

RESEARCH ARTICLE

4D nucleome equation predicts gene expression controlled by long-range enhancer-promoter interaction

Zihao Wang^{1,2}, Songhao Luo^{1,2}, Zhenquan Zhang^{1,2}, Tianshou Zhou^{1,2*}, Jiajun Zhang^{1,2*}

1 Guangdong Province Key Laboratory of Computational, Sun Yat-sen University, Guangzhou, People's Republic of China, **2** School of Mathematics, Sun Yat-Sen University, Guangzhou, People's Republic of China

* mcszhtsh@mail.sysu.edu.cn (TZ); zhjjajun@mail.sysu.edu.cn (JZ)



OPEN ACCESS

Citation: Wang Z, Luo S, Zhang Z, Zhou T, Zhang J (2023) 4D nucleome equation predicts gene expression controlled by long-range enhancer-promoter interaction. *PLoS Comput Biol* 19(12): e1011722. <https://doi.org/10.1371/journal.pcbi.1011722>

Editor: Maxwell Wing Libbrecht, Simon Fraser University, CANADA

Received: April 14, 2023

Accepted: November 28, 2023

Published: December 18, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1011722>

Copyright: © 2023 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The source code and data used to produce the results and analyses presented in this manuscript are available from

Abstract

Recent experimental evidence strongly supports that three-dimensional (3D) long-range enhancer-promoter (E-P) interactions have important influences on gene-expression dynamics, but it is unclear how the interaction information is translated into gene expression over time (4D). To address this question, we developed a general theoretical framework (named as a 4D nucleome equation), which integrates E-P interactions on chromatin and biochemical reactions of gene transcription. With this equation, we first present the distribution of mRNA counts as a function of the E-P genomic distance and then reveal a power-law scaling of the expression level in this distance. Interestingly, we find that long-range E-P interactions can induce bimodal and trimodal mRNA distributions. The 4D nucleome equation also allows for model selection and parameter inference. When this equation is applied to the mouse embryonic stem cell smRNA-FISH data and the E-P genomic-distance data, the predicted E-P contact probability and mRNA distribution are in good agreement with experimental results. Further statistical inference indicates that the E-P interactions prefer to modulate the mRNA level by controlling promoter activation and transcription initiation rates. Our model and results provide quantitative insights into both spatiotemporal gene-expression determinants (i.e., long-range E-P interactions) and cellular fates during development.

Author summary

Gene expression is an essential biological process in all organisms. Numerous experimental studies have reported that the long-range enhancer-promoter (E-P) interaction on three-dimensional (3D) chromatin architecture plays important roles in regulating gene expression and cell functions, but the quantitative and qualitative impact of E-P interaction on gene expression over time is unclear. We develop a theoretically and numerically efficient model (called the 4D nucleome equation) to couple E-P interaction with gene expression and use this equation to characterize dynamic behavior. Then, we obtain the

GitHub (<https://github.com/cellfate/4DNucleomeEquation>).

Funding: This work was supported by the National Key R&D Program of China (2021YFA1302500 to JZ), the Natural Science Foundation of P. R. China (12171494 to JZ; 11931019 to TZ; 11775314 to TZ; 62373384 to TZ; 12301646 to ZZ), the Guangdong Basic and Applied Basic Research Foundation (2022A1515011540 to JZ; 2023A1515011982 to ZZ), Key-Area Research and Development Program of Guangzhou, P. R. China (2019B110233002 to JZ; 202007030004 to TZ), the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University (2020B1212060032 to JZ), the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (23qnpy48 to ZW; 23qnpy49 to ZZ), and the China Postdoctoral Science Foundation (2023M734061 to ZW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

theoretical distribution of mRNAs and predict the gene expression profiles under E-P regulations. Interestingly, we find that E-P interactions can induce bimodal and trimodal shapes of mRNA distribution. When applying this framework to mouse embryonic stem cell data to investigate the dynamical behaviors of E-P interaction and gene expression, we reproduce the experimentally measured E-P contact frequencies and mRNA distributions under different E-P interactions. Our results support the picture of an essential mechanism for explaining phenotypic diversity and cellular decision-making.

Introduction

Gene expression is tightly related to three-dimensional (3D) genome conformation that may change over 1D time [1–4]. Specific DNA sequences—promoters and enhancers—orchestrate transcription in a highly complex and multilayered manner to ensure accurate spatiotemporal gene expression [1–4]. Many experimental studies have shown the importance of the roles of distal enhancers in ensuring reliable cell functioning and cellular decision-making [5–10].

However, the mechanism of how 3D chromatin organization (in particular 3D enhancer-promoter (E-P) interactions) in time (4D) shapes gene-expression dynamics still remains elusive.

Hierarchic genomic structures support various possible E-P topologies and the connection of upstream stochastic E-P interaction to downstream gene transcription [11–16]. Many efforts have been made to explore the essential factors of E-P interaction affecting transcription. Recent live-imaging measurements have provided clear evidence that E-P genomic distance effectively controls gene activities [6,10,17]. And a collection of experimental evidence has established that the E-P interaction strength significantly impacts gene expression levels [6,7,18,19]. For example, the *sna* shadow enhancer, whose strength is determined by chromatin inheritance, generates more bursts than the *sna* primary enhancer in *Drosophila* embryos [6], and the hormone or heavy metal exposure, which externally alters the E-P interaction strength [20,21], boosts the mRNA level of *c-Fos* gene with the increase of dose concentrations [18]. These experimental observations indicate that the E-P genomic distance and the E-P interaction strength are important factors impacting gene expression profiles. However, biological experiments alone are not sufficient to unravel the complete picture of the dynamics of E-P interaction-regulated gene expression, and it is necessary to develop biologically reasonable mathematical models to investigate the underlying mechanism.

The conventional modeling of gene expression kinetics is based on simple models such as the two-state model and multistate model [22–30], in which an implicit hypothesis is that chromatin behavior is frozen. Recent work adds the effects of chromatin structure to the model [10,31], but still ignores the spatiotemporal dynamic regulation of chromatin topologies [31–35]. So far, we still lack a mechanistic mathematical model that couples stochastic chromatin organizations and stochastic gene expression processes. Addressing this issue faces two challenges. First, genomic structures are stochastic at almost every level of organization, and this stochasticity is suggestively linked to gene transcription and finally affects transcriptional outcomes [14,32]. Overall, the temporal disconnection, as well as the stochasticity of E-P topology and gene expression, lies at the heart of a broad challenge in the physical biology of both establishing a comprehensive theoretical framework of the information transmission from the upstream chromatin organization to the downstream gene expression and forecasting mRNA profiles from the dynamics of underlying molecular processes. Second, the regulation of gene expression by the chromatin spatial structure is a multiscale system. Many experimental studies have indicated distinct timescale differences between upstream chromatin dynamics and

downstream gene expression [36,37]. For instance, E-P interaction occurs on a timescale of seconds to minutes [8,38,39], whereas the gene produces over a longer timescale compatible with E-P interaction, precisely the timescale of minutes to hours as suggested in multiple studies [9,36,40].

Here, we developed a general theoretical framework (formulated as the so-called 4D nucleome equation) to investigate how E-P interaction (characterized by E-P genomic distance and E-P interaction strength) affects gene expression dynamics. Specifically, this framework considered upstream chromatin motion on a fast timescale and downstream mRNA production on a slow timescale as well as the temporal connection between the upstream and the downstream. This framework involving space and time is referred as 4D gene expression kinetics [41,42]. With the timescale separation method, we derived analytical mRNA distribution and studied how E-P interaction qualitatively impacts the characteristics of mRNA distribution. Importantly, we found that the E-P interaction can flexibly regulate mRNA patterns by inducing multiple shapes of mRNA distributions including bimodal distribution with two non-origin peaks and even trimodal distribution. And our theoretical analysis and simulations revealed a power-law scaling of gene expression levels in the E-P genomic distance. Finally, by performing statistic inference on the mouse embryonic stem cells (mESCs) data, our framework exhibited good predictions of mRNA distribution and E-P contact probability under different E-P interactions. We emphasize that our 4D nucleome equation provides a general modeling framework for studying how chromatin dynamics affect gene expression kinetics and our results suggest a possible mechanism for explaining phenotypic diversity and cellular decision-making in realistic cases.

Materials and methods

In this section, we introduce a theoretical framework for predicting gene expression regulated by long-range E-P interaction in 4D (Fig 1A). First, we model chromatin as a polymer [43–46] discretized into a collection of successive monomers, and assume that there are N monomers on chromatin, spatial positions of which are denoted by $\mathbf{r} = [r_1, \dots, r_N]^T$. Each monomer represents a segment of DNA, whose length depends on the levels of coarse-grained chromatin. Second, we model the gene expression process as a discrete multistate model, and assume that the gene has K different gene states (each state includes promoter’s state (ON or OFF) and the number of mRNA), which altogether constitute the vector $\mathbf{s} = [s_1, \dots, s_K]^T$.

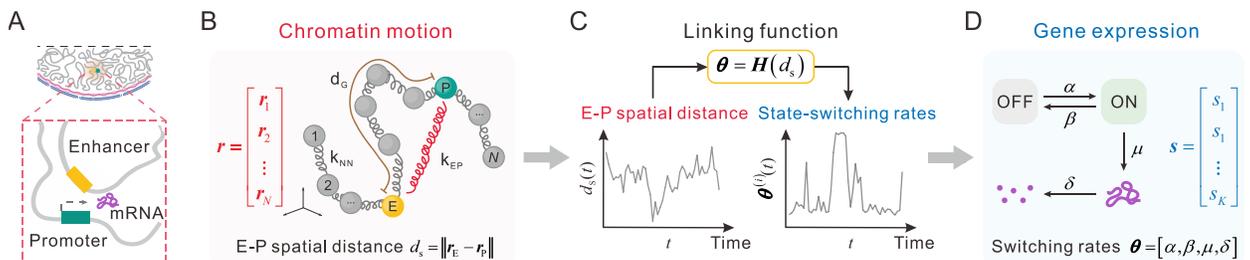


Fig 1. Schematic representation of gene expression regulated by E-P interaction. (A) E-P interaction in the cell nucleus plays a key role in regulating gene expression. (B) A polymer model involving E-P interaction (the red spring with coefficient k_{EP}) is proposed to simulate chromatin dynamics, where $\mathbf{r} = [r_1, \dots, r_N]^T$ represents monomers positions in 3D space, d_G is E-P genomic distance, and $d_s = \|\mathbf{r}_E - \mathbf{r}_P\|$ is E-P spatial distance. (C) A link function vector $\boldsymbol{\theta} = \mathbf{H}(d_s)$, where \mathbf{H} is a function vector, bridges the temporal disconnection between upstream E-P topology and downstream gene expression. (D) The two-state telegraph model is used to imitate gene expression, where $\mathbf{s} = [s_1, \dots, s_K]^T$ is the vector of the gene’s states and $\boldsymbol{\theta} = [\alpha, \beta, \mu, \delta]$ is the vector of state-switching rates.

<https://doi.org/10.1371/journal.pcbi.1011722.g001>

4D nucleome equation governing probabilistic evolutionary behavior in phase space

Let $\mathbf{p}(\mathbf{r}, \mathbf{s}; t)$ be a vector of the joint probability density functions that monomers are in position \mathbf{r} and the gene is in state \mathbf{s} at time t . Specifically, $\mathbf{p}(\mathbf{r}, \mathbf{s}; t) = [p(\mathbf{r}, s_1; t), \dots, p(\mathbf{r}, s_K; t)]^T$. Noting $\mathbf{p}(\mathbf{r}, \mathbf{s}; t) = p(\mathbf{r}; t)\mathbf{p}(\mathbf{s}|\mathbf{r}; t)$, we have

$$\frac{\partial \mathbf{p}(\mathbf{r}, \mathbf{s}; t)}{\partial t} = \frac{\partial p(\mathbf{r}; t)}{\partial t} \mathbf{p}(\mathbf{s}|\mathbf{r}; t) + \frac{\partial \mathbf{p}(\mathbf{s}|\mathbf{r}; t)}{\partial t} p(\mathbf{r}; t). \tag{1}$$

On one hand, the motion of monomers has a continuous trajectory in the region Ω (a connected and bounded domain). The derivative $\partial \mathbf{p}(\mathbf{r}; t) / \partial t$ in Eq [1] can be formally written as $\partial p(\mathbf{r}; t) / \partial t = -\nabla_r \cdot [\mathbf{F}(\mathbf{r}, \mathbf{s}; t) p(\mathbf{r}; t)]$, where ∇_r is the gradient operator and $\mathbf{F}(\mathbf{r}, \mathbf{s}; t)$ is a velocity field. Next, if we consider isotropic diffusion and friction across the region of monomers' positions, then $\mathbf{F}(\mathbf{r}, \mathbf{s}; t)$ takes a generalized Fokker–Planck approximation, i.e., $\mathbf{F}(\mathbf{r}, \mathbf{s}; t) = \mathbf{V}(\mathbf{r}, \mathbf{s}; t) - \nabla_r(D \log p(\mathbf{r}; t))$, where the first term represents the deterministic part of the velocity field and the second term is a stochastic ingredient of the velocity under isotropic diffusion with diffusion coefficient D . We further assume that changes in gene state do not contribute to chromatin motion, and $\mathbf{V}(\mathbf{r}, \mathbf{s}; t)$ can be approximated by $\mathbf{V}(\mathbf{r}; t)$. Thus, we can obtain

$$\frac{\partial p(\mathbf{r}; t)}{\partial t} = -\nabla_r \cdot (\mathbf{V}(\mathbf{r}; t) p(\mathbf{r}; t)) + \nabla_r^2 (D p(\mathbf{r}; t)), \tag{2}$$

where ∇_r^2 is the Laplace operator.

On the other hand, the stochastic gene expression process regulated by chromatin dynamics can be modeled by the master equation of the form

$$\frac{\partial \mathbf{p}(\mathbf{s}|\mathbf{r}; t)}{\partial t} = \mathbf{W}(\mathbf{r}; t) \mathbf{p}(\mathbf{s}|\mathbf{r}; t), \tag{3}$$

where $\mathbf{W}(\mathbf{r}; t)$ is a monomer position-dependent state transition matrix, and the elements of \mathbf{W} are related to gene state switching rates and chromatin position.

We assume that tiny changes in monomers coordinates do not alter gene state, implying that the derivative of conditional probability, $\nabla_r \mathbf{p}(\mathbf{s}|\mathbf{r}; t)$, approximately equals zero partly because the time interval of gene state transition is generally longer than that of chromatin motion [36]. Thus, substituting Eq [2] and Eq [3] into Eq [1] yields the following equation

$$\frac{\partial \mathbf{p}(\mathbf{r}, \mathbf{s}; t)}{\partial t} = \underbrace{-\nabla_r \cdot (\mathbf{p}(\mathbf{r}, \mathbf{s}; t) \mathbf{V}(\mathbf{r}; t)^T)}_{\text{chromatin dynamics}} + \underbrace{\nabla_r^2 (D \mathbf{p}(\mathbf{r}, \mathbf{s}; t)) + \mathbf{W}(\mathbf{r}; t) \mathbf{p}(\mathbf{r}, \mathbf{s}; t)}_{\text{gene-expression dynamics}} \tag{4}$$

The first two terms on the right-hand side of Eq [4] represent chromatin's spatiotemporal diffusion process. The first term is the deterministic component, and the second is the stochastic component accounting for random fluctuations. The last term captures the gene states' random switching process. It should be noted that Eq [4] is a comprehensive description of the gene expression process toward the 4D reality, so we called it the 4D nucleome equation. In the Results section, we will derive the mRNA distribution based on Eq [4].

Stochastic model simulating trajectory evolutionary behavior in configuration space

Modeling chromatin dynamics. Successive monomers in the chromatin are connected with harmonic springs of stiffness k_{NN} (Fig 1B). Each monomer represents a nucleosome or a

DNA segment (this length does not affect our qualitative results to be obtained) with the 3D position denoted by $\mathbf{r}_i = (r_{i1}, r_{i2}, r_{i3})$, where $i = 1, \dots, N$. We employ one nucleosome to represent an enhancer or promoter and posit that there are only one enhancer and one promoter on the chromatin. To mimic E-P interaction and to simplify without loss of generality, we add a harmonic spring with stiffness k_{EP} between the enhancer and the promoter to materialize this abstract concept.

Based on the above assumptions, we model the chromatin dynamics according to the Langevin equation $d\mathbf{r} = \mathbf{V}(\mathbf{r}; t)dt + \sqrt{2D}d\mathbf{B}(t)$, which is equivalent to Eq [2], where $B(t)$ is a vector of independent Brownian motions. And $\mathbf{V}(\mathbf{r}; t) = -\nabla_r U(\mathbf{r}; t)/\gamma$, where γ is the friction coefficient and $U(\mathbf{r}; t)$ is the total potential of chromatin conformation. Note that

$U(\mathbf{r}; t) = U_{NN}(\mathbf{r}; t) + U_{EP}(\mathbf{r}; t)$, where without loss of generality, we set $U_{NN}(\mathbf{r}; t) =$

$(1/2) \sum_{j=1}^{N-1} k_{NN}(\mathbf{r}_j - \mathbf{r}_{j+1})^2$ which represents the potential for the chain connection, and

$U_{EP}(\mathbf{r}; t) = (1/2)k_{EP}(\mathbf{r}_E - \mathbf{r}_P)^2$ which represents the potential for the E-P interaction with E, $P \in \{1, \dots, N\}$ representing the index of the monomer occupied by enhancer and promoter respectively. We let $d_G = |E-P|$ represent the E-P genomic distance, which can be directly measured by an experimental method. However, the k_{EP} representing the E-P interaction strength cannot be directly measured by experiments but can be estimated from experimental data, e.g., Hi-C data [47] or Capture-C data [10].

Modeling gene expression dynamics. Gene expression process can be characterized by a two-state model with an active ON and a silent OFF state (Fig 1D). The switching rates from OFF to ON and vice versa are α and β , respectively. The transcription rate is μ , and the mRNA degradation rate is δ . We define $\theta = [\alpha, \beta, \mu, \delta]$ as the gene-state switching rates set.

Bridging chromatin dynamics and gene expression. After having identified chromatin conformations and transcriptional reactions independently, we next build a biologically reasonable link between them (Fig 1C). E-P interaction carries regulatory information to orchestrate gene expression, and we hypothesize that the time-varying E-P spatial distance $d_s = \|\mathbf{r}_E - \mathbf{r}_P\|$ encodes the information to regulate the transcription-associated reaction rates θ . That is $\theta = \mathbf{H}(d_s)$, where \mathbf{H} is a link function vector. Therefore, the two-state model mentioned above becomes a variable two-state model in which the rates θ depend on the E-P spatial distance d_s .

Recent experimental results show that E-P proximity increases the likelihood of gene expression [14]. It is reasonably assumed that if d_s is smaller, then α and μ are larger, but β is smaller (in the case that δ is constant). For simplification, we assume β is independent of d_s and $\alpha(\mu)$ changes between $\alpha_{\max}(\mu_{\max})$ and $\alpha_{\min}(\mu_{\min})$ with the fluctuating d_s . We choose a Hill function [48] to reflect the effect of changes in the E-P spatial distance d_s on the rates θ , although the specific shape of the corresponding response curve can also be captured by other functions. Specifically, we assume that the rates α and μ depend nonlinearly on d_s , i.e., each is a Hill-like function vector \mathbf{H} (see S1 Text).

Under the above settings, we propose a stochastic simulation algorithm to simulate the time evolution of the entire system (see S1 Text). In Result section, we will study how E-P interaction strength k_{EP} and E-P genome distance d_G regulate gene expression and how they affect the shape and characterization of mRNA distributions. In short, our modeling strategy provides a possible framework for characterizing 3D chromatin motion and tracking gene-expression processes over 1D time [41,42].

Model-data approach to infer E-P interactions and gene-expression kinetics

We use the above model to fit the mRNA distribution of single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) in mESCs [10]. The data includes 6 cell lines C_k

($k = 1, \dots, 6$). In each cell line, the enhancer is placed at different positions from the promoter to drive the expression of enhanced green fluorescent protein (eGFP), so we can get the E-P genomic distance and corresponding mRNA distribution with the bin number n_k as well as the corresponding steady-state probability distribution $Q_k(X = x_i)$ ($1 \leq i \leq n_k$).

For each cell line C_k , we theoretically calculate the steady-state probability distribution with parameters Γ (including E-P interaction parameters and gene expression parameters). Then, we discretize the steady-state distribution $P_k(X = x_i; \Gamma)$ ($1 \leq i \leq n_k$) that is comparable with the experiment data $Q_k(X = x_i)$. Note that the cross entropy of the cell line C_k is given by

$$H_k(\Gamma) = - \sum_{i=1}^{n_k} Q_k(X = x_i) \log(P_k(X = x_i; \Gamma)), \tag{5}$$

and the best-fit parameters by minimizing the total cross entropy function

$$\arg \min_{\Gamma} H(\Gamma) = \arg \min_{\Gamma} - \sum_{k=1}^6 \sum_{i=1}^{n_k} Q_k(X = x_i) \log(P_k(X = x_i; \Gamma)). \tag{6}$$

In fact, the minimum cross-entropy method is the same as maximum likelihood estimation. To solve this optimization problem, we use the *fmincon* function (a nonlinear programming solver) in the LBFGS method of MATLAB to find the minimum value of the optimization problem given a set of initial values and parameter intervals (S1 Text).

Results

Analytical results of steady mRNA distribution

To study the qualitative and quantitative effects of E-P interaction (including E-P genomic distance d_G and E-P interaction strength k_{EP}) on gene expression (including the mRNA distribution, the mean mRNA expression level and the coefficient of variation (CV) defined as the ratio of standard deviation over mean), we use the 4D nucleome equation (Eq [4]) to solve the steady mRNA distribution $P(x)$ depending on the E-P spatial distance d_s .

Note that the entire gene expression system contains two modules—upstream chromatin conformation on a fast timescale and downstream gene expression on a slow timescale. If the typical timescales for the former and latter are denoted by τ_{up} and τ_{down} respectively, Eq [4] can be rewritten as the rescaled equation

$$\frac{\partial \mathbf{p}(\mathbf{r}, \mathbf{s}; t)}{\partial t} = - \frac{1}{\tau_{up}} [\nabla_{\mathbf{r}} \cdot (\mathbf{p}(\mathbf{r}, \mathbf{s}; t) \tilde{\mathbf{V}}(\mathbf{r}; t)^T) + \nabla_{\mathbf{r}}^2 (\tilde{\mathbf{D}} \mathbf{p}(\mathbf{r}, \mathbf{s}; t))] + \frac{1}{\tau_{down}} [\tilde{\mathbf{W}}(\mathbf{r}; t) \mathbf{p}(\mathbf{r}, \mathbf{s}; t)], \tag{7}$$

where $\tilde{\mathbf{V}} = \tau_{up} \mathbf{V}$, $\tilde{\mathbf{D}} = \tau_{up} \mathbf{D}$, $\tilde{\mathbf{W}} = \tau_{down} \tilde{\mathbf{W}}$. Deriving the mRNA distribution directly from the above multiscale system is a particular challenging task. Here we resort to a timescale separation method. Before this, however, we consider the E-P interaction and gene expression dynamics independently.

When focusing only on the motion of upstream chromatin conformation, we find that the stationary probability density function of chromatin conformation \mathbf{r} can be expressed as $p^{\text{stat}}(\mathbf{r}) = \prod_{i=1}^3 p_i(\mathbf{r})$ since every monomer moves independently in each dimension. Notably, we find that E-P spatial distance d_s obeys the following exact Maxwell-Boltzmann distribution (Fig 2A and S1 Text)

$$p_{DS}(d_s) = \sqrt{\frac{2}{\pi}} \Theta^{-3} d_s^2 \exp\left(-\frac{d_s^2}{2\Theta^2}\right), \tag{8}$$

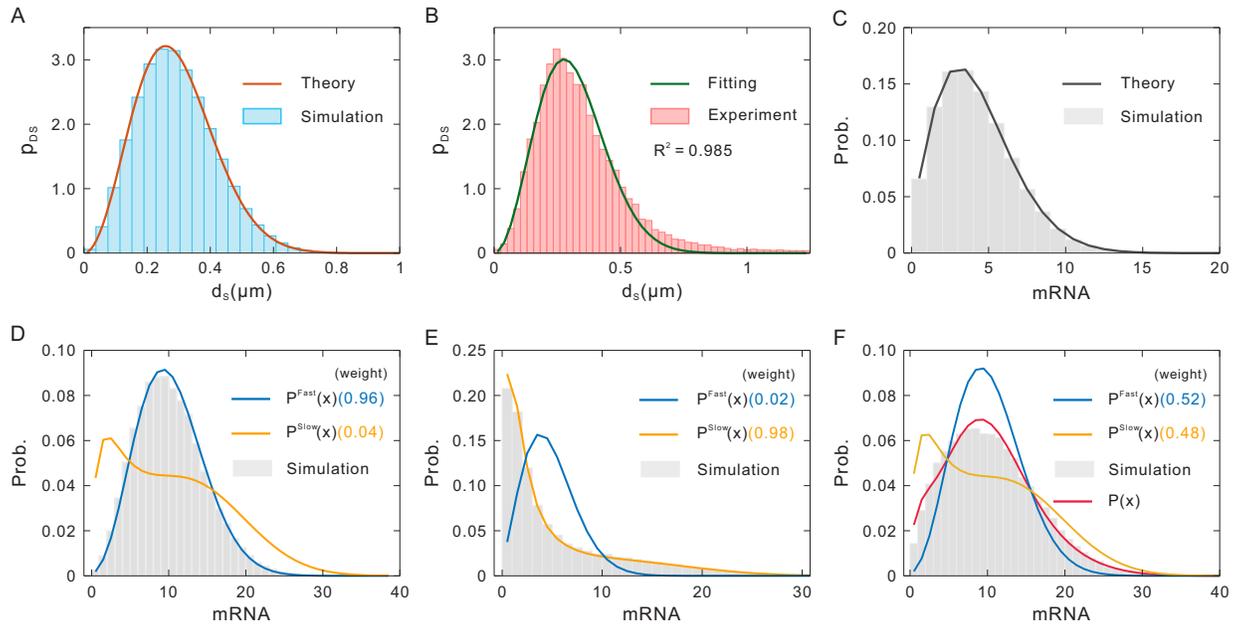


Fig 2. Distributions of gene expression model. (A) The distribution of E-P spatial distance (Maxwell-Boltzmann distribution given by Eq [8]) where red lines represent theoretical results, and histograms represent numerical results. (B) Experimental measurement of 3D spatial distance between Sox2 and its enhancer in living ESCs (histogram, [49]). The blue line is obtained by fitting Eq [8]. (C) The mRNA distribution conditional on frozen chromatin structure (Poisson-Beta distribution) where $\alpha = 0.30, \beta = 0.50, \mu = 1$ and $\delta = 0.1$. (D) Upstream chromatin fluctuations are much faster compared with the downstream state switching rates ($\omega = 0.04$), where the values in parentheses indicate the weight of the interpolation distribution. Parameter values are set as $\gamma = 0.1, k_{EP} = 0.3$. (E) Upstream chromatin fluctuations are slow ($\omega = 59.95$). Parameter values are set as $\gamma = 100, k_{EP} = 0.1$. (F) Upstream chromatin fluctuations and downstream state switching rates are comparable ($\omega = 0.91$). Parameter values are set as $\gamma = 5, k_{EP} = 0.3$. Other parameters in (D-F) are set as $\alpha_{max} = 0.60, \alpha_{min} = 0.30, \beta = 0.50, \mu_{max} = 3, \mu_{min} = 0.1, \delta = 0.1$.

<https://doi.org/10.1371/journal.pcbi.1011722.g002>

where the lumping parameter $\Theta = \sqrt{D\gamma(k_{NN}/d_G + k_{EP})^{-1}}$ determines how changes in E-P interaction strength k_{EP} or E-P genomic distance d_G alter the shape of the $p_{DS}(d_s)$. Experimental measurement of 3D spatial distance between Sox2 and its enhancer in living ESCs [49] or between the *even-skipped (eve)* locus and its enhancer in the fly embryo [50] can be well fitted by Eq [8], confirming the validity of theoretical analysis (Fig 2B).

When focusing only on the downstream gene expression, that is, given a d_s (or a frozen E-P topology), we find that the stationary probability distribution of mRNA abundance $P_{mRNA}(x)$ is a Poisson-Beta distribution (Fig 2C, [51]). That is,

$$P_{mRNA|d_s}(x) = \text{Poisson}(x; \tilde{\mu}t) \wedge \text{Beta}(t; \tilde{\alpha}, \tilde{\beta})$$

$$= \int_0^1 \frac{(\tilde{\mu}t)^x}{x!} e^{-\tilde{\mu}t} \frac{t^{\tilde{\alpha}-1}(1-t)^{\tilde{\beta}-1}}{B(\tilde{\alpha}, \tilde{\beta})} dt, \tag{9}$$

where \wedge represents the mixture of two distributions, and the gene-state switching rates are in the unit of the mRNA decay rate, i.e., $\tilde{\alpha} = \alpha/\delta, \tilde{\beta} = \beta/\delta, \tilde{\mu} = \mu/\delta$ (Fig 2C).

Now, we turn to derive mRNA distribution by using the timescale separation method. First, we assume that the upstream process is much faster than the downstream process, i.e., we only consider the limit case of $\tau_{up}/\tau_{down} = \zeta \ll 1$. By rescaling t to t/τ_{down} , we can then assume the formal solution $p(\mathbf{r}, \mathbf{s}; t) = \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t) + \zeta \mathbf{p}^{(1)}(\mathbf{r}, \mathbf{s}; t) + O(\zeta^2)$. Considering the leading order of ζ and the stationary condition, we have $0 = \nabla_r \cdot (\mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t) \tilde{\mathbf{V}}(\mathbf{r}; t)^T) + \nabla_r^2 (\tilde{\mathbf{D}}\mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t))$.

Since we have neglected the regulation of gene state by chromatin structure, the zero-th order solution $\mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t)$ can be decoupled as $\mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t) = \mathbf{p}^{\text{stat}}(\mathbf{r})\mathbf{p}(\mathbf{s}; t)$, implying that the quasi-stationary of chromatin motion is well separated from the downstream transcription. In general, the stationary mRNA probability distribution is given through (Fig 2D)

$$\begin{aligned}
 P^{\text{Fast}}(x) &= P_{\text{mRNA}|(DS)}(x) \int_0^{+\infty} \mathbf{H}(d_s) P_{DS}(d_s) dd_s \\
 &= \text{Poisson}(x; \langle \tilde{\mu}(d_s) \rangle_t) \wedge \text{Beta}(t; \langle \tilde{\alpha}(d_s) \rangle, \langle \tilde{\beta}(d_s) \rangle),
 \end{aligned}
 \tag{10}$$

where $\langle DS \rangle$ is the expectation of E-P spatial distance and $P_{\text{mRNA}|(DS)}(x)$ (the distribution conditional on the averaged d_s) is given through Eq [9].

Second, we consider the opposite limit, i.e., $\tau_{up}/\tau_{down} = \zeta \gg 1$. In this limit and similar to the analysis in the case of $\tau_{up}/\tau_{down} = \zeta \ll 1$, we find that $\mathbf{p}^{(0)}(\mathbf{r}, \mathbf{s}; t) = p(\mathbf{r}; t)\mathbf{p}^{\text{stat}}(\mathbf{s}|\mathbf{r})$, where the equation group $\tilde{\mathbf{W}}(\mathbf{r})\mathbf{p}^{\text{stat}}(\mathbf{s}|\mathbf{r}) = \mathbf{0}$ determines $\mathbf{p}^{\text{stat}}(\mathbf{s}|\mathbf{r})$. Therefore, the chromatin spatial positions are not independent of the downstream gene expression. But the stationary mRNA probability distribution can be calculated according to (Fig 2E)

$$\begin{aligned}
 P^{\text{Slow}}(x) &= \int_0^{+\infty} P_{\text{mRNA}|d_s}(x|\mathbf{H}(d_s)) P_{DS}(d_s) dd_s \\
 &= \text{Poisson}(x; \tilde{\mu}(d_s)t) \wedge \text{Beta}(t; \tilde{\alpha}(d_s), \tilde{\beta}(d_s)) \wedge_{d_s} \text{MaxBoltz}(d_s)
 \end{aligned}
 \tag{11}$$

where $P_{\text{mRNA}|d_s}(x)$ (i.e., the distributions conditional on the d_s) is given by Eq [9].

Finally, in the intermediate regime, we find that $P(x)$ can be well fit with the following formula (Fig 2F)

$$\begin{aligned}
 P(x) &\approx \frac{1}{1 + \omega} P^{\text{Fast}}(x) + \frac{\omega}{1 + \omega} P^{\text{Slow}}(x) \\
 &= \frac{1}{1 + \omega} \text{Poisson}(x; \langle \tilde{\mu}(d_s) \rangle_t) \wedge \text{Beta}(t; \langle \tilde{\alpha}(d_s) \rangle, \langle \tilde{\beta}(d_s) \rangle) \\
 &\quad + \frac{\omega}{1 + \omega} \text{Poisson}(x; \tilde{\mu}(d_s)t) \wedge \text{Beta}(t; \tilde{\alpha}(d_s), \tilde{\beta}(d_s)) \wedge_{d_s} \text{MaxBoltz}(d_s),
 \end{aligned}
 \tag{12}$$

where ω is a scaling factor defined as the ratio of minimum variable transition rate and maximum chromatin motion velocity. $\omega = \min\{\alpha_{\min}, \mu_{\min}\} / \max V(\mathbf{r})$, where $V(\mathbf{r}) = [(k_{NN}/d_G + k_{EP})d_s] / (b\gamma)$, b is E-P encounter distance and d_s can be selected as the maximum E-P spatial distance that can be theoretically reached (in fact, the cumulative density function of $p_{DS}(d_s)$ reaches 0.99).

It is worth mentioning that Eq [12] provides high-accuracy approximations of mRNA steady distribution and is a useful and simple formula for predicting the dynamics of mRNA over a broad range of timescale separation. We can further trace the respective contributions of the system's key parameters (e.g., d_G and k_{EP}) to the patterns of mRNA distribution, to better reveal the essential mechanism of gene expression.

E-P interaction controls the mean level and variability of gene expression

To understand how E-P interaction modulates gene expression, we use the above theoretical analysis to explore the qualitative impact of E-P genomic distance d_G and E-P interaction strength k_{EP} on gene expression. We calculate the mean ($\langle \text{mRNA} \rangle$) and the CV of the mRNA probability distribution in steady state, based on Eq [12].

Fig 3A and 3B depict how changes in d_G alter $\langle \text{mRNA} \rangle$ and CV. We find that when k_{EP} is fixed, $\langle \text{mRNA} \rangle$ monotonously decreases and CV monotonically increases with increasing d_G .

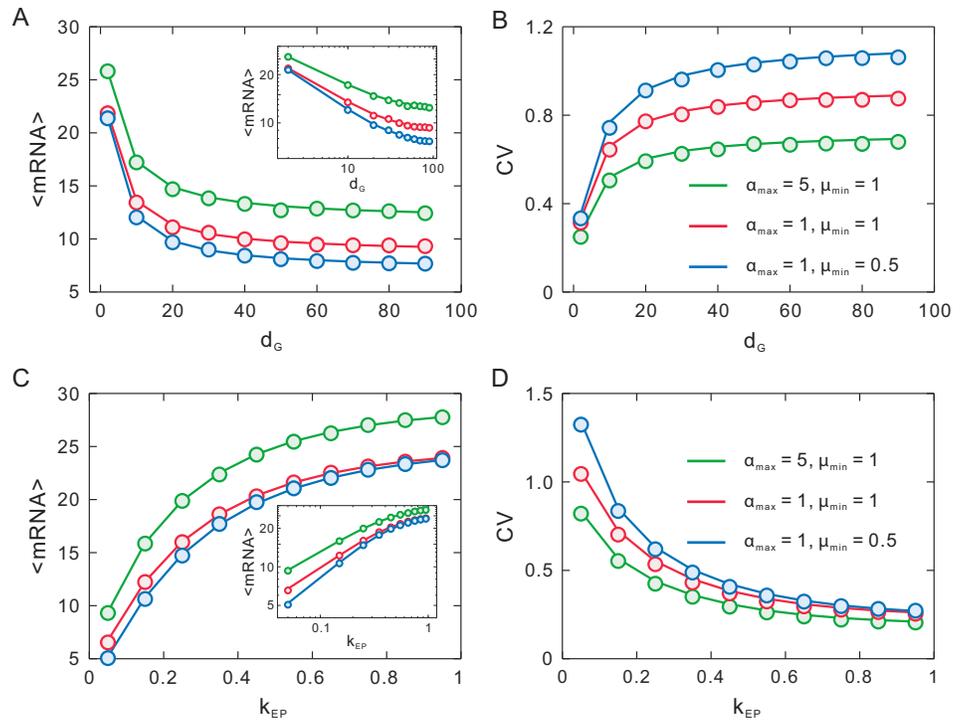


Fig 3. Influence of E-P interaction on mRNA mean and CV. (A-B) The effect of E-P genomic distance d_G on mRNA mean and CV, where solid lines represent theoretical results, and color points represent numerical results. (C-D) The effect of E-P interaction strength k_{EP} on mRNA mean and CV. The inset in (A) and (C) shows the log-log plot. Other parameter values are set as $\alpha_{min} = 0.06, \beta = 0.20, \mu_{max} = 3, \delta = 0.1, \gamma = 50$.

<https://doi.org/10.1371/journal.pcbi.1011722.g003>

Theoretically, a smaller d_G corresponds to a shorter E-P spatial distance and a more frequent E-P interaction, thus boosting the expression level. Fig 3C and 3D reveal the distinct effects of different k_{EP} on gene expression. It shows that the tendencies are completely different from those in the case of changing d_G . $\langle \text{mRNA} \rangle$ is monotonously upregulated and CV is downregulated by k_{EP} . Indeed, Eq [8] has shown the opposite effect of d_G and k_{EP} on E-P spatial distance and then on gene expression. Log-log plot represents power-law behaviors of $\langle \text{mRNA} \rangle$ regulated by d_G or k_{EP} (Fig 3A and 3C inset). Moreover, it can be seen that the enhancement or the reduction of $\langle \text{mRNA} \rangle$ (at larger d_G or k_{EP}) tends to be saturated, indicating that the effect of E-P interaction on gene expression is not infinite but limited.

Notably, the qualitative behaviors of $\langle \text{mRNA} \rangle$ are consistent with many experimental observations [6,10,17,18,52]. For example, a larger d_G between *sna* shadow enhancer and promoter generates fewer mRNAs than a smaller d_G [6,17], and a stronger enhancer (corresponding to a larger k_{EP}) produces more mRNAs than a weaker enhancer [6] in living *Drosophila* embryos. And with the increase of dose concentrations (corresponding to the increment of k_{EP}), the mRNA of the MS2 reporter gene grows and reaches saturation [52]. Another example is that in mESCs, the mRNA level decreases and tends to remain unchanged in response to the enlargement of E-P genomic distance [10].

In addition, we study the combined effect of state-switching rates and E-P interaction on gene expression. The downstream gene-expression model is a variable ON-OFF model with variable α and μ . We find that increasing α (or μ , no matter increase α_{max} (or μ_{max}) and α_{min} (or μ_{min})) enlarges $\langle \text{mRNA} \rangle$ and reduces CV (Fig 3). This is because larger α can elongate the

ON state residence time, and larger μ makes the time interval of mRNA generation become shorter, and each case leads to enlargement in $\langle \text{mRNA} \rangle$ and diminishment in CV.

E-P interaction can induce bimodal and multimodal mRNA distributions

Having clarified the qualitative effect of E-P interaction on the $\langle \text{mRNA} \rangle$ and CV of mRNA, we next analyze how E-P interaction impacts the shape of mRNA distribution, including the peak numbers and peak probabilities. We adjust the E-P interaction strength k_{EP} and E-P genomic distance d_G and calculate the theoretical distributions according to Eq [12].

Each peak is defined as the observed local maximum value to identify the peak number of mRNA distributions. Based on this, we draw the boundary lines. As a result, we find that the unimodal (U), bimodal (B), and trimodal (T) distributions can appear (Fig 4A). Besides, the heatmap in Fig 4A is the bimodal coefficient $BC = 1/(K - S^2)$, where $S = v_3/v_2^{3/2}$ is the skewness of mRNA distribution, $K = v_4/v_2^2$ is the kurtosis, and v_i is the i -th central moment, $i = 2,3,4$.

Fig 4B shows that when d_G is fixed, with the increase of k_{EP} , the distribution can appear in five modes: unimodal with the origin peak (OP), bimodal with one OP and one non-origin peak (NOP), trimodal with one OP and two NOPs, bimodal with two NOPs and unimodal with the NOP. In fact, we can see the evolutionary process of the peak numbers and peak probabilities. The OP goes under with amplifying the k_{EP} , and a NOP begins to grow and a bimodal then occurs. When the distance between the two peaks of the bimodal becomes larger, a peak emerges at the larger mRNA and the bimodal distribution turns to trimodal distribution. Subsequently, the OP vanishes and gets back to the bimodal. The distance between the two NOPs of the bimodal distribution can become smaller, and the peak corresponding to the fewer mRNA disappears when the bimodal becomes unimodal. Fig 4D perfectly demonstrates the above process, and the simulations (histograms) are in good agreement with theoretical predictions (solid lines). In addition, when we fix k_{EP} , the influence of d_G on the distribution

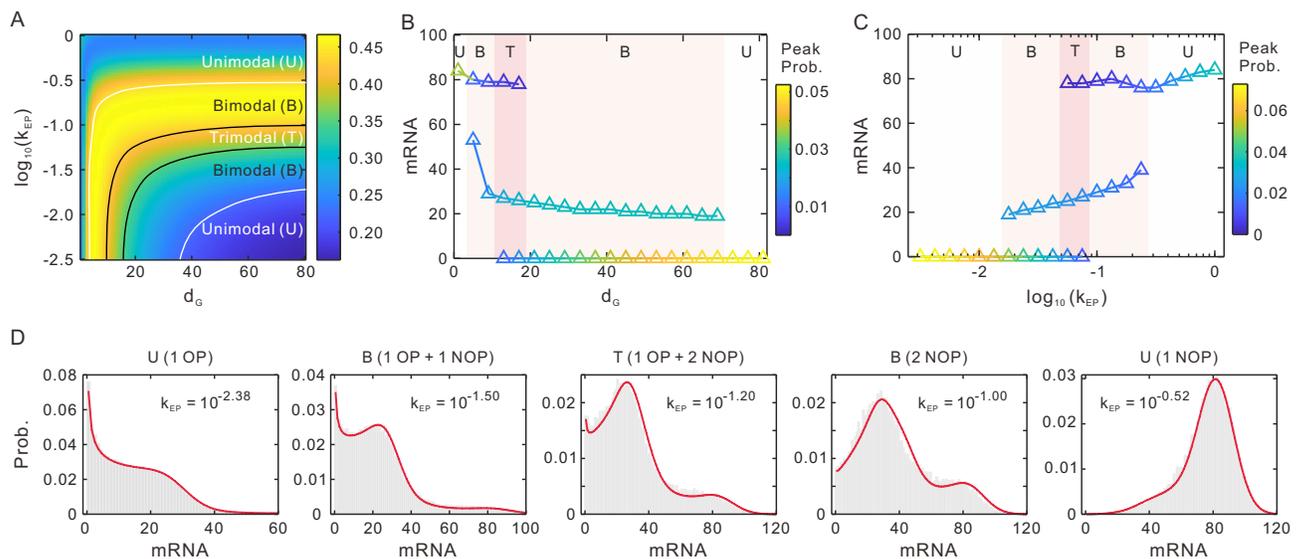


Fig 4. E-P interaction can induce multiple shapes of mRNA distribution. (A) Effects of E-P interaction on mRNA distribution. The lines stand for the boundaries of different peak numbers. The plain is divided into distinctive regions representing unimodal (U), bimodal (B) and trimodal (T), respectively. The heatmap is for the bimodal coefficient. (B) Dependence of the most probable mRNA numbers on k_{EP} , where $d_G = 70$. The color bars represent different peak numbers. The color bar represents peak probability. (C) Dependence of the most probable mRNA numbers on d_G , where $k_{EP} = 10^{-1.76}$. (D) The example of unimodal/bimodal/trimodal mRNA distribution in (B). OP (origin peak), NOP (non-origin peak). The solid lines represent theoretical results, and histograms represent numerical results. Parameter values are set as $d_G = 70$, $\alpha_{max} = 2$, $\alpha_{min} = 0.06$, $\beta = 0.10$, $\mu_{max} = 9$, $\mu_{min} = 3$, $\delta = 0.1$, $\gamma = 50$.

<https://doi.org/10.1371/journal.pcbi.1011722.g004>

modes is opposite to that of k_{EP} (Comparing Fig 4B and 4C). It should be pointed out that although the trimodal distribution may be absent under some parameter values, the pattern of distribution along k_{EP} or d_G still changes from unimodal to bimodal and then back to unimodal. During the bimodal phase, the peak at the smaller mRNA gradually disappears, resulting in the transition from bimodal to unimodal (see S2 and S3 Figs).

These results indicate that E-P interaction can produce multiple modes of mRNA distribution. More importantly, it can lead to mRNA distributions with two NOPs and even with three peaks. S1 Fig shows more bimodal and trimodal cases under different parameter values. Notably, only the downstream ON-OFF model (Eq [9]) can produce neither bimodal with two NOPs nor trimodal. In fact, the phenomenon of bimodal with two NOPs has been confirmed in biological experiments, such as the random activity of latent HIV-1 promoters [53]. In Ref [54], the researchers studied the transcriptional behavior of GREB1 changes with estrogen dose. Generally, addition of estrogen (17 β -estradiol, E2) may increase contacts between the GREB1 gene and the estrogen-receptor- α -bound enhancer [55], thus altering the E-P interaction strength. When increasing the E2 concentration (increasing the E-P interaction strength), the mRNA level grows and the mRNA distribution changes from the single peak to a bimodal distribution. A new peak emerges at a higher mRNA level and gradually increases and the peak near the origin gradually decreases, which is consistent with our model (see S2 Fig). In a word, these phenomena highlight the importance of E-P interaction in regulating gene expression, which may explain the source of multimodality. And the distinctive modes of mRNA distribution for achieving fast responses to stimulations and phenotypic switching would be essential for environmental adaptation and cellular decision-making.

Analysis of mouse embryonic stem cell data indicates that E-P interaction regulates promoter activation and transcription initiation

To check the effectiveness of our model, we used different E-P distances that forming distinctive cell lines C_k ($k = 1, \dots, 6$) and corresponding mRNA distribution measured by smRNA-FISH in mESCs [10] to infer gene expression dynamics (Materials And Methods). Assume that the mRNA distributions in different cell lines C_k are different only due to the different E-P genomic distances, and that an enhancer is inserted into the upstream or downstream of a promoter with the same genomic distance has no significant difference. The E-P genomic distances for different cell lines are NaN, 112.1710 kb, 39.4530 kb, 23.1110 kb, 17.0190 kb and 6.0600 kb respectively, where the NaN means the eGFP transcription is driven by the Sox2 promoter alone. Then, we suppose that one monomer in the coarse-grained polymer model represents 5kb, the E-P genomic distances for our model in different cell lines C_k are NaN, 22, 8, 5, 3, and 1, respectively. It should be pointed out that the d_G is the value after rounding.

First, we preliminarily attempt to use the two-state model without considering E-P regulatory to infer gene expression dynamics. We find that the changing α and μ in different cell lines get the minimize cross entropy (Table A in S1 Text). Then, we involve the long-range E-P regulatory into the model and define three variable two-state models (the variable α two-state model, the variable μ two-state model, and the variable α and μ two-state model) to simulate the transcription process (S1 Text). For each cell line C_k , we calculate the steady-state probability distributions by using Eq [12] and use Eq [6] to estimate the optimized parameter values (Materials And Methods). We find that the two-state model with variable α and μ obtains the minimum cross entropy, which indicates E-P communication prefers to regulate the promoter activation (α) and transcription initiation (μ) rates (Table B in S1 Text). The multiscale model fit converged to a good set of parameters $\Gamma = [k_{NN}, k_{EP}, \gamma, \alpha_{min}, \alpha_{max}, \beta, \mu_{min}, \mu_{max}]$ describing

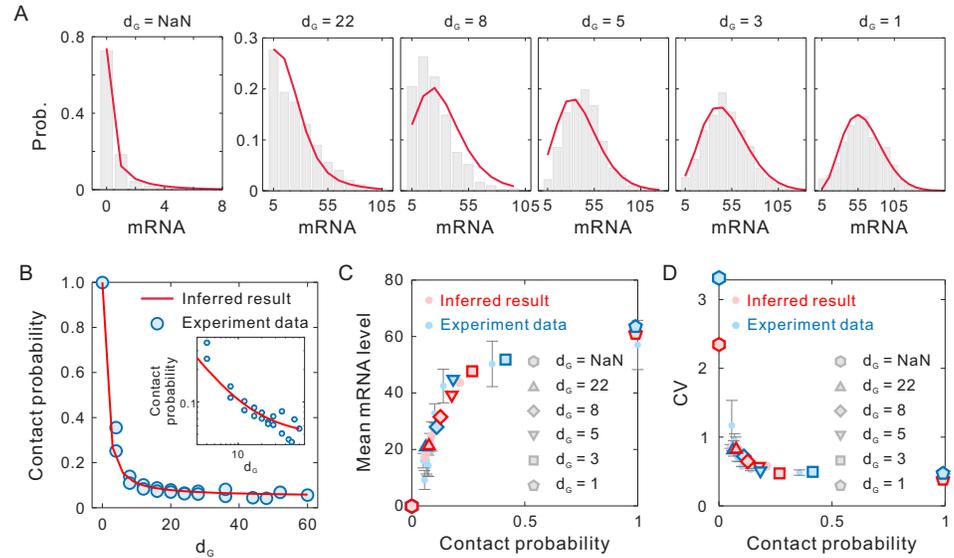


Fig 5. Results obtained by analyzing experiment data on mESCs in response to the varying E-P genomic distance. (A) Distributions of the mRNA numbers of smRNA-FISH in different cell lines. The solid line shows the best fit of our stochastic multiscale model to the experimental data shown in the histograms. $d_G = \text{NaN}$ represents promoter-only control cell line. Parameter values are listed in Table E in S1 Text. (B) The relationship between d_G and contact probability. The blue circles represent the experimental data from the ectopic Sox 2 transgene. The solid line is obtained based on the best-fit values of parameters k_{NN} and k_{EP} . The inset shows the log-log plot. (C) Mean eGFP mRNA level plotted against contact probability between the ectopic Sox2 promoter and SCR insertions. The shadow blue dots show the experimental data presented as mean values \pm standard deviation, whereas the shadow red dots show the theoretical results obtained by best-fit parameter values. The blue polygon shows the experimental data in (A), and the red polygon shows the corresponding theoretical results. (D) CV of eGFP level against contact probability. The meanings of symbols are the same as those in (C).

<https://doi.org/10.1371/journal.pcbi.1011722.g005>

chromatin and gene expression dynamics (see Table C in S1 Text). In addition, different monomer lengths do not alter the fitting results (see Table D in S1 Text).

As a result, Fig 5A shows that the 4D nucleome equation can well fit the mRNA distributions under different E-P interaction. It should be pointed out that the weight of slow chromatin dynamics ($1/(1+\omega)$) plays an important role in modifying the fitted distribution (see S5 Fig and S1 Text). Based on the best fitting parameter values and Eq [8], we can further study the relationship between E-P genomic distance and E-P contact probability (see S1 Text). Fig 5B shows that with increasing E-P genomic distance, the contact probability steeply decays and tends to be flat, which is in good agreement with the experimental results confirmed using Capture-C analysis [10]. Log-log plot represents a power-law behavior between E-P genomic distance and contact probability (Fig 5B inset), also in accordance with many experimental data in consecutive and nonconsecutive TAD borders [56], even on the genome-wide scale [57,58]. Second, Fig 5C shows the relationship between contact probability and mean gene expression level. We see that the values and the change tendency of the contact probability and average mRNA level obtained by theoretical calculations (light red circle) are basically consistent with the experimental data (light blue circle). If we focus on the six special cell lines shown in Fig 5A, we find that the inferred results (red polygon) and experimental data (blue polygon) are relatively consistent in values and change tendency. Third, for the relationship between CV of eGFP levels and contact probability (Fig 5D), we find that the theoretical result and experimental data are also relatively consistent in values and change tendency. In addition, we consider changing the upstream-downstream nonlinear Hill function in the model to a linear

form (i.e., using a linear function within a reasonable range of E-P spatial distances). By using the same fitting method, we find that the linear model fits the experimental data well (see [S6 Fig](#)). Therefore, in our model, non-linear dependence of the activation rate on E-P contacts does not seem to be a dominant cause of the nonlinear relationship between connection probability and mean expression observed in the experiment. Rather, the coupling of the fast-slow chromatin dynamics may be an important factor contributing to the nonlinearity.

Finally, we verify that the inferred results are consistent with previous observations. The inferred $k_{\text{NN}} = 0.7758$ is one order of magnitude larger than $k_{\text{EP}} = 0.0969$, indicating that the ectopic enhancer and promoter do not have a prominent specific interaction compared to the connection between successive monomers. This result is consistent with experimental studies [10]. Meanwhile, the fold change between α_{max} and α_{min} (18 folds) is larger than that between μ_{max} and μ_{min} (8 folds), indicating that E-P interaction is inclined to adjust OFF state dwell time (related to α) and further verifying that E-P interaction mainly regulates burst frequency [6,7,10,59,60]. According to the inferred results, the ON-state dwell time is in the timescale of minutes, and the OFF-state dwell time is from minutes to hours [61–63] (considering that the time unit is mRNA lifetime, expected to be around 1.5 hours [64]). The number of mRNA per burst is about tens of mRNAs, consistent with a previous experimental observation [63].

Taken together, our model results are consistent with previous experimental observations and suggest a possible essential mechanism of gene expression regulated by E-P interaction.

Discussion

Imaging studies, high-resolution chromatin conformation maps, and genome-wide occupancy profiles of architectural proteins have revealed that genome topology encoding E-P interaction information is tightly correlated with gene expression. In this paper, we have proposed a general yet tractable multiscale model, which integrates the E-P interaction information into gene expression, to shed light on the pivotal role of E-P spatial interaction in the control of gene expression profiles.

This theoretical framework characterizes the dynamic process of stochastic gene expression regulated by chromatin movement in the nucleus by using a 4D nucleome equation. First, the fast-scale chromatin motion processes (including E-P interaction) can be described by a generalized Rouse model, which is suitable for modeling the situation that the environmental effects of entanglement and crowding are negligible [65]. Second, the slow-scale gene expression process is described by a two-state model that captures essential events occurring in transcriptional processes. Third, an input-output relation was proposed to link upstream chromatin configurations to downstream gene expression. This equation unifies two stochastic processes with two different time scales, which can be taken as a good starting point for analyzing how chromatin motion affects gene expression in complex cases.

Overcoming time-scale differences and biological-process complexity to solve mRNA distribution analytically in this 4D nucleome equation is challenging. However, the analytical methodology developed here can be used in quantitatively studying distribution characteristics of gene expression involving E-P interaction. In fact, we have used the timescale separation method to obtain analytical mRNA distributions. In addition, our approach also allows us to make meaningful predictions about how upstream chromatin dynamics affect downstream gene-expression phenotypes. First, we have shown that E-P genomic distance d_G and E-P interaction strength k_{EP} , two key parameters in our model, have the opposite change trends on mRNA mean levels and CVs, which were qualitatively verified by experimental data. Second, our model demonstrated that the combination of k_{EP} and d_G is a flexible regulation strategy to explain possible implications of bimodal distributions with two NOPs and trimodal

distributions. The peaks are more sensitive to the change of d_G when k_{EP} is small, but to the change of k_{EP} when d_G is large (Figs 4, S2, and S3). S7 Fig also shows different parameters in modulating mRNA distribution. System's parameters would collectively regulate the resulting mRNA distribution, and the parameters related to E-P communication shows a consistent pattern of modulating mRNA distribution across parameter changes. These sensitivity phenomenon provide insights into complex mechanisms of biological processes and are essential for cellular decision-making and environmental adaptation. Third, our mathematical model shows well predictions on the experimental mESCs data, including the mRNA distributions and the relationship between contact probability and mean mRNA level (and CV) under different E-P interactions. The inference results indicate that the slow chromatin dynamics plays an important role in regulating the distribution of mRNAs (S5 Fig). And in particular, for the slow dynamics, the regulation of upstream to downstream is inseparable, which may cause the gene expression levels of the slow limit are not linear dependence on E-P contacts and eventually lead to the nonlinearity observed in the experiment between transcriptional levels and E-P contact probabilities. These results provide a possible mechanism for E-P interaction translated into gene-expression dynamics.

Our model has several advantages compared to previous work that studied the effect of chromatin structure on gene expression. Xiao *et al.* [31] proposed a mechanism of how E-P signals were involved in the accumulation and removal of the transcription factors favoring transcription to explain the hypersensitivity of transcription to changes in contact frequency. Zuin *et al.* [10] used the E-P encounter probability obtained from the data analysis as the regulator of the upstream to the downstream and assumed that the OFF to ON process was a multi-step process with cumulative effects. However, we note that the corresponding models did not use dynamically fluctuating E-P spatial distances to regulate downstream transcription processes. In our model, random movement of the enhancer and promoter causes fluctuations in E-P spatial distance and regulates gene transcription in real time. In fact, this direct and dynamic regulation leads to the variable rate from OFF to ON, which is practically an alternative to the multi-step process described above. We point out that in the case of fast chromatin dynamics, upstream-to-downstream regulation is performed through the expectation of the E-P spatial distance, which can be equated, in a general way, to the contact probability in Ref. [10]. However, the slow chromatin dynamics also plays an important role in regulating the distribution of mRNAs (see S1 Text). Moreover, our model has the ability to infer the gene-expression dynamics using less information that can be directly measured and leading to better fitting on mESC data to some extent (S4 Fig), especially for the E-P genomic distance in different cell lines, rather than the E-P encounter probability.

Significantly, our modeling framework can also be extended to more complex situations. For example, some experimental studies reported that gene products could affect chromatin structure [66,67]. We can incorporate this feedback into our model by modifying the gene-state-dependent drift function $V(r,s;t)$ in Eq [4]. In addition, our modeling of chromatin motion is not limited to one-pair E-P interaction. The potential extensions include the cases of multiple enhancers to one promoter [68], one enhancer to multiple promoters [6], or super-enhancers [69]. Finally, using multistate models of gene expression to extract insights from enormous experimental data and complex biological phenomena is impressive. Our two-state model is not the default option, and we may adjust the form of the downstream gene expression model to include more complex biological processes such as mRNA splices [70] and cell cycle [71]. However, we still need to balance the complexity and solvability of the model.

Finally, we point out that our theoretical model, which aims to develop a general modeling framework to study 4D gene-expression kinetics, may provide an opportunity for a dialogue

between theoretical studies and biological experiments. We envision that our modeling framework will be helpful for the biophysical analysis of broader *in vivo* cellular processes.

Supporting information

S1 Text. It consists of two parts: (1) model description and simulation; (2) fitting experimental data using model. In the first section, we supplement the details of the model and give the simulation algorithm as well as the statistics analyses of the simulation data. In the second section, we present the procedure for fitting experimental data using our model and comparing the experimental data with theoretical results.

(PDF)

S1 Fig. Bimodal and trimodal cases under different parameters. (A) The bimodal distribution (with a origin peak and a non-origin peak) of gene expression where $k_{EP} = 0.5$, $\gamma = 1$, $\alpha_{max} = 0.10$, $\alpha_{min} = 0.01$, $\beta = 0.03$, $\mu_{max} = 5$, $\mu_{min} = 2$, $\delta = 0.1$. (B) The bimodal distribution (with two origin peaks) of gene expression where $k_{EP} = 0.20$, $\gamma = 100$, $\alpha_{max} = 5$, $\alpha_{min} = 0.30$, $\beta = 0.50$, $\mu_{max} = 8$, $\mu_{min} = 2$, $\delta = 0.1$. (C) The trimodal distribution of gene expression where $k_{EP} = 0.1$, $\gamma = 300$, $\alpha_{max} = 0.5$, $\alpha_{min} = 0.05$, $\beta = 0.02$, $\mu_{max} = 10$, $\mu_{min} = 2$, $\delta = 0.1$.

(TIF)

S2 Fig. E-P interaction can induce bimodal mRNA distribution with two non-origin peaks.

(A) Effects of E-P interaction on mRNA distribution. The black lines stand for the boundaries of different peak numbers. The plain is divided into distinctive regions representing unimodal (U) and bimodal (B). The heatmap is the bimodal coefficient. (B) Dependence of the most probable mRNA numbers on k_{EP} where $d_G = 60$. The color regions represent different peak numbers. The color bar represents the peak probabilities. (C) The example of unimodal/bimodal mRNA distribution. NOP (non-origin peak). The solid lines represent theoretical results and histograms represent numerical results. Parameter values are set as $d_G = 60$, $\alpha_{max} = 5$, $\alpha_{min} = 0.3$, $\beta = 0.5$, $\mu_{max} = 8$, $\mu_{min} = 2$, $\delta = 0.1$, $\gamma = 50$.

(TIF)

S3 Fig. E-P interaction can induce bimodal mRNA distribution with one origin peak and one non-origin peak.

(A) Effects of E-P interaction on mRNA distribution. The black lines stand for the boundaries of different peak numbers. The plain is divided into distinctive regions representing unimodal (U) and bimodal (B). The heatmap is the bimodal coefficient. (B) Dependence of the most probable mRNA numbers on k_{EP} where $d_G = 60$. The color regions represent different peak numbers. The color bar represents the peak probabilities. (C) The example of unimodal/bimodal mRNA distribution. OP (origin peak), NOP (non-origin peak). The solid lines represent theoretical results and histograms represent numerical results. Parameter values are set as $d_G = 60$, $\alpha_{max} = 0.10$, $\alpha_{min} = 0.01$, $\beta = 0.03$, $\mu_{max} = 5$, $\mu_{min} = 2$, $\delta = 0.1$, $\gamma = 1$.

(TIF)

S4 Fig. Comparisons the fitting results of our model and Zuin's model to experimental data. (A) KS distances of different cell lines (the KS distance is defined in Eq [16] in the [S1 Text](#)). The blue dots and shadow line stand for the KS distance between our model and experimental data, and the dashed blue line represents the mean KS distance for all cell lines. The red shows the KS distance between Zuin's model and experimental data. (B) Distribution of mRNA numbers of smRNA-FISH in cell lines C_6 . The blue line shows the best fit of our model whereas the red line shows the best fit of Zuin's model to the experimental data shown in the histograms. (C) CDF of the distributions in (B). The $KS_6^W = 0.0284$ and $KS_6^Z = 0.12$. The

small arrows represent the number of mRNAs corresponding to the KS distance. (TIF)

S5 Fig. The contribution of slow chromatin dynamics to data fitting. (A) The weight of slow chromatin dynamics in fitting distribution ($1/(1+\omega)$). (B) KS distances of different cell lines. The blue dots and shadow line stand for the KS distance between the fast distribution and experimental data, and the red shows the KS distance between interpolation distribution and experimental data. The dashed lines represent the mean KS distance for all cell lines. (C) The distribution of Cell line 2. (D) The distribution of Cell line 6. (TIF)

S6 Fig. Comparing the fitting results of model's Hill and linear dependence to experimental data. (A) KS distances of different cell lines. The red (/green) dots and shadow line stand for the KS distance between Hill (/linear) (Eq [3]/Eq [4] in [S1 Text](#)) dependence and experimental data. The blue shows the KS distance between Zuin's model and experimental data. The dashed lines represent the mean KS distance for all cell lines. (B) The relationship between d_G and contact probability. The blue circles represent the experimental data from the ectopic Sox 2 transgene. The solid red (/green) line is obtained based on the Hill (/linear) dependence. The inset shows the log-log plot. (C) Mean eGFP mRNA level plotted against contact probability between the ectopic Sox2 promoter and SCR insertions. The blue polygon shows the experimental data, and the red (/green) polygon shows the corresponding theoretical results based on the Hill (/linear) dependence. (D) CV of eGFP level against contact probability. The meanings of symbols are the same as those in (C). (TIF)

S7 Fig. Comparisons of parameter's changes on the multimodality of mRNA distribution. (A) Effects of E-P interaction strengths k_{EP} and friction coefficient γ on the pattern of mRNA distribution. Parameters are $d_G = 60$, $\alpha_{max} = 5$, $\alpha_{min} = 0.3$, $\beta = 0.5$, $\mu_{max} = 8$, $\mu_{min} = 2$, $\delta = 0.1$. (B) Effects of k_{EP} and the ratio of μ_{max}/μ_{min} on the pattern of mRNA distribution. The $\mu_{min} = 0.1, 0.3, 0.5$. Parameters are $\alpha_{max} = 0.10$, $\alpha_{min} = 0.01$. (C) Effects of k_{EP} and the ratio of $\alpha_{max}/\alpha_{min}$ on the pattern of mRNA distribution. The $\alpha_{min} = 0.01, 0.04, 0.08$. Parameters are $\mu_{max} = 5$, $\mu_{min} = 2$. Each line divides the entire region into two parts, with U indicating a single peak area and M indicating a multimodal area. Other parameter values are set as $d_G = 60$, $\beta = 0.03$, $\delta = 0.1$, $\gamma = 1$. The solid line divides the entire region into two parts, with the part marked U indicating unimodal and the part marked M indicating multimodal (including bimodal and trimodal). (TIF)

Author Contributions

Conceptualization: Jiajun Zhang.

Investigation: Zihao Wang, Songhao Luo, Zhenquan Zhang.

Methodology: Zihao Wang, Songhao Luo, Zhenquan Zhang, Jiajun Zhang.

Project administration: Jiajun Zhang.

Supervision: Jiajun Zhang.

Visualization: Zihao Wang, Songhao Luo, Zhenquan Zhang.

Writing – original draft: Zihao Wang, Songhao Luo, Zhenquan Zhang, Tianshou Zhou, Jiajun Zhang.

Writing – review & editing: Zihao Wang, Songhao Luo, Zhenquan Zhang, Tianshou Zhou, Jiajun Zhang.

References

1. Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: How enhancer-promoter communication is sculpted in 3D. *Mol Cell*. 2019; 74(6):1110–22. <https://doi.org/10.1016/j.molcel.2019.05.032> PMID: 31226276
2. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet*. 2020; 21(2):71–87. <https://doi.org/10.1038/s41576-019-0173-8> PMID: 31605096
3. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet*. 2019; 20(8):437–55. <https://doi.org/10.1038/s41576-019-0128-0> PMID: 31086298
4. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018; 19(10):621–37. <https://doi.org/10.1038/s41580-018-0028-8> PMID: 29946135
5. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*. 2003; 12(14):1725–35. <https://doi.org/10.1093/hmg/ddg180> PMID: 12837695
6. Fukaya T, Lim B, Levine M. Enhancer control of transcriptional bursting. *Cell*. 2016; 166(2):358–68. <https://doi.org/10.1016/j.cell.2016.05.025> PMID: 27293191
7. Bartman CR, Hsu SC, Hsiung CC-S, Raj A, Blobel GA. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol Cell*. 2016; 62(2):237–47. <https://doi.org/10.1016/j.molcel.2016.03.007> PMID: 27067601
8. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat Genet*. 2018; 50(9):1296–303. <https://doi.org/10.1038/s41588-018-0175-z> PMID: 30038397
9. Rodriguez J, Ren G, Day CR, Zhao K, Chow CC, Larson DR. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell*. 2019; 176(1–2):213–26. <https://doi.org/10.1016/j.cell.2018.11.026> PMID: 30554876
10. Zuin J, Roth G, Zhan Y, Cramard J, Redolfi J, Piskadlo E, et al. Nonlinear control of transcription through enhancer-promoter interactions. *Nature*. 2022; 604:571–7. <https://doi.org/10.1038/s41586-022-04570-y> PMID: 35418676
11. Li J, Hsu A, Hua Y, Wang G, Cheng L, Ochiai H, et al. Single-gene imaging links genome topology, promoter–enhancer communication and transcription control. *Nat Struct Mol Biol*. 2020; 27(11):1032–40. <https://doi.org/10.1038/s41594-020-0493-6> PMID: 32958948
12. Su J-H, Zheng P, Kinrot SS, Bintu B, Zhuang X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell*. 2020; 182(6):1641–59. <https://doi.org/10.1016/j.cell.2020.07.032> PMID: 32822575
13. Gizzi AMC, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol Cell*. 2019; 74(1):212–22. <https://doi.org/10.1016/j.molcel.2019.01.011> PMID: 30795893
14. Bohrer CH, Larson DR. The stochastic genome and its role in gene expression. *Cold Spring Harb Perspect Biol*. 2021; 13:a040386. <https://doi.org/10.1101/cshperspect.a040386> PMID: 33903158
15. Hübner MR, Eckersley-Maslin MA, Spector DL. Chromatin organization and transcriptional regulation. *Curr Opin Genet Dev*. 2013; 23(2):89–95. <https://doi.org/10.1016/j.gde.2012.11.006> PMID: 23270812
16. Oudelaar AM, Higgs DR. The relationship between genome structure and function. *Nat Rev Genet*. 2021; 22(3):154–68. <https://doi.org/10.1038/s41576-020-00303-x> PMID: 33235358
17. Yokoshi M, Segawa K, Fukaya T. Visualizing the role of boundary elements in enhancer-promoter communication. *Mol Cell*. 2020; 78(2):224–35. <https://doi.org/10.1016/j.molcel.2020.02.007> PMID: 32109364
18. Senecal A, Munsy B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep*. 2014; 8(1):75–83. <https://doi.org/10.1016/j.celrep.2014.05.053> PMID: 24981864
19. Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA, et al. Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol Cell*. 2019; 73(3):519–32. <https://doi.org/10.1016/j.molcel.2018.11.004> PMID: 30554946

20. Stavreva DA, Coulon A, Baek S, Sung MH, John S, Stixova L, et al. Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. *Genome Res.* 2015; 25(6):845–57. <https://doi.org/10.1101/gr.184168.114> PMID: 25677181
21. Hakim O, Sung MH, Voss TC, Splinter E, John S, Sabo PJ, et al. Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* 2011; 21(5):697–706. <https://doi.org/10.1101/gr.111153.110> PMID: 21471403
22. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci USA.* 2008; 105:17256–61. <https://doi.org/10.1073/pnas.0803850105> PMID: 18988743
23. Kumar N, Platini T, Kulkarni RV. Exact distributions for stochastic gene expression models with bursting and feedback. *Phys Rev Lett.* 2014; 113:268105. <https://doi.org/10.1103/PhysRevLett.113.268105> PMID: 25615392
24. Wang Z, Zhang Z, Zhou T. Exact distributions for stochastic models of gene expression with arbitrary regulation. *Sci China Math.* 2020; 63(3):485–500. <https://doi.org/10.1007/s11425-019-1622-8>
25. Zhang J, Zhou T. Markovian approaches to modeling intracellular reaction processes with molecular memory. *Proc Natl Acad Sci USA.* 2019; 116(47):23542–50. <https://doi.org/10.1073/pnas.1913926116> PMID: 31685609
26. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 2006; 4(10):e309. <https://doi.org/10.1371/journal.pbio.0040309> PMID: 17048983
27. Peccoud J, Ycart B. Markovian modeling of gene-product synthesis. *Theor Popul Biol.* 1995; 48:222–34. <https://doi.org/10.1006/tpbi.1995.1027>
28. Zhang J, Zhou T. Promoter-mediated transcriptional dynamics. *Biophys J.* 2014; 106(2):479–88. <https://doi.org/10.1016/j.bpj.2013.12.011> PMID: 24461023
29. Harper CV, Finkenstädt B, Woodcock DJ, Friedrichsen S, Semprini S, Ashall L, et al. Dynamic analysis of stochastic transcription cycles. *PLoS Biol.* 2011; 9(4):e1000607. <https://doi.org/10.1371/journal.pbio.1000607> PMID: 21532732
30. Zhou T, Zhang J. Analytical results for a multi-state gene model. *SIAM J Appl Math.* 2012; 72:789–818. <http://dx.doi.org/10.1137/110852887>.
31. Xiao JY, Hafner A, Boettiger AN. How subtle changes in 3D structure can create large changes in transcription. *eLife.* 2021; 10:e64320. <https://doi.org/10.7554/eLife.64320> PMID: 34240703
32. Sood V, Misteli T. The stochastic nature of genome organization and function. *Curr Opin Genet Dev.* 2022; 72:45–52. <https://doi.org/10.1016/j.gde.2021.10.004> PMID: 34808408
33. Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. *Cell.* 2016; 164(6):1110–21. <https://doi.org/10.1016/j.cell.2016.02.007> PMID: 26967279
34. Elimelech A, Birnbaum RY. From 3D organization of the genome to gene expression. *Curr Opin Syst Biol.* 2020; 22:22–31. <https://doi.org/10.1016/j.coisb.2020.07.006>
35. Brouwer I, Lenstra TL. Visualizing transcription: Key to understanding gene expression dynamics. *Curr Opin Chem Biol.* 2019; 51:122–9. <https://doi.org/10.1016/j.cbpa.2019.05.031> PMID: 31284216
36. Lammers NC, Kim YJ, Zhao J, Garcia HG. A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Curr Opin Cell Biol.* 2020; 67:147–57. <https://doi.org/10.1016/j.ceb.2020.08.001> PMID: 33242838
37. Johnstone CP, Wang NB, Sevier SA, Galloway KE. Understanding and engineering chromatin as a dynamical system across length and timescales. *Cell Syst.* 2020; 11(5):424–48. <https://doi.org/10.1016/j.cels.2020.09.011> PMID: 33212016
38. Heist T, Fukaya T, Levine M. Large distances separate coregulated genes in living *Drosophila* embryos. *Proc Natl Acad Sci USA.* 2019; 116(30):15062–7. <https://doi.org/10.1073/pnas.1908962116> PMID: 31285341
39. Lim B, Heist T, Levine M, Fukaya T. Visualization of transvection in living *Drosophila* embryos. *Mol Cell.* 2018; 70(2):287–96. <https://doi.org/10.1016/j.molcel.2018.02.029> PMID: 29606591
40. Lammers NC, Galstyan V, Reimer A, Medin SA, Wiggins CH, Garcia HG. Multimodal transcriptional control of pattern formation in embryonic development. *Proc Natl Acad Sci USA.* 2020; 117(2):836–47. <https://doi.org/10.1073/pnas.1912500117> PMID: 31882445
41. Dekker J, Belmont A, Guttman M, Leshyk V, Lis J, Lomvardas S, et al. The 4D nucleome project. *Nature.* 2017; 549(7671):219–26. <https://doi.org/10.1038/nature23884> PMID: 28905911
42. Marti-Renom MA, Almouzni G, Bickmore WA, Bystricky K, Cavalli G, Fraser P, et al. Challenges and guidelines toward 4D nucleome data and model standards. *Nat Genet.* 2018; 50(10):1352–8. <https://doi.org/10.1038/s41588-018-0236-3> PMID: 30262815
43. Doi M, Edwards SF, Edwards S. *The theory of polymer dynamics*: Oxford University Press; 1988.

44. Barbieri M, Chotalia M, Fraser J, Lavitas L, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA*. 2012; 109(40):16173–8. <https://doi.org/10.1073/pnas.1204799109> PMID: 22988072
45. Shukron O, Holcman D. Transient chromatin properties revealed by polymer models and stochastic simulations constructed from Chromosomal Capture data. *PLoS Comput Biol*. 2017; 13(4):e1005469. <https://doi.org/10.1371/journal.pcbi.1005469> PMID: 28369076
46. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*. 2014; 157(4):950–63. <https://doi.org/10.1016/j.cell.2014.03.025> PMID: 24813616
47. Lu L, Liu X, Huang W-K, Giusti-Rodríguez P, Cui J, Zhang S, et al. Robust Hi-C maps of enhancer-promoter interactions reveal the function of non-coding genome in neural development and diseases. *Mol Cell*. 2020; 79(3):521–34. e15. <https://doi.org/10.1016/j.molcel.2020.06.007> PMID: 32592681
48. Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, et al. The Hill equation: A review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol*. 2008; 22(6):633–48. <https://doi.org/10.1111/j.1472-8206.2008.00633.x> PMID: 19049668
49. Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife*. 2019; 8:e41769. <https://doi.org/10.7554/eLife.41769> PMID: 31124784
50. Brueckner DB, Chen H, Barinov L, Zoller B, Gregor T. Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. 2023. <https://doi.org/10.1101/2023.01.18.524527>.
51. Vu TN, Wills QF, Kalari KR, Niu N, Wang L, Rantalainen M, et al. Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics*. 2016; 32(14):2128–35. <https://doi.org/10.1093/bioinformatics/btw202> PMID: 27153638
52. Larson DR, Fritsch C, Sun L, Meng X, Lawrence DS, Singer RH. Direct observation of frequency modulated transcription in single cells using light activation. *eLife*. 2013; 2:e00750. <https://doi.org/10.7554/eLife.00750> PMID: 24069527
53. Tantale K, Garcia-Oliver E, Robert M-C, L'hostis A, Yang Y, Tsanov N, et al. Stochastic pausing at latent HIV-1 promoters generates transcriptional bursting. *Nat Commun*. 2021; 12(1):1–20. <https://doi.org/10.1038/s41467-021-24462-5>.
54. Fritsch C, Baumgärtner S, Kuban M, Steinshorn D, Reid G, Legewie S. Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. *Mol Syst Biol*. 2018; 14(2):e7678. <https://doi.org/10.15252/msb.20177678> PMID: 29476006
55. Hevener AL, Ribas V, Moore TM, Zhou Z. ERα in the control of mitochondrial function and metabolic health. *Trends Mol Med*. 2021; 27(1):31–46. <https://doi.org/10.1016/j.molmed.2020.09.006>.
56. Cattoni DI, Gizzi AMC, Georgieva M, Di Stefano M, Valeri A, Chamousset D, et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat Commun*. 2017; 8(1):1753. <https://doi.org/10.1038/s41467-017-01962-x> PMID: 29170434
57. Ing-Simmons E, Vaid R, Bing XY, Levine M, Mannervik M, Vaquerizas JM. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat Genet*. 2021; 53(4):487. <https://doi.org/10.1038/s41588-021-00799-x> PMID: 33795866
58. Bonev B, Cohen NM, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D genome rewiring during mouse neural development. *Cell*. 2017; 171(3):557–72. <https://doi.org/10.1016/j.cell.2017.09.043> PMID: 29053968
59. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, et al. Genomic encoding of transcriptional burst kinetics. *Nature*. 2019; 565(7738):251–4. <https://doi.org/10.1038/s41586-018-0836-1> PMID: 30602787
60. Ochiai H, Hayashi T, Umeda M, Yoshimura M, Harada A, Shimizu Y, et al. Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Sci Adv*. 2020; 6(25):eaaz6699. <https://doi.org/10.1126/sciadv.aaz6699> PMID: 32596448
61. Molina N, Suter DM, Cannavo R, Zoller B, Gotic I, Naef F. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc Natl Acad Sci USA*. 2013; 110(51):20563–8. <https://doi.org/10.1073/pnas.1312310110> PMID: 24297917
62. Halpern KB, Tanami S, Landen S, Chapal M, Szlak L, Hutzler A, et al. Bursty gene expression in the intact mammalian liver. *Mol Cell*. 2015; 58(1):147–56. <https://doi.org/10.1016/j.molcel.2015.01.027> PMID: 25728770

63. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011; 332(6028):472–4. <https://doi.org/10.1126/science.1198817> PMID: 21415320
64. Horvathova I, Voigt F, Kotrys AV, Zhan Y, Artus-Revel CG, Eglinger J, et al. The dynamics of mRNA turnover revealed by single-molecule imaging in single cells. *Mol Cell*. 2017; 68(3):615–25. <https://doi.org/10.1016/j.molcel.2017.09.030> PMID: 29056324
65. Zhang Y, Dudko OK. First-passage processes in the genome. *Annu Rev Biophys*. 2016; 45(1):117–34. <https://doi.org/10.1146/annurev-biophys-062215-010925> PMID: 27391924
66. Heinz S, Texari L, Hayes MG, Urbanowski M, Chang MW, Givarkes N, et al. Transcription elongation can affect genome 3D structure. *Cell*. 2018; 174(6):1522–36. <https://doi.org/10.1016/j.cell.2018.07.047> PMID: 30146161
67. Busslinger GA, Stocsits RR, Van Der Lelij P, Axelsson E, Tedeschi A, Galjart N, et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*. 2017; 544(7651):503–7. <https://doi.org/10.1038/nature22063>
68. He B, Chen C, Teng L, Tan K. Global view of enhancer–promoter interactome in human cells. *Proc Natl Acad Sci USA*. 2014; 111(21):E2191–E9. <https://doi.org/10.1073/pnas.1320308111> PMID: 24821768
69. Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2015; 47(1):8–12. <https://doi.org/10.1038/ng.3167> PMID: 25547603
70. Kan Z, States D, Gish W. Selecting for functional alternative splices in ESTs. *Genome Res*. 2002; 12(12):1837–45. <https://doi.org/10.1101/gr.764102> PMID: 12466287
71. Johnson DG, Walker CL. Cyclins and cell cycle checkpoints. *Annu Rev Pharmacol Toxicol*. 1999; 39:295–312. <https://doi.org/10.1146/annurev.pharmtox.39.1.295> PMID: 10331086