increase in TF expression level. The amount of RNA produced, however, is limited by the off-rate. Even at high occupancy it cannot reach the same level as for a low off-rate. Intuitively, under a high TF off-rate and high TF concentration the TF binding site will be occupied most of the time, but individual molecules remain bound for only a short time, before they dissociate and are replaced by a new molecule. The TF modification necessary to induce transcription is, thus, lost, before productive transcription can start.

The model is formally defined by a system of ordinary differential equations (ODEs) corresponding to the model topology shown in **Figure 2.27 A**. It comprises three promoter states: unbound (A), TF bound (B) and bound by a modified TF (C). The modification is assumed to be set only on promoter bound TFs and to be quickly removed from dissociated TFs. The concentration of free TF is assumed to be much larger than the concentration of promoter binding sites, so that it is not reduced, when a few molecules associate with binding sites. It can be absorbed by the binding rate constant yielding the pseudo binding rate constant $k_{on}^{*} = TF \cdot k_{on}$. For simplicity the total number of promoters is normalized so that the three promoter states sum up to 1: $A + B + C = 1$. RNA is only produced from state C with production rate k_t and removal rate k_m.

$$\begin{aligned}
\frac{dB}{dt} &= k_{on}^{*} \cdot (1 - B - C) - (k_{off} + k_1) \cdot B \\
\frac{dC}{dt} &= k_1 \cdot B - k_{off} \cdot C \\
\frac{dR}{dt} &= k_t \cdot C - k_m \cdot R
\end{aligned} \tag{2.16}$$

The effect of different values of the binding rates on the relative amounts of promoter states and on the amount of RNA that is produced can be seen in the steady state levels:

$$\begin{aligned}
B &= \frac{1}{(1 + \frac{k_{off}}{k_{on}}) \cdot (1 + \frac{k_1}{k_{off}})} \\
C &= \frac{1}{(1 + \frac{k_{off}}{k_{on}}) \cdot (1 + \frac{k_{off}}{k_1})} \\
R &= \frac{k_t}{k_m \cdot (1 + \frac{k_{off}}{k_{on}^{*}}) \cdot (1 + \frac{k_{off}}{k_1})}
\end{aligned} \tag{2.17}$$

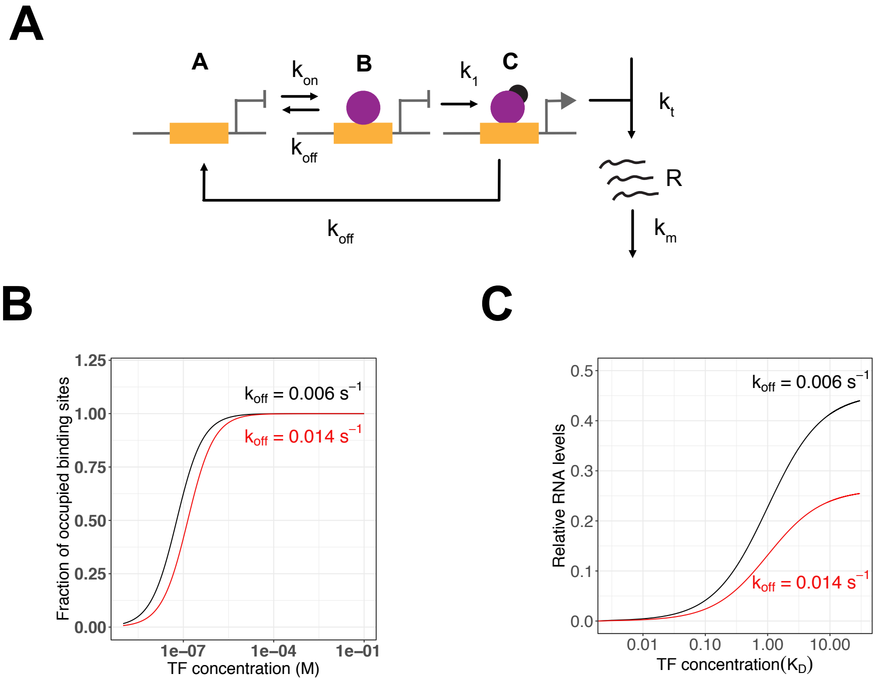


Figure 2.27: Kinetic proofreading model of TF residence time and transcriptional activation. **(A)** Model topology. There are three promoter states: unbound (A), TF bound (B) and bound by a modified TF (C). RNA (R) is only produced from the active promoter state C. The respective rate constants for transitions between promoter states and RNA production and removal are indicated. **(B-C)** Steady state values in dependence of the ligand concentration for long (black) and short (red) residence times. **(B)** Binding site occupancy saturates with increasing TF concentration at comparable levels for long and short residence times. A typical value for k_{on} of $1 \cdot 10^5 M^{-1} s^{-1}$ was assumed. **(C)** RNA production is limited by short residence times. RNA levels are shown relative to their theoretical maximal levels, if the promoter was continuously in state C.

Interestingly, the steady state RNA levels are limited by two terms in the denominator. The first term $(1 + \frac{k_{off}}{k^*_{on}})$ is related to the binding site occupancy θ, which can be determined from the sum of steady state levels of promoter states B and C.

$$\theta = \frac{1}{1 + \frac{k_{off}}{k^*_{on}}} = \frac{TF}{TF + \frac{k_{off}}{k_{on}}} \tag{2.18}$$

The second term $(1 + \frac{k_{off}}{k_1})$ contains the ratio of the unbinding rate constant and the rate constant for the modification of bound TF. Hence, if a TF has residence times shorter than the time needed for it to be modified ($k_{off} > k_1$), RNA levels are limited not only by a low occupancy, but also by the ratio of k_{off} and k_1. In the case that TF modification is much faster than unbinding ($k_{off} << k_1$) the second term approaches 1 and RNA levels are limited only by the binding site occupancy (first term) in thermodynamic equilibrium. Importantly, short residence times (low values of k_{off}) can be compensated by high TF concentrations in the occupancy term (via a large value of k^*_{on}), but TF concentration cannot compensate for the effect of the second term, which only depends on the (fixed) value of the TF modification rate. Thus, if transcription was regulated by such a mechanism, it would more specifically discriminate TF target binding from non-specific binding (with short residence times), because the off-rate would limit RNA production twice: by reduced occupancy and by the failure to modify bound TFs. Transcriptional regulation would be more robust to fluctuations in TF expression levels, because one transcription limiting factor would be concentration independent. It is noted that additional proof-reading steps, for instance two consecutive TF modifications, would increase specificity even more.

For illustration **Figure 2.27 B, C** show the occupancy θ and the relative steady state RNA levels as a function of TF concentration in two scenarios with long (167 s, $k_{off} = 0.006 s^{-1}$) and short residence times(71 s, $k_{off} = 0.014 s^{-1}$), when the average time for a bound TF molecule to be modified is 200 s ($k_1 = 0.005\ s^{-1}$). The on-rate was assumed to be $1 \cdot 10^5 M^{-1} s^{-1}$ [72]. The binding site occupancy is given by

$$\theta(TF) = \frac{TF}{\frac{k_{off}}{k_{on}} + TF} \tag{2.19}$$

For RNA levels the TF concentration is given in units of K_D so that differences in transcription at the same occupancy can be observed.

$$R_{rel}(TF) = \frac{1}{(1 + \frac{1}{TF}) \cdot (1 + \frac{k_{off}}{k_1})} \tag{2.20}$$

These steady state RNA levels are shown relative to the maximum level they could reach, if the promoter was only in state C ($R_{max} = \frac{k_t}{k_m}$).

For short residence times (**Figure 2.27 B, red line**) higher TF concentrations are needed to reach a saturation of binding sites, but for both short and long residence times full occupancy is eventually reached. In contrast, the maximum steady state RNA levels for short residence times are strongly reduced for long residence times, even at very high TF concentrations (**Figure 2.27 C**). Thus, TF residence times can limit RNA production, even at full binding site occupancy and independently of TF expression levels. The rate constant for these calculation were chosen in a range that appears reasonable for TF residence times (tens of seconds) and times for TF modification, but a significant proof reading effect is expected in all cases, in which the residence time is shorter than the modification time. The specificity for long residence time binding could be further increased by an analogous two-step TF modification mechanism, in which two consecutive modifications are needed before the TF can induce RNA production. The RNA steady state levels in this case are $R = \frac{k_t}{k_m \cdot (1+\frac{k_{off}}{k_{on}^*}) \cdot (1+\frac{k_{off}}{k_1}) \cdot (1+\frac{k_{off}}{k_2})}$ with rate constant k_2 for the second modification. Each proof-reading step increases the specificity by a factor depending on the ratio of residence time and the time for TF modification.

The TF modification model could provide a kinetic proofreading mechanism that is more adequate for the reporter gene promoter studied here than a mechanism that relies on nucleosome remodeling. Since there are multiple TF binding sites, an active promoter state characterized by the presence of a TF molecule and the absence of a promoter nucleosome could be sustained by multiple TF molecules that independently bind to the binding sites for short times. In the TF modification model, in contrast, the necessity to remain bound long enough for the TF modification to take place would apply to each of multiple TF molecules bound to the cluster of binding sites.

There are good reasons, why a kinetic proofreading mechanism of transcriptional induction by TF binding would be functionally advantageous (increased specificity, independence of expression levels) and mechanistically plausible (known TF modifications). However, comparing the residence times (section 2.4.1) and activation capacities (section 2.3) of the different synthetic TFs studied here, there was no strict dependency of successful activation on long residence times. The dCas9 based TF architectures had the longest residence times and the strongest activation potential measured by bulk RNA levels, but the loop complexes with much shorter residence times showed a nearly as strong activation. The inability of the opto and optoloop complexes to activate transcription cannot be attributed to short residence times, which are in the same range as for the activation competent opto-rTetR complex. There was, however, a strong difference between the activation potential of the ADs VPR and VP16 and VPR had a stabilizing effect on weakly bound complexes as shown in section 2.4.1. Overall, the effect of the different residence times of the synthetic TFs on their activation potential remained unclear. The comparison is made difficult by their different architectures, components and stoichiometries (e.g. two ADs for loop based complexes, compared to one in direct fusions). To isolate the effect of residence time from these confounding factors a system in which minimal changes to the synthetic TF lead to strong residence time changes was needed.

2.4.3 A minimal mutation demonstrates reduced transcriptional activation mediated by residence time

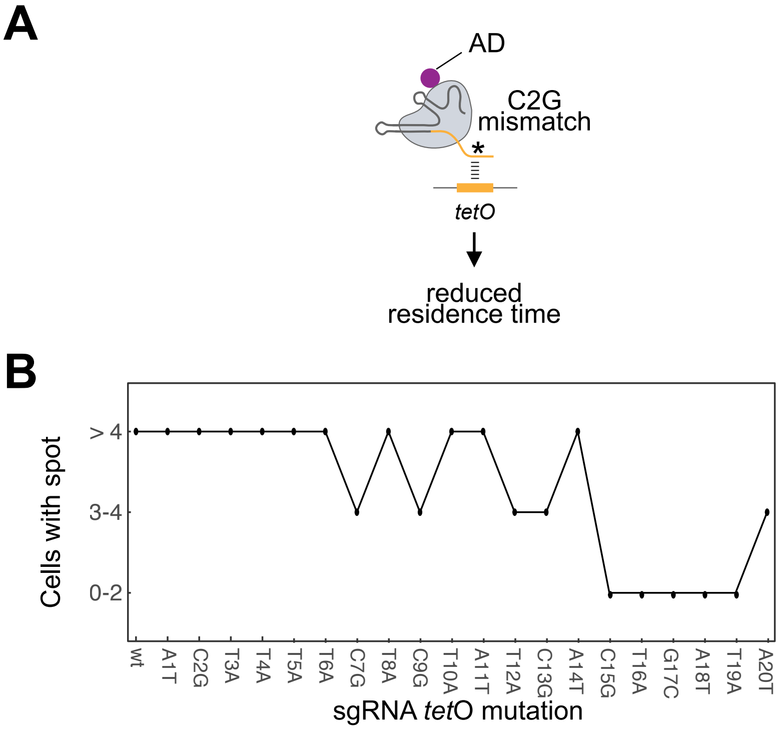


Figure 2.28: Reducing the binding strength of the DBD by a single nucleotide mutation. **(A)** Experimental strategy: mutation of a single nucleotide in the targeting region of the sgRNA leads to reduced residence times. Depending on expression levels and the resulting affinity a high occupancy can be preserved. **(B)** Screening for a mutation that preserves targeting to the *tetO* sites by mutating each base in the sgRNA targeting sequence by A<>T and C<>G replacement. Recruitment of dCas9-EGFP-VPR to the *tetO* cluster was analyzed by microscopy. The number of cells with detectable spot per inspected region for each mutated sgRNA is shown.

The effect of binding stabilization and prolonged residence times on transcription remained unclear. In order to reduce the residence time of a complex with minimal changes to its composition a single base mutation was introduced into the tetO-2xPP7-guide RNA. Each of the 20 bases in the targeting region of the guide RNA was systematically mutated and the resulting set of guide RNAs in complex with dCas9-EGFP-VPR was screened for the remaining ability to bind to the cluster of *tetO* binding sites (**Figure 2.28 A**).

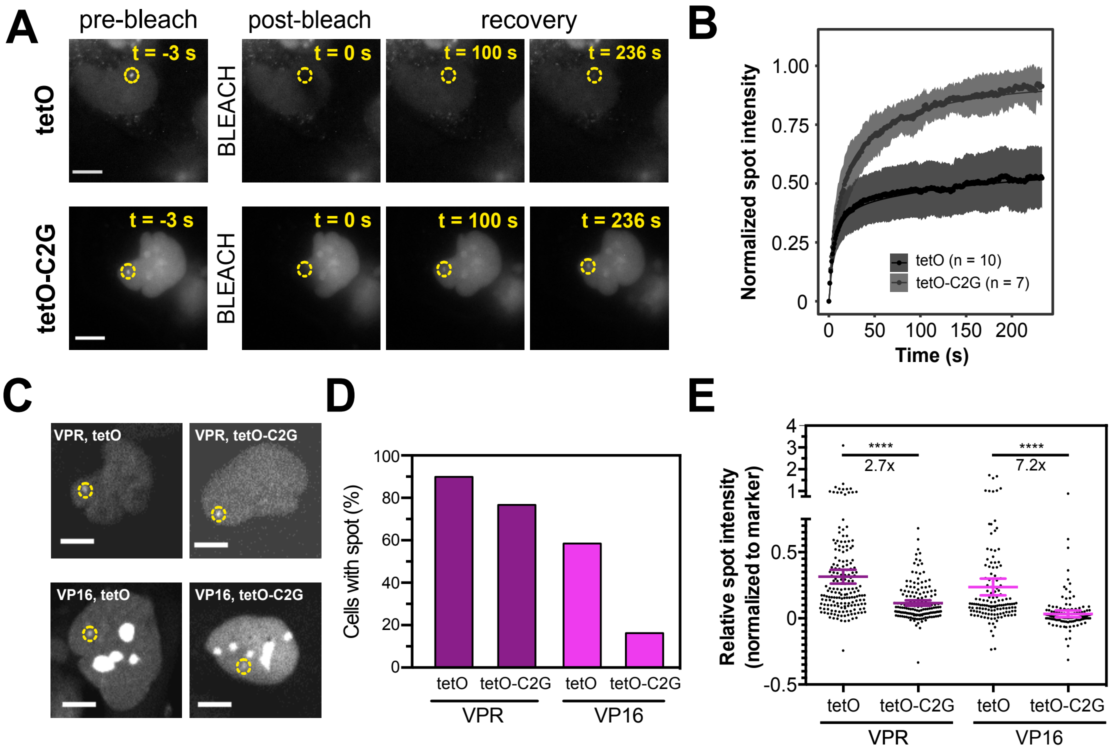


Figure 2.29: Reduced residence time and binding site occupancy by C2G sgRNA mutation. **(A)** FRAP image series of dCas9-EGFP-VPR with (tetO-C2G) and without (tetO) mismatch mutation. Scale bar: 10 µm. **(B)** Average recovery curves, fit (solid line) and 95% CI (shaded area). **(C-E)** Recruitment of dCas9-EGFP-VP16/-VPR to *tetO* binding site cluster with and without sgRNA mutation. Spot area was defined by the tagBFP-LacI marker. **(C)** Exemplary images. Note that there is strong nucleolar enrichment of dCas9-EGFP-VP16 complicating the detection of spots with weak enrichment. Scale bar: 10 µm. **(D)** Fraction of cells that exhibit visible enrichment of dCas9-EGFP-VP16/-VPR intensity in the spot area. **(E)** Quantification of dCas9-EGFP-VP16/-VPR intensity at reporter spot normalized to tagBFP-LacI marker intensity. Dots: single cell values, bars: mean, error bars: 95% CI, n = 127 -175 cells per condition, ****: $p < 0.0001$ two-sided unpaired Student's t-test.

The effect of mismatch mutations in the guide RNA has been explored by several studies, but these typically used a readout that is based on the cutting of a target sequence [123, 201, 202]. The readout used here, in contrast, directly measures the binding and not the activity of the dCas9 complexes. The effect of the mutants ranged from unchanged targeting frequency to a complete loss of targeting (**Figure 2.28 B**). This was assessed by counting the number of cells with visible dCas9 recruitment to the *tetO* cluster. One mismatch guide RNA, tetO-2xPP7_C2G, which could in complex with dCas9-EGFP-VPR still successfully target the reporter gene was selected and its residence time was determined by FRAP.

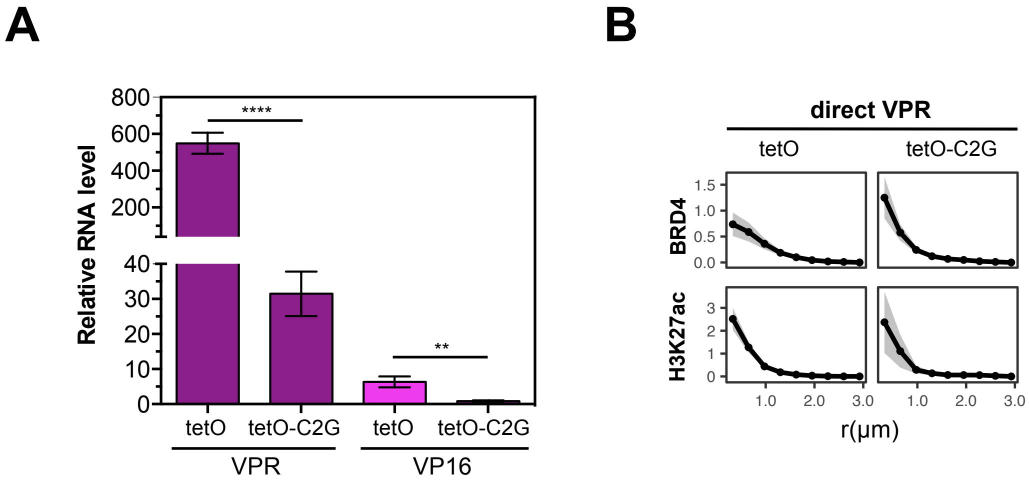


Figure 2.30: Reduced binding mediated by C2G sgRNA led to strongly reduced or abolished transcription, but did not prevent co-activation. **(A)** Bulk RNA induced by dCas9-EGFP-VP16/-VPR measured by qRT-PCR. The data for the non-mutated sgRNA corresponds to Figure 2.21 and is shown for comparison. RNA levels are relative to mock transfection. (qRT-PCR experiments were performed and analyzed by Lukas Frank.) Error bars: $\pm 1 s.d.$, n = 3 replicates, **: $p < 0.01$, ****: $p < 0.0001$, two-sided unpaired Student's t-test. **(B)** Intensity enrichment profiles for co-activation marks BRD4 and H3K27ac. Profiles for non-mutated sgRNA correspond to **Figure 2.22** and are shown for comparison. Note that enrichment values cannot be compared quantitatively but represent a qualitative readout for the presence or absence of intensity enrichment. n = 28 -184 cells per condition, shaded area: 95% CI.

Due to the lower copy number of *tetO* sites compared to the *lacO* sites the widefield FRAP approach was adapted by using a 100x objective for higher magnification (**Figure 2.29 A**). The C2G mutation reduced the residence time on the tetO sites from 127 s to

57 s and also the immobile fraction was reduced from 36 % to 7 % (**Figure 2.29 B**).

The local concentration of dCas9-EGFP-VPR at the reporter cluster with the mutated guide RNA (**Figure 2.29 C**), was reduced. The fraction of cells (with tagBFP-LacI marker) with visible dCas9 fusion recruitment was reduced from 90 % to 76 % for VPR and 59 % to 17 % for VP16 (**Figure 2.29 D**).

The relative spot intensity normalized to the marker tagBFP-LacI , was reduced to 37 % of the intensities for the matching guide RNA, whereas it was reduced to 14 % for dCas9-EGFP-VP16 (**Figure 2.29 E, Table 8**). In many cells this corresponded to undetectable enrichment of dCas9-EGFP-VP16, even though in some cells weak recruitment could still be seen. These observations suggests that direct DNA binding is strongly reduced, but that this effect can be compensated, if there are multivalent interactions that stabilize the complex.

As expected the reduced residence time and occupancy also affected the transcription levels. For dCas9-EGFP-VPR the guide RNA mutation led to a strong reduction of reporter RNA induction from 550-fold to 31-fold over mock in RT-PCR, whereas for dCas9-EGFP-VP16 it completely abolished transcriptional induction (6.4-fold to 0.9-fold, **Figure 2.30 A**). The (non-quantitative) intensity enrichment profiles for dCas9-EGFP-VPR (**Figure 2.30 A**) confirmed that under reduced residence times co-activation via histone acetylation and BRD4 recruitment still takes place. It is concluded that a reduction of direct DNA binding by the DBD can strongly reduce transcription and that this effect can partly be compensated by multivalent interactions of the AD. Strikingly, dCas9-EGFP-VP16 did accumulate visibly at the reporter spot at least in a subpopulation of cells (14 %) but could not induce transcription. The residence time could not be measured for dCas9-EGFP-VP16 because the weak signal enrichment at the reporter spot in conjunction with strong nucleolar enrichment prevented a robust quantification of spot intensity in FRAP image series. Still, it is reasonable to assume that residence times will be reduced by the C2G mutation of the guide RNA for VP16 just as for VPR. Thus, the inability of dCas9-EGFP-VP16 to induce transcription is a first indication that short residence times can prevent activation, even if the activation domain gets enriched at the promoter.

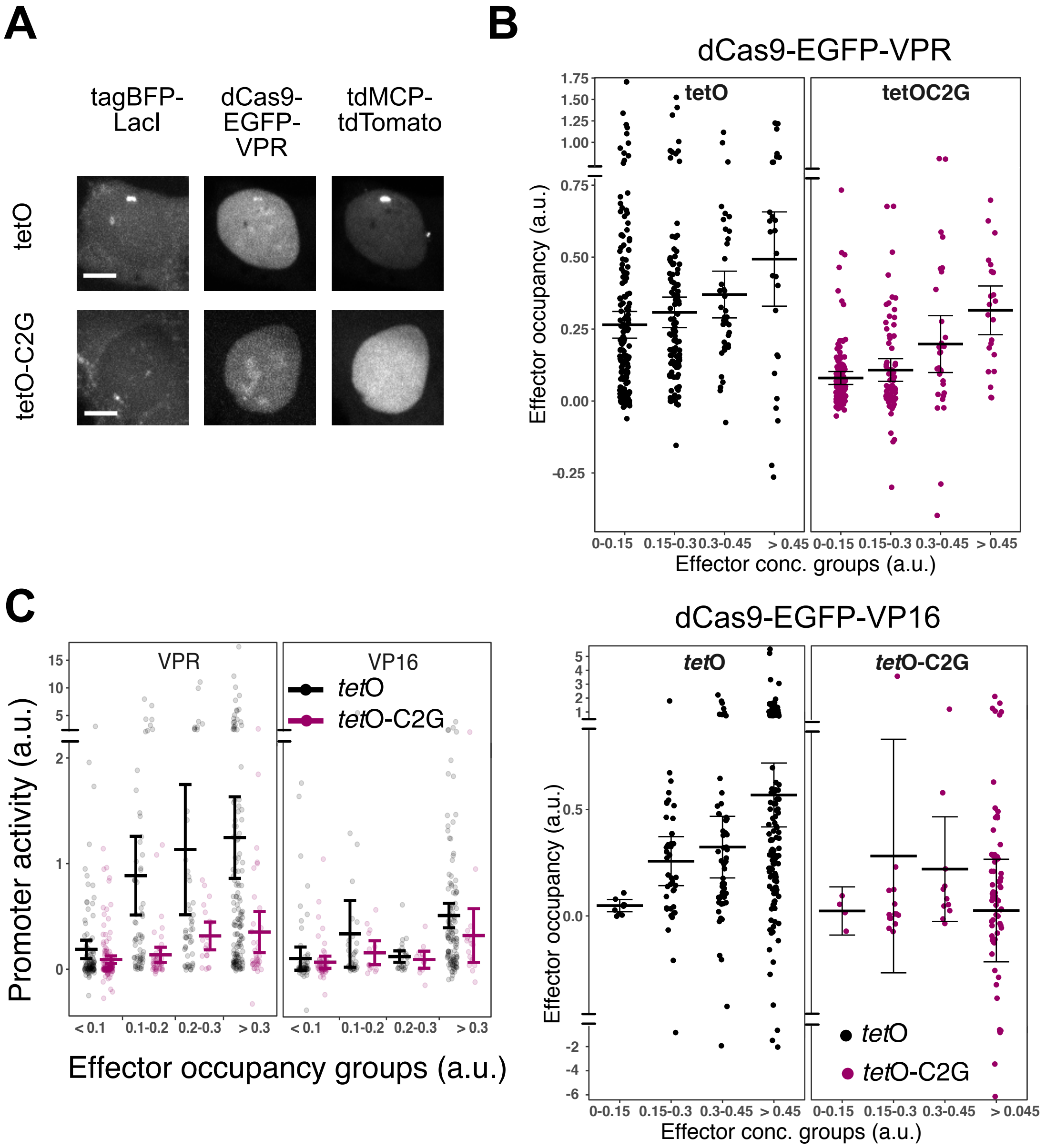


Figure 2.31: Effect of residence time was independent of occupancy. Binding site occupancy of dCas9-EGFP-VP16/-VPR and nascent RNA labeled with tdMCP-tdTomato were measured in single cells by microscopy. Grouping of cells allowed to compare cells with similiar TF expression and occupancy levels. **(A)** Representative images of array marker, TF complex and nascent RNA with and without sgRNA mutation. Scale bar: 10 µm. **(B)** Effector spot intensities normalized to the marker intensities for cells grouped by their nuclear effector intensity. **(C)** Nascent RNA levels normalized to the marker channel for cells grouped by the effector occupancy (effector spot intensity normalized to marker). Dots: single cell values, bars: mean, error bars: 95 % CI.

The guide RNA mutation reduced both residence time and occupancy. In principle, for high affinities (due to high on-rates) and at sufficiently high TF expression levels it should be possible to reduce the TF residence time while maintaining full occupancy, if binding site saturation is in the plateau region (compare **Figure 2.27 B**). This was not the case for the C2G mutation as the binding site occupancy was reduced (**Figure 2.29 C-E**).

In order to study the effect of the residence time independent of the reduced occupancy, the transcriptional induction was measured in single cells by microscopy to make use of the heterogeneous TF expression levels caused by the transient transfection. Cells were transfected with the respective synthetic TF constructs and the matching (tetO) or mutated (tetO-C2G) sgRNAs and nascent RNA levels were measured as the accumulated tdMCP-tdTomato intensity at the reporter gene cluster spot labeled by tagBFP-LacI (**Figure 2.31 A**). Intensities of the EGFP-labeled dCas9 activation complex and of tdTomato labeled RNA were normalized to the tagBFP intensity to reduce inter-cellular heterogeneity caused by variable reporter gene copy numbers. This approach allowed to group cells according to their TF expression level or according to the binding site occupancy.

The rising occupancy of binding sites with increasing TF expression levels could be followed in cells grouped by their nuclear TF intensity (**Figure 2.31 B**). There was a continuous rise of local TF intensity at the reporter gene spot with increasing nuclear concentration in all conditions, except for dCas9-EGFP-VP16 in complex with the mutated tetO-C2G guide RNA. The irregular dependency of the measured binding site occupancy on expression levels for the combination of mutated RNA and VP16 is likely due to its weak enrichment and accumulation in nucleoli that distorts intensity measurements. Note that negative values can appear, if the local background is greater than the intensity in the spot. The long residence times of the dCas9 complexes (see section 2.4.1) raises the expectation that binding should saturate and reach a plateau of occupancy at low expression levels at least for the matching tetO-guide RNA. This was not the case here. Potentially, a part of the increasing local concentration could be caused by indirectly recruited molecules, which does not saturate at increasing expression levels. Comparing occupancy levels between matching and mutated guide RNA, there was no clear trend for VP16, but for VPR occupancy with the mismatch guide RNA was lower for cells with the same TF expression level. Importantly, there was a wide range of occupancy values within and across expression level groups and there was a large overlapping range of occupancy values for the matching and mutated guide RNA. This allowed to compare promoter activities for guide RNAs with different residence times in cells with comparable occupancies (**Figure 2.31 C**). As expected, nascent RNA levels rose with rising occupancy for VPR and in most cases also for VP16. Importantly, nascent RNA levels were lower for the mutated tetO-C2G guide RNA compared to the matching tetO-guide RNA in every occupancy group. This held true for both VPR and VP16. This leads to the conclusion that nascent RNA production is regulated not only by TF occupancy, but also independently by TF residence time. While a kinetic proofreading mechanism, based for instance on activating TF modifications, was not directly verified, it could provide an attractive explanation for the observed residence time dependency.

In summary, a role for TF residence time independent of binding site occupancy was established. In extreme cases this can lead to a situation, in which the high-turnover TF is visibly enriched at the promoter, but unable to activate transcription. Moreover, TFs with strong multivalent interactions, like VPR, seem to be more robust to a reduction of direct DNA binding, e.g. by DBD mutations. TF-TF interaction could be responsible for the stabilization. The reduced residence time and binding site occupancy lead to reduced transcription also for these TFs, but not to a complete shut-down of transcription as for low-interaction TFs, like VP16.

2.5 Multivalent interactions of the AD and transcription dynamics

The strength of transcriptional activation varies between the activation domains (ADs) of TFs [127]. It remains unclear which properties of the AD determine activation strength [203], which is partly due to the lack of sequence conservation [82] that would allow to classify ADs into families as has been done for the DBDs. Their intrinsically disordered structure does not rely on single amino acid residues that could explain their activity [86]. However, the propensity to form phase-separated compartments has emerged as an important TF property [39] and could represent a key feature that determines activation strength [100]. Potential mechanisms how multivalent interactions could enhance transcription range from reduction of target search times [99], enrichment of factors in transcription compartments [40] to facilitated interactions of genomic loci in these compartments. Besides the activation strength that describes the amount of RNA that is induced by an AD also the speed with which RNA production is induced could vary between ADs. This speed could be crucial, for instance, for the induction of antiviral genes by the innate immune system that needs to be faster than the induction of competing transcription programs by viral TFs to avoid spreading of an infection. The ability of ADs to induce transcription fast has been much less characterized for ADs than their strength at later time points. Here, the protocols to acquire time-resolved transcription induction time courses that have been established in section 2.1 were applied to compare the activation strength and speed of a well-characterized set of ADs. Their phase-separation propensities that reflects the strength of their multivalent interactions have been determined (section 2.3.2). The real time acquisition of transcription time courses by microscopy allows not only to measure activation strength and speed, but also distinguish between single cells with and without phase-separated compartments. Moreover, the serendipitous discovery that CIBN-LacI enhances optodroplet formation (section 2.3.2) has opened a way to increase the phase-separation propensity of low-interaction ADs.

2.5.1 Multivalent interaction strength of AD correlates with fast and strong transcriptional induction

Does the propensity to self-interact make an AD a better activator? VPR and its two sub-components Rta and p65 have already been reported to lead to stronger transcriptional activation than VP64 that is related to VP16 [127], but the activation strength and dynamics of these ADs and STAT2 had not been determined for the reporter cell line studied here. The induction dynamics and strength were measured by acquiring light-induced time courses of nascent reporter RNA labeled by tdMCP-tdTomato (**Figure 2.32 A**). The AD was recruited to the promoter by the light-induced binding of PHR-EGFP-AD to CIBN-rTetR. PHR-EGFP-NLS containing only the nuclear localization signal (NLS) that is also present in the other ADs was included as a negative control. The acquired image series allowed to judge activation strength, but also the presence or absence of TF optodroplets (**Figure 2.32 B**). Cells were individually classified as responders or non-responders based on visible enrichment of tdMCP-tdTomato at the reporter gene spot. This revealed that a smaller fraction of cells responded (**Figure 2.32 C**) to recruitment of STAT2 and VP16 (42-67%) than to Rta, p65 and VPR (67-92%). A small fraction of cells challenged with the mock effector PHR-EGFP-NLS was annotated as responders due to weak intensity enrichment at the reporter spot in at least one time frame. However, nascent RNA levels showed only a minimal rise in response to the mock effector (**Figure 2.32 D, E**). Single-cell RNA time courses (**Figure 2.32 D**) exhibited a large variability within ADs, but as a general trend there was a more shallow, continuous rise for VP16 and STAT2, while several curves reached a plateau early for p65, Rta and VPR. Average curves confirmed that the droplet forming ADs Rta, p65 and VPR showed a steeper initial rise and reached higher plateau values of nascent RNA (**Figure 2.32 E**). This held true, if only including cells classified as responders or if averaging over all cells including non-responders.

It is noted that due to the large number of conditions tested (with 2-3 replicates each) not all experiments could be performed in one session. Instead experiments were distributed on two sessions. It cannot be excluded that the intensity measurements in the different session could show a systematic difference, e.g. due to aging of the excitation light source. However, the comparison within one experimental session (VP16 vs. p65 and STAT2 vs. Rta and VPR) consistently showed higher plateau values for the droplet forming ADs. The activation dynamics should not be influenced by (potentially) different fluorescence detection efficiencies.

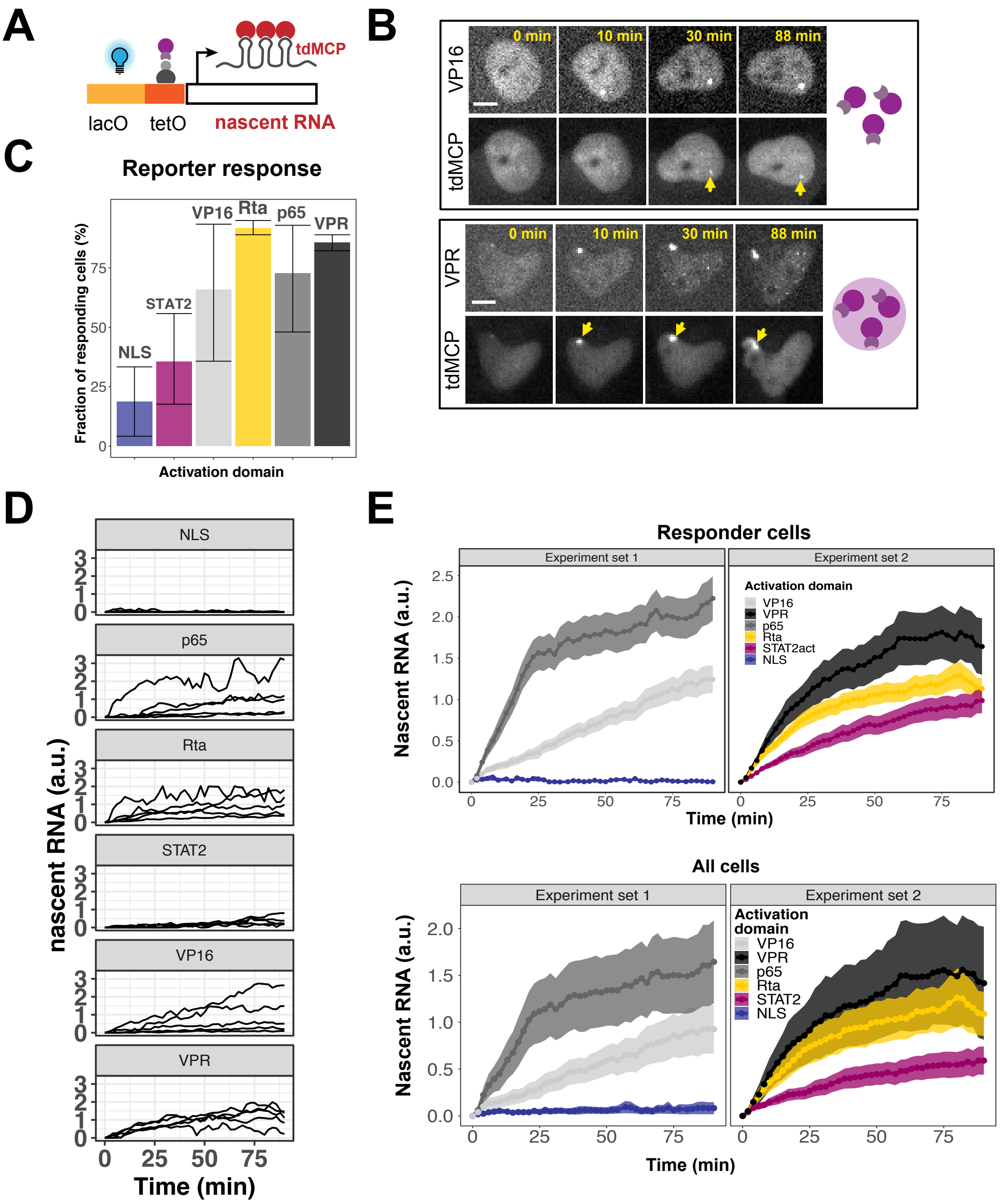


Figure 2.32: ADs induced different levels of promoter activity **(A)** Experimental strategy: PHR-EGFP-AD recruitment to CIBN-rTetR at the promoter was triggered by light and nascent RNA levels were followed by tdMCP-tdTomato intensity at the promoter spot. **(B)** Representative image series of PHR-EGFP-AD and tdMCP-tdTomato for ADs with low (VP16) and high (VPR) propensity to form optodroplets. Scale bar: 10 µm. **(C)** Fraction of cells classified as responders. Error bars: minimum and maximum of two to three replicate experiments. (Continued on next page).

Average normalized time courses showed that the three droplet forming ADs VPR, p65 and Rta activated moderately faster than VP16 and STAT2 (**Figure 2.33 A**). There was a steeper rise of nascent RNA levels in the first 30 min. The shift to earlier activation can also be seen in the full population of single-cell trajectories of the droplet forming ADs compared to VP16 and STAT2 (**Figure 2.33 B**). To compare the activation dynamics of single-cells the time at which the single-cell trajectory reached half of its maximum value was determined. The time to half-maximal activation was longer for activators with low self-interaction VP16 and STAT2 (42/38 min) than for self-interacting VPR, p65 and Rta (28/26/28 min) (**Table 9**). There was, thus, a correlation of the ability of an AD to self-interact and its activation speed. This is visualized by the relation of the critical concentration at which a PHR-AD fusion forms optodroplets (section 2.3.2) and the time to half-activation of the reporter gene (**Figure 2.33 C**). The higher propensity to form optodroplets of PHR-EGFP-Rta, -p65, and -VPR (characterized by a lower critical concentration) coincided with a faster activation (indicated by a shorter time to half-maximum activation).

The observed favorable effect of multivalent TF interactions on activation speed and strength could be mediated by one or several of the following mechanisms: (1) formation of a phase-separated compartment at the promoter leading to the increase of local TF concentration, (2) stabilization of direct chromatin binding leading to longer residence times and higher binding site occupancy or (3) interactions with other (co-)activator molecules containing intrinsically disordered regions (IDRs).

There were subtle differences in the dynamics of TF recruitment. For all PHR constructs recruitment was somewhat more prolonged than would be expected from the PHR-CIBN binding kinetics that should saturate within the first minute (**Figure 2.33 D**). Instead the initial recruitment took around 10 min which could be explained by the pulsatile light exposure (150 ms every 2 min and stray light from neighboring positions).

Figure 2.32: (Continued) **(D)** Exemplary nascent RNA time courses of single cells. **(E)** Mean absolute intensity time courses. Activation time courses of the different PHR-EGFP-AD constructs were acquired in two separate microscopy sessions and absolute values should be only compared within one set. Induction curves are shown for responder cells (top) and for all cells also including cells classified as non-responders (bottom).

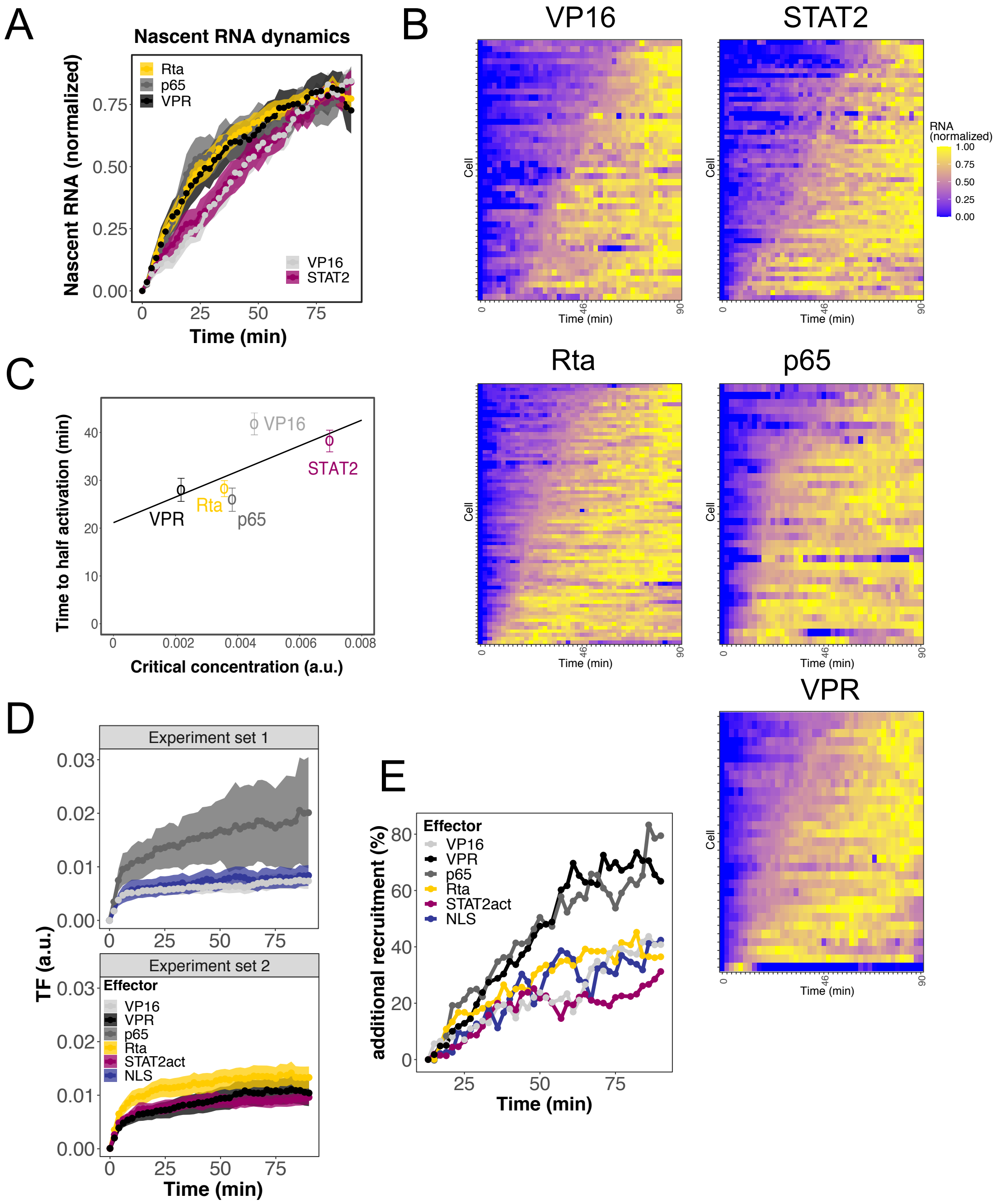


Figure 2.33: Normalized time courses reveal different activation dynamics of ADs. **(A)** Mean normalized intensity time courses of nascent RNA for cells classified as responders. Shaded area: 95% CI. **(B)** Color-coded normalized single-cell RNA time courses sorted by their time to half-activation. **(C)** Correlation of TF multivalent interactions (characterized by small critical values) and activation speed (characterized by short time to half-activation). Error bars: s.e.m. **(D)** Mean TF recruitment time courses. Shaded area: 95 % CI. **(E)** Additional TF recruitment after the 12 min time point. Additional recruitment is indicated relative to this time point.

Nevertheless, the major part of TF recruitment took place in the first minutes. This first recruitment phase was followed by a prolonged phase of ongoing slower recruitment, which was more pronounced for VPR and p65 (around 70 % additional recruitment in the second phase compared to around 40% for the other ADs, **Figure 2.33 E**). This could potentially be caused by an increasing binding site accessibility as a result of the transcriptional activation process (as has been discussed already for reactivation experiments in section 2.1.6). The stronger additional gain of TF binding for VPR and p65 could mean that these ADs lead to a stronger induction of this (co-)activation pathway. It seems less likely that the additional recruitment could be caused by the multivalent interactions of ADs (e.g. by fusion of optodroplets), because optodroplet-forming PHR-EGFP-Rta showed similar recruitment as the VP16, STAT2 and NLS constructs.

2.5.2 Formation of phase-separated transcription compartments does not enhance transcription

There was a stronger and faster activation of transcription and a larger fraction of responding cells for ADs that showed a high propensity to phase separate in the optodroplet assay. This correlation could be directly caused by the phase-separated compartments. It could, however, also be caused by the underlying stronger multivalent interactions, so that the actual formation of (opto-)droplets is not necessary for the transcription enhancing effect. To address the involvement of phase-separated compartments, the induction time course data was re-analyzed by comparing cells with and without droplet formation for the same PHR-AD construct. Surprisingly, no improved activation was found, but rather a potential repressive effect of droplet formation. In averaged normalized time courses activation was moderately delayed in cells with droplet formation for VPR, as well as for Rta, while it remained unchanged for p65 (**Figure 2.34 A**). The mean time to half-maximal activation was increased from 25 min without droplets to 30 min for VPR (Rta: 25 to 31 min, p65: unchanged 26 min) (**Figure 2.34 B**).

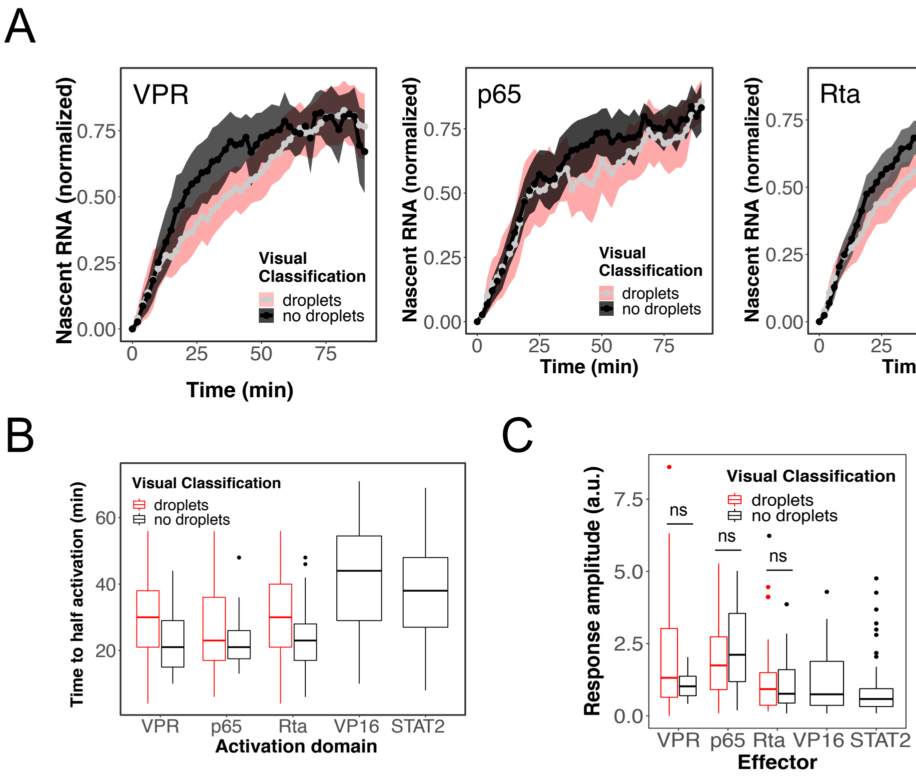


Figure 2.34: Cells with macroscopic phase-separated PHR-AD optodroplets do not activate faster. **(A)** Mean normalized nascent RNA time courses of experiments shown in Figure 2.33 separated into cells with and without optodroplets for VPR (left), Rta (center) and p65 (right). Shaded area: 95% CI. **(B)** Distribution of times to half-activation with and without optodroplets. **(C)** Distribution of plateau values of absolute nascent RNA trajectories with and without optodroplets. ns: not significant ($p > 0.05$), two-sided unpaired Welch's t-test.

While the time to half activation was clearly influenced by the effector ($p < 0.0001$, two-way ANOVA), there was a less clear trend towards slower activation effect for droplet containing cells ($p < 0.1$). The effect of droplet formation on the activation time course plateau value was mixed, increasing the amplitude for VPR and reducing it for p65 and Rta (**Figure 2.34 C**). Across the ADs there was no increase of plateau values in cells with droplet formation ($p > 0.1$, two-way ANOVA accounting for effector and presence of droplets). Overall, there was no enhancement of activation in droplet forming cells and the observed slower activation dynamics raised the possibility of an inhibitory effect of droplet formation.

2.5.3 Inducing the formation of a phase-separated compartment has a repressive effect on transcription

The previous experiments had shown that the actual formation of phase-separated TF compartments did not increase transcription for ADs exhibiting strong multivalent interactions and that it could maybe even have a mildly repressive effect on the induction dynamics. This showed that macroscopically visible optodroplets were not the driver of the higher activation potential of high-interaction ADs VPR, p65 and Rta. Nevertheless, it remained unclear, if droplets really were a side-effect of the AD's strong multivalent interactions with inhibitory consequences, or if there still was a functional role for phase-separated TF compartments. In order to go beyond correlation, three strategies were developed to enforce droplet formation for an AD that typically does not form optodroplets. VP16 was chosen as an AD with weak phase separation potential. The strategies relied on either direct fusions of an IDR to the AD or on bridging factors to increase the strength and valency of TF interactions. The first approach used CIBN-LacI as a bridging factor, the second approach used a PHR-GFP binding protein (GBP) fusion as a bridging factor and the third approach increased multivalent interaction by fusion of the high interaction FUS domain.

In the first approach the ability of CIBN-LacI to lower the critical concentration for optodroplet formation by PHR-EGFP-VP16 (section 2.3.2) was exploited. It served as a bridging factor, presumably increasing interaction valency by binding to PHR via the CIBN domain and by forming dimers or tetramers of the LacI domain. It was co-transfected with CIBN-rTetR and PHR-EGFP-VP16 (**Figure 2.35 A**) and transcription activation time courses were recorded. An additional control condition was included to account for a potential interference by disturbing the chromatin structure at *lacO* sites: EGFP-LacI recruited to *lacO* sites in combination with PHR-EGFP-VP16 recruited to CIBN-rTetR. The approach inevitably leads to additional TF recruitment to the promoter because both *tetO* and *lacO* sites are targeted. But recruitment of VP16 to the distal *lacO* sites of the reporter gene by CIBN-LacI alone induced transcription only in a small fraction of cells (18 %) and at a low level (**Figure 2.36 A, B**). CIBN-LacI successfully and strogly increased the fraction of cells with VP16 optodroplets (**Figure 2.35 B, C**). For VPR, which has a strong propensity for optodroplet formation on its own, no additional droplet formation was observed.

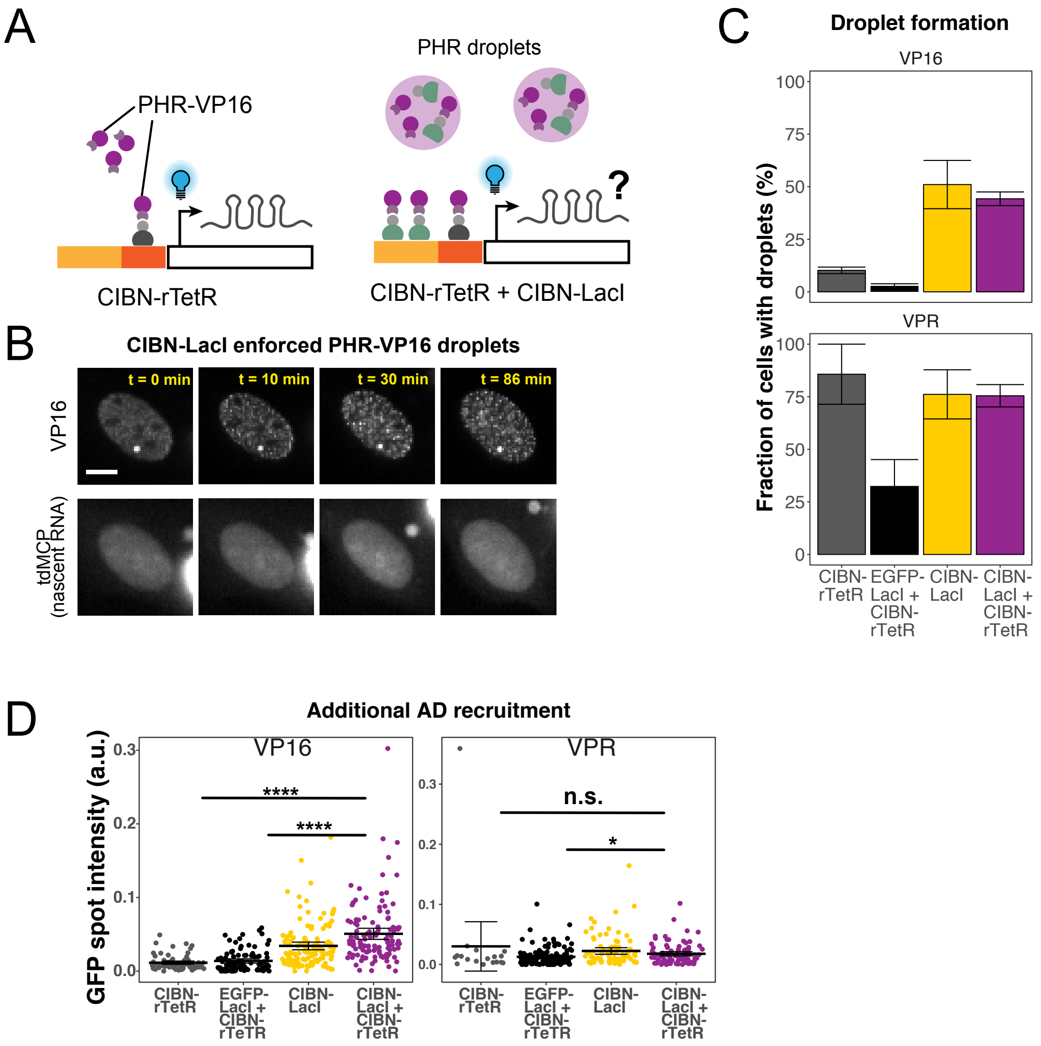


Figure 2.35: Inducing VP16 optodroplets by interactions with CIBN-LacI. **(A)** Experimental strategy: PHR-EGFP-VP16 was recruited to CIBN-rTetR and CIBN-LacI. The combination of PHR-EGFP-VP16 and CIBN-LacI leads to increased optodroplet formation (*see* Figure 2.20). Recruitment to CIBN-rTetR or CIBN-LacI alone or the combination of CIBN-rTetR with activation-inert EGFP-LacI were included as controls. **(B)** Image series of PHR-EGFP-VP16 recruited to CIBN-LacI and CIBN-rTetR. In this cell no transcription was induced. Scale bar: 10 µm. **(C)** Fraction of cells in time course experiments that showed visible optodroplets. Error bars: minimum and maximum of experiments. **(D)** Amount of EGFP-labeled activator recruited to the promoter cluster. Dots: single cell values, bars: mean, error bars: 95% CI, n.s.: not significant, *: $p < 0.05$,****: $p < 0.0001$, two-sided unpaired Student's t-test.

Importantly, CIBN-LacI by increasing the phase separation propensity of PHR-EGFP-VP16 not only led to the formation optodroplets in the nucleus but also to larger amounts of the TF at the reporter gene cluster (**Figure 2.35 D**). This was also the case in comparison with EGFP-LacI and CIBN-rTetR suggesting that the additional TF recruitment cannot be explained by targeting more binding sites. Rather, there was a phase-separated compartment formed at the promoter cluster. In this setting activation in the presence and absence of optodroplets could now be compared using the same AD.

The additional presence of CIBN-LacI reduced the fraction of responding cells compared to recruitment to CIBN-rTetR alone from 70 to 24% and led to a 5.4-fold reduction of the response amplitude in the average across responding and non-responding cells (**Figure 2.36 A-E**). Confining the comparison to responding cells there was a 2.8-fold reduction.

However, the transcriptional response was also inhibited to a lesser degree by the binding of EGFP-LacI to *lacO* sites alone. This raised the question whether repression by optodroplets could be detected independent of repression by *lacO* binding. Comparing activation via CIBN-rTetR in the presence of CIBN-LacI to EGFP-LacI the responder fraction was reduced from 42 to 24% and the RNA plateau value was reduced 2.1-fold ($p < 0.01$, two-sided Student's t-test) across all cells and 1.8-fold ($0.05 < p < 0.1$) for responder cells (**Table 10**). Thus, there was repression by CIBN-LacI induced droplets, but this was achieved at least in part by a reduction of responding cells, whereas the response amplitude of responder cells was less affected. Notably, the activation time courses with CIBN-LacI only started to diverge from the ones without after the first few minutes indicating that the response of poised promoters could possibly be not affected. For VPR a similar trend could be observed, even though its phase separation propensity was not further increased. The responder fraction was reduced from 78 to 32%, the response amplitude was reduced across all cells 5.6-fold ($p < 0.05$), and for responder cells 6.3-fold ($p < 0.05$). Again comparing these quantities to the mildly repressive EGFP-LacI+CIBN-rTetR condition the effect of CIBN-LacI was less pronounced but still detectable. The responder fraction decreased from 43 to 32%, the response amplitude across all cells 1.2-fold ($p > 0.05$) and for responder cells 2.7-fold ($p < 0.05$).

The reduction in RNA production was confirmed for bulk RNA levels by qRT-PCR after 90 min illumination (VP16: 2.3-fold reduction or 1.7-fold reduction compared to the EGFP-LacI control, $p < 0.01$, **Figure 2.36 F**, qRT-PCR experiments and analysis performed by Lukas Frank). These results demonstrated that phase separation of a transcriptional activator does not necessarily increase transcription, but can inhibit it, even if there is an increased local concentration of the activator at the promoter.

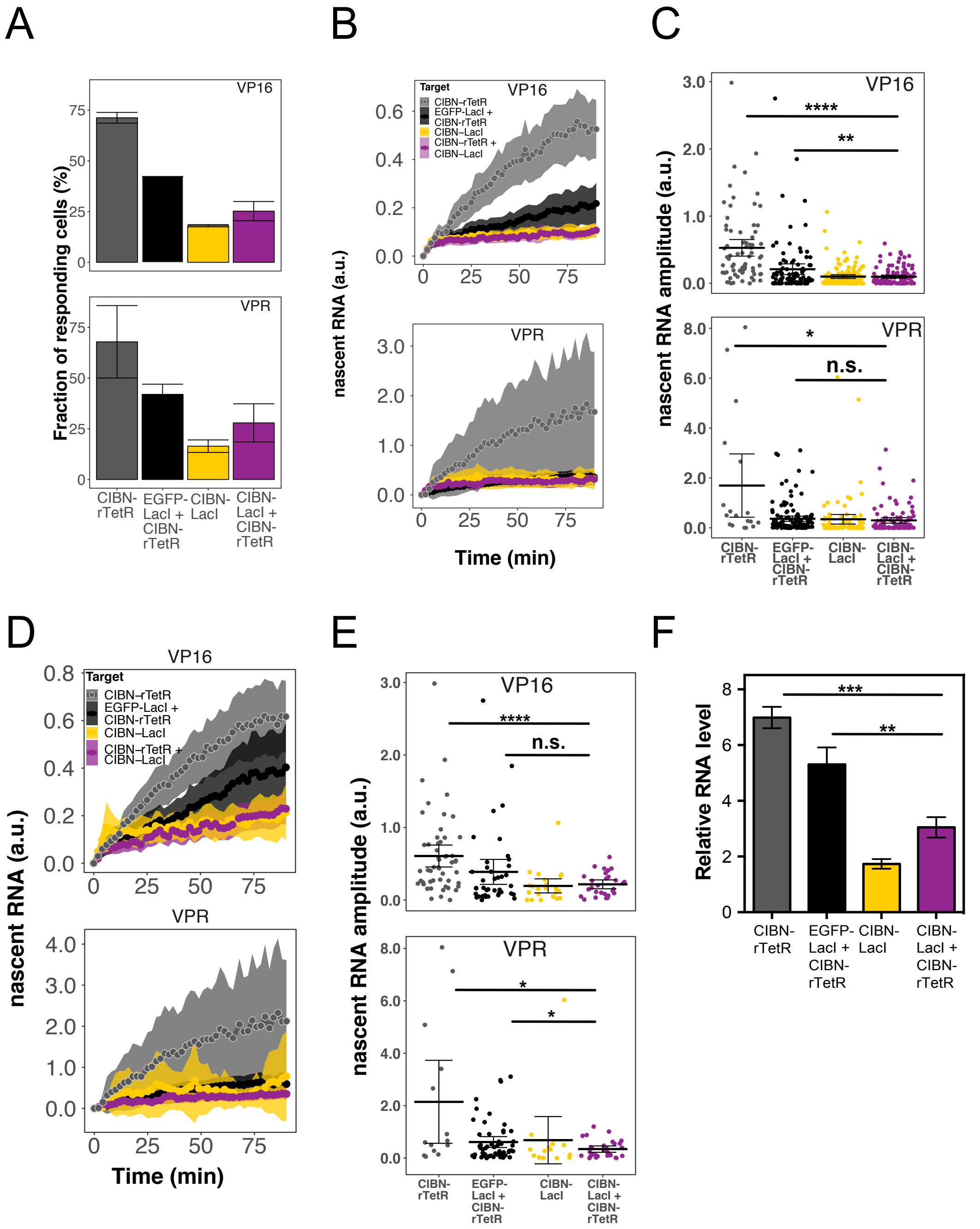


Figure 2.36: CIBN-LacI driven optodroplets suppressed transcription. **(A)** Average fraction of cells with visible enrichment of tdMCP-tdTomato at the reporter gene cluster. Error bars: maximum and minimum of two experiments. **(B)** Mean absolute nascent RNA time course for all cells including non-responders. Shaded area: 95 % CI, n= 74 -128 cells per condition. **(C)** Plateau values of individual nascent RNA trajectories. Dots: single cell values, bars: mean, error bars: 95% CI, n.s.: not significant, *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$, two-sided unpaired Welch's t-test. **(D)** Mean absolute nascent RNA time course for responder cells. Shaded area: 95 % CI, n= 14-55 cells per condition. (Continued on next page.)

Since the optodroplet induction by CIBN-LacI clearly indicated a repressive role for phase-separated transcription compartments, but was complicated by the interference by chromatin binding, two additional experimental strategies were followed to induce VP16 optodroplet formation without binding at the *lacO* sites (**Figure 2.37 F**). First, the GFP binding protein (GBP) was used to construct an alternative bridging factor PHR-GBP. This construct binds EGFP in PHR-EGFP-VP16 with high affinity resulting effectively in a complex with two PHR domains which both have the ability to participate in multivalent interactions. Cells were activated by PHR-EGFP-VP16 in the presence of PHR-GBP. It was anticipated that this would increase the phase separation propensity.

Second, a high interaction AD was constructed using VP16 and the N-terminal portion of FUS (FUSN), a protein domain that can drive oncogenic transformation, when it is fused to a TF in a genomic translocation event [111], and which is known for its high phase separation propensity [8]. This PHR-EGFP-FUSN-VP16 construct was also expected to display a strong propensity for optodroplet formation. PHR-EGFP-FUSN was included as a negative control to confirm that the FUSN domain itself does not activate transcription. In all conditions PHR was recruited to CIBN-rTetR and nascent RNA levels were recorded using tdMCP-tdTomato.

Both experimental strategies successfully induced formation of optodroplets (**Figure 2.37 B,C**). Nearly all cells in the PHR-GBP condition had optodroplets, whereas the FUSN constructs induced optodroplets in around half of transfected cells. There was a striking reduction in transcriptional induction in the presence of PHR-GBP, so that there were virtually no responder cells (**Figure 2.37 D**). As expected, FUSN did not activate transcription on its own, whereas responder fractions remained constantly high between VP16 and FUSNVP16.

Figure 2.36: (Continued) **(E)** Plateau values of individual nascent RNA trajectories for responder cells. Dots: single cell values, bars: mean, error bars: 95% CI, n.s.: not significant, *: $p < 0.05$, ****: $p < 0.0001$, two-sided unpaired Welchs's t-test. **(F)** Bulk RNA levels after 90 min illumination measured by qRT-PCR. (qRT-PCR experiments were performed and analyzed by Lukas Frank.) Error bars: $\pm 1 s.d.$, * *: $p < 0.01$, ***: $p < 0.001$, two-sided unpaired Student's t-test.

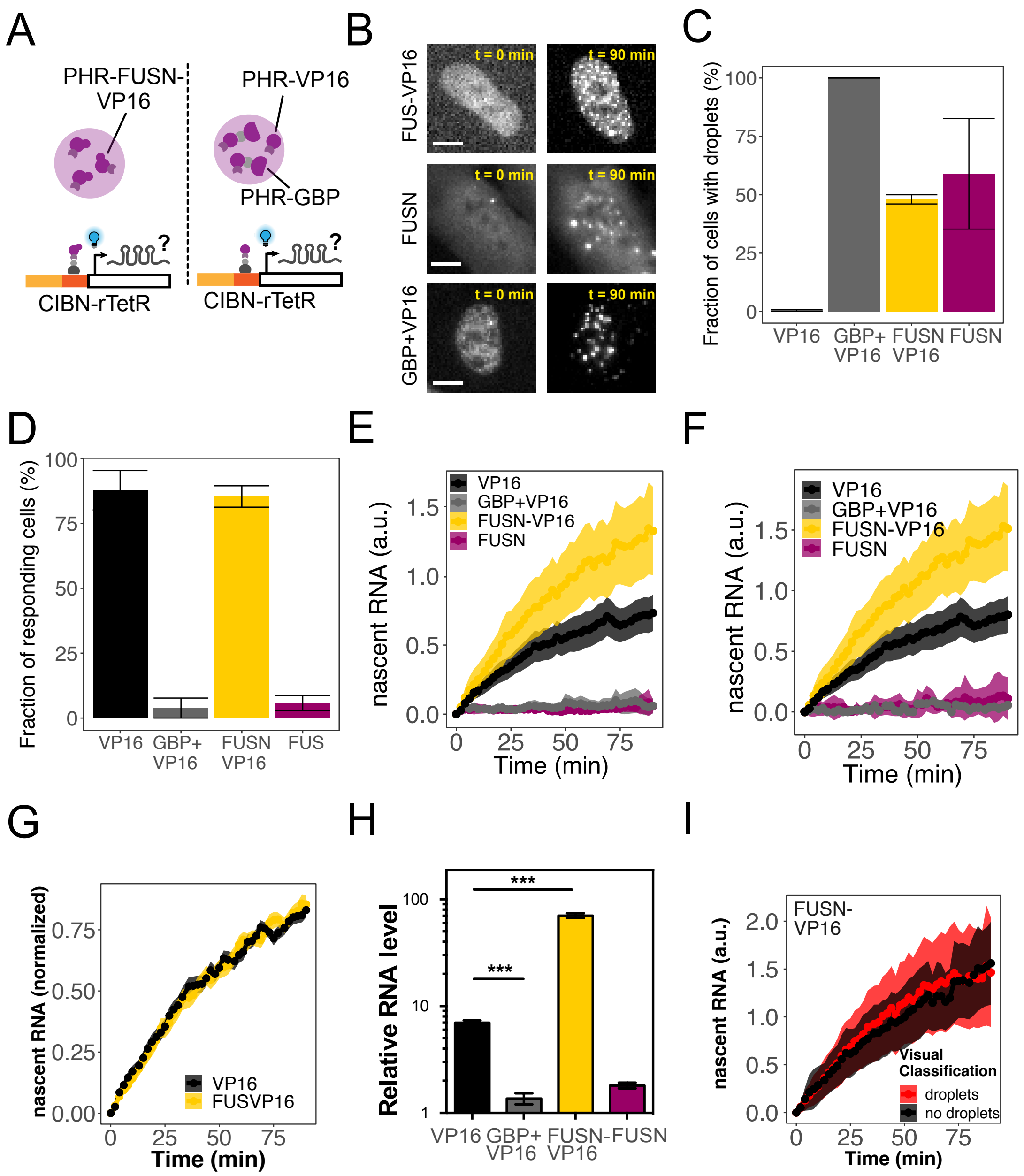


Figure 2.37: Forcing optodroplet formation of PHR-EGFP-VP16. **(A)** Forcing droplet formation by fusion to a strong multivalent interactor or by using a bridging factor: VP16 was forced to form optodroplet by fusion with a domain with strong multivalent interactions in PHR-EGFP-FUSN-VP16 (left). Alternatively, PHR-GFP binding protein (GBP) was co-transfected and bound to EGFP in PHR-EGFP-VP16 thereby creating a complex with two PHR domains that could take part in multivalent interactions. **(B)** Exemplary images at start and end of time course showing the formation of optodroplets. **(C)** Fraction of cells in time course experiments that showed visible optodroplets. Error bars: minimum and maximum of two experiments. **(D)** Average fraction of cells with visible enrichment of tdMCP-tdTomato at the reporter gene cluster. Error bars: maximum and minimum of two experiments. (Continued on next page.)

Nascent RNA time courses comprising all cells (**Figure 2.37 E**) and only responder cells (**Figure 2.37 F**) confirmed the inability of GBP+VP16 and of FUSN alone to induce transcription.

Contrarily, the fusion AD FUSVP16 showed a strong 1.8-fold increase of the time course plateau value compared to VP16 alone ($p < 0.001$, two-sided Welch's t-test). Normalized RNA time courses of responder cells (**Figure 2.37 G**) revealed no difference in the activation dynamics of VP16 and FUSVP16. These observations were also confirmed for bulk RNA levels by qRT-PCR after 90 min illumination (qRT-PCR experiments and analysis performed by Lukas Frank, **Figure 2.37 H**). PHR-GBP had a strongly repressive effect leading to a 3.8-fold reduction of RNA levels ($p < 0.001$, two-sided Student's t-test), a level only 1.9-fold above mock transfection levels, FUSN alone did not activate transcription and FUSNVP16 led to a strong 10-fold increase of bulk RNA ($p < 0.001$). It is noted that the presence of optodroplets in some cases led to problems with the identification of gene cluster sites in microscopy experiments in cases in which optodroplet formation occured as early as promoter recruitment and had similar fluorescence enrichment. This was most prevalent for the GBP+VP16 condition. However, typically the promoter spot appeared earlier with higher intensity and could be clearly identified. Moreover, the qRT-PCR data did not rely on identifying the spot region and agreed well with the microscopy data. Similar to VPR, p65 and Rta the strong activation by FUSNVP16 could be attributed directly to its strong multivalent interactions (which may help in recruiting co-activators or stabilize DNA binding) or it could rely on the formation of a phase-separated transcription compartment.

Figure 2.37: (Continued) **(E)** Mean absolute nascent RNA time course for all cells including non-responders. Shaded area: 95 % CI, n = 24 - 154 cells per condition. **(F)** Mean absolute nascent RNA time course for responder cells. Shaded area: 95 % CI, n = 94-130 cells per condition (GBP+VP16: 1 , FUSN: 3 responder cells). **(G)** Normalized mean nascent RNA time courses for responder cells (VP16 and FUSN-VP16). Shaded area: 95 % CI. **(H)** Bulk RNA levels after 24 h illumination measured by qRT-PCR. (qRT-PCR experiments were performed and analyzed by Lukas Frank.) ***: $p < 0.001$, two-sided unpaired Student's t-test. **(I)** Normalized mean nascent RNA time courses for FUSNVP16 separated into cells with and without optodroplets. Shaded area: 95 % CI.

In order to dissect the contribution of optodroplet formation to a strong activation the FUSNVP16 microscopy experiment was re-analyzed. The nascent RNA time courses were grouped by their classification as droplet containing or not (**Figure 2.37 H**). There was no difference in the absolute RNA levels or in the activation dynamics. This allowed the following conclusions: Increasing multivalent interaction by fusion of FUSN strongly boosted the activation capacity of low-interaction VP16. This was, however, not dependent on the formation of macroscopically observable compartments, but probably due to TF complex stabilization or interactions with the transcription (co-activation) machinery. In contrast, enforcing droplets using CIBN-LacI or PHR-GBP as bridging factors led to the formation of a different kind of compartment which had a repressive effect on transcription. This could be due to the biophysical properties of these condensates, e.g. a decreased mobility of activator molecules in the highly inter-connected molecular network, or to the sequestration of activator molecules at non-promoter localized optodroplets. Overall, these findings reveal a beneficial effect of multivalent interactions of the AD, which is, however, not dependent on condensate formation, and a potential to repress transcription by certain condensate compositions.

2.6 The role of co-activators in transcriptional induction by ADs with strong multivalent interactions

The previous sections have demonstrated that there can be enrichment of co-activators at the promoter without inducing productive transcription, that co-activators can potentially be recruited by TFs which are only weakly or indirectly bound to their target DNA site and that there is a correlation of TF multivalent interactions, coactivator recruitment and the strength of transcriptional induction. All these findings hint towards a co-activation pathway that is partly uncoupled from the induction of the core transcriptional machinery and relies more on multivalent interactions than on firmly and specifically bound promoter sequences. In this section the existence of an independent H3K27ac-BRD4 co-activation pathway that relies on multivalent interactions of the TF rather than core transcriptional processes is investigated. The experiments relied in part on the transcription incompetent VPR opto and optoloop complexes that allowed to observe co-activation in real time without activating transcription and which were characterized in further detail. Recruitment of the histone acetyl transferase p300 by a dCas9-fusion and inhibition of BRD4 binding by the small molecule inhibitor JQ1 allowed to study the relative dependencies on co-activation of low interaction AD VP16 and high interaction VPR.

2.6.1 Co-activation marks boost transcriptional activation, but are non-essential

The ability of the VPR opto and optoloop complex to induce H3K27 acetylation and BRD4 recruitment without activation of productive transcription had given a first indication that these co-activation marks are not sufficient to induce transcription. However, in its original publication [194] the opto complex with VP64 as an AD had been successfully employed to activate several endogenous genes in a light-dependent manner. Among these genes were *HBG1/2*, *IL1RN* and *ASCL1*. In order to verify that the opto complex did not fail to activate due to technical problems in the U2OS 2-6-3 cell line it was tested in a different reporter cell line containing a turboRFP gene under the control of a minimal CMV promoter and five *tetO* sites.

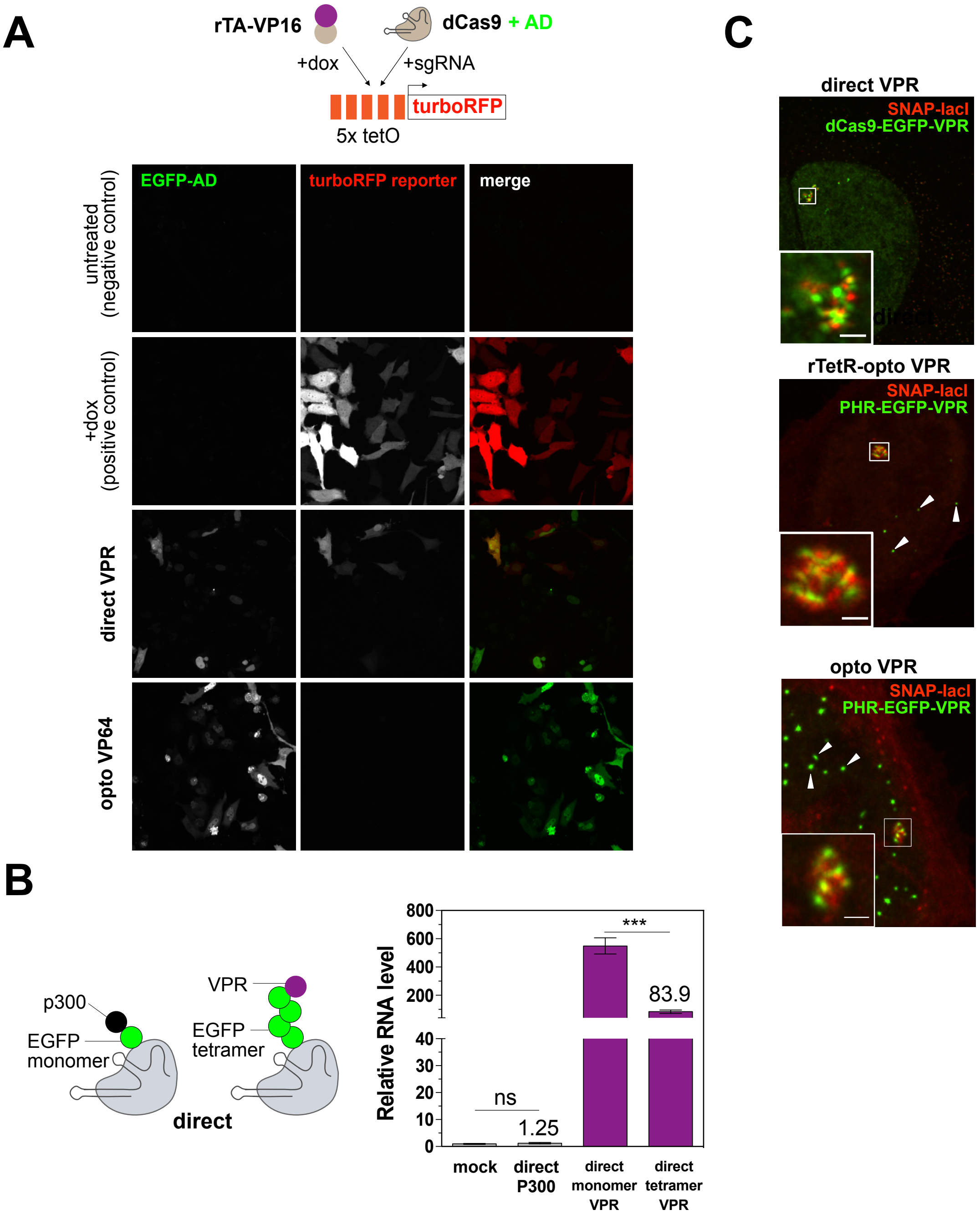


Figure 2.38: Co-activation in the absence of transcription. **(A)** Confirmation for inability of opto complex to activate transcription despite recruitment of co-activation marks. The opto VP64 complex was employed in a turboRFP-reporter cell line and failed to induce mCherry expression, while positive controls rTA+dox and dCas9-EGFP-VPR induced the reporter gene. (Images were acquired by Lukas Frank and are shown for completeness.) **(B)** Left: Two synthetic TFs that recapitulate properties of the inert opto and optoloop complex. dCas9-EGFP-p300Core locally induces co-activation marks (*see* Figure 2.22). dCas9-EGFP$_4$-VPR has a similiar size. Right: Bulk RNA levels induced by the different dCas9 complexes measured by qRT-PCR. (Continued on next page.)

VP64 was chosen as an AD to resemble the activation complex of the original publication [194]. The cell line stably expresses a rTA-VP16 activator that allows to activate the reporter gene by doxycycline (dox) addition. The reporter was activated and produced turboRFP, if exposed to dox, or if dCas9-EGFP-VPR was targeted to the *tetO* sites, but the opto VP64 complex failed to activate transcription (**Figure 2.38 A**). This demonstrated that the failure of the opto complex was not confined to the U2OS 2-6-3 cell line, but that the class of genes that can be activated by this system is generally constrained.

The VPR opto and optoloop complex both have a large size of 340 and $348 kDa$, respectively, counting the protein components starting from the DBD up to the first AD. This is due to the large size of its subcomponents, notably dCas9, PHR and VPR. This architecture could negatively affect transcriptional induction, for instance, if the AD is located at a large distance from the location at which the polymerase holoenzyme complex assembles on the DNA. To test, whether this large complex size is responsible for the inability to activate a synthetic TF of comparable size was constructed. An EGFP-tetramer was used as a spacer between dCas9 and the VPR AD. Indeed, dCas9-EGFP$_4$-VPR showed strongly reduced transcriptional activation measured by qRT-PCR (**Figure 2.38 B**). Yet, transcription was still strongly induced above mock transfection levels. Thus, a large complex size can hamper transcriptional induction, but is not sufficient to explain the complete failure of the opto and optoloop complexes. The failed activation can also not be attributed to the optogenetic recruitment of the AD, since the PHR-CIBN module works well for activating via the PHR-AD+CIBN-rTetR (opto rTetR) combination.

Figure 2.38: (Continued) Bars represent mean RNA levels relative to mock transfected cells. (qRT-PCR experiments were performed and analyzed by Lukas Frank.) Error bars: $\pm 1 s.d.$, n.s.: not significant, ***: $p < 0.001$, two-sided unpaired Student's t-test. Data for dCas9-EGFP-VPR corresponds to **Figure 2.21** and is shown for comparison. **(C)** Super-resolution imaging of transcriptionally inert complexes by SRRF. Maximum projections are shown. Arrows indicate optodroplets. The direct VPR image is a maximum projection of the single z-slices shown in **Figure 2.5 A**. Scale bar: $1.0\,\mu\text{m}$.

A potential inhibition by the formation of a phase-separated PHR-AD compartment at the promoter would also be expected for this TF architecture. Comparing the structure of the reporter gene cluster after recruitment of dCas9-EGFP-VPR, VPR rTetR opto or VPR opto complexes did not exhibit obvious differences that could explain different activation behavior (**Figure 2.38 C**). Notably, PHR-EGFP-VPR did not form a homogeneous droplet at the gene cluster, but rather individual clusters at *tetO* sites could be observed just as for dCas9-EGFP-VPR. One potential explanation for the lack of transcriptional activation by opto and optoloop complexes is that the combination of short AD residence times (of the PHR-CIBN interaction) and large complex size make the interaction with the core transcriptional machinery inefficient. This hypothesis could not be tested by further experiments, but is compatible with the activation potential of all synthetic TFs tested here. The discrepancy of failed target gene induction reported here and successful activation of e.g. *IL1RN* [194] could be based on the recruitment of co-activators. Despite their failure to activate transcription the VPR opto and optoloop complexes were able to enrich H3K27ac and BRD4 at the promoter (section 2.3.3). These co-activation marks could be sufficient to activate certain genes, while failing to activate genes with a heterochromatic promoter state as the U2OS 2-6-3 reporter [144]. Indeed, recruitment of dCas9-EGFP-p300 has been reported to be sufficient for *IL1RN* induction [204]. To test this, dCas9-EGFP-p300 which had proven capable of inducing histone acetylation and BRD4 enrichment (section 2.3.3) was recruited to the *tetO* sites and qRT-PCR confirmed that there was no induction of RNA production (**Figure 2.38 B**). The failure to activate supported the hypothesis that co-activation marks are insufficient for the transcriptional activation of the reporter gene studied here. The transcriptional activation by VP16 based TFs (which do not induce co-activation marks, section 2.3.3) shows that it is neither necessary. It could, however, make the difference between a weak and a strong AD.

2.6.2 VPR shows stronger, faster and partly H3K27ac-independent BRD4 recruitment compared to VP16

The different activation speeds and strengths of VP16 and VPR could stem from differences in the dynamics with which they induce co-activation marks. Therefore, BRD4-mCherry levels at the reporter gene cluster were recorded after inducing recruitment of PHR-EGFP-VP16 or -VPR to the promoters by light via the optoloop complex (**Figure**

2.39 A). The bromodomain of BRD4 binds acetylated histones [30], but in this setting also histone acetylation independent recruitment could be probed using JQ1, an inhibitor of the BRD4 bromodomain [205]. Recruitment to both *lacO* and *tetO* sites was necessary to achieve strong enough BRD4 accumulation for robust quantification. For VP16 recruitment to only one type of binding site did not lead to visible BRD4 enrichment. Even with the VP16 optoloop complex at *lacO* and *tetO* sites BRD4 levels could scarcely be detected (**Figure 2.39 B**). The optoloop complex allowed to avoid the use of CIBN-LacI that could have led to the formation of inhibitory optodroplets and at the same time did not activate transcription that could interfere via potential feedback mechanisms. BRD4 accumulation was strong and fast for VPR, but required longer for VP16 (**Figure 2.39 B-D**). 92% of cells responded to VPR with BRD4 accumulation compared to 27% to VP16. The plateau value of BRD4 accumulation was 4.1-fold higher for VPR compared to VP16. Curves were normalized to compare the dynamics of the accumulation. The amplitude of the averaged normalized time courses after subtraction of the start value was lower for VP16 (**Figure 2.39 D, Table 11**). This was due to the fact that the BRD4 spot intensities of single-cell curves did not reach their maximum synchronously at the end of the 60 min experiment, as they did for most cells in the VPR experiment. For VPR there was an initial steep rise of the BRD4 curve over the first 10 min followed by a phase of slower BRD4 accumulation. VP16 did not induce such a biphasic behaviour. The time to half-maximal accumulation of BRD4 was 15 min for VP16 and 13 min for VPR. This precedes the times to half maximal transcriptional activation of nascent RNA time courses that had been measured for PHR-EGFP-VP16/ -VPR recruited to CIBN-rTetR.

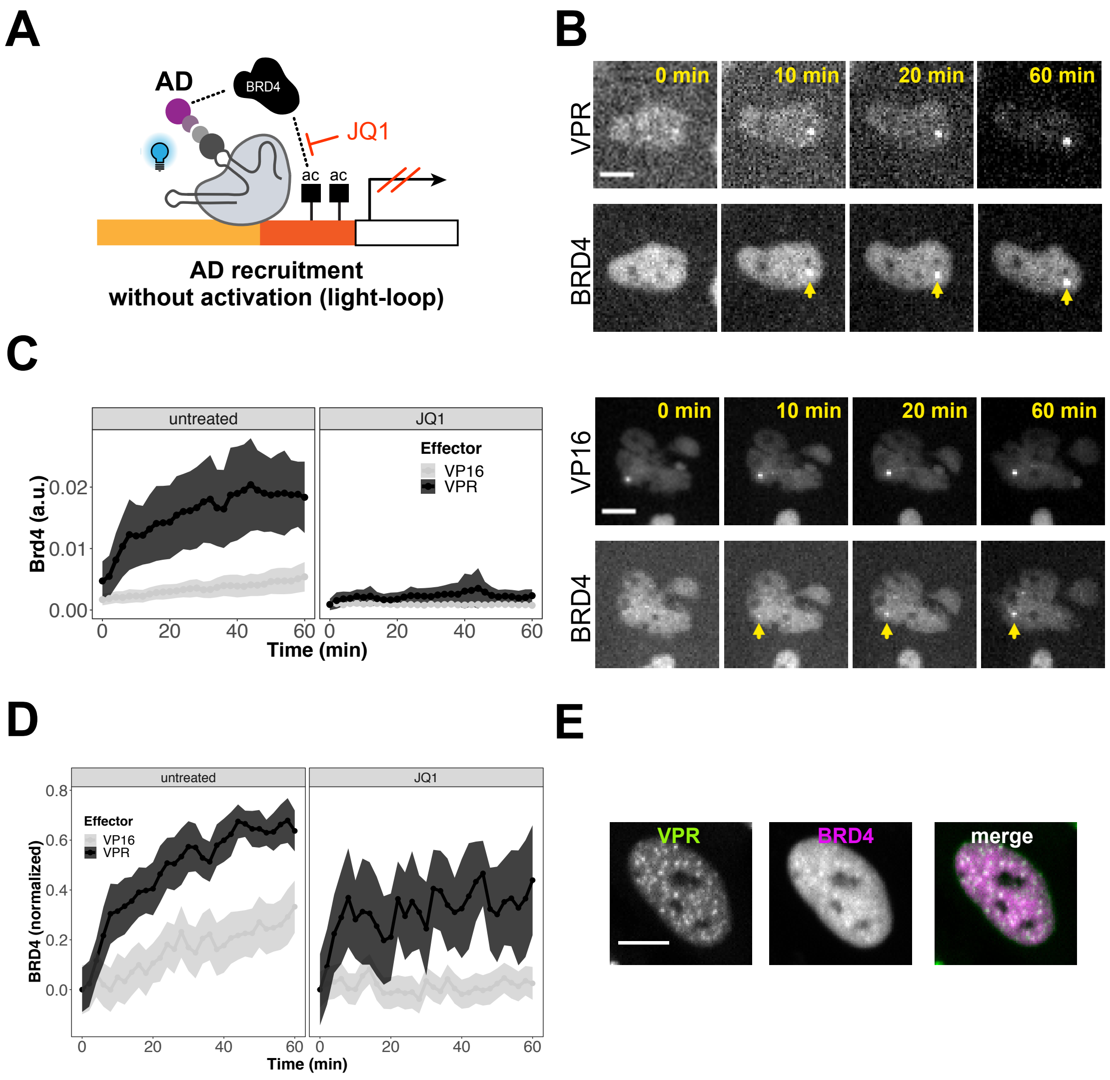


Figure 2.39: BRD4 recruitment kinetics for VP16 and VPR. **(A)** Experimental strategy: Targeting PHR-EGFP-VP16/-VPR to both *lacO* and *tetO* sites by the optoloop complex yields sufficient signal to quantify recruitment kinetics of BRD4. BRD4 could be recruited by binding to acetylated chromatin or by direct interactions with the AD. JQ1 treatment 3 h before start of imaging inhibits the bromodomain of BRD4 that binds to acetylated histones. **(B)** Exemplary image series of mCherry-BRD4 co-recruitment induced by VP16 or VPR optoloop complex. Scale bar: 10 µm. **(C)** Mean absolute intensity time courses of mCherry-BRD4 at the reporter locus with and without JQ1 treatment including responder and non-responder cells. Shaded area: 95 % CI, n = 10-85 cells per condition. **(D)** Mean normalized intensity time courses. The initial value was subtracted after normalization. Curves with a maximum value at variable time points in the cell population will thereby reach lower levels. Shaded area: 95 % CI. **(E)** Co-localization of VPR optodroplets (green, in the opto complex) with mCherry-BRD4 (magenta), which could be observed occasionally. Scale bar: 10 µm.

BRD4 could, thus, be a factor that mediates the different activation dynamics of VP16 and VPR. The fast accumulation of BRD4 at the reporter gene cluster suggested that it could be partly caused by direct interactions with the AD and not fully depend on prior histone acetylation. To test this hypothesis, cells were treated with a final concentration of 1 µM JQ1 three hours before the start of illumination to inhibit binding of acetylated lysines by the BRD4 bromodomains. This prevented BRD4 accumulation for VP16 and reduced it for VPR (**Figure 2.39 C**). The experiments with and without JQ1 treatment were performed on different days, so that the absolute intensity values may not be fully comparable. However, it is unlikely that the drastic reduction of BRD4 recruitment could be explained by a reduced detection efficiency. The fraction of (responder) cells with visible BRD4 enrichment was reduced to 10% for VPR and 0% for VP16. Comparing the BRD4 dynamics with and without JQ1 treatment in normalized time courses (**Figure 2.39 D**) showed that for VPR the initial steep rise remained, but the ongoing BRD4 recruitment was lost. This raised the possibility that the first phase of fast BRD4 co-recruitment was caused by direct interactions of VPR with BRD4, while a second fraction of BRD4 molecules binds once acetylation of histones has created binding sites for the BRD4 bromodomains. In a small subset of cells co-localization of BRD4 with PHR-EGFP-VPR optodroplets was observed, for instance for the VPR opto complex (**Figure 2.39 E**). It is noted that these were isolated cases and that they could be caused by the overexpression of mCherry-BRD4. Nevertheless, it demonstrated that BRD4 enrichment in phase-separated AD compartments is generally possible.

2.6.3 Activation by VPR and VP16 has different sensitivities to perturbations of the H3K27ac-BRD4 co-activation axis

VPR had demonstrated stronger multivalent interactions (section 2.3.2), stabilization of TF-DNA binding (section 2.4.1), more interactions with co-activators (section 2.3.3) and stronger transcriptional activation (section 2.5.1) compared to VP16. However, it remained unclear, whether the stronger induction of the H3K27ac-BRD4 co-activation pathway was merely a side effect or if it was a driver of the higher activation capacity. The functional role of the co-activation pathway for the induction of productive transcription was probed by two different strategies relying on the inhibition of co-activation and on circumventing TF-mediated co-activator recruitment.

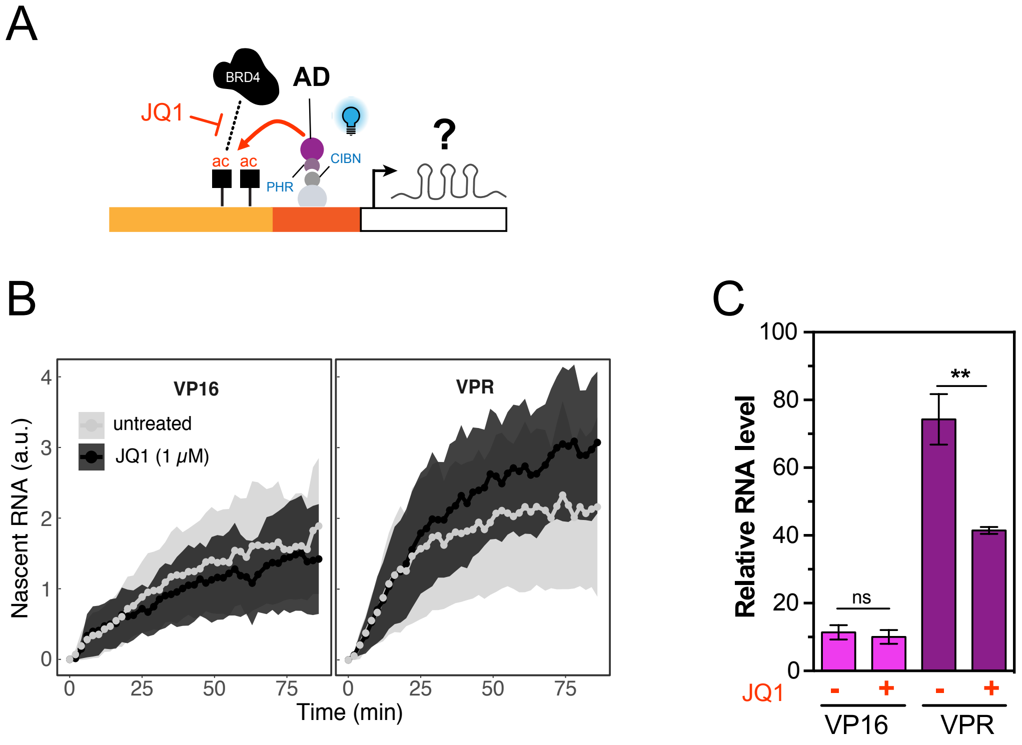


Figure 2.40: Inhibition of the bromodomain of BRD4 has no effect on activation by VP16 and a weak or no effect for VPR. **(A)** Experimental strategy: Bromodomain inhibition by JQ1 (for 3 h before the start of imaging) inhibits recruitment of BRD4 to actylated histones induced by PHR-AD. This allows to follow RNA production in the absence of this co-activation mark. **(B)** Mean absolute intensity time courses of nascent RNA in responder cells show no effect of bromodomain inhibition for either AD. Shaded area: 95 % CI, n = 7 - 20 cells per condition. **(C)** Quantification of bulk RNA after 90 min illumination by qRT-PCR shows weak repressive effect for VPR. Error bars: $\pm 1 s.d.$, n.s.: not significant, **: $p < 0.01$, two-sided unpaired Student's t-test. (qRT-PCR experiments were performed and analyzed by Lukas Frank.)

Accumulation of BRD4 was inhibited by JQ1 and light-induced nascent RNA time courses were acquired. This treatment allows H3K27ac to be set, but interrupts this pathway at the binding of BRD4 to acetylated histones (**Figure 2.40 A**). As shown in the previous section JQ1 efficiently reduced BRD4 recruitment induced by VPR and abolished it for VP16. Treatment with 1 µM JQ1 three hours before light-activation did not reduce nascent RNA levels induced by the recruitment of PHR-EGFP-VP16/-VPR to CIBN-rTetR measured by microscopy (**Figure 2.40 B, Table 12**). The transcriptional induction dynamics were not affected either. Bulk RNA measured by RT-PCR showed a moderate 1.8-fold reduction only for VPR, but not for VP16 (**Figure 2.40 A**). Consequently, BRD4 accumulation appeared to accompany transcriptional activation, but not to be a strict prerequisite for the induction of the reporter gene. It may, however, give a moderate boost to transcription induced by VPR. One limitation of the approach is that the major part of BRD4 accumulation relying on histone acetylation was inhibited, but that the remaining fast BRD4 recruitment (probably due to direct interactions with VPR) could be sufficient to boost transcription. In any case, the results show that strong transcriptional induction did not depend on large BRD4 quantities at the promoter cluster. The lack of an effect of BRD4 inhibition on activation by VP16 is consistent with the view that VP16 does not rely on this co-activation pathway to induce transcription.

In an opposing approach it was investigated whether pre-set co-activation marks could help activation by VP16 more than activation by VPR. The reporter promoter was primed with dCas9-EGFP-p300 (containing the core domain of p300) recruited constitutively to the *lacO* sites and local hyper-acetylation of histones was induced (**Figure 2.41 A**). Transcription was then induced by PHR-EGFP-VP16/-VPR recruited to CIBN-rTetR upon illumination. For VP16 a 1.5-fold enhanced transcriptional response was observed compared to mock-priming by dCas9-EGFP for bulk RNA levels using RT-PCR (**Figure 2.41 C**). For VPR this effect could not be seen.

When observing nascent RNA by microscopy an enhancement could be seen for both VP16 and VPR (**Figure 2.41 D**). This boost was stronger for VP16 compared to VPR (2.9-fold amplitude increase vs. 1.8-fold; **Figure 2.41 E, Table 12**). The dynamics were not strongly altered (**Figure 2.41 F**) exhibiting only a moderate decrease of times to half-activation. This is surprising, since a poised promoter state that already contains some activation marks would be expected to accelerate the activation process. It was concluded that pre-existing histone acetylation can increase the amplitude of the transcriptional response and does so more for VP16, which has a lower ability to co-recruit histone acetyl transferases. However, histone acetylation does not appear to be rate limiting for gene induction placing it in a parallel branch of the activation mechanism.

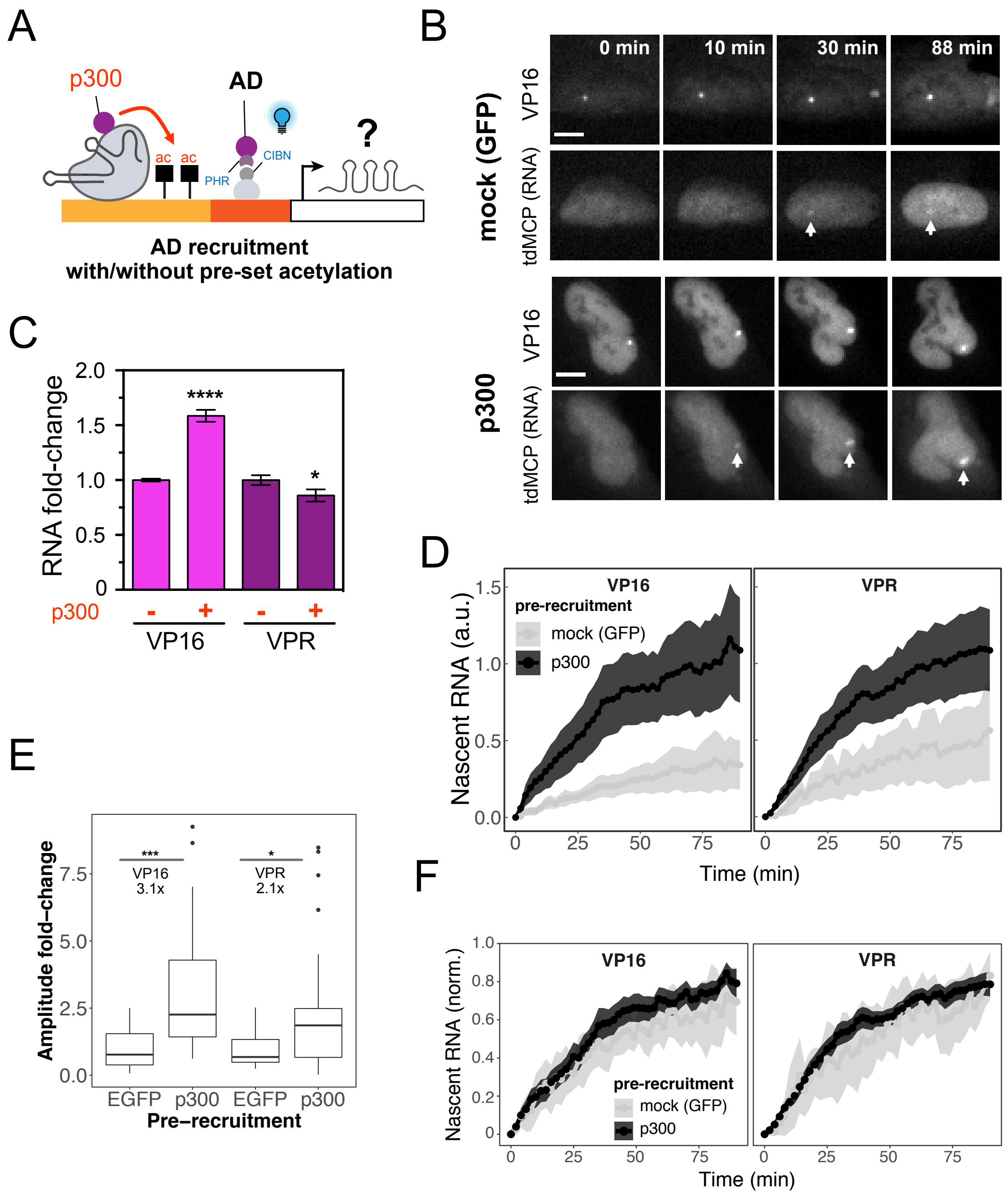


Figure 2.41: Promoter priming by p300 pre-recruitment benefits primarily VP16 **(A)** Experimental strategy: Constitutive pre-recruitment of dCas9-EGFP-p300Core to the *lacO* sites primes the promoters by inducing H3K27ac. Transcriptional time courses can then be followed after light-induced recruitment of PHR-AD to CIBN-rTetR at the *tetO* sites. **(B)** Exemplary image series for activation by PHR-EGFP-VP16 with mock (dCas9-EGFP) or p300 (dCas9-EGFP-p300) pre-recruitment. Scale bar 10 µm. (Continued on next page.)

Overall, the differential sensitivities of VP16 and VPR to perturbations of the co-activation pathway indicated that a part of VPR's superior activation capacity could be caused by the H3K27ac-BRD4 axis and that giving weaker ADs the ability to induce this pathway could increase their activation strength. The effect sizes of inhibition or creating a pre-poised promoter state were small reiterating the results of the previous sections that histone acetylation and BRD4 are dispensable for the core transcriptional process that lead to RNA PolII recruitment and productive RNA elongation.

Figure 2.41: (Continued) **(C)** Bulk RNA measured by qRT-PCR after 90 minute illumination. Values are shown as fold-changes compared to mock (dCas9-EGFP) prerecruitment. Error bars: $\pm 1 s.d.$, *: $p < 0.05$, ***: $p < 0.001$, two-sided unpaired Student's t-test. (qRT-PCR experiments were performed and analyzed by Lukas Frank.) **(D)** Mean absolute nascent RNA time courses of responder cells. Shaded area: 95 % CI, n = 9 - 54 cells per condition. **(E)** Distribution of nascent RNA plateau values as fold-change compared to the mean of the respective mock pre-recruitment. *: $p < 0.05$, ***: $p < 0.001$, two-sided unpaired Student's t-test. **(F)** Mean normalized nascent RNA time courses. Shaded area: 95 % CI.

Chapter 3

Discussion

This thesis set out to explore the link between dynamic TF binding properties and transcriptional activation to address the following questions: How do TF binding kinetics, multivalent interactions of the AD and formation of phase-separated compartments determine the kinetics of transcriptional induction and the amount of RNA produced? **Figure 3.1** summarizes the methods that were developed and applied to answer this question and the main observations that were made. The research question was motivated by the expectation that, in addition to the stoichiometric binding of TFs to their binding sites in equilibrium, also the binding kinetics and weak multivalent interactions between TFs could regulate transcription. The recent proposition that phase-separated transcriptional condensates control the expression of strongly induced genes is very different from the traditional view that gene activation is determined by the TF binding site occupancy. The dynamic binding properties of TFs that were tested here as regulators of transcription were formation of transcriptional condensates, TF residence time and multivalent interactions. Synthetic TFs with vastly different binding and AD-AD interaction strengths were constructed from a versatile toolbox containing DBD and AD modules. These TFs were then employed in time-resolved experiments to address the impact of their dynamic properties on gene activation. Quantities of the transcription process that were characterized were the levels of nascent and mature mRNA, the dynamics of their production and the amounts and dynamics of co-activator recruitment, specifically of H3K27ac and BRD4.

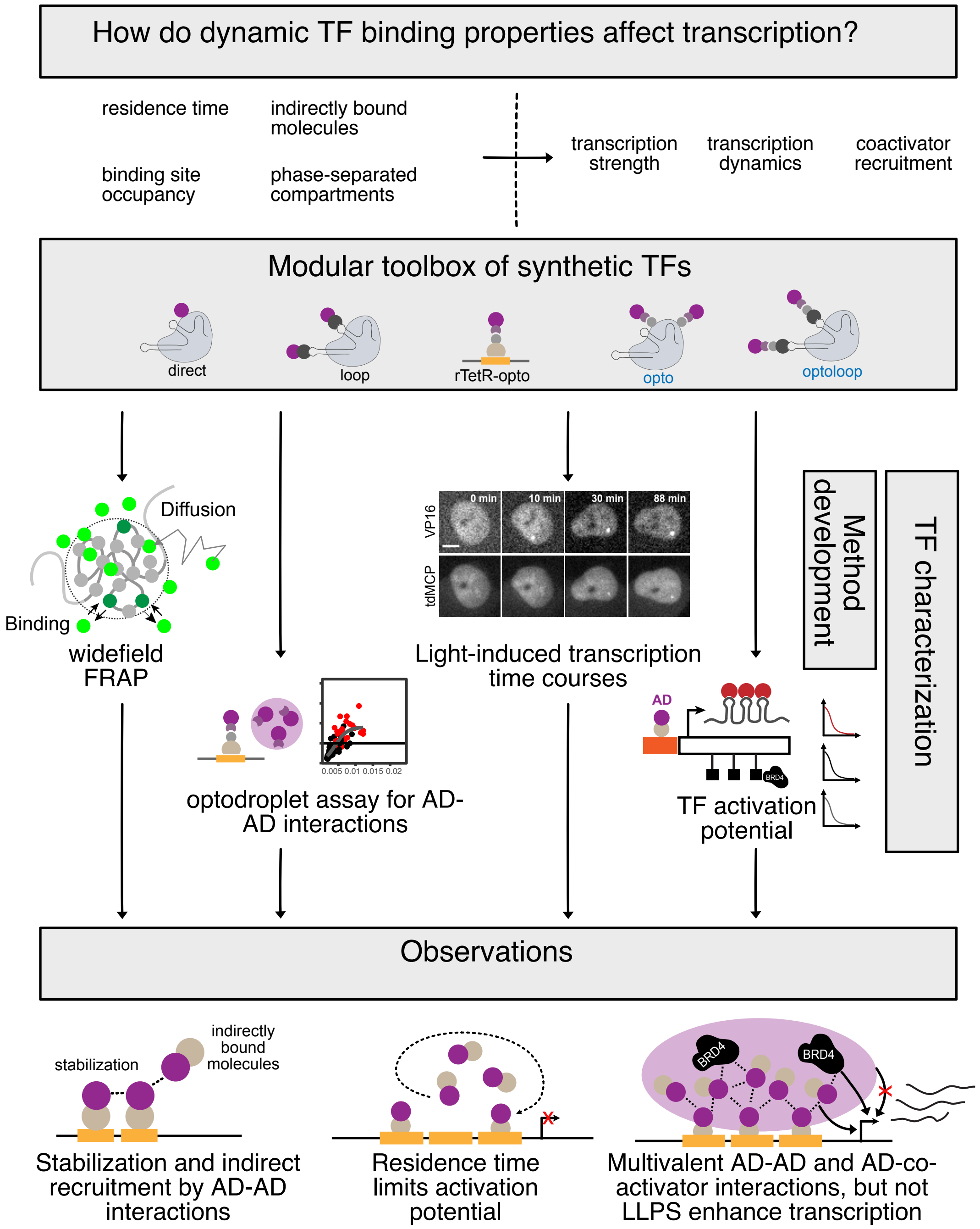


Figure 3.1: Overview of main research question, methods that were developed and applied and experimental observations of this thesis.

Two automated workflows were developed including analysis software packages to measure TF binding kinetics at a cluster of binding sites by widefield FRAP and to record transcription induction time courses, each capable of measuring tens of cells in multiple conditions. These methods are expected to be of additional value in the future for different research questions, as was demonstrated by two proof-of-concept applications. Two additional types of experiments completed the characterization of TFs: The strength of multivalent AD-AD interactions was measured by the propensity of PHR-AD constructs to form optodroplets. To systematically characterize the activation potential of synthetic TFs, enrichment of nascent RNA and coactivators at the reporter gene cluster was evaluated. With the approach implemented here three main results were obtained: (1) Multivalent interactions of the AD stabilized the binding of their TFs and led to recruitment of indirectly bound TF molecules. (2) The residence time regulated transcription levels independent of the occupancy of TF binding sites. (3) There was a positive correlation of the propensity of TFs to form optodroplets and consequently the strength of their multivalent AD-AD interactions with the strength and speed of transcriptional induction. However, this was not caused by the actual formation of phase-separated TF compartments, which could even inhibit transcription.

In the following sections, these results will be discussed regarding the ability of the promoter to process dynamic information by a multistep activation mechanism. The role of multivalent AD-AD and AD-coactivator interactions will be confronted with the role of liquid-liquid phase separation. Moreover, the relationship of transcriptional activation and the H3K27ac-BRD4 co-activation pathway will be discussed. The interplay of AD and DBD properties in promoter activation will be integrated into a model and potential applications for the design of synthetic TFs and for the understanding of pathogenic processes will be discussed.

3.1 Dynamic processing of the TF binding signal at the promoter

In an equilibrium binding view of transcriptional regulation the amount of produced RNA can (in prokaryotic cells) be predicted from the amount of TF that is bound at the promoter [69]. TF binding to its cognate sites in the promoter proximal region or in enhancers is in this view followed by an ordered sequence of events. These comprise specific interactions of the TF with general transcription factors (GTFs) like TATA-box binding protein (TBP) [93, 203] resulting in PIC assembly and the start of productive transcription. The discovery of transcriptional bursting [44, 179], refratory promoter states [47, 48] and the differential encoding of cellular signals in the nuclear translocation dynamics of TFs, like p53 [4, 174] and NFκB [2], have challenged that view.

In eukaryotes transcriptional regulation must be more complex than the binding site occupancy directly regulating the number of escaping polymerases. The alternative model - multistate transcriptional regulation - is expected to be reflected in the dynamics of transcriptional induction of a previously silenced gene as the promoter passes through multiple activation states. Growing evidence suggests that besides binding site occupancy also the TF binding kinetics regulate transcription [76, 77]. It prompted this study to focus on the link of binding kinetics and transcriptional activation capacity. This work contributed to a better understanding of the dynamic regulation of transcription by developing methods for the acquisition of induction time courses and for measuring TF binding kinetics. It, further, applied these methods to characterize the binding properties of a toolbox of synthetic TFs and to link residence time and activation capacity. The widefield FRAP approach allowed to robustly measure binding kinetics at a binding site cluster over several minutes for 14 conditions in hundreds of cells in total. The synthetic TFs covered a wide range of complex architectures ranging from direct DNA binding domain-activation domain (DBD-AD) fusion constructs to multicomponent systems comprising dCas9 as a DBD, guide RNA binding tdPCP as a relay module and PHR-CIBN interactions for light-controlled AD recruitment. While some of these complexes did not accurately represent the architecture of (single-component) endogenous TFs, they provided a versatile system to modulate TF properties, like AD residence time, complex size or the timing of promoter binding. Another limitation of the multicomponent complexes was that the apparent off-rate measured by FRAP will not necessarily be determined by the interaction of interest, but by the fastest exchanging component. However, for the constructs studied here the binding turnover of the DBD module was always lower compared to the tdPCP-AD or PHR-AD module, so that the obtained parameters should give a good representation of the tdPCP-PP7 or the PHR-CIBN interactions. The FRAP experiments measured the AD off-rate, since it always contained the EGFP label. Potential regulation mechanisms, like the TF-modification kinetic proofreading model presented in section 2.4.2, depend on the residence time of the AD making the apparent off-rate measured by FRAP an interesting quantity.

TF residence time alone was not a sufficient predictor for activation capacity across variable TF architectures, as is highlighted by high activation levels induced by tdPCP-EGFP-VPR binding to guide RNA stem loops of dCas9 (loop complex) with high exchange rates. This is likely due to the strong differences in size, spatial orientation of sub-components and stoichiometry of the respective architectures. Much more comparable conditions were achieved by a single base mutation of a guide RNA targeting the dCas9-EGFP-AD activator to the *tetO* sites of the reporter gene. This minimal perturbation conserved binding site occupancy at a level that still ensured visible targeting in a substantial fraction of cells but led to shorter residence times. While the reduction in the local concentration of dCas9-EGFP-VPR at the array was less than three-fold, the reduction in RNA levels was greater than 17-fold. Moreover, the mutation led to a complete loss of transcriptional induction for VP16, even though visible promoter recruitment was observed in 17% of cells. This suggests that short residence times can prevent transcription induction even if binding sites are occupied. Since residence time and occupancy are coupled via the dissociation constant ($K_D = \frac{k_{off}}{k_{on}} = \frac{1}{\tau_{residence} \cdot k_{on}}$), it is difficult to change one without affecting the other. Grouping of cells by their local TF concentration at the promoter spot revealed that the short residence time activation complex induced less RNA production for the same binding site occupancy. This analysis allowed to conclude, that the residence time is a regulator of transcription in its own right and independent of its influence on the binding site occupancy. The results presented here confirmed the observations of previous studies that longer residence times correlate with stronger activation [76, 160]. The results also demonstrate for the first time, that this effect cannot be attributed to solely a resulting change of binding site occupancy, but that there must be an activation mechanism that is sensitive to the residence time itself. Importantly, here the residence time was measured directly at the target binding sites, while previous studies [76, 160] used single particle tracking that also measures binding at off-target sites. Correlations found in these studies could also be caused by globally increased chromatin binding of TFs, e.g. due to TF modifications [76]. Several studies used pulse-chase experiments to determine TF residence times and correlate it with activation capacity. They found enhanced activation for TFs with residence times of hours to days. The measurement of residence times by these methods are intrinsically limited to longer times because they rely on inducing the slow expression of a chaser molecule [78] or on expression of a reporter gene [206]. While residence times measured by these methods deviate from the seconds to minutes time scale determined here and in other imaging based studies, they might represent the long-term binding of the (typically small) immobile fraction. Unbinding of these apparently immobile molecules is outside the imaging duration and could have additional functional consequences.

As proposed by a recent theoretical study [80] nucleosome remodeling could be a necessary step between TF binding and transcriptional activation that could make the activation process susceptible to a reduction of TF residence time. The model of residence time dependent activation in this thesis relies on an alternative mechanism. In this mechanistic model the TF itself stores the information of a sufficiently long residence at the promoter. This could be a more adequate mechanism for a reporter gene that is controlled by a large number of binding sites, so that always a certain fraction of these sites is occupied. The modification of the TF ensures that the residence time of the individual TF molecules is read out. It is noted that this mechanism can be adapted to include inhibitory TF modfications that limit activation by long residence time activators. Thereby, TF activation capacity could be limited to residence times inside a certain range.

It remains an open question in which range of absolute values residence time changes are able to affect transcription levels. In the set of synthetic TFs studied here, residence times varied widely. For instance, both dCas9-EGFP-VPR (with a fully matching guide RNA) and tdPCP-EGFP-VPR (in the loop complex) with residence times of 124 s and 47 s, respectively, were able to elicit a strong transcriptional response, whereas dCas9-EGFP-VPR with the C2G-mismatch guide RNA and a residence time of 57 s exhibited a strongly reduced induction. It must be noted that two AD molecules can be recruited in the loop complex and that there were additional differences in the immobile fractions. The TF modification model clearly showed that the residence time relative to the time for TF modification determines transcription levels. If this model holds true, times for modifications should be in the range of tens of seconds to few minutes. If rates of TF modification turnover were in the same range as the rate for histone acetylation, they would be in the range of few minutes [183, 207] which is compatible with a proofreading function at the observed TF residence times.

Control of transcription levels by TF residence time could have advantages over control by binding site occupancy, since it does not depend on (variable) expression levels and prevents activation at off-target sites that could become occupied at sufficiently high TF concentrations. This view is supported by the fact that the residence time of certain TFs is regulated by signal transduction pathways. For instance, acetylation of p53 triggered by radiation induced DNA damage increases its residence time and thereby induces a response of the transcriptome [76].

It was not in the scope of this study to identify the exact mechanism that conveys residence time dependence to the promoter activation process, but the established molecular tools allow to test some of the assumptions of the TF-modification proofreading mechanism. For example, synthetic TFs could be purified before and after promoter recruitment and probed for PTMs like acetylation by Western Blots . This could be either dox-controlled using rTetR-EGFP-AD or light-controlled using PHR-EGFP-AD. Once an associated PTM was identified, site-directed-mutagenesis of the TF sequence could test, whether these modifications are really limiting for transcription.

The TF modification proofreading model had postulated the existence of different states of the promoter that were characterized by the occupancy of binding sites with TF and modified TF. The dynamics of reporter gene induction in the U2OS 2-6-3 cell line have already been described by a multi-step process potentially containing a positive feedback step [12]. Here, the development of an automated imaging and analysis pipeline for light-induced transcription time courses provided the opportunity to gain additional insights into the dynamic aspects of gene activation.

There were two main limitations of the light-induced time course experiments which were linked to cellular heterogeneity: First, the heterogeneous reporter gene copy number hampered the analysis of stochastic gene activation, because the heterogeneity could not clearly be attributed to stochastic events or different copy numbers. Second, there was also heterogeneity between microscopy experiments with the same AD performed on different days. This could be partially due to variable detection efficiencies for the fluorescence intensities, but also biological differences could play a role. Biological replicates of qRT-PCR experiments harvested on the same day showed very good agreement, but the variability was greater for experiments with the same condition performed on different days. As a consequence, the transcriptional response should be compared only within sets of experiments that were performed together, while experiments from different experimental sessions might display a greater variability.

The light-induced time course experiments proved very versatile: In two proof-of-concept applications they revealed that the information contained in heterogeneous single-cell RNA time courses can help to distinguish between feedback based and simple multistep activation mechanisms. They further showed that previous activation events could potentially poise the promoter for sensitized re-activation. These first indications for a network of promoter states (also suggested by the model of kinetic proofreading) let it appear plausible that the promoter possesses a certain capacity for signal processing. Besides distinguishing between long and short residence time binding, promoters could also have a differential responsiveness to a sustained or a pulsatile binding of TFs like p53 and NFκB [2, 4, 174].

The light-induced time course experiments also revealed that ADs differ in their activation speeds and that perturbations of the co-activation pathway can affect the amplitude, but not the dynamics of the transcriptional response. It is noted that the differences in activation speed were moderate amounting to a difference of around 10 min in half-activation time between slow and fast ADs. However, given that typical half-activation times were around 30 min this is a relatively strong relative difference. Moreover, especially for slow ADs the RNA time course had not reached a plateau value at the end of the 90 min experiments in many cells, so that activation speed of slow ADs and the difference to fast ADs was likely underestimated. That the dynamics were not affected by co-activators was surprising, because the fast BRD4 recruitment was suggestive for a mechanism, in which co-activator accumulation precedes and causes full transcriptional activation. Moreover, studies of transcriptional bursting had found that histone acetylation was correlated with burst frequency [52, 208], which suggests a role of histone acetylation in rate limiting steps of promoter state transitions. Even though this study did not find a rate limiting role for H3K27ac and BRD4, it could show that the accumulation of both marks is an early event after TF binding. It also placed H3K27ac functionally upstream of BRD4 recruitment, because histone acetylation by dCas9-p300 was sufficient to also induce BRD4 accumulation. This did not exclude an additional way of BRD4 recruitment by direct interactions with an AD with strong multivalent interactions (VPR).

3.2 The role of multivalent interactions and phase-separated transcription compartments

This thesis found that TFs with high multivalent interactions of the AD activated transcription faster and more strongly, but saw no supportive effect in cells, in which the multivalent interactions actually led to the formation of phase-separated compartments. There was rather a trend towards slower activation in these cells and forcing droplet formation using bridging factors even had a strongly repressive effect. The actual physical state of TF condensates at promoters and enhancers remains debated [107]. Early studies postulated that TF associated in condensates with the characteristics of liquid-liquid phase separation based on in vitro experiments, fast exchange of molecules with the other phase measured by FRAP and broad disruption of molecular interactions with 1,6-hexanediol. However, these criteria are insufficient to prove the liquid like nature of phase-separated compartments [108]. An in vivo study [103] has since then suggested that TFs form hubs that contain more molecules than those directly bound to DNA, but are limited in size in contrast to true LLPS droplets that can in principle grow indefinitely as the TF concentration exceeds the critical value. Surface condensation could explain the formation of such compartments with a limited size [109]. Here, all macroscopically observable TF droplets were dependent on PHR-fusions supporting the view that LLPS is not typical for TFs at normal expression levels.

The PHR-CIBN system, in contrast, does form optodroplets with true LLPS characteristics like concentration dependent formation, high exchange rates, independence of chromatin as a scaffold and droplet fusion, as has been demonstrated in this thesis. This is in line with the original description of optodroplets [140]. This could be seen as a limitation, because it may not recapitulate the true physical state of TFs in living cells. However, it also represents an opportunity to investigate what effects it would have, if TFs did form true LLPS compartments in the nucleus. PHR optodroplet formation, moreover, proved useful in determining the propensities of ADs to favor or disfavor phase separation of the PHR-AD fusion constructs. This was used in an assay for multivalent interactions similar to the approach used for heterochromatin protein 1 (HP1) [168]. Three ADs, VPR, p65 and Rta had a high phase separation propensity in the optodroplet assay displaying low critical concentration values and a high percentage of cells with visible (macroscopic) optodroplets. In contrast, phase separation propensity was low for ADs of VP16 and especially STAT2, which did not form optodroplets at all. The propensity to form optodroplets depended also on the localizer and its ability to function as a bridging factor. Even though CIBN-rTetR or CIBN-LacI were not fluorescently labeled, so that the influence of their expression levels on droplet formation could not be measured, the optodroplet propensity of VP16 and VPR was determined in the presence of these two different localizers. This serendipitously led to the discovery that LacI can increase optodroplet formation.

It was the ADs with strong multivalent interactions, VPR, p65 and Rta that showed the strongest and fastest activation, but comparing cells with and without droplets for one construct revealed no increase in transcription levels in the presence of phase-separated compartment and hinted rather at a somewhat delayed activation. Indeed, enforcing droplet formation using CIBN-LacI or PHR-GBP as bridging factors had a repressive effect. This suggests that multivalent interactions could benefit transcriptional activation in a way that is independent of the formation of phase-separated compartments. It very clearly showed that it is not sufficient to increase the local amount of TF by tethering a phase-separated TF droplet at the promoter. It also challenges a model that assumes that nucleation of TF droplets at sites with an accumulation of phosporylated RNA PolII CTD represents a positive feedback mechanism [101]. Instead, multivalent interactions could boost transcription by stabilizing the DNA binding of TFs or by fostering co-activation that could be dependent on multivalent TF-BRD4 or TF-HAT interactions. The effects of the fusion of multivalent FUSN to VP16 fit well into this view. The increased multivalent interactions led to a strong increse of activation capacity, but this did not depend on the actual presence of PHR-FUSNVP16 optodroplets. This is particularly interesting given the oncogenic potential of TF-FUSN fusion proteins that lead to a transcriptional reprogramming of the cell [111]. A liquid droplet can be assumed to be absent at the promoter, if no optodroplets were observed in the rest of the nucleus. It was outside the scope of this work to determine the exact physical state of TFs bound at the promoter. Consequently, some of the mechanisms suggested to enhance transcription in phase-separated compartments could also work in sub-microscopic TF hubs of a smaller size than optodroplets that could not be observed in this study. Among these potential mechanisms is a reduction of TF search time [99] or the facilitation of interactions of active genomic loci.

Here, it was also found that multivalent interactions of the AD stabilize DNA binding and make the TF more resistant to a reduction of direct DNA binding as was the case for dCas9-VPR, when the *tetO* guide RNA was mutated. Hence, multivalent interactions did not only regulate the strength of transcriptional induction, but could also have an impact on the robustness and specificity of target site recognition. It is to be expected that a TF with a high interaction AD will show more binding to a target site with a single base mismatch than a TF without these interactions. In contrast to the results for the reporter gene cluster studied here, a single particle tracking study found no residence time prolonging effect of the AD for NFκB [160]. However, it is noted that these experiments measure residence times globally on chromatin, whereas the FRAP experiments here were conducted at a cluster of repetitive binding sites, which is likely to promote cooperative binding via multivalent interactions.

The exact mechanism by which the bridging factors CIBN-LacI and PHR-GBP inhibited transcription could not be elucidated. It is possible that the formation of large amounts of optodroplets leads to a sequestration of transcriptional regulatory factors, like GTFs, co-activators or RNA-PolII, a phenomenon originally termed TF squelching [209]. It could also be imagined that the growth of TF droplets dilutes its activating constituents and possibly reduces chromatin contacts of the promoters in the droplet. However, despite a potentially slightly reduced activation speed single component activator optodroplets without bridging factors like those based on PHR-VPR, -Rta, -p65 and FUSNVP16 did not lead to such a drastic repression. It seems, thus, more likely that the exact composition of phase-separated compartments determines whether they have a neutral or a repressive effect. Bridging factors could reduce the internal mobility of molecules inside the droplet, exclude other activator molecules or lead to more gel- or solid like states. There have been examples of repressive phase-separation before, even though it has been mainly thought to enhance the associated processes. In the neuronal cytoplasm a repressive effect of phase-separated compartments on the translation of their constituent mRNA molecules has been observed [210]. In nucleoli inhibition of ribosomal RNA was associated with the formation of a phase-separated nucleolar subcompartment containing RNA Pol I. While these are different subcellular contexts, they might share a mechanism that prevents the processing of the contents of phase-separated compartments.

Co-activators seemed to play a supportive role in the activation of the reporter gene studied here but they were neither necessary, nor sufficient to induce transcription, nor was transcription a prerequisite for the acquisition of co-activation marks. VP16 was able to induce enrichment of BRD4 when recruited to both *tetO* and *lacO* sites via the optoloop complex. However, BRD4 accumulation occurred slowly and only to low levels. The radial profile analysis revealed that even after 24 hours, most VP16 complexes recruited to *tetO* alone, lacked detectable BRD4 and H3K27ac enrichment, despite being able to induce transcription. Conversely, VPR complexes induced robust H3K27ac and BRD4 enrichment, even in the absence of transcription (opto VPR and optoloop VPR).

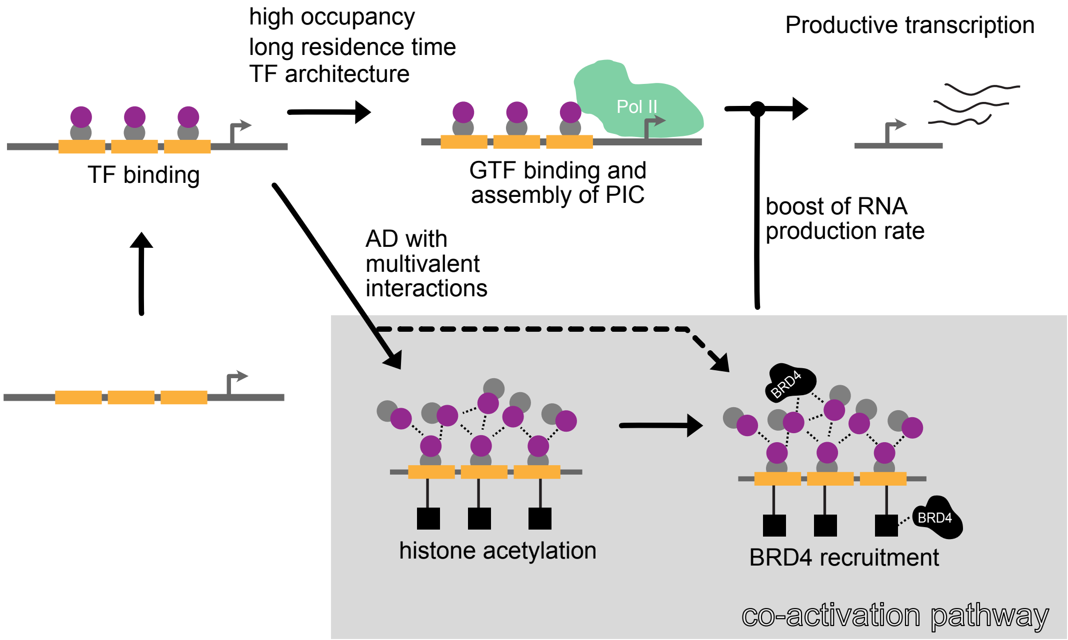


Figure 3.2: Model for independent pathways to transcriptional induction and co-activation. Transcription factor (TF) binding triggers both pathways. Recruitment of general transcription factors (GTFs) and RNA PolII for the establishment of the pre-initiation complex (PIC) and start of productive transcription require sufficiently strong TF-DNA binding and a TF architecture that has the right spatial orientation of the AD to interact with PIC components. Histone acetylation and sequentially BRD4 recruitment can be induced by local enrichment of the AD at the promoter and are independent of the strength of TF-DNA binding. BRD4 can also be recruited by direct interactions with the AD. The co-activators are not necessary for productive transcription, but can boost RNA production rates.

A fraction of BRD4 molecules was identified that interacted with VPR directly and in a bromo domain-independent manner. A similar result was found by ChIP-seq analysis after JQ1 treatment [211], which showed that BRD4 peaks were mainly lost at transcription start sites (TSS), but not in the gene body. There is the intriguing possibility that ADs with high self-interaction potential could also be more prone to interact with co-activators, which could boost transcriptional activation. For this mode of co-activator recruitment a strong direct DNA binding would not be necessary as illustrated by the occasional incorporation of BRD4 into optodroplets. Similar to VPR, the co-activator

BRD4 contains large intrinsically disordered regions [40], which could possibly undergo such direct weak interactions. Although not explicitly shown in the experiments, a similar mode of co-recruitment could be envisioned for histone acetyl transferases interacting with VPR to explain the robust correlation of VPR with H3K27ac. Similarly, the Mediator complex could be recruited by multivalent interactions with the AD, as it has been shown to co-localize with TFs and BRD4 [40, 104]. Inhibiting the co-activation pathway by JQ1 treatment or pre-activating it with dCas9-p300 had only a moderate effect on transcription, but showed that it contributes to VPR's activation potency and can boost activation by VP16, which on its own would not induce co-activation. There were two limitations to these experiments: First, small amounts of co-activators below the detection limit could also be recruited by VP16, as BRD4 recruitment just above the detection limit was found, when VP16 was recruited to both *tetO* and *lacO* sites. Second, inhibition of BRD4 recruitment by JQ1 was not complete for VPR, probably due to bromodomain independent VPR-BRD4 interactions. This leaves the possibility that small amounts of co-activators are strictly necessary for RNA production. Nevertheless, a detectable activation of the H3K27ac-BRD4 pathway is not. Hence, co-activation could be seen as an independent parallel branch of the transcriptional activation pathway (Figure 3.2) that is not strictly necessary, but can increase the amount of RNA that is produced, once direct TF binding has led to an active promoter state (e.g. via GTF recruitment). While the pathway to productive transcription requires TF binding with sufficiently high binding site occupancy and long residence times and a TF architecture that has the right spatial orientation of the AD, co-activation could proceed via multivalent interactions with the AD. There is an ordered sequence of co-activation events, in which histone acetylation is set first and BRD4 binds to acetylated histones via its bromodomain second. But direct interactions of BRD4 with the AD can also partly bypass this step. BRD4 helps in the transition from initiation to productive elongation by setting S2 phosphorylation of the RNA Pol II CTD either directly [32] or via P-TEFb. For BRD4 independent activation P-TEFb is recruited directly and probably in smaller amounts, for instance by direct binding to ubiquitinylated VP16 [200].

3.3 An integrated view of TF binding, multivalent interactions and activation

The experiments performed in this thesis with synthetic TFs containing various ADs and DBDs of different strengths allow to formulate a set of rules to estimate the relative strength of TFs. They are depicted in the scheme in **Figure 3.3**. It shows different situations at the gene promoter in dependence on the two TF properties that were studied in this work: TF residence time on the x-axis and multivalent interactions of the AD on the y-axis. It was found that AD and DBD, even though they can be modularly combined to yield functional transcription factors, were not independent but mutually influenced their binding and activation properties. It was found that TFs with multivalent ADs had longer residence times (**Figure 3.3**, grey area: strong interactions shift the TF complex to longer residence times). Strong DNA binding, comprising long residence times and high occupancy, in combination with weak multivalent interactions of the AD caused moderate levels of transcription that were limited by the lack of an additional indirectly bound activator fraction and possibly of co-activators that interacted with the AD (**Figure 3.3 A**). For optimal activation of transcription both strong target site binding and multivalent interactions of the AD were needed. This state was characterized by a high binding site occupancy, an additional fraction of indirectly bound TFs and the recruitment of co-activators (**Figure 3.3 B**). A reduction of direct DNA binding and the consequently shortened TF residence time led to a decrease of RNA production (**Figure 3.3 C**), while co-activator levels were not affected.

For ADs with low multivalent interactions reduced DNA binding led to an even stronger reduction of local TF concentrations and to a complete stop of transcription (**Figure 3.3 D**). Based on the analysis of guide RNA mutations in single cells the residence time played an important role in transcriptional activation independent of binding site occupancy. TFs with strong multivalent interactions like dCas9-VPR also depended on binding site residence time, but the reduced DNA binding could be partly compensated by the stabilizing TF-TF interactions. Multivalent interactions, thus, had a buffering effect. Multivalent activator interactions led to a faster and stronger induction of transcription and a larger fraction of responding cells (**Figure 3.3 B-C**).

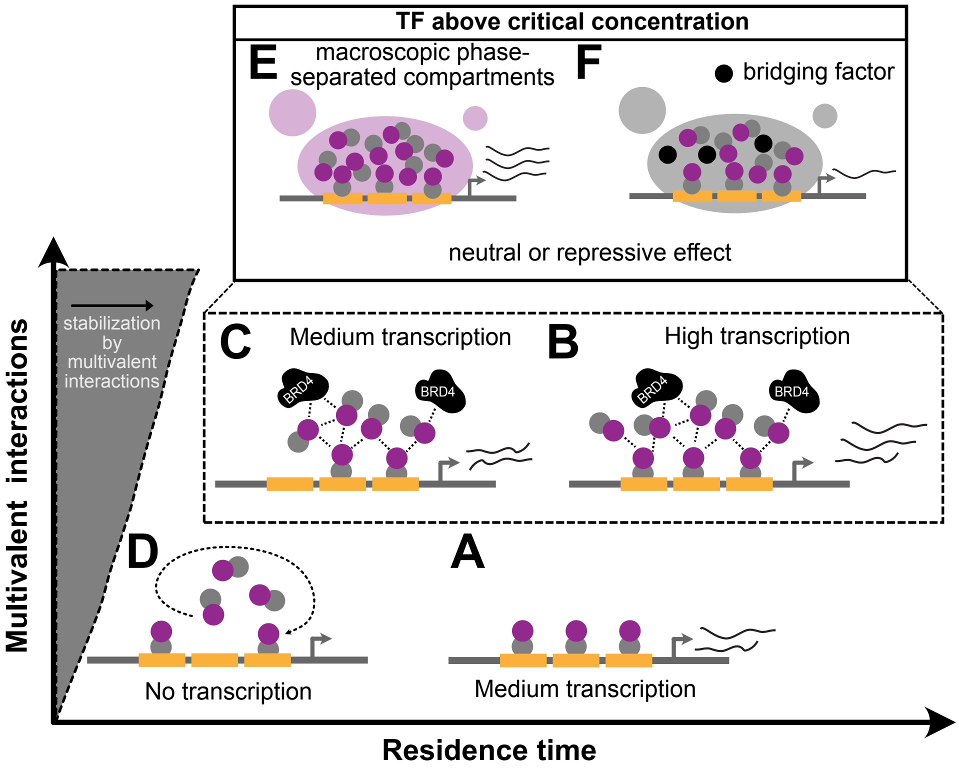


Figure 3.3: Integrative model of TF binding, multivalent interactions and transcriptional (co-)activation. The situation at the promoter is depicted for different values of the TF residence time and strengths of multivalent interactions. Strong multivalent interactions stabilize TF-DNA binding (shift of promoter situation to higher residence times indicated by grey area). For strong multivalent interactions (dashed box) the TF can undergo phase separation, if it is expressed above the critical concentration (solid box outside the coordinate system). **(A)** Medium level transcription, if all binding sites are occupied by a long residence TF without multivalent interactions. **(B)** Full activation, additional indirectly bound TFs and co-activator recruitment for a TF with long residence time and strong multivalent interactions. **(C)** Reduced activation at reduced residence time. Multivalent interactions keep local TF concentration relatively high and allow for co-activator recruitment. **(D)** No transcriptional induction by a short residence TF without multivalent interactions. **(E-F)** Formation of phase-separated transcription compartments for TFs with strong multivalent interactions and high expression levels. **(E)** No additional activation in cells with visible TF droplets. **(F)** Repression of transcription for droplets that are enforced by multivalent bridging factors.

This effect did not depend on the formation of macroscopic phase-separated compartments, which would form, if the nuclear TF concentration exceeded a critical value (box, **Figure 3.3 E-F**). Rather, a trend towards slower activation was observed. Enhanced droplet formation via addition of bridging factors had an inhibitory effect (**Figure**

3.3 F). It could be caused by a repressive organization of the promoter-tethered droplet or could result from sequestration of components of the transcription machinery/co-factors in optodroplets at other nuclear locations.

In summary, it is proposed that transcriptional induction requires a well-positioned activation domain with reasonable residence times at the promoter that can interact with the core transcriptional machinery, while co-activators can also be recruited to the promoter by more loosely bound TFs lacking a transcription-competent configuration. It was shown that the toolbox of modular synthetic transcription factors provides a versatile framework to resolve fundamental concepts of transcription factor AD-DBD interdependency. It is anticipated that further identification of these concepts will help in designing strong and specific synthetic transcription factors and potentially open ways to specifically interfere with the transcriptional induction for some classes of TFs and target genes.

3.4 Conclusions and Perspective

The results of this thesis provide ample evidence that the AD and the DBD are functionally linked with respect to the capacity of a given TF to activate transcription: Interactions of the AD can influence the binding stability of the DBD and the mode of recruitment by the DBD can influence the efficiency of the activation via binding site occupancy, residence time or spatial orientation of the bound TF.

The results of this thesis show that short TF residence times can limit transcriptional activation. This finding indicates that transcriptional induction is another biological process that relies on a kinetic proofreading mechanism [196]. This mechanism has already been found experimentally for DNA damage sensing [212], translation [213] and antigen recognition by the T-cell receptor [214]. Here, it was shown for the first time that TF residence time regulates transcription independent of (and in addition to) binding site occupancy. Identifying the molecular mechanism that underlies the residence time dependency could in the future allow to understand which steps of transcriptional induction are limiting for expression levels. These could, for instance, be modification of the promoter-bound TF as proposed here, or nucleosome remodeling suggested by others [80].

The ability of intrinsically disordered ADs to form phase-separated droplets in vitro

[39] has been invoked to rationalize the local enrichment at highly transcribed genes, e.g. via super-enhancers [100]. However, the functional consequences of multivalent interactions between TFs, their local enrichment and recruitment of co-activators for the transcriptional activation process have remained unclear. The experiments of this study demonstrate that the formation of a liquid activator compartment at the promoter is not necessarily linked to enhanced transcription. PHR-AD optodroplets had a neutral or in some conditions even an inhibitory effect. Consequently, the formation of phase separated transcription compartments is probably not a necessary mechanism by which TFs with strong multivalent interactions achieve high levels of gene activation. The formation of PHR-AD optodroplets proved to be a useful assay to detect multivalent interactions of the AD, even though the presence of optodroplets was not necessary for strong and fast activation.

Amplification of phase-separating TFs like MYC [39, 215] or increase of the multivalent interactions of endogenous TFs by fusion events [8] could be disease mechanisms in cancer. Interestingly, fusing the FUSN domain, implicated in several oncofusion proteins, to VP16 created a very strong AD with high multivalent interactions. This shows that the toolbox of synthetic TFs used here could also be useful to study the consequences of oncogenic TF fusions. Inhibition of co-activators has been suggested as a therapeutic approach that predominantly targets super-enhancers [38], which potentially rely on multivalent interactions of their activators [40]. Besides overcoming technical challenges, like the short half-life of the BRD4 inhibitor JQ1, such approaches should check, if and how interactions between TFs and co-activators specifically enhance transcription. The results of this thesis underline that the simple accumulation of regulators at a target gene must not necessarily be functionally relevant. A better understanding of the exact mechanism that links TF multivalent interactions with strong activation and co-activation could help to improve such therapeutic interventions or open new ways to specifically target TF ADs, which have been considered to be undruggable [216]. For instance, the repressive effect of TF droplets enforced by bridging factors discovered in this thesis could be harnessed to design specific inhibitors for individual TFs that are deregulated in tumor cells. Such bridging factors would require a domain that interacts with the TF of interest (e.g. the IDR from the same TF) and a multimerization domain. The experimental framework developed in this thesis could be used to develop and test such repressive bridging factors and verify their specific effect for certain activation domains before moving towards more disease relevant model systems.

The findings of this thesis also have implications for the construction of synthetic TFs. Reprogramming of cells by CRISPR-based synthetic TFs has been suggested as a therapeutic tool [60, 217]. The two TF characteristics that were studied here, residence time and multivalent interactions, can be adjusted in engineered constructs to tune the strength and induction speed of synthetic activators. The results presented here indicate that the generally high specificity of dCas9 binding [218] could be reduced by multivalent interactions of a fused AD which stabilize weakly bound dCas9-AD molecules at off-target sites. On the other hand, the endogenous kinetic proofreading mechanism for TF binding identified in this study could also be adapted for the construction of synthetic reaction networks that lead to the detection or activation of genomic loci with high specificity.

In summary, the results of this thesis show that transcription is not only regulated by equilibrium binding of TFs at their target genes. Additional TF properties, like the residence time, can increase the specificity of transcription regulation that is robust to variations of TF expression levels. The findings obtained here raise a number of questions that need to be further investigated: What is the mechanism by which the general transcription machinery is able to read out these TF properties in terms of activation strength? How are TF features connected to other transcriptional phenomena like transcriptional bursting and induction of transcription in a multistep process? Moreover, the results of this thesis highlight the need to critically check claims that the formation of phase-separated TF compartments would amplify gene expression or increase transcription activation. Finally, the discovery of inhibitory liquid-like TF compartments opens up the possibility to design synthetic inhibitors that sequester TFs in an inactive state as a new strategy for the inhibitory targeting of TFs.

Chapter 4

Materials and Methods

4.1 Materials

4.1.1 Plasmids and cloning

Plasmids were constructed using classical restriction enzyme based cloning, site-directed mutagenesis (SDM) or Gibson assembly. **Table 4.1** summarizes plasmids constructed in this thesis. **Table 4.2** lists additional plasmids that were used. Construct expression was typically under the control of a CMV promoter with exceptions for EGFP-LacI under a SV2 promoter and guide RNA constructs under a U6 promoter.

Table 4.1: Plasmids constructed in this thesis.

Name	Description
CIBN-rTetR	rTetR from Tet-On 3G system (TakaraBio). Contains T2A-Puro resistance marker
dCas9-EGFP4-VPR	EGFP tetramer sequence provided by Fabian Erdel
PHR-EGFP	Contains same NLS as VP16 constructs
PHR-EGFP-VP16	FP exchange from Addgene #103822
PHR-EGFP-VPR	VPR domain from Addgene #63798
PHR-EGFP-p65	p65 domain from Addgene #63798
PHR-EGFP-Rta	Rta domain from Addgene #63798
PHR-EGFP-STAT2	STAT2 activation domain (1-10 and 722-857) from [219]
tdMCP-tdTomato	From Addgene #40649 and #54642 with TATA-box of the promoter removed
tdPCP-EGFP-VP16	TATA-box of the promoter removed

dCas9-EGFP-VP16

Table 4.2: Additional plasmids.

Name	Description	Source
rTetR-EGFP	Contains same NLS as VP16 constructs	[193]
rTetR-EGFP-VP16		[193]
rTetR-EGFP-VPR		[193]
EGFP-LacI		[220]
CIBN-LacI		[12]
SNAPtag-LacI	LacI in pSNAPf vector (New England Biolabs)	[193]
dCas9-EGFP		[168, 169]
dCas9-EGFP-VPR		[168, 169]
dCas9-EGFP-p300	p300 core domain from Addgene #61357	[193]
PHR-EGFP-FUSN	FUSN from [221] Addgene #122148	[193]
PHR-EGFP-FUSN-VP16	FUSN from [221] Addgene #122148	[193]
PHR-GBP	GBP from ref. [222]	[193]
tdPCP-EGFP	Tandem PCP from Addgene #40650	[193]
tdPCP-EGFP-VPR	TATA-box of the promoter removed	[193]
tdPCP-CIBN	TATA-box of the promoter removed	[193]
mCherry-BRD4	Murine BRD4 from ref. [145]	[193]
pU6-tetO-sgRNA-2xPP7	sgRNA targeting *tetO* sites	[193]
pU6-lacO-sgRNA-2xPP7	sgRNA targeting *lacO* sites	[193]

4.1.2 Microscopy instrumentation

FRAP, time course and optodroplet formation experiments were performed on a Zeiss AxioObserver widefield microscope (Zeiss, **Table 4.3**). Imaging of smRNA FISH, in SRRF mode and for (profile) enrichment analysis were performed on a Dragonfly 505 multi-modal spinning disc microscope (Andor, **Table 4.4**).

Table 4.3: Zeiss AxioObserver widefield microscope equipment.

Component	Use	Manufacturer
AxioObserver Z1	microscope	Zeiss
Plan-Apochromat 20x/0.8 M27	air objective	Zeiss
Plan-Apochromat 100x/1.46 Oil DIC M27	objective	Zeiss
AxioCam MRm Rev.3 monochrome	camera	Zeiss
HXP 120V	fluorescence light source	Leistungselektronik JENA GmbH
F46-002	GFP: 470/40;495;525/40	AHF
43 HE (modified)	tdTomato: 535/30;570;572/25	Zeiss
63 HE (modified)	mCherry: 550/25;590;629/62	Zeiss
Zen 2012 pro	multi-position time-lapse z-stack acquisition	Zeiss
70 mW bleach laser	473 nm laser	Rapp OptoElectronic GmbH
UGA40	laser scanning system	Rapp OptoElectronic GmbH

Table 4.4: Andor Dragonfly 505 spinning disc equipment.

Component	Use	Manufacturer
Nikon Ti2-E	microscope	Nikon
CFI Plan Fluor 40x Oil 1.30/0.20	objective	Nikon
CFI SR HP Apochromat Lambda S 100x	objective	Nikon
Laser diodes	405 nm (tagBFP); 488 nm (GFP); 561 nm (tdTomato/mCherry); 673 nm (Alexa 633)	Andor
Quad band dichroid	405/488/561/640	Andor

450/50	emission filter (tagBFP)	Andor
525/50	emission filter (GFP)	Andor
600/50	emission filter (tdTomato/mCherry)	Andor
700/75	emission filter (Alexa 633)	Andor
iXon Ultra 888 EM-CCD	camera	Andor
incubation chamber	5% CO2; 37 ℃	Okolab
Fusion software	multi-position/time course/SRRFstream acquisition	Andor

4.1.3 Software

Image analysis was performed in the R programming language [166] in RStudio [223] using the *EBImage* package [167]. The *ReacTran* package [191] was used for numerical simulations of reaction-diffusion processes in FRAP. Two software packages were developed in this thesis and are made available with the respective publications: *NSSQ* provides functions for the analysis of transcription induction time courses [165] and *NSQFRAP* contains functions for image processing and fitting of recovery curves for FRAP experiments [193]. The packages are available at https://github.com/RippeLab/NSSQ and at https://github.com/RippeLab/NSQFRAP. Additional packages that were used were: *plyr*, *reshape2* and *ggplot2*. FIJI/ImageJ [224] was used for image inspection, file conversion to the tiff format, cropping and contrast adjustments of exemplary images.

4.2 Methods

4.2.1 Cell culture

U2OS 2-6-3 cells were cultured in low glucose (1g/l) DMEM medium (Gibco) supplemented with 2 mM L-glutamine (PAN-Biotech), penicilin/streptomycin (100 U ml^{-1}, 0.1 mg ml^{-1}, PAN-Biotech) and 10% fetal calf serum (FCS, tetracycline-free, PAN-Biotech) using standard cell culture methods. Cells were sub-cultured between 1:5 and 1:15 every three days. For microscopy experiments cells were seeded in 8-well LabTek chambered coverglasses (Thermo Fisher Scientific) at a density of $1 \cdot 10^4$ cells per well on day 0. On day 1 the medium was replaced with FluoroBrite imaging medium (Gibco) with identical supplementation and cells were transfected with plasmid mix using Xtreme Gene 9 (Roche) according to the manufacturer's instructions. Briefly, 20 µl OptiMem (Gibco), 0.6 µl transfection reagent and 200 ng plasmid mix were used. Plasmid mixes contained equal amounts (by mass) of all constructs. For transfections with guide RNA expression plasmids 100 ng were guide RNA plasmid and the remaining 100 ng were split equally between the remaining plasmids. Cells were imaged on day 2 (24 h post-transfection) for time course or enrichment analysis experiments or on day 3 (48 h post-transfection) for FRAP experiments. Cells transfected with light-sensitive constructs were protected from light exposure. They were stored in colored light-impermeable styrofoam boxes and handled under red LED light. For long-term activation experiments, doxycycline (Sigma-Aldrich) was added directly after transfection (final concentration 5 µg/ml) and cells were illuminated by diffuse white LED light.

4.2.2 Light-induced time course experiments

A detailed protocol for performing light-induced transcription time course experiments has been published [165]. Combinations of transfected constructs are given in **Table 4.5**. Transfected cells were protected from light exposure and all handling before the start of imaging was done in the dark using red LED safelights. For conditions in which BRD4 inhibition was required JQ1(+) (Sigma-Aldrich) was added to a final concentration of 1 µM 3 h before the start of image acquisition. 15 min before the start of imaging doxycycline (Sigma-Aldrich) was added to individual LabTek wells to a final concentration of 5 µg ml^{-1}. Imaging on the AxioObserver Z1 widefield microscope (Zeiss) comprised a GFP (20% intensity, 150 ms exposure) and a tdTomato channel (30% intensity, 150 ms exposure), three z-slices with 1.0 µm distance, 16 positions in a 4x4 grid with an overlap of -50%. The focus plane was found by red transmitted light that did not trigger PHR activation. Time courses typically had a duration of 90 min with 2 min intervals. To achieve higher signal strength pixels were binned 2x2 and the NIR mode of the camera was activated. The focus was kept constant over the time course with the Definite Focus hardware autofocus (Zeiss). Several LabTeks were imaged in an alternating order, so that PHR molecules in neighboring wells that had been exposed to stray light, had time to revert to their inactive conformation before the start of the next time course. These settings were modified for recording mCherry-BRD4 recruitment. The mCherry channel (30 % intensity, 250 ms exposure) was used and the duration of the time course was reduced to 60 min. Reactivation experiments were performed with 3 min intervals and for durations of 130 min, 150 min and 180 min with a dark period without imaging of 60 min. Doxycycline was added in the beginning to all wells and only a single slide was used, so that the results of the reactivation experiments could be biased by some pre-activation (see discussion in section 2.1.6).

Table 4.5: Combinations of transfected plasmids.

Experiment	Constructs	Treatment
Reactivation	CIBN-rTetR + PHR-EGFP-VP16 + tdMCP-tdTomato	-
AD comparison	CIBN-rTetR + PHR-EGFP-VP16/-STAT2/-p65/-Rta/-VPR + tdMCP-tdTomato	-
Activation with enforced droplets	CIBN-rTetR (+ CIBN-LacI)(+PHR-GBP) (+EGFP-LacI) + PHR-EGFP-VP16/-FUSN/-FUSN-VP16 + tdMCP-tdTomato	-
BRD4 recruitment dynamics	dCas9 + tetO-sgRNA-2xPP7 + lacO-sgRNA-2xPP7 + tdPCP-CIBN + PHR-EGFP-VP16/-VPR +mCherry-BRD4	± JQ1
BRD4 independent transcription	CIBN-rTetR + PHR-EGFP-VP16 + tdMCP-tdTomato	JQ1
Activation of hyperacetylated promoter	dCas9-EGFP-p300/-EGFP + tetO-sgRNA-2xPP7 + CIBN-rTetR + PHR-EGFP-VP16/-VPR + tdMCP-tdTomato	-

4.2.3 Analysis of time course experiments

The analysis of transcription induction image series comprised segmentation of the nucleus and transcription spot, object tracking and intensity quantification. Image analysis was performed with the *NSSQ* package according to the workflow depicted in **Figure 2.2**. Nuclei with successful PHR-EGFP-AD recruitment to the gene cluster spot were selected manually in the central z-slice of the third time frame. The following steps were fully automated. At each selected position the nucleus was segmented in each time frame by local thresholding, nuclei were tracked by finding the closest object in consecutive time frames starting from the third frame and moving outwards. For each time point the best z-position was selected as the z-slice with the largest average intensity gradient thereby providing a “belated auto-focus” routine. The gene cluster spot was segmented inside the segmented nucleus region by a intensity quantile based threshold. The brightest spot object in the third time frame was selected and tracked forward and backward in time. Additional ring shaped masks were created around the nucleus and the spot for local background measurements. The resulting time series of segmented nucleus and spot were visually inspected in the reader channel (tdMCP-tdTomato or mCherry-BRD4) for expression of the reader construct, normal morphology, successful segmentation and accumulation of reader in the spot area. Cells that did not meet these criteria were sorted out, while the remaining cells were annotated as either responders or non-responders depending on the visible accumulation of reader in the spot. The average fluorescence intensity in the segmented regions was measured and saved for each individual cell. The integrated intensity at the gene cluster was calculated by subtraction of background values and multiplication by the spot area according to equation 2.1. The value of the first time point was subtracted so that all time courses start at 0. To account for the small time-shift of image acquisition between grid positions the intensity at the start of each imaging cycle was estimated by linear interpolation. These interpolated values could be used for averaging and for normalization. Normalized time courses $R_{norm}(t)$ were calculated by subtracting the value of the first time point and dividing by the intensity maximum of the full time course:

$$R_{norm}(t) = \frac{R(t) - R(0)}{max(R(t) - R(0))} \tag{4.1}$$

Times to half-maximum activation were determined for single cells as the time at which their normalized time course reached or exceeded a value of 0.5. The time course amplitude was determined as the mean of the last five time points of $R(t)$ corresponding to the plateau value.

4.2.4 Measurement of PHR binding kinetics

To measure dissociation of PHR-AD from CIBN-rTetR after a short or long activation period, cells transfected with CIBN-rTetR, PHR-iRFP-VP16 and EGFP-LacI were imaged on the Dragonfly spinning disc microscope. Channel settings were 2% intensity and 30 ms exposure for GFP and 50% intensity and 200 ms exposure for iRFP. z-stacks of five slices with a distance of $1.0\,\mu\mathrm{m}$ were recorded at 12 positions on a 4x3 grid with 1% overlap. Time courses comprised an activation phase of 8 min (short) or 60 min (long) with the GFP channel and a dark phase without of 30 min or 120 min with 2.5 min intervals.

Nuclei and spot regions were selected manually, segmented by local thresholding and tracked manually using functions from the *NSSQ* package. Time courses were normalized to the start value.

4.2.5 SRRF imaging of decondensed reporter locus

Cells transfected with SNAPtag-LacI, dCas9-EGFP-VPR and tetO-sgRNA-2xPP7 (or with the respective components of the rTetR opto VPR or opto VPR complexes) were labeled with SNAP-Cell 647-SiR substrate at a concentration of $3\,\mu\mathrm{M}$ in medium for 30 min. This was followed by three washes with medium, one wash for 30 min and three washes with PBS. Fixation took 10 min in 4% para-formaldehyde (PFA) and was followed by one PBS wash. Cells were stored in PBS for imaging. SRRF measurements [170] used the SRRF stream module of the Fusion software (Andor) on the Dragonfly spinning-disc microscope. The 100x objective was used in combination with an additional 2x magnification lense for over-sampling. One SRRF image was recorded using 200 frames with 2.5 ms exposure, 100% laser intensity of the 488 nm or 633 nm laser. The on-line image analysis used 5x5 sub-pixels and a 1.5 pixel ring radius for radiality calculations. Radiality images were mean-projected.

4.2.6 Single molecule RNA FISH

Single molecule RNA FISH was done with the RNAscope fluorescence multiplex reagent kit (ACD Bio) [225] using custom designed target probes (ACD Bio) for the MS2 sequence between bases 851 and 2163 of the reporter RNA. Transfected and untransfected cells were illuminated for 24 h and fixed with 4% PFA for 12 min. Cells were washed three times with PBS, treated with 3% hydrogen peroxide, washed once with PBS and exposed to Protease III (1:15) for 10 min followed by three PBS washes. Target, amplification and C1 detection probes labeled with Alexa 488 were hybridized consecutively following the manufacturer instructions by adding the reagents directly to the LabTeks wells in amounts sufficient to cover the full surface. Between hybridization steps washing buffer (400 µl) was added directly to the wells. Cells were stored in PBS supplemented with 0.1% sodium azide for conservation and imaged in PBS on the Dragonfly spinning-disc microscope with and without SRRF mode using the 100x objective. Nascent RNA numbers were estimated by comparing the integrated intensity of single molecule spots with the reporter gene cluster in a single z-slice. RNA numbers in nucleus and cytoplasm were estimated by counting spots in a region of defined area in each compartment and extrapolating to the whole compartment area in a SRRF mode image.

4.2.7 Optodroplet propensity measurements

Cells transfected with PHR-EGFP-AD and CIBN-localizer were imaged with the same settings on the AxioObserver widefield microscope as described for the transcription induction time courses with only the GFP channel for six cycles at 25 positions. All conditions were imaged together, so that absolute intensities are comparable. The image analysis pipeline based on the *NSSQ* package from the induction time course experiments was adapted using a single channel and the gene cluster spot was removed from the nucleus mask. Besides manual control of segmentation quality also the presence or absence of visible optodroplets was annotated. The coefficient of variation (CV) was calculated as the ratio of standard deviation and mean of the intensity in the nucleus mask. From a plot of CV against nuclear intensity a value of 0.25 was determined as a threshold that yielded good separation between cells annotated as droplet containing and not. The CV was fitted as a function of nuclear intensity by a logistic function in order to obtain a smoothed curve:

$$D(c) = A + \frac{B}{1 + exp(-k \cdot (c - c_0))} \tag{4.2}$$

Here, $D(c)$ is the droplet abundance measured by the CV and c the nuclear concentration. The nuclear intensity at which the fit crossed the threshold value of 0.25 was determined as the critical concentration.

4.2.8 Simulations and fitting of stochastic promoter state models

Stochastic gene induction time courses of the multistep and the feedback model depicted in **Figure 2.9 A** were simulated with the Gillespie algorithm [181] using a base set of parameter values shown in the same figure. Briefly, each simulation run progressed in time by steps drawn from a exponential distribution and by updating the particle numbers according to random reactions drawn according to the reaction propensities. Each simulation run resulted in a single-cell trajectory. In each iteration step two random values r_1 and r_2 were drawn from a uniform distribution between 0 and 1 using the R-function *runif()* and then transformed to determine the time step size and the reaction. Each reaction was assigned an interval. The size of these intervals was determined from the reaction propensities which were calculated from the reaction rate constants and the numbers of the involved educt particles. Reaction propensities for the Multistep model were:

$$\begin{aligned}
a_1 &= k_a \cdot I_0 \\
a_2 &= k_a \cdot I_1 \\
a_3 &= k_a \cdot I_2 \\
a_4 &= k_a \cdot I_3 \\
a_5 &= k_a \cdot I_4 \\
a_6 &= k_a \cdot I_5 \\
a_7 &= k_a \cdot I_6 \\
a_r &= k_r \cdot A \\
a_0 &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_r
\end{aligned} \tag{4.3}$$

and for the Feedback model:

$$\begin{aligned} a_1 &= k_a \cdot I_0 \\ a_2 &= k_f \cdot I_1 \cdot A \\ a_r &= k_r \cdot A \\ a_0 &= a_1 + a_2 + a_r \end{aligned} \tag{4.4}$$

The time step size τ was determined as

$$\tau = \frac{-ln(r_2)}{a_0} \tag{4.5}$$

RNA levels were calculated deterministically by integrating the differential equation that describes the production of RNA from active promoters and its decay:

$$\frac{dR}{dt} = k_t \cdot A - k_d \cdot R \tag{4.6}$$

Integration yielded the RNA trajectory:

$$R(t) = k_t \cdot e^{-k_d \cdot t} \int_0^t A(t) e^{k_d \cdot t} dt \tag{4.7}$$

The integral was solved numerically by summing over the values of the time course $A(t)$ multiplied by the interval duration. Single-cell trajectories were then used to calculate population statistics comprising the mean trajectory, the (empirical) cumulative distribution function (ECDF) for each time point and the correlation matrix that represents the correlation of RNA levels at each pair of time points. For simulations of the base parameter set and for variations of this parameter set 50 time courses were simulated, while for the models with a varying fraction of initially activate promoters 100 simulations were run. The fraction of initially active promoters $f_A = \frac{A(t=0)}{N}$ was drawn from a uniform distribution from 0 to 0.1 (0.12 for the multistep model) for each run. Tentative model discrimination using simulated ground-truth data from simulations of the base parameter sets or using experimental data was performed with a particle swarm fitting algorithm. The implementation of this algorithm was adapted from [186]. In this approach each parameter set is represented by a “particle” that can move in paramter space. The velocity and direction of the particle in parameter space is determined by evaluating the objective function for the parameter set and moving the particle in the direction of the particle with the lowest objective function in combination with a random contribution. In this way particles converge in a region of the parameter space with minimal values of the objective function. The initial particle positions $\vec{p}$ were selected by latin hypersquare sampling and the initical velocity $\vec{v}$ was drawn from a uniform distribution covering the range of half the distance of the particle position and the parameter (upper or lower) boundary. In each iteration cycle the parameter sets corresponding to the particle positions were used to simulate 20 promoter state trajectories. In order to reduce the computational cost the average of these trajectories of state A were transformed and fitted to the (simulated) data using the *nls* function of R according to equation 4.7. This resulted in the fit parameters k_t and k_d, which were used to convert the individual simulated trajectories of state $A(t)$ to RNA trajectories $R(t)$. The objective function was calculated from the RNA trajectories as the weighted average of the sums of squared residuals of the average time course, the standard deviation, the correlation matrix and the empirical cumulative distribution function (ECDF).

$$\begin{aligned}
R_{mean}(t) &= mean(R(t)) \\
R_{sd}(t) &= sd(R(t)) \\
C_{ij} &= cor(R(t=t_i), R(t=t_j)) \\
ECDF(r,t) &= \frac{N_{r \geq R(t)}}{N} \\
OF &= a \cdot \sum_t (R_{mean,data}(t) - R_{mean,simulation}(t))^2 + \\
&\quad b \cdot \sum_t (R_{sd,data}(t) - R_{sd,simulation}(t))^2 + \\
&\quad c \cdot \sum_{i,j} (C_{ij,data} - C_{ij,simulation})^2 + \\
&\quad d \cdot \sum_{r,t} (ECDF(r,t) - ECDF(r,t))^2 \qquad (4.8)
\end{aligned}$$

For the mutual fitting of the Multistep and the Feedback models the weigthing factors (a, b, c, d) were set to $(0, 0, 1, 1)$ and for the fitting of experimental data of induction by VP16 to $(0.01, 0, 1, 1)$. For each particle two optimal parameter sets were determined. The particle optimum $\vec{p}_{best}$ corresponded to the position in parameter space that the particle had visited in all previous iterations with the minimal objective function. The global optimum $\vec{p}_{global}$ for each particle corresponded to the best parameter set of the current iteration cycle of the particle itself and two connected “neighbor” particles. The particles here were numbered and consecutive particles (and the last and first) were considered connected corresponding to a ring-like connection structure. The particles were then moved into the direction of their two optima by updating and adding their velocity $\vec{v}_{i+1}$ to their current position:

$$\vec{v}_{i+1} = w \cdot \vec{v}_i + \phi \cdot r_1 \cdot (\vec{p}_i - \vec{p}_{best}) + \phi \cdot r_2 \cdot (\vec{p}_i - \vec{p}_{global}) \tag{4.9}$$

The random variables r_1 and r_2 were drawn from a uniform distribution between 0 and 1, w was set to 0.721 and ϕ to 1.193 following ref. [186]. The new particle positions $\vec{p}_{i+1}$ were then used for the next iteration round $(i+1)$. If a particle position was outside the defined parameter range they were set to the respective parameter boundary. Moreover, to save simulation time, parameter sets (particle positions) that lead to unstable simulation behavior were replaced by new parameter sets from latin hypersquare sampling. This detection of unstable parameter sets was based on the steady state level of promtoter state A that was not allowed to be smaller than 10 to avoid stochastic “extinction” events of this molecular species. The steady state levels of A was determined from the (deterministic) ODE system. It was

$$A_{steadystate} = \frac{N}{7 \cdot \frac{k_r}{k_a} + 1} \tag{4.10}$$

for the Multistep model and

$$\begin{aligned} A_{steadystate,1} &= 0 \\ A_{steadystate,2} &= \frac{N - \frac{k_r}{k_f}}{1 + \frac{k_r}{k_a}} \end{aligned} \tag{4.11}$$

for the Feedback model. For the Feedback model a linear stability analysis of steady states additionally showed that the value λ_1 had to fulfill the following condition in order for the system not to approach the inactive first steady state:

$$\begin{aligned} \lambda_1 &= 0.5 \cdot \Bigg(- k_f \cdot A_{steadystate,2} - k_a + \\ &\sqrt{(k_f \cdot A_{steadystate,2} - k_a)^2 - 4 \cdot k_r \cdot k_f \cdot A_{steadystate,2}} \Bigg) \leq 0 \end{aligned} \tag{4.12}$$

Fitting was done using 30 particles over 30 iteration cycles. The resulting estimated parameter sets were used to simulate 50 trajectories of the active promoter state $A(t)$, which were fitted to the data to determine k_t and k_d and thereby the RNA trajectories, which were used for presentation of average time courses, correlation matrices and cumulative distributions.

4.2.9 FRAP: measurements

Widefield FRAP experiments were performed with the Zeiss AxioObserver Z1 microscope and the 20x objective. For the comparison of dCas9-EGFP-VPR with wt- and

C2G-tetO-sgRNA the 100x objective was used due to the smaller amount of constructs recruited to *tetO* instead of *lacO* sites. Imaging settings of the GFP channel were 20% excitation light intensity and 150 ms exposure. The 476 nm bleaching laser was controlled by a distinct software via the UGA40 controller. Laser positions imaged by the ZEN software of the Zeiss microscope were mirrored in the UGA40 software and the positioning was calibrated by a built-in routine using a coverslip labeled with yellow textmarker as an uniformly fluorescent calibration slide. Cells were illuminated with blue excitation light for at least 1 min by imaging in live mode to ensure full recruitment of light-responsive constructs before starting the imaging time course. On-spot experiments had a duration of 240 s with 1 s intervals, while off-spot experiments had a duration of 60 s with 0.3 s intervals. The bleach position was selected in the UGA40 software before starting the time course acquisition and the 100% (70 mW) 1 s bleach pulse was triggered manually from the UGA40 software around 3 s after the start of the time course. The bleach position was selected directly on the labeled gene cluster spot for on-spot experiments and outside the spot in the nucleus avoiding nucleolar regions for off-spot experiments.

Confocal FRAP measurements of EGFP-LacI were performed by Anne Rademacher on a Leica TCS SP5 II confocal microscope (Leica) using a 63x Plan-Apochromat objective. Briefly, imaging settings were 128x128 px images with zoom factor 9 (193.73 nm/px), 1400 Hz line (frame time 115 ms), 70 pre-bleach, 2 bleach and 1200 (on-spot) or 300 (off-spot) post-bleach frames and bleaching using 100% intensity of the 458nm, 476 nm, 488 nm, 496 nm Argon laser lines in a circular region with 1.0 µm diameter.

4.2.10 FRAP: image analysis

FRAP image series comprised 240 frames for on-spot and 200 frames for off-spot experiments. Image series were segmented, tracked and quantified using functions of the *NSQFRAP* package by the workflow depicted in **Figure 2.15**. This analysis comprised two steps. In the first step image series were automatically processed, while the second step comprised manual curation and intensity measurements. First, the bleach frame(s) were detected by their high intensity values and the coordinates of the bleach region were retrieved by segmenting the intensity difference of last pre- and first postbleach images. Then the image was cropped around these coordinates and the nucleus was segmented in all time frames using local thresholding with a iteratively reduced threshold until all frames contained a nuclear mask of the expected size covering the bleach position. For the dCas9-EGFP construct, which exhibits mainly cytoplasmic localization, images were blurred before segmentation, so that the whole cell instead of the nucleus was segmented. The bleach region was segmented, as well. For on-spot experiments the gene cluster spot was segmented inside the bleach region in the first (prebleach) frame using a quantile-based thresholding. Spot and bleach region masks were shifted for each time frame according to the movements of the center of the nucleus mask. Local background masks of the spot and the nucleus were created as a ring-shaped area around their reference masks. In the second step the segmented and tracked images were visually inspected and for on-spot experiments with failed tracking the spot tracking was manually corrected. If correction was not possible, cells were excluded from further analysis. After this curation step intensities in the segmented mask regions were measured for each time point. Mean intensities were measured inside mask regions and the intensity profile was measured around the center of the bleach region. The intensity profiles and the intensity time courses were normalized as described by equations 2.7 and 2.8.

4.2.11 FRAP: fitting of reaction-diffusion model

The reaction-diffusion model is described in section 2.2.2 and defined by equations 2.9 for on-spot and 2.10 for off-spot measurements. Numerical simulation runs for given parameter sets were implemented in R using the *ReacTran* package [191]. The nucleus was represented by a cylinder with 50 concentric grid cells. The initial conditions were determined by fitting the intensity profile of the bleach region in the first post-bleach frame to a gaussian with a central plateau as described by equation 2.11. Simulations resulted in concentration profiles for free f and bound c molecules across the nucleus radius. The concentration profiles were normalized in the same way as the image data:

$$y_{norm}(r,t) = (1 - f_i) \cdot \frac{f(r,t) + c(r,t)}{f(r, t = 240s) \cdot (1 + spotRatio \cdot bgRatio)} \tag{4.13}$$

These normalized profiles were converted to average normalized intensities by integrating the intensities up to a radius of $1.0\,\mu\text{m}$ for on-spot simulations. Single-cell FRAP curves were fitted individually. For each ligand-target combination first the off-spot experiments were fitted with the diffusion-only model to determine the effective diffusion coefficient D_{eff}, which was then used for simulations of the full reaction-diffusion system for on-spot fitting. The off-spot fitting procedure used the *nls* R-function providing multiple start values of the two fit parameters. Start values for D_{eff} were varied between 0.1 and $5\,\mu m^2/s$ and for the non-specifically bound immobile fraction f_i between 0.1 and 0.5. The individual on-spot recovery curves were fitted using a grid of parameter values, on which the sum of squared residuals of simulation vs. data was evaluated. There were three free fit parameters: the dissociation rate constant k_{off}, a correction factor $bgRatio$ that was multiplied with $spotRatio$ (determined from ratio of spot and background intensity according to equation 2.15) and an immobile fraction f_i. The immobile fraction was estimated separately in order to reduce the fitting parameter space. The recovery curve without the first 30 s of was fitted to a single exponential (equation 2.13) using the *nls* function with multiple start values. The immobile fraction was calculated from the off-set A and scaling factor B as $f_i = 1 - A - B$. The estimated value of f_i was then fixed and a grid of parameter values for the remaining two free parameters was created. The off-rate k_{off} was varied between $1 \cdot 10^{-4}s^{-1}$ and $0.1s^{-1}$ in seven irregular steps, and the correction factor bgRatio was varied between 1 and 3.45 in steps of 0.35. Simulations of the full reaction-diffusion system were run at each position using the median of D_{eff} values estimated for the respective constructs in off-spot experiments. The best fit was determined as the parameter pair with the smallest sum of squared residuals. In a second step the fitting was refined by varying k_{off} with a factor between 0.2 and 8 and $bgRatio$ in steps of 0.03 around the best fit parameter values. The parameter set with the best fit of all evaluations was then used. Fitted recovery curves of single cells as well as data points were averaged after fitting and used for plotting. The residence time of complexes was determined as $\frac{1}{k_{off}}$ after averaging.

4.2.12 Bulk RNA measurements by qRT-PCR

qRT-PCR experiments were performed and analyzed by Lukas Frank. Details on the full qRT-PCR methods can be found associated with reference [193]. Briefly, cells were grown and transfected in 6-well plates and exposed to diffuse white light depending on the activator system. RNA was purified after 24 h by phenol-chloroform extraction and DNase treated. After cDNA synthesis relative RNA levels were determined by qRT-PCR using SYBR Green PCR Mastermix (Applied Biosystems). Primers for beta-actin were 5'-TCCCTGGAGAAGAGCTACGA-3' (fwd) and 5'-AGCACTGTGTTGGCGTACAG-3' (rev), and for CFP-SKL of the reporter RNA 5'-GTCCGGACTCAGATCTCGA-3' (fwd) and 5'-

TTCAAAGCTTGGACTGCAGG-3’ (rev). Levels of reporter RNA were normalized to beta-actin and to levels of mock transfected cells.

4.2.13 Immunofluorescence staining

To detect enrichment of H3K27ac at the reporter gene cluster cells transfected with the respective activator constructs were illuminated by diffuse white LED light for 24 h. Cells were then washed once with PBS and fixed with 4% PFA for 12 min followed by permeabilization with 0.1% Triton-X in PBS for 5 min. 10% goat serum in PBS was used for blocking for 15 min. Incubation with the primary antibody (H3K27ac, 1:1000, rabbit, Abcam ab 4729, Lot GR183922-1) in PBS with 10% goat serum took 1 h and was followed by three washes with 0.002% NP40 in PBS for 5 min. Secondary antibody (goat anti-rabbit Alexa 633, 1:1000, Abcam, Lot 1073053) incubation in PBS with 10% goat serum took 30 min. Cells were washed twice with PBS for 5 min and stored and imaged in PBS.

4.2.14 Enrichment analysis: image acquisition

Images of fluorescent reporter spot signals were acquired on the Dragonfly 505 spinning disc microscope with the 40x oil immersion objective. Pin-hole size was 40 µm. These experiments were performed with two different purposes: (1) the qualitative detection of transcription marks with imaging parameters that were adapted individually for each sample and (2) quantitative intensity measurements to compare local concentrations using fixed imaging parameters. Before imaging cells were stored 24 hours after transfection and were constantly illuminated by diffuse white LED light in this period. Cells for quantitative intensity comparison did not express light-responsive proteins and were not illuminated. For each sample 81 positions (9x9 grid, 1% overlap) were imaged in 15 z-planes (distance 1.0 µm). EM-Gain was 200 for all channels. GFP was imaged with excitation intensities between 2 and 20%, 10 to 200 ms exposure; tdTomato and mCherry with 3 to 40%, 40 to 300 ms; tagBFP with 6 to 14%, 100 to 300 ms; Alexa 633 with 5 to 50%, 30 to 300 ms. For the comparison of local concentrations of tdPCP-EGFP-VP16 and tdPCP-EGFP-VPR channel settings were: GFP 3%, 50 ms; tagBFP: 14%, 150 ms. For the comparison of dCas9-EGFP-VP16 and dCas9-EGFP-VPR occupancy with or without guide RNA mutations channel settings were: GFP 3%, 25 ms;

tagBFP 14%, 100 ms; and for the comparison of promoter activity: tdTomato 10%, 80 ms; GFP 3%, 35 ms; tagBFP 14%, 300 ms.

4.2.15 Enrichment analysis: absolute intensities and relative profiles

Images were processed in R and RStudio using the packages *EBImage* and *NSSQ*. Maximum projections of the 15 z-planes for each position were used to manually select nuclei with successful effector recruitment in the EGFP-AD channel. The nucleus was segmented in sum-projected images by local thresholding. The three consecutive z-planes with the highest contrast (determined from the gradient inside the nucleus area) were selected and mean-projected to yield a single image for each channel for quantification. The spot position was selected manually in each segmented cell and a disc shaped spot mask with a diameter of 1.6 µm (5 pixels) for EGFP-AD and of 3.8 µm (12 pixels) for tdMCP-tdTomato labeling nascent RNA created. Average intensities in all channels were measured in the nucleus and spot area and in ring shaped areas around them. Radial profiles were measured in all channels by creating masks of concentric rings of pixel-wise increasing radius around the selected spot position (see **Figure 2.22 A**) and measuring average intensities. Radial profiles extended to a radius of 2.9 µm (9 pixels). The minimum value was subtracted from the profiles and they were divided by the local background intensity for normalization. Then for each condition all single-cell profiles were averaged and the minimum value subtracted. The resulting enrichment score profile gives qualitative information about the presence or absence of accumulated intensity in the spot center for each channel. In conditions with strong recruitment most cells will have the maximum intensity at the spot center yielding an enrichment score profile with high amplitude. In conditions without recruitment the maximum will be distributed over the profile between cells. The resulting normalized profile will have a low amplitude after minimum subtraction. The enrichment profiles represent a qualitative readout for intensity enriched at the spot center compared to the periphery, which is useful for comparing cells imaged with different imaging parameters, but cannot be related directly to the absolute amount of activator that is recruited. For quantitative comparisons of local concentrations average spot intensities were measured in images acquired with the same parameters. The average intensity in the spot background region was subtracted from the average spot intensity in the EGFP-AD or in the tdMCP-tdTomato channel. The background corrected intensity was normalized to the tagBFP-marker intensity.

4.2.16 Statistical tests and data reporting

Average time courses for transcription induction experiments and recovery curves were displayed with the 95% confidence interval (CI) calculated from the t-distribution. Error bars are one standard deviation for qRT-PCR plots and the standard error of the mean (s.e.m.) for half activation times. Statistical tests comprised unpaired Student's t-tests for qRT-PCR experiments, Welch's t-tests for comparisons of local concentrations and half-activation times, as well as two-way ANOVA (type II) for the comparison of activation dynamics in dependence of the AD and presence of droplets and for the comparison of nascent RNA levels in dependence of the occupancy group and the type of sgRNA. In cases, in which the data points spanned a wide range of values due to outlier values, axis breaks were introduced in order to visualize all of the data in a meaningful way. These breaks were clearly indicated and mentioned in the figure legends. Exemplary images displayed in the figures were contrast adjusted without non-linear transformations using the same adjustments for all frames of a time series.

Bibliography

1. Zhang, Z. *et al.* Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes & Development* **30,** 2106–2118 (2016).
2. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IkB-NF-kB Signaling Module: Temporal Control and Selective Gene Activation. *Science* **298,** 1241–1245 (2002).
3. Sampattavanich, S. *et al.* Encoding Growth Factor Identity in the Temporal Dynamics of FOXO3 under the Combinatorial Control of ERK and AKT Kinases. *Cell Systems* **6,** 664–678 (2018).
4. Harton, M. D., Koh, W. S., Bunker, A. D., Singh, A. & Batchelor, E. p53 pulse modulation differentially regulates target gene promoters to regulate cell fate decisions. *Molecular Systems Biology* **15,** e8685 (2019).
5. Ferraro, T. *et al.* Transcriptional Memory in the Drosophila Embryo. *Current Biology* **26,** 212–218 (2016).
6. Kamada, R. *et al.* Interferon stimulation creates chromatin marks and establishes transcriptional memory. *Proceedings of the National Academy of Sciences* **115,** E9162–E9171 (2018).
7. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
8. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171,** 163–178.e19 (2017).
9. Srivastava, D. & DeWitt, N. In Vivo Cellular Reprogramming: The Next Generation. *Cell* **166,** 1386–1396 (2016).

10. Morel, D., Jeffery, D., Aspeslagh, S., Almouzni, G. & Postel-Vinay, S. Combining epigenetic drugs with other therapies for solid tumours - past lessons and future promise. *Nature Reviews. Clinical Oncology* **17,** 91–107 (2020).

11. Lee, T. I. & Young, R. A. Transcription of Eukaryotic Protein-Coding Genes. *Annual Review of Genetics* **34,** 77–137 (2000).

12. Rademacher, A., Erdel, F., Trojanowski, J., Schumacher, S. & Rippe, K. Real-time observation of light-controlled transcription in living cells. *Journal of Cell Science* **130,** 4213–4224 (2017).

13. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology* **16,** 1128–1133 (2009).

14. Ardehali, M. B. & Lis, J. T. Tracking rates of transcription and splicing in vivo. *Nature Structural & Molecular Biology* **16,** 1123–1124 (2009).

15. Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. *Cell* **136,** 688–700 (2009).

16. Hochberg-Laufer, H. *et al.* Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLOS Genetics* **15,** e1008459 (2019).

17. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322,** 1845–1848 (2008).

18. Krebs, A. R. *et al.* Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell* **67,** 411–422.e4 (2017).

19. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development* **26,** 2119–2137 (2012).

20. Lu, H., Zawel, L., Fisher, L., Egly, J. M. & Reinberg, D. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358,** 641–645 (1992).

21. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of Biological Chemistry* **270,** 12335–12338 (1995).

22. Zawel, L., Kumar, K. P. & Reinberg, D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes & Development* **9,** 1479–1490 (1995).

23. Teves, S. S. *et al.* A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes. *eLife* **7,** e35621 (2018).

24. Karlic, R., Chung, H.-R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences* **107,** 2926–2931 (2010).

25. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* **21,** 381–95 (2011).

26. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351,** 720–724 (2016).

27. Weinert, B. T. *et al.* Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell* **174,** 231–244.e12 (2018).

28. Black, J. C., Choi, J. E., Lombardo, S. R. & Carey, M. A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Molecular Cell* **23,** 809–818 (2006).

29. Gorisch, S. M., Wachsmuth, M., Toth, K. F., Lichter, P. & Rippe, K. Histone acetylation increases chromatin accessibility. *J Cell Sci* **118,** 5825–34 (2005).

30. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proceedings of the National Academy of Sciences* **100,** 8758–63 (2003).

31. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19,** 535–45 (2005).

32. Devaiah, B. N. *et al.* BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. *Proceedings of the National Academy of Sciences* **109,** 6927–6932 (2012).

33. Jackson, D. A., Iborra, F. J., Manders, E. M. & Cook, P. R. Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei. *Molecular Biology of the Cell* **9,** 1523–1536 (1998).

34. Li, J. *et al.* Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178,** 491–506 (2019).

35. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *Embo J* **12,** 1059–65 (1993).

36. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–19 (2013).

37. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–47 (2013).

38. Lovén, J. *et al.* Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* **153,** 320–334 (2013).

39. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175,** 1842–1855.e16 (2018).

40. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361,** eaar3958 (2018).

41. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural & Molecular Biology* **25,** 833 (2018).

42. Dufourt, J. *et al.* Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nature Communications* **9,** 5194 (2018).

43. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol* **4,** e309 (2006).

44. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Current Biology* **16,** 1018–1025 (2006).

45. Molina, N. *et al.* Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proceedings of the National Academy of Sciences* **110,** 20563–20568 (2013).

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

47. Cesbron, F., Oehler, M., Ha, N., Sancar, G. & Brunner, M. Transcriptional refractoriness is dependent on core promoter architecture. *Nature Communications* **6,** 6753 (2015).

48. Li, C., Cesbron, F., Oehler, M., Brunner, M. & Höfer, T. Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. *Cell Systems* **6** (2018).

49. Zoller, B., Little, S. C. & Gregor, T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell* **175,** 835–847 (2018).

50. Das, D., Dey, S., Brewster, R. C. & Choubey, S. Effect of transcription factor resource sharing on gene expression noise. *PLOS Computational Biology* **13,** e1005491 (2017).

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

52. Nicolas, D., Zoller, B., Suter, D. M. & Naef, F. Modulation of transcriptional burst frequency by histone acetylation. *Proceedings of the National Academy of Sciences* **115,** 7153–7158 (2018).

53. Sanchez, A. & Golding, I. Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science* **342,** 1188–1193 (2013).

54. Padovan-Merhar, O. *et al.* Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Molecular Cell* **58,** 339–352 (2015).

55. Bain, D. L., Heneghan, A. F., Connaghan-Jones, K. D. & Miura, M. T. Nuclear receptor structure: implications for function. *Annual Review of Physiology* **69,** 201–220 (2007).

56. Hirai, H., Tani, T. & Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *The International Journal of Developmental Biology* **54,** 1589–1596 (2010).

57. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development* **25,** 2227 (2011).

58. Lupo, A. *et al.* KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Current Genomics* **14,** 268–278 (2013).

59. Lewis, M. *et al.* Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science* **271,** 1247–1254 (1996).

60. Pandelakis, M., Delgado, E. & Ebrahimkhani, M. R. CRISPR-Based Synthetic Transcription Factors In Vivo: The Future of Therapeutic Cellular Programming. *Cell Systems* **10,** 1–14 (2020).

61. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* **10,** 252–263 (2009).

62. Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome Biology* **1,** gb-2000-1-1-reviews001 (2000).

63. Ramos, J. L. *et al.* The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews* **69,** 326–356 (2005).

64. Stewart, A. J., Hannenhalli, S. & Plotkin, J. B. Why Transcription Factor Binding Sites Are Ten Nucleotides Long. *Genetics* **192,** 973–985 (2012).

65. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Research* **21,** 447–455 (2011).

66. Hippel, P. H. v. & Berg, O. G. On the specificity of DNA-protein interactions. *Proceedings of the National Academy of Sciences* **83,** 1608–1612 (1986).

67. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Research* **29,** 2860–2874 (2001).

68. Gerland, U., Moroz, J. D. & Hwa, T. Physical constraints and functional characteristics of transcription factor–DNA interaction. *Proceedings of the National Academy of Sciences* **99,** 12015–12020 (2002).

69. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development* **15,** 116–124 (2005).

70. González, A., Jafari, S., Zenere, A., Alenius, M. & Altafini, C. Thermodynamic model of gene regulation for the Or59b olfactory receptor in Drosophila. *PLOS Computational Biology* **15,** e1006709 (2019).

71. Suter, D. M. Transcription Factors and DNA Play Hide and Seek. *Trends in Cell Biology* **30,** 491–500 (2020).

72. Normanno, D. *et al.* Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nature Communications* **6,** 7357 (2015).

73. Slutsky, M. & Mirny, L. A. Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. *Biophysical Journal* **87,** 4021–4035 (2004).

74. Xiong, K. & Blainey, P. C. Molecular sled sequences are common in mammalian proteins. *Nucleic Acids Research* **44,** 2266–2273 (2016).

75. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science* **332,** 475–478 (2011).

76. Loffreda, A. *et al.* Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nature Communications* **8,** 313 (2017).

77. Clauß, K. *et al.* DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Research* **45,** 11121–11130 (2017).

78. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484,** 251–255 (2012).

79. Hu, B. *et al.* Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Research* **43,** e132 (2015).

80. Shelansky, R. & Boeger, H. Nucleosomal proofreading of activator–promoter interactions. *Proceedings of the National Academy of Sciences* **117,** 2456–2461 (2020).

81. Liu, J. *et al.* Intrinsic Disorder in Transcription Factors. *Biochemistry* **45,** 6873–6888 (2006).

82. Piskacek, M., Havelka, M., Rezacova, M. & Knight, A. The 9aaTAD Transactivation Domains: From Gal4 to p53. *PLOS ONE* **11,** e0162842 (2016).

83. Mitchell, P. J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245,** 371–378 (1989).

84. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **6,** 197–208 (2005).

85. Gerber, H. P. *et al.* Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263,** 808–811 (1994).

86. Hope, I. A., Mahadevan, S. & Struhl, K. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* **333,** 635–640 (1988).

87. Tanaka, M. & Herr, W. Reconstitution of transcriptional activation domains by reiteration of short peptide segments reveals the modular organization of a glutamine-rich activation domain. *Molecular and Cellular Biology* **14,** 6056–6067 (1994).

88. Lu, X., Ansari, A. Z. & Ptashne, M. An artificial transcriptional activating region with unusual properties. *Proceedings of the National Academy of Sciences* **97,** 1988–1992 (2000).

89. Staller, M. V. *et al.* A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Systems* **6,** 444–455.e6 (2018).

90. Staller, M. V., Ramirez, E., Holehouse, A. S., Pappu, R. V. & Cohen, B. A. Human transcriptional activation domains require hydrophobic and acidic residues. *bioRxiv,* 2020.10.28.359026 (2020).

91. Ravarani, C. N. *et al.* High-throughput discovery of functional disordered regions: investigation of transactivation domains. *Molecular systems biology* **14,** e8190 (2018).

92. Drögemüller, J. *et al.* Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Scientific Reports* **5,** 16428 (2015).

93. Hall, D. B. & Struhl, K. The VP16 Activation Domain Interacts with Multiple Transcriptional Components as Determined by Protein-Protein Cross-linking in Vivo. *Journal of Biological Chemistry* **277,** 46043–46050 (2002).

94. Ferreira, M. E. *et al.* Mechanism of Transcription Factor Recruitment by Acidic Activators. *Journal of Biological Chemistry* **280,** 21779–21784 (2005).

95. Gao, M., Yang, J., Liu, S., Su, Z. & Huang, Y. Intrinsically Disordered Transactivation Domains Bind to TAZ1 Domain of CBP via Diverse Mechanisms. *Biophysical Journal* **117,** 1301–1310 (2019).

96. Blau, J. *et al.* Three functional classes of transcriptional activation domain. *Molecular and Cellular Biology* **16,** 2044–2055 (1996).

97. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular Cell* **79,** 459–471.e4 (2020).

98. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal* **114,** 2262–2270 (2018).

99. Kent, S. *et al.* Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. *Cell Reports* **33,** 108248 (2020).

100. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).

101. Wei, M.-T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nature Cell Biology* **22,** 1187–1196 (2020).

102. Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nature Communications* **9,** 2985 (2018).

103. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361,** eaar2555 (2018).

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

105. Quintero-Cadena, P., Lenstra, T. L. & Sternberg, P. W. RNA Pol II Length and Disorder Enable Cooperative Scaling of Transcriptional Bursting. *Molecular Cell* (2020).

106. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572,** 543–548 (2019).

107. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & Development* **33,** 1619–1634 (2019).

108. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176,** 419–434 (2019).

109. Morin, J. A. *et al.* Surface condensation of a pioneer transcription factor on DNA. *bioRxiv,* 2020.09.24.311712 (2020).

110. Lindén, M. *et al.* FET family fusion oncoproteins target the SWI SNF chromatin remodeling complex. *EMBO reports* **20,** e45766 (2019).

111. Riggi, N. *et al.* Expression of the FUS-CHOP Fusion Protein in Primary Mesenchymal Progenitor Cells Gives Rise to a Model of Myxoid Liposarcoma. *Cancer Research* **66,** 7016–7023 (2006).

112. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).

113. Ma, J. & Ptashne, M. A new class of yeast transcriptional activators. *Cell* **51,** 113–119 (1987).

114. Gersbach, C. A., Gaj, T. & Barbas, C. F. Synthetic Zinc Finger Proteins: The Advent of Targeted Gene Regulation and Genome Modification Technologies. *Accounts of Chemical Research* **47,** 2309–2318 (2014).

115. Wu, H., Yang, W. P. & Barbas, C. F. Building zinc fingers by selection: toward a therapeutic application. *Proceedings of the National Academy of Sciences* **92,** 344–348 (1995).

116. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences* **95,** 14628–14633 (1998).

117. Zhang, M. *et al.* TALE: A tale of genome editing. *Progress in Biophysics and Molecular Biology* **114,** 25–32 (2014).

118. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326,** 1509–1512 (2009).

119. Jansen, R., Embden, J. D. A. v., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43,** 1565–1575 (2002).

120. Jinek, M. *et al.* A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).

121. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

122. Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K. & Severinov, K. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA. *Nucleic Acids Research* **44,** 2837–2845 (2016).

123. Anderson, E. M. *et al.* Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. *Journal of Biotechnology* **211,** 56–65 (2015).

124. Josephs, E. A. *et al.* Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Research* **43,** 8924–8941 (2015).

125. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

126. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature Biotechnology* **34,** 528–530 (2016).

127. Chavez, A. *et al.* Highly-efficient Cas9-mediated transcriptional programming. *Nature Methods* **12,** 326–328 (2015).

128. Mollinari, C. *et al.* Transdifferentiation: a new promise for neurodegenerative diseases. *Cell Death & Disease* **9,** 1–9 (2018).

129. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51,** 987–1000 (1987).

130. Pang, Z. P. *et al.* Induction of human neuronal cells by defined transcription factors. *Nature* **476,** 220–223 (2011).

131. Kramer, M., Fischer, A., Groth, J., Weber, W. & Radziwill, G. Optogenetic Switches. *OptoBase, https://www.optobase.org/switches/* (2020).

132. Briggs, W. R. The LOV domain: a chromophore module servicing multiple photoreceptors. *Journal of Biomedical Science* **14,** 499–504 (2007).

133. Zoltowski, B. D., Motta-Mena, L. B. & Gardner, K. H. Blue Light-Induced Dimerization of a Bacterial LOV–HTH DNA-Binding Protein. *Biochemistry* **52,** 6653–6661 (2013).

134. Bubeck, F. *et al.* Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9. *Nature Methods* **15,** 924–927 (2018).

135. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nature Communications* **5,** 4404 (2014).

136. Liu, H. *et al.* Photoexcited CRY2 Interacts with CIB1 to Regulate Transcription and Floral Initiation in Arabidopsis. *Science* **322,** 1535–1539 (2008).

137. Ma, L. *et al.* Structural insights into the photoactivation of Arabidopsis CRY2. *Nature Plants* **6,** 1432–1438 (2020).

138. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods* **7,** 973–975 (2010).

139. Che, D. L., Duan, L., Zhang, K. & Cui, B. The Dual Characteristics of Light-Induced Cryptochrome 2, Homo-oligomerization and Heterodimerization, for Optogenetic Manipulation in Mammalian Cells. *ACS Synthetic Biology* **4,** 1124–1135 (2015).

140. Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168,** 159–171 (2017).

141. Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175,** 1481–1491 (2018).

142. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. *Molecular Cell* **71,** 468–480 (2018).

143. Stavreva, D. A. *et al.* Transcriptional Bursting and Co-bursting Regulation by Steroid Hormone Release Pattern and Transcription Factor Mobility. *Molecular Cell* **6,** 1161–1177 (2019).

144. Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* **116,** 683–698 (2004).

145. Rafalska-Metcalf, I. U., Powers, S. L., Joo, L. M., LeRoy, G. & Janicki, S. M. Single Cell Analysis of Transcriptional Activation Dynamics. *PLOS ONE* **5,** e10272 (2010).

146. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401,** 921–923 (1999).

147. Braastad, C. D., Hovhannisyan, H., van Wijnen, A. J., Stein, J. L. & Stein, G. S. Functional characterization of a human histone gene cluster duplication. *Gene* **342,** 35–40 (2004).

148. Kim, T.-K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162,** 948–959 (2015).

149. Dao, L. T. M. *et al.* Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics* **49,** 1073–1081 (2017).

150. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical Journal* **16,** 1055–69 (1976).

151. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching. *Biophysical Journal* **86,** 3473–3495 (2004).

152. Sprague, B. L. *et al.* Analysis of Binding at a Single Spatially Localized Cluster of Binding Sites by Fluorescence Recovery after Photobleaching. *Biophysical Journal* **91,** 1169–1191 (2006).

153. Mueller, F., Wach, P. & McNally, J. G. Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. *Biophysical Journal* **94,** 3323–3339 (2008).

154. Mueller, F., Morisaki, T., Mazza, D. & McNally, J. G. Minimizing the Impact of Photoswitching of Fluorescent Proteins on FRAP Analysis. *Biophysical Journal* **102,** 1656–1665 (2012).

155. Torres-Padilla, M.-E. & Chambers, I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* **141,** 2173–2181 (2014).

156. Tanay, A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Research* **16,** 962–972 (2006).

157. Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annual Review of Cell and Developmental Biology* **35,** 357–379 (2019).

158. Hager, G. L., McNally, J. G. & Misteli, T. Transcription Dynamics. *Molecular Cell* **35,** 741–753 (2009).

159. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology* **11,** 823 (2015).

160. Callegari, A. *et al.* Single-molecule dynamics and genome-wide transcriptomics reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional activation. *PLoS Genetics* **15,** e1007891 (2019).

161. Chapela, P. J., Broaddus, R. R., Hawkins, S. M., Lessey, B. A. & Carson, D. D. Cytokine Stimulation of MUC4 Expression in Human Female Reproductive Tissue Carcinoma Cell Lines and Endometrial Cancer. *Journal of Cellular Biochemistry* **116,** 2649–2657 (2015).

162. Wu, B., Chao, J. A. & Singer, R. H. Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophysical Journal* **102,** 2936–2944 (2012).

163. Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O. & Hillen, W. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Research* **32,** 842–847 (2004).

164. Yamada, M., Suzuki, Y., Nagasaki, S. C., Okuno, H. & Imayoshi, I. Light Control of the Tet Gene Expression System in Mammalian Cells. *Cell Reports* **25,** 487–500.e6 (2018).

165. Trojanowski, J., Rademacher, A., Erdel, F. & Rippe, K. *Light-Induced Transcription Activation for Time-Lapse Microscopy Experiments in Living Cells* in *Imaging Gene Expression: Methods and Protocols* (ed Shav-Tal, Y.) 251–270 (Springer New York, New York, NY, 2019).

166. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2017).

167. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26,** 979–981 (2010).

168. Erdel, F. *et al.* Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Molecular Cell* **78,** 236–249.e7 (2020).

169. Frank, L., Weinmann, R., Erdel, F., Trojanowski, J. & Rippe, K. *Transcriptional activation of heterochromatin by recruitment of dCas9 activators.* in *Enhancers and Promoters: Methods and Protocols* (eds Borggrefe, T. & Daniele Giaimo, B.) in press (2021).

170. Gustafsson, N. *et al.* Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications* **7,** 12471 (2016).

171. Harding, R. M., Boyce, A. J. & Clegg, J. B. The Evolution of Tandemly Repetitive DNA: Recombination Rules. *Genetics* **132,** 847–859 (1992).

172. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

173. Rademacher, A., Erdel, F., Trojanowski, J. & Rippe, K. *Tracing Reversible Light-Induced Chromatin Binding with Near-infrared Fluorescent Proteins* in *Photo-switching Proteins : Methods and Protocols* (ed Niopek, D.) 171–188 (Springer US, New York, NY, 2020).

174. Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science* **336,** 1440–1444 (2012).

175. Friedrich, D. *et al.* Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency. *Molecular Systems Biology* **15,** e9068 (2019).

176. Zhang, Q. *et al.* NF-kB Dynamics Discriminate between TNF Doses in Single Cells. *Cell Systems* **5,** 638–645 (2017).

177. Cirillo, L. A. *et al.* Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell* **9,** 279–289 (2002).

178. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72,** 73–84 (1993).

179. Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332,** 472–474 (2011).

180. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* **99,** 12795–12800 (2002).

181. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22,** 403–434 (1976).

182. Markaki, Y. *et al.* Functional nuclear organization of transcription and DNA replication: a topographical marriage between chromatin domains and the interchromatin compartment. *Cold Spring Harb Symp Quant Biol* **75,** 475–92 (2010).

183. Katan-Khaykovich, Y. & Struhl, K. Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes & Development* **16,** 743–52 (2002).

184. Kügler, P. Moment Fitting for Parameter Inference in Repeatedly and Partially Observed Stochastic Biological Models. *PLOS ONE* **7,** e43001 (2012).

185. Kennedy, J. & Eberhart, R. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks* **4,** 1942–1948 vol.4 (1995).

186. Clerc, M. Standard Particle Swarm Optimisation. *https://hal.archives-ouvertes.fr/hal-00764996* (2012).

187. Hipp, L. *et al.* Single-molecule imaging of the transcription factor SRF reveals prolonged chromatin-binding kinetics upon cell stimulation. *Proceedings of the National Academy of Sciences* **116,** 880–889 (2019).

188. Leica. FRAP Device for Widefield Microscopy. *https://www.leica-microsystems.com/company/news/news-details/article/frap-device-for-widefield-microscopy/* (2004).

189. Kitamura, A. & Kinjo, M. Determination of diffusion coefficients in live cells using fluorescence recovery after photobleaching with wide-field fluorescence microscopy. *Biophysics and Physicobiology* **15,** 1–7 (2018).

190. Fukano, T., Hama, H. & Miyawaki, A. Similar diffusibility of membrane proteins across the axon-soma and dendrite-soma boundaries revealed by a novel FRAP technique. *Journal of Structural Biology* **147,** 12–18 (2004).

191. Soetaert, K. & Meysman, F. Reactive transport in aquatic ecosystems: Rapid model prototyping in the open source software R (vol 32, pg 49, 2012). *Environmental Modelling & Software* **35,** 194–194 (2012).

192. Sprague, B. L. & McNally, J. G. FRAP analysis of binding: proper and fitting. *Trends in Cell Biology* **15,** 84–91 (2005).

193. Trojanowski, J., Frank, L., Rademacher, A., Grigaitis, P. & Rippe, K. Transcription activation is enhanced by multivalent interactions independent of liquid-liquid phase separation. *bioRxiv,* 2021.01.27.428421 (2021).

194. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* **11,** 198–200 (2015).

195. Elf, J., Li, G.-W. & Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **316,** 1191–1194 (2007).

196. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences* **71,** 4135–9 (1974).

197. Wilson, B. J., Tremblay, A. M., Deblois, G., Sylvain-Drolet, G. & Giguere, V. An acetylation switch modulates the transcriptional activity of estrogen-related receptor alpha. *Mol Endocrinol* **24,** 1349–58 (2010).

198. Evans, P. M. *et al.* Kruppel-like factor 4 is acetylated by p300 and regulates gene transcription via modulation of histone acetylation. *Journal of Biological Chemistry* **282,** 33994–4002 (2007).

199. Salghetti, S. E., Caudy, A. A., Chenoweth, J. G. & Tansey, W. P. Regulation of Transcriptional Activation Domain Function by Ubiquitin. *Science* **293,** 1651–1653 (2001).

200. Kurosu, T. & Peterlin, B. M. VP16 and Ubiquitin: Binding of P-TEFb via Its Activation Domain and Ubiquitin Facilitates Elongation of Transcription of Target Genes. *Current Biology* **14,** 1112–1116 (2004).

201. Fu, B. X., St. Onge, R. P., Fire, A. Z. & Smith, J. D. Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Research* **44,** 5365–5377 (2016).

202. Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Scientific Reports* **7,** 40638 (2017).

203. Melcher, K. The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins by K. Yamamoto. *Journal of Molecular Biology* **301,** 1097–1112 (2000).

204. Shrimp, J. H. *et al.* Chemical Control of a CRISPR-Cas9 Acetyltransferase. *ACS Chemical Biology* **13,** 455–460 (2018).

205. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).

206. Gurdon, J. B., Javed, K., Vodnala, M. & Garrett, N. Long-term association of a transcription factor with its chromatin binding site can stabilize gene expression and cell fate commitment. *Proceedings of the National Academy of Sciences* **117,** 15075–15084 (2020).

207. Waterborg, J. H. Dynamics of histone acetylation in Saccharomyces cerevisiae. *Biochemistry* **40,** 2599–605 (2001).

208. Chen, L.-F. *et al.* Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell Reports* **26,** 1174–1188 (2019).

209. Wright, A. P. H., McEwan, I. J., Dahlman-Wright, K. & Gustafsson, J.-Å. High Level Expression of the Major Transactivation Domain of the Human Glucocorticoid Receptor in Yeast Cells Inhibits Endogenous Gene Expression and Cell Growth. *Molecular Endocrinology* **5,** 1366–1372 (1991).

210. Tsang, B. *et al.* Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proceedings of the National Academy of Sciences* **116,** 4218–4227 (2019).

211. Uppal, S. *et al.* The Bromodomain Protein 4 Contributes to the Regulation of Alternative Splicing. *Cell Reports* **29,** 2450–2460 (2019).

212. Tlusty, T., Bar-Ziv, R. & Libchaber, A. High-Fidelity DNA Sensing by Protein Binding Fluctuations. *Physical Review Letters* **93,** 258103 (2004).

213. Hopfield, J. J., Yamane, T., Yue, V. & Coutts, S. M. Direct experimental evidence for kinetic proofreading in amino acylation of tRNAIle. *Proceedings of the National Academy of Sciences* **73,** 1164–1168 (1976).

214. Yousefi, O. S. *et al.* Optogenetic control shows that kinetic proofreading regulates the activity of the T cell receptor. *eLife* **8,** e42475 (2019).

215. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. *Proceedings of the National Academy of Sciences* **80,** 1707–1711 (1983).

216. Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer* **19,** 611–624 (2019).

217. Weltner, J. *et al.* Human pluripotent reprogramming with CRISPR activators. *Nature Communications* **9** (2018).

218. Polstein, L. R. *et al.* Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Research* **25,** 1158–1169 (2015).

219. Frahm, T., Hauser, H. & Köster, M. IFN-type-I-mediated signaling is regulated by modulation of STAT2 nuclear export. *Journal of Cell Science* **119,** 1092–104 (2006).

220. Jegou, T. *et al.* Dynamics of telomeres and promyelocytic leukemia nuclear bodies in a telomerase-negative human cell line. *Molecular Biology of the Cell* **20,** 2070–82 (2009).

221. Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175,** 1467–1480.e13 (2018).

222. Rothbauer, U. *et al.* A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Molecular and Cellular Proteomics* **7,** 282–9 (2008).

223. Team, R. RStudio: Integrated Development for R (2016).

224. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9,** 676–82 (2012).

225. Wang, F. *et al.* RNAscope. *The Journal of Molecular Diagnostics* **14,** 22–29 (2012).

Supplemental Tables

Table 6: FRAP parameters estimated from recovery curves. Averages of parameter values estimated from single-cell recovery curves are shown. Values in brackets correspond to the 95% CI.

[a]: not determined. For simulations of the reaction-diffusion model the value of dCas9-EGFP-VPR with the wt-*tetO* sgRNA was used.

Complex	DNA	Residence time (s)	f_i (%)	D_{eff} ($\mu m^2/s$)	n
EGFP-LacI	*lacO*	108 (91-134)	8 (5-11)	2.3 (1.5-3.0)	13
EGFP-LacI (confocal)	*lacO*	97 (69-167)	29 (14-45)	3.3 (2.0-4.5)	7
dCas9-EGFP	*lacO*	74 (32->240)	44 (34-54)	1.8 (0-4.3)	11
dCas9-EGFP-VP16	*lacO*	>240 (>240)	37 (15-59)	1.4 (0.8-2.0)	6
dCas9-EGFP-VPR	*lacO*	204 (112->240)	19 (4-34)	1.8 (0.7-2.9)	10
dCas9-EGFP-VPR	*tetO*	124 (75->240)	36 (25-47)	0.6 (0.4-0.9)	10
dCas9-EGFP-VPR	*tetO*-C2G	57 (34-184)	7 (0-16)	- [a]	7
rTetR-EGFP-VP16	*tetO*	132 (74->240)	5 (0-14)	4.3 (3.0-5.4)	6
rTetR-EGFP-VPR	*tetO*	57 (42-86)	7 (2-12)	3.1 (1.6-4.5)	14
dCas9 + tdPCP-EGFP	*lacO*	12 (9-18)	3 (0-5)	3.4 (0.9-6.2)	12
dCas9 + tdPCP-EGFP-VP16	*lacO*	33 (17->240)	4 (0-10)	2.9 (0.6-5.2)	7
dCas9 + tdPCP-EGFP-VPR	*lacO*	47 (33-83)	7 (1-13)	2.2 (1.7-2.8)	11
CIBN-dCas9-CIBN + PHR-EGFP-VP16	*lacO*	28 (18-58)	10 (1-19)	2.0 (1.2-2.9)	11

CIBN-dCas9-CIBN + PHR-EGFP-VPR	*lacO*	49 (37-72)	14 (8-21)	1.3 (0.6-1.9)	14
tdPCP-CIBN + dCas9 PHR-EGFP-VP16	*lacO*	29 (20-53)	2 (0-4)	2.1 (1.4-2.8)	12
tdPCP-CIBN + dCas9 PHR-EGFP-VPR	*lacO*	60 (45-91)	10 (3-16)	1.7 (0.4-3.0)	12
CIBN-rTetR + PHR-EGFP-VP16	*tetO*	42 (33-58)	4 (2-7)	2.3 (1.7-2.8)	18
CIBN-rTetR + PHR-EGFP-VPR	*tetO*	71 (60-88)	6 (2-10)	1.6 (1.1-2.1)	16

Table 7: Binding site occupancy for loop complexes. The occupancy corresponds to the relative EGFP-AD intensity normalized to the tagBFP-LacI marker. The percentage of visible arrays corresponds to the cells with a marker spot, where there is also enrichment of EGFP-AD.

AD	Occupancy ($a.u.$)	n	Visible Array (%)
tdPCP-EGFP-VP16	0.53 (0.41-0.66)	166	78
tdPCP-EGFP-VPR	0.99 (0.77-1.22)	164	93

Table 8: Binding site occupancy of dCas9-EGFP-AD for mutated sgRNA. The occupancy corresponds to the relative EGFP-AD intensity normalized to the tagBFP-LacI marker. The percentage of visible arrays corresponds to the cells with a marker spot, where there is also enrichment of EGFP-AD.

AD	sgRNA	Occupancy ($a.u.$)	n	Visible Array (%)
VP16	tetO-2xPP7 (wt)	0.24 (0.17-0.30)	138	59
VP16	tetO-2xPP7-C2G (mut)	0.03 (0.01-0.06)	127	17
VPR	tetO-2xPP7 (wt)	0.32 (0.26-0.37)	175	90
VPR	tetO-2xPP7-C2G (mut)	0.12 (0.09-0.14)	163	76

Table 9: Activation kinetics for different ADs. Average values of half-activation time ($t_{1/2}$) and nascent RNA plateau values of single-cell time courses. Responder fractions correspond to cells with PHR-EGFP-AD recruitment that show visible tdMCP-tdTomato enrichment at the spot. Values in brackets correspond to the 95% CI.

PHR-EGFP-AD	Cell subgroup	n	Responders (%)	$t_{1/2}$ (min)	RNA max. ($a.u.$)
VP16	All cells	64	67	42 (37-46)	1.2 (0.89-1.6)
VPR	All cells	37	84	28 (23-33)	1.7 (1.1-2.4)
VPR	Cells without droplets	15	87	25 (17-34)	1.1 (0.83-1.4)
VPR	Cell with droplets	22	82	30 (24-36)	2.2 (1.0-3.3)
p65	All cells	52	67	26 (21-31)	2.1 (1.6-2.6)
p65	Cells without droplets	23	78	26 (19-33)	2.3 (1.5-3.0)
p65	Cell with droplets	29	59	26 (18-34)	1.9 (1.2-2.7)
Rta	All cells	77	92	28 (25-31)	1.2 (0.92-1.5)
Rta	Cells without droplets	33	94	25 (21-29)	1.2 (0.73-1.7)
Rta	Cell with droplets	44	91	31 (26-36)	1.2 (0.85-1.6)
STAT2	All cells	132	42	38 (34-43)	0.95 (0.66-1.2)

Table 10: Activation kinetics for enforced AD droplets. CIBN-rTetR was used as a localizer, if not specified otherwise. Average values defined as in **Table 9**.

[a]: not determined due to low number of responder cells.

PHR-EGFP-AD	Condition	n	Responders (%)	$t_{1/2}$ (min)	RNA max. $(a.u.)$
VP16	No additional factors	74	70	35 (30-39)	0.61 (0.46-0.76)
VP16	GFP-LacI	97	42	34 (28-39)	0.39 (0.22-0.56)
VP16	CIBN-LacI	118	24	31 (23-38)	0.22 (0.16-0.28)
VP16	CIBN-LacI but no CIBN-rTetR	126	18	20 (13-28)	0.20 (0.09-0.29)
VPR	No additional factors	18	78	31 (22-40)	2.2 (0.56-3.7)
VPR	GFP-LacI	128	43	33 (28-37)	0.61 (0.40-0.82)
VPR	CIBN-LacI	93	32	31 (23-38)	0.34 (0.22-0.46)
VPR	CIBN-LacI but no CIBN-rTetR	86	16	26 (15-38)	0.68 (0.00-1.6)
VP16	No additional factors	154	84	34 (31-36)	0.78 (0.63-0.93)
VP16	PHR-GBP	24	4	n. d.[a]	n. d.[a]
FUSN-VP16	No additional factors	108	87	37 (34-41)	1.5 (1.1-1.8)
FUSN	No additional factors	57	5	n. d.[a]	n. d.[a]

Table 11: BRD4 recruitment kinetics. Average values of half-activation time ($t_{1/2}$) and mCherry-BRD4 plateau values of single-cell time courses (including non-responder cells) induced by PHR-EGFP-AD recruitment to the optoloop complex at *tetO* and *lacO* sites. Responder fractions correspond to cells with PHR-EGFP-AD recruitment that show visible mCherry-BRD4 enrichment at the spot. Values in brackets correspond to the 95% CI.

PHR-EGFP-AD	Condition	n	Responders (%)	$t_{1/2}$ (min)	BRD4 max. $(a.u.)$
VP16	–	37	27	13 (7-20)	0.007 (0.005-0.010)
VP16	JQ1	85	0	–	0.002 (0.002-0.003)
VPR	–	13	92	13 (8-18)	0.024 (0.016-0.032)
VPR	JQ1	10	10	–	0.005 (0.002-0.008)

Table 12: Activation kinetics after chromatin perturbations. Average activation parameters as in **Table 9** after priming promoter chromatin by JQ1 treatment or recruitment of dCas9-effector to the *lacO* sites.

PHR-EGFP-AD	Condition	n	Responders (%)	$t_{1/2}$ (min)	RNA max. $(a.u.)$
VP16	–	29	69	35 (30-39)	1.7 (0.91-2.5)
VP16	JQ1	12	58	28 (18-39)	1.4 (0.63-2.2)
VPR	–	15	87	29 (20-39)	2.1 (1.0-3.2)
VPR	JQ1	21	81	33 (27-38)	3.0 (2.0-4.0)
VP16	dCas9-EGFP	52	25	35 (25-44)	0.35 (0.18-0.51)
VP16	dCas9-EGFP-p300	49	53	31 (27-36)	1.1 (0.76-1.4)
VPR	dCas9-EGFP	27	33	37 (19-54)	0.51 (0.23-0.79)
VPR	dCas9-EGFP-p300	72	75	31 (27-34)	1.1 (0.82-1.4)

Danksagung

Ich möchte mich ganz herzlich bei Karsten Rippe bedanken. Dafür, im richtigen Moment das Projekt in die richtige Richtung gelenkt zu haben, und für Diskussionen über wissenschaftliches Arbeiten. Sowohl über die Inhalte, als auch über das ganze notwendige Drumherum. Außerdem danke ich den Mitgliedern meines Thesis Advisory Committee, Thomas Höfer und Michael Brunner, für kritisches und produktives Feedback und Ursula Kummer, die sich bereit erklärt hat ebenfalls Teil meiner Prüfungskommission zu sein.

Ich danke dem ganzen Rippe Lab für die freundschaftliche Arbeitsatmosphäre, genauso wie für Unternehmungen außerhalb des Labors. Mit Euch als Kollegen und Freunde bin ich immer gerne ins BioQuant gekommen. Besonders möchte ich mich bei Lukas bedanken, mit dem die Zusammenarbeit im “OptoFRAP” Projekt nicht nur produktiv war, sondern auch viel Spaß gemacht hat. Danke an Markus, Lara und Robin für kritisches Feedback, an Sabrina und Caro für praktische Hilfe im Labor und an alle für entspannte und unterhaltsame Kaffeepausen, Ausflüge, Curry-Donnerstage, Retreats, Cocktailabende, Partys, Spieleabende und Weihnachtsfeiern.

Vielen Dank an Fabian Erdel, Congxin Li und Manuel Gunkel für hilfreiche Diskussionen und praktische Hilfe. Ebenso vielen Dank an die DKFZ Light Microscopy Facility für die Unterstützung von Manuela Brom, Damir Krunic und Felix Bestvater.

Ich danke besonders meinen Eltern, die mich mit ihrem Interesse und ihrem Rat immer unterstützt haben. Und Philipp, mit dem ich über alles außer Biologie reden kann und der immer für mich da ist. Vielen Dank auch an meine Nichtlabor-Freunde, die mich regelmäßig in die Welt außerhalb des Labors zurückholen: Beccy, Ines, Kathi, Katja, Leo, Markus, Nik, Raphi, Robert, Sebastian, Sarina und Vale.