Transcriptional kinetic synergy: a complex landscape revealed by integrating modelling and synthetic biology

Rosa Martinez-Corral[1,*], Minhee Park[2,3,4,*], Kelly Biette[1,5,*], Dhana Friedrich[1], Clarrissa Scholes[1,6], Ahmad S. Khalil[2,3,7], Jeremy Gunawardena[1], Angela H. DePace[1]

1 Department of Systems Biology, Harvard Medical School,Boston, MA 02115, USA.
2 Biological Design Center, Boston University, Boston, MA 02215, USA.
3 Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.
4 Present address: Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA.
5 Present address: Novartis Institutes for Biomedical Research, Cambridge, MA 02139, USA.
6 Present address: Mammoth Biosciences, San Francisco, CA 94080, USA.
7 Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.
* Equal contribution
ORCID IDs:
RMC: 0000-0003-3600-3601, MP: 0000-0002-2101-5485; DF: 0000-0003-2144-9280;
CS: 0000-0001-9208-3234; ASK: 0000-0002-8214-0546; JG:0000-0002-7280-1152;
AHD: 0000-0001-5723-0438

1 Abstract

Eukaryotic genes are combinatorially regulated by a diversity of factors, including specific DNA-binding proteins called transcription factors (TFs). Physical interactions between regulatory factors have long been known to mediate synergistic behaviour, commonly defined as deviation from additivity when TFs or sites act in combination. Beyond binding-based interactions, the possibility of synergy emerging from functional interactions between TFs was theoretically proposed, but its governing principles have remained largely unexplored. Theoretically, the interplay between the binding of TFs and their effects over transcription has been challenging to integrate. Experimentally, probing kinetic synergy is easily confounded by physical interactions. Here we circumvent both of these limitations by focusing on a scenario where only one TF can be specifically bound at any given time, which we build using a synthetic biology approach in a mammalian cell line. We develop and analyze a mathematical model that explicitly incorporates the details of the binding of the TFs and their effects over transcription. The model reveals that synergy depends not only on the biochemical activities of the TFs, but also on their binding kinetics. We find experimental evidence for this result in a reporter-based system where fusions of mammalian TFs with engineered zinc fingers bind to a single, shared site. A complex synergy landscape emerges where TF activity, concentration and binding affinity shape the expression response. Our results highlight the relevance of an integrated understanding of TF function in eukaryotic transcriptional control.

2 Introduction

Transcription regulation is a fundamental process in all domains of life. In order to increase or decrease gene expression, cells rely on the activity of transcription factors (TFs), proteins that bind the DNA and modulate the binding and activity of the RNA polymerase complex. In eukaryotes, a hallmark of transcription is the usage of multiple TFs, of the same and different type, that bind to gene regulatory regions called enhancers and promoters (Field and Adelman, 2020).

Such combinatorial control enables binding and response specificity (Wunderlich and Mirny, 2009; Georges et al., 2010), and expands the regulatory capabilities of the finite set of TFs encoded by an organism. Despite a wealth of studies that have characterised TF binding sites and binding profiles in model genes, genome-wide and random sequences (Smith et al., 2013; Vandel et al., 2019; Inukai et al., 2017), the mechanism by which a combination of regulatory proteins determine the level of expression of a target gene is still poorly understood (Reiter et al., 2017; King et al., 2020; Nie et al., 2020). This is a key aspect for predicting gene expression from sequence data (Ouyang et al., 2009; de Boer et al., 2020) and challenges our ability to anticipate the effect of mutations and pharmacological interventions, a long-sought goal of biomedicine and synthetic biology.

When TFs interact to regulate expression, the combined response is not simply predicted by the responses from each of the TFs alone. Deviation from the null expectation (i.e. the response when TFs do not interact) is generally referred to as "synergy" and has gone by multiple definitions. Often, the null is taken to be the additive output from each of the individual TFs or sites, and synergy has been used to refer to deviation from this additive expectation (Carey et al., 1990; Herschlag and Johnson, 1993). As a result of TF interactions, a small change in input TF concentration can lead to large changes in expression. This may result into a nonlinear response to increasing TF concentrations, which has also been taken as evidence for synergy (Carey, 1998). In some cases, the term synergy has also been used interchangeably with binding cooperativity (below), or as a special form of it (Veitia, 2003; Michida et al., 2020). More broadly, synergy may also refer to effects beyond a single gene (Okawa and Del Sol, 2019). Here we use "synergy" in the narrow sense, to refer to the expression of a target gene when TFs act in combination, as compared to the expression from the individual TFs.

Synergy between TFs has generally been attributed to direct or indirect binding interactions (Vashee et al., 1998; Ambrosetti et al., 2000; Spitz and Furlong, 2012; Goldstein et al., 2017). Starting with the pioneering studies on the regulation of lambda and lac in bacteria (Johnson et al., 1979; Oehler et al., 1990), it is well understood that direct interactions between adjacently-bound TFs and/or polymerase can modulate the free energies of binding, therefore enhancing or reducing binding to a given site when neighboring sites are already bound. This phenomenon is typically known as binding cooperativity (Ackers et al., 1982; Estrada et al., 2016). Assuming that polymerase binding is directly related to expression, binding cooperativity can lead to nonlinear input-output responses and synergy (Frank et al., 2012). Similarly, indirect interactions through a shared molecule or complexes like Mediator can mediate such nonlinear responses (Carey et al., 1990; Malik and Roeder, 2010), a phenomenon that can be harnessed for engineering gene circuits with complex dynamical features (Bashor et al., 2019). Binding cooperativity can also emerge indirectly through allosteric mechanisms via nucleosomes (Mirny, 2010) or the DNA (Narasimhan et al., 2015).

The effect of protein-protein interactions on gene expression has been extensively modelled using the thermodynamic formalism (Ackers et al., 1982; Bintu et al., 2005b; Bashor et al., 2019). Under this view, the recruitment of RNA polymerase is assumed to be the rate-limiting step in transcription, determined by the binding pattern of the regulatory factors on the DNA. The regulatory machinery is assumed to be at steady state, and binding and unbinding is assumed to occur at thermodynamic equilibrium. At a steady state of thermodynamic equilibrium, the binding configuration of the system is determined by the free energies of the states following the Boltzman distribution, by the principles of statistical mechanics. Under the thermodynamic view, if two activators independently interact with RNA polymerase, then the predicted fold-change activation is the product of the individual fold-changes. Hence, multiplicative activation has also been taken as equivalent to synergistic activation in the literature (Bintu et al., 2005a).

Despite the success of the thermodynamic framework in many scenarios (Segal et al., 2008; Phillips, 2015; González et al., 2019), its assumptions contrast with several experimental observations of eukaryotic gene regulation (Wong and Gunawardena, 2020). The first challenge lies in the assumption that RNA polymerase is the only rate-limiting step in transcription. In contrast, it is well known that there are multiple steps involved in the regulation of eukaryotic transcription, which has led to its conceptualization in terms of a multi-step cycle (Fuda et al., 2009). Eukaryotic DNA is often wrapped around nucleosomes. Displacing nucleosomes or post-translationally modifying

histones may be a pre-requisite for RNA polymerase binding (Mao et al., 2010; Hansen and O'Shea, 2013; Cui et al., 2020). Moreover, RNA polymerase may pause shortly after starting transcription, and regulation of pause-release and the elongation rate may be another point of control (Jonkers and Lis, 2015; Core and Adelman, 2019). In addition, RNA polymerase has been found to be already bound on many inactive genes, suggesting that under certain scenarios activation does not rely on regulating polymerase binding, but modulating a subsequent step (Oven et al., 2007). Furthermore, the splicing machinery can also affect transcription (Tellier et al., 2020).

Similarly, the assumption of equilibrium binding is challenged by several observations. Given the relationship between free energies and binding affinities (ratios between binding on-rates and off-rates), at a steady state of thermodynamic equilibrium only the affinities, but not the individual rates, matter. Importantly, this is no longer true either at a steady state away from equilibrium, or away from steady state (even if the system is relaxing towards an equilibrium steady state) in which case kinetics become relevant (reviewed in Wong and Gunawardena (2020)). Mehta et al. (2018) reported data consistent with both the binding on and off rates of a TF being subject to regulation, suggesting that their actual values might matter, and not only their ratio. Likewise, our group has suggested that experimental data on the *Drosophila* model gene *Hunchback* cannot be explained by an equilibrium model (Park et al., 2019a). Along the same lines, low affinity binding sites have sometimes been observed to drive higher expression than higher affinity sites (Ramos and Barolo, 2013), again contrary to the occupancy-based view of the thermodynamic formalism where more activator binding is linked to more expression. Moreover, live-cell imaging of transcription has revealed widespread "bursting" in gene expression (Tunnacliffe and Chubb, 2020), a phenomenon that can only be understood kinetically.

As a result, kinetic models of transcription are also abundant in the literature, where a series of rate-limiting steps are involved in the regulation of gene expression (e.g Suter et al., 2011; Hansen and O'Shea, 2013; Rybakova et al., 2015). Under a kinetic view, the possibility of "kinetic synergy" was theoretically proposed: if there are multiple rate-limiting steps in transcription, and two TFs modulate different steps, then their combined effect may result in synergy, in the sense of "greater-than-additivity" (Herschlag and Johnson, 1993; Scholes et al., 2017). This mechanism of synergy would enable combinatorial control even in the absence of physical interactions between TFs, wherein TFs would not even need to be simultaneously present at the regulatory site.

Multiple lines of evidence make kinetic synergy very plausible. First, transcriptional activators can increase transcription by different mechanisms. Blau et al. (1996) found that TF activation domains can either stimulate transcription initiation, elongation, or both, and more recent studies have continued to reveal diversity among TFs (e.g. Fu et al., 2004; Rahl et al., 2010; Baluapuri et al., 2019). Along the same lines, Danko et al. (2013) reported differences in RNA polymerase II pausing depending upon treatment with E2 or TNF-alpha signals, which were attributed to the TFs that transduce those signals (ERα and NF-κB) acting at different rate-limiting steps upon their target genes. Moreover, a diversity of TFs are often associated to higher levels of expression as compared to regulation by a single TF type (Smith et al., 2013; Vanhille et al., 2015; Singh et al., 2020). In line with this, Keung et al. (2014) found evidence of synergistic activation between the viral activator VP16 and some chromatin regulators in a reporter system. Similarly, the activity of many *Drosophila* TFs and cofactors was found to be highly context-dependent (Stampfel et al., 2015), suggesting that activation may require a particular combination of biochemical functions.

However, it is unclear to what extent kinetic synergy arises in experimental systems and what functional capabilities it provides. A recent theoretical study by our group showed that similar to binding cooperativity, kinetic synergy can implement logic and analog computations (Scholes et al., 2017), and that it can generate a wide diversity of behaviors. However, that study did not explicitly account for TF binding, which as highlighted above can be important for the effect of TFs. To our knowledge, there have been little attempts to model the interplay between TF binding and the transcription cycle. Li et al. (2018) proposed a model that explicitly incorporated binding and transitions over the cycle, but assumed a time-scale separation between TF binding and the rest of the processes, with quasi-equilibrium in TF binding. However, both TF residence times and the half-life of certain biochemical steps in the transcription cycle may occur on similar timescales, on the order of several seconds or a few minutes (Methods, section 5.2). On the experimental

side, a big challenge for assessing kinetic synergy is the ability to disentangle the kinetic from the binding interactions between regulators. Here we address both of these points.

We focus on the case of TFs binding to a single, shared site. This setup eliminates the possibility of binding interactions, and is amenable to modelling and experimental interrogation. We propose a model for this scenario that explicitly accounts for TF binding and TF effects over the polymerase cycle which unifies the two main modelling strategies in transcription outlined above. Experimentally, we build this system using a synthetic biology approach. We engineer synthetic TFs where activation domains of a set of functionally diverse mammalian TFs are fused to a computationally designed zinc-finger (ZF) DNA binding domain predicted to bind only to an artificial site upstream of a reporter (Fig. 1A) (Khalil et al., 2012; Keung et al., 2014; Park et al., 2019b). We propose a comprehensive measure of synergy where we compare the expression output when both TFs are present, to that when only one of them is present. By exploring the synergistic behaviour of the model in parameter space, we find that a diversity of behaviors can emerge in this setup, for which we find experimental evidence. A qualitative comparison between modelling and experiments reveals a complex synergy landscape, shaped by the interplay between the activation effect of the TFs with respect to the basal state, and their binding, which we can experimentally tune. This highlights the relevance of considering genomic context, binding and biochemical function together when characterizing TFs, and illuminates how functional interactions between TFs may contribute to eukaryotic transcriptional control.

3 Results

3.1 Mathematical model

We study how kinetic synergy emerges in a scenario where two TFs bind to a shared site in a regulatory sequence, such that only one TF can be specifically bound at any given time. Fig. 1A schematizes this situation for a general 3-state transcription cycle, where TF A promotes the first step (illustrated as the assembly of the RNA polymerase complex), and TF B promotes a process downstream.

In order to model this system, we exploit the linear framework formalism, a graph-based approach to Markov Processes that can be used to model a diversity of biological processes in a biophysically realistic and mathematically tractable way (Gunawardena, 2012; Ahsendorf et al., 2014). We have previously applied this framework to study how binding interactions modulate gene expression by implicitly averaging over the states of the polymerase cycle (Estrada et al., 2016; Biddle et al., 2019; Park et al., 2019a). In contrast, in a previous study of kinetic synergy, we modelled the effect of TFs on a detailed transcription cycle but effectively combined their binding with their enzymatic effects (Scholes et al., 2017). Here we propose a model that unifies both approaches and doesn't make assumptions about the binding reactions being on a different timescale than the polymerase cycle reactions, improving previous approaches in the literature (Li et al., 2018) (Methods, 5.1).

The system is represented by a graph (Fig. 1B, $\mathcal{N} \times \mathcal{P}$), whose vertices are the biological states of interest, and the edges are the transitions between them, assumed to follow Markovian dynamics with infinitesimal transition rates corresponding to the graph edge labels. Structurally (i.e. ignoring edge labels) the graph for the complete system is the graph product between two simpler graphs: a binding graph and a polymerase cycle graph. The binding graph for our system of interest is represented in Fig. 1B (Binding graph ($\mathcal{N}$)), and consists of a binding site that can either be empty, bound by TF A, or bound by TF B. For the polymerase cycle (Fig. 1B Polymerase cycle graph ($\mathcal{P}$)), we consider the simplest cycle, with 3 states (labelled 1,2,3). The first transition is assumed to be reversible, and the other two irreversible in agreement with the macroscopic irreversibility of posttranslational modifications like phosphorylation, or the synthesis of mRNA. mRNA is assumed to be produced when the system transitions from state 3 to state 1. This simple graph can be interpreted in terms of empty transcription start site (TSS), assembled RNA polymerase, and C-terminal phosphorylated or elongating polymerase, although mapping onto specific states isn't required to interpret the results. Given these two graphs, taking all pairwise combinations of their vertices (graph product) gives the complete graph $\mathcal{N} \times \mathcal{P}$ (Fig. 1B).

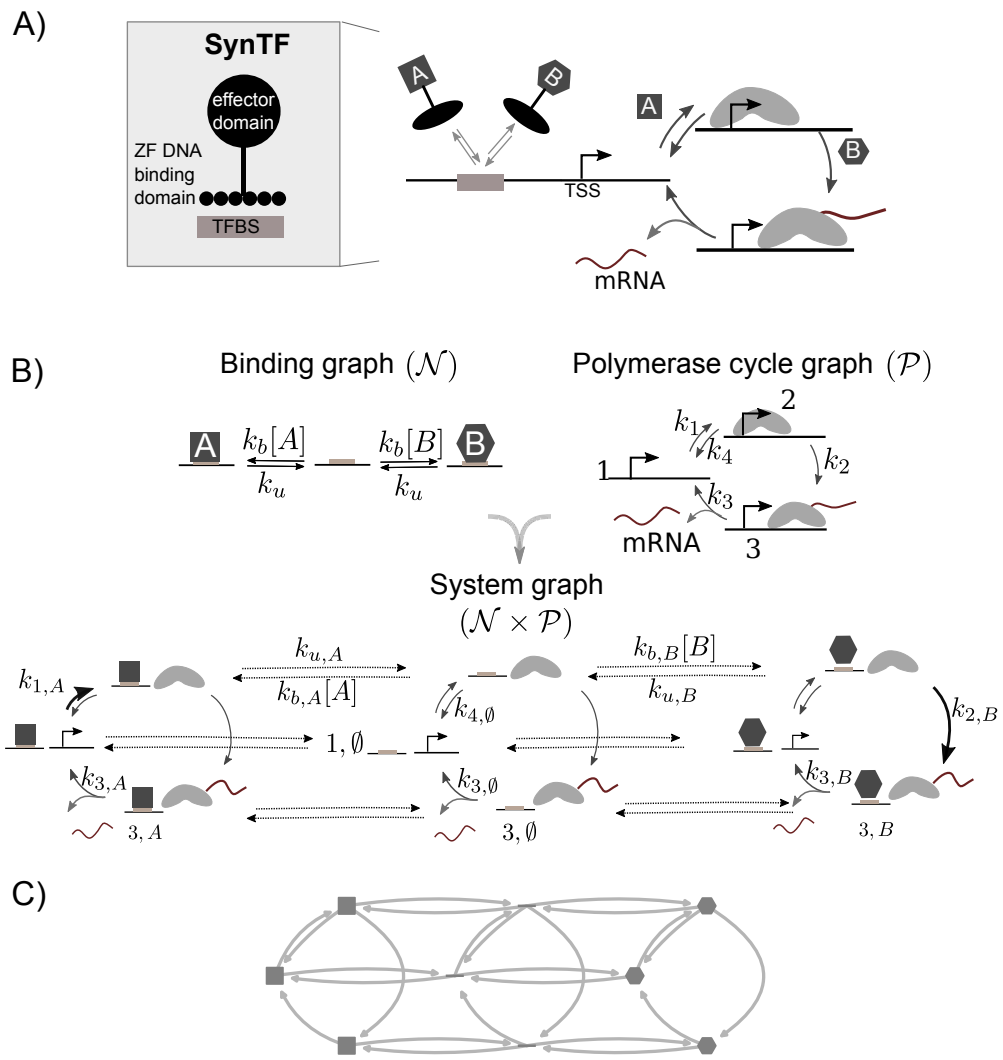

Figure 1: A model for kinetic synergy between two TFs sharing a site. A) Cartoon schematizing the strategy of this work to examine kinetic synergy: two synthetic TFs regulate a reporter (not shown) through a shared binding site. As an example, TF A controls the first step in the transcription cycle, and TF B controls the second step. B) Model used in this work. The graph product of a binding graph $\mathcal{N}$, and a 3-state polymerase cycle graph $\mathcal{P}$, gives rise to the complete linear framework graph of the system ($\mathcal{N} \times \mathcal{P}$). Only a subset of nodes and edges are labelled, for clarity. The horizontal edges from the central cycle to the outer cycles denote binding of each of the TFs, and the reverse edges denote unbinding. The three cycles allow us to account for the effect of the TFs, since the rates can be different depending upon the state of the binding site. As an example, the darker arrows denote the activator effect of A and B on the first and second transitions, respectively ($k_{1,A} > k_{1,\emptyset}$, $k_{2,B} > k_{2,\emptyset}$). C) Schema of the full graph, as used in Figure 2.

TF binding on-rates (horizontal from the central cycle to the right and left) have units of inverse of concentration and time, and binding off rates have units of inverse of time. The genomic context is modelled by the values of the basal rates over the polymerase cycle in the absence of TFs (central cycle). To incorporate the effect of a TF on a transition rate, we assume that the TF only has effect while it is bound. The effect is then incorporated with the parameter value for that transition on the cycle where the TF is bound. As an example, the darker arrows on the left and right cycles

in Fig. 1B, $\mathcal{N} \times \mathcal{P}$, represent the activator effect of A and B on the first and second transitions, respectively. In this case, $k_{1,A} > k_{1,\emptyset}$, and $k_{2,B} > k_{2,\emptyset}$. Similarly, repression could be included as well by a smaller value for a transition rate than the corresponding basal rate. For simplicity here we examine synergy between "pure" activators only, defined by either not affecting or increasing the clockwise rates ($k_{1,x} \geq k_{1,\emptyset}$, $k_{2,x} \geq k_{2,\emptyset}$,$k_{3,x} \geq k_{3,\emptyset}$, $x \in \{A, B\}$) and not affecting or decreasing the counterclockwise rate ($k_{4,A} \leq k_{4,\emptyset}$, $k_{4,B} \leq k_{4,\emptyset}$).

We interpret the system in probabilistic terms, and assume each vertex of the graph holds the probability of the system being in that state. The transition rates determine the time-evolution of the probabilities, which eventually reach a steady state (Methods, 5.1). We assume first-order mRNA degradation. By taking the mRNA degradation rate as a constant that normalises the time unit of the transition rates, the steady-state mRNA (m^*) is given by:

$$m^* = k_{3,\emptyset} P^*_{3,\emptyset} + k_{3,A} P^*_{3,A} + k_{3,B} P^*_{3,B} \tag{1}$$

where $P^*_{i,X}$ are the steady-state probabilities of states $3_\emptyset$, 3_A, 3_B, and the $k_{3,x}$ are normalised (Methods, 5.1).

The focus of the analysis is to compare this quantity when both TFs are present to that when only one is present (synergy, below). Notice that when only one or none of the TFs is present, m^* can be computed in the same way. In that case, the steady-state probabilities for those states corresponding to the absent TF being bound will be 0, and the rest will be redistributed according to the parameter values. The value of m^* in the absence of TFs corresponds to basal expression.

3.2 A measure of synergy

Our interest is to understand how synergy emerges in this system. As shown by Scholes et al. (2017), if two TFs act on more than one step in the cycle, the overall effect may not be greater than additive even if they interact kinetically. In order to quantify synergy more comprehensively, we consider a two-dimensional quantity that compares the steady-state expression when both TFs are present ($m^*(A, B)$) to the steady-state expression when either of them is alone, but at twice as much concentration ($m^*(2A)$, $m^*(2B)$). In this way, the total concentration of TF is the same in the combined as in the individual situation. Enhanced expression in combination with respect to the strongest TF (the TF with a higher level of expression on its own), or reduced with respect to the weakest, must arise as a result of the functional interactions of the TFs over the cycle.

Positive synergy corresponds to higher expression in combination as compared to individually, and can be regarded as "true" synergy in the sense of enhanced expression in combination: expression is greater than that of the strongest TF even if half the molecules are substituted by those of a weaker TF. We note however that the output does not have to be greater than additive to fall into this class. Negative synergy corresponds to lower expression in combination, with expression lower than that of the weakest TF alone. Asymmetric synergy results when expression is increased only with respect to the weakest TF. In this case, it may be unclear whether there are any synergistic interactions. Potentially, these can still be detected depending on the extent to which the expression is reduced or increased with respect to the strongest or weakest TF, respectively. Thus, we propose to quantify synergy as a point in 2D, by comparing the effects of adding one TF to the other. This is quantified by $S_{A,B}$ (effect of B on A) and $S_{B,A}$ (the effect of A on B) as follows:

$$S_{A,B} = \log_2\left(\frac{m^*(A,B)}{m^*(2A)}\right) \tag{2}$$

$$S_{B,A} = \log_2\left(\frac{m^*(A,B)}{m^*(2B)}\right) \tag{3}$$

If A is taken to be the strongest TF, positive (green), asymmetric (blue) and negative (red) synergy map to 3 quadrants of a two-dimensional synergy space, as depicted in Fig. 2A.

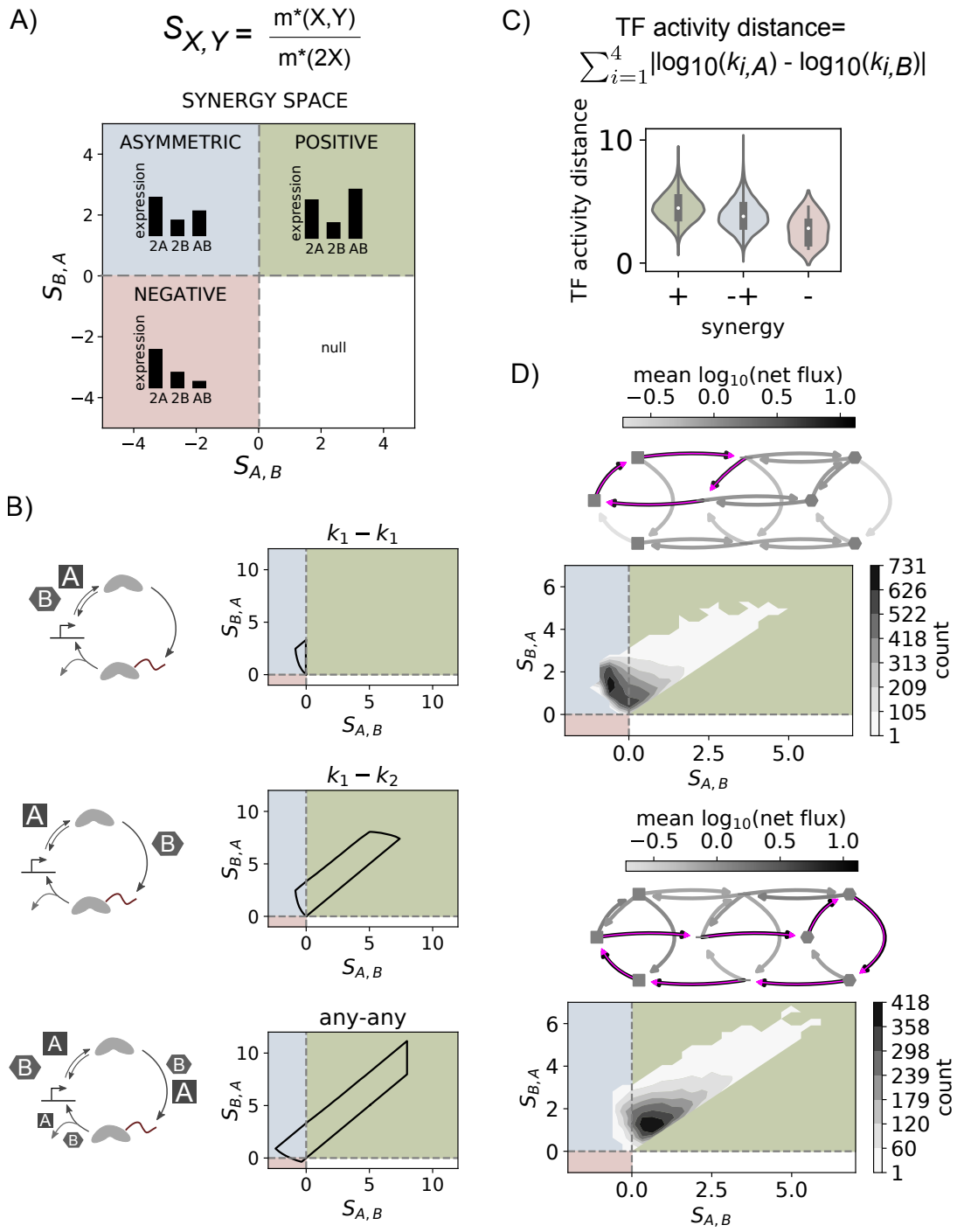

Figure 2: Positive, negative or asymmetric synergy emerge in the model depending upon parameter values. A) Synergy space. See text for details. B) Regions of the synergy space spanned by 3 regulatory strategies. Top: A and B act on the first step exclusively. Middle: A on one of the first two steps, B on the other one. Bottom: A and B act on any step (to various degrees). Constraints for the boundary search (see Methods): parameter values between 1 and 10^4, TF rates at most 1000 times larger than the basal rates for the clockwise (k_1,k_2, k_3) or 0.001 times smaller for k_4. Fold change in m^* for each TF individually (at concentration 2) with respect to the basal condition with no TF between 1 and 10. Fig. 2 Supplement 1B shows the results for more constrained parameters. *(continued on next page)*

Figure 2 ***(previous page)*****:** C) Distribution of TF activity distances per synergy quadrant for a random sample of parameter sets under the same constraints as in the bottom panel in B (synergies are plotted as a scatterplot in Fig. 2 Supplement 1C). D) The two most prevalent dominant flux paths for the points used in the analysis in C. The arrow diagrams represent the model states and transitions, as schematized in Fig. 1C; arrow color denotes the average probability net flux for that transition over all the parameter sets that share the dominant path highlighted in magenta. Notice that reversible edges may appear in both directions if some points have net flux in one direction and other points in the other. The distributions underneath show contours for the two-dimensional histogram of synergy values corresponding to those parameter sets that share the same dominant path.

3.3 Positive, negative or asymmetric synergy can theoretically emerge from two activators

We begin by exploring the theoretically possible synergistic behaviours between two activators ($k_{1,x} \geq k_{1,\emptyset}$, $k_{2,x} \geq k_{2,\emptyset}$, $k_{3,x} \geq k_{3,\emptyset}$, $k_{4,A} \leq k_{4,\emptyset}$, $x \in \{A, B\}$). In line with our experimental system where TFs share the binding domain, we assume that both TFs have the same binding kinetics (given by a binding rate k_b and an unbinding rate k_u) but different activation capabilities (given by the $k_{i,x}$, $i \in \{1, 2, 3, 4\}$, $x \in \{A, B\}$). We assume the concentration unit is incorporated in the binding on-rates, such that both A and B are present at a concentration of 1 unit each when they are both present together, and at concentration 2 when they are alone. In order to define the boundaries of the synergy space region that can be covered by the model under biologically-plausible parameter values and constraints (Methods, 5.2), we numerically sampled the parameter space using a biased sampling algorithm (Methods, 5.3). We explored the synergy space when TFs act on the same step, exclusively complementary steps, or all steps (Fig. 2B).

As a control, we first explored the case where both TFs enhance the first step. Fig. 2B-top shows that as expected, only asymmetric synergy appears in this case. Intuitively, if TF A drives stronger expression than TF B but both act on the same step, then mixing A with B can only reduce expression with respect to the strongest one, and increase it with respect to the weakest.

Next, we explored the case where TFs have complementary activities, with each TF enhancing either the first or second transition. Fig. 2B-middle shows that this control strategy mostly results in positive synergy, but also covers a region of the asymmetric synergy quadrant (notice that the result is restricted to the upper diagonal region of the positive quadrant due to the definition of TF A as the strongest of the pair). A very similar result is obtained for any other pair of complementary rates (Fig. 2 Supplement 1A). The appearance of asymmetric synergy in this case shows that even if TFs have complementary activities, that may not be enough to enhance expression beyond that of the strongest TF when half of its concentration is substituted by the weaker TF.

TFs have sometimes been classified as initiating or elongating factors. However, they are often found to interact with a wide range of cofactors and regulators (Dingar et al., 2015; Kim et al., 2017; Carnesecchi et al., 2020), and it is therefore likely that they modulate multiple processes albeit with different strengths. Hence, we next considered a more general scenario where each TF can enhance any of the transitions to different extents (Fig. 2B, bottom). In this case, a slightly higher region of the positive and asymmetric synergy quadrants are occupied, and slightly negative synergy can also emerge. We interpret this as indication that under some parameter values, TFs can interfere with each other's action and reduce the expression as compared to when only one of them is present.

For all these cases, the synergy space region that can be spanned by the model becomes smaller for more constrained parameter values, representing the assumption that the system has a smaller basal expression and TFs are weaker (Fig. 2 Supplement 1B).

3.4 The activity of the TFs over the cycle is not the only determinant of synergy

The original proposition of kinetic synergy stemmed from the assumption that synergy would emerge from TFs acting on different rate-limiting steps in transcription (Herschlag and Johnson, 1993). In the case of TFs with potentially overlapping effects, to what extent is positive synergy linked to TFs working exclusively, or nearly exclusively, on separate steps, so that they complement each other to enhance the cycle? In order to address this question, we looked at the correspondence between parameter values and synergy. For this, we generated a random sample of points that span a wide region of the synergy space (plotted in Fig. 2 Supplement 1C, Methods, 5.4). In order to quantify the degree of complementarity between the pair of TFs in a given parameter set, we use the following measure, which we call TF activity distance: the sum, over all the polymerase cycle transitions, of the absolute differences between the logarithms of the transition rates associated to each TF (Fig. 2C). Similar TF parameter values result in a small distance value, whereas TFs with big differences in their rates, and therefore more complementary to each other, result in a larger distance. As shown in Fig. 2C, positive synergy tends to emerge at higher distances than asymmetric and negative synergies, suggesting a higher complementarity is indeed linked to positive synergy. However, the distances that lead to asymmetric synergy and those that lead to positive synergy overlap, suggesting that the different functions of the TFs are not the only determinants of synergy output. When binning the distributions by the basal expression (steady state m^* in the absence of TFs) and binding and unbinding rates, these factors appear to be important as well: higher basal expression and higher binding and unbinding rates correlate with less distant TFs producing positive synergy (Fig. 2 Supplement 1D). In addition, the basal expression and binding rates also modulate the correlation between the distance of two TFs and the extent of positive synergy that they exhibit (Fig. 2 Supplement 1E).

Intuitively, for positive synergy to emerge, we would expect that each of the TFs binds and unbinds appropriately as to be able to exert its effect and not interfere with the binding and the effect of the other TF. In order to test the extent to which this is indeed linked to synergy, we looked at the steady-state probability fluxes in the graph. Given the irreversible nature of the transitions of the polymerase cycle, a net probability flux remains even when the system is at steady state. The flux of probability of the system is intimately linked to the production of mRNA, since mRNA is produced as the system transitions through the polymerase cycle. Formally, the flux from node i to node j, J_{ij} is given by $J_{i,j} = k_{ij}\, P_i$, with $k_{i,j}$ the transition rate between i and j, and P_i the probability of node i. In the case of irreversible edges, this equals the net flux. In the case of reversible edges, the net flux $\overline{J_{i,j}}$ can be defined as $\overline{J_{i,j}} = J_{i,j} - J_{j,i}$, with $J_{i,j} > J_{j,i}$.

For the same sample of points (parameter sets) as in Fig. 2C, we computed the net fluxes in the presence of A and B. Then, for each point, by starting at the polymerase-empty state with no TF bound (state $1, \emptyset$ in Fig 1B, $\mathcal{N} \times \mathcal{P}$) we followed the transition with a higher net flux, and repeated the same iteratively until reaching state $1, \emptyset$ again or any other node already encountered. This generates what we call the dominant path of net fluxes over the graph. After computing the dominant path for each of the paramater sets, we quantified how many parameter sets share the same dominant path. For this analysis, we pulled together those pairs of paths that are mirror images of each other, since they are equivalent.

Out of all the parameter sets sampled, the majority correspond to one of either two paths, represented in Fig. 2D. The most predominant involves the binding of one TF, transition over the first step (binding of polymerase), unbinding of the TF, and reversion to the empty state. The two-dimensional density plot below the flux diagram shows that the majority of the points with this dominant path of fluxes correspond to asymmetric synergy. In contrast, the second most frequent dominant path involves cycling over the whole graph, with the first transitions occurring under one TF, and the last occurring under the other. In this case, the majority of the points are associated to positive synergy. The rest of the dominant paths that make up to 90% of all the dominants paths in the sample of points are shown in Fig. 2 Supplement 2. The density plots show that dominant paths are not uniquely associated to individual synergy classes, but there are clear biases, with positive synergy being mostly associated to dominant paths that traverse the whole graph, and asymmetric synergy linked to dominant paths that show nonproductive cycling. This

agrees with the expectation that positive synergy should emerge when TFs act productively to enhance progression over the polymerase cycle, but also suggests that an intricate balance between all the transitions in the system is required for positive synergy to emerge.

3.5 Observation of kinetic synergy using a synthetic platform

In order to probe the extent of kinetic synergy experimentally achievable under this setup, we built a reporter platform based upon synthetic TF fusions (Khalil et al., 2012; Park et al., 2019b). We selected 5 activation domains of mammalian TFs with a described diversity of functions in the literature, depicted in Fig. 3A. SP1 is a ubiquituous mammalian transcription factor whose mechanism of action has classically been linked to the recruitment of the transcriptional machinery (O'Connor et al., 2016). cMyc is also a ubiquitous regulator. It interacts with a diverse range of proteins, but its mechanism of action has been predominantly linked to processes downstream the recruitment of the transcriptional machinery, including pause-release (Rahl et al., 2010) and elongation via interaction with the elongation factor Spt5 (Baluapuri et al., 2019). BRD4 has also been described to have elongating activity, through the interaction with positive transcription elongation factor b (Yang et al., 2005; Moon et al., 2005). In addition, it has been involved in phase-separation at super-enhancers (Vasile et al., 2018), suggesting that BRD4 may mediate other steps in the transcription cycle through the biophysics of the interactions it mediates. Finally we chose the activation domain of HSF1, which has been described to have both initiating and pause-release stimulating activity, and a mutant version of it, which we call HSF1-m, which was described to be elongation-deficient (Brown et al., 1998).

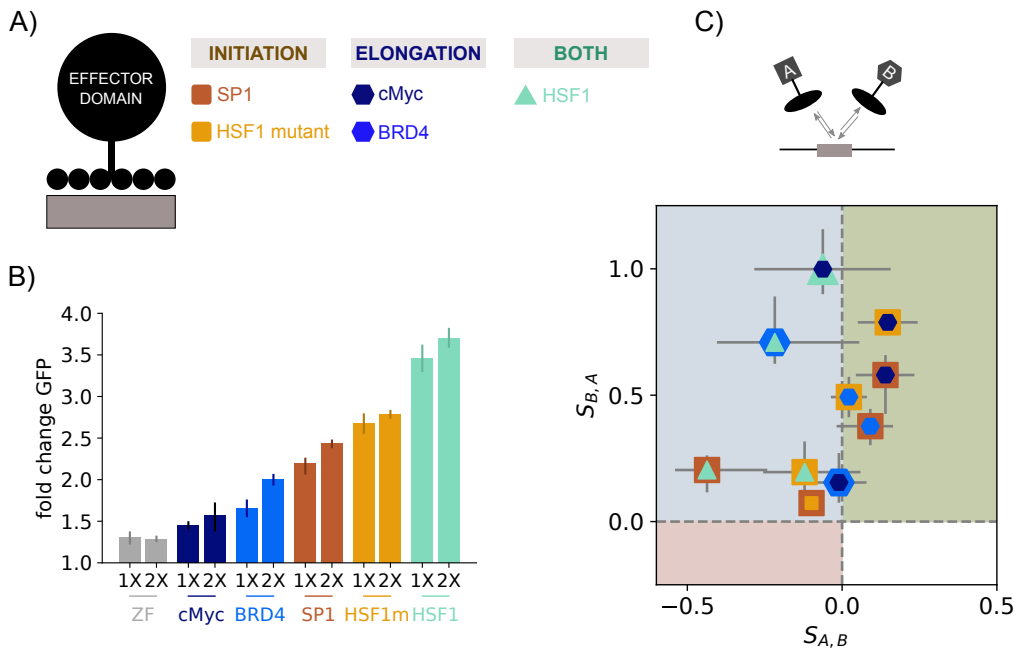

Figure 3: Observation of kinetic synergy between 2 transcriptional activators on a synthetic platform. A) Schema of the synthetic TFs. B) Expression upon transfection with either 10 (1X) or 20 (2X) ng of one TF, or the ZF binding domain alone (grey). Error bars denote the confidence interval for the mean GFP fold changes, obtained from bootstrapping of the average GFP fold change values from the technical replicates for each condition. C) Experimental synergy between two activators, defined as in Eqs. 2-3 ($\log_2$ of the ratio of average fold-change expression when 10 ng of each TF is transfected, over the average fold-change expression when 20 ng of one is transfected). TF A is the strongest of the pair in the single TF expression, as shown in the 2X conditions of panel B. Error bars denote ranges from at least two biological replicates, with 3 technical replicates each. Barplots corresponding to this data are shown in Fig. 3 Supplement 1D.

We engineered synthetic TFs (synTFs) composed of an activation domain from the above-described TFs fused to a synthetic zinc finger (ZF) DNA binding domain (Methods, 5.5), designed to target a 20-bp binding site that does not natively exist in the mammalian genome sequence (Fig. 1A, Fig. 3A) (Park et al., 2019b). This allows us to specifically recruit the activation domains to a reporter to assess their effects on transcription, while minimizing confounding effects from native TFs acting on the reporter. We then stably integrated into HEK293FT cells a reporter, composed of a single target binding site upstream of a minimal CMV (minCMV) promoter driving the expression of a destabilized EGFP (d2EGFP) (Methods, 5.7). Given its rapid turnover (Li et al., 1998), destabilized EGFP serves as a convenient genomic reporter of the mRNA expression level (Raj et al., 2006). The expression of the synTFs was induced by transient transfection of the synTFs, whose expression level can be controlled by the amount of the plasmids transfected (Fig. 3 Supplement 1A). We chose to transfect synTF at 10 ng each to ensure that the concentration (i.e. expression level) of synTF is the limiting factor. Reporter expression outcome was assessed by quantifying GFP fluorescence using flow cytometry 48 hours later (Methods, 5.8, 5.9).

Fig. 3B shows reporter activation by each of the synTFs. We observed similar activation strengths varying from about 1.5 fold change in GFP fluorescence to 4 fold change, with slight increases upon doubling the amount of TF transfected. Such fold change up-regulation is in the range of physiological induction in mammalian signalling pathways (Strasen et al., 2018; Wong et al., 2019; Friedrich et al., 2019). A similar dose-dependent increase in reporter signal is also observed at the mRNA level (Fig. 3 Supplement 1B).

In order to assess the extent of synergy between pairs of TFs, we compared the fold-change in GFP expression when TFs were transfected in pairs at 10 ng each, to the fold-change in GFP expression when only one is transfected at 20 ng. Fig. 3C shows that both positive and asymmetric synergy appears (See Fig. 3 Supplement 1D for details). Consistent with the correlation in the model between TF activity distance and synergy class, the pairs exhibiting positive synergy (Fig 3C, green quadrant) correspond to those where each TF predominantly has been described to have either initiating or elongating factor activities. No TF was capable of increasing the expression from that driven by HSF1, which is the strongest synTF in the set and is described to have both initiating and elongating activities (Brown et al., 1998). However, different TFs reduced its expression to different extents, suggesting some functional interactions are occurring (e.g. compare the $S_{A,B}$ coordinate for SP1-HSF1 and cMyc-HSF1 in Fig. 3C). For the pairs of TFs described to predominantly act upon the same step, almost no synergy was detected (SP1-HSF1m, cMyc-BRD4).

Fig. 3B shows a small activation effect of the empty ZF-binding domain alone. However, the combination with a full synTF only leads to asymmetric synergy (Fig. 3 Supplement 1C), with all TFs except HSF1 being reduced by the same extent, and HSF1 being reduced even further. This suggests that although the ZF may have a small effect perhaps through the chromatin or by increasing the ability of the basal transcriptional machinery to bind, the positive synergy observed between pairs of TFs is most likely due to their activation domains, since the ZF only reduces expression when mixed with any of the TFs.

3.6 Kinetic synergy depends upon the binding affinity

The analysis of the model in the previous sections suggests that the distance between the TFs is not the only determinant of synergy. The distributions of Fig. 2 Supplement 1D-E and the analysis of the dominant flux paths in Fig. 2 point to binding kinetics as an important contributor to synergy as well. To further investigate this point, we explored how the synergy exhibited by a pair of TFs changes in the model as a function of affinity, assumed to be inversely proportional to the unbinding rate. To this end, we randomly sampled parameter sets for binding and unbinding and the basal rates over the polymerase cycle (k_b, k_u, $k_{1,\emptyset}$,$k_{2,\emptyset}$,$k_{3,\emptyset}$,$k_{4,\emptyset}$). For each of these basal sets, we sampled parameter values for pairs of TFs ($k_{1,A}$, $k_{2,A}$, $k_{3,A}$, $k_{4,A}$, $k_{1,B}$, $k_{2,B}$, $k_{3,B}$, $k_{4,B}$). For each pair, we varied the unbinding rate over a 2 order magnitude range, 10 fold up and down the basal value, and tracked the corresponding behavior over the synergy space. Given that binding affinity changes expression from each TF alone, we only considered those parameter sets where the strongest TF is independent of the binding, so that synergy is consistently defined throughout.

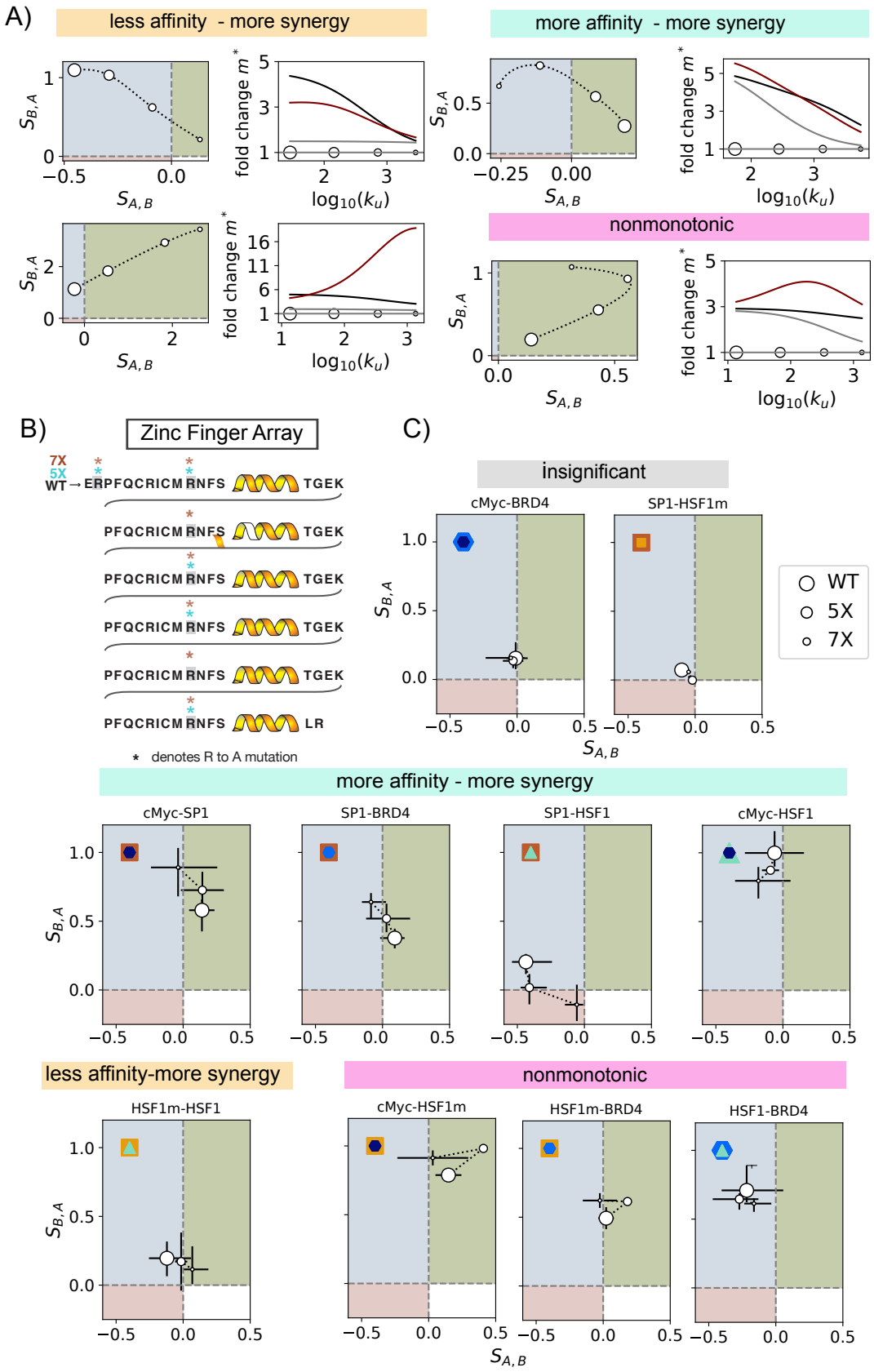

Figure 4: Synergy between a pair of TFs depends upon binding affinity. *(continued on next page)*

Figure 4 ***(previous page)*****:** A) Model examples for 4 sets of parameter values demonstrating the diversity in how synergy changes as a function of the unbinding rate. For each example, the left plot shows the behavior in synergy space. The right plot shows the corresponding fold change in expression as compared to no TF present, for either one of the TFs at concentration 2, or both TFs at concentration 1, as a function of the unbinding rate. The circles on the bottom of the right plot and those on the left plot correspond to the same values of synergy. B) Mutations in the ZF domains introduced to modulate affinity. C) Experimental synergy values for pairs of TFs with the WT binding domains, or those with reduced affinity (smaller marker size). Error bars denote ranges.

In order to classify the behavior over the synergy space systematically, we focused on the positive and asymmetric synergy behaviors, and used a 4-bit string that captures the affinity extremes: the first position denotes if $S_{A,B}$ is positive (p) or negative (n) at highest affinity, and the second position denotes the sign at the lowest affinity. The third and fourth positions denote whether $S_{A,B}$ and $S_{B,A}$ increase (i) or decrease (d), respectively. We disregard those situations where there is no change. As a result, there are theoretically 12 possible behaviors. We found that for some basal sets of parameters, all 12 possible behaviors emerged depending upon the pair of TF parameter values. One such example is shown in Fig. 4 Supplement 1, and selected examples from across basal parameter sets and TFs are highlighted in Fig. 4A.

As expected from typical occupancy-based hypothesis, we found instances where reducing affinity led to a reduction in synergy (Fig. 4A, more affinity-more synergy), changing from positive to asymmetric. In contrast, we also found examples where even if the expression from the individual TFs also decreases with less affinity, synergy increases and can change from asymmetric to positive as affinity is reduced (Fig. 4A, less affinity-more synergy). As seen in Fig. 4 Supplement 1, we found many instances of nonmonotonic behaviour, where synergy was maximal at intermediate affinities. Similar results were found when modulating the binding on-rate k_b (Fig. 4 Supplement 2). Notably, this can also be interpreted as modulating the baseline concentration of the TFs at 1X concentration. These results strengthen the conclusion from the analysis of the dominant paths of net probability fluxes, and show that synergy critically depends on an intricate balance between the TF binding and their functional effects on the transcription cycle.

In order to test whether binding affinity can experimentally change the synergy produced by pairs of TFs, we generated affinity variants of the synTFs. Specifically, this was accomplished by introducing alanine mutations to a series of arginine residues in the ZF framework known to mediate nonspecific interactions with the DNA phosphate backbone (Elrod-Erickson et al., 1996; Pavletich and Pabo, 1991)(Fig. 4B) . These mutations have been shown to reduce binding affinity in a dose-dependent manner (Khalil et al., 2012; Bashor et al., 2019).

As shown in Fig. 4C, we observed a range of behaviours upon tuning binding affinity experimentally, in agreement with the model. For a subset of TF pairs, synergy was reduced upon affinity reduction (more affinity-more synergy). Notably, for cMyc-SP1 and SP1-BRD4, the mutations changed synergy from positive to asymmetric. Synergy was also reduced for the other pairs shown on the middle row, leading to negative synergy appearing for SP1-HSF1. In contrast, for other pairs synergy increased with lower affinity and became positive (HSF1m-HSF1), or changed non-montonically, with maximal synergy at intermediate affinities for cMyc-HSF1m and HSF1m-BRD4. Notably, for the pairs of synTFs with the activation domains described to act upon the same step (cMyc-BRD4, SP1-HSF1m), their synergy was almost negligible, largely insensitive to the affinity and biased towards the asymmetric quadrant. This agrees with the model result in Fig. 2B top, where only asymmetric synergy appears for pairs of TFs acting on the same step, and the idea that positive synergy is linked to complementarity between the activities of the TFs on the transcription cycle.

3.7 A complex synergy landscape emerges when TFs have different activation and binding

Until now we have considered that the affinity is the same between the two TFs of the pair, which differ in their activation capability. Our analysis so far suggests that kinetic synergy emerges from the interplay between the complementarity in the activation activities of the TFs and the binding kinetics, so that TFs bind and unbind appropriately to have an optimal effect in the progression of the cycle. Therefore, we hypothesized that by combining pairs with different activation domains and affinities, synergy might be further enhanced. In agreement with this, the synergy space covered by the model expands slightly when weak TFs have different binding affinities, as compared to when their affinity is the same (Fig. 5A). Experimentally, we find results along the same lines. Fig. 5B shows the synergy between pairs of TFs with different binding affinities. For essentially all pairs, there is some combination of affinities that maximizes synergy beyond WT affinity, and that the synergy region covered by a given pair of TFs expands. Similar to Fig. 4C, we still find that for the pairs of TFs corresponding to both described to have predominantly initiating or elongating activities, the synergy space covered is more confined than the rest.

This data agrees with the results of the previous sections that show that the biochemical activities of the TFs are important for the synergy they exhibit, but their binding kinetics have a notable impact. This suggests a scenario where the combinatorial effect of TFs can be flexibly tuned by the combined effect of their biochemical activities and binding.

4 Discussion

Combinatorial control is widespread in biological systems. Even though well recognized and appreciated for enabling many of the complex behaviours of living organisms, it represents a central challenge for our ability to understand biology and intervene for biomedical or engineering purposes. In the context of eukaryotic transcription, combinatorial control occurs at multiple scales, with many enhancers controlling the activity of a gene, and in turn many TFs binding to a given enhancer (Spitz and Furlong, 2012). Here we have focused on this latter scale, where we have investigated the emergence of kinetic synergy. Theoretically proposed almost 30 years ago (Herschlag and Johnson, 1993), its experimental interrogation has been challenging largely due to the confound of physical interactions. To circumvent this limitation, we have used a synthetic biology approach. We have built a platform where only one TF can be specifically bound at any given time. By combining modelling and experiments, we have shown how kinetic synergy can emerge in this scenario. Importantly, we find that synergy between a pair of TFs is not an intrinsic feature of the pair, but depends upon the balance between their binding and their functional effects.

A measure of synergy

Deviation of additivity has been a common approach to quantify synergy, under the assumption that if TFs do not interact, then their combined effect should be the sum of the effects obtained when each TF is present alone (Carey, 1998). Deviation of multiplicativity has also been taken as a measure of synergy (Bintu et al., 2005a). However, in Scholes et al. (2017) we showed that when TFs interact functionally on a 2-step cycle, additivity or multiplicativity is only expected under very restricted circumstances. In order to provide a more comprehensive measure of synergy, here we propose to compare expression when both TFs are present together, to expression when only one of them is present, under the same total TF concentration. By having the same total TF concentration in both cases, changes in the expression when there are two TFs as compared to only one must be due to their interactions, and evidences synergy. In addition to positive synergy, asymmetric and negative synergy are defined in this approach. This enables the quantitative characterization of the synergy between a pair of TFs as a function of a variable of the system, by looking at the corresponding trajectory in synergy space.

We believe this measure could also be used to quantitatively characterize the expression of combinations of TFs in a more natural scenario where they bind to separate sites. As long as they are at relatively low concentration that ensure the binding is not saturated, by keeping the total con-

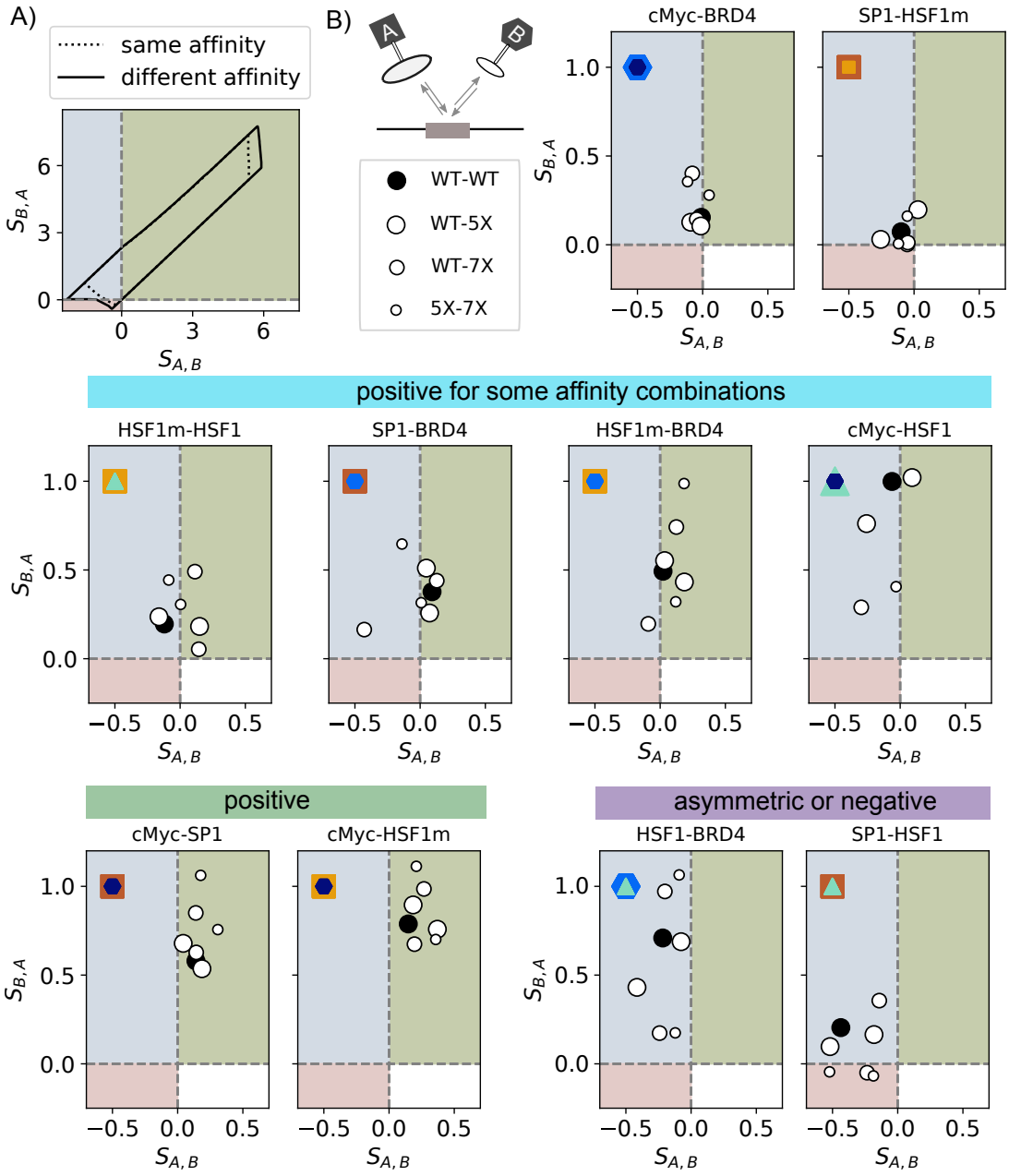

Figure 5: Different binding affinities between pairs of TFs expands the synergy space. A) Region of the synergy space spanned by the model under parameter constraints determining weak basal expression and weak TFs: basal expression parameter values between 1-100 for clockwise rates, 100-1000 for $k_{4,\emptyset}$. TF parameter values at most 100X greater (0.01X smaller for k_4). Fold change from each TF alone at 2X concentration limited to 5. k_b and k_u are either same for both TFs (dotted line), or different (solid line). For the case of same binding, it is the same result as the dotted line in Fig. 2 Supplement 1B, right. B) Experimental average data corresponding to the indicated affinity combinations. For each affinity pair, for the purpose of quantifying $S_{A,B}$ and $S_{B,A}$, TF A is the one driving higher expression (in some cases it changes for different affinity combinations).

centration of TF constant in the combined or the individual scenario, the effect of the combination in comparison to the individual will become apparent. This measure is likely to be more sensitive than assessing deviations from additivity or multiplicativity.

A synthetic biology approach to experimentally probe kinetic synergy

One of the main challenges to study kinetic synergy experimentally is the potential for direct or indirect binding interactions between the TFs if each TF binds to a distinct binding site. In order to circumvent this confound, we have exploited a synthetic approach where TFs have been engineered to bind to the same, shared site. By forcing the TFs to be separate in time, their functional interactions are evidenced. This work gives yet another example of the power of synthetic biology to answer fundamental biological questions (Crocker et al., 2017; Park et al., 2019a).

The modular and flexible design of the platform allows us to test many individual or pairs of TFs. In addition, the binding affinity of the TFs is tunable. Moreover, the target binding site on the reporter could be moved around to test the effect of binding site position, or more sites could be introduced to test combinations of TFs binding to different sites. However, our approach makes two simplifying assumptions that could be refined in future studies. First, the concentration of TF is controlled by transfection, and second, mRNA expression is assessed at the protein level by measuring GFP fluorescence. Although this introduces potential inaccuracies in our interpretation of the data, it enables the throughput required to measure the effects of the TFs individually and in combination as needed for the study, which quickly grow combinatorially.

A model that explicitly accounts for the interplay between TF binding and polymerase activity

In order to reason about the single binding site experiment, we have developed a model with details of both the binding of the TFs and the progression over the polymerase cycle. This model brings together the two main modelling frameworks of transcription in the literature, where either the binding is taken implicitly (Scholes et al., 2017), or the polymerase cycle is not detailed (Estrada et al., 2016). In contrast to other attempts in the literature (Li et al., 2018), we don't make assumptions about the timescales of the binding and unbinding of the TFs with respect to those of the biochemical transitions over the polymerase cycle. This provides greater generality. In addition, the model can easily be extended to include more polymerase states and more binding sites for other TFs or coregulators, if such details become relevant in future studies. One of the simplifying assumptions of the model is that TFs only exert their effect while they are bound. We note that this doesn't necessarily have to be the case, since they may act through other cofactors that can remain bound even if the TF unbinds. This could be easily incorporated at the expense of more states and parameters. However, it wouldn't fundamentally change our conclusions, since there would also be an interplay between the kinetics of the binding of these other components and those of the effects on the cycle.

We have explored the behavior of the model in parameter space under the assumption that the system is at steady state. This is a widely used assumption and reasonable for our experimental setup. However, one of the contexts where combinatorial control is most relevant is development, and many developmental processes may be too fast to allow for a steady state to be reached. In this case, it may become important to incorporate explicitly the time delay that emerges from polymerase travelling along the gene body, which we have not accounted for. Although at steady state this is likely to be effectively incorporated by the parameter of the last transition rate in the polymerase cycle, it could have important implications when considering how synergy emerges in transient regimes.

Kinetic synergy can emerge when two TFs time-share a binding site

We have found that extensive positive synergy is theoretically possible in the case where two activators share their binding site. Our analysis shows that this is due to TFs productively enhancing the polymerase cycle when acting in combination, by binding and unbinding appropriately as to allow each TF to exert its effect. As a result, synergy between a pair of TFs is strongly influenced

by the binding kinetics. Theoretically, both the binding on-rate and off-rate can modulate the synergy exhibited by a pair of TFs. Importantly, lower affinity can increase the synergy observed for a pair of TFs, even if this can reduce expression from the TFs acting individually. In some cases, the compromise is evidenced as a nonmonotonic effect of affinity upon synergy. The experimental data supports these theoretical findings.

These results highlight that synergy between a pair of TFs is not intrinsic to the TFs. Rather, it depends upon the context where they are: what are the basal rates in the absence of the TFs, and the binding kinetics, which can be given by the DNA binding site sequence, the DNA binding domain, or TF concentration. This extends to the kinetic case an observation already made for binding-based synergy, which was found dependent upon concentration (Carey et al., 1992; Frank et al., 2012).

Previous analyses had suggested that assessing synergy might be a way to elucidate the mechanism of action of TFs (Scholes et al., 2017). However, we show that this is confounded by the effect of the binding kinetics. Of note, parameter constraints that generate positive synergy in the model also generate asymmetric synergy. In this case, even if TFs may have complementary activities, their binding patterns may be imbalanced and may not allow productive interaction. In the case where either of the TFs works exclusively on one of two complementary steps, this contrasts with the finding of exclusively greater-than-additive behaviour by Scholes et al. (2017), highlighting the importance to account for the binding kinetics.

The model also shows that when TFs have overlapping activities, negative synergy can emerge even if individually they act to enhance the cycle. Again, this arises due to an imbalance between the timescales of their binding and functional effects, where in combination they interfere with each other. However, the extent of this effect is small and requires very fine tuned parameter sets, as evidenced by the low numbers of points in this region obtained from pure random sampling. In agreement with this, there is only one pair of TFs for which we observe negative synergy experimentally.

We notice that the extent of positive synergy experimentally observed is small compared to the regions covered by the model. In the model, we have found that the region of the synergy space is reduced as more constraints on the parameters are imposed, especially when constraining the extent to which a TF can enhance a given rate and their individual fold changes. This suggests that the synTFs have relatively weak effects, in agreement with the small fold-change activation that they produce.

Implications for gene regulation in natural scenarios

In endogenous enhancers, some TFs do have overlapping binding sites as in our setup (Han et al., 1998; Pan and Nussinov, 2011; Cheng et al., 2013). However, most typically, each TF has its own binding site. Even in this case, binding kinetics may still be important. The residence time of the TF on the DNA has to be long enough for it to be able to exert its function. However, there can be interferences either directly or through recruited cofactors, such that output may be maximized at intermediate affinities. This could be another reason behind the widespread presence of relatively low-affinity binding sites in eukaryotes (Crocker et al., 2016), and the observation of fast TF binding kinetics (Paakinaho et al., 2017; Li et al., 2019; Donovan et al., 2019). Moreover, tuning binding site affinity might be an effective way to modulate expression beyond fully adding or removing a binding site, which could have evolutionary implications (Kurafeiski et al., 2019). Along the same lines, kinetic synergy relaxes the need for strict arrangements between binding sites, another feature typical of eukaryotic transcriptional control (Junion et al., 2012; Smith et al., 2013).

TF activity has often been studied under a modular view, where the activity of the activation domain is independent of that of the binding domain, which is assumed to be important only to target the TF to specific sites on the genome (Ptashne, 1988). Evidence against this model includes allosteric interactions between the DNA binding domain and the activation domain (Li et al., 2017), and the observation that the activation domain may be involved in DNA recognition (Brodsky et al., 2020). Adding to this, our study highlights the importance of considering TFs as

a unit, where the binding and activation domains should be understood together. Beyond single TFs, our study also shows the relevance of incorporating an integrated view of transcriptional activity to the study of gene regulation, where the effect of a TF has to be understood in terms of the other components of the system.

5 Methods

5.1 Modelling details and the linear framework

In this work we have used the linear framework formalism to model the interplay between the TF binding and their effects on the transcription cycle. This framework was introduced in (Gunawardena, 2012) and we have previously exploited it to study other problems in gene regulation. Ahsendorf et al. (2014); Estrada et al. (2016) can be consulted for details. We outline the main features here.

In this setting, a biological system is represented by a finite, directed, labelled graph G with no self-loops. The graph represents a coarse-grained version of the system of interest, with the nodes being the states of interest, and the edges the transitions between them. The edge labels are the infinitesimal transition rates, with units of inverse of time. For transitions that represent the binding of a TF, the label includes the TF concentration, which is assumed to remain constant over time (i.e. TF is in excess and binding does not reduce the free TF concentration available for binding).

If we assume that each vertex can hold the probability of the system for being in that state, then the graph defines the time-evolution of the probability for each of the states of the system. Mathematically, this is represented by a linear differential equation (Master equation):

$$\frac{d\vec{P}}{dt} = \mathcal{L}(G)\vec{P}. \tag{4}$$

where $\vec{P}$ is a column vector of state probabilities at any given time, with dimension n, and $\mathcal{L}(G)$ is the graph Laplacian: a square matrix $n \times n$ where the off-diagonals contain the transition rates between nodes, and the diagonals are the negative of the column sums.

For a strongly connected graph, the system has a unique steady state, where $d\vec{P}/dt = 0$. The steady-state probability values for each state are computed by summing over the products of the rate labels for each of the spanning trees rooted at that state, and normalising appropriately (see Estrada et al. (2016) for details).

The mRNA is assumed to evolve according to:

$$\frac{dm}{dt} = k_{3,\emptyset}P_{3,\emptyset} + k_{3,A}P_{3,A} + k_{3,B}P_{3,B} - \delta_m\, m \tag{5}$$

where the $P_{i,X}$ are the probabilities of states $3_\emptyset$, 3_A, 3_B at a given time (Fig. 1B). By assuming steady state, setting $dm/dt = 0$, and dividing by δ_m, we obtain the expression for the steady state mRNA (* denotes steady state):

$$m^* = \frac{k_{3,0}}{\delta_m}P^*_{3,\emptyset} + \frac{k_{3,A}}{\delta_m}P^*_{3,A} + \frac{k_{3,B}}{\delta_m}P^*_{3,B} \tag{6}$$

$$m^* = \overline{k_{3,0}}P^*_{3,\emptyset} + \overline{k_{3,A}}P^*_{3,A} + \overline{k_{3,B}}P^*_{3,B} \tag{7}$$

This gives Eq. 1 of the main text, where the overbars are dropped for simplicity. In the parameter exploration, we directly sample on the normalised rates.

5.2 Biologically plausible ranges for parameter values

We considered a biologically plausible range for the normalised parameter values to be between 1 and 10^4, according to the following reasoning:

The events from the binding of the polymerase complex until the production of an mRNA molecule involve many biochemical reactions, including the binding interactions associated with the assembly of the pre-initiation complex, the phosphorylation of the C-terminal domain of RNA polymerase and other post-translational modifications (Schröder et al., 2013), as well as the biochemistry associated to elongation. Our 3-state cycle is therefore a coarse-grained representation of all these processes. In order to determine biologically plausible parameter ranges, we searched for measurements of reaction rates for these processes, and normalised those to typical rates of mRNA degradation, taken to have typical half-lives between 10 min (0.0167 s^{-1}) and 5 h (5.55×10^{-5} s^{-1}) (Sharova et al., 2009; Chan et al., 2018).

For a reaction at a rate of 1 s^{-1} (1 s half-life), normalizing by the mRNA degradation rates would result into a normalised range of 600-18000.

For a rate of 0.1 s^{-1} (10 s half-life), the normalised range would be 60-1800.

For a rate of 0.016 s^{-1} (1 min half-life), the normalised range would be 10-300.

And for a rate of 0.001667 s^{-1} (10 min half-life) the normalised range would be 1-30.

These values are consistent with measurements of various transcription-associated biochemical reactions: the *in vitro* rate of pre-initiation complex assembly was found to vary over ranges on the order 10^{-3} s^{-1} (Ferguson et al., 2001) to 0.1 s^{-1}(Kugel and Goodrich, 1998), and the rate of promoter opening/escape was reported to be 0.002 s^{-1} (Kugel and Goodrich, 1998). Pause stability is estimated to be from 3 s to 20 min (Wissink et al., 2019). And the TF residence time can be from just a few seconds to a few minutes (Paakinaho et al., 2017; Mehta et al., 2018).

Therefore, we took a range of 1-10^4 for our parameter values. We note that we also checked the results with smaller ranges, around slower rates for the polymerase cycle, but found that it didn't affect the results qualitatively, only reduced the synergy region as shown in Fig. 2 Supplementary 1B.

5.3 Synergy space boundary for a regulatory strategy

In order to determine the region of the synergy space that can be spanned by a given regulatory strategy, we used a biased random sampling algorithm, modified from that in Estrada et al. (2016). Parameters were chosen from a given range of normalised rate values, and TFs were assumed to at most modify the basal rates by a certain factor (see figure captions for the values corresponding to each figure). A maximum fold change for expression in the presence of one TF alone (at 2X concentration) was also pre-specified, such that parameter sets that generate expression outside this range were discarded. The steps of the algorithm are as follows:

1. Define constraints and two-dimensional grid of synergy values. Initialize with the hyperparameters (below).
2. Randomly sample parameter values from their range (in log scale) until 10 points are found that fall in different cells of the grid.
3. Until convergence: at each iteration, search the surrounding parameter space of each boundary point (see below) and keep the new parameter sets that generate synergy values not already found (empty cells). Convergence is determined by 2000 iterations where no new points occupying empty cells are found.

In order to search the surrounding parameter space of a given parameter set (point in synergy space), we followed 3 strategies (each point was modified using the 3 strategies at each iteration, provided there were sufficient points for steps 2 (10) and 3 (100)):

1. Randomly select a few parameter values and modify them.

2. "Pull" towards a target point in the direction determined by the centroid and the point being modified, away from the boundary: for 500 trials or until convergence, slightly modify the parameter set, and keep the new one if it generates a point in synergy space closer to the target.
3. "Pull" in the direction (approximately) perpendicular to the tangent between the point being modified and its neighbor, as in 2.

The algorithm depends on various hyperparameters: probability of selecting a parameter value for mutation (0.2, 0.5), probability of replacing an already-existing boundary parameter (0.2,0.6), width of the interval around a parameter value to sample for new parameter values (in log (base 10) scale: [-2,2],[-1.5,1.5],[-1,1]). Searches were run for all 12 combinations of hyperparameters, and results were merged together.

5.4 Random sample of points in synergy space

In order to randomly sample parameter values in the synergy space we followed a rejection sampling approach. Parameters were sampled logarithmically from its predefined range (1-10^4). The constraints on the maximum fold change effect on the polymerase cycle rates by the TFs were checked, as well as the constraint on the expression fold change by each of the TFs at 2X. We collected 1 million parameter sets. Then, in order to have a more uniform distribution of points over the synergy space, we binned the synergy space into a grid with bins every 0.025 $S_{A,B}$ and $S_{B,A}$, and kept one parameter set per bin. We repeated the procedure 10 times. The resulting points are those in Fig. 2 Supplement 1C.

5.5 Construct design and cloning

The reporter construct consists of a single synthetic zinc finger binding site (CGGCGTAGCCGATGTCGCGC) upstream of a minimal CMV promoter (taggcgtgtacggtgggaggcctatataagcagagctcgtttagtgaaccgtcagatcgcctgga) driving d2EGFP (EGFP destabilized with degradation domain (fusion with aa 422-461 of mouse ornithine decarboxylase)).

synTF fusion proteins containing an activation domain of interest fused to an N-terminal zinc-finger binding domain with a GGGGS flexible linker were driven under control of a ubiquitin promoter and contain a 5' sv40 nuclear localization sequence, C-terminal HA and rabbit globin polyA 3' UTR. Zinc finger binding domains were selected after bioinformatically screening synTF target sequences against the mammalian genome (Park et al., 2019b). To tune the affinity of the zinc finger, 5, or 7 arginine residues in the zinc finger array were mutated to alanine as denoted in Figure 4B. The following protein domains were selected and conjugated as respective activation domains according to previous studies:

SP1 (Residues 263 – 499) [PMID: 8278363]
NITLLPVNSVSAATLTPSSQAVTISSSGSQESGSQPVTSGTTISSASLVSSQASSSSFFTNANSY
STTTTTSNMGIMNFTTSGSSGTNSQGQTPQRVSGLQGSDALNIQQNQTSGGSLQAGQQKE
GEQNQQTQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVPNSGP
IIIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSN

cMyc (Residues 1-70) [PMID: 12177005]
MDFFRVVENQQPPATMPLNVSFTNRNYDLDYDSVQPYFY
CDEEENFYQQQQQSELQPPAPSEDIWKKFEL

BRD4 (Residues 1308-1362) [PMID: 24860166]
PQAQSSQPQSMLDQQRELARKREQERRRREAMAATIDMNFQSDLLSIFEENLF

HSF1 (Residues 370-529) [PMID: 9606196]
PEKCLSVACLDKNELSDHLDAMDSNLDNLQTMLSSHGFSVDTSALLDLFSPSVTVPDMSLP
DLDSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYTAQPLFLLDPGSVDTGSNDLPVLF

ELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

HSF1 mutant (Residues 370-529, F418A, F492A, F500A) [PMID: 9606196]
PEKCLSVACLDKNELSDHLDAMDSNLDNLQTMLSSHGFSVDTSALLDLASPSVTVPDMS
LPDLDSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYTAQPLFLLDPGSVDTGSNDLP
VLAELGEGSYASEGDGFAEDPTISLLTGSEPPKAKDPTVS

5.6 Cell culture

HEK293FT cells (Thermo Fisher Scientific) were used as a background cell line in this study. Cells were cultured in DMEM with L-glutamine, 4.5g/L Glucose and Sodium Pyruvate (Thermo Fisher Scientific) supplemented with 10% FBS (Clontech), GlutaMAX supplement (Thermo Fisher Scientific), MEM Non-Essential Amino Acids solution (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Thermo Fisher Scientific). Cells were maintained at 37°C with 5% CO2 in a humidified incubator, with splitting every 2-3 days.

5.7 Genomic integration of reporter constructs

Reporter lines were generated by site-specific integration of reporter constructs into HEK293FT cells using CRISPR/Cas9 mediated homologous recombination into the AAVS1 (PPP1R2C) locus as previously described (Park et al., 2019b). Briefly, 60,000 cells were plated in a 48-well plate and transfected the following day by PEI with a mixture of the following: 70ng of gRNA_AAVS1-T2 plasmid (Addgene 41820), 70 ng of VP12 humanSpCas9-Hf1 plasmid (Addgene 72247), and 175 ng of donor reporter plasmid. Donor reporter plasmids contain flanking arms homologous to the AAVS1 locus, a puromycin resistance cassette, and constitutive mCherry expression. After transfection, cells were cultured in 2 mg/mL puromycin selection for at least 2 weeks with splitting 1:10 every 3 days. Monoclonal populations for reporter cell lines were isolated by sorting single cells from this population into a 96-well plate and growing cell lines from each well. A minimum of 6 monoclonal cell lines that express high level of mCherry protein were transiently transfected with a strong synTF activator (HSF1 or VP16) and a monoclonal cell line to be used going forward was selected based on the fold-change of GFP expression relative to basal GFP level.

5.8 Transient transfection

Stable reporter cell lines were transfected with synTF plasmid constructs using polyethylenimine (PEI, Polysciences) as described in (Park et al., 2019b). 60,000-100,000 cells/well were plated in 48-well plates and transfected the following day with a total of 10ng per synTF, unless otherwise noted, with single stranded filler DNA (Thermo Fisher Scientific) up to 200ng total. 50ng of pCAG-iRFP720 (Addgene, #89687) was used as a transfection control plasmid. Two days after transfection, cells were collected and prepared for flow cytometry, unless otherwise noted.

5.9 Flow cytometry and data analysis

For each measurement, cells were harvested and run on an Attune NxT (Thermo Fisher Scientific) or LSR II (BD) Flow Cytometer equipped with a high-throughput auto-sampler. A minimum of 10,000 events were collected for each well and were gated by forward and side scatter for live cells and single cells, as described in (Park et al., 2019b). Cells were then gated by iRFP for transfection-positive populations. We observed a large transfection efficiency, which allowed us to obtain data in some repeat experiments where for technical reasons it was not possible to gate by iRFP transfection control. The geometric mean of GFP fluorescence distribution was calculated in FlowJo (Treestar Software). GFP expression fold-change was determined by normalizing with mean GFP intensity of the reporter only control. Flow cytometer laser/filter configurations used in this study were: EGFP (488 nm, 510/10), mCherry (561 nm, 615/25), iRFP-720 (638 nm, 720/30). All flow cytometry measurements were performed in technical replicates.

5.10 Western blotting

A reporter cell line was transfected with indicated amounts of ZF-HSF1 (0, 10, 20, 50, 100, 150, 200ng) in a 48-well plate at a cell density of 1x105 per well. After 2 days, cells were rinsed with PBS and lysed with 200 μL of NuPAGE LDS sample buffer (Thermo Fisher Scientific), followed by 5 seconds of sonication. Whole cell lysates were mixed NuPAGE Sample Reducing agent (10X, Thermo Fisher Scientific) at 95°C for 5 minutes. Samples were then loaded into a 4-12% NuPAGE Bis-Tris Mini Protein precast gel (Thermo Fisher Scientific) and were run at 200V for 30 minutes in NuPAGE MES SDS Running Buffer. Separated proteins were transferred to a PVDF membrane using P0 protocol of iBlot2 system (Thermo Fisher Scientific). Membranes were blocked for 1hr at room temperature in blocking solution (5% w/v nonfat dry milk in 1X PBST) with gentle rocking. The membranes were probed with anti-HA (1:4000; Abcam ab9110) and anti-GAPDH (1:1000; Abcam ab9485) antibodies at room temperature for 1 hour with gentle rocking. The membranes were washed in PBST three times for 5 minutes each, and incubated with a goat anti-rabbit IgG-HRP antibody (1:2000; Abcam ab6721). The target proteins were visualized by chemiluminescence using SuperSignal West Pico PLUS substrate (Thermo Fisher Scientific) and an iBright Western Blot Imaging Systems (Thermo Fisher Scientific). Quantification of band intensities was carried out using FIJI (Schindelin et al., 2012).

5.11 Quantitative Real-Time PCR

1×10^5 Hek293 reporter cells were seeded one day prior to transfection in 6cm culture dishes. Transfection was performed with the indicated amounts of synTF plasmid as described above for flow cytometry experiments using polyethylenimine (PEI) (polyscience) or Lipofectamine 3000 (Thermo Fisher Scientific). Two days post transfected, cell pellets were harvested and mRNA was extracted using the RNeasy Mini Kit (Qiagen). 500 ng extracted total RNA was reverse transcribed into cDNA for each sample. Reverse transcription was performed using Protoscript II reverse transcriptase (New England Biolabs) and oligo-dT primers (New England Biolabs). Quantitative real-time PCR was performed in triplicates using iTaq^{TM}Universal SYBR®Green reagent (Bio-Rad) on a CFX96 PCR machine (Bio-Rad). Primers were used in a final concentration of 243.2 nM. β-actin expression was used as a reference gene for relative quantification of RNA levels. Used primer sequences are (5'-3'):
Actin_fwd: GGCACCCAGCACAATGAAGATCAA;
Actin_rev: TAGAAGCATTTGCGGTGGACGATG;
eGFP_fwd: AAGTTCATCTGCACCACCG; eGFP_rev: TCCCTTGAAGAAGATGGTGCG;
zf_fwd:TTTTCGAGAAGACACGGCCT; zf_rev: GCTGCTGTGGTCGGAGAAAT

6 Acknowledgements

We thank members of the DePace, Khalil and Gunawardena labs for helpful discussions. This work was supported by NSF grant MCB-1713855 (A.S.K.)/ MCB-1715184 (A.H.D), NIH grant 1R01GM122928 (A.H.D. and J.G.), NIH grant R01EB029483 (A.S.K.), NSF CAREER IOS-1452557 (A.H.D.), EMBO Fellowship 683-2019 (R.M.C.) and the Giovanni Armenise Harvard Foundation (A.H.D.). A.S.K. also acknowledges funding from the NIH Director's New Innovator award (1DP2AI131083-01), an NSF CAREER award (MCB-1350949), and a DARPA Young Faculty award (D16AP00142). A.H.D. also gratefully acknowledges funding from the Lynch Foundation Systems Biology Fellowship, the McKenzie Family Charitable Trust Systems Biology Fellowship Fund, and the John and Virginia Kaneb Fellowship award at Harvard Medical School.

7 Supplementary Figures

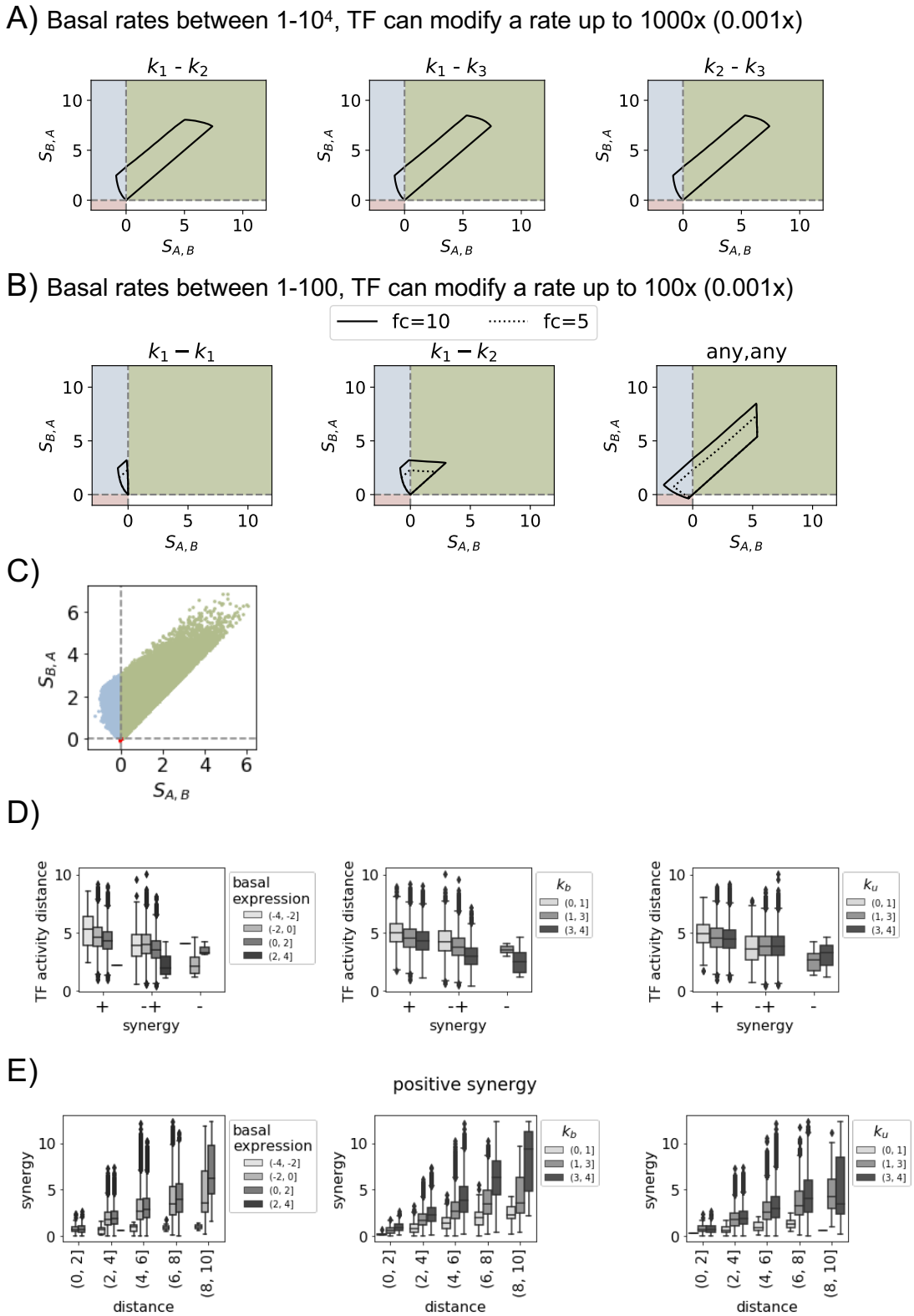

Figure 2 Supplement 1 Characterization of the model behavior in synergy space. *(caption on next page)*

(from previous page) A) Region of the synergy space spanned by the model when each TF acts uniquely on one of two steps, indicated in the title of each plot. The first plot is the same as in Fig. 2B middle. Parameter values in the range between 1 and 10^4. Parameters for the TFs at most 1000 times larger than the basal parameters for the clockwise rates (k_1,k_2, k_3) or 0.001 times smaller for k_4. Fold change in m^* for each TF individually with respect to the basal condition with no TF bound between 1 and 10. B) Region of the synergy space from each of the 3 regulatory strategies in Fig. 2B for more constrained parameters: parameter values for the transitions over polymerase states in the range between 1 and 100 for the clockwise rates, 100-10000 for $k_{4,\emptyset}$; parameters for the TFs at most 100 times larger than the basal parameters for the clockwise rates (k_1,k_2, k_3) or 0.01 times smaller for k_4. Fold change in m^* for each TF individually with respect to the basal condition with no TF bound between 1 and 10 (solid line) or 1 and 5 (dashed line). C) Random sample of points where both TFs act on any step, randomly sampled under the same constraints as in Fig. 2B bottom (Methods, 5.4) and used for the distributions in Fig. 2C and panels D and E of this Figure. D) Distribution of TF activity distances as a function of synergy class, binned by basal expression (expression in the absence of TFs, left), binding on-rate (middle), binding off-rate (right), for the points in panel C. E) Distribution of synergy values as a function of TF activity distance and binned by basal expression, binding on-rate or binding off-rate for the points with positive synergy in panel C (green points).

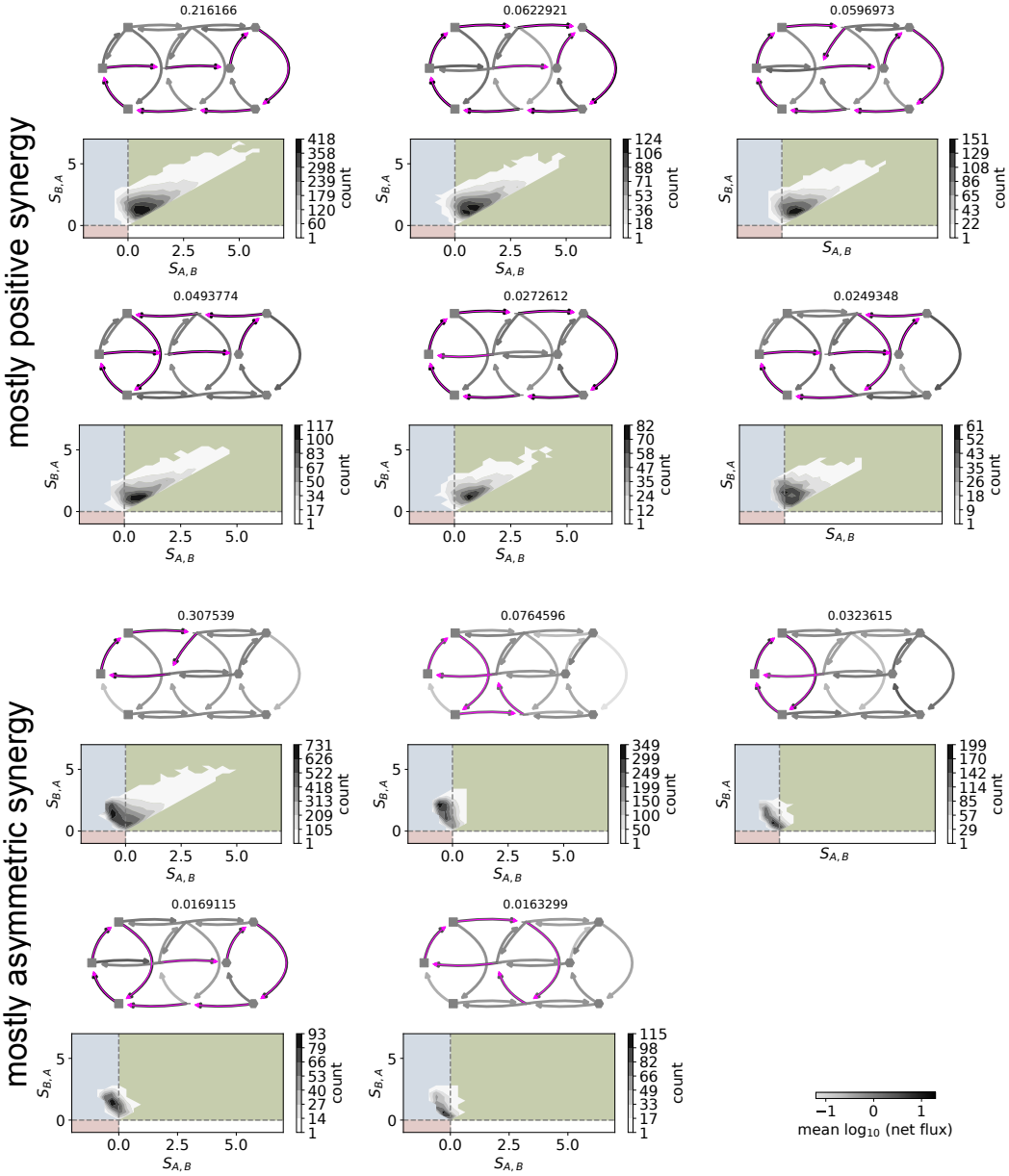

Figure 2 Supplement 2 Top 90% net flux dominant paths for each parameter set in Fig. 2 Supplement 1C. For each path, the representation is as in Fig. 2D. The top diagram shows the dominant path (magenta arrows) and the arrow color encodes the average net flux for the transitions. The lower plot shows the distribution of synergy values that correspond to that dominant path. For each of the two groups, dominant paths are sorted according to their frequency, indicated on top of each. Note that the first path of each group corresponds to the plot in Fig. 2D.

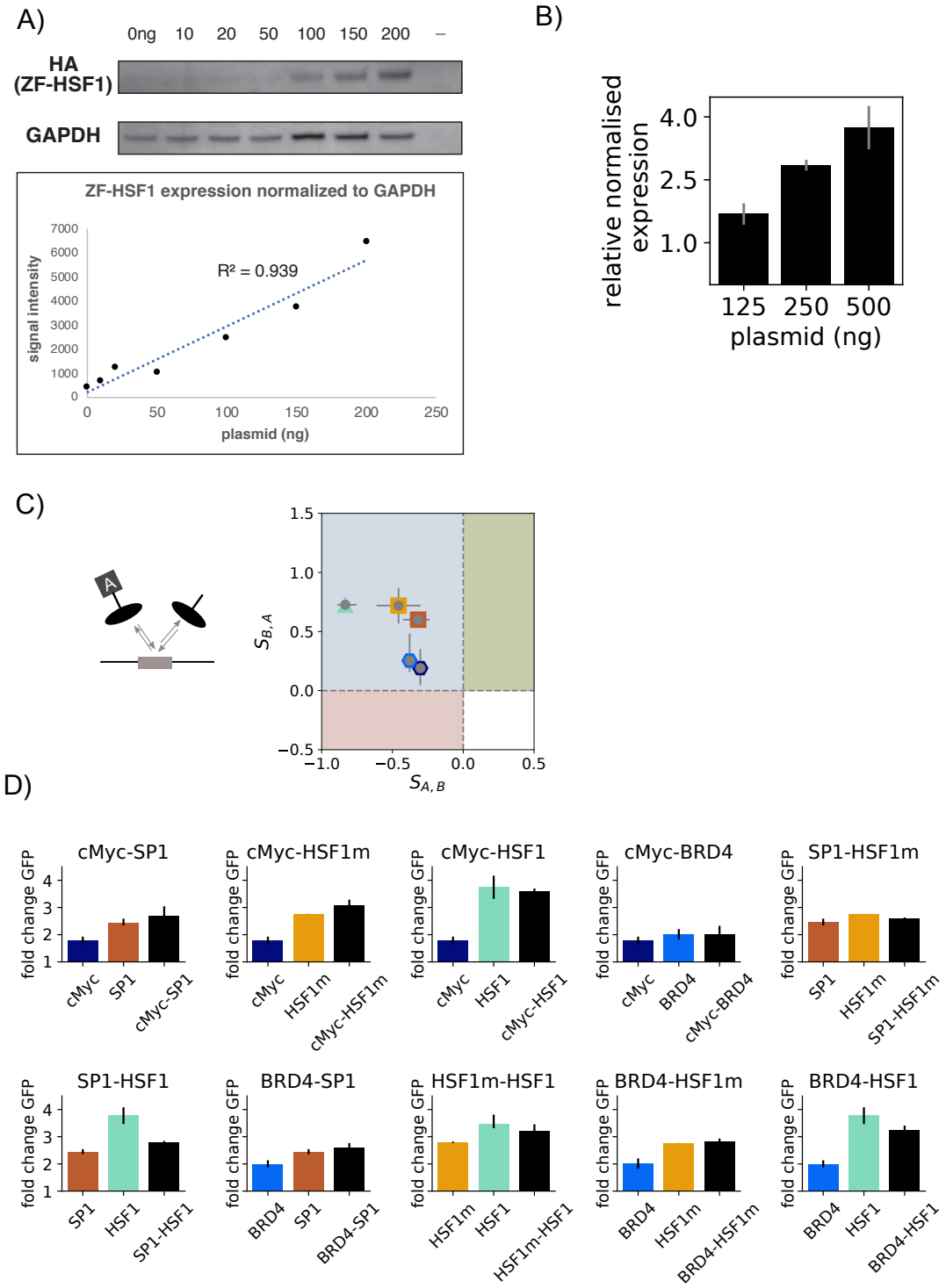

Figure 3 Supplement 1 A) Western Blot and corresponding quantification of the ZF-HSF1 protein as a function of ng of plasmid transfected (Methods, 5.10). B) Normalised GPF expression as measured by qPCR in response to SP1 synTF. The concentrations of plasmid used are scaled by the number of cells (Methods, 5.11) so that 125 and 250 ng are approximately equivalent to 10 and 20 ng in the flow cytometry experiments. Error bars denote SEM from technical replicates. C) The combination of a synTF with the ZF alone only generates asymmetric synergy, where expression is between that of the ZF and that of the full TF. Error bars denote the ranges of the data. D) Details of the fold change in expression for the conditions that generate the synergy plot in Fig. 3C.

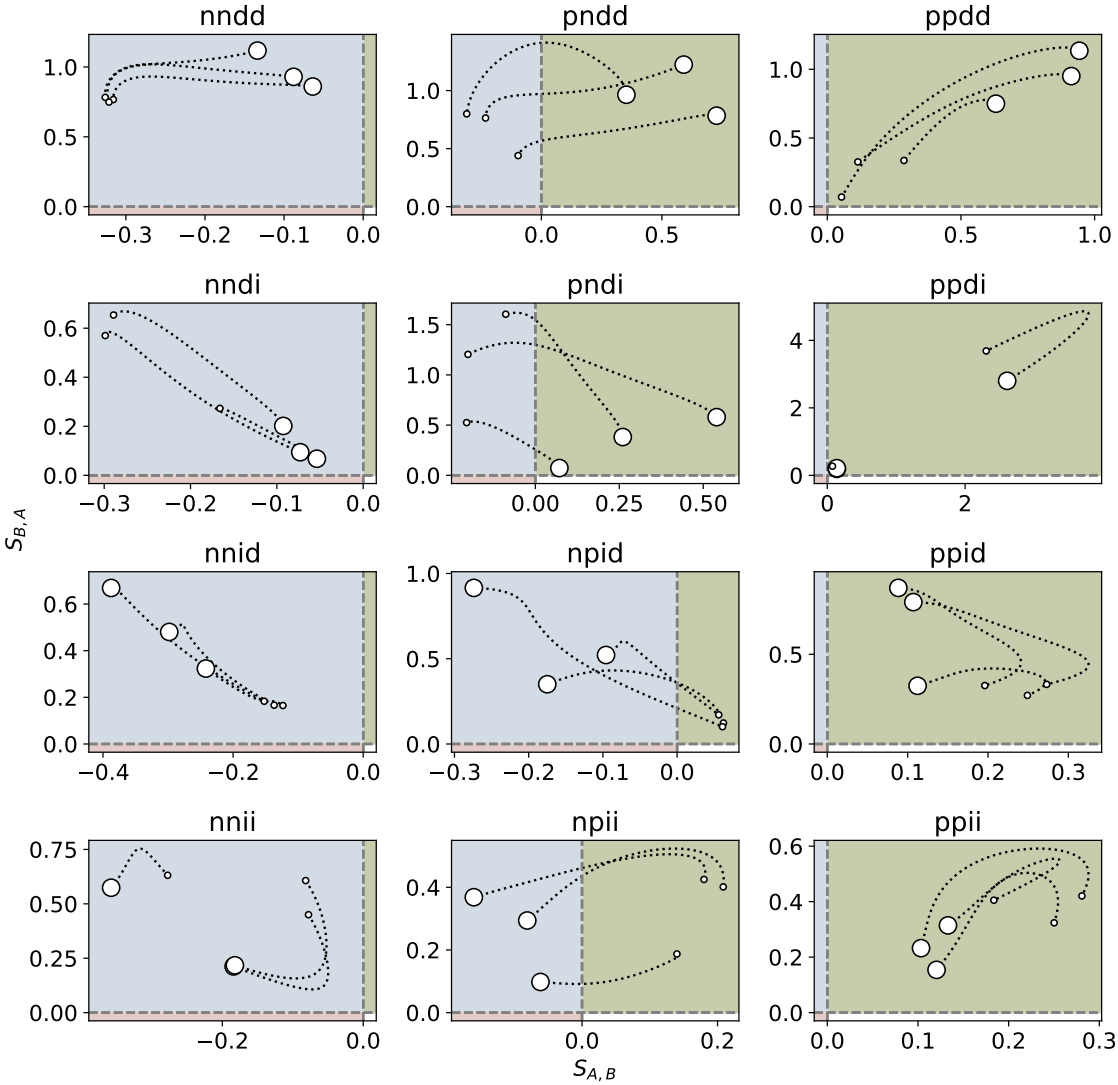

Figure 4 Supplement 1 Example of the 12 possible changes in synergy when the unbinding rate (k_u) is modulated (both $S_{A,B}$ and $S_{B,A}$ decrease: nndd, ppdd, pndd. $S_{A,B}$ decreases but $S_{B,A}$ increases: nndi, pndi, ppdi. $S_{A,B}$ increases but $S_{B,A}$ decreases: nnid, npid, ppid. And both increase: nnii, npii, ppii) for a set of basal rates and different pairs of TF parameter values. Each line corresponds to a given pair of TF parameter values as the unbinding rate is modified. All the lines share the same basal and binding rates. The biggest marker denotes highest affinity (lowest unbinding rate), and the smallest marker denotes lowest affinity (higher unbinding rate). Result is selected from a rejection-based sampling random search of parameter values, with parameter constraints as follows: basal rates for the polymerase cycle between 1-10^4. Binding and unbinding rate between 10-1000 (so that unbinding rate can be modified 10 fold up or down and remain in the range 1-10^4). Parameter values at most 1000X the respective basal ones (0.001X for counter-clockwise $k_{4,\emptyset}$). Fold change expression by each TF individually at 2X concentration between 1 and 5.

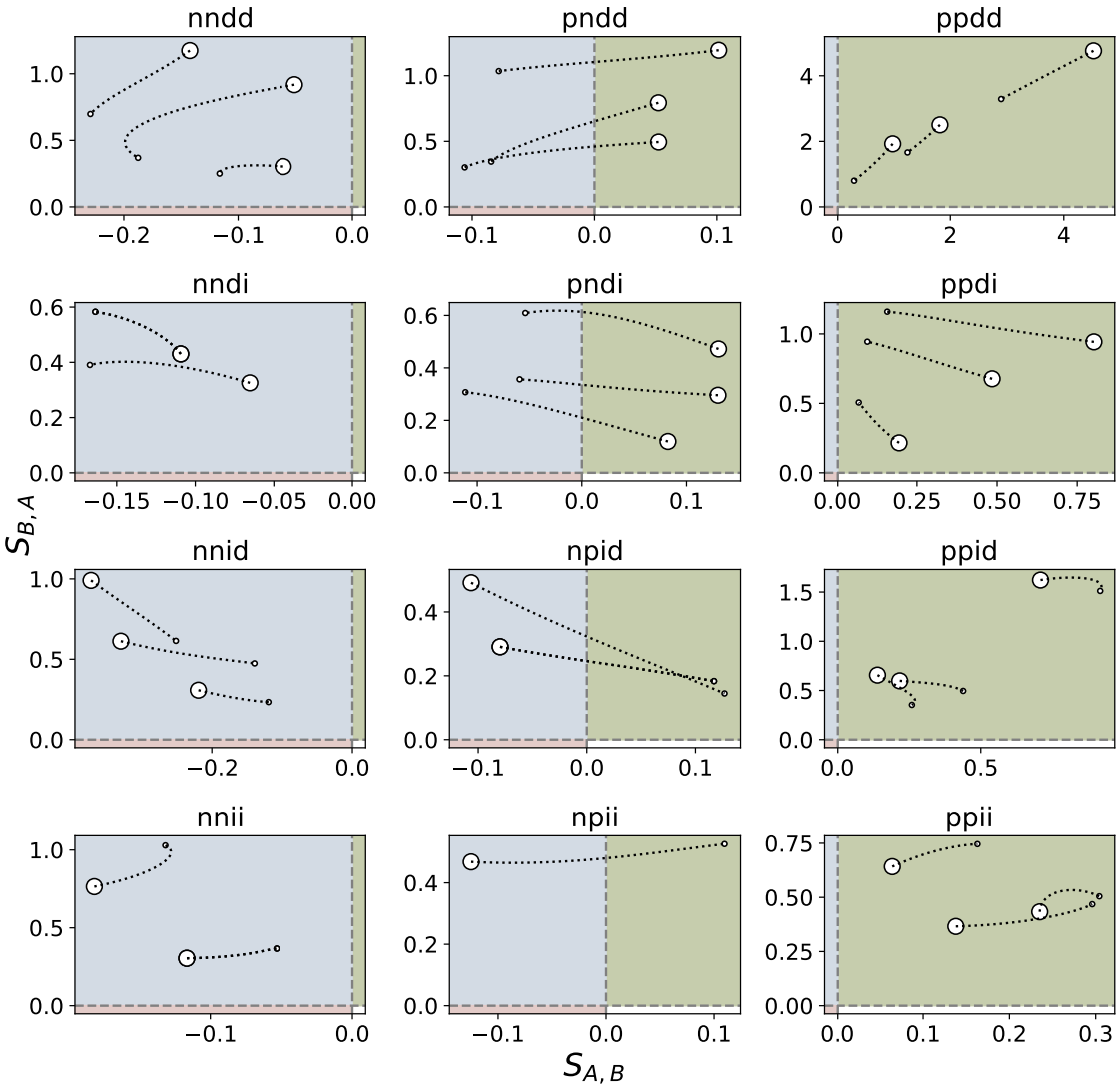

Figure 4 Supplement 2 Example of the 12 possible changes in synergy when the binding rate (k_b) is modulated, similar to Figure 4 Supplement 1 (same parameter constraints for the random search). Each line corresponds to a pair of TF parameter values, but all of them share the same basal and binding rates. The biggest marker denotes highest affinity (largest binding rate), and the smallest marker denotes lowest affinity (lowest binding rate).

References

Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences*, 79(4):1129–1133.

Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014). A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. *BMC Biology*, 12(1):102.

Ambrosetti, D. C., Schöler, H. R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the Fibroblast growth factor-4 enhancer. *Journal of Biological Chemistry*, 275(30):23387–23397.

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S. M., Adhikari, B., Schwarz, J. D., Narain, A., Vogt, M., Wang, S. Y., Düster, R., Jung, L. A., Vanselow, J. T., Wiegering, A., Geyer, M., Maric, H. M., Gallant, P., Walz, S., Schlosser, A., Cramer, P., Eilers, M., and Wolf, E. (2019). MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Molecular Cell*, 74(4):674–687.e11.

Bashor, C. J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J. J., and Khalil, A. S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science*, 364(6440):593–597.

Biddle, J. W., Nguyen, M., and Gunawardena, J. (2019). Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. *eLife*, 8.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development*, 15(2):125–35.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–24.

Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. *Molecular and Cellular Biology*, 16(5):2044–2055.

Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., and Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular Cell*, 79:1–13.

Brown, S. A., Weirich, C. S., Newton, E. M., and Kingston, R. E. (1998). Transcriptional activation domains stimulate initiation and elongation at different times and via different residues. *EMBO Journal*, 17(11):3146–3154.

Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell*, 92(1):5–8.

Carey, M., Kolman, J., Katz, D. A., Gradoville, L., Barberis, L., and Miller, G. (1992). Transcriptional synergy by the Epstein-Barr virus transactivator ZEBRA. *Journal of Virology*, 66(8):4803–4813.

Carey, M., Lin, Y.-s., Green, M. R., and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature*, 345(May):361–364.

Carnesecchi, J., Sigismondo, G., Domsch, K., Baader, C. E. P., Rafiee, M. R., Krijgsveld, J., and Lohmann, I. (2020). Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity. *Nature Communications*, 11(1).

Chan, L. Y., Mugler, C. F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife*, 7:1–32.

Cheng, Q., Kazemian, M., Pham, H., Blatti, C., Celniker, S. E., Wolfe, S. A., Brodsky, M. H., and Sinha, S. (2013). Computational Identification of Diverse Mechanisms Underlying Transcription Factor-DNA Occupancy. *PLoS Genetics*, 9(8).

Core, L. and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & Development*, 33:1–23.

Crocker, J., Preger-Ben Noon, E., and Stern, D. L. (2016). *The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution*, volume 117. Elsevier Inc., 1 edition.

Crocker, J., Tsai, A., and Stern, D. L. (2017). A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. *Cell Reports*, 18(1):287–296.

Cui, G., Dong, Q., Duan, J., Zhang, C., Liu, X., and He, Q. (2020). NC2 complex is a key factor for the activation of catalase-3 transcription by regulating H2A.Z deposition. *Nucleic Acids Research*, pages 1–17.

Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., Siepel, A., and Kraus, W. L. (2013). Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Molecular Cell*, 50(2):212–222.

de Boer, C. G., Vaishnav, E. D., Sadeh, R., Abeyta, E. L., Friedman, N., and Regev, A. (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature Biotechnology*, 38(1):56–65.

Dingar, D., Kalkat, M., Chan, P. K., Srikumar, T., Bailey, S. D., Tu, W. B., Coyaud, E., Ponzielli, R., Kolyar, M., Jurisica, I., Huang, A., Lupien, M., Penn, L. Z., and Raught, B. (2015). BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors. *Journal of Proteomics*, 118:95–111.

Donovan, B. T., Huynh, A., Ball, D. A., Patel, H. P., Poirier, M. G., Larson, D. R., Ferguson, M. L., and Lenstra, T. L. (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *The EMBO journal*, pages 1–18.

Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996). Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. *Structure*, 4(10):1171–1180.

Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information Integration and Energy Expenditure in Gene Regulation. *Cell*, 166(1):234–244.

Ferguson, H. A., Kugel, J. F., and Goodrich, J. A. (2001). Kinetic and mechanistic analysis of the RNA polymerase II transcription reaction at the human interleukin-2 promoter. *Journal of Molecular Biology*, 314(5):993–1006.

Field, A. and Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annual Review of Biochemistry*, 89(1):213–234.

Frank, T. D., Carmody, A. M., and Kholodenko, B. N. (2012). Versatility of cooperative transcriptional activation: A thermodynamical modeling analysis for greater-than-additive and less-than-additive effects. *PLoS ONE*, 7(4).

Friedrich, D., Friedel, L., Finzel, A., Herrmann, A., Preibisch, S., and Loewer, A. (2019). Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency . *Molecular Systems Biology*, 15(12):1–20.

Fu, D., Wen, Y., and Ma, J. (2004). The co-activator CREB-binding protein participates in enhancer-dependent activities of bicoid. *Journal of Biological Chemistry*, 279(47):48725–48733.

Fuda, N. J., Ardehali, M. B., and Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, 461(7261):186–92.

Georges, A. B., Benayoun, B. A., Caburet, S., and Veitia, R. A. (2010). Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *The FASEB Journal*, 24(2):346–356.

Goldstein, I., Paakinaho, V., Baek, S., Sung, M.-H., and Hager, G. L. (2017). Synergistic gene expression during the acute phase response is characterized by transcription factor assisted loading. *Nature Communications*, 8(1):1849.

González, A., Jafari, S., Zenere, A., Alenius, M., and Altafini, C. (2019). Thermodynamic model of gene regulation for the Or59b olfactory receptor in Drosophila. *PLOS Computational Biology*, 15(1):e1006709.

Gunawardena, J. (2012). A Linear Framework for Time-Scale Separation in Nonlinear Biochemical Systems. *PLoS ONE*, 7(5):e36321.

Han, W., Yu, Y., Su, K., Kohanski, R. A., and Pick, L. (1998). A Binding Site for Multiple Transcriptional Activators in the fushi tarazu Proximal Enhancer Is Essential for Gene Expression In Vivo. *Molecular and Cellular Biology*, 18(6):3384–3394.

Hansen, A. S. and O'Shea, E. K. (2013). Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Molecular systems biology*, 9(704):704.

Herschlag, D. and Johnson, F. B. (1993). Synergism in transcriptional activation: a kinetic view. *Genes & Development*, 7(2):173–179.

Inukai, S., Kock, K. H., and Bulyk, M. L. (2017). Transcription factor–DNA binding: beyond binding site motifs. *Current Opinion in Genetics and Development*, 43:110–119.

Johnson, A. D., Meyer, B. J., and Ptashne, M. (1979). Interactions between DNA-bound repressors govern regulation by the λ phage repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 76(10):5061–5065.

Jonkers, I. and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):167–177.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E. H., Birney, E., and Furlong, E. E. M. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, 148(3):473–486.

Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J., and Khalil, A. S. (2014). Using Targeted Chromatin Regulators to Engineer Combinatorial and Spatial Transcriptional Regulation. *Cell*, 158(1):110–120.

Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., and Collins, J. J. (2012). A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. *Cell*, 150(3):647–658.

Kim, B. R., Coyaud, E., Laurent, E. M., St-Germain, J., Van De Laar, E., Tsao, M. S., Raught, B., and Moghal, N. (2017). Identification of the SOX2 interactome by BioID reveals EP300 as a mediator of SOX2-dependent squamous differentiation and lung squamous cell carcinoma growth. *Molecular and Cellular Proteomics*, 16(10):1864–1888.

King, D. M., Hong, C. K. Y., Shepherdson, J. L., Granas, D. M., Maricque, B. B., and Cohen, B. A. (2020). Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. *eLife*, 9:1–24.

Kugel, J. F. and Goodrich, J. A. (1998). Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 95(16):9232–9237.

Kurafeiski, J. D., Pinto, P., and Bornberg-Bauer, E. (2019). Evolutionary potential of cis-regulatory mutations to cause rapid changes in transcription factor binding. *Genome Biology and Evolution*, 11(2):406–414.

Li, C., Cesbron, F., Oehler, M., Brunner, M., and Höfer, T. (2018). Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. *Cell Systems*, 6(4):409–423.e11.

Li, J., Dong, A., Saydaminova, K., Chang, H., Wang, G., Ochiai, H., Yamamoto, T., and Pertsinidis, A. (2019). Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell*, 178(2):491–506.e28.

Li, J., White, J. T., Saavedra, H., Wrabl, J. O., Motlagh, H. N., Liu, K., Sowers, J., Schroer, T. A., Thompson, E. B., and Hilser, V. J. (2017). Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *eLife*, 6:1–17.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C. C., and Kain, S. R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. *Journal of Biological Chemistry*, 273(52):34970–34975.

Malik, S. and Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics*, 11(11):761–772.

Mao, C., Brown, C. R., Falkovskaia, E., Dong, S., Hrabeta-Robinson, E., Wenger, L., and Boeger, H. (2010). Quantitative analysis of the transcription control mechanism. *Molecular Systems Biology*, 6(431):1–12.

Mehta, G. D., Ball, D. A., Eriksson, P. R., Chereji, R. V., Clark, D. J., McNally, J. G., and Karpova, T. S. (2018). Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Molecular Cell*, 72(5):875–887.e9.

Michida, H., Imoto, H., Shinohara, H., Yumoto, N., Seki, M., Umeda, M., Hayashi, T., Nikaido, I., Kasukawa, T., Suzuki, Y., and Okada-Hatakeyama, M. (2020). The Number of Transcription Factors at an Enhancer Determines Switch-like Gene Expression. *Cell Reports*, 31(9):107724.

Mirny, L. a. (2010). Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22534–22539.

Moon, K. J., Mochizuki, K., Zhou, M., Jeong, H. S., Brady, J. N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell*, 19(4):523–534.

Narasimhan, K., Pillay, S., Huang, Y. H., Jayabal, S., Udayasuryan, B., Veerapandian, V., Kolatkar, P., Cojocaru, V., Pervushin, K., and Jauch, R. (2015). DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Research*, 43(3):1513–1528.

Nie, Z., Guo, C., Das, S. K., Chow, C. C., Batchelor, E., Simons, S. S., and Levens, D. (2020). Dissecting transcriptional amplification by MYC. *eLife*, 9:1–32.

O'Connor, L., Gilmour, J., and Bonifer, C. (2016). The role of the ubiquitously expressed transcription factor Sp1 in tissue-specific transcriptional regulation and in disease. *Yale Journal of Biology and Medicine*, 89(4):513–525.

Oehler, S., Eismann, E., Krämer, H., and Müller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *The EMBO Journal*, 9(4):973–979.

Okawa, S. and Del Sol, A. (2019). A general computational approach to predicting synergistic transcriptional cores that determine cell subpopulation identities. *Nucleic Acids Research*, 47(7):3333–3343.

Ouyang, Z., Zhou, Q., and Wong, W. H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21521–21526.

Oven, I., Brdičková, N., Kohoutek, J., Vaupotič, T., Narat, M., and Peterlin, B. M. (2007). AIRE Recruits P-TEFb for Transcriptional Elongation of Target Genes in Medullary Thymic Epithelial Cells. *Molecular and Cellular Biology*, 27(24):8815–8823.

Paakinaho, V., Presman, D. M., Ball, D. A., Johnson, T. A., Schiltz, R. L., Levitt, P., Mazza, D., Morisaki, T., Karpova, T. S., and Hager, G. L. (2017). Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nature Communications*, 8(May):1–14.

Pan, Y. and Nussinov, R. (2011). The Role of Response Elements Organization in Transcription Factor Selectivity: The IFN-β Enhanceosome Example. *PLoS Computational Biology*, 7(6):e1002077.

Park, J., Estrada, J., Johnson, G., Vincent, B. J., Ricci-Tam, C., Bragdon, M. D., Shulgina, Y., Cha, A., Wunderlich, Z., Gunawardena, J., and DePace, A. H. (2019a). Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. *eLife*, 8:e41266.

Park, M., Patel, N., Keung, A. J., and Khalil, A. S. (2019b). Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell*, 176(1-2):227–238.e20.

Pavletich, N. and Pabo, C. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007):809–817.

Phillips, R. (2015). Napoleon Is in Equilibrium. *Annual Review of Condensed Matter Physics*, 6(1):85–111.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature*, 335(6192):683–689.

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., and Young, R. A. (2010). C-Myc regulates transcriptional pause release. *Cell*, 141(3):432–445.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, 4(10):1707–1719.

Ramos, A. I. and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632):20130018.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development*, 43:73–81.

Rybakova, K. N., Bruggeman, F. J., Tomaszewska, A., Moné, M. J., Carlberg, C., and Westerhoff, H. V. (2015). Multiplex Eukaryotic Transcription (In)activation: Timing, Bursting and Cycling of a Ratchet Clock Mechanism. *PLoS Computational Biology*, 11(4):1–24.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682.

Scholes, C., DePace, A. H., and Sánchez, Á. (2017). Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. *Cell Systems*, 4(1):97–108.e9.

Schröder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D. A., Kaehlcke, K., Cho, S., Pollard, K. S., Capra, J. A., Schnölzer, M., Cole, P. A., Geyer, M., Bruneau, B. G., Adelman, K., and Ott, M. (2013). Acetylation of RNA Polymerase II Regulates Growth-Factor-Induced Gene Transcription in Mammalian Cells. *Molecular Cell*, 52(3):314–324.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, 451(7178):535–540.

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Research*, 16(1):45–58.

Singh, G., Mullany, S., Moorthy, S. D., Zhang, R., Mehdi, T., Shchuka, V. M., Tian, R., Moses, A. M., and Mitchell, J. A. (2020). A flexible repertoire of transcription factor binding sites and diversity threshold determines enhancer activity in embryonic stem cells. *bioRxiv*.

Smith, R. P., Taher, L., Patwardhan, R. P., Kim, M. J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature Genetics*, 45(9):1021–1028.

Spitz, F. and Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9):613–626.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature*, 528(7580):147–151.

Strasen, J., Sarma, U., Jentsch, M., Bohn, S., Sheng, C., Horbelt, D., Knaus, P., Legewie, S., and Loewer, A. (2018). Cell-specific responses to the cytokine TGF β are determined by variability in protein levels . *Molecular Systems Biology*, 14(1):1–17.

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science*, 332(6028):472–474.

Tellier, M., Maudlin, I., and Murphy, S. (2020). Transcription and splicing: A two-way street. *WIREs RNA*, 11(5):1–25.

Tunnacliffe, E. and Chubb, J. R. (2020). What Is a Transcriptional Burst? *Trends in Genetics*, 36(4):288–297.

Vandel, J., Cassan, O., Lèbre, S., Lecellier, C. H., and Bréhélin, L. (2019). Probing transcription factor combinatorics in different promoter classes and in enhancers. *BMC Genomics*, 20(1):1–19.

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T., Fernandez, N., Ballester, B., Andrau, J. C., and Spicuglia, S. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nature Communications*, 6.

Vashee, S., Willie, J., and Kodadek, T. (1998). Synergistic activation of transcription by physiologically unrelated transcription factors through cooperative DNA-Binding. *Biochemical and Biophysical Research Communications*, 247(2):530–535.

Vasile, E., Hnisz, D., Klein, I. A., Young, R. A., Manteiga, J. C., Malik, S., Lee, T. I., Abraham, B. J., Schuijers, J., Cisse, I. I., Roeder, R. G., Dall'Agnese, A., Hannett, N. M., Day, D. S., Boija, A., Coffey, E. L., Shrinivas, K., Li, C. H., Sabari, B. R., Sharp, P. A., Zamudio, A. V., Chakraborty, A. K., and Guo, Y. E. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 361(6400):eaar3958.

Veitia, R. A. (2003). A sigmoidal transcriptional response: Cooperativity, synergy and dosage effects. *Biological Reviews of the Cambridge Philosophical Society*, 78(1):149–170.

Wissink, E. M., Vihervaara, A., Tippens, N. D., and Lis, J. T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nature Reviews Genetics*, 20(12):705–723.

Wong, F. and Gunawardena, J. (2020). Gene Regulation in and out of Equilibrium. *Annual Review of Biophysics*, 49(1):199–226.

Wong, V. C., Mathew, S., Ramji, R., Gaudet, S., and Miller-Jensen, K. (2019). Fold-Change Detection of NF-κB at Target Genes with Different Transcript Outputs. *Biophysical Journal*, 116(4):709–724.

Wunderlich, Z. and Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics*, 25(10):434–440.

Yang, Z., Yik, J. H., Chen, R., He, N., Moon, K. J., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*, 19(4):535–545.