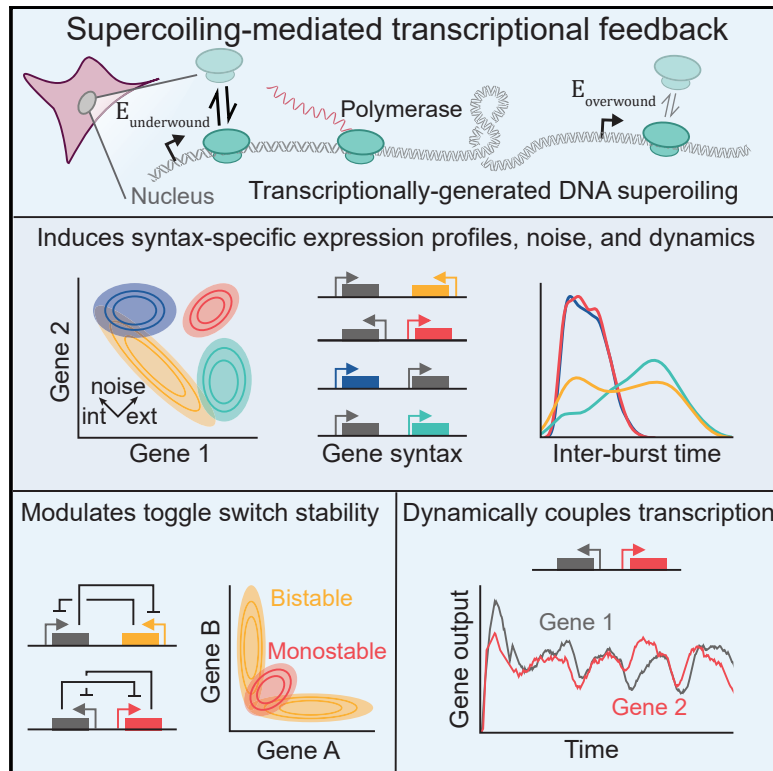


# Supercoiling-mediated feedback rapidly couples and tunes transcription

## Graphical abstract



## Authors

Christopher P. Johnstone,  
Kate E. Galloway

## Correspondence

katiegall@mit.edu

## In brief

Supercoiling-mediated feedback couples the transcription of proximal genes. Here, Johnstone and Galloway provide a framework for integrating biochemical gene regulation with the biophysical effects of DNA supercoiling. This unified model provides design principles for improving the performance of gene networks, developing novel regulatory functions, and accessing previously inaccessible regulatory dynamics.

## Highlights

- Supercoiling dynamics confer rapid, tunable coupling between adjacent genes
- Syntax—the relative order and orientation of genes—alters expression levels
- Supercoiling-dependent feedback tunes transcriptional variance and bursting
- Supercoiling coordinates the dynamics of transcriptional networks and gene circuits



## Article

# Supercoiling-mediated feedback rapidly couples and tunes transcription

Christopher P. Johnstone<sup>1</sup> and Kate E. Galloway<sup>1,2,\*</sup><sup>1</sup>Department of Chemical Engineering, MIT, 25 Ames St., Cambridge, MA 02139, USA<sup>2</sup>Lead contact\*Correspondence: [katiegal@mit.edu](mailto:katiegal@mit.edu)<https://doi.org/10.1016/j.celrep.2022.111492>**SUMMARY**

Transcription induces a wave of DNA supercoiling, altering the binding affinity of RNA polymerases and reshaping the biochemical landscape of gene regulation. As supercoiling rapidly diffuses, transcription dynamically reshapes the regulation of proximal genes, forming a complex feedback loop. However, a theoretical framework is needed to integrate biophysical regulation with biochemical transcriptional regulation. To investigate the role of supercoiling-mediated feedback within multi-gene systems, we model transcriptional regulation under the influence of supercoiling-mediated polymerase dynamics, allowing us to identify patterns of expression that result from physical inter-gene coupling. We find that gene syntax—the relative ordering and orientation of genes—defines the expression profiles, variance, burst dynamics, and inter-gene correlation of two-gene systems. Furthermore, supercoiling can enhance or weaken biochemical regulation. Our results suggest that supercoiling couples behavior between neighboring genes, providing a regulatory mechanism that tunes transcriptional variance in engineered gene networks and explains the behavior of co-localized native circuits.

**INTRODUCTION**

Cells coordinate complex behaviors through precise spatiotemporal control of gene expression. To rapidly advance gene- and cell-based therapies, synthetic biology aims to harness the power of native biology by constructing synthetic gene regulatory networks capable of dynamically prescribing cellular processes, states, and identities (Chen et al., 2012; Beitz et al., 2022; Purnick and Weiss, 2009; Elowitz and Lim, 2010). Synthetic networks process diverse inputs into complex logical and temporal responses (Weinberg et al., 2017; Xie et al., 2011; Tabor et al., 2009). From oscillators to pulse generators, synthetic circuits can precisely coordinate dynamic patterns of gene expression across populations of cells to control cell fate (Gardner et al., 2000; Elowitz and Leibler, 2000; Stricker et al., 2008; Danino et al., 2010; Ma et al., 2022; Park et al., 2019; Bashor et al., 2008; Galloway et al., 2013). However, rational *de novo* design of synthetic circuits remains challenging. Despite extensive biomolecular modeling, integration of single genetic elements into systems often leads to emergent behaviors, requiring iterative design-build-test cycles to achieve the desired performance (Jones et al., 2020; Frei et al., 2020; Qian et al., 2017). Compounding the challenge, transcription exhibits significant extrinsic and intrinsic noise (To and Maheshri, 2010; Zopf et al., 2013; Desai et al., 2021). In particular, the stochastic nature of transcription makes coordinating expression across multiple genetic elements challenging (Rodriguez and Ren, 2019; Rodriguez and Larson, 2020; Quarton et al., 2020). Spatial variation

in the nucleus and biochemical dynamics in condensates may contribute to bursting but provide limited parameters for tuning transcriptional noise (Henninger et al., 2020; Guo and Yang, 2019). Alternatively, mechanical sources of gene regulation offer one potential mechanism by which to understand and harness transcriptional noise to improve the predictable design of gene circuits (Johnstone and Galloway, 2021; Ancona et al., 2019; Kim et al., 2019; El Houdaigui et al., 2019; Meyer and Beslon, 2014).

The mechanical forces of DNA supercoiling powerfully shape transcriptional variance (Desai et al., 2021; Chong et al., 2014). In the process of transcription, RNA polymerases induce a leading wave of positive DNA supercoiling (Wu et al., 1988; Liu and Wang, 1987), reshaping the local structure of chromatin (Achar et al., 2020; Teves and Henikoff, 2014; Naughton et al., 2013; Guo et al., 2021). At the kilobase scale, chromatin structure correlates with gene regulation (Hsieh-Tsung-Han et al., 2020; RowleyNichols et al., 2017; Krietenstein et al., 2020). In yeast and human cells, transcription-induced supercoiling demarks bounds of gene activity (Achar and Jagadheesh, 2020; Naughton et al., 2013; Kouzine et al., 2013). In particular, transcriptional activity dictates the strength of contact domains, indicating a role for transcription in forming and maintaining interactions at the kilobase scale (Rowley and Corces, 2018; RowleyNichols et al., 2017). Together these data suggest that the process of transcription drives formation of supercoiling-linked, kilobase-scale structures that feed back into transcriptional regulation of gene expression. As supercoiling rapidly diffuses across long



distances (Loenhout and Dekker, 2012), transcriptional activity at one site may affect the overall activity and dynamics of transcription of proximal genes (Sevier and Levine, 2018; Sevier and Hormoz, 2022; Tripathi et al., 2021). Understanding how supercoiling induces coupling between neighboring genes provides the opportunity to improve the predictable design of transgenic systems from simple reporters to sophisticated dynamic circuits.

Here we develop a model of transcriptional regulation that integrates DNA supercoiling to examine how the orientation and placement of neighboring genes affects expression. Extending from a model of supercoiling-dependent polymerase motion (Sevier and Levine, 2018), our model includes the effects of supercoiling on polymerase binding and initiation. Specifically, we model RNA polymerase binding and initiation as a function of DNA supercoiling, such that underwound DNA favors RNA polymerase binding and overwound DNA limits binding. To extract experimentally testable predictions, we apply our model to simple two-gene systems that include a constitutive reporter and an inducible gene. Using these two-gene systems, we find that DNA supercoiling strongly influences the profile of gene expression and that influence is defined by syntax—the relative orientation and position of genetic elements—and the enclosing boundary conditions. In addition to regulating the output of simulated genes, supercoiling-dependent feedback tunes the size and frequency of transcriptional bursts. To investigate how these tunable parameters may affect synthetic gene circuits, we applied our model to a canonical gene circuit, a synthetic toggle switch constructed with different syntaxes. We find that circuit syntax affects the stability of states and sets biochemical parameters required for bistability, including repressor cooperativity and RNA stability. Finally, we explored how DNA supercoiling might support transcriptional coordination within the native genome of zebrafish (*Danio rerio*) to enable somite segmentation. We find that DNA supercoiling acts as a mechanism for coordinating expression between divergently expressed genes in the segmentation network. Supercoiling-dependent feedback supports tight regulation of these proximal segmentation genes, providing a molecular mechanism for the precise coordination of gene expression observed during somite formation (Zinani et al., 2021). Thus, supercoiling-mediated feedback represents a testable regulatory mechanism that can both explain native behaviors and guide synthetic designs.

## RESULTS

Simulating the behavior of native and synthetic circuits under the influence of transcription-induced feedback requires a model that integrates explicitly modeled RNA polymerase (RNAP) motion and RNA- and protein-mediated feedback mechanisms. Our method combines three modeling levels: an ordinary differential equation system that simulates the continuous progression of polymerases loaded onto DNA, a core stochastic system that models supercoiling-dependent polymerase initiation, and a user-specified stochastic layer that allows for simulation of other modes of transcriptional regulation such as the promoter-repressive and -activating interactions that are often included

in synthetic circuits (see STAR Methods for details on model development).

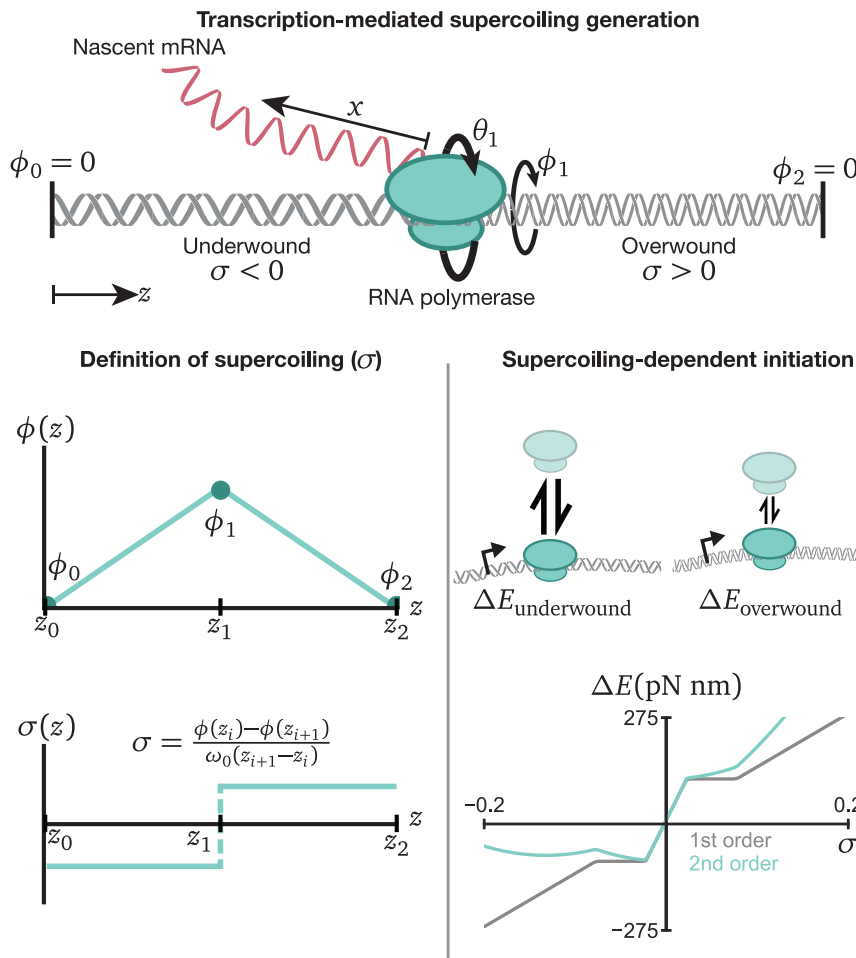
As RNA polymerases move along chromatin, positive supercoiling is generated ahead of the polymerase and negative supercoiling is generated behind. Supercoiling generation occurs because the drag of the nascent mRNA rotating with the polymerase is balanced by torque arising from the over- and under-winding of the DNA (Figure 1). On the length scale of tens of kilobases, supercoiling diffusion is negligible, causing the supercoiling density,  $\sigma$ , to take on a constant value between polymerases. Here, we model supercoiling as affecting gene expression in two ways: excessive supercoiling can both stall polymerases due to excess torque and enhance or suppress polymerase binding at promoter regions due to supercoiling-dependent initiation (Figure 1; STAR Methods).

### Gene syntax and boundary conditions define DNA supercoiling dynamics, expression profiles, and noise

In order to characterize the behavior of supercoiling-mediated feedback, we simulated a series of two-gene systems. These experimentally accessible circuits allow us to test and understand the core design considerations—syntax, relevant boundary conditions, and other experimentally tunable parameters—within a well-defined and controlled system. Our two-gene systems consist of a reporter gene that is constitutively active and an adjacent, inducible gene placed in either a tandem orientation with the reporter upstream, tandem orientation with the reporter downstream, convergent orientation, or divergent orientation (Figure 2A).

Varying syntax, we examined how boundary conditions affect the expression profiles of the reporter and inducible genes. The type of boundary condition determines how transcriptionally generated supercoiling propagates to adjacent genes. Experimentally, plasmid constructs and genomically integrated cassettes allow for interrogation of circular and linear boundary conditions, respectively. Plotting reporter expression as a function of adjacent gene induction, we observed that circular boundary conditions show a monotonic increase in reporter output that scales by syntax (Figure 2B). With circular boundary conditions, both the convergent and divergent syntaxes show a large enhancement of expression relative to the uninduced case, whereas the tandem cases show only a slight enhancement. This phenomenon occurs because, on a circle, the convergent and divergent syntaxes differ only in the relative lengths of the intergenic and inter-promoter regions; as such, negative supercoiling density will accumulate in the inter-promoter region, enhancing expression in the convergent and divergent syntaxes. In contrast, linear boundary conditions show diverse behaviors. The upstream-tandem and divergent syntaxes maintain high expression levels, while the convergent and downstream-tandem syntaxes show decreasing reporter expression with increasing induction of the neighboring gene.

Transcriptional noise substantially contributes to variance of gene expression (Quarton et al., 2020), often confounding circuit designs. Thus, designing circuits to respond to or suppress noise may improve circuit performance. To investigate how syntax affects expression profiles and noise, we simulated the ensemble behavior for circular and linear two-gene systems at



**Figure 1. Four key variables define the location of each polymerase: its linear distance  $z$  along the genome, the length of the nascent mRNA transcript  $x$ , the rotation of the polymerase  $\theta$ , and the local DNA excess twist  $\phi$**

The relaxed DNA twist frequency  $\omega_0$  has value  $\omega_0 = 1.85$  radians/nm. The tradeoff between RNAP rotation and DNA rotation generates supercoiling upstream and downstream, with the drag generated by the nascent mRNA primarily balancing the torque caused by generated supercoils. In the limit of fast supercoiling relaxation relative to polymerase motion, the supercoiling density is constant in the region between polymerases and can be calculated from the slope of the linearly interpolated  $\phi(z)$  graph. Using an energy model responsive to local supercoiling, we can derive supercoiling-dependent initiation terms to model differential polymerase loading rates.

To observe supercoiling across the two-gene systems, we averaged the supercoiling density across the simulated ensemble. To examine how induction of the adjacent gene changes supercoiling density, we compared the profiles for when the adjacent gene is uninduced (0-fold induction) and induced (1-fold induction). As expected, the uninduced cases uniformly show that positive supercoiling accumulates downstream of the constitutively active reporter gene while negative supercoiling accumulates upstream (Figure 2E; for other induction levels, see Figure S1A). Differences in basal expression

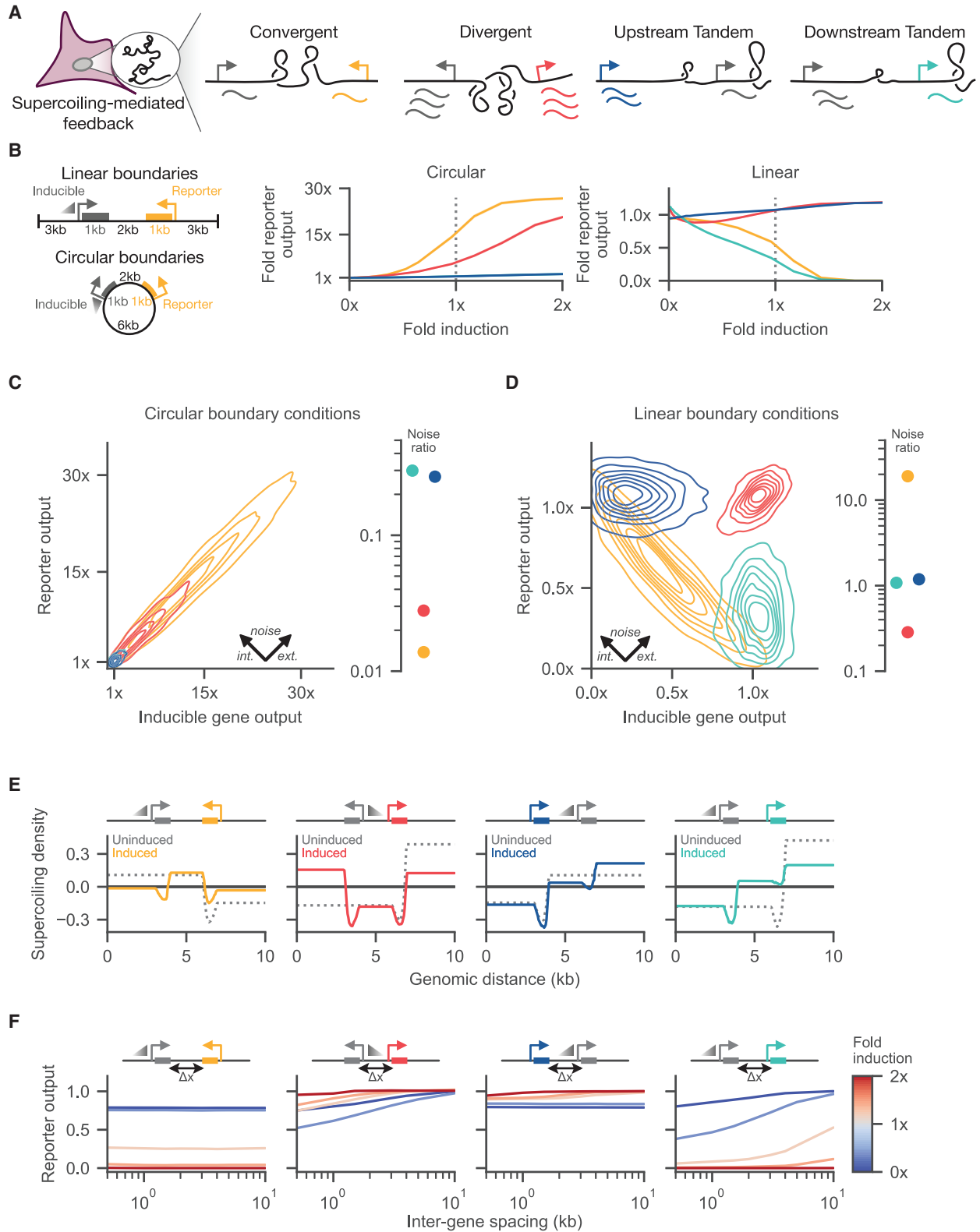
of the reporter correspond to differences in supercoiling profile across the different syntaxes. While each uninduced case generates an equal magnitude of positive and negative supercoiling upstream and downstream, the distance from the reporter gene to the boundaries differs depending on syntax, leading to the shorter segment having the higher magnitude of supercoiling density. Upon induction of the adjacent gene, supercoiling accumulates within the intergenic regions in a syntax-specific manner.

equal transcriptional induction (gray dotted line in Figure 2B). To examine different forms of noise, we decompose the population variance into an extrinsic noise component that describes how “all” genes co-vary within a cell and an intrinsic noise component that describes the inter-gene variance within a cell. Then, we define the noise ratio as the intrinsic noise divided by the extrinsic noise. For the circular-boundary condition simulations, we found that all four ensembles are dominated by extrinsic noise (Figure 2C). In contrast, while the tandem syntaxes with linear boundary conditions show approximately equal intrinsic and extrinsic noise, the linear convergent and divergent populations showed diverging shifts in the variance distribution (Figure 2D). Despite similar levels of extrinsic noise, the divergent syntax minimizes intrinsic variation between the two genes while the convergent syntax maximizes inter-gene variation. Moving forward, we used the linear set of boundary conditions to analyze system behaviors as we observe the richest set of behaviors under these boundary conditions.

Supercoiling density determines both supercoiling-dependent initiation and polymerase stalling. Changing the inter-gene spacing directly tunes the transcriptional activity required to reach a specified supercoiling density and thus reach different expression profiles. To understand the spacing-driven deviations in reporter behavior, we simulated our linear two-gene circuits with different inter-gene spacings from 500 bp to 10 kb and plotted the reporter output of each circuit (Figure 2F). We found that the convergent and tandem-upstream syntaxes show minimal spacing-dependent changes in reporter output; the reporter output in these two syntaxes is mostly dependent on the level of adjacent induction. In contrast, the reporter output in the divergent and tandem-downstream syntaxes depends on inter-gene spacing. With weak adjacent induction in the divergent

To understand the mechanisms that support syntax-specific expression profiles (Figure 2D), we examined the ensemble supercoiling density of our two-gene systems. Putatively, differences in supercoiling across the two-gene systems give rise to differences in RNAP initiation and thus affect gene expression.

Supercoiling density determines both supercoiling-dependent initiation and polymerase stalling. Changing the inter-gene spacing directly tunes the transcriptional activity required to reach a specified supercoiling density and thus reach different expression profiles. To understand the spacing-driven deviations in reporter behavior, we simulated our linear two-gene circuits with different inter-gene spacings from 500 bp to 10 kb and plotted the reporter output of each circuit (Figure 2F). We found that the convergent and tandem-upstream syntaxes show minimal spacing-dependent changes in reporter output; the reporter output in these two syntaxes is mostly dependent on the level of adjacent induction. In contrast, the reporter output in the divergent and tandem-downstream syntaxes depends on inter-gene spacing. With weak adjacent induction in the divergent



(legend on next page)

case (blue curves), decreasing the inter-gene spacing *decreases* reporter output. We attribute this decrease to the effect of polymerase stalling; a small inter-gene spacing increases the local supercoiling density which can lead to polymerase stalling. At high adjacent induction (red curves), strong supercoiling-driven polymerase initiation overcomes the stalling effect, leading to a relatively distance-insensitive divergent reporter output. Finally, in the tandem-downstream syntax, in addition to the strong suppression of reporter output at high adjacent induction, a short inter-gene spacing further reduces reporter output by increasing the supercoiling density. Thus, depending on syntax, supercoiling-dependent feedback can be tuned by changing inter-gene spacing.

### DNA supercoiling dynamics confer rapid, tunable coupling between adjacent genes

To explore the impact of DNA supercoiling beyond the ensemble steady-state behavior, we investigated the dynamic behavior of our system. To understand the initiation dynamics of the system, we initialized two-gene constructs with only the constitutive reporter gene active. After a settling period (2.8 h), we induced transcription of the adjacent gene (Figure 3A). We observe that expression dynamics vary extensively by syntax. In the convergent syntax, we observe anti-correlated dynamics, with “either-or” promoter activity (Video S1). In contrast, the divergent syntax supports high levels of expression from both genes (Video S2). When comparing the two tandem orientations, we observe strong biasing of expression toward the upstream gene with stochastic bursts of expression from the downstream gene (Videos S3 and S4). While this upstream dominance does not completely disable the downstream gene, activation of the upstream gene reduces the average expression of the downstream reporter.

To ensure that this observed behavior was not an artifact of enabling adjacent induction later in the simulation, we need to verify that the two-gene ensembles are ergodic (not kinetically trapped or otherwise dependent on initial condition). To start, we examined the ensemble expression distribution at the simulation endpoint (Figure S1B), and found that we recapitulate the population behavior observed in Figure 2D. As an additional check, we plotted the ensemble-averaged supercoiling density both before (at 2.3 h) and after (at 11 h) adjacent gene induction (Figure 3B).

If the system is ergodic, then the ensemble behavior observed before and after the adjacent gene is turned on (snapshots of one population separated in time) should match ensemble behavior of simulations where the adjacent gene is time-independently active or inactive (snapshots of separate populations). Figure 3B does replicate the behavior observed in Figure 2E, indicating that the two-gene circuit populations are ergodic.

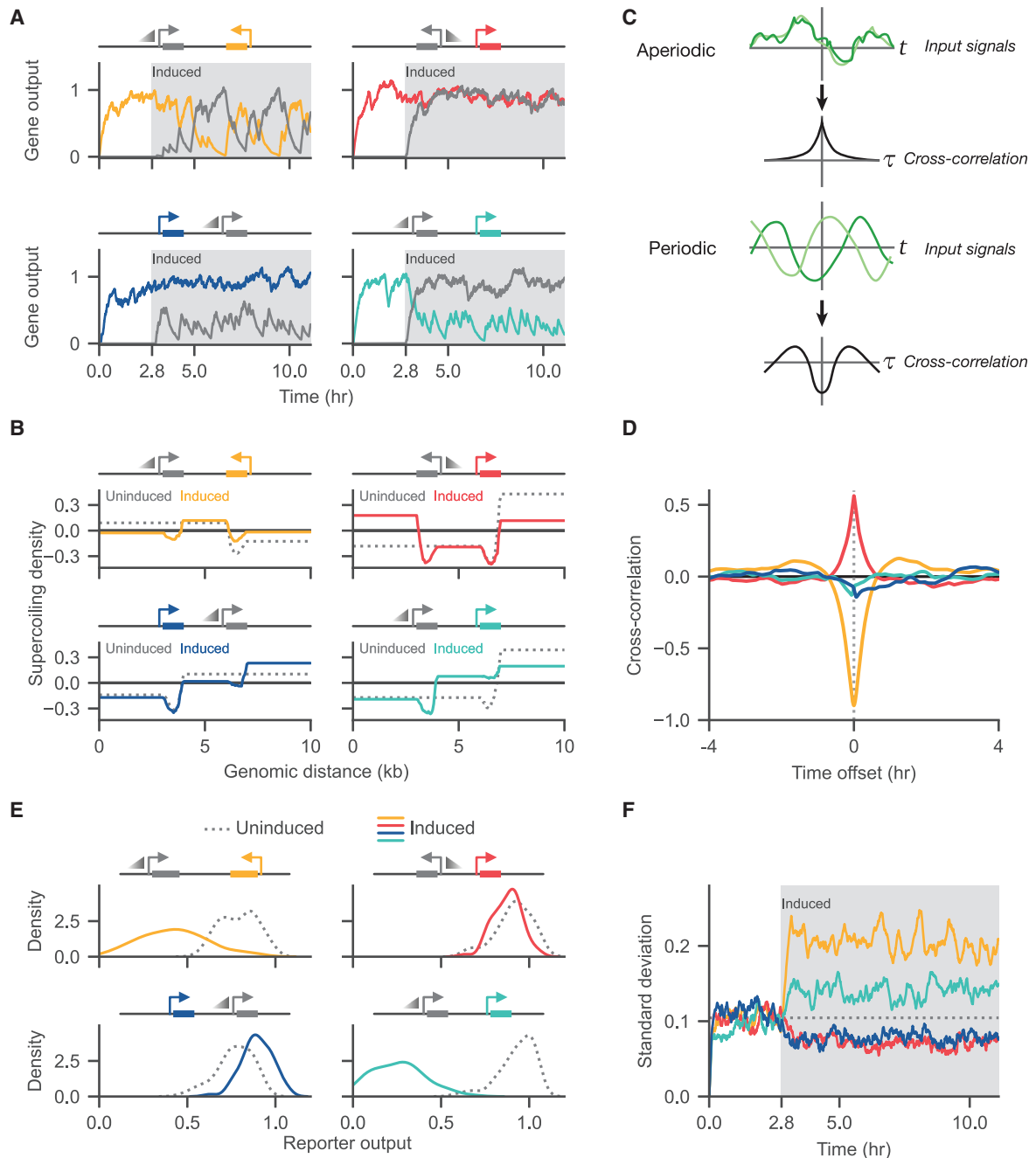
To expand our understanding of the emergent supercoiling-dependent dynamics, we examined the temporal correlation of transcription between both genes. To quantify the correlation and extract temporal patterns, we computed the normalized cross-correlation between the gene outputs following induction. The normalized cross-correlation of two signals is itself a function of a time offset; the normalized cross-correlation at some offset time  $\tau$  can be thought of as the Pearson correlation coefficient between the two signals where one has been shifted by  $\tau$  (Figure 3C; see “Computing cross-correlation” in STAR Methods for more detail). In particular, periodic but out-of-phase signals appear as strong negative and positive peaks on a cross-correlation plot. The time offset of peaks corresponds to the phase offset between the signals.

The four syntaxes show starkly different cross-correlation behaviors (Figure 3D). The convergent system shows a large anti-correlation at zero time offset with positive correlation peaks at offsets around  $\pm 2$  h. This combination suggests a periodic but out-of-phase behavior between the two genes with a period of around 2 h, confirming that our ensemble behaves similarly to the example simulation in Figure 3A. In contrast, the divergent syntax shows a strong positive peak at zero time offset, showing strong aperiodic, but correlated, behavior. The tandem syntax shows weak correlation between genes.

We then examined how adjacent induction affects the reporter output distributions. Visualizing the entire ensemble reporter output distribution per condition in Figure 3E at equal induction, we observe that adjacent induction changes both the mean and variance of the reporter output. In particular, the widths of the distributions change before and after induction, suggesting a change in the noise profile. At high adjacent induction, we observe that the upstream-tandem and divergent cases show enhanced transcription, with the downstream-tandem and convergent cases effectively turning off (Figure S1C).

### Figure 2. Supercoiling-dependent feedback induces syntax-specific expression profiles

- (A) Two-gene circuits serve as a testbed for investigating supercoiling-mediated feedback. All four syntaxes include a reporter gene (colored) and an inducible gene (gray).
- (B) Two classes of boundary conditions are simulated. Linear boundary conditions are simulated with adjacent “walls” that prevent supercoiling propagation, whereas circular boundary conditions allow supercoiling generated at one gene to freely affect other genes in either direction around the circle. At right, mean reporter expression is plotted as a function of the level of induction of the adjacent gene for circular and linear boundary conditions. Reporter output is normalized to the uninduced expression case by dividing mRNA counts by a constant value per boundary condition case (10 mRNAs for the circular case and 250 mRNAs for the linear case).
- (C) Gene expression distributions of simulations with circular boundary conditions are shown, where the adjacent induced gene is equally induced relative to the reporter gene. For each of the four syntaxes, the expression variance can be decomposed into intrinsic and extrinsic noise components; the ratio of intrinsic to extrinsic noise is shown on the right. Reporter output is normalized by dividing mRNA counts by a constant value (10 mRNAs).
- (D) Gene expression distributions and the intrinsic to extrinsic noise ratios are shown for linear boundary conditions. Reporter output is normalized by dividing counts by a constant value (250 mRNAs).
- (E) The mean supercoiling density of linear constructs are shown as a function of induction of the inducible gene. Induction (colored line) displays syntax-specific behavior compared with the uninduced case (dashed line).
- (F) Reporter output is shown as a function of inter-gene spacing ( $\Delta x$ ) at five different induction levels. Reporter output is normalized by dividing mRNA counts by a constant value (250 mRNAs). See also Figure S1.



**Figure 3. Supercoiling-dependent feedback induces dynamic coupling and altered variance between genes**

(A) mRNA counts over time from example individual simulations for the four syntaxes are shown. Simulations are initialized with only the reporter gene (colored) active, with the adjacent gene (gray) enabled with equal basal expression after 10,000 s (2.8 h).

(B) The average ensemble supercoiling density is shown both before and after adjacent gene induction.

(C) The cross-correlation of two signals  $f(t)$ ,  $g(t)$  at a time offset  $\tau$  can be calculated by “sliding” one mean-centered signal relative to the other mean-centered signal and integrating the product of the resulting signals.

(D) The cross-correlation between the two genes is shown for the equal-induction case across the four syntaxes. The convergent and divergent syntaxes showed the strongest cross-correlation, with the convergent case showing periodic behavior and the divergent case showing strong correlated expression.

(E) Distributions of the reporter output before (dotted) and after (solid) induction of the adjacent gene show changes in both the mean and standard deviation due to adjacent expression.

(F) Ensemble noise behavior for the four simulated syntaxes is shown by plotting the standard deviation of the reporter gene across the ensemble of simulations as a function of time. See also [Figure S1](#) and [Videos S1, S2, S3, and S4](#).

As noise affects the properties of native and synthetic gene networks, we quantified the width of these distributions by plotting the standard deviation of the ensemble reporter output as a function of time (Figure 3F). Prior to induction of the second gene, all four systems display similar standard deviations. Syntax differences emerge upon induction of the second gene. We found that syntax strongly modulates the noise behavior of the reporter. In particular, the downstream-tandem and convergent syntaxes show a strong *increase* in noise levels while the upstream-tandem and divergent syntaxes show a small *decrease* in noise levels. These changes in noise may result from differences in transcriptional burst sizes and frequencies.

### Burst dynamics vary in different models of DNA supercoiling

Transcription occurs in bursts of activity, and native and synthetic mechanisms can modify burst dynamics (Desai et al., 2021; Chong et al., 2014; Popp et al., 2021). In our base model, transcriptional bursting arises from the stochastic addition of polymerases. We hypothesized that burst dynamics may explain the distribution dynamics observed in Figures 3E and 3F. While our base model provides a species-agnostic approach to investigate supercoiling-mediated feedback, the mechanical regulation of eukaryotic chromatin may introduce additional complexity and tunability to bursting dynamics. For example, nucleosomes can serve as a reservoir of negative supercoiling by stabilizing the wrapped double-loop of DNA (Le et al., 2019). Furthermore, at hypernegative or hyperpositive supercoiling densities, local chromatin structure may be disrupted causing structures such as R-loops and G-quadruplexes to preferentially form. These structures can block polymerase binding and motion (Stolz et al., 2019). Even when including our second-order correction term as detailed in Equation 7, our base model predicts favorable supercoiling-dependent polymerase initiation at the hypernegative value of  $\sigma = -0.1$  (Figure 4A). To address these complexities, we implemented two alternate models that extend our torque-response and polymerase-initiation energy equations.

To simulate the effects of structures forming at hypernegative or hyperpositive supercoiling densities, we penalize polymerase initiation by adding a near-infinite positive energy penalty to polymerase initiation at hypernegative and hyperpositive supercoiling densities (Figure 4A) roughly matching the density at which structures such as R-loops form ( $\sigma < -0.06$ ,  $\sigma > 0.125$ ; Stolz et al., 2019). Putatively, these barriers enable us to simulate the formation of stable structures within the DNA that strongly penalize RNAP binding while remaining agnostic to any specific molecular structure. We find that penalization of extreme supercoiling globally reduces supercoiling density (Figure S2A) but maintains the qualitative syntax-dependent behavior of the steady-state, linear systems. However, syntax-specific differences are mostly eliminated for circular templates (Figures S2B and S2C).

To model nucleosomes, we updated our torque-response function to phenomenologically match *in vitro* experimental measurements (Le et al., 2019). Nucleosomes putatively “buffer” the effects of positive supercoiling by unbinding and releasing stored negative supercoiling. We accounted for this buffering

by extending a zero-torque plateau within the region  $0 \leq \sigma \leq 0.031$  (Figure 4A). We found that the nucleosome buffering does not significantly affect the steady-state expression profile or supercoiling density observed in Figure 2 (Figures S3A–S3C).

With these three models, we simulated the four linear syntaxes and recorded polymerase initiation events per gene. We then examined the distribution of both burst size, which we define as the number of polymerase initiation events during a burst, and inter-burst time, the amount of time separating two consecutive bursts for the reporter gene (Figure 4B; see “Burst threshold choice” in STAR Methods for more detail). Upon induction of the adjacent gene, we find burst dynamics differ by syntax (Figure 4C). In both the base model and the nucleosome-buffering model, the downstream-tandem and convergent syntaxes show a reduced burst size compared with the divergent and upstream-tandem syntaxes. We attribute this reduction in burst size to accumulated positive supercoiling at the site of the reporter gene’s promoter. In contrast, induction of the adjacent gene in the divergent syntax increases the burst size of the reporter gene, putatively due to enhanced loading of polymerases facilitated by accumulated negative supercoiling. With penalization of hypernegative supercoiling density, the divergent syntax shows a decrease in burst size, potentially due to a shift in the tradeoff between polymerase stalling and supercoiling-dependent initiation.

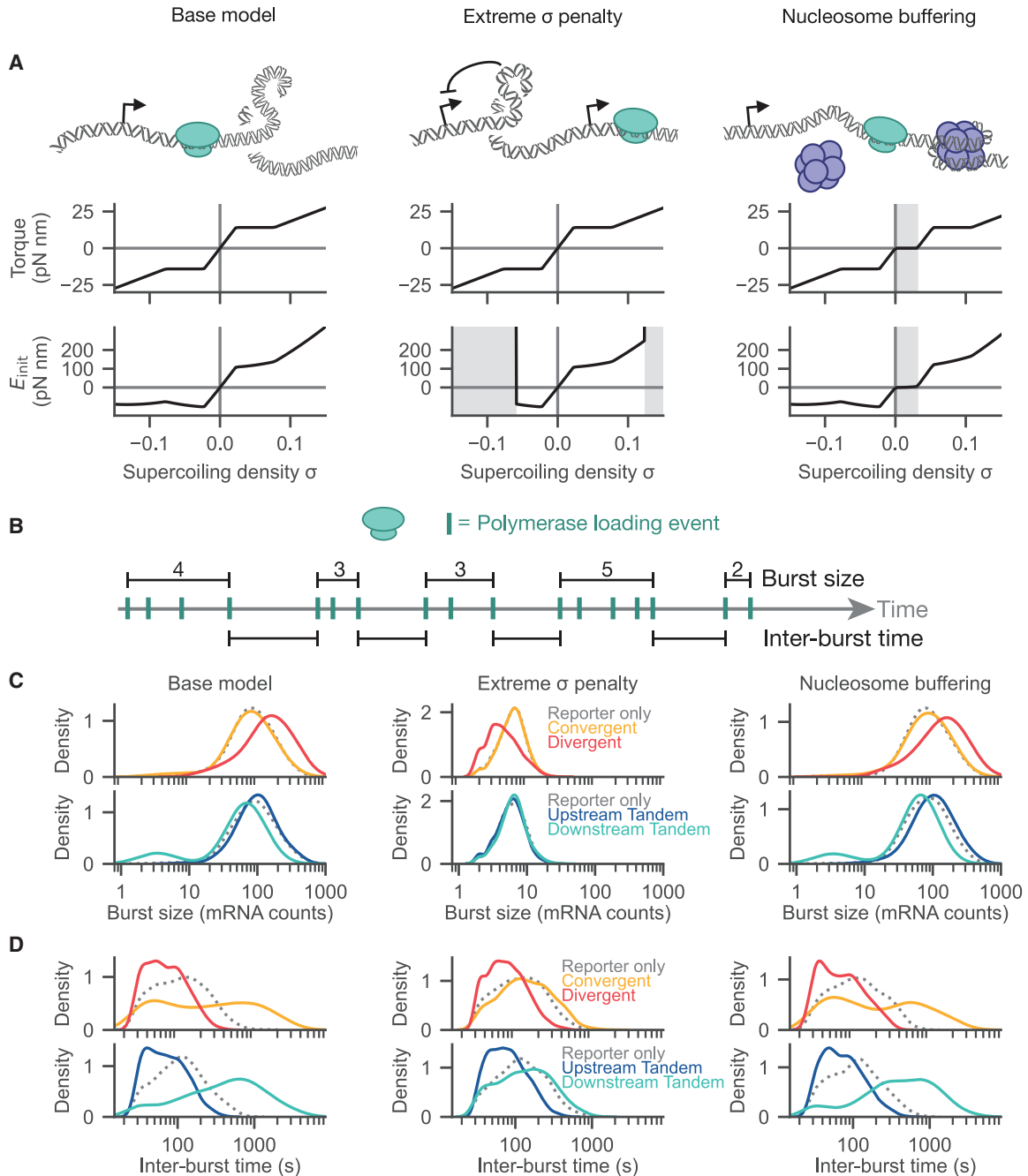
Examining the inter-burst time distributions, induction of the neighboring gene shifts the upstream-tandem and divergent syntaxes to shorter inter-burst times (Figure 4D). The downstream-tandem and convergent syntaxes shift to longer inter-burst times upon induction of the adjacent gene. Interestingly, this qualitative observation holds independently of the choice of torque-energy model, suggesting that changes in inter-burst time—and its inverse, burst frequency—is a syntax-dependent effect that is largely insensitive to additional chromatin perturbations.

From these observations, we can now understand the population-level behaviors observed in Figures 3E and 3F. The increase of noise and reduction of mean expression value in the downstream-tandem and convergent cases occurs concomitant with an increase in inter-burst time (Figure 4D). This is expected; if bursts occur rarely, stochastic fluctuations will have an outsized effect on each individual simulation, enhancing population variance. In contrast, the decrease of noise in the upstream-tandem and divergent syntaxes is matched by a decrease in the inter-burst time (Figure 4D). Because bursts happen more frequently, the ensemble reporter output is more stable as we approach the limit of constant transcription. Taken as a whole, syntax provides a powerful design parameter for inducing and tuning time-dependent behaviors between genes and shaping output gene distributions.

### Optimizing toggle-switch performance and stability through circuit syntax

From oscillators to pulse generators, synthetic circuits aim to precisely coordinate dynamic patterns of gene expression. However, emergent dynamics, mediated through DNA supercoiling, may support or impede the performance of dynamic circuits. To examine how supercoiling-mediated feedback influences a





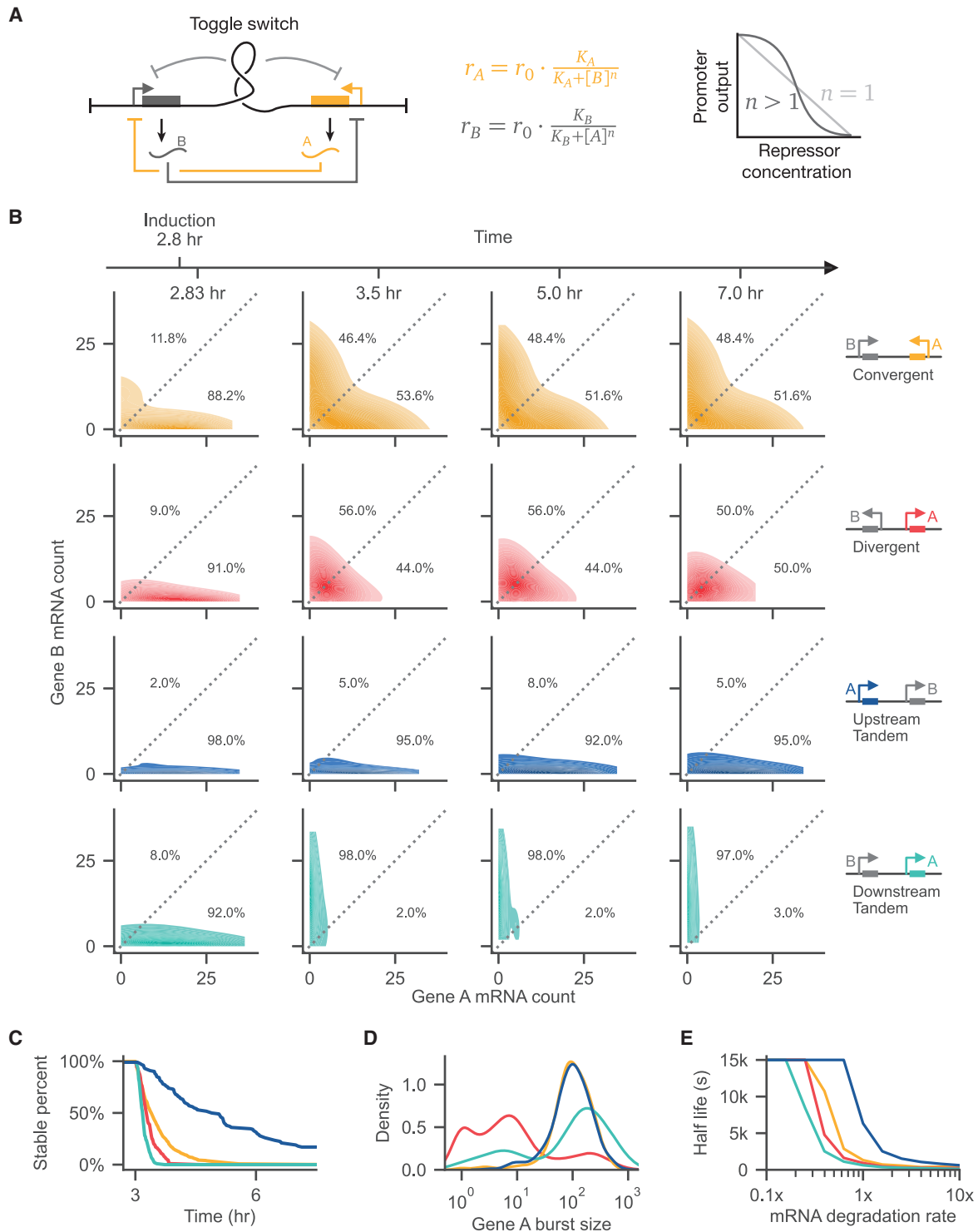
**Figure 4. Burst dynamics vary in different models of DNA supercoiling**

(A) Two model perturbations are compared with the base model, one where polymerase initiation is impossible once the supercoiling density is extreme, and one where nucleosomes provide buffering against positive supercoiling. Each model tunes either the function relating torque and supercoiling density (middle) or the polymerase initiation energy function (bottom); changes relative to the base model are marked with a gray background.

(B) Bursts are defined as a group of consecutive polymerase loading events. The size of a burst is defined as the number of loaded polymerases, whereas inter-burst time is defined as the gap between successive bursts.

(C) The ensemble distribution of burst size is shown for the different orientations for each of the polymerase conditions.

(D) The ensemble distribution of inter-burst time is shown for the different orientations for each of the polymerase conditions. See also [Figures S1–S3](#).



**Figure 5. Toggle switches implemented as a mutually inhibitory pair of genes show syntax-specific stability**

(A) Schematic of a synthetic toggle switch composed of mutual transcriptional repressors, A and B, which are expressed from a promoter negatively regulated by the opposite gene. Repression follows a Hill function (center), which shows cooperativity based on the value of  $n$ . Reactions where  $n$  is greater than 1 show cooperativity. Simulated toggle switches are regulated both by a mutually inhibitory interaction at the mRNA level and via supercoiling-dependent phenomena. (B) The ensemble mRNA count distributions are shown as a function of syntax at four selected time points. All plots represent simulations where the Hill coefficient has been set to  $n = 2.0$ .

(legend continued on next page)

dynamic circuit, we applied our model to the classic repressor-mediated toggle switch (Gardner et al., 2000). Toggle switches are well characterized both theoretically (Gardner et al., 2000) and experimentally (Gardner et al., 2000; Yeung et al., 2017; Zhu et al., 2021). The behavior of a simple toggle under the additional influence of supercoiling-mediated feedback provides an ideal testbed to understand how syntax influences circuit performance.

A toggle switch can be constructed with two genes that mutually repress each other (Figure 5A). Ideally, such a toggle switch exhibits bistability, generating two stable basins. If modeled with continuous, noise-free equations, a toggle switch will remain within one of the basins based on the initial conditions (Gardner et al., 2000). However, if we treat the mRNA concentration discretely with a stochastic simulation, the system escapes the stable basin with a certain probability, depending on the size of fluctuations relative to the steady-state values. How does supercoiling-mediated feedback affect this probability of escape? How might supercoiling-mediated feedback interact with mutual transcriptional repression and alter toggle-switch behavior? To answer these questions, we simulate the behavior of a two-gene toggle switch with our model for various circuit syntaxes. To establish the conventional dual repressor system used for toggle switches, we abstracted the regulatory interaction of the repressors using a Hill function to define the base promoter initiation rates:

$$r_A = r_0 \frac{K_A}{K_A + [B]^n} \quad r_B = r_0 \frac{K_B}{K_B + [A]^n} \quad (\text{Equation 1})$$

for  $r_0 = 1/160 \text{ s}^{-1}$  and  $n = 2.0$ . Here, we do not explicitly model protein production. Rather, transcriptional repression directly depends on the mRNA counts of repressors. This parsimonious model allows us to understand the behavior of the system without introducing additional rate constants. We chose the half-max value  $K$ —the mRNA count at which the promoter activity is half that of  $r_0$ —to approximately match the mean steady-state value of either stable state to ensure that the toggle switch operates in the regime of maximum sensitivity (see "Toggle-switch parameter selection" in STAR Methods for details). In order to compare the behavior of the toggle switches, we initialized toggle switches of different syntaxes within one of the stable basins (gene A) and induced the second gene (gene B) after 2.8 h (see Figure S4A for example runs). The system then evolves under simultaneous mutual inhibition from expression of the repressors as well as from supercoiling-dependent feedback.

Circuit syntax specifies unique toggle-switch dynamics that can be understood by visualizing the distribution of mRNA counts over time (Figure 5B). Initially, nearly all of the simulations in the ensemble lie along the axis corresponding to the initially active gene, gene A. As time progresses, each ensemble approaches and fluctuates around an equilibrium. The convergent

syntax approaches an equilibrium where approximately half of the population distributes into each state. In this syntax, activation of either gene causes positive supercoiling to accumulate in the intergenic region, enhancing negative feedback between genes and thus between states. With divergent syntax, the toggle distributions converge toward monostability with low differential expression of either gene. Accumulation of negative supercoiling between genes enhances polymerase loading, weakening negative feedback. Finally, the vast majority of the simulations in the tandem syntax remain in or transition to the upstream-active state, demonstrating upstream dominance. We find that these results qualitatively hold for varying values of  $n$ , the repressor cooperativity coefficient (Figure S4E). We find that even in the absence of cooperativity,  $n = 1.0$ , the convergent syntax shows some level of bistability (Figure S4D), indicating that supercoiling-mediated feedback introduces a degree of nonlinearity that can reinforce toggle-switch function.

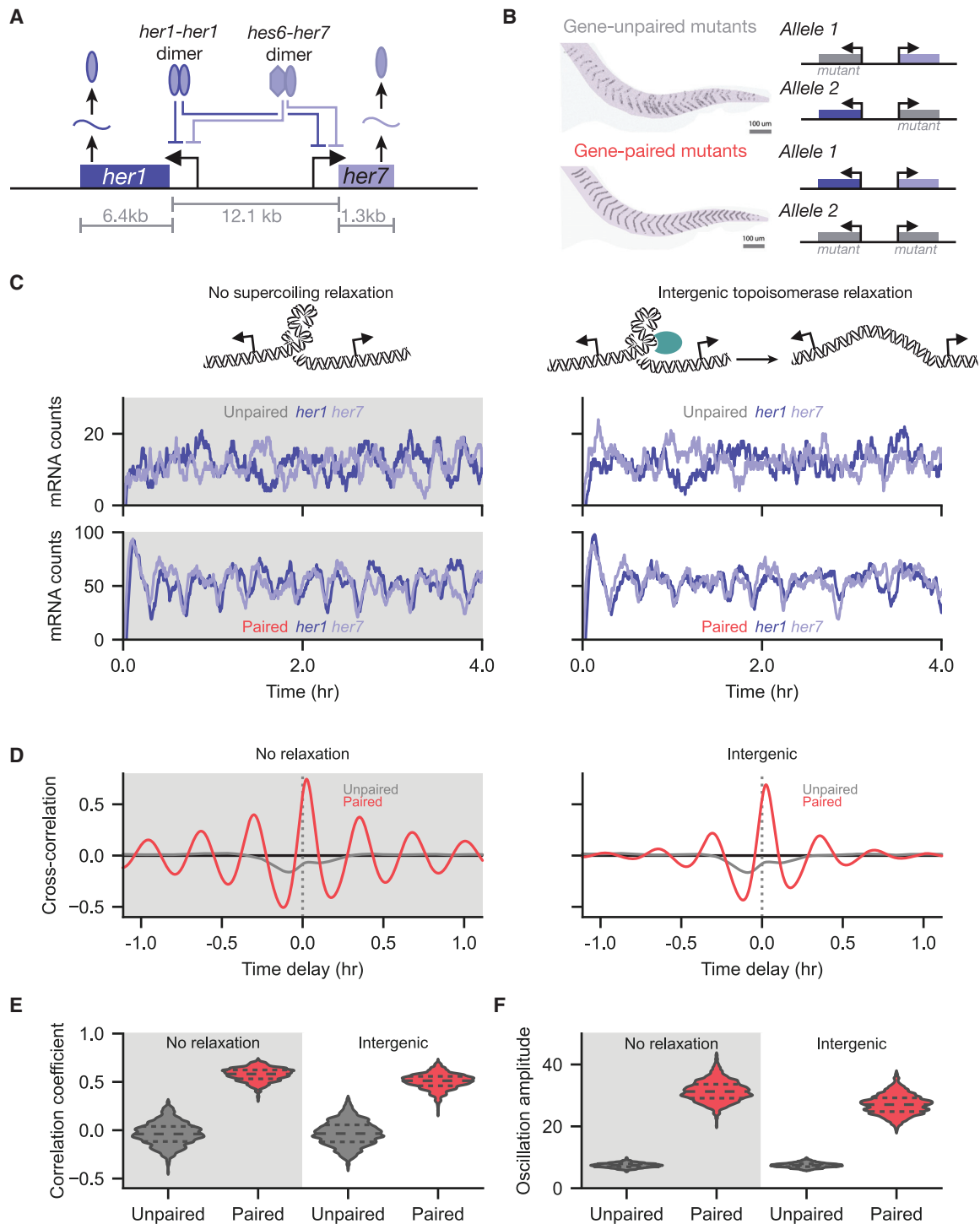
To quantify these distribution results, we computed the stable fraction of the ensemble, defined as the fraction of simulations that have never left the initial starting basin at a certain simulation time. The stable fraction monotonically decreases toward zero with time, as simulations that cross into the other stable basin are no longer counted as stable even if they return to the original basin. We observe substantial syntax differences in the dynamics of the stable fraction of the ensemble (Figure 5C). While the tandem orientations represent the extremes of stability, the convergent and divergent syntaxes exhibit intermediate stabilities. As expected, different burst dynamics characterize toggle-switch behaviors, varying by syntax. In particular, we observe that the divergent syntax displays *reduced* burst size compared with the other syntaxes (Figure 5D). In contrast, the inter-burst time distributions do not significantly vary with syntax (Figure S4C), suggesting that burst size, not burst frequency, is a key parameter tuned by supercoiling-mediated feedback in this context. We hypothesize that the divergent toggle switch may be governed by a conflicting interaction at the promoter level between supercoiling-mediated feedback and mutual inhibition. Overall, these trends suggest that the observed toggle-switch behavior emerges through correlated (or anti-correlated) transcription and changes in burst size, suggesting that toggle-switch behavior could be tuned orthogonally by supercoiling-mediated feedback.

For any stochastic system, the steady-state number of molecules influences the stability of the system. As the reservoir of molecules grows larger, the size of fluctuations relative to the total concentration decreases. For toggle switches, we expect that, as the number of steady-state mRNA molecules grows, we should approach the theoretical, continuous solution that predicts that no state switching occurs. To examine this expectation, we modified the simulated mRNA degradation rate, scaling  $K$  as described in "Toggle-switch parameter selection" in STAR Methods, and plotted the resulting half-lives (Figure 5E). As the mRNA degradation rate goes to zero, we increase the

(C) The stability, measured as the percentage of simulations in the ensemble that have never escaped the initial starting basin, of the four starting states of the system plotted as a function of time.

(D) Expression burst size distributions of the initially active gene A are plotted as a function of circuit syntax.

(E) The half-life at different values of the mRNA degradation rate are shown. As the mRNA degradation rate principally sets the average number of mRNA molecules, high degradation rates lead to systems with low overall mRNA concentration and concordant stochastic instability. See also Figure S4.



**Figure 6. Supercoiling-mediated feedback supports robust transcriptional coordination within the zebrafish segmentation clock circuit**  
 (A) Schematic of the mutually inhibitory *her1-her7* system. Either a *her1-her1* dimer or a *hes6-her7* dimer can bind to either promoter, preventing transcription of the downstream gene.  
 (B) Coupling between *her1-her7* genes on the same allele supports proper zebrafish somite formation (Zinani et al., 2021). Disruption of this intra-allele coupling through unpaired mutations in *her1-her7* leads to loss of proper segmentation.

(legend continued on next page)

reservoir size and observe that the half-life for all syntaxes approaches the simulation upper-limit on the half-life (Figure 5E). Interestingly, increasing mRNA degradation rates reduces the asymmetry in the half-lives between the tandem-upstream and tandem-downstream syntaxes. These results suggest that state switching increases as mRNA degradation rate increases as expected.

### DNA supercoiling tightly coordinates expression of proximal segmentation genes

DNA supercoiling provides a mechanism for the precise coordination of co-localized genes. Through colocalization, native circuits may incorporate transcription-linked feedback mechanisms to reduce noise and tune cell-state specific output in tightly regulated, dynamic processes such as somite formation. In zebrafish, proper somite segmentation requires precise coordination of two clock genes, *her1* and *her7*. *her1* and *her7* form an inhibitory feedback loop encoded in a divergent syntax (Figure 6A). Proper somite formation requires one intact allele of *her1* and *her7*, provided these genes are expressed from the same locus (Zinani et al., 2021). Mutant zebrafish embryos where *her1* and *her7* are only expressed from separate loci eliminate any supercoiling-mediated coupling while retaining the dimer-mediated inhibitory feedback loop. Zinani et al. (2021) found that, in the gene-unpaired embryos, transcriptional coordination between genes is lost and proper somite segmentation is disrupted (Figure 6B). Consequently, physical colocalization represents an important feature supporting transcriptional coordination between genes and proper somitogenesis, which may be mediated by supercoiling-mediated feedback.

Based on our above results from two-gene systems, we hypothesized the feedback from DNA supercoiling supports coordination between discrete transcripts expressed from divergent promoters. Using our full computational model, we replicated the previously developed stochastic reaction network. Importantly, *her1* and *her7* are regulated in a binary fashion; the promoters are either completely off when bound by a dimer or expressed at their basal rate when unbound. We simulated two cases: an *unpaired* system where the simulated genes were separated by a large distance (1 Mb) to prevent supercoiling interactions, and a *gene-paired* system where *her1* and *her7* were spaced at their genomically active locations. In the paired system, linear boundary conditions were used with boundaries chosen at the nearest adjacent genes on each side in the zebrafish genome.

Strikingly, the gene-paired case shows strong periodic levels of mRNA expression (Figure 6C). In fact, such levels of periodicity are not observed even in the original computational model presented by Zinani et al. (2021) (Figures S5A and S5B); we confirmed that this is not simply an artifact of the uniform time resampling performed in order to compare our model behavior to the literature model (Figure S5C). The level of periodicity appears sensitive to our second-order polymerase initiation model. Per-

forming simulations with a weakened second-order penalty term (see STAR Methods; Figures 2 and S7) reduces the amount of periodicity observed (Figures S5D to S5G).

Examining the time-dependent nature of the *her1-her7* system, we plotted the ensemble cross-correlation for paired and unpaired genes (Figure 6D). Here, we found that, in addition to the enhanced positive correlation peak at a time delay of  $\tau = 0$  seconds, the paired case showed exceptionally strong, nearly symmetric cross-correlation at positive and negative time offsets. Such cross-correlation is the hallmark of a periodic signal. Thus, both individual examples (Figure 6C) and ensemble behavior (Figure 6D) show that supercoiling-mediated feedback provides a strong mechanistic driver of inter-gene coordination in the *her1-her7* clock circuit that is inaccessible to solely dimer-mediated regulation.

In order to confirm that these results apply across the ensemble, we examined the ensemble correlation between the counts of *her1* and *her7* mRNA (Figure 6E). We found that while the unpaired case shows minimal correlation, the gene-paired case shows strong correlation between the two clock genes. We attribute this strong, periodic correlation to the additional biophysical coupling conferred by the divergent syntax. Notably, we observed that our model predicts an increase in the amplitude of oscillations (Figure 6F). *In vivo*, loss of gene pairing reduces oscillation amplitude, leading to improper segmentation (Zinani et al., 2021).

Because this periodic behavior depends on biophysical coupling, we investigated whether the periodicity was robust to topoisomerase activity, nucleosome buffering, or hypernegative chromatin structure formation. We found that intergenic topoisomerase relaxation (see STAR Methods) does not abrogate the periodic behavior, with strong correlated oscillations still visible (Figures 6D–6F). However, intergenic topoisomerase activity does reduce the cross-correlation after a few periods, indicating that topoisomerase activity contributes to variance in oscillation frequency while still supporting strongly correlated *her1* and *her7* expression. Intragenic topoisomerase activity showed similar behaviors (Figure S5I). Finally, we found that while penalization of hypernegative supercoiling densities eliminates periodic behavior (Figures S2D to S2G), periodicity is maintained in the presence of nucleosome buffering (Figures S3D to S3G). Thus, we propose that supercoiling-mediated feedback offers a mechanism to support robust oscillations in the *her1* and *her7* network for proper somite formation.

## DISCUSSION

Transcription induces significant variance in gene expression. At a single-cell level, individual genes are expressed stochastically, with most genes experiencing relatively long periods of quiescence punctuated by bursts of polymerase activity. Phenomenological models of this process based on stochastic probability

(C) Example simulations of *her1* and *her7* mRNA levels are shown as a function of gene syntax and topoisomerase activity.

(D) The ensemble cross-correlation between the *her1* and *her7* mRNA counts is shown across pairing and topoisomerase conditions. The large maxima at  $\tau = 0$  combined with large roughly symmetric minima observed in the gene-paired cases signal the strong cyclic behavior observed experimentally in zebrafish.

(E) The distribution of correlation coefficients between the *her1* and *her7* mRNA counts is shown for the various pairing and topoisomerase conditions.

(F) The oscillation amplitude over the ensemble is shown for the various conditions. See also Figure S5.

distributions can provide some insights, but defining the mechanically regulated physical factors that influence RNA polymerase dynamics will improve existing models of gene regulation and support enhanced design of transgenic systems. Importantly, we sought to develop a model that would define a set of experimentally testable predictions as well as lay the groundwork for future modeling across multigenic loci and circuits.

In this work, we developed a model of supercoiling-mediated feedback that captures emergent coupling between neighboring genes to influence expression levels as well as dynamics. This model allowed us to tractably compute polymerase activity at the scale of synthetic circuits (Figure 2). Within supercoiling-mediated feedback, we included both supercoiling-dependent polymerase motion terms and supercoiling-dependent polymerase initiation terms. This computational framework lays the groundwork for understanding how DNA supercoiling functions as a regulatory mechanism that can be integrated with canonical biochemical models of gene regulation. Using this model, we extracted insights into how mechanical and biochemical regulation combine to generate diverse profiles of expression and support or impede the performance of gene networks.

We find that induction of neighboring genes significantly influences the transcriptional activity of both genes (Figure 2). Syntax-specific differences in DNA supercoiling dynamics, expression profiles, and noise emerge due to physical coupling. We find that such a system regulated by this biophysical coupling is responsive to expression level changes in adjacent genes, with both mean expression and population variance changing as a function of gene orientation (Figure 3). The observed supercoiling-mediated feedback is itself dependent on inter-gene spacing, mRNA degradation rate, and other variables tunable in an experimental setting. Generally, accumulated negative supercoiling leads to correlated bursting, which occurs concomitant to a decrease in intrinsic noise. In contrast, accumulated positive supercoiling can lead to anti-correlated bursting, which instead enhances intrinsic noise. While the tandem syntax does not lead to large supercoiling accumulation in the intergenic region, we observe upstream dominance, where the upstream gene is more highly expressed than the downstream gene.

Our prediction of burst dynamics (Figure 4) complements theoretical and experimental investigations of cooperative interactions of RNA polymerases arising from the beneficial cancellation of positive and negative supercoiling generated by adjacent polymerases (Sevier and Hormoz, 2022; Kim et al., 2019). We predict that syntax causes a 2- to 3-fold change in burst size (Figure 4C) but can shift inter-burst time by an order of magnitude (Figure 4D). In addition, the syntax-specific trends in inter-burst time are not strongly affected by other potential mechanical regulators of eukaryotic chromatin—nucleosomes and strained structures such as R-loops—suggesting that syntax may robustly control gene expression through differences in burst dynamics. While we do not directly examine transcription elongation rates, we similarly predict syntax-specific differences in expression dynamics but observe distinct syntax-specific behaviors in our model (Sevier and Hormoz, 2022; Tripathi et al., 2021). We also find that intergenic distance only weakly affects

supercoiling feedback (Tripathi et al., 2021). In alignment with experimental work, we find that positive supercoiling accumulates in the intergenic region of convergently oriented native genes (Guo et al., 2021). When combined with sequencing methods that precisely measure nascent mRNA transcription (Mellor et al., 2016), these methods may provide a window into experimental systems in order to test theoretical predictions of our work and others.

The fast timescale of supercoiling-mediated feedback offers access to a uniquely tunable and orthogonal form of gene regulation. In contrast to regulatory mechanisms dependent on relatively long timescales, such as mRNA- and protein-mediated systems, supercoiling-mediated feedback occurs at the time-scale of seconds. Polymerases can stall and unstick each other within seconds, while local polymerase loading rates can vary over the course of minutes. By combining the fast dynamic feedback with slower classic feedback mechanisms, circuit regulation can be selectively stabilized or destabilized. We found that specification of syntax within a simple two-gene toggle switch generated diverse behaviors, including a reasonably stable switch, a hypersensitive toggle with hysteresis, and an asymmetric system that preferentially decays toward a single target state (Figure 5). As a rapid mechanism for coordinating transcriptional dynamics, supercoiling-dependent feedback may support intergenic coordination in native systems. Examining the zebrafish segmentation clock, we find that addition of supercoiling-mediated feedback recapitulates the synchronized, periodic expression of the clock genes, *her1-her7* (Figure 6).

Our model integrates supercoiling-mediated biophysical feedback with classic gene regulation motifs that are well studied in native and synthetic contexts. This unified framework brings us closer to an understanding of how supercoiling contributes to transcriptional regulation. We offer testable predictions about the performance of genetic circuits. The predicted changes in reporter output, supercoiling density, and burst dynamics observed in Figures 2–4 are experimentally accessible with modern sequencing and single-cell imaging technology (Guo et al., 2021; Mellor et al., 2016; Patel et al., 2022). Experimental verification of our theoretical results will aid in constructing a mechanistic understanding of how transcription-induced supercoiling couples expression. Harnessing these insights will enable gene regulation at the level of transcription, providing a robust method to control expression dynamics, levels, and noise.

### Limitations of the study

In deriving our model, we made several simplifying assumptions. Our derivation of the energy function for supercoiling-dependent polymerase initiation adds relevant molecular detail to our model. The second-order correction term reflects the asymptotic relationship observed in *in vitro* assays between torque and supercoiling density for underwound DNA (Le et al., 2019). Inclusion of this correction supports the periodic behavior of the native *her1-her7* clock circuit, suggesting this term may accurately capture regulation *in vivo*. For simplicity, we model the dynamic processes of RNAP binding, initiation, and pause release as a single reaction. In real biological systems, each of these processes may vary across the genome by sequence

and by the presence of nucleosomes, transcription factors, and other DNA-binding proteins. Formation of supercoiled structures shows sequence bias *in vitro* that may affect *in vivo* structures and gene regulation (Kim et al., 2018).

While we do approximate the behavior of the system in the presence of hypernegative/hyperpositive chromatin structures and nucleosomes, we do not explicitly model the presence of these and instead perturb our torque and polymerase binding energy functions. This implicit simulation method may not accurately capture the discrete nature of these phenomena, especially in the case of positive supercoiling waves displacing nucleosomes. Furthermore, while real topoisomerases dynamically relax chromatin and are often recruited to sites of active transcription (Baranello et al., 2016), we use a simplified model that instantly relaxes DNA in a non-specific manner. We find that our systems are mostly insensitive to this simple model of topoisomerase activity; future work using a different topoisomerase activity model may reveal additional roles for supercoiling relaxation.

Our model also excludes polymerase collision, premature termination, and the impact of 3D structures and loop domains formed by protein complexes such as CCCTC-binding factor (CTCF) and other structural maintenance of chromosomes (SMC) proteins. More broadly, we assume that regions of simulated chromatin are uniformly accessible and have equal torque responses during polymerase elongation; these assumptions may fail at the boundaries of chromatin domains. Finally, we also neglect the speed of supercoiling diffusion. While this is expected to be a negligible effect at the scale considered in this work, supercoiling diffusion remains slow at the scale of hundreds of kilobases to megabases.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - Supercoiling-dependent transcription model
  - Supercoiling-dependent initiation model
  - Additional discrete reaction model
  - Boundary conditions
  - Burst threshold choice
  - Toggle-switch parameter selection
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - ODE and stochastic simulation
  - Summarizing stochastic simulations
  - Computing cross-correlation

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2022.111492>.

## ACKNOWLEDGMENTS

Research reported in this manuscript was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35-GM143033. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center (Reuther et al., 2018) for providing HPC resources that have contributed to the research results reported within this paper. We also thank Adam Beitz, Patrick Han, Ross Jones, Sneha Kabaria, Brittany Lende, Kasey Love, Christian Otero, Emma Peterman, Jack Suggs, and Nathan Wang for helpful suggestions throughout and for their effort in editing this manuscript.

## AUTHOR CONTRIBUTIONS

C.P.J. and K.E.G. conceptualized the work. C.P.J. developed the mathematical model, wrote the simulation and data analysis software, and visualized and synthesized results. K.E.G. supervised the project and acquired financial support for this project. C.P.J. and K.E.G. discussed results and reviewed and edited the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 2, 2022

Revised: August 4, 2022

Accepted: September 21, 2022

Published: October 18, 2022

## REFERENCES

- Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. *Nature* 577, 701–705. <https://doi.org/10.1038/s41586-020-1934-4>.
- Ancona, M., Bentivoglio, A., Brackley, C.A., Gonnella, G., and Marenduzzo, D. (2019). Transcriptional bursts in a nonequilibrium model for gene regulation by supercoiling. *Biophys. J.* 117, 369–376. <https://doi.org/10.1016/j.bpj.2019.04.023>. pmid: 31103229.
- Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B.N., Chung, H.J., Chan-Salis, K.Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., et al. (2016). RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* 165, 357–371. <https://doi.org/10.1016/j.cell.2016.02.036>.
- Bashor, C.J., Helman, N.C., Yan, S., and Lim, W.A. (2008). Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* 319, 1539–1543. <https://doi.org/10.1126/science.1151153>.
- Beitz, A.M., Oakes, C.G., and Galloway, K.E. (2022). Synthetic gene circuits as tools for drug discovery. *Trends Biotechnol.* 40, 210–225. <https://doi.org/10.1016/j.tibtech.2021.06.007>.
- Chen, Y.Y., Galloway, K.E., and Smolke, C.D. (2012). Synthetic biology: Advancing biological Frontiers by building synthetic systems. *Genome Biol.* 13, 240. <https://doi.org/10.1186/gb-2012-13-2-240>.
- Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in Bacteria. *Cell* 158, 314–326. <https://doi.org/10.1016/j.cell.2014.05.038>.
- Danino, T., Mondragón-Palomino, O., Tsimring, L., and Hasty, J. (2010). A synchronized quorum of genetic clocks. *Nature* 463, 326–330. <https://doi.org/10.1038/nature08753>.
- Desai, R.V., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H., et al. (2021). A DNA-repair pathway can affect transcriptional noise to promote cell fate transitions. *Science* 373, eabc6506, pmid: 34301855. <https://doi.org/10.1126/science.abc6506>.
- El Houdaigui, B., Forquet, R., Hindré, T., Schneider, D., Nasser, W., Reverchon, S., and Meyer, S. (2019). Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.* 47, 5648–5657. <https://doi.org/10.1093/nar/gkz300>.

- Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. <https://doi.org/10.1038/35002125>.
- Elowitz, M., and Lim, W.A. (2010). Build life to understand it. *Nature* 468, 889–890. <https://doi.org/10.1038/468889a>.
- Frei, T., Cella, F., Tedeschi, F., Gutiérrez, J., Stan, G.B., Khammash, M., and Siciliano, V. (2020). Characterization and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* 11, 4641. <https://doi.org/10.1038/s41467-020-18392-x>.
- Galloway, K.E., Franco, E., and Smolke, C.D. (2013). Dynamically reshaping signaling networks to program cell fate via genetic controllers. *Science* 341. <https://doi.org/10.1126/science.1235005>.
- Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. <https://doi.org/10.1038/35002131>.
- Guo, M.S., Kawamura, R., Littlehale, M.L., Marko, J.F., and Laub, M.T. (2021). High-resolution, genome-wide mapping of positive supercoiling in chromosomes. *eLife* 10, e67236. <https://doi.org/10.7554/eLife.67236>.
- Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548. <https://doi.org/10.1038/s41586-019-1464-0>.
- Henninger, Jonathan, E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M.M., Andrews, J.O., Zamudio, A.V., Lazaris, C., et al. (2020). RNA-mediated feedback control of transcriptional condensates. *Cell*, 0092–8674. <https://doi.org/10.1016/j.cell.2020.11.030>.
- Hsieh, Tsung-Han, S., et al. (2020). Resolving the 3D landscape of transcription-linked Mammalian chromatin folding. *Molecular Cell*, 1097–2765. <https://doi.org/10.1016/j.molcel.2020.03.002>.
- Johnstone, C.P., and Galloway, K.E. (2021). Engineering cellular symphonies out of transcriptional noise. *Nat. Rev. Mol. Cell Biol.*, 369–370. <https://doi.org/10.1038/s41580-021-00359-5>.
- Jones, R.D., Qian, Y., Siciliano, V., DiAndreth, B., Huh, J., Weiss, R., and Del Vecchio, D. (2020). An Endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* 11, 5690. <https://doi.org/10.1038/s41467-020-19126-9>.
- Kim, S., Beltran, B., Irnov, I., and Jacobs-Wagner, C. (2019). Long-distance cooperative and Antagonistic RNA polymerase dynamics via DNA supercoiling. *Cell* 179, 106–119.e16. <https://doi.org/10.1016/j.cell.2019.08.033>.
- Kim, S.H., Ganji, M., Kim, E., van der Torre, J., Abbondanzieri, E., and Dekker, C. (2018). DNA sequence encodes the position of DNA supercoils. *eLife* 7, e36557. <https://doi.org/10.7554/eLife.36557>.
- Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20, 396–403. <https://doi.org/10.1038/nsmb.2517>.
- Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.H.S., Parsi, K.M., Yang, L., Maehros, R., Mirny, L.A., et al. (2020). Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* 78, 554–565.e7. <https://doi.org/10.1016/j.molcel.2020.03.003>.
- Le, T.T., Gao, X., Park, S.H., Lee, J., Inman, J.T., Lee, J.H., Killian, J.L., Badman, R.P., Berger, J.M., and Wang, M.D. (2019). Synergistic coordination of chromatin torsional mechanics and topoisomerase activity. *Cell* 179, 619–631.e15. <https://doi.org/10.1016/j.cell.2019.09.034>.
- Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* 84, 7024–7027. <https://doi.org/10.1073/pnas.84.20.7024>. pmid: 2823250.
- van Loenhout, M.T.J., de Grunt, M.V., and Dekker, C. (2012). Dynamics of DNA supercoils. *Science* 338, 94–97. <https://doi.org/10.1126/science.1225810>. pmid: 22983709.
- Ma, Y., Budde, M.W., Mayalu, M.N., Zhu, J., Lu, A.C., Murray, R.M., and Elowitz, M.B. (2022). Synthetic mammalian signaling circuits for robust cell population control. *Cell*. <https://doi.org/10.1016/j.cell.2022.01.026>.
- Marko, J.F. (2007). Torque and dynamics of linking number relaxation in stretched supercoiled DNA. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76, 021926–023755. <https://doi.org/10.1103/PhysRevE.76.021926>.
- Mellor, J., Woloszczuk, R., and Howe, F.S. (2016). The interleaved genome. *Trends Genet.* 32, 57–71. <https://doi.org/10.1016/j.tig.2015.10.006>.
- Meyer, S., and Beslon, G. (2014). Torsion-mediated interaction between adjacent genes. *PLoS Comput. Biol.* 10, e1003785–e1003758. <https://doi.org/10.1371/journal.pcbi.1003785>.
- Muniz, L., Nicolas, E., and Trouche, D. (2021). RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* 40, e105740–e114189. <https://doi.org/10.15252/embj.2020105740>. pmid: 34254686.
- Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* 20, 387–395. <https://doi.org/10.1038/nsmb.2509>. pmid: 23416946.
- Park, M., Patel, N., Keung, A.J., and Khalil, A.S. (2019). Engineering epigenetic regulation using synthetic read-write modules. *Cell* 176, 227–238.e20. <https://doi.org/10.1016/j.cell.2018.11.002>.
- Patel, H.P., Coppola, S., Pomp, W., Brouwer, I., and Lenstra, T.L. (2022). DNA supercoiling restricts the transcriptional bursting of neighboring eukaryotic genes. Preprint at arXiv. <https://doi.org/10.1101/2022.03.04.482969>.
- Popp, A.P., Hettich, J., Christof, J., and Gebhardt, M. (2021). Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Res.*, 0305–1048. <https://doi.org/10.1093/nar/gkab443>.
- Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from Modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422. <https://doi.org/10.1038/nrm2698>.
- Qian, Y., Huang, H.H., Jiménez, J.I., and Del Vecchio, D. (2017). Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* 6, 1263–1272. <https://doi.org/10.1021/acssynbio.6b00361>.
- Quarton, T., Kang, T., Papakis, V., Nguyen, K., Nowak, C., Li, Y., and Bleris, L. (2020). Uncoupling gene expression noise along the central dogma using genome engineered human cell lines. *Nucleic Acids Res.* 48, 9406–9413. <https://doi.org/10.1093/nar/gkaa668>.
- Rackauckas, C., and Nie, Q. (2017). DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Softw.* 5. <https://doi.org/10.5334/jors.151>.
- Reuther, A., Kepner, J., Byun, C., Samsi, S., Arcand, W., Bestor, D., Bergeron, B., Gadepally, V., Houle, M., Hubbell, M., et al. (2018). Interactive supercomputing on 40,000 cores for machine learning and data analysis. In 2018 IEEE High Performance Extreme Computing Conference (HPEC). 2018 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1–6. <https://doi.org/10.1109/HPEC.2018.8547629>.
- Revyakin, A., Ebricht, R.H., and Strick, T.R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc. Natl. Acad. Sci. USA* 101, 4776–4780, issn: 0027-8424, 1091-6490. <https://doi.org/10.1073/pnas.0307241101>. pmid: 15037753.
- Rodriguez, J., and Larson, D.R. (2020). Transcription in living cells: molecular mechanisms of bursting. *Annu. Rev. Biochem.* 89, 189–212. <https://doi.org/10.1146/annurevbiochem-011520-105250>. pmid: 32208766.
- Rodriguez, J., Ren, G., Day, C.R., Zhao, K., Chow, C.C., and Larson, D.R. (2019). Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* 176, 213–226.e18. <https://doi.org/10.1016/j.cell.2018.11.026>.
- Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. <https://doi.org/10.1038/s41576-018-0060-8>.



- Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* 67, 837–852.e7. <https://doi.org/10.1016/j.molcel.2017.07.022>.
- Sevier, S.A., and Hormoz, S. (2022). Collective polymerase dynamics emerge from DNA supercoiling during transcription. Preprint at bioRxiv. <https://doi.org/10.1101/2021.11.24.469850>.
- Sevier, S.A., and Levine, H. (2018). Properties of gene expression and chromatin structure with mechanically regulated elongation. *Nucleic Acids Res.* 46, 5924–5934. <https://doi.org/10.1093/nar/gky382>.
- Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019). Interplay between DNA sequence and negative superhelicity drives Rloop structures. *Proc. Natl. Acad. Sci. USA* 116, 6260–6269. <https://doi.org/10.1073/pnas.1819476116>.
- Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–519. <https://doi.org/10.1038/nature07389>.
- Tabor, J.J., Salis, H.M., Simpson, Z.B., Chevalier, A.A., Levskaya, A., Marcotte, E.M., Voigt, C.A., and Ellington, A.D. (2009). A synthetic genetic edge detection program. *Cell* 137, 1272–1281. <https://doi.org/10.1016/j.cell.2009.04.048>.
- Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* 21, 88–94. <https://doi.org/10.1038/nsmb.2723>. pmid: 24317489.
- To, T.-L., and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* 327, 1142–1145. <https://doi.org/10.1126/science.1178962>.
- Tripathi, S., Brahmachari, S., Onuchic, J.N., and Levine, H. (2021). DNA supercoiling-mediated collective behavior of co-transcribing RNA polymerases. <https://doi.org/10.1101/2021.03.04.433986>.
- Tsitouras, C. (2011). Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption. *Comput. Math. Appl.* 62, 770–775. <https://doi.org/10.1016/j.camwa.2011.06.002>.
- Weinberg, B.H., Pham, N.T.H., Caraballo, L.D., Lozanoski, T., Engel, A., Bhatia, S., and Wong, W.W. (2017). Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* 35, 453–462. <https://doi.org/10.1038/nbt.3805>.
- Wu, H.-Y., Shyy, S., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440. [https://doi.org/10.1016/0092-8674\(88\)90163-8](https://doi.org/10.1016/0092-8674(88)90163-8).
- Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* 333, 1307–1311. <https://doi.org/10.1126/science.1205527>.
- Yeung, E., Dy, A.J., Martin, K.B., Ng, A.H., Del Vecchio, D., Beck, J.L., Collins, J.J., and Murray, R.M. (2017). Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst.* 5, 11–24.e12. <https://doi.org/10.1016/j.cels.2017.06.001>.
- Zhu, R., del Rio-Salgado, J.M., Garcia-Ojalvo, J., and Elowitz, M.B. (2021). Synthetic multistability in mammalian cells. Preprint at bioRxiv. <https://doi.org/10.1101/2021.02.10.430659>.
- Zinani, O.Q.H., Keseroğlu, K., Ay, A., and Özbudak, E.M. (2021). Pairing of segmentation clock genes drives robust pattern formation. *Nature* 589, 431–436. <https://doi.org/10.1038/s41586-020-03055-0>.
- Zopf, C.J., Quinn, K., Zeidman, J., and Maheshri, N. (2013). Cell-cycle dependence of transcription Dominates noise in gene expression. *PLoS Comput. Biol.* 9, e1003161–e1007358. <https://doi.org/10.1371/journal.pcbi.1003161>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Simulation datasets	This work	<a href="https://doi.org/10.5281/zenodo.7041641">https://doi.org/10.5281/zenodo.7041641</a>
Software and algorithms		
Adobe Illustrator CC	Adobe Systems	<a href="https://www.adobe.com">https://www.adobe.com</a>
Julia 1.6.1		<a href="https://julialang.org">https://julialang.org</a>
DifferentialEquations.jl 6.19.0	<a href="#">Rackauckas and Nie, 2017</a>	<a href="https://diffeq.sciml.ai/stable">https://diffeq.sciml.ai/stable</a>
Python 3		<a href="https://www.python.org">https://www.python.org</a>
Simulation and figure-generation code	This work	<a href="https://doi.org/10.5281/zenodo.7054394">https://doi.org/10.5281/zenodo.7054394</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources, datasets, and code should be directed to and will be fulfilled by the lead contact, Kate E. Galloway ([katiegal@mit.edu](mailto:katiegal@mit.edu)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- All unprocessed and preprocessed simulation data reported in this paper has been deposited at Zenodo and is publicly available with the DOI: <https://doi.org/10.5281/zenodo.7041641>.
- All original code has been deposited at Zenodo and is publicly available with the DOI: <https://doi.org/10.5281/zenodo.7054394>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### METHOD DETAILS

#### Supercoiling-dependent transcription model

Supercoiling is defined as the amount of excess twist  $\varphi$  relative to relaxed DNA. Relaxed DNA rotates one full revolution per  $\approx 10$  basepairs (1bp  $\approx 0.34$  nm); thus its relaxed twist is  $\omega_0 = 1.85$  radians/nm. Supercoiling density can generally be defined as a varying function of genomic location  $\sigma(z)$ . However, in a region of constant supercoiling density, we can use the excess DNA twist  $\varphi_i, \varphi_{i+1}$  at the endpoints  $z_i, z_{i+1}$  to define the supercoiling density as:

$$\sigma = \frac{\varphi_i - \varphi_{i+1}}{\omega_0(z_{i+1} - z_i)} \quad (\text{Equation 2})$$

For a polymerase with  $\varphi_1 > 0$  between endpoints with  $\varphi_0 = \varphi_2 = 0$ , [Equation 2](#) implies that the supercoiling density is *positive* in front of the polymerase and *negative* behind the polymerase ([Figure 1](#)). Following on the work of Sevier et al. ([Sevier and Levine, 2018](#)), we assume that on the length scales of synthetic and native circuits of interest ( $O(\approx 10\text{kb})$ ), the supercoiling density is constant in all regions between polymerases and other barriers. Because supercoiling diffusion and plectoneme hopping ([Loenhout and Dekker, 2012](#)) occur at rates faster than transcription (supercoiling diffusion:  $\mathcal{D} \approx O(0.5 \frac{\text{kb}^2}{\text{s}})$  versus transcription rate:  $v_0 \approx 0.05 \frac{\text{kb}}{\text{s}}$ ) ([Muniz et al., 2021](#)), the supercoiling generated by a polymerase will diffuse outward far more rapidly than polymerases can move. As in previous reported work, we make a pseudo-steady assumption for inter-RNAP supercoiling—assuming that [Equation 2](#) holds between polymerases—over the relatively small ( $\sim 10$  kb) genomic distances considered in this work in order to simplify the resulting model.

How does transcription both drive the process of supercoiling generation and react to changes in local supercoiling? Under the assumption of supercoiling relaxation, each polymerase is defined by four variables—the one-dimensional genomic location of the polymerase  $z_i$ , the length of the nascent RNA  $x_i$ , the excess twist at the location of the polymerase  $\varphi_i$ , and the rotation angle of the polymerase ([Figure 1](#)). Then, two governing equations define the motion of all polymerases ([Sevier and Levine, 2018](#)). First, we equate linear polymerase motion with the rotational motion required to track the DNA groove:

$$\underbrace{\omega_0 \frac{dz_i}{dt}}_{\text{RNAP velocity}} = \underbrace{\frac{d\theta_i}{dt}}_{\text{RNAP rotation}} + \underbrace{\frac{d\varphi_i}{dt}}_{\text{supercoiling generation}} \quad (\text{Equation 3})$$

where the change in  $\theta_i$  represents polymerase rotation and the change in  $\varphi_i$  represents local rotation of the DNA. The second equation provides a torque balance between DNA-mediated torques on the left hand side and torque caused by drag acting on the nascent RNA:

$$\underbrace{\tau(\sigma(z_i, \varphi_{i-1}, \varphi_{i+1}))}_{\text{torque acting on RNAP}} + \underbrace{\chi \frac{d\varphi_i}{dt}}_{\text{supercoiling restoring force}} = \underbrace{\eta \chi_i^n \frac{d\theta_i}{dt}}_{\text{nascent RNA drag}} \quad (\text{Equation 4})$$

To develop a final system of ordinary differential equations, we still must define the torque response function  $\tau(\sigma)$  and the polymerase velocity function  $\frac{dz}{dt} = v(\tau)$ . With these two functions, Equations 3 and 4 can be solved as in [Sevier and Levine \(2018\)](#). Here, we use Marko's torque-response model of supercoiling which accounts for the thermodynamic behavior of both non-buckled, twisted DNA and buckled, plectonemic DNA (see [Equations \(S1\) and \(S4\) in supplemental section B \(Marko, 2007\)](#)). The resulting  $\tau(\sigma)$  function exhibits a phase transition, where the torque response is nearly constant at intermediate values of  $\sigma$  where the DNA is transitioning from a locally-twisted phase to a plectonemic-phase ([Figure S6A](#)).

For the velocity response of a polymerase experiencing a torque  $\tau_f$  in front and  $\tau_b$  behind, we model polymerase stalling as:

$$v(\tau_f, \tau_b) = \frac{v_0}{(1 + e^{(|\tau_f| - \tau_s)/\tau_w})(1 + e^{(|\tau_b| - \tau_s)/\tau_w})} \quad (\text{Equation 5})$$

where the stall torque  $\tau_s = 12$  pN nm and stall-width  $\tau_w = 3$  pN nm define a sigmoidal stall-response curve. As shown in [Figures S6E and S6F](#), our results are only weakly dependent on the specific choice of  $\tau_s$  and  $\tau_w$ . Importantly, our selected phenomenological term will stall polymerase motion if *either* the torque upstream or downstream exceeds the stall torque  $\tau_s$ . Some models choose a stalling equation that only stalls if the difference between the upstream and downstream torque exceeds a stall torque ([Tripathi et al., 2021](#)); we chose this form, reasoning that the DNA unwinding and rewinding process opposed, respectively, by upstream and downstream torque could independently stall. When simulated, the difference between these stalling models is small in practice; polymerases at the start or end of the burst encounter both high upstream and downstream torques *and* a high torque difference, whereas polymerases in the middle of a burst experience both lower adjacent torques and a lower torque difference.

Taking the above equations together, we can simulate the coupled motion of an arbitrary number of polymerases as a single coupled ODE system. We further examine different experimental systems by implementing different boundary conditions that allow us to simulate both plasmid systems and genomically-integrated systems (see [Boundary conditions](#)).

### Supercoiling-dependent initiation model

While the described differential equation system can simulate polymerase motion, we need a way to model the addition of polymerases to simulated genes. A simple strategy is to assume a supercoiling-independent initiation rate and use a stochastic simulation method to randomly add polymerases to transcriptional start sites at a certain fixed rate. However, this simple model assumes that polymerases can bind equally well to initiation sites independent of local supercoiling, missing supercoiling-dependent binding dynamics ([Revyakin et al., 2004](#)). In order to include supercoiling in a polymerase initiation model, we relate the basal expression rate to a corresponding base energy term. We can then additively introduce extra energy costs for polymerase binding under different local supercoiling conditions. Under the approximation that the direct energetic cost of locally melting the DNA to fit in the RNAP groove dwarfs the relative change in unwinding energy caused by supercoiling, the majority of the energetic cost comes from inserting supercoiling ahead and behind the inserted polymerase. Under this assumption, the first-order supercoiling energetic correction can be written as:

$$E_{sc} = 1.2 \cdot 2\pi \cdot \tau(\sigma) \quad (\text{Equation 6})$$

Is this a good approximation? We can estimate the energetic cost of local melting, and find that neglecting local melting leads to a minor change in the resulting energy as seen in [Figure S6B](#). A full derivation of Equation 6 is given in [Methods S1](#).

While this first-order energetic term introduces much-needed behavior to the modeled system—where locally high positive supercoiling decreases the RNAP initiation rate and locally negative supercoiling increases the RNAP initiation rate—at extreme values of  $\sigma$ , this energetic term gives aphysical predictions. In particular, under highly negative supercoiling densities, the energetics of polymerase loading become increasingly favorable, with loading sometimes occurring more than two orders of magnitude faster when

compared with relaxed DNA. To correct for this behavior, we add a second-order (quadratic) term that constrains polymerase loading at highly positive or negative local supercoiling:

$$E_{sc} = 1.2 \cdot 2\pi \cdot \tau(\sigma) + \alpha \cdot \tau_0 \cdot \sigma^2 \quad (\text{Equation 7})$$

for  $\tau_0$ , the relevant scale factor in the  $\tau(\sigma)$  equation (Equation (S1) and (S3) from (Marko, 2007)) and  $\alpha$ , a positive tunable parameter. As the  $\tau(\sigma)$  equation is linear in  $\sigma$  outside of the phase-transition region, this added  $\sigma^2$  term can be contextualized as an additional term in the Taylor expansion of the physically-realistic  $E_{sc}(\sigma)$  equation. This form of the binding energy enables us to qualitatively match the experimentally observed asymptotic behavior between torque and supercoiling for underwound DNA (Le et al., 2019).

For these three models of supercoiling-dependent initiation, we found that the supercoiling-independent initiation model predicted only small changes in reporter output (Figure S7A). Comparing the first- and second-order models, we found that a critical value of  $\alpha$  existed,  $\alpha \approx 0.2$ , above which the second-order model demonstrated emergent non-monotonic behavior (Figure S7). At low values of  $\alpha$ , the second-order model behaves similarly to the first-order model, so we used  $\alpha = 0.025$  for this work. Increasing  $\alpha$  beyond this chosen value appears to scale down reporter output without qualitatively modifying behavior (Figure S7).

When simulating the ODE model, the rate of stochastic polymerase initiation,  $r_{\text{initiation}}$ , varies continuously based on the local supercoiling density  $\sigma$  at the transcription start site as:

$$r_{\text{initiation}} = r_{\text{base rate}} \cdot e^{-E_{sc}/(k_B T)} \quad (\text{Equation 8})$$

### Additional discrete reaction model

Many of the native and synthetic systems of interest include mechanisms of gene regulation that rely on other regulatory species. In order to analyze these types of systems using our supercoiling model, we extended our model to simultaneously simulate arbitrary discrete stochastic equations—such as those commonly used in the literature to model protein production, degradation, dimerization, and more. This addition allowed us to model discrete events otherwise not accounted for in the continuous model.

Importantly, this framework allows us to simulate the activity of topoisomerases. While the total amount of supercoiling (the integral of supercoiling density) is conserved, topoisomerases can be modeled as stochastic events that redistribute supercoiling in certain regions. Here, we simulated the removal of supercoiling in either intergenic or intragenic regions. Removal of supercoiling in intergenic regions is performed by updating the rotation of polymerases on adjacent genes to make the intergenic supercoiling density zero. After relaxation of supercoiling in intragenic regions, polymerases on the gene experience a constant supercoiling density which conserves overall supercoiling. With the exception of the segmentation gene network simulations in Figure 6 and related supplemental figures, we found that inclusion of topoisomerase relaxation did not appreciably change the observed results. Thus, unless otherwise stated, simulations were performed without topoisomerase relaxation.

In addition, we allowed the base initiation rate of genes to vary as an arbitrary function of all species concentrations in the model ( $S_i$ ), such that Equation 8 becomes:

$$r_{\text{initiation}} = r(S_i) \cdot e^{-E_{sc}/(k_B T)} \quad (\text{Equation 9})$$

By combining discrete reactions with the ability to dynamically change polymerase initiation rates, we are able to simulate a wide range of phenomena. For example, a cooperative repressive interaction between some repressor protein  $R$  and a promoter could be modeled using a repressive Hill function:

$$r_{\text{initiation}}(R) = \frac{1}{1 + \left(\frac{R}{K}\right)^n} \cdot e^{-E_{sc}/(k_B T)}$$

More generally, we can use stochastic formulations of other regulatory mechanisms and test how these mechanisms behave in concert with supercoiling-mediated feedback.

### Boundary conditions

Key to our simulations is calculating the supercoiling density across the domain using Equation 2. For simulations using linear boundary conditions, we use the left and right edges as boundaries, assigning excess twist  $\varphi = 0$  at both boundary locations. Then, the supercoiling density can be defined between every polymerase. The location of the boundary conditions for the simulations is described in Table S3.

For simulations using circular boundary conditions, we must define how generated supercoiling “wraps around” the edges of the simulation. To do this, we choose an arbitrary origin, and order polymerases based on their (clockwise) position from the origin. In Table S3, the first boundary location is used as this origin location and the second boundary location is the length of the circle, relative to the origin. As in the linear case, when there are zero polymerases loaded, the supercoiling density is uniformly 0. In addition, for circular boundary conditions, the supercoiling density is also uniformly 0 when a single polymerase is present; when assuming fast supercoiling relaxation, a single polymerase on a circle can never accumulate negative or positive supercoiling.

For two or more polymerases, we create a list of excess twists, duplicating the endpoints as  $\varphi_n, \varphi_1, \dots, \varphi_n, \varphi_1$ . We additionally project the locations of the wrapped polymerases past the origin (e.g., the position of the wrap-around  $\varphi_n$  is placed at  $z_{\text{total length}} - z_n$ ), then compute the supercoiling density as in the linear case.

### Burst threshold choice

We calculated burst size and inter-burst time by using a burst threshold. From polymerase initiation times, we calculate the time between successive polymerase additions. Intra-addition times greater than the burst threshold form the boundaries between different expression bursts. We define the burst size to be the number of polymerases included in a burst, and the inter-burst time to be those intra-polymerase-addition times greater than the burst threshold. For the main text, we used a burst threshold of 30 s; specifically, this means that bursts ended if 30 s passed without a new polymerase being added.

In Figures S1D and S1E, we reanalyze the data presented in Figures 4C and 4D for different burst thresholds. We find that using a burst threshold of twenty or sixty seconds does not significantly affect the qualitative results observed. However, using a ten second burst threshold does dramatically shift the resulting burst size and inter-burst time distributions, with the inter-burst time distribution becoming concentrated around ten seconds. This indicates that a ten-second burst time is too short and incorrectly separates bursts.

### Toggle-switch parameter selection

The half-max value  $K$ , the mRNA count at which the promoter activity is half that of  $r_0$ , is chosen here to approximately match the mean steady-state expression of the steady states. The mean steady-state value is identified using simulations where only one of the toggle switch genes is enabled; this allows us to directly account for the influence of supercoiling-mediated feedback on the steady state mRNA concentration. With this choice of  $K$ , we ensure that the toggle switch operates in the regime of maximum sensitivity (e.g., the stable basin steady-state value is in the middle of the sigmoidal repression curve).

In Figure 5E, we tune the mRNA degradation rate, which directly impacts the mean steady-state value. If the mRNA degradation rate is doubled, we expect that the mean steady-state value should decrease to half its original value. To compare between these otherwise disparate conditions, we scaled the  $K$  value alongside the mRNA degradation rate, dividing by the fold increase in the mRNA degradation rate.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### ODE and stochastic simulation

The core ordinary differential equations were simulated using a Tsitouras's explicit Runge-Kutte 4-5 order method (Tsitouras, 2011). Normally, one implements stochastic simulations using a time-jumping method such as Gillespie's method. However, because the propensity of our stochastic events changes *continuously* with the continuous simulation, we need a stochastic solver that can be applied within the continuous integrator loop. Here, we used DifferentialEquations.jl, a performant Julia package that allows layered differential and stochastic equations (Rackauckas and Nie, 2017).

### Summarizing stochastic simulations

Stochastic simulations inherently sample from an underlying population distribution. As single simulations may not adequately represent the population behavior, we simulated ensembles of simulations and chose an ensemble size that was large enough to show the desired population behavior. These simulations then were post-processed, by either taking the average over each ensemble or by directly showing population distributions, summarized by a kernel density estimate (smoothed histograms provided by the Python package seaborn). The ensemble size for each data plot is summarized in Table S1.

### Computing cross-correlation

We computed the cross-correlation between the gene outputs following induction (Figure 3C), normalized by the geometric mean of the auto-correlation of these outputs. Strong negative or positive peaks evenly spaced around  $\tau = 0$  is a hallmark of periodic behavior, with the time offset of the peak encoding the phase offset between the signals. For two signals  $f(t)$  and  $g(t)$ , the normalized cross-correlation at a certain time offset  $\tau$  is bounded between  $\pm 1$  and can be thought of as the correlation coefficient between the unshifted version of one of the signals ( $f(t)$ ) and the other signal shifted in the time axis by  $\tau$  ( $g(t + \tau)$ ). Mathematically, we use:

$$\text{cross - correlation}(\tau) = \frac{(f \star g)(\tau)}{\sqrt{(f \star f)(0) \cdot (g \star g)(0)}} \quad (\text{Equation 10})$$

for

$$(f \star g)(\tau) = \sum_t (f(t) - \langle f(t) \rangle)(g(t + \tau) - \langle g(t) \rangle) \quad (\text{Equation 11})$$

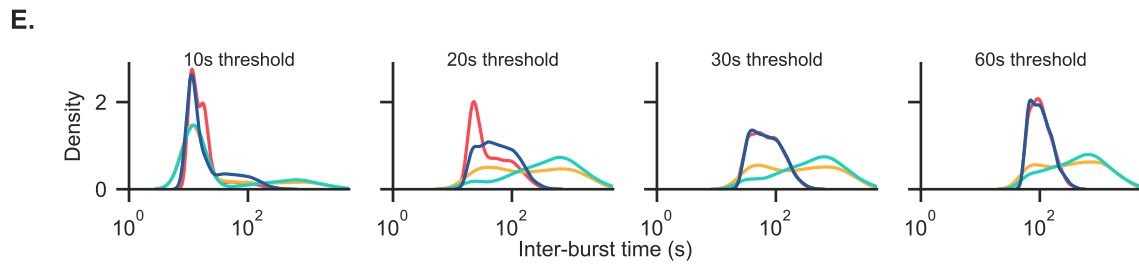
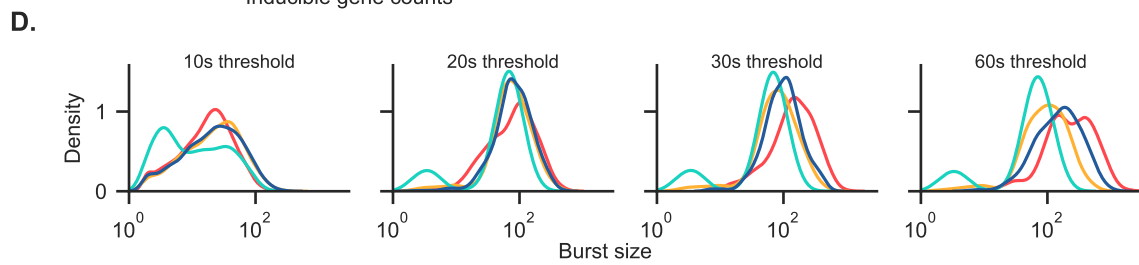
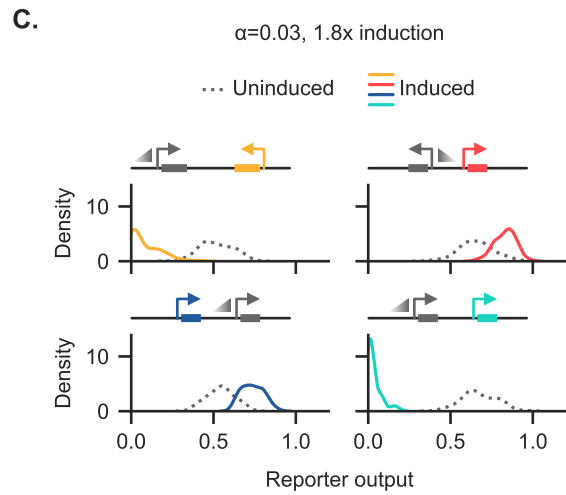
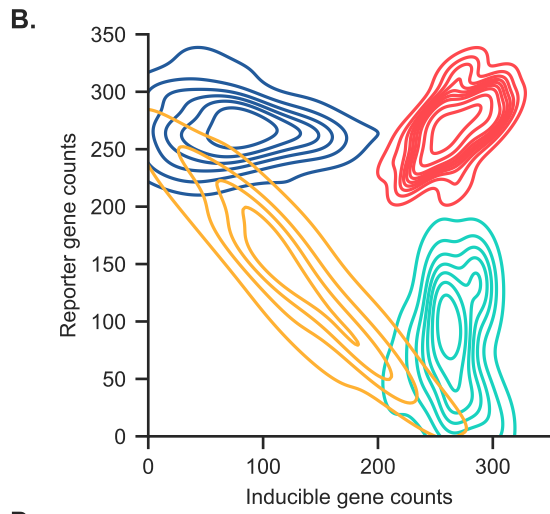
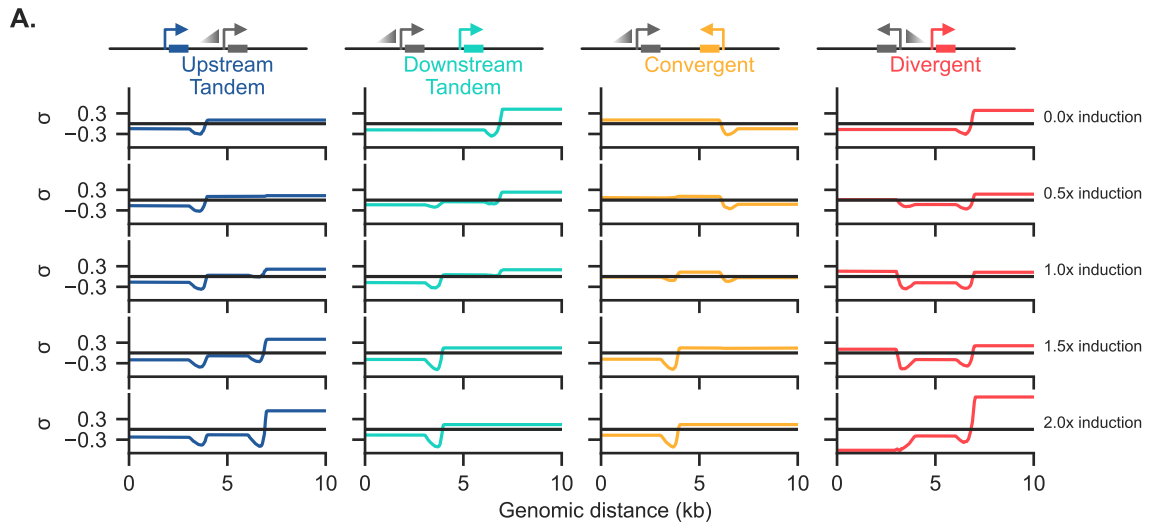
In fact, for  $\tau = 0$ , the normalized cross-correlation of the two signals is exactly the Pearson correlation coefficient. The shape of the cross-correlation curve can also reveal periodic and other time-dependent correlative behaviors.

**Cell Reports, Volume 41**

**Supplemental information**

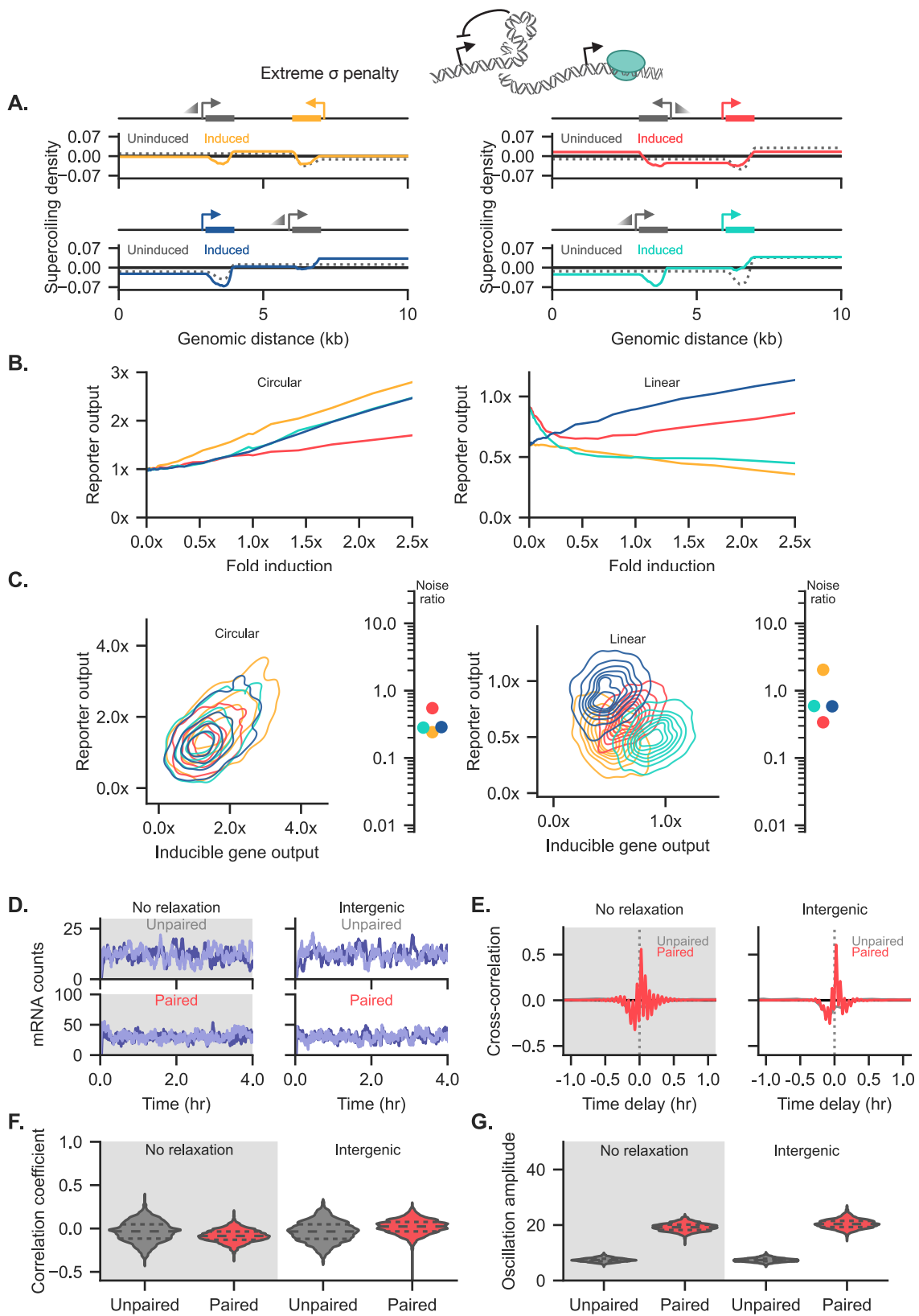
**Supercoiling-mediated feedback  
rapidly couples and tunes transcription**

**Christopher P. Johnstone and Kate E. Galloway**

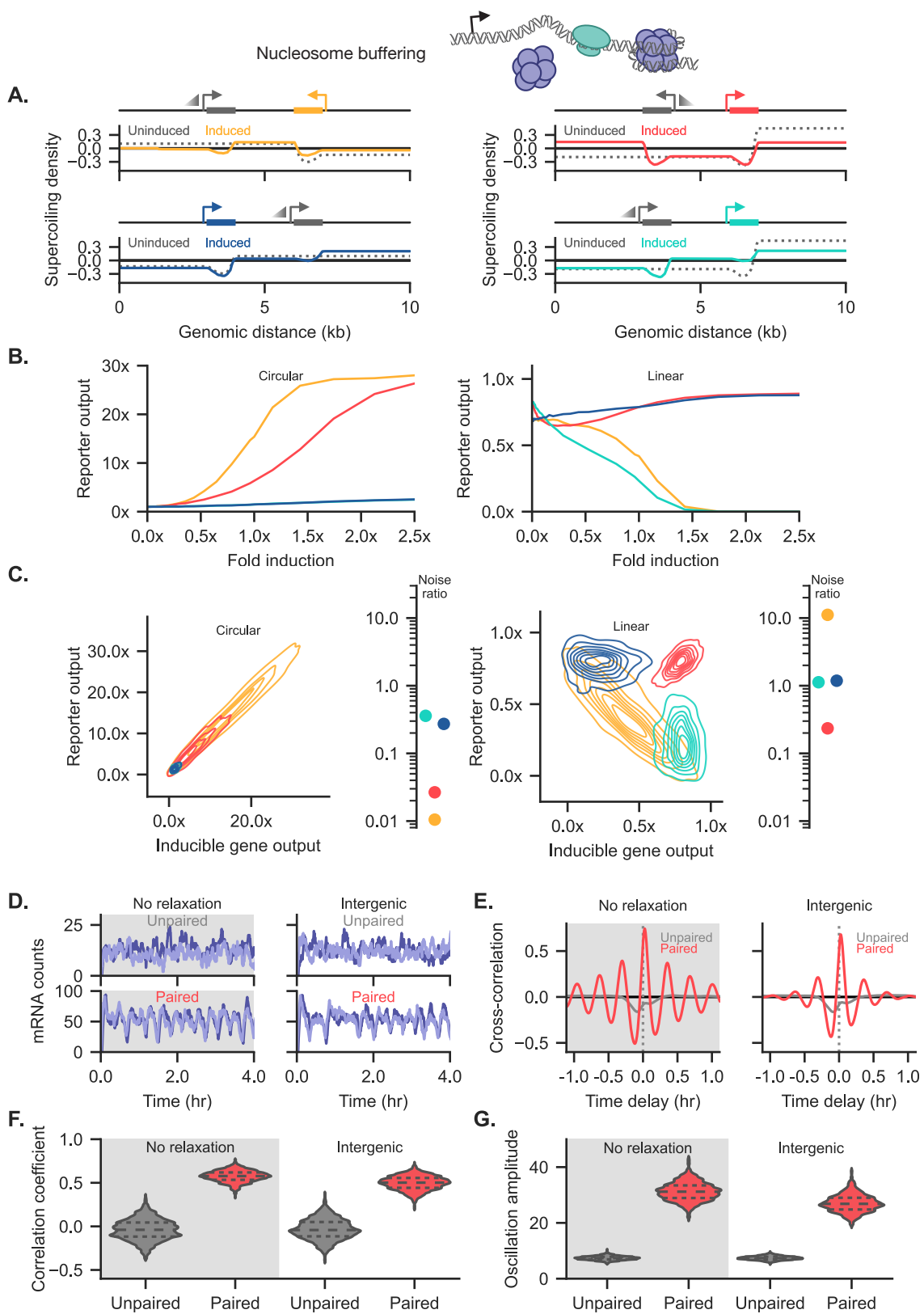


**Figure S1:** Related to figs. 2 to 4. **a)** Supercoiling density profiles are plotted for a wide range of adjacent gene induction levels across the four syntaxes considered. At low adjacent induction of the adjacent gene, positive and negative supercoiling mainly accumulates upstream and downstream, respectively, of the reporter gene. At 1-fold (equal base rates) induction, the supercoiling distributions are nearly symmetric in the convergent and divergent syntaxes. At high adjacent induction, the adjacent gene expression dominates the generated supercoiling profiles. **b)** The distribution of inducible and reporter mRNA counts are shown for the simulation ensembles presented in fig. 3. The different initial starting condition does not affect the final distribution state when compared to fig. 2d, indicating that the final state reached is independent of initial state. **c)** The ensemble distribution of the four linear circuit syntaxes is shown for  $\alpha = 0.03$  and high induction of the adjacent gene (1.8 fold). **d),e)** Burst size and inter-burst time distributions are shown at equal-induction as a function of threshold time and circuit syntax. At the lowest threshold time, the burst size and inter-burst time distributions qualitatively differ, with only a small percentage of bursts having an inter-burst time larger than 20 seconds. This lack of larger inter-burst times suggests that a burst threshold of at least 20 seconds gives useful predictions.

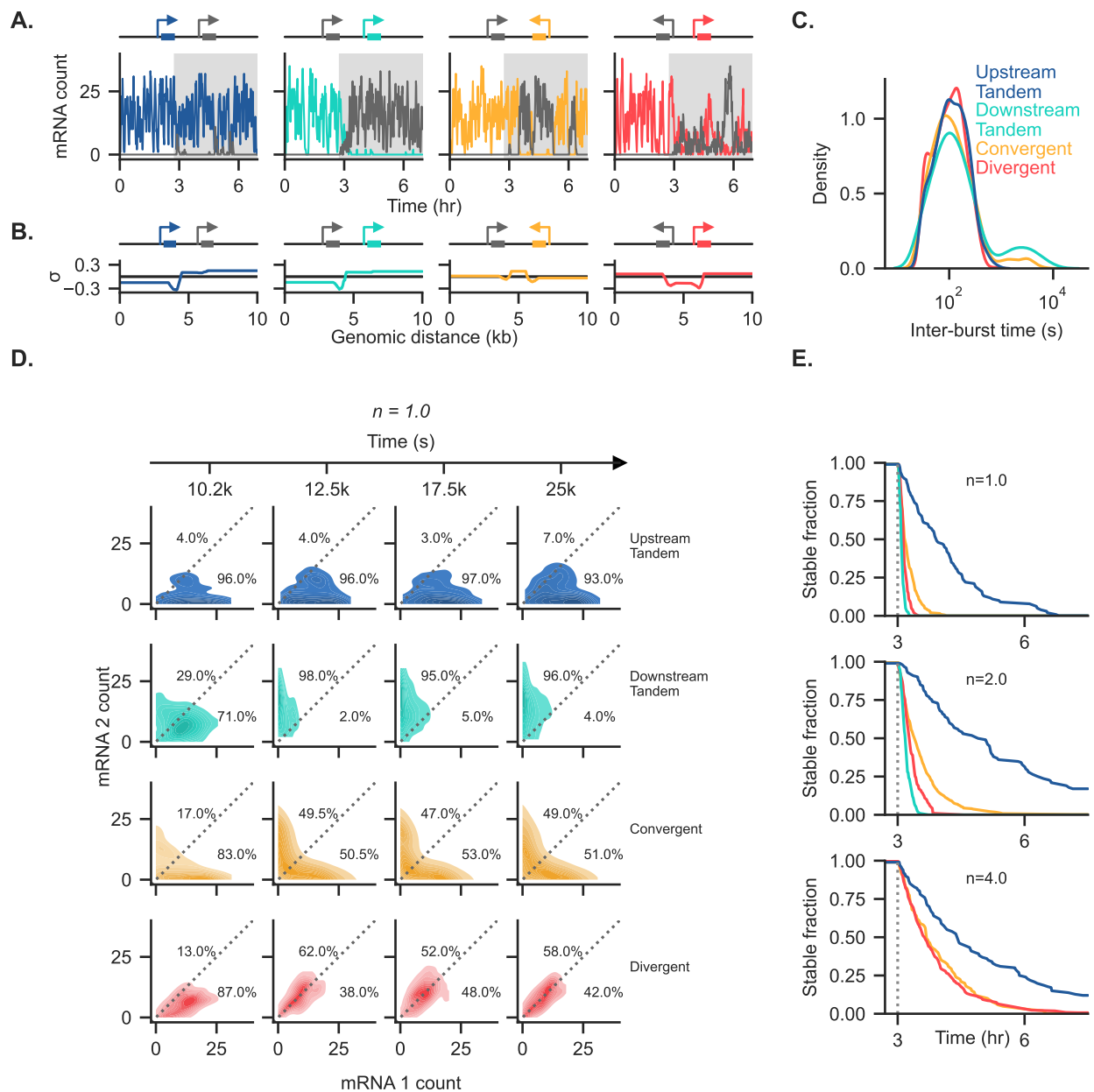




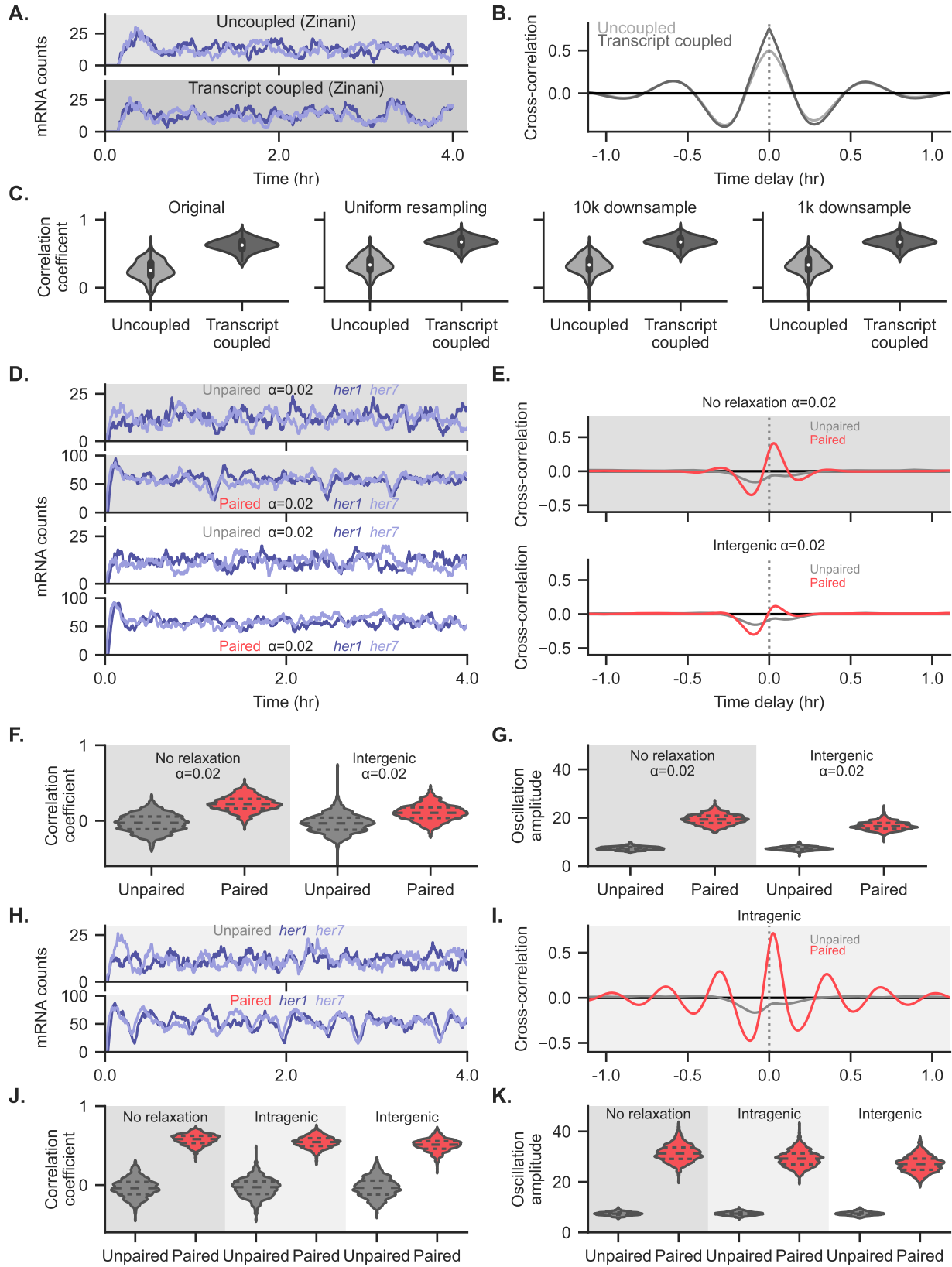
**Figure S2:** Related to Figures 2, 4 and 6. Simulation of an RNAP binding penalty at highly positive and negative supercoiling densities reduces syntax-specific differences and impedes transcriptional coupling between *her1* and *her7*. **a)** The average supercoiling density for both the induced and induced case is shown. The average supercoiling density is smaller than observed in figs. 3b and S3a due to the inhibition of polymerase initiation at extreme values of supercoiling density. **b)** Reporter output as a function of adjacent gene induction and circuit syntax is shown for both circular and linear boundary conditions. **c)** The population distribution is shown as a function of boundary condition and circuit syntax for the equally-induced case. The noise ratio of each is defined as the intrinsic noise (off-diagonal component) divided by the extrinsic noise (on-diagonal component). **d)** Four example simulation traces of the zebrafish segmentation gene network are shown, for two topoisomerase conditions and the gene-unpaired and gene-paired orientations. No clear periodicity is observed. **e)** The ensemble cross-correlation for the zebrafish segmentation gene network is shown. In this restricted polymerase initiation case, strong periodicity is not observed. **f)** The ensemble distribution of the correlation coefficient between *her1* and *her7* is shown. Restricting polymerase initiation eliminates *her1/her7* correlation, even in the gene-paired case. **g)** The ensemble oscillation amplitude is shown for each topoisomerase and gene pairing condition.



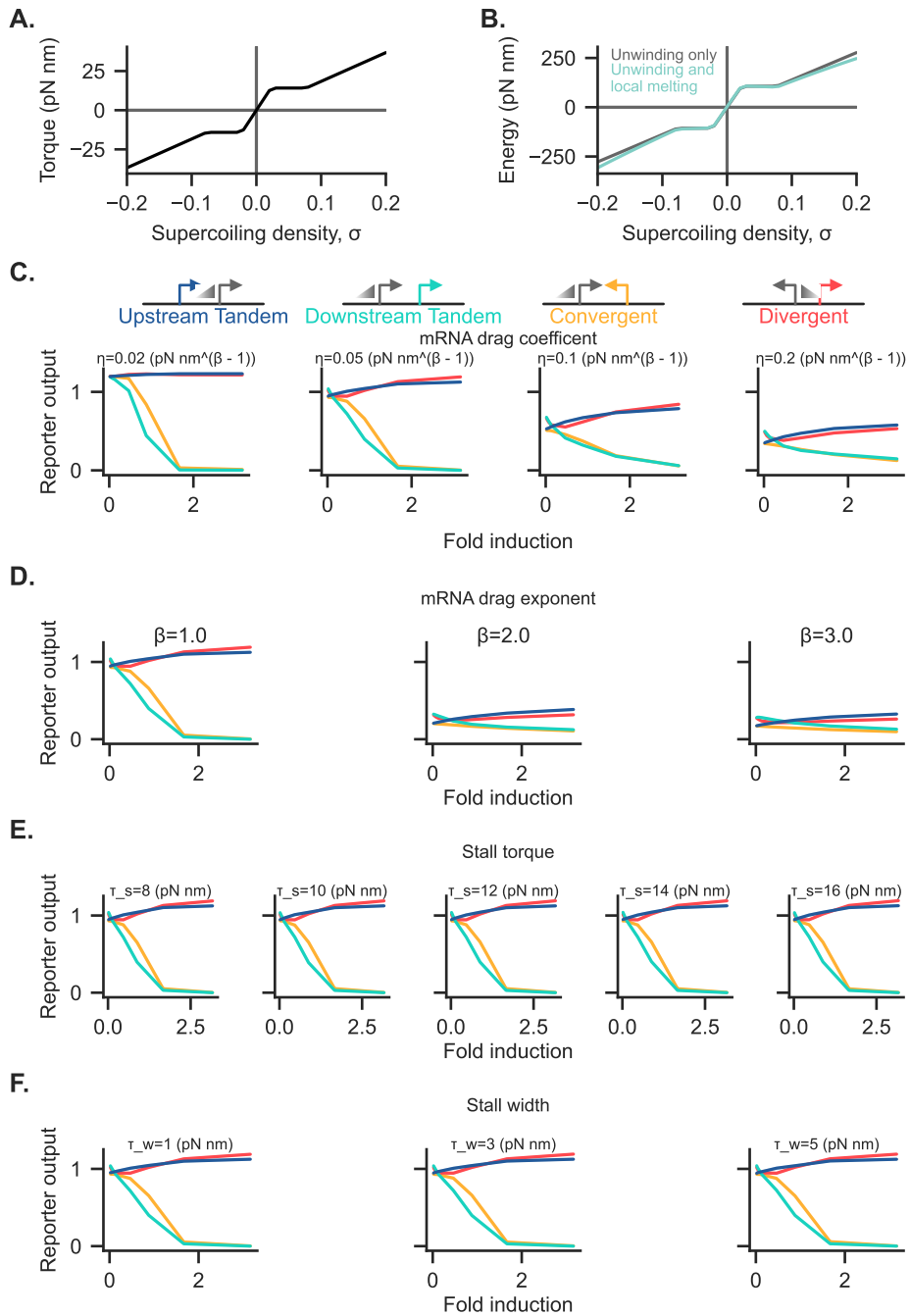
**Figure S3:** Related to Figures 2, 4 and 6. Nucleosome buffering of positive supercoiling displays weak effects on the expression profiles of two-gene systems at steady state and on the *her1-her7* somite network. **a)** The average supercoiling density for both the induced and induced case is shown. **b)** Reporter output as a function of adjacent gene induction and circuit syntax is shown for both circular and linear boundary conditions. **c)** The population distribution is shown as a function of boundary condition and circuit syntax for the equally-induced case. The noise ratio of each is defined as the intrinsic noise (off-diagonal component) divided by the extrinsic noise (on-diagonal component). **d)** Four example simulation traces of the zebrafish segmentation gene network are shown, for two topoisomerase conditions and the gene-unpaired and gene-paired orientations. **e)** The ensemble cross-correlation for the zebrafish segmentation gene network with nucleosome buffering is shown. Robust periodicity in the gene-paired case remains in the presence of nucleosomes. **f)** The ensemble distribution of the correlation coefficient between *her1* and *her7* is shown. In the presence of nucleosomes, the gene-paired case has a significantly increased correlation coefficient. **g)** The ensemble oscillation amplitude is shown for each topoisomerase and gene pairing condition.



**Figure S4:** Related to Figure 5. **a)** Four example simulation runs for the toggle switches are placed in the four orientations. **b)** The mean ensemble supercoiling density is plotted for the four different circuit syntaxes. **c)** The distribution of inter-burst times for the four different toggle switch architectures is shown. **d)** The ensemble mRNA count distributions are shown as a function of syntax at four selected time points for  $n = 1.0$ . Even as early as 200 seconds after induction of the second gene, we see that the tandem-downstream and divergent syntaxes are quickly approaching their equilibrium ensemble distributions. **e)** The basin stability for the four toggle-switch architectures is shown explicitly as a function of Hill coefficient,  $n$ .

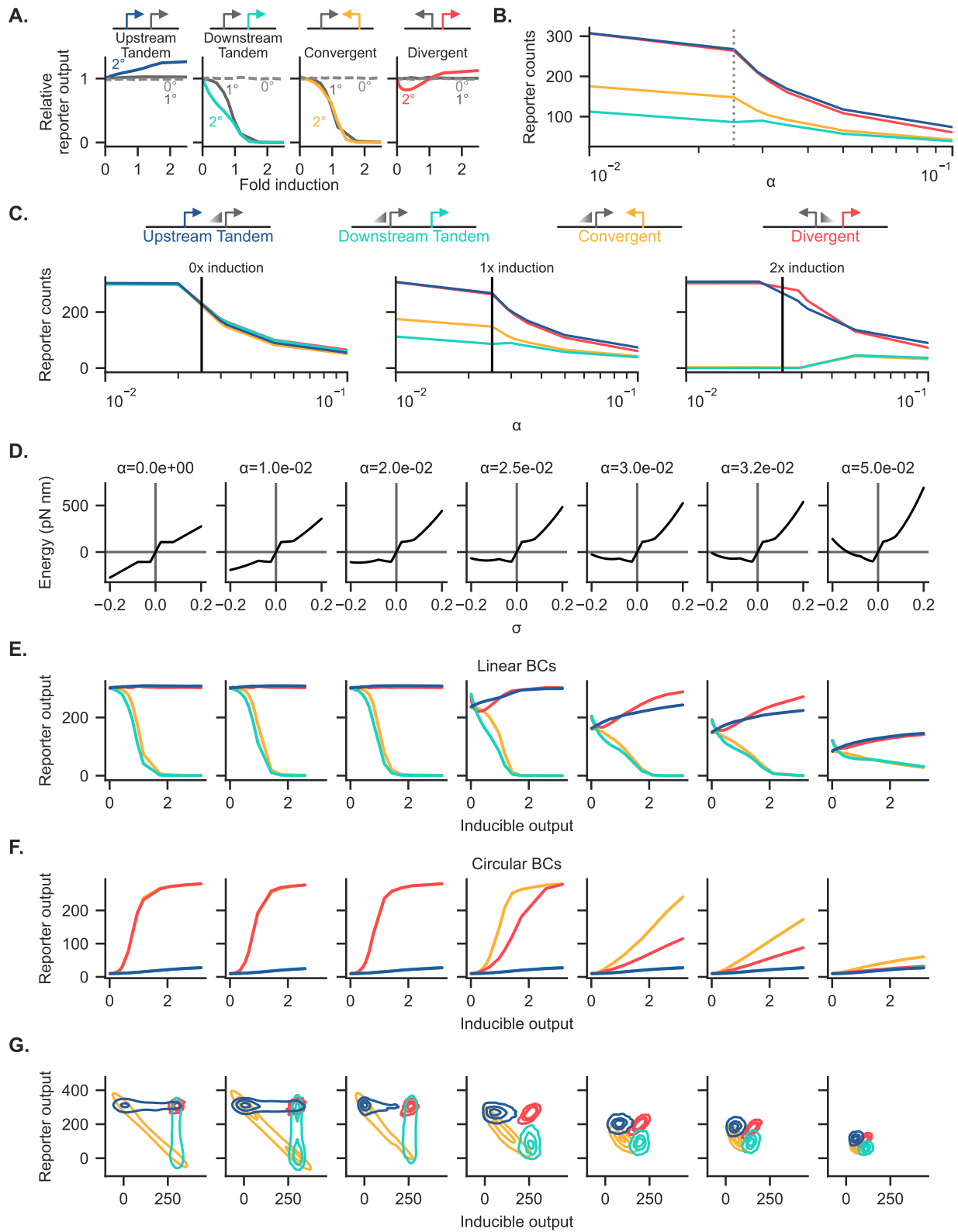


**Figure S5:** Related to Figure 6. **a)** Using the authors' MATLAB implementation of the model in Zinani et al. (2021), we simulated ensembles of 10,000 runs in order to compare to our simulation work. Unlike our implementation, which does not contain extra delays not due to polymerase motion and stalling, the original model uses a fixed mRNA production delay time of 9 minutes and a fixed protein production delay time of 1 minute. Example runs of the original model are shown. The transcript-coupled case shows correlated but relatively aperiodic behavior. **b)** The ensemble cross-correlation shows that the uncoupled and transcript coupled cases in the original model show very similar correlation structures, showing that both have similar amounts of periodicity in the original model. **c)** Using Zinani's original MATLAB code, we computed the correlation coefficient distributions using both their original sampling method (far-left) and with the dataset resampled uniformly in time. The qualitative results between these two sampling methods remain the same, with the original sampling method slightly under-reporting *her1-her7* correlation due to its reliance on Gillespie-sampled timepoints. We further found that down-sampling the uniformly-sampled dataset did not noticeably change the correlation distributions. **d)-f)** Re-simulation of the *her1-her7* system presented in fig. 6 is shown for  $\alpha = 0.02$ , a choice that occurs in the first regime of supercoiling-dependent initiation behavior. Strikingly, the strong correlated behavior seen in fig. 6 is absent at this choice of  $\alpha$ , suggesting that supercoiling-dependent initiation is an important phenomena in the synchronization of *her1* and *her7* activity. **d)** The number of *her1* and *her7* mRNAs are shown in the two different simulation contexts. **e)** For  $\alpha = 0.02$ , the biophysically-coupled case no longer shows strong periodic behavior in the cross-correlation plots, instead showing similar asymmetric cross-correlation behavior. **f)** In contrast to fig. 6e, for  $\alpha = 0.02$ , the gene paired system now shows a weaker correlated expression. **g)** The mean oscillation amplitude is shown for the gene unpaired and gene paired cases with intragenic topoisomerase relaxation. The gene paired case still shows an increase in average oscillation amplitude. **h)** Example mRNA traces are shown for the zebrafish segmentation network with intragenic topoisomerase relaxation. **i)** The cross correlation of clock gene expression is shown for intragenic topoisomerase relaxation. **j)** The correlation coefficient between *her1 her7* is shown, compared across the three different topoisomerase activity conditions. **k)** The size of oscillations is compared across the three different topoisomerase activity conditions.





**Figure S6:** Related to STAR Methods. Reporter output of two-gene circuits with linear boundary conditions (as in fig. 2) demonstrates similar qualitative behavior when underlying model parameters are changed. **a)** The torque response predicted by Marko's model (eq. (S1)) exhibits phase behavior, with a transition occurring between a over- and under-twisted phase to a largely plectonemic phase. During the transition, the torque response is constant. **b)** For the supercoiling-dependent initiation model, the energy that it takes to introduce supercoiling upstream and downstream by unwinding the DNA dominates the energy expression. Including a term representing the energy of local melting of the double helix at the location of the polymerase only gives a minor correction. **c)** Varying the nascent mRNA drag coefficient from the chosen value  $\eta = 0.05$  rescales but does not dramatically change reporter output behavior. Larger drag coefficients increase generated supercoiling by introducing additional torque into the system. **d)** Varying the exponent on the nascent mRNA length in the drag coefficient from the choice of linear drag  $\beta = 1.0$  also rescales but does not dramatically change reporter output behavior. Choices of  $\beta > 1.0$  tends to increase generated supercoiling due to longer nascent mRNAs having an increased drag profile. **e)** Varying the RNAP stall torque from the chosen  $\tau_s = 12$  pN nm has a minor effect on reporter output. **f)** Varying the width of the stalling function from the chosen  $\tau_w = 3$  pN nm has no noticeable effect on reporter output.



**Figure S7:** Related to STAR Methods. As the value of the quadratic energy correction coefficient  $\alpha$  varies, two regimes of supercoiling-dependent behaviors emerge. **a)** Reporter output is shown, normalized to the zero-fold induction case for the four tested syntaxes for the three different polymerase initiation models. **b)** For the second-order model, we chose the value of  $\alpha$  that defined the transition between regimes. **c)** For  $\alpha \leq 0.02$ , supercoiling-dependent initiation appears to dominate, leading to both the maintenance of reporter output in the linear divergent case and near-equality of the convergent and divergent cases with circular boundary conditions. For  $\alpha \geq 0.025$ , emergent non-monotonic behavior appears, leading to complex behaviors with linear boundary conditions and a separation between the response of the convergent and divergent syntaxes with circular boundary conditions. **d)** The additional polymerase binding energy as a function of supercoiling density is plotted for various values of  $\alpha$ . **e)** Reporter behavior of the four syntaxes under linear and circular boundary conditions are shown, in comparison to fig. 2b. **f)** Reporter behavior of the four syntaxes under linear and circular boundary conditions are shown, in comparison to fig. 2b. **g)** The overall distribution of the four syntaxes as a function of alpha is shown, in comparison to fig. 2d.

Plot description	$n$	Figures
Each point comprising the plotted curves is an average over $n$ simulations.	750	figs. <a href="#">S2b</a> and <a href="#">S3b</a>
	2000	figs. <a href="#">2b</a> , <a href="#">S6c</a> to <a href="#">S6f</a> , <a href="#">S7a</a> , <a href="#">S7c</a> , <a href="#">S7e</a> and <a href="#">S7f</a>
	5000	fig. <a href="#">2f</a>
	10000	figs. <a href="#">5c</a> , <a href="#">5e</a> and <a href="#">S4e</a>
Each line shows the average supercoiling density profile over an ensemble of $n$ simulations.	100	figs. <a href="#">2e</a> , <a href="#">S1a</a> and <a href="#">S4b</a>
	250	figs. <a href="#">3b</a> , <a href="#">S2a</a> and <a href="#">S3a</a>
The standard deviation of gene output for an ensemble of $n$ simulations is shown as a function of time.	250	fig. <a href="#">3f</a>
The average cross-correlation over an ensemble of $n$ simulations is plotted.	250	fig. <a href="#">3d</a>
	2000	figs. <a href="#">6d</a> , <a href="#">S2e</a> , <a href="#">S3e</a> , <a href="#">S5e</a> and <a href="#">S5i</a>
	10000	fig. <a href="#">S5b</a>
Distribution profiles are shown over an ensemble of $n$ simulations.	100	fig. <a href="#">4c</a> (base model), fig. <a href="#">4d</a> (base model), figs. <a href="#">S1b</a> , <a href="#">S1d</a> , <a href="#">S1e</a> and <a href="#">S4c</a>
One-dimensional distributions are shown as a kernel density estimate (smoothed histogram) or violin plot.	250	figs. <a href="#">3e</a> and <a href="#">S1c</a> , fig. <a href="#">4c</a> (model perturbations), fig. <a href="#">4d</a> (model perturbations)
	750	figs. <a href="#">S2c</a> and <a href="#">S3c</a>
Two-dimensional distributions are shown as filled or unfilled isocurves.	2000	figs. <a href="#">2c</a> , <a href="#">2d</a> , <a href="#">6e</a> , <a href="#">6f</a> , <a href="#">S2f</a> , <a href="#">S2g</a> , <a href="#">S3f</a> , <a href="#">S3g</a> , <a href="#">S5f</a> , <a href="#">S5g</a> , <a href="#">S5j</a> , <a href="#">S5k</a> and <a href="#">S7g</a>
	10000	figs. <a href="#">5b</a> , <a href="#">5d</a> , <a href="#">S4d</a> and <a href="#">S5c</a>
An example simulation was randomly selected from an ensemble of $n$ simulations and plotted.	250	fig. <a href="#">3a</a>
	2000	figs. <a href="#">6c</a> , <a href="#">S2d</a> , <a href="#">S3d</a> , <a href="#">S5d</a> and <a href="#">S5h</a>
	10000	figs. <a href="#">S4a</a> and <a href="#">S5a</a>

**Table S1:** Key ensemble size and interpretive description of each plot. Related to STAR Methods.

Drag coefficient $\eta$ (pN nm <sup><math>\beta-1</math></sup> )	Drag exponent $\beta$	Stall torque $\tau_s$ (pN nm)	Stall width $\tau_w$ (pN nm)	Figures
1/20	1	12	3	All except figs. S6c to S6f
1/50 $\leftrightarrow$ 1/5	1	12	3	Figure S6c
1/20	1 $\leftrightarrow$ 3	12	3	Figure S6d
1/20	1	8 $\leftrightarrow$ 16	3	Figure S6e
1/20	1	12	1 $\leftrightarrow$ 5	Figure S6f

**Table S2:** Key dynamic constants for all simulations. Related to STAR Methods.

Type	Boundary locations	Reporter gene	Inducible gene	Figures
Upstream tandem	0, 10000	3000 $\rightarrow$ 4000	6000 $\rightarrow$ 7000	Figures 2b to 2e, 3, 4, S1a to S1e, S2a to S2c, S3a to S3c, S6c to S6f, S7 and S7a
Downstream tandem	0, 10000	6000 $\rightarrow$ 7000	3000 $\rightarrow$ 4000	
Convergent	0, 10000	7000 $\rightarrow$ 6000	3000 $\rightarrow$ 4000	
Divergent	0, 10000	6000 $\rightarrow$ 7000	4000 $\rightarrow$ 3000	
Type	Boundary locations	Reporter gene	Inducible gene	Figures
Upstream tandem	0, 8000 + $\Delta x$	3000 $\rightarrow$ 4000	4000 + $\Delta x$ $\rightarrow$ 5000 + $\Delta x$	Figure 2f
Downstream tandem	0, 8000 + $\Delta x$	4000 + $\Delta x$ $\rightarrow$ 5000 + $\Delta x$	3000 $\rightarrow$ 4000	
Convergent	0, 8000 + $\Delta x$	5000 + $\Delta x$ $\rightarrow$ 4000 + $\Delta x$	3000 $\rightarrow$ 4000	
Divergent	0, 8000 + $\Delta x$	4000 + $\Delta x$ $\rightarrow$ 5000 + $\Delta x$	4000 $\rightarrow$ 3000	
Type	Boundary locations	Gene A	Gene B	Figures
Upstream tandem	0, 10000	3500 $\rightarrow$ 4500	5500 $\rightarrow$ 6500	Figures 5 and S4c to S4e
Downstream tandem	0, 10000	5500 $\rightarrow$ 6500	3500 $\rightarrow$ 4500	
Convergent	0, 10000	3500 $\rightarrow$ 4500	6500 $\rightarrow$ 5500	
Divergent	0, 10000	4500 $\rightarrow$ 3500	5500 $\rightarrow$ 6500	
Type	Boundary locations	<i>her1</i>	<i>her7</i>	Figures
Gene-unpaired	0, 10 <sup>9</sup> (free ends)	6405 $\rightarrow$ 6017	(10 <sup>9</sup> - 9864) $\rightarrow$ (10 <sup>9</sup> - 8549)	Figures 6, S2d to S2g, S3d to S3g and S5d to S5k
Gene-paired	0, 34393	12422 $\rightarrow$ 6017	24529 $\rightarrow$ 25844	

**Table S3:** Simulated circuit locations and distances. Locations are all given in basepairs (with 0.34 nanometers per base pair). For the given genes, the first number given is the simulated promoter location and start of the gene body, and the second number is a simulated termination site. Related to STAR Methods.

## Methods S1: Model derivation

Our model builds upon several prior papers and relevant mathematical frameworks. To ensure access and provide definitions needed for implementation of our model, we have aggregated all of these relevant equations and simulation constants together. From this foundation, we discuss how we have extended our model to examine how DNA supercoiling generates coupling between proximal genes.

### Literature supercoiling model

Why do we care about supercoiling at all? From statistical mechanics, the thermal energy unit  $k_B T$ , defined by the product of the Boltzmann constant and the temperature of the system, defines an energy accessible by fluctuations; useful chemical reactions and interactions often occur above this fluctuation energy. Converting into units relevant to supercoiling at  $T = 300$  K,  $k_B T = 4.1$  pN nm, Many of the later quantities we will solve for will have magnitude around 10-100 pN nm—well above the thermal energy—so can determining system behavior.

When we derived the supercoiling density,  $\sigma$ , in eq. (2), we defined  $\phi$  to be the local excess twist. While this term is most easily identified with the local DNA excess twist, it is truly a linking number constraint that takes into account both twist, local stretching of bonds that causes more or fewer complete rotations to occur over a set distance, and writhe, where mesoscopic plectonemes store supercoiling. These two forms of supercoiling are physically interconvertable, but are not energetically identical. This energetic difference is accounted for in the underlying statistical mechanical model (fig. S6a).

The statistical mechanics model depends on the following constants:

Variable	Value	Reference
$k_b$ , Boltzmann constant	$0.01381 \frac{\text{pN nm}}{\text{K}}$	-
$\omega_0$ , relaxed DNA twist frequency	$1.85 \frac{1}{\text{nm}}$	(Stuart A Sevier and Levine 2018), others
A, bend persistence length	50nm	(Marko 2007)
C, twist persistence length (twisted)	$95 \pm 10\text{nm}$	(Marko 2007)
P, twist persistence length (plectonemic)	$24 \pm 3\text{nm}$	(Marko 2007)
$f$ , applied force on DNA	1 pN	(Stuart A Sevier and Levine 2018)

**Table S4:** Key values for the statistical mechanics model

Especially key measurable constants are the bend persistence length  $A$ , the twist persistence length of stretched DNA  $C$ , and the twist persistence length of the plectonemic state  $P$ . Because plectonemes convert twist into writhe, we expect  $P < C$ .

Marko defines the following rescaled versions of  $P$  and  $C$ ,  $p, c$ :

$$c = k_B T C \omega_0^2 \quad p = k_B T P \omega_0^2$$

which in turn can be used to define helper constants  $g, c_s$ :

$$g = f - \sqrt{\frac{k_B T f}{A}} \quad c_s = c \left[ 1 - \frac{C}{4A} \sqrt{\frac{k_B T}{A f}} \right]$$

With these defined, Marko's statistical-mechanics model calculates the torque as a function of supercoiling density (fig. S6a) (Marko 2007):

$$\tau(\sigma) = \begin{cases} \tau_s \sigma & |\sigma| < |\sigma_s| \\ \tau_0 \text{sgn}(\sigma) & |\sigma_s| \leq |\sigma| \leq |\sigma_p| \\ \tau_p \sigma & |\sigma_p| \leq |\sigma| \end{cases} \quad (\text{S1})$$

for critical supercoiling values  $\sigma_s, \sigma_p$  separating the phase regimes:

$$|\sigma_s| = \frac{1}{c_s} \sqrt{\frac{2pg}{1-\frac{p}{c_s}}} \quad |\sigma_p| = \frac{1}{p} \sqrt{\frac{2pg}{1-\frac{p}{c_s}}} \quad (\text{S2})$$

with corresponding critical torque values  $\tau_s, \tau_p$ :

$$\tau_s = \frac{c_s}{\omega_0} \quad \tau_0 = \sqrt{\frac{2pg}{\omega_0^2(1-\frac{p}{c_s})}} \quad \tau_p = \frac{p}{\omega_0} \quad (\text{S3})$$

Using the torque equation eq. (S1), Marko expands the free energy of the DNA per unit length,  $S$ , as a power series in  $\sigma$ :

$$S = \begin{cases} -g + \frac{1}{2}c_s\sigma^2 & |\sigma| < |\sigma_s| \\ \frac{-g}{1-\frac{p}{c_s}} + \sqrt{\frac{2pg}{1-\frac{p}{c_s}}}|\sigma| & |\sigma_s| < |\sigma| < |\sigma_p| \\ \frac{1}{2}p\sigma^2 & |\sigma| > |\sigma_p| \end{cases} \quad (\text{S4})$$

While the cubic term can also be calculated and accounts for asymmetry between under- and over-winding, Marko (2007) argues that this effect is small for realistic physiological conditions.

### Literature dynamic model

Following Stuart A Sevier and Levine (2018), we adopt a dynamics model for modeling the motion of polymerases over the DNA. This model depends on the following constants:

Variable	Value	Reference
$\chi$ , DNA twist mobility	0.01381 s pN nm	(Stuart A Sevier and Levine 2018)
$\eta$ , mRNA drag coefficient	1/20 pN	(Stuart A Sevier and Levine 2018)
$n$ , Drag scaling exponent	1	(Stuart A Sevier and Levine 2018)
$v_0$ , Maximum polymerase velocity	20 $\frac{\text{nm}}{\text{s}}$	(Stuart A Sevier and Levine 2018)
$\tau_s$ , Polymerase stall torque	12 pN nm	(Stuart A Sevier and Levine 2018)
$\Delta\tau_s$ , Polymerase stall width	3 pN nm	This work
$\delta$ , Polymerase radius	15 nm	(Stuart A Sevier and Levine 2018)

**Table S5:** Key dynamic model values

Briefly, for each polymerase, this model tracks the position  $z_i$ , transcript length  $x_i$ , and excess DNA twist  $\phi_i$  of each RNAP. We create a *linking number constraint* by specifying  $\phi$  at any genomic location, which means that each RNAP sets an linking number constraint. We use dynamic equations to model the motion of RNAP against either fixed or free boundary conditions.

As mentioned in the main text, the dynamic equations are:

$$\omega_0 \frac{dz_i}{dt} = \frac{d\theta_i}{dt} + \frac{d\phi_i}{dt} \quad \tau(z_i, \phi_{i-1}, \phi_{i+1}) = \eta x_i^n \frac{d\theta_i}{dt} - \chi \frac{d\phi_i}{dt}$$

By solving the left equation for  $\frac{d\theta_i}{dt}$ , the polymerase angular velocity, and substituting it into the right equation, we get:

$$\tau(z_i, \phi_{i-1}, \phi_{i+1}) = \omega_0 \frac{dz_i}{dt} \eta x_i^n - \eta x_i^n \frac{d\phi_i}{dt} - \chi \frac{d\phi_i}{dt}$$

Solving for  $\frac{d\phi_i}{dt}$ , we get a dynamic equation for the differential linking number constraint:

$$\frac{d\phi_i}{dt} = \omega_0 \frac{dz_i}{dt} \frac{\eta x_i^n}{\chi + \eta x_i^n} - \frac{\tau(z_i, \phi_{i-1}, \phi_{i+1})}{\chi + \eta x_i^n} \quad (\text{S5})$$

To model polymerase stalling, we use a parameterized function discussed in eq. (5) that explicitly depends on a stall torque and a torque width  $\tau_w$  over which stalling begins to occur.

Equations B.2 and 5 along with the definition of linear velocity  $\frac{dx_i}{dt} = \frac{dz_i}{dt}$  together represent a complete coupled set of ODEs that can be simulated, as long as the following constants are defined:

### Model extension: supercoiling-dependent initiation

Reflecting on the desired properties of a model based on literature observations, we extended the models from Sevier and Marko to account for supercoiling-dependent initiation. The rest of this appendix details this process, in addition to the initial model formulation and useful simplifications.

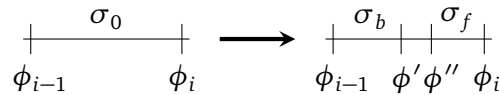
#### Literature observations and modeling goals

A naked DNA bead experiment (Revyakin, Ebright, and Strick (2004)) found that positive and negative supercoiling affected RNAP polymerase binding rates, with negative supercoiling encouraging binding and positive supercoiling discouraging binding. When a polymerase bound, approximately 1.2 turns of DNA unwound (13 base pairs, or 4.42nm). For a typical bacterial promoter, RNAP binding occurred reversibly. The mean time between RNAP binding events was measured and related to a kinetic model. The mean time between binding events defines a promoter-on rate. Once the DNA torque reaches the constant-torque regime, the promoter-on rate becomes constant.

Ideally, our model would replicate these results, with negative supercoiling promoting RNAP binding and negative supercoiling inhibiting RNAP binding, but would only depend on the local supercoiling density. Successful phenomenological models such as El Houdaigui et al. (2019) often use sigmoidal curves that are most sensitive in the range from  $\sigma = -0.1$  to  $\sigma = 0.1$ ; ideally, our model would be sensitive in a similar regime.

#### Dual linking number constraint model

To model the unwinding induced by RNAP binding, we add two new linking number constraints, spaced 13 base pairs apart, and analyze the energy required to perform this unwinding. To start, all of these have specific linear locations  $z$  and rotations  $\phi$ ; we will relax this restriction later.



**Figure S8:** Diagram of the addition of two linking number constraints

We then put an angular restriction on the two new linking number constraints. By subtracting off the excess angle  $\phi$  that the undisturbed DNA has prior to RNAP binding, we can define:

$$\Delta\phi' = \phi' - \left( \phi_{i-1} + \frac{z' - z_{i-1}}{z_i - z_{i-1}} (\phi_i - \phi_{i-1}) \right) \quad (\text{S6})$$

$$\Delta\phi'' = \phi'' - \left( \phi_{i-1} + \frac{z'' - z_{i-1}}{z_i - z_{i-1}} (\phi_i - \phi_{i-1}) \right) \quad (\text{S7})$$



For unwinding of the region bound to the RNAP to occur, we impose two constraints that ensure that the 13 base pairs that interact with the RNAP are fully unwound:

$$|\Delta\phi'| + |\Delta\phi''| = 1.2 \cdot 2\pi \quad (\text{S8})$$

$$\Delta\phi' - \Delta\phi'' = 1.2 \cdot 2\pi \quad (\text{S9})$$

For the intermediate region, this implies that there are two different ways to unwind the DNA; either to rotate the leading linking number constraint backwards or rotate the trailing linking number constraint forwards. There are also many intermediate solutions where *both* of the bounding linking number constraints move. The first restriction is there to ensure that we use the solution that minimizes total amount of rotation needed.

Given a complete energy model that related  $\tau$  over the complete range of unwinding  $\Delta\phi', \Delta\phi''$  values, we could directly write the binding energy as

$$\Delta E = \bar{\tau}' \Delta\phi' + \bar{\tau}'' \Delta\phi'' \quad (\text{S10})$$

and minimize the energy cost with respect to  $\Delta\phi'$  or  $\Delta\phi''$ , giving a unique solution for any specific geometry. As Marko's statistical mechanical model is likely not valid in the limit of complete unwinding, we instead move forward with this model by estimating the energy cost in both the "exterior" region and within the "interior", 13-bp region.

#### *Energy estimation in the exterior region*

If we assume that the energy surface is locally linear (e.g. in the exterior region, there is a small change in supercoiling density), then we can relate the instantaneous torque to the change in energy:

$$\tau = \frac{1}{\omega_0} \frac{\partial S(\sigma)}{\partial \sigma}$$

for  $S$ , the energy per unit length. This means that:

$$\Delta E = (z_2 - z_1) \tau \omega_0 \Delta\sigma = (z_2 - z_1) \frac{\partial S}{\partial \sigma} \Delta\sigma \quad (\text{S11})$$

By expanding the definition of  $\Delta\sigma$  for an arbitrary region bounded by two linking number constraints, we can actually recover the normal energy-torque definition:

$$\Delta E_{\text{external}} = (z_2 - z_1) \tau \omega_0 \frac{\Delta\phi_2 - \Delta\phi_1}{\omega_0(z_2 - z_1)} = \tau(\Delta\phi_2 - \Delta\phi_1) \quad (\text{S12})$$

Note that this implies that the energy cost of linking number constraint rotation is independent of the region width; under a locally linear assumption, the energy cost depends only on the local instantaneous torque and the rotation angle. When is this valid? The change in supercoiling density in the exterior region is relatively low,  $O(0.05)$  as long as the region of interest is at least  $250bp$  away from the nearest linking number constraint on one of its sides. For promoter regions analyzed here, space between gene bodies ensures that this assumption holds.

#### *Energy estimation in the interior region (local melting)*

We can generally write the energy it takes to unwind the DNA in the interior region as an integral over the linear free energy density,  $S$ :

$$\Delta E_{\text{internal}} = \int_{\sigma_0}^{\sigma_0 + \Delta\sigma_{\text{RNAP}}} \Delta z_{\text{RNAP}} \frac{\partial S}{\partial \sigma} d\sigma$$

Given the well-defined structure of the RNAP-DNA complex and the fact that under all realistic physiological conditions  $\Delta\sigma_{\text{RNAP}} \gg \sigma_0$ , we assume that the end-state energy is a constant, but unknown  $S_{\text{unwound}}$ :

$$\Delta E_{\text{internal}} = \Delta z_{\text{RNAP}} (S_{\text{unwound}} - S(\sigma_0)) \quad (\text{S13})$$

While  $S_{\text{unwound}}$  is unknown, it is a constant with respect to all promoters, so this energetic term is implicitly already included in any promoter base rate. This means that the external energy depends only on the free energy density, evaluated at the local, initial supercoiling density  $\sigma_0$ :

$$\Delta E_{\text{internal}} = -\Delta z_{\text{RNAP}} S(\sigma_0) \quad (\text{S14})$$

### Order of magnitude energy analysis

Combining the results of eqs. (S12) and (S14) and plugging in eq. (S9), we have an estimated binding energy:

$$\Delta E = \tau(\sigma_0) 1.2 \cdot 2\pi - \Delta z_{\text{RNAP}} S(\sigma_0) \quad (\text{S15})$$

When we plot this energetic term in fig. S6b, we see that the internal energetic term is minimal when compared to the external energetic term (local melting) at the supercoiling densities considered here, so we use the simplified form

$$\Delta E = \tau(\sigma_0) \cdot 1.2 \cdot 2\pi \quad (\text{S16})$$

as presented in eq. (6).

### Implementing the zebrafish clock gene network

In order to implement the zebrafish clock gene network, we used the stochastic reaction network and rate constants presented by Zinani et al. (2021). In brief, we add additional stochastic protein-production and protein-degradation reactions that correspond to the production of *her1* and *her7*. The protein production reactions are first-order with respect to the corresponding mRNA concentrations and the protein degradation reactions are first-order with respect to the protein concentration, and take the form:

$$r_{\text{her1 production}} = k_{\text{protein production}} \cdot [\text{her1 mRNA}] \quad (\text{S17})$$

$$r_{\text{her7 production}} = k_{\text{protein production}} \cdot [\text{her7 mRNA}] \quad (\text{S18})$$

$$r_{\text{her1 degradation}} = k_{\text{protein degradation}} \cdot [\text{her1}] \quad (\text{S19})$$

$$r_{\text{her7 degradation}} = k_{\text{protein degradation}} \cdot [\text{her7}] \quad (\text{S20})$$

As in Zinani et al. (2021), we assume the starting amount of the cofactor *hes6* is 100 proteins.

The *her1* and *her7* promoters can then be in a bound or unbound state by either the *her1-her1* homodimer or the *hes6-her7* heterodimer. Zinani et al. (2021) use a single combined dimerization and binding equation, with rates:

$$r_{\text{promoter binding (11)}} = k_{\text{dimer association}} [\text{unbound promoter}] \frac{[\text{unbound her1}][(\text{unbound her1}) - 1]}{2} \quad (\text{S21})$$

$$r_{\text{promoter unbinding (11)}} = k_{\text{dimer disassociation}} [\text{bound promoter}_{11}] \quad (\text{S22})$$

$$r_{\text{promoter binding (67)}} = k_{\text{dimer association}} [\text{unbound promoter}] [\text{unbound hes6}] [\text{unbound hes7}] \quad (\text{S23})$$

$$r_{\text{promoter unbinding (67)}} = k_{\text{dimer disassociation}} [\text{bound promoter}_{67}] \quad (\text{S24})$$

Note that the bound/unbound promoter species variables take on values zero or one, for each promoter. For example, if the *her1* promoter is currently bound by a *her1-her1* homodimer, then the unbound variable [unbound *her1* promoter] = 0, but the bound variables take values [bound *her1*<sub>11</sub>] = 1, [bound *her1*<sub>67</sub>] = 0.