bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Collective polymerase dynamics emerge from DNA Supercoiling Mediated Transcription

Stuart A. Sevier [1,2], Sahand Hormoz [1,2,3*]
[1] *Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA*
[2] *Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA 02215, USA*
[3] *Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA*

All biological processes ultimately come from physical interactions. The mechanical properties of DNA play a critical role in transcription. RNA polymerase can over or under twist DNA (referred to as DNA supercoiling) when it moves along a gene resulting in mechanical stresses in DNA that impact its own motion and that of other polymerases. For example, when enough supercoiling accumulates, an isolated polymerase halts and transcription stops. DNA supercoiling can also mediate non-local interactions between polymerases that shape gene expression fluctuations. Here, we construct a comprehensive model of transcription that captures how RNA polymerase motion changes the degree of DNA supercoiling which in turn feeds back into the rate at which polymerases are recruited and move along the DNA. Surprisingly, our model predicts that a group of three or more polymerases move together at a constant velocity and sustain their motion (forming what we call a polymeton) whereas one or two polymerases would have halted. We further show that accounting for the impact of DNA supercoiling on both RNA polymerase recruitment and velocity recapitulates empirical observations of gene expression fluctuations. Finally, we propose a mechanical toggle switch whereby interactions between genes are mediated by DNA twisting as opposed to proteins. Understanding the mechanical regulation of gene expression provides new insights into how endogenous genes can interact and informs the design of new forms of engineered interactions.

PACS numbers:

Numerous physical processes contribute to gene expression [1]. For example, transcription, an essential step in gene expression where DNA is converted to RNA, can be thought of as a mechanical process. During transcription, an RNA polymerase can twist DNA and generate stresses on DNA [2, 3] that impact its own motion and that of other polymerases [4] affecting gene expression [5]. Therefore, to understand the dynamics of gene expression, we need to account for the mechanical nature of transcription [6].

Gene expression occurs in a stochastic ('bursty') manner [7]. Early insights into gene expression fluctuations used simple phenomenological frameworks that did not account for DNA mechanics or the physical forces involved in transcription [8–12]. These models were followed by theoretical [13–15] and experimental studies [16–18] that found over and under twisting of DNA, or DNA supercoiling, to be a robust mechanism of generating transcriptional bursting. In addition to influencing fluctuations, experimental [19] and theoretical [13] observations have shown that DNA twisting can halt isolated polymerases and stop transcription whereas multiple polymerases can undergo effective elongation [4]. Despite the advances that these studies have made, we still lack a coherent framework that captures how DNA twisting by RNA polymerase during transcription feeds back into the recruitment and motion of other RNA polymerases which in turn change the degree of twisting of DNA.

Here, we construct a model of mechanical aspects of gene expression (referred to as mechanical epigenetics) that captures the feedback cycle between DNA twisting and RNA polymerase recruitment and motion. Within our framework multiple polymerases interact non-locally via twisting of DNA generated by the movement of the polymerases along the gene. We show that these interactions play a key role in setting the average velocity at which polymerases can move along a gene. Surprisingly, we demonstrate that three or more interacting polymerases undergo sustained motion whereas isolated polymerases are halted by mechanical forces (a collective phenomenon that we call a polymeton). We also show that incorporating the impact of DNA twisting on both recruitment of RNA polymerase and their interaction can correctly predict observed fluctuations in gene expression. Finally, we use mechanical coupling to propose a computational model of a toggle switch whereby interactions between two genes are mediated by DNA twisting as opposed to proteins.

*To whom correspondence should be addressed. Email: sahand_hormoz@hms.harvard.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

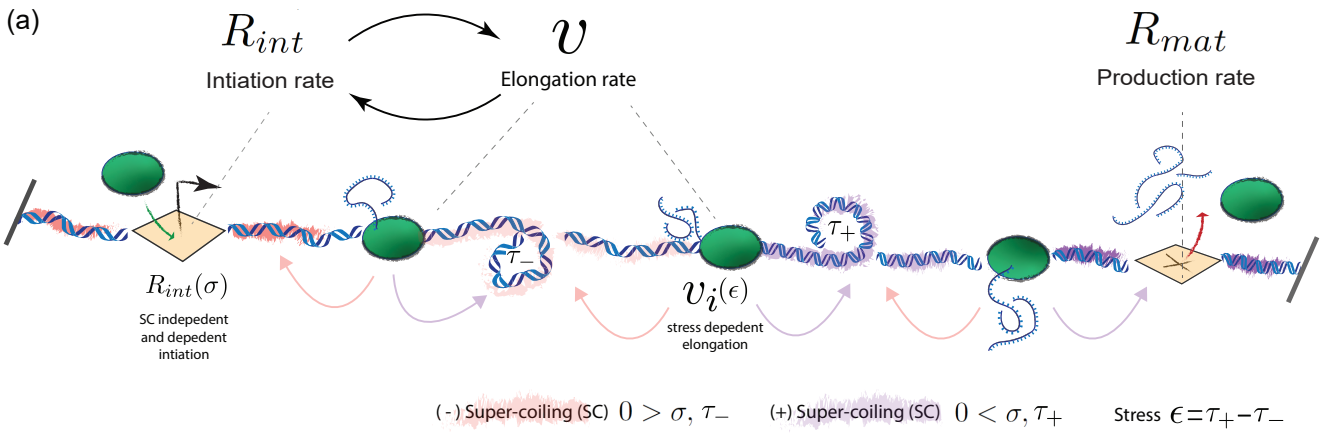

FIG. 1: Schematic of the model of the role of supercoiling during transcription. Gene expression output is determined by both polymerase recruitment (initiation) as well as polymerase transport along the gene (elongation). During elongation, polymerases produce supercoiling, the over and under twisting of DNA (shown in purple and red respectively), causing a corresponding change in the torque and in turn the torsional stress on the DNA. DNA supercoiling and torque can be transported between polymerases creating non-local interactions between polymerases that influence elongation and initiation. These interactions also impact both the average rate of mRNA production and its fluctuations.

Model description.

Two major factors determine gene expression output. The first is the rate at which polymerases are recruited at the transcription start site (TSS), referred to as the initiation rate. Second is the rate at which the recruited polymerases are transported from the start site to the termination site, referred to as the elongation rate. The rates of initiation and elongation together determine the rate of gene expression output.

Fluctuations in the output are not necessarily equal to fluctuations created during initiation because polymerase velocity and spacing can change during transport. To understand the fluctuations in the output and how it relates to the fluctuations in the input, we need to model the transport of polymerases and quantify its contribution to the fluctuations in the output.

Our model describes the position and velocity of the polymerases during transport. We assume that there are N polymerases between the transcription start site and the termination site. The ith polymerase has position x_i and velocity v_i. The density function of the polymerases $\rho = \sum_i \delta(s - x_i(t))$ encodes the position of all the polymerases along the gene, where position along the gene is parameterized by s (in bp). Similarly , the flux of polymerases along the gene is defined as:

$$j(s,t) = \sum_i v_i(t)\delta(s - x_i(t)) \tag{1}$$

Flux j at position s corresponds to the number of polymerases crossing s per unit time. The unit of flux is the same as that of the initiation and output rates. The flux captures the transport of polymerases along the gene and can be used to link initiation and elongation to output (Figure 1).

At steady state, the initiation rate is equal to output rate when averaged over a sufficiently long period of time. In addition, these rates should equal the average flux at any point along the gene when there is no depletion of polymerases along the gene. However, the fluctuations in these three rates are not necessarily equal. To understand how variations in the flux control output fluctuations, we need to incorporate the physical factors that modify polymerase velocity and the initiation rate.

DNA supercoiling (SC), the over and under-twisting of DNA [20], can control both the initiation rate and the velocity [4, 5] of polymerases. Importantly, as polymerase moves along the gene it can change the degree of supercoiling . This is because to transcribe, polymerase has to rotate to follow the helical grooves of DNA and/or DNA itself has to to twist as its pulled through a polymerase [20]. Polymerase rotation does not change the degree of supercoiling whereas DNA twisting does [13]. Therefore, we need to incorporate the feedback between polymerase velocity and DNA supercoiling in our model.

Following a physical construction of the twin-domain model of transcription [13], we define $\dot{\theta}_i$ to be the rate of rotation (angular velocity) of the ith polymerase. $\phi(s)$ is the degree of twisting of DNA at position s. We define $\phi(s = 0) = 0$. Even with zero supercoiling, as s increases ϕ also increases because of the natural helical form of DNA at a rate of $\omega_o = 1.67 rad/bp$ (for relaxed DNA). $\dot{\phi}(s)$ is the rate of twisting per unit time at position s. $\partial_s\phi$ evaluated position s_0 captures how the degree of twist changes when moving along the gene from position s_0 to a point infinitesimally away $s_0 + ds$. $\partial_s\phi$ is referred to as the local twist density.

Thus, $v\partial_s\phi$ determines the rate at which a polymerase

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

encounters twist when moving at velocity v along the gene. To follow the grooves of DNA, a combination of polymerase rotation or DNA twisting (in the opposite direction) must occur [13]. Therefore, for the ith polymerase, the equation

$$v_i \partial_s \phi(s,t) = \dot{\theta}_i(t) - \dot{\phi}(s,t). \tag{2}$$

encodes the local interplay between DNA twist density encountered during elongation, polymerase rotation and further DNA twisting.

The supercoiling density σ is defined as the change along the gene of the over or under twisting of DNA past its natural twist density ($\omega_0 = 1.67(rad/bp)$) which can be expressed as

$$\sigma(s,t) = \omega_o^{-1} \partial_s \phi(s,t) - 1 \tag{3}$$

We can write equation 2 using the supercoiling density σ instead of the local twist density $\partial_s \phi$.

$$v_i(t)\,(1+\sigma(t))\,\omega_o = \dot{\theta}_i(t) - \dot{\phi}(s,t) \tag{4}$$

We can use mechanics to relate the rate of rotation of polymerase $\dot{\theta}_i$ and the rate of twisting of DNA $\dot{\phi}(s)$ to the torque applied to the polymerase or DNA at position s. At each position s there exists some amount of torque $\tau(s)$ in response to the amount of twisting $\phi(s)$ at that position. The torsional stress $\epsilon_i(t) = \partial_s \tau(s = x_i, t)$ (also called the local torque per unit length) exerts a damping force on the polymerase that rotates it. Intuitively torsional stress can be thought of as the difference of the torque in front and behind the polymerase. Specifically, the rate of rotation of ith polymerase is given by,

$$(\gamma + \eta x_i^{\alpha}(t))\,\dot{\theta}_i(t) = \epsilon_i(t), \tag{5}$$

where the drag coefficient has a constant term γ and a term that increases with the distance x_i of the polymerase from the transcription start site. The drag coefficient increases with the distance of the polymerase from the start site because the nascent RNA attached to the polymerase increases in length as the polymerase moves along the gene. α is a phenomenological parameter that captures how the drag coefficient changes with the distance from the start site [21].

Similarly, the rate of rotation of DNA at position s away from transcription is related to the torsional stress experienced by the DNA at that position.

$$\zeta \dot{\phi}(s,t) = \epsilon(s,t) \tag{6}$$

where ζ is the constant drag coefficient of DNA and $\epsilon(s,t) = \partial_s \tau(s,t)$.

By substituting equation 4 into equation 5 to eliminate $\dot{\theta}$ we obtain an equation for the rate of rotation of DNA at the position of the ith polymerase

$$\dot{\phi}_i(t) = -\omega_o\,(1+\sigma)\,v_i(t) + \epsilon_i(t)/\left(\gamma + \eta x_i(t)^{\alpha}\right) \tag{7}$$

This equation directly relates dynamics of DNA twisting to the rate of polymerase elongation v_i and positioning x_i.

To write equation 7 for any position s on the gene, we combine the discrete drag coefficients of individual polymerases and the continuous drag coefficient of DNA and define,

$$D(s,t) = \frac{1}{\zeta}\left(1 - \sum_i \delta(s - x_i(t))\right) + \sum_i \frac{\delta(s - x_i(t))}{\gamma + \eta x_i^{\alpha}(t)}. \tag{8}$$

Using the above form of the the drag coefficient and our previous definition of flux j (Eq.1), we can write down an equation for the dynamics of DNA twisting that combines the contribution of polymerases with that of DNA itself

$$\dot{\phi}(s,t) = -\omega_o\,(1+\sigma(s,t))\,j + D(s,t)\epsilon(s) \tag{9}$$

By applying equation 3 we arrive at a supercoiling density transport equation for transcription

$$\partial_t \sigma(s,t) = \partial_s\left(-(1+\sigma(s,t))\,j + D(s,t)\epsilon(s,t)\right) \tag{10}$$

This result stands in contrast to previous models [15, 22] which assume a constant diffusion coefficient and neglect the fact that polymerases serve as both a source of supercoiling as well as barriers to its free diffusion.

We can write equation 10 explicitly in terms of σ (and thus stress) by specifying the local torque as

$$\tau(s,t) = C\sigma(s,t) \tag{11}$$

where C is the twist modulus of DNA ($75nm \cdot k_b t$ [23]). Here we do not account for DNA bending or writhing in our simple model though their analytical inclusion is straightforward [24–27] . Their inclusion can lead to DNA buckling resulting in altered twist transport and torsional responses. To account for buckling in our simulations, we will utilize a phenomenological relationship between torque and supercoiling (see S.I.).

The flux term in the above equation and the diffusion constant depend on the position and the velocity of the polymerases. To close the set of equations in our description, we need to relate the position of the polymerases to their velocity and relate their velocity to the supercoiling density σ. The position of the polymerase ith change in time as $\dot{x}_i = v_i$.

We assume a phenomenological relationship between the supercoiling density and the velocity of polymerase i using the following functional form

$$v_i(t) = \frac{v_{max}}{1 + e^{\kappa(\epsilon_i(t) - \epsilon_c)}};\ \epsilon_i = \partial_s \tau(x_i) \tag{12}$$

where ϵ_i is the torsional stress experienced by the ith polymerase as previously defined. ϵ_c the torsional stress cut-off (around $0.2N$ for E.Coli [19] assuming that the length scale of the polymerase is 0.6 angstroms [28]).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

v_{max} is the maximum velocity that polymerases can travel when $\epsilon_i << \epsilon_c$. Conversely, if $\epsilon_i >> \epsilon_c$ then the polymerase stalls. This phenomenological form is motivated by the empirical observations made in [19].

Taken together, equations eqs. (10) to (12) provide a closed form description of the generation and transport of supercoiling density along the gene due transcription as well as the position and velocity of each polymerase. To solve these equations, we need to also specify the initial conditions and the boundary conditions. $s = 0$ and $s = L$ are the positions of the two boundaries of the system. The boundaries of the gene itself (marked by the position of the transcription start site and termination site) are contained within the boundaries of the system. For an open system (DNA that is free to rotate at the boundaries) we have $\sigma(s = 0) = \sigma(s = L) = 0$. Similarly, for DNA that forms a closed loop $\sigma(0) = \sigma(L)$, $\partial_s\sigma(0) = \partial_s\sigma(L)$ so that the supercoiling density and torques match at the beginning and end of the system but are not necessarily zero. In our simulations, we use fixed boundary conditions where the gradient of supercoiling density at the two boundaries is zero $\partial_s\sigma(s = 0) = \partial_s\sigma(s = L) = 0$ at all times but the supercoiling densities are not $\sigma(0) \neq \sigma(L)$.

The above equations with the boundary conditions specified can be simulated to study the generation and transport of supercoiling density along the gene. In our simulations, we assume the limit of fast diffusion of supercoiling density between the polymerases. This limit corresponds to $\zeta \to 0$ in Equation 8. With this assumption, the supercoiling density is constant along DNA regions between polymerases allowing for us to directly use equations 7 to keep track of supercoiling in alignment with previous efforts [14].

Finally, we incorporate the role of topoisomerases into our simple model. Topoisomerases play an important role in regulating transcription in both bacteria and eukaryotes [29]. General classes of topoisomerases are formed by their ability to relieve either positive or negative supercoiling density as well as their mechanisms of actions which use either single or double strand breaks to modify the supercoiling density [30].

If no mechanism to relieve supercoiling is included (such as the one provided by topoisomerase) large amounts of supercoiling density accumulate and stalling occurs for reasonable choices of parameters as observed in experiments [31]. To incorporate topoisomerase action in our model, at random time-points, we scale the supercoiling at every point along the gene by the same constant factor by re-assigning the twist at each polymerase $\phi_i \to a\phi_i$ ($a = 0.1$ in our simulations. The value of this constant is not important because the rate of topoisomerase action is the free parameter in our simulations).

The rate at which topoisomerase acts is set by the difference of the supercoiling density at the the two boundaries:

$$\Omega(t) = \sigma(L, t) - \sigma(0, t) \tag{13}$$

$\Omega(t)$ roughly corresponds to the accumulation of supercoiling density along the gene over time. In our simulations, we use the following form for the rate of topoisomerase action as a function of $\Omega(t)$:

$$R_{topo}(t) = \lambda \frac{1}{1 + \Omega(t)}. \tag{14}$$

As described in more detail in the S.I., we tried multiple phenomenological forms relating $\Omega(t)$ to the rate of topoisomerase action: a constant rate independent of $\Omega(t)$; a rate proportional to $\Omega(t)$; and a rate that approaches zero with increasing $\Omega(t)$. The rate that approaches zero with increasing $\Omega(t)$ (corresponding to a constant rate of removal of supercoiling density) was most consistent with empirical observations [32] and was the only tested form able to produce results consistent with empirical observations.

Simulations

We simulated the model described above to understand how initiation, transport, and supercoiling work together to determine gene expression output (the simulations are described in detail in supplemental section A). To do this, we simulated a single isolated gene of length 1000 bp contained within a total stretch of DNA of length $L = 3000bp$. The start site of the gene is located at $s = 1000bp$. At $s = 0$ and $s = L$ DNA will be prevented from freely rotating causing supercoiling density to build up at the boundaries corresponding to the fixed boundary condition.

We begin with a constant initiation rate that does not depend on the supercoiling density at the transcription start site. Therefore, the loading process of polymerases at the transcription start site ($s = 1000bp$) is a Poisson process with rate R_{int}. We will later consider the initiation rate that is a function of the supercoiling density at the transcription start site.

As the polymerases move along the gene, the supercoiling density changes according to equation 7 which governs the local twist change at each transcription site. We set the drag coefficient $\gamma = 10^{-1}[pNs]$, and the phenomenological parameters $\eta = 10^{-4}[pNs/bp^2]$ and $\alpha = 2$. While there is little empirical data to determine the precise values of these parameters, biophysical considerations (as well as the observation that short genes do not induce supercoiling while long ones do [33]) imply a drag greater than $1pNnm$ for a nascent transcript of length $1kbp$ or greater rotating at $10rad/s$. We will regardless show that our results are robust to changes in the values of these parameters (see Discussion and S.I.). We assume

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

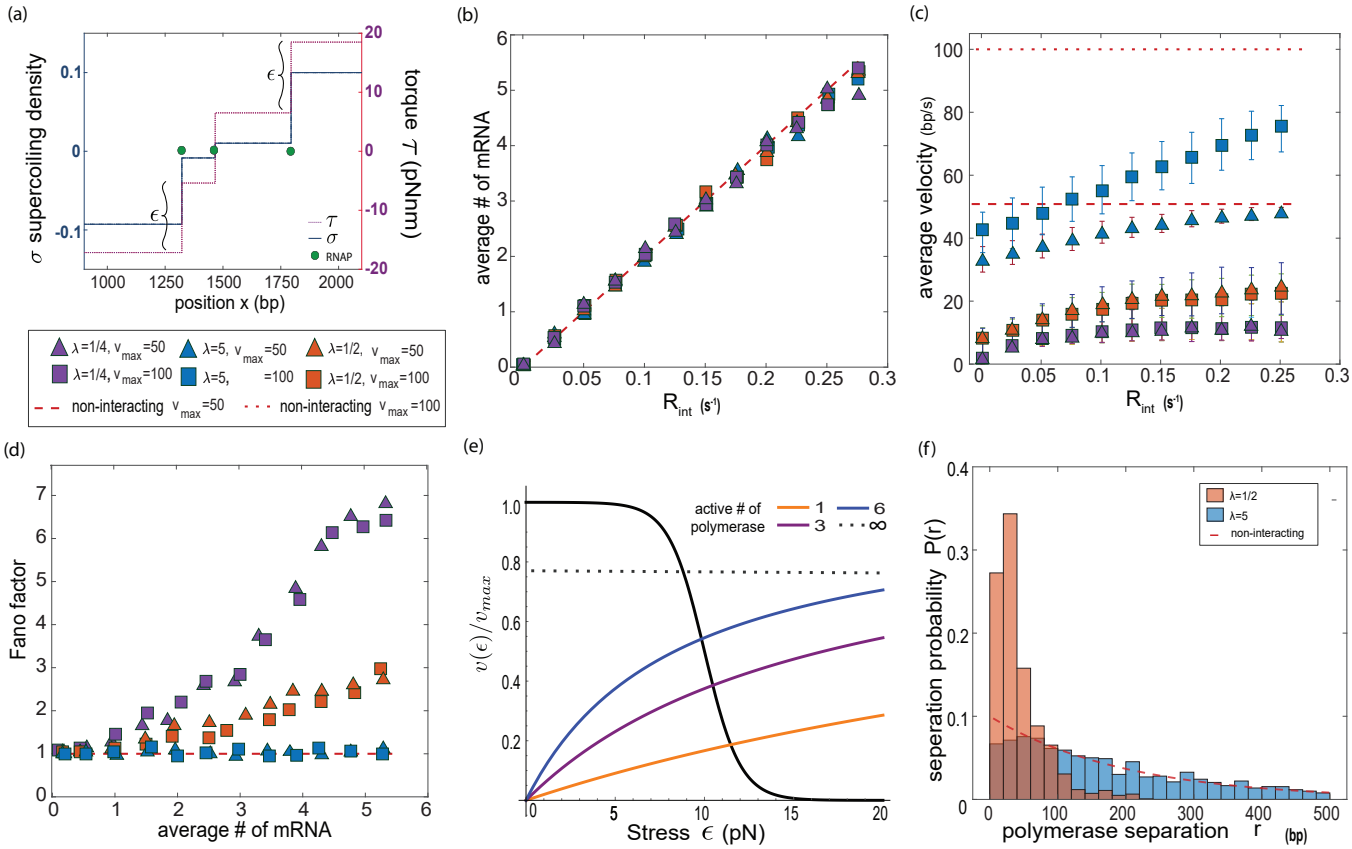

FIG. 2: DNA supercoiling mediated interactions between polymerases do not alter the average mRNA production rate but change its fluctuations by altering the separation distance between polymerases and their elongation rates.(a)Transcribing polymerases generate supercoiling density σ and accompanying torque during elongation. (b)The average number of mRNAs produced is insensitive to elongation kinetics of the polymerases and is solely determined by the initiation rate. Changing the elongation kinetics by altering the maximum polymerase elongation rate v_{max} or the rate of topoisomerase action λ does not change the production rate.(c)Average elongation rates depend on the initiation rates and therefore the average number of polymerases present on the gene, demonstrating a cooperative interaction between the polymerases. The elongation rate plateaus to a value that is predominantly set by the rate of topoisomerase action. (d) An approximately linear relationship between the mean mRNA production rate and its fluctuations (Fano factor) emerges with the slope determined by the rate of topoisomerase action. High rate of topoisomerase action results in fluctuations that resemble the Poisson statistics of non-interacting polymerases (red dashed line). (e) Analytical expressions for the elongation rate and the rate of topoisomerase action as a function of the stress. The rate at which supercoiling is added is proportional to the elongation rate. The steady state value of stress and in turn the elongation rate is set when the rate of addition of supercoiling equals the rate of its removal by topoisomerase 16 (f) Supercoiling mediated interactions change inter-polymerase separation distances. The separation distance between the polymerase nearest the TSS and its closest neighbor, r, decreases with increasing accumulation rate of supercoiling density. The altered separation distances results in higher fluctuations in gene expression. The distribution of separation distances deviates from that of non-interacting polymerases that follow Poisson statistics (exponentially distributed separation distances, shown by red line). Simulation details described in Supplemental Section A.

that the torque is related to the supercoiling density using the functional form shown in S.I. section C. This choice is motivated by empirical observations [23] and only contains one free parameter which sets the torque at which DNA buckles. We also ensured that our results are robust to the choice of this parameter (see figure S1).

When a polymerase reaches the transcription termination site, a mature mRNA is produced that then is removed at a constant rate $\mu = 10^{-2}s^{-1}$. The simulations were started with no polymerases on the gene and ran for a sufficiently long period of time to reach steady state when the number of RNA polymerases stabilized. We observed that the simulations reached steady state typically after 10 minutes out of a total of 1 hour of simulation time (time-scale set by μ) .

Figure 2a shows a snapshot of the simulation with 3 polymerases moving along the gene. For each simulation run we computed the number of mRNA molecules averaged across multiple simulation samples after each simulation reached steady state. The average number of mRNA is plotted as a function of the initiation R_{int} for different values of parameters λ and v_{max} (Figure 2b).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As expected the number of mRNA molecules is proportional to R_{int} but does not depend on the values of λ and v_{max}. This is because at steady state the rate at which polymerases are loaded must equal to the rate of production of mRNA.

The average velocity at which the polymerases move along the gene (shown in Figure 2c) also depends on R_{int} but saturates to a value that does not necessarily correspond to v_{max}, especially when the rate of topoisomerase action λ is low or the ends of DNA are free in accordance with experimental observation [4]. Importantly, this behavior recapitulates two additional empirical observations. One is that the polymerase velocity changes as the rate of topoisomerase action changes [4, 16]. Second is that the velocity of polymerases can be smaller than the bare velocity ($v_{max}/2$) defined as the velocity of a single polymerase moving along a linear piece of DNA with open boundaries where there is no accumulation of supercoiling density [4, 16].

To gain an intuition for this behavior, consider the simplest case where all the polymerase velocities are equal to v. In addition, we assume that there is sufficiently large drag ($\gamma + \eta x^{\alpha} >> 1$) on each polymerase so that we can ignore polymerase rotation ($\dot{\theta} = 0$). In this case the supercoiling density generated by each polymerase is cancelled by the supercoiling density generated by the neighboring polymerases except for at the boundaries where there is no cancellation. Then the rate of supercoiling density generation is $-\omega_o v$ at the boundary closest to the start site and $\omega_o v$ at the other boundary (Equation 10). Therefore, we can write down an equation for the rate of change Ω (eq.13) as

$$\dot{\Omega} = 2\omega_o v - \lambda \frac{\Omega}{1+\Omega}, \tag{15}$$

where the second term on the right hand site is the rate of removal of supercoiling density by topoisomerase action. At steady state, $\dot{\Omega} = 0$.

If we make the additional simplifying assumption that the change in supercoiling density across each polymerase $\Omega = N\epsilon$ is proportional to the torsional stress on that polymerase ϵ, Equation 15 can be written as a function of N and ϵ as

$$Cv(\epsilon) - \lambda \frac{N\epsilon}{1+N\epsilon} = 0 \tag{16}$$

Above equation sets the value of torsional stress ϵ for each polymerase and in turn their velocity. Figure 2d shows the contribution of each term in Equation 16 as a function of ϵ.

Increasing the polymerase loading rate R_{int} increases N. However, the rate of removal of supercoiling density due to topoisomerase actions saturates with increasing N (Figure 2e). The value at which this saturation occurs sets the value of torsional stress on each polymerase and thereby their average velocity. The dashed curve shows the smallest possible ϵ that satisfies equation 16 when $N \to \infty$ and in turn sets the values of elongation rate for large R_{int} in Figure 2c. Because of the saturation of the rate of topoisomerase action, the average polymerase velocity can be smaller than the maximum allowed velocity v_{max}. Other choices for the functional form relating the rate of topoisomerase action to $\Omega(t)$ such as $R_{topo} \sim Cnst.$ or $R_{topo} \sim \Omega(t)$ will not result in a velocity that saturates to a value other than v_{max} for large N.

Next, we computed the fluctuations in the number of mRNA molecules. We computed the Fano factor of the number of mRNA molecules (variance divided by mean) over the entire duration of the simulation after steady-state was reached. Surprisingly, the Fano factor deviates from simple Poisson statistics at sufficiently high initiation rate and low rate of topoisomerase action (Figure 2e). This is because as polymerases move along the gene, they interact with each other and change their separation distances (referred to as clustering) from the initial exponentially-distributed separation distances set by the Poisson loading process (Figure 2f). With interactions, the distribution of separation distances peaks at a non-zero value and has a narrower range as evident Figure 2f. This effect is larger for lower rate of topoisomerase action because the interactions are mediated by the accumulation of supercoiling density. Clustering of polymerases due to interactions is a plausible explanation for the universally observed 'bursting' dynamics of gene expression and the relationship between the average number of transcript observed in individual cells and the fluctuations of the number of transcripts across cells in a population [34].

To gain an intuition for how clustering occurs, consider the dynamics of two neighboring polymerases. The relative distance between the two polymerases changes as

$$\dot{r} = v_1(\epsilon_1) - v_2(\epsilon_2) \tag{17}$$

where the velocities of each polymerase is determined by equation 12. The torsional stress experienced by the first and second polymerase is a function of supercoiling density in front of, in between, and behind the two polymerases $\sigma_F, \sigma_M, \sigma_B$ respectively. In particular, $\epsilon_1 \propto \sigma_F - \sigma_M$ and $\epsilon_2 \propto \sigma_M - {}_B$. Because supercoiling density always accumulates in the front and all polymerases move in the same direction, $\sigma_F > \sigma_M > \sigma_B$. Importantly, the supercoiling density between the two polymerases is inversely proportional to their separation distance, $\sigma_m \propto 1/r$. If the separation distance between two polymerases is large then σ_m is small which in turn generally implies $\epsilon_1 > \epsilon_2$ and $v_1 < v_2$. Therefore, in the case of large separation, $\dot{r} < 0$ and the separation distance between the two polymerases shrinks. Conversely, if the separation distance between the two polymerases is small then σ_m is large which in turn generally implies $\epsilon_1 < \epsilon_2$ and $v_1 > v_2$ and an increasing separation distance

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\dot{r} > 0$. Taken together, as polymerases move along gene, because of the interactions, they converge to a preferred separation distance (Figure 2f) and then move with a constant velocity (Figure 3b).

Surprisingly in our simulations, we observed that a minimum number of three interacting polymerases moves a larger distance before stalling compared with one or two polymerases, as shown in Figure 3b. To gain an intuition for why a minimum of three interacting polymerases are required for sustained motion, consider how stress accumulates as one polymerase moves along the gene. We can write an equation for the rate of change of the velocity of the polymerase $\dot{v}$ in terms of the dynamics of the local stress by applying a simple chain rule to the stress dependent velocity (eq.12)

$$\dot{v}_i(\epsilon_i) = \frac{d}{d\epsilon_i} v_i(\epsilon_i)\dot{\epsilon} \tag{18}$$

An isolated polymerase moves at velocity v_{max} initially. As it moves along the gene, the torsional stress across the polymerase accumulates at the rate that is proportional to its velocity, $\dot{\epsilon} \propto v$. Substituting this into Eq. 18, implies $\dot{v} \propto v$. Therefore, the velocity of the polymerase decays exponentially to zero. When topoisomerase relieves the stress, the polymerase can start to move again . When two polymerases move along the gene, their velocities converge to the same value as described above. As with the case of an isolated polymerase, the two polymerases also accumulate torsional stress at a rate proportional to their velocity. This occurs even though supercoiling does not accumulate in the region between the two polymerases because the negative supercoiling density generated by the leading polymerase cancels the positive supercoiling density generated by the trailing polymerase. However, supercoiling density does accumulate outside the two polymerases because there is no cancellation. Therefore, the velocities of the two polymerases also decays exponentially as is the case with a single polymerase (shown in example traces in Figure 3b).

This picture changes qualitatively with the addition of a third polymerase. The velocities of the three polymerases also converge to a constant value as they move along the gene. However, the supercoiling density generated by the middle polymerases is exactly cancelled by the two polymerases on either side. Therefore, the middle polymerase has no torsional stress accumulation. This puts the middle polymerase in a privileged position. As the outer two polymerases slow down from the accumulation of stress, the middle polymerase continues to move. When this happens, the middle polymerase accumulates torsional stress because there is no exact cancellation of supercoiling density from the other two polymerases. Importantly, the accumulation of supercoiling density in the middle relieves torsional stress on the outer polymerases, resulting in their collective motion.

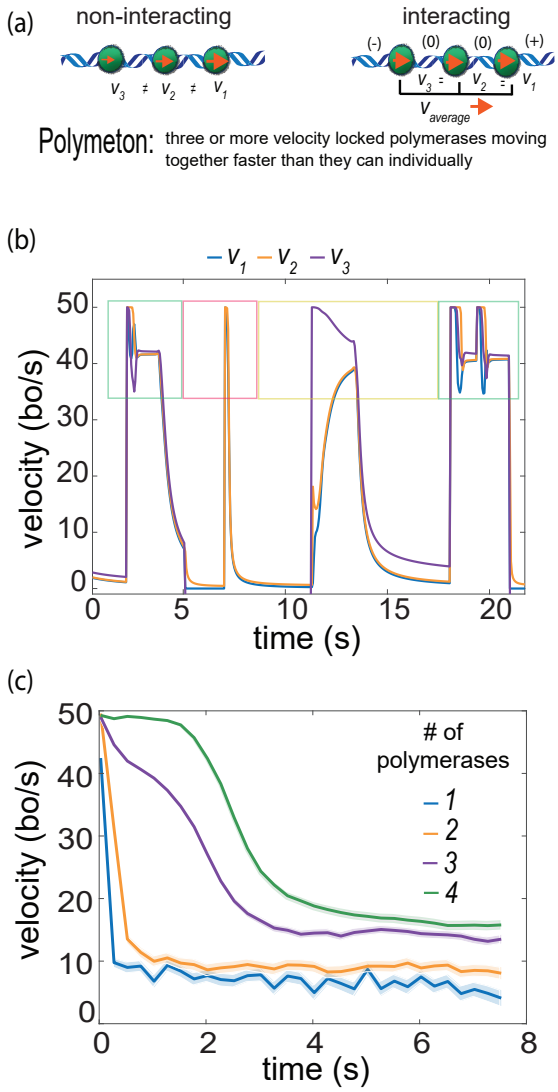

FIG. 3: Supercoiling mediated interactions lead to cooperative elongation rates through the creation of Polymetons. (a) Supercoiling density differences across different polymerases lead to velocity differences that in turn change the polymerase separation distances. Polymetons are velocity-locked groups of three or more polymerases where supercoiling density does not accumulate for the middle polymerase. (b)Representative velocity trajectories that show cycles of elongation and stalling due to supercoiling density accumulation and release via topoisomerase action. Three polymerases moving as a polymeton are initially present on the gene labeled with numbers that increase from the termination site to the start site (green square).At approximately t=5s, the leading polymerase reaches the termination site and is removed leaving behind two polymerases that rapidly stall (orange square). Initiation of a new polymerase at the start site increases the elongation rate of the two stalled polymerases mediated through supercoiling induced interactions (yellow square). Finally, the three polymerases again form a polymeton and elongate with approximately similar velocities (green square on the right). (c) The average elongation rate as a function time following topoisomerase action binned by the number polymerases present on the gene. There is a clear jump when the number of polymerases present changes from two to three showing the formation of polymetons.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

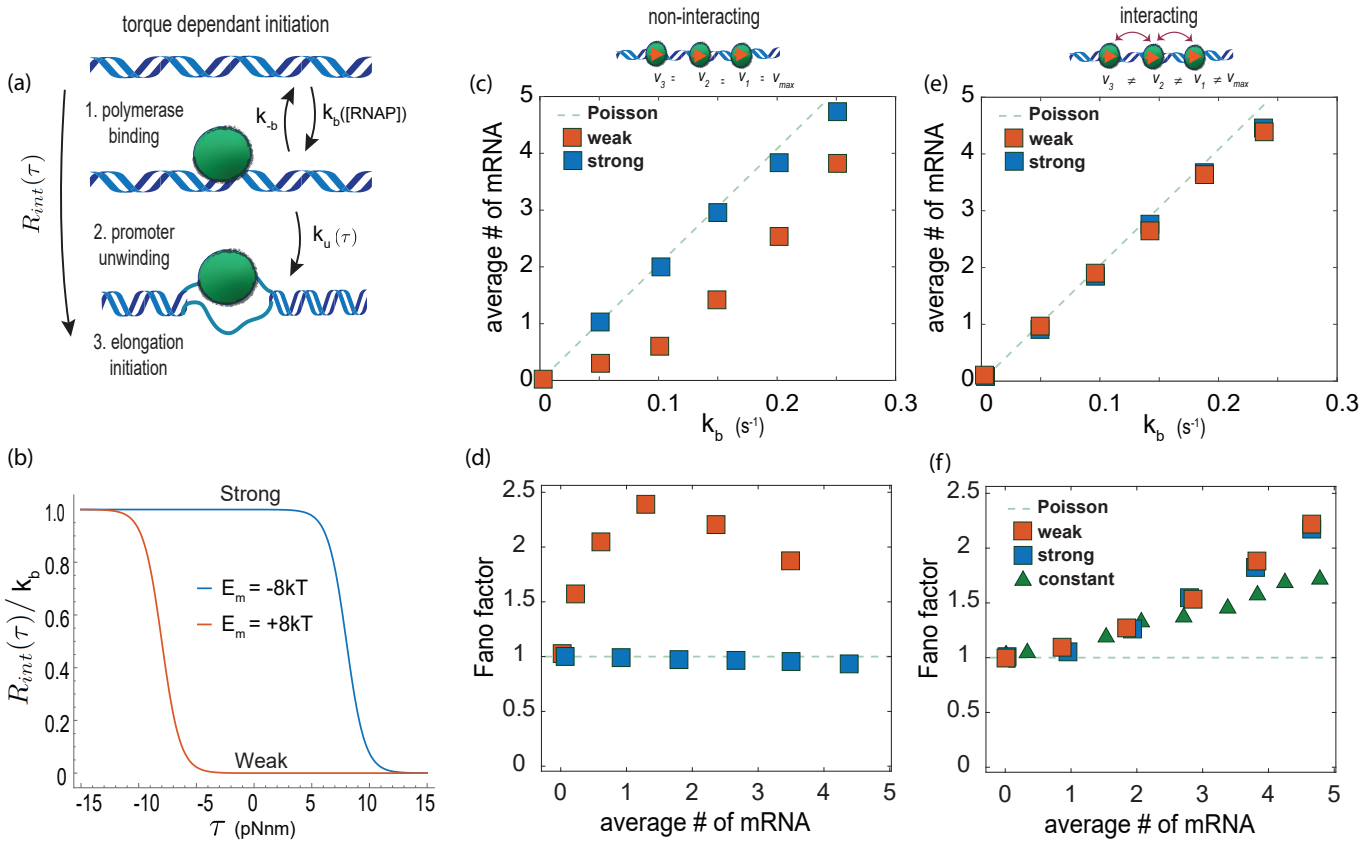

FIG. 4: Torque-dependent initiation alongside super-coiling mediated interactions recapitulates the expected mean mRNA production rate and its fluctuations. (a) Torque dependent initiation is a two step process: a reversible polymerase binding step that does not depend on torque but on the free RNA polymerase concentration, followed by an irreversible promoter unwinding step which depends on the torque at the promoter site. (b) A weak promoter ($E_b > 0$) is more sensitive to the torque at the promoter site than a strong promoter ($E_b < 0$) requiring negative torque to initiate transcription. (c-d) Mean expression and expression fluctuations (Fano factor) for torque dependent initiation with increasing polymerase binding rate k_b both in the absence of supercoiling mediated interactions between polymerases. The weak promoter exhibits a production rate that scales non-linearly with k_b (c) and non-monotonic fluctuations in expression (d). (e-f) Same as in (c-d) but with the addition of super-coiling mediated interactions between the polymerases. Both strong and weak promoters exhibit linear production rates in k_b (e) with fluctuations that increase monotonically with the average production rate (e). The green triangles in (f) show the fluctuations when interactions are present but the initiation rate is not dependent on torque.

Taken together, three polymerases can sustain their motion for significantly longer periods of time before stalling (as shown in the traces in Figure 3b).

In summary, our results show that there is an emergent phenomena where three or more polymerases can sustain their collective motions for a longer period of time than one or two polymerases. We will refer to three or more interacting polymerases undergoing sustained motion as polymetons. Polymetons emerge in our system following a topoisomerase action when three or more simultaneously elongating polymerases move at near constant velocity (fig.3b). While the precise form of this behavior depends on the details of the model, the qualitative behavior that middle polymerases occupy a privileged position and can assist in sustaining the motion of the group should not depend on the precise form of the model.

Torque dependant initiation

Next, we incorporate torque dependent initiation into the model. Up to this point, polymerase initiation in the model occurs stochastically as a Poisson process with a constant rate R_{int}. However, there is experimental evidence that the initiation rate should depend on the supercoiling density at the promoter site. First, there are in vivo measurements showing that increasing the level of negative supercoiling at a promoter by inducing the production of neighboring genes can alter its output [5, 35]. Second, in vitro single molecule experiments have directly measured promoter unwinding kinetics as a function of the torque in DNA [28] demonstrating that initiation is sensitive to the torque at the promoter.

Polymerase initiation is basically an ordered process

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of polymerase binding to DNA at the promoter site, unwinding of the DNA, and polymerase leaving the promoter (referred to as clearance). Polymerase-DNA binding follows standard chemical kinetics whereby the binding rate increases proportionally with polymerase concentration and promoter affinity, K_o, which captures the intrinsic affinity of polymerase for the promoter, $k_b = [RNAP]K_o$. The unbinding rate of polymerase from the promoter, k_{-b}, is not sensitive to polymerase concentration [36]. Precise kinetic observations [28, 36] have measured these rates for specific promoters and could thus be used as inputs into a model of initiation. In our model, we explicitly vary k_b to capture the behavior of genes with different promoter affinities and polymerase concentrations. Inducing or repressing correspond to varying k_b.

Following promoter binding, the polymerase locally unwinds the DNA at the promoter site. This step has a strong dependence on the torsional state of DNA and can become the rate limiting step in initiation. Precise characterization of promoter unwinding and clearance by polymerases for varying levels of DNA supercoiling have been made [28]. The rate of unwinding depends on the level of torque at the promoter site: the rate of unwinding decreases as torque is increased following a simple Arrhenius law form as measured by [28]. Consequentially, we model the unwinding rate k_u as

$$k_u = e^{-(\tau_{prom}(t)\delta + E_m)} \tag{19}$$

δ captures the dependence of the unwinding rate on the torque τ. E_m captures the strength of the promoter and can vary from one promoter to another. E_m sets the value at which torque unwinds specific promoters and allows transcription. Some promoters require negative torque to unwind and allow transcription (referred to as weak promoters) whereas strong (consensus) promoters can unwind for positive values of torque [28] (rrnBP1 and lacCONS promoters respectively).

Importantly, the torque at the promoter site $\tau_{prom}(t)$ is given by the supercoiling density which is in turn set by the polymerases as they move along the gene and interact (Eq. 10). Promoter clearance occurs rapidly following the promoter unwinding [36] and thus is ignored here. Collectively, our model of initiation is composed of a two step process of reversible polymerase-DNA binding followed by irreversible promoter unwinding which results in initiation (Figure 4a). Here we utilize kinetic unwinding data of the Lac promoter [28] to model a strong promoter requiring no free parameters (see figure 4b). The inferred values are $E_m = -8$ and $\delta = 2$ (values for $k_{\pm b}$ are given in [28]). This allows us to make a direct comparison between torsion dependent initiation and gene expression for the same promoter [10].

The torque dependent initiation rate constructed above can be incorporated into our existing framework. We first examine the role of torque dependent initiation on gene output without polymerase interaction. Figure 4c and 4d show that a strong promoter is not affected by the changes in torque at the promoter site generated by the elongating polymerases even if they are not interacting. In this case, the gene mRNA output increases with the binding rate k_b but exhibits Poisson fluctuations (Fano Factor= 1). A weak promoter displays a non-linear relationship between the average mRNA output and the binding rate k_b. However, the fluctuations in the output displays a non-monotonic dependence on the mean expression in disagreement with general experimental observations [9] as well as specific observations for the rrnBP1 and lac promoters [10].

Importantly, when the polymerases are allowed to interact as they move along the gene, the dependence of mRNA output and its fluctuation changes. Both weak and strong promoters show a linear dependence of average mRNA output and binding rate k_b (Figure 4e). This output matches what would be expected if polymerase initiation did not depend on the torque but followed a simple Poisson process. The change in the behavior of output is because the torque at the promoter site changes when polymerases interact with each other. In addition, interacting polymerases also change how the mRNA output fluctuates (Figure 4f). Both weak and strong promoters now show non-Poisson fluctuations that monotonically increase with the average output, again consistent with empirical observations [10].

Taken together, these results indicate that torque dependent initiation alone is insufficient to explain the expected bursting behavior of genes. However, incorporating polymerase (velocity) interactions recovers the expected bursting behavior in genes pointing to fluctuations incurred during elongation as the overriding source of bursting in gene expression.

Additionally the insensitively of strong promoters to negative torque (supercoiling) [28] calls into question the widespread use of initiation as the source of transcriptional bursting used in models [5, 15, 22].

Finally, the inter-play between torque dependent initiation and elongation could be exploited to engineer novel gene regulatory mechanisms or identify existing ones. For instance imagine two identical genes convergently oriented towards one another both with weak promoters (Figure 5A). The gene that spontaneously initiates transcription first generates positive supercoiling density at the promoter site of the other gene because of the geometry of their arrangement. This supercoiling density in turn changes the initiation rate of the other gene. Therefore, the two genes repress each other resulting in DNA supercoiling-mediated toggle switch.

To demonstrate this effect, we constructed a system of two convergently oriented, identical genes with torque dependent initiation and weak promoters. This system demonstrates bi-stable expression where when one gene is on, the other is off (Figure 5b-c). Importantly, the bistability of this system is not set by proteins and therefore

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

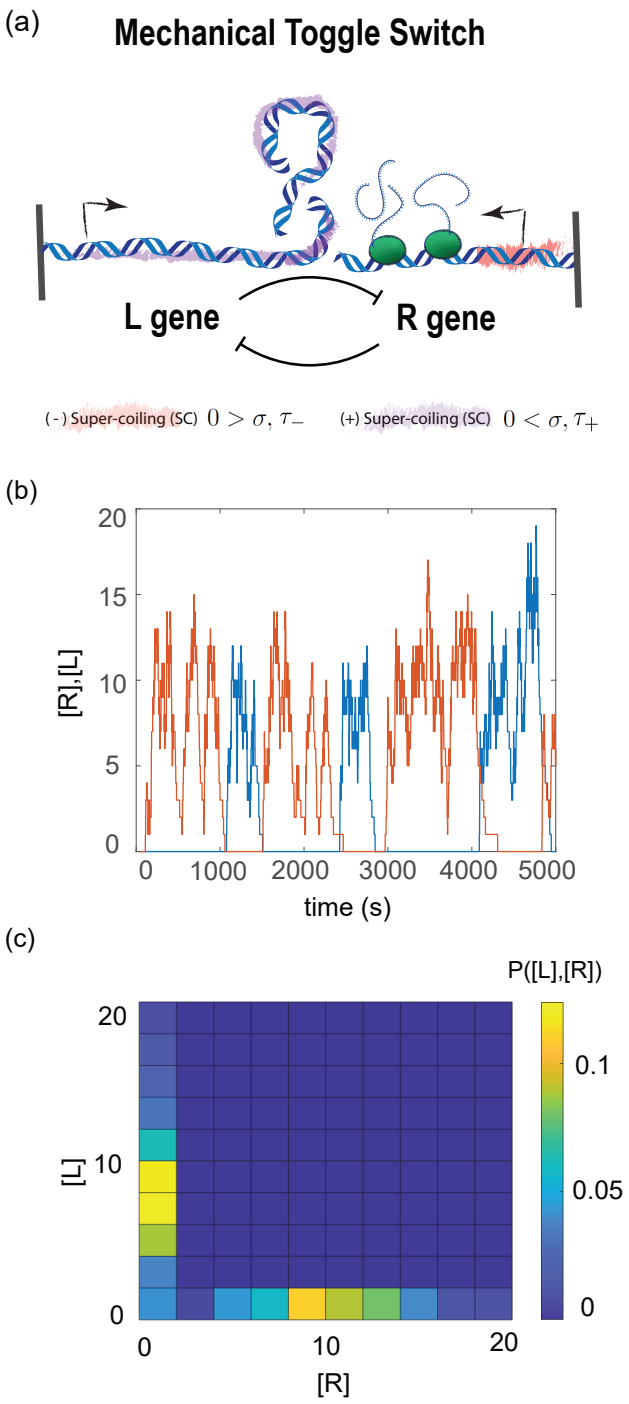

FIG. 5: (Mechanical toggle switch demonstrates that DNA supercoiling can be used to engineer interactions across genes without using proteins. a)Two neighboring genes with their promoters oriented towards each other can regulate each other's expression through supercoiling mediated initiation and elongation resulting in a mechanical toggle switch. (b) Simulated production levels of two identical genes with weak promoters with the orientation shown in (a) as a function of time. The two genes mutually repress one another leading to alternating periods of mRNA production by the [L] and [R] gene. (c) Alternating periods of mRNA production leads to a bi-modal distribution of mRNA expression levels.

the timescale of the oscillations are independent of protein life times. This simple system highlights the potential of novel forms of gene regulation mediated through supercoiling.

Discussion

In the framework introduced above, non-local interactions between polymerases emerge leading to cooperative elongation rates, clustering and bursting, consistent with observations of cooperative elongation kinetics [4] and bursting [9]. A collective phenomenon, referred to as Polymetons, emerges whereby groups of three or more polymerases can sustain their elongation due to the privileged status of the interior polymerase. In addition, we introduced a mechanical model of transcription initiation. The addition of torque-dependent initiation alone is insufficient to reproduce observed relationships between the mean levels and fluctuations in gene expression. We found that including both torque-dependent initiation and elongation is sufficient to match experimental observations. This result calls into question the validity of using models of torque-dependent initiation alone that do not also incorporate interactions to explain gene expression fluctuations. Finally, we constructed an example where mechanical epigenetics can be used to build novel synthetic regulatory circuits using two neighboring genes to form a mechanical toggle switch.

Our model relies on some simplifying assumptions. First, we assume that supercoiling density diffuses infinitely fast along the gene. This assumption was motivated because the mechanical state of DNA can relax much more quickly compared with the rate at which polymerase moves along DNA. Second, the conversion of DNA twist into writhe (bending) was not explicitly incorporated into our model. Instead, we used a phenomenological formulation of the relationship between supercoiling density and torque which implicitly includes DNA buckling [23]. Similarly, a simple phenomenological framework was used to model the drag on elongating polymerases [21]. Both of these phenomenological models are motivated by physical models of how polymers behave [21, 23]. Third, we constructed a simple model of topoisomerase action. Our model was motivated by simple kinetic observations of topoisomerase acting DNA plasmids [32]. While not derived from microscopic knowledge of topoisomerase action, the phenomenological form used in our model was constrained because alternative constructions resulted in qualitative behaviors that disagreed with experimental observations (see S.I. figure S2). Finally, the existence of cooperative elongation kinetics and polymetons was found to be robust under various phenomenological parameters used for DNA buckling, nascent mRNA drag and topoisomerase action (see S.I. figure S1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The twist transport equation developed here (eq.9) can be used in future work where DNA writhe is computationally or analytically incorporated allowing for important effects such as histone occupancy dynamics to be included. These effects are particularly important for understanding gene expression dynamics in eukaryotic cells. The current form of the model is most directly applicable to studying the effects of DNA supercoiling on gene expression in prokaryotes.

Further measurements will offer insights into the respective contributions of initiation and elongation to gene expression statistics. For instance, strategically placing single-molecule reporter systems such as MS2 and PP7 in a single gene, one near the promoter and the other near the termination site [17, 37] would allow for the de-convolution of flux due to initiation and its evolution during elongation.

During the development of this article two related works with some overlapping results appeared [38, 39]. The first article [38] used an existing model from [14] and the second article makes no explicit connection to DNA mechanics and is largely focused on explaining the recent experimental results indicating that increasing initiation rates can increase elongation rates [4]. In both articles the connection between elongation kinetics, clustering and the relationship between mean expression and expression output are not made. In fact in [39] it is stated that their model hinders the formation of clustered convoys of polymerases [18]. Importantly, both articles do not report the existence of polymetons and do not examine the effects of torque-dependent initiation on gene expression fluctuations.

In conclusion we have shown that DNA supercoiling serves as a powerful mediating force between polymerases by altering initiation and elongation kinetics. These effects are inescapable physical attributes of transcription and offer a widespread non-local feedback mechanism between polymerases. This framework challenges the traditional decoupling of transcriptions initiation and elongation and its implications for gene expression fluctuations. Understanding the role that DNA mechanics plays in gene expression can provide new insights into how genes are regulated through mechanical epigenetics.

The authors acknowledge funding from NIH NHLBI R01HL158269 grant.

[1] T. L. Lenstra, J. Rodriguez, H. Chen, and D. R. Larson, "Transcription dynamics in living cells," *Annual Review of Biophysics*, vol. 45, no. 1, pp. 25–47, 2016. PMID: 27145880.

[2] T. B. K. Le, M. V. Imakaev, L. A. Mirny, and M. T. Laub, "High-resolution mapping of the spatial organization of a bacterial chromosome," *Science*, vol. 342, no. 6159, pp. 731–734, 2013.

[3] G. K. Marinov, A. E. Trevino, T. Xiang, A. Kundaje, A. R. Grossman, and W. J. Greenleaf, "Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate breviolum minutum," *Nature Genetics*, vol. 53, pp. 613–617, Apr. 2021.

[4] S. Kim, B. Beltran, I. Irnov, and C. Jacobs-Wagner, "Long-distance cooperative and antagonistic RNA polymerase dynamics via DNA supercoiling," vol. 179, no. 1, pp. 106–119.e16.

[5] B. El Houdaigui, R. Forquet, T. Hindré, D. Schneider, W. Nasser, S. Reverchon, and S. Meyer, "Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling," vol. 47, no. 11, pp. 5648–5657. Publisher: Oxford Academic.

[6] C. P. Johnstone, N. B. Wang, S. A. Sevier, and K. E. Galloway, "Understanding and engineering chromatin as a dynamical system across length and timescales," *Cell Systems*, vol. 11, no. 5, pp. 424–448, 2020.

[7] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, "Real-time kinetics of gene activity in individual bacteria," *Cell*, vol. 123, no. 6, pp. 1025–1036, 2005.

[8] A. Raj and A. v. Oudenaarden, "Nature, nurture, or chance: Stochastic gene expression and its consequences," vol. 135, no. 2, pp. 216–226. Publisher: Elsevier.

[9] A. Sanchez and I. Golding, "Genetic determinants and cellular constraints in noisy gene expression," vol. 342, no. 6163, pp. 1188–1193.

[10] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, "General properties of transcriptional time series in escherichia coli," vol. 43, no. 6, pp. 554–560. Number: 6 Publisher: Nature Publishing Group.

[11] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, "Transcriptional burst frequency and burst size are equally modulated across the human genome," *Proceedings of the National Academy of Sciences*, vol. 109, no. 43, pp. 17454–17459, 2012.

[12] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, "Real-time observation of transcription initiation and elongation on an endogenous yeast gene," *Science*, vol. 332, no. 6028, pp. 475–478, 2011.

[13] S. A. Sevier, D. A. Kessler, and H. Levine, "Mechanical bounds to transcriptional noise," vol. 113, no. 49, pp. 13983–13988. Publisher: National Academy of Sciences Section: Physical Sciences.

[14] S. A. Sevier and H. Levine, "Properties of gene expression and chromatin structure with mechanically regulated elongation," vol. 46, no. 12, pp. 5924–5934.

[15] M. Ancona, A. Bentivoglio, C. A. Brackley, G. Gonnella, and D. Marenduzzo, "Transcriptional bursts in a nonequilibrium model for gene regulation by supercoiling," vol. 117, no. 2, pp. 369–376.

[16] S. Chong, C. Chen, H. Ge, and X. S. Xie, "Mechanism of transcriptional bursting in bacteria," vol. 158, no. 2, pp. 314–326. Publisher: Elsevier.

[17] K. Fujita, M. Iwaki, and T. Yanagida, "Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA," vol. 7, no. 1, p. 13788.

[18] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand, "A single-molecule view of transcription reveals convoys of RNA polymerases and

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

multi-scale bursting," vol. 7, no. 1, p. 12248. Number: 1 Publisher: Nature Publishing Group.

[19] J. Ma, L. Bai, and M. D. Wang, "Transcription under torsion," *Science*, vol. 340, no. 6140, pp. 1580–1583, 2013.

[20] L. F. Liu and J. C. Wang, "Supercoiling of the dna template during transcription," *Proceedings of the National Academy of Sciences*, vol. 84, no. 20, pp. 7024–7027, 1987.

[21] M. Laleman, M. Baiesi, B. P. Belotserkovskii, T. Sakaue, J.-C. Walter, and E. Carlon, "Torque-induced rotational dynamics in polymers: Torsional blobs and thinning," *Macromolecules*, vol. 49, no. 1, pp. 405–414, 2016.

[22] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo, "Stochastic model of supercoiling-dependent transcription," *Phys. Rev. Lett.*, vol. 117, p. 018101, Jun 2016.

[23] J. F. Marko, "Torque and dynamics of linking number relaxation in stretched supercoiled dna," *Phys. Rev. E*, vol. 76, p. 021926, Aug 2007.

[24] C. W. Wolgemuth, T. R. Powers, and R. E. Goldstein, "Twirling and whirling: Viscous dynamics of rotating elastic filaments," *Phys. Rev. Lett.*, vol. 84, pp. 1623–1626, Feb 2000.

[25] T. R. Powers, "Dynamics of filaments and membranes in a viscous fluid," *Rev. Mod. Phys.*, vol. 82, pp. 1607–1631, May 2010.

[26] T. R. Powers, "Dynamics of filaments and membranes in a viscous fluid," *Rev. Mod. Phys.*, vol. 82, pp. 1607–1631, May 2010.

[27] H. Wada and R. R. Netz, "Non-equilibrium hydrodynamics of a rotating filament," *Europhysics Letters (EPL)*, vol. 75, pp. 645–651, aug 2006.

[28] A. Revyakin, R. H. Ebright, and T. R. Strick, "Promoter unwinding and promoter clearance by rna polymerase: Detection by single-molecule dna nanomanipulation," *Proceedings of the National Academy of Sciences*, vol. 101, no. 14, pp. 4776–4780, 2004.

[29] J. C. Wang, "Dna topoisomerases," *Annual Review of Biochemistry*, vol. 65, no. 1, pp. 635–692, 1996. PMID: 8811192.

[30] N. G. Bush, K. Evans-Roberts, and A. Maxwell, "DNA topoisomerases," vol. 6, Oct. 2015.

[31] R. S. Joshi, B. Piña, and J. Roca, "Topoisomerase II is required for the production of long Pol II gene transcripts in yeast," *Nucleic Acids Research*, vol. 40, pp. 7907–7915, 06 2012.

[32] Y. Wang, S. Rakela, J. W. Chambers, Z.-C. Hua, M. T. Muller, J. L. Nitiss, Y.-C. Tse-Dinh, and F. Leng, "Kinetic study of DNA topoisomerases by supercoiling-dependent fluorescence quenching," vol. 4, no. 19, pp. 18413–18422. Publisher: American Chemical Society.

[33] T. B. Le and M. T. Laub, "Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries," *The EMBO Journal*, vol. 35, no. 14, pp. 1582–1595, 2016.

[34] A. Sanchez and I. Golding, "Genetic determinants and cellular constraints in noisy gene expression," *Science*, vol. 342, no. 6163, pp. 1188–1193, 2013.

[35] E. Yeung, A. J. Dy, K. B. Martin, A. H. Ng, D. Del Vecchio, J. L. Beck, J. J. Collins, and R. M. Murray, "Biophysical constraints arising from compositional context in synthetic gene networks," *Cell Systems*, vol. 5, no. 1, pp. 11–24.e12, 2017.

[36] L. J. Friedman and J. Gelles, "Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation," *Cell*, vol. 148, no. 4, pp. 679–689, 2012.

[37] J. Liu, D. Hansen, E. Eck, Y. J. Kim, M. Turner, S. Alamos, and H. G. Garcia, "Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between rna initiation, elongation, and cleavage," *PLOS Computational Biology*, vol. 17, pp. 1–26, 05 2021.

[38] S. Tripathi, S. Brahmachari, J. N. Onuchic, and H. Levine, "Dna supercoiling-mediated collective behavior of co-transcribing rna polymerases," *bioRxiv*, 2021.

[39] P. Chatterjee, N. Goldenfeld, and S. Kim, "Dna supercoiling drives a transition between collective modes of gene synthesis," *bioRxiv*, 2021.

[40] A. Crut, D. A. Koster, R. Seidel, C. H. Wiggins, and N. H. Dekker, "Fast dynamics of supercoiled dna revealed by single-molecule experiments," *Proceedings of the National Academy of Sciences*, vol. 104, no. 29, pp. 11957–11962, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPORTING INFORMATION APPENDIX (SI)

Simulation details and parameters

Simulations were conducted in MATLAB. Euler's method was used for the numerical integration of the twist angles given in equation 7 of the main text. Integration steps of size $\Delta T = 1/200(s)$ where used. The same method was applied for the integration of polymerase motion. See [14] for more details. Stochasticity for the initiation of polymerases, degradation of produced mRNA and topoisomerase action where modelled as Poisson processes. This was implemented in the simulation by allowing for each process to occur within each time-step with a probability specified by the rate for that process.

Model predictions are robust to the parameterization of drag

Simulation parameters for stress dependent polymerase velocity (equation 12 of main text) were $v_{max} = 50, \kappa = 1/2pN^{-1}, \epsilon_c = 12pN$. The drag coefficient associated with polymerase rotation (equation 5 of main text) has a constant term γ and a term that increases with the distance x_i of the polymerase from the transcription start site. This form follows from polymer biophysics [21]. However, the precise parameter values for mRNA and polymerase have not been empirically established. $\gamma = 10^{-1}$, $\eta = 10^{-4}$, $\alpha = 2$ were used unless otherwise noted. Figure S1 examines the impact of varying these parameters on the main results of the paper. Figure S1a shows that for any parameter values that result in large drag ($\eta x^{\alpha} >> 1$) the velocity plateaus at a value less than $v_{max} = 50$. This is because a large drag causes significant supercoiling density build-up that impacts the elongation velocities. However if ($\eta x^{\alpha} < 1$) supercoiling density does not build up and elongation velocities can proceed at rates similar to v_{max}. The size of ηx^{alpha} (and thus the drag associated with rotation) is dependent on the length of the gene. Consequentially, even if the values for α and η are small for long enough genes supercoiling density will eventually build-up and impact elongation velocities. All values of the parameters α and η generate a mono-tonic relationship between the average mRNA number and mRNA fluctuations (Fano factor) as shown in S1b. These results show that the model robustly generates the correct behavior for a wide range of parameter values.

Phenomenological form of the torque of supercoiled DNA

DNA mechanics relax on time-scales much faster than polymerase motion [40]. Due to this fact it is not necessary to explicitly simulate the diffusion of twist between polymerases. Instead we can simply numerically integrate the twist angles of equation 7 of the main text assuming infinitely fast propagation of twist relaxation between polymerases. Thus, throughout the article we employed a steady-state relationship [23] between the level of supercoiling density and corresponding torque in DNA $\tau(\sigma)$. Our framework allows for buckling to change the torque inside DNA as a function of the supercoiling density. Following the phenomenological approach given by Marko [23] the torque in a given piece of DNA held at a constant force f is specified by the supercoiling density as

$$\tilde{\tau}(\sigma) = \begin{cases} S\sigma, & |\sigma| < \sigma_s^* \\ \tau_0 \; sign(\sigma), & \sigma_s^* < |\sigma| < \sigma_p^* \\ P\sigma, & \sigma_p^* < |\sigma| \end{cases} \tag{20}$$

where the coefficients S, τ_o, P and transition values σ_s^*, σ_p^* are given by DNA mechanical constants and are a function of applied force (see [23]). The average force f inside DNA sets the transition values. Throughout the main text a value of $f = 1pNnm$ was used. This is a sensible value given the basic polymeric nature of DNA at these scales. To check the robustness of the results of the main text against varying the force f we ran the simulation at both $f = 1/10pN$ as well as $f = 10pN$. Both values show similar results to simulations run with $f = 1pN$ (see figure S1).

Different phenomenological forms of topoisomerase action

As mentioned in the main text , we tried multiple phenomenological forms relating $\Omega(t)$, the overall supercoiling density in the system (see equation 13 of main text), to the rate of topoisomerase action: a constant rate λ independent of $\Omega(t)$; a rate proportional to $\Omega(t)$; and a rate that approaches zero with increasing $\Omega(t)$ (equation 14 of main text). The rate that approaches zero with increasing $\Omega(t)$ was used throughout the main text. This was because it was the only form which produced results that matched empirical observations for both the elongation velocity and the relationship between average mRNA output and mRNA fluctuations. Topoisomerase action which occurred with a constant rate as well as action which occurred with rates linearly proportional to $\Omega(t)$ were also tried. The results are shown in figure S2 for the average elongation rates as a function of initiation rates R_{int} and fluctuations in mRNA production as a function of average number of mRNAs. Figure S2a

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shows that a constant rate of topoisomerase action or one that scales linearly with $\Omega(t)$ both limit the average velocities below v_{max}. However, both of these forms also demonstrated non-monotonic relationships between the rate of initiation R_{int} and average velocity (Figure S2a) for varying rates λ. Additionally, both forms showed a non-monotonic relationship between average mRNA output and mRNA fluctuations (Figure S2b). These results collectively led to the elimination of these phenomenological forms of topoisomerase action for our model.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

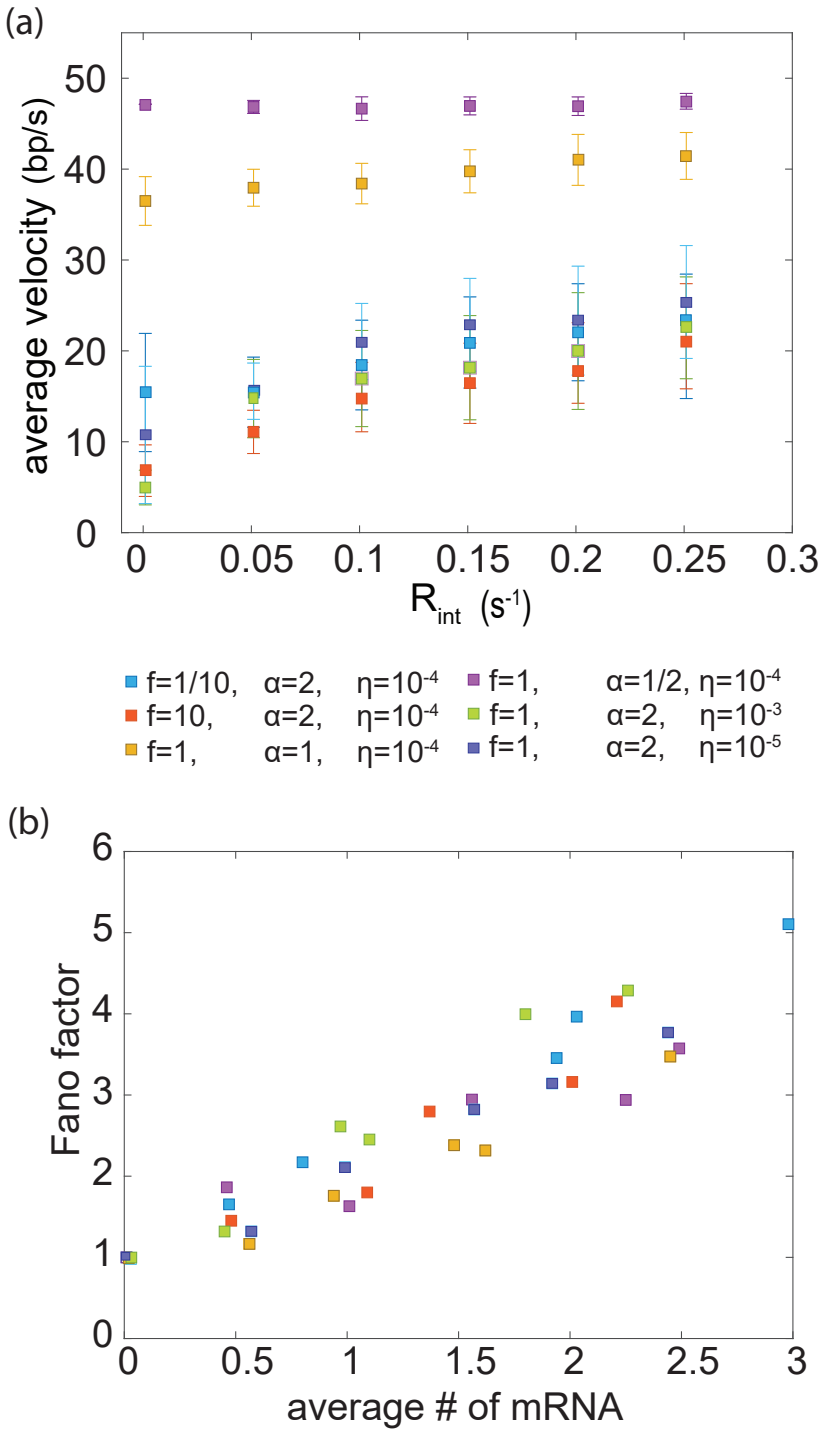

FIG. 6: Elongation velocities and the relationship between mean mRNA output and mRNA fluctuations are robust against simulation parameters. (a)The average elongation rate plateaus at a value less than the maximum elongation rate of polymerases as long as sufficient supercoiling at any DNA force value f. If the drag associated with polymerase rotation is not significant (η, α are small) then only long genes will create significant supercoiling build up. (b) All parameters values correctly demonstrate a monotonic relationship between the mean mRNA output and mRNA output fluctuations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469850; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

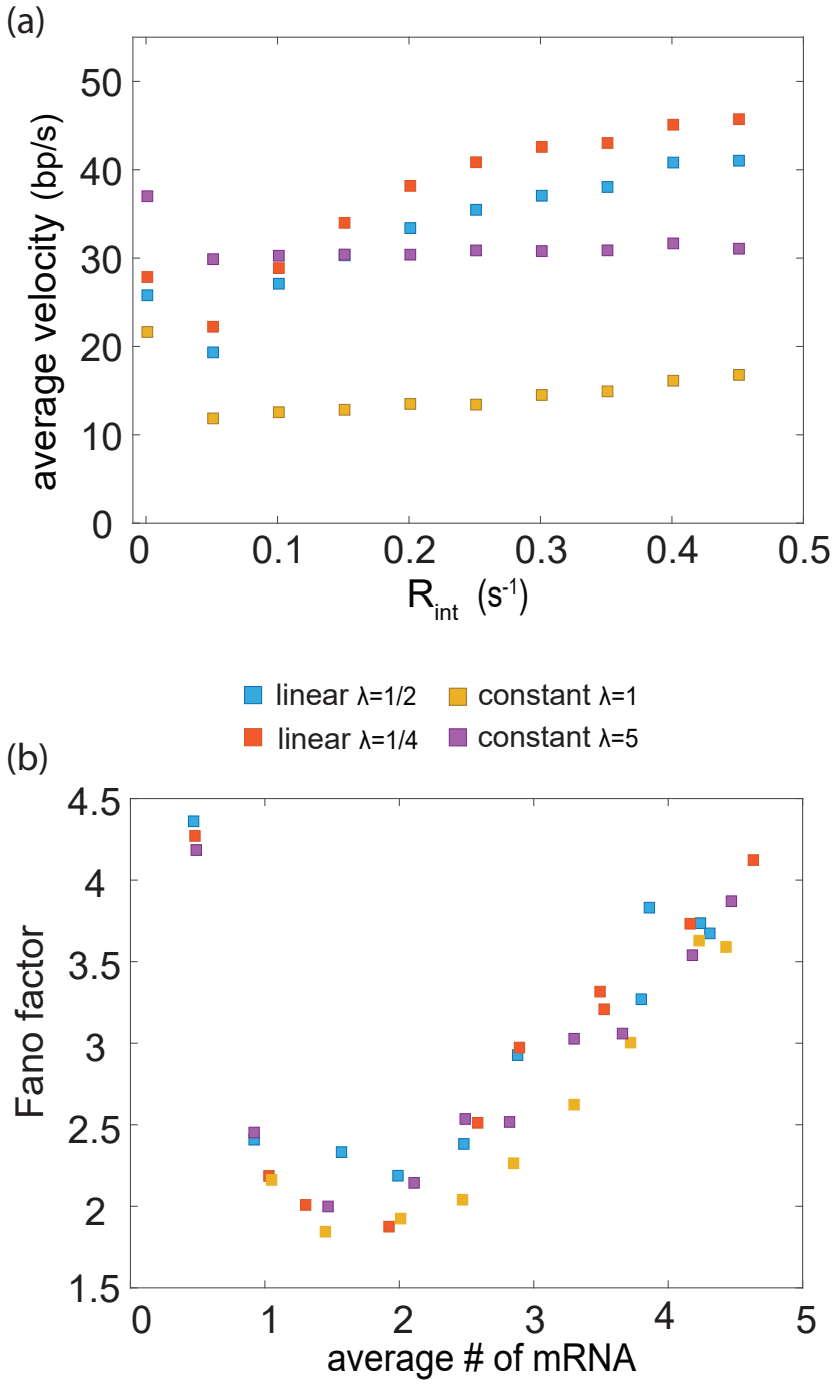

FIG. 7: Constant rates of topoisomerase action or rates that linearly depend on supercoiling density do not recapitulate empirical observations. (a)The average elongation rate shows a non-monotonic dependence on the initiation rate for both a constant and linear rate of topoisomerase action. (b) Both a constant and linear rate of topoisomerase action result in a non-monotonic relationship between the average mRNA output and the fluctuations in mRNA. These two results demonstrate that neither a constant nor a rate of topoisomerase action that linearly depends on supercoiling density can recapitulate empirical observations and were therefore not used in the model.