Extrinsic Noise or Intrinsic Coupling: Dissecting Correlated Fluctuations in Gene Transcription

Shihe Zhang,[1,2] Huihan Bao,[1,2] and Heng Xu[1,2]

[1] *School of Physics and Astronomy, Shanghai Jiao Tong University, 200240 Shanghai, China*
[2] *Institute of Natural Sciences, Shanghai Jiao Tong University, 200240 Shanghai, China*

(Dated: September 9, 2025)

Understanding stochastic gene transcription requires distinguishing between intrinsic molecular randomness and extrinsic environmental variability. Traditionally, these components are identified as the independent and correlated fluctuations between identical gene copies within the same cell. However, intrinsic gene-gene interactions can introduce additional correlations, challenging this standard approach. Here, we develop a new noise decomposition framework based on a theoretical model of stochastic transcription for an intrinsically coupled gene pair under fluctuating environments. By analytically deriving correlated fluctuations of nascent RNA, we disentangle intrinsic coupling from extrinsic noise based on their distinct relationships with the mean and variance of gene activation probabilities across varying environments. Applying this framework to single-cell transcription data from *Drosophila* embryos, we uncover previously unidentified couplings between sister alleles and between alternative promoters of an endogenous gene. Our findings offer a versatile approach for deciphering intrinsic gene-gene interactions in stochastic transcription within fluctuating environments.

Introduction—Gene transcription, the production of RNA in biological cells, is a well-known stochastic process driven by both the intrinsic randomness of biomolecular reactions and extrinsic environmental variability [1–4]. Understanding the mechanisms of stochastic transcription kinetics requires accurately distinguishing these two components within the total fluctuation (or noise). Formally, this is done by decomposing the variance of cellular RNA number m using the law of total variance [4–10]:

$$\sigma_m^2 = \underbrace{\left\langle \sigma_{m|\mathrm{E}}^2 \right\rangle}_{\sigma_{\mathrm{int}}^2} + \underbrace{\sigma_{\langle m|\mathrm{E}\rangle}^2}_{\sigma_{\mathrm{ext}}^2}. \tag{1}$$

Here, the intrinsic variance σ_{int}^2 measures the average fluctuation in m under fixed environmental conditions (E), whereas the extrinsic variance σ_{ext}^2 captures the variability in the conditional average of m due to fluctuating environments.

To experimentally achieve such partitioning, the standard approach is the dual reporter assay, which monitors two distinguishable copies of the same gene embedded in the same cell (via, e.g., fluorescent RNA reporter) [1,5,6,9,11–13]. Assuming each gene copy transcribes independently, the covariance between their RNA numbers reflects extrinsic variance from the shared environment, while the remaining uncorrelated variance is intrinsic.

However, while this simple approach has been the gold standard for two decades, the assumption of gene copy independence appears overly idealistic, as genes within the same cell inevitably interact. For instance, gene copies may compete for transcription factors [14,15], and their products may exert feedback regulations [16, 17]. Moreover, homologous alleles of the same gene located on different chromosomes within the same cell can cross-regulate each other via direct trans-homolog enhancer-promoter interactions, a phenomenon known as transvection [18,19]. Similar enhancer-sharing mechanisms also occur between genes on the same chromosome or between promoters of the same gene [20,21]. In principle, these interactions can introduce couplings between genes, even in the absence of extrinsic fluctuations, thereby violating the independence assumption.

In many biological systems, particularly in eukaryotic cells where genes typically exist in homologous pairs, these effects may be nonnegligible [13,21]. While some indirect effects, such as feedback regulations, can be separated using short-lived reporters (e.g., nascent RNA reporters) or time-correlation analysis [13,22], most direct couplings at the gene activation level are inseparable from the system. Attempting to eliminate or ignore these intrinsic couplings is often both technically cumbersome and scientifically counterproductive, particularly when studying endogenous genes in realistic (rather than oversimplified synthetic) systems. Therefore, a new framework is needed to disentangle correlated transcriptional fluctuations arising from intrinsic gene-gene coupling and extrinsic noise. The goal of this Letter is to develop such a framework.

The model—We start with a simple telegraph model for nascent transcription of a pair of intrinsically coupled identical genes in a eukaryotic cell [23–26] [Fig. 1(a), see Supplemental Material [27]]. In this model, each gene copy ($i = a, b$) fluctuates between an active (ON) and an inactive (OFF) transcription state, with baseline transition rates k_{ON} and k_{OFF}. If one gene is ON, it modulates the activation rate of its sister copy by a factor γ, i.e., $k_{\mathrm{ON}} \to \gamma k_{\mathrm{ON}}$. Here, $\gamma = 1$ indicates independent gene copies, while $\gamma > 1$ and $0 < \gamma < 1$ denote positive and negative couplings, respectively. When a gene is active, it initiates new transcription events as a Poisson process

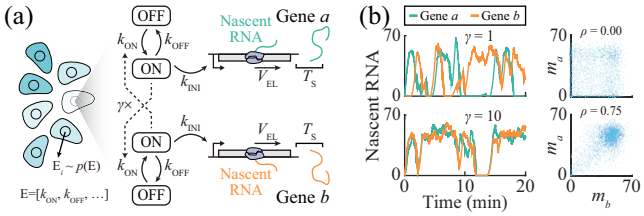

FIG. 1. A stochastic model of coupled gene transcription. (a) Model schematic. (b) Simulated results for the model with and without intrinsic coupling under a fixed environmental condition. Left: Time traces of gene transcription. Right: Scatter plot of nascent RNA levels for both gene copies. $\gamma \neq 1$ leads to correlated fluctuation. The kinetic parameters are set as: $k_{\rm ON} = 0.5$ min^{-1}, $k_{\rm OFF} = 0.5$ min^{-1}, $k_{\rm INI} = 50$ min^{-1}, $T_{\rm RES} = 2$ min, $g(\tau) = -\tau$, and $\gamma = 1$ and 10 for independent and coupled gene pairs, respectively.

with rate $k_{\rm INI}$, followed by a constant-speed ($V_{\rm EL}$) RNA synthesis (of length L) and a post-synthesis residence duration $T_{\rm S}$ before the nascent RNA is released.

To account for extrinsic variability, we assume that all kinetic parameters $\mathrm{E} = [k_{\rm ON}, k_{\rm OFF}, \cdots]$ may vary across cells and over time following a probability distribution $p(\mathrm{E})$ [Fig. 1(a), see Supplemental Material [27]]. Based on prior studies, many of these extrinsic variations occur on longer time scales than nascent RNA dynamics, justifying an adiabatic approximation, with quasi-static E over the duration of nascent RNA production [49,50]. Among the parameters in E, $k_{\rm ON}$ and $k_{\rm OFF}$ are likely the primary targets of extrinsic fluctuations [4,50]. Simulations of this model demonstrate that two gene copies with $\gamma \neq 1$ can generate correlated outputs even in the absence of extrinsic fluctuation [Fig. 1(b)].

Solving coupled intrinsic fluctuation—At a given time t, the state of the system is characterized by four random variables: the gene states n_i (0 and 1 for OFF and ON states, respectively) and the amount of nascent RNA m_i ($m_i \geq 0$) at each gene copy [23, 51]. Notably, m_i represents the cumulative signal (in units of a complete RNA molecule) from all nascent RNA molecules initiated within the residence time window $T_{\rm RES} = L/V_{\rm EL} + T_{\rm S}$ prior to t, i.e., $m_i(t) = \sum_{t-T_{\rm RES} \leq t_j \leq t} g(t_j - t)$ [23] (Supplemental Material [27]). Here, $g(\tau)$ represents the contribution of a nascent RNA molecule initiated at time τ (relative to t, $-T_{\rm RES} \leq \tau \leq 0$) to the total nascent RNA signal at t. Note that the system can be nondimensionalized by rescaling τ and all kinetic rates with respect to $T_{\rm RES}$. Without loss of generality, we set $T_{\rm RES} = 1$ for this study. The probability distribution of the system state $\mathbf{P}(m_a, m_b) = \begin{bmatrix} P(n_a{=}0,n_b{=}0,m_a,m_b) \\ P(n_a{=}1,n_b{=}0,m_a,m_b) \\ P(n_a{=}0,n_b{=}1,m_a,m_b) \\ P(n_a{=}1,n_b{=}1,m_a,m_b) \end{bmatrix}$ and its statistical moments can be obtained analytically (Appendix A) [23,52,53].

Specifically, under a fixed environmental condition, the steady-state marginal distribution of n_a and n_b is:

$$\tilde{\mathbf{P}}_0 = \frac{1}{1 + 2\left(\frac{k_{\rm ON}}{k_{\rm OFF}}\right) + \gamma\left(\frac{k_{\rm ON}}{k_{\rm OFF}}\right)^2} \begin{bmatrix} 1 \\ k_{\rm ON}/k_{\rm OFF} \\ k_{\rm ON}/k_{\rm OFF} \\ \gamma(k_{\rm ON}/k_{\rm OFF})^2 \end{bmatrix}, \quad (2)$$

which yields the covariance between n_a and n_b [Fig. 2(a)]:

$$C_{n_a,n_b|\mathrm{E}} = (\gamma - 1)\left(\frac{\left(\frac{k_{\rm ON}}{k_{\rm OFF}}\right)}{1 + 2\left(\frac{k_{\rm ON}}{k_{\rm OFF}}\right) + \gamma\left(\frac{k_{\rm ON}}{k_{\rm OFF}}\right)^2}\right)^2. \quad (3)$$

As expected, $\gamma > 1$ and $0 < \gamma < 1$ lead to positive and negative covariance, respectively. For each $\gamma \neq 1$, Eq. (3) reveals an extremum of $C_{n_a,n_b|\mathrm{E}} = \left(\sqrt{\gamma} - 1\right)/4\left(\sqrt{\gamma} + 1\right)$, corresponding to equal probabilities of the ON and OFF states for a single gene copy, i.e., $\langle n|\mathrm{E}\rangle = 0.5$. Equivalently, $C_{n_a,n_b|\mathrm{E}}$ can be expressed as a function of $\langle n|\mathrm{E}\rangle$ and γ [Fig. 2(b)], i.e., $C_{n_a,n_b|\mathrm{E}} = f(\langle n|\mathrm{E}\rangle, \gamma)$, with (Supplemental Material [27]):

$$f(n,\gamma) = (\gamma - 1)\left[\frac{(1-n)\left(2n - 1 + \sqrt{4(\gamma-1)n(1-n)+1}\right)}{2\gamma(1-n) + 2n - 1 + \sqrt{4(\gamma-1)n(1-n)+1}}\right]^2. \quad (4)$$

Because $\langle n|\mathrm{E}\rangle$ is experimentally more accessible than $k_{\rm ON}$ and $k_{\rm OFF}$, this reformulation helps connect our theory to experimental data in subsequent derivations. In addition to covariance, the total variance of n is simply $\sigma^2_{n|\mathrm{E}} = \langle n|\mathrm{E}\rangle - \langle n|\mathrm{E}\rangle^2$, independent of γ.

These gene-state quantities are closely related to experimentally measurable nascent RNA statistics (Appendix A). For example, the mean nascent RNA signal is proportional to the ON-state probability:

$$\langle m|\mathrm{E}\rangle = k_{\rm INI}\bar{g}\langle n|\mathrm{E}\rangle, \quad (5)$$

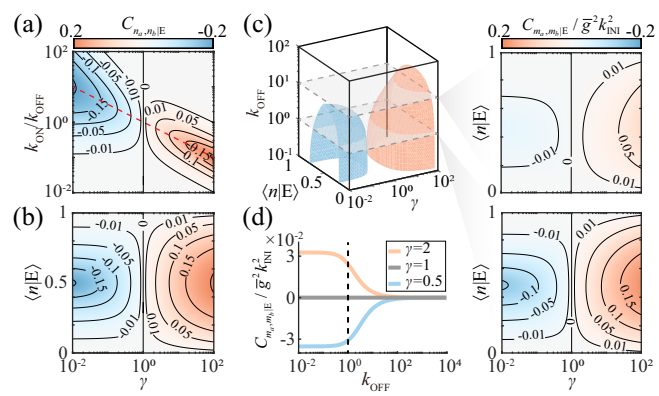

FIG. 2. Intrinsic covariance in coupled gene transcription. (a–b) Covariance of gene state as a function of γ and $k_{\rm ON}/k_{\rm OFF}$ (a) or γ and $\langle n|\mathrm{E}\rangle$ (b) under a fixed environmental condition. (c) Covariance of nascent RNA as a function of γ, $\langle n|\mathrm{E}\rangle$, and $k_{\rm OFF}$. Red and blue surfaces, $C_{m_a,m_b|\mathrm{E}} = \pm 0.04$, respectively. (d) Decay of nascent-RNA covariance with $k_{\rm OFF}$ for representative γ values ($\langle n|\mathrm{E}\rangle = 0.5$).

where $\bar{g} = \int_{-1}^{0} g(\tau)d\tau$. Furthermore, both the covariance and total variance at nascent RNA level follow trends similar to those at the gene-state level [Fig. 2(c) and S1(a)]. Specifically, in the limit of slow gene-state transitions ($k_{\rm ON}$ & $k_{\rm OFF} \to 0$), the covariances and total variance at these two levels are simply related as:

$$C_{m_a,m_b|\rm E} = \bar{g}^2 k_{\rm INI}^2 C_{n_a,n_b|\rm E}, \tag{6}$$

$$\sigma^2_{m|\rm E} = \bar{g}^2 k_{\rm INI}^2 \sigma^2_{n|\rm E} + \overline{g^2} k_{\rm INI} \langle n|\rm E\rangle, \tag{7}$$

where $\overline{g^2} = \int_{-1}^{0} g^2(\tau)d\tau$. Notably, these approximations hold well up to $k_{\rm ON}$ & $k_{\rm OFF} \sim 1$ [Fig. 2(d) and S1(b), see Appendix A]. Therefore, for experiments measuring short-lived nascent RNA signals, whose lifetimes satisfy $k_{\rm ON}$ & $k_{\rm OFF} \lesssim 1$ in many biological systems [54], Eqs. (6) and (7) remain valid.

Partitioning correlated noise in fluctuating environments—Under changing environments, the correlated fluctuation for an intrinsically coupled gene pair is a combination of intrinsic and extrinsic noise components. To establish a general framework for noise decomposition in this context, we refine Eq. (1) to explicitly separate the independent ($\sigma^2_{\rm int,ind}$) and coupled ($\sigma^2_{\rm ind,cp}$) components of intrinsic fluctuations:

$$\sigma^2_m = \underbrace{\left\langle \sigma^2_{m|\rm E} - C_{m_a,m_b|\rm E} \right\rangle}_{\sigma^2_{\rm int,ind}} + \underbrace{\left\langle C_{m_a,m_b|\rm E} \right\rangle}_{\sigma^2_{\rm int,cp}} + \underbrace{\sigma^2_{\langle m|\rm E\rangle}}_{\sigma^2_{\rm ext}}, \tag{8}$$

where $\sigma^2_{\rm int,ind}$ remains the uncorrelated component of the total fluctuation, while $\sigma^2_{\rm ind,cp}$ is defined as the average covariance between the nascent RNA levels of the two gene copies under fixed environmental conditions. Notably, both $\sigma^2_{\rm ind,cp}$ and $\sigma^2_{\rm ext}$ contribute to experimentally measured total covariance, i.e.:

$$C_{m_a,m_b} = \left\langle C_{m_a,m_b|\rm E} \right\rangle + \sigma^2_{\langle m|\rm E\rangle}. \tag{9}$$

Hypothetically, $\sigma^2_{\rm ind,cp}$ can be determined from Eq. (9) if the average expression level under every environmental condition ($\langle m|\rm E\rangle$) can be precisely measured.

Practically, however, determining the absolute values of $\langle m|\rm E\rangle$ and $\sigma^2_{\langle m|\rm E\rangle}$ in fluctuating environments can be challenging. To distinguish $\sigma^2_{\rm ind,cp}$ and $\sigma^2_{\rm ext}$ in a general sense, we notice that, if environmental fluctuations primarily affect gene-state transitions, then,

$$\sigma^2_{\langle m|\rm E\rangle} \approx \bar{g}^2 k_{\rm INI}^2 \sigma^2_{\langle n|\rm E\rangle}, \tag{10}$$

i.e., the extrinsic fluctuation is simply proportional to the variance of ON-state probabilities over fluctuating environments ($\sigma^2_{\langle n|\rm E\rangle}$), independent of their mean ($\langle n\rangle$). In contrast, for $k_{\rm ON}$ & $k_{\rm OFF} \lesssim 1$, the coupled intrinsic fluctuation is a function of both $\langle n\rangle$ and $\sigma^2_{\langle n|\rm E\rangle}$ (Supplemental Material [27]):

$$\left\langle C_{m_a,m_b|\rm E} \right\rangle \approx \bar{g}^2 k_{\rm INI}^2 \left[f(\langle n\rangle, \gamma) + \frac{1}{2}\frac{\partial^2 f}{\partial n^2}\bigg|_{\langle n\rangle} \sigma^2_{\langle n|\rm E\rangle} \right]. \tag{11}$$

Therefore, the dependence of total covariance on $\langle n\rangle$ and $\sigma^2_{\langle n|\rm E\rangle}$ varies with the relative contributions of $\sigma^2_{\rm ind,cp}$ and $\sigma^2_{\rm ext}$ [Fig. 3(a)]. Specifically, for an independent gene pair ($\gamma = 1$), C_{m_a,m_b} is solely a function of $\sigma^2_{\langle n|\rm E\rangle}$. In contrast, for a fully coupled or anti-coupled gene pair ($\gamma \to \infty$ or $\gamma \to 0$), C_{m_a,m_b} depends exclusively on $\langle n\rangle$ (Supplemental Material [27]). In all other cases, C_{m_a,m_b} depends on both $\langle n\rangle$ and $\sigma^2_{\langle n|\rm E\rangle}$, with its gradient offering insights into γ. In particular, with given $\langle n\rangle$, the partial slope of C_{m_a,m_b} with respect to $\sigma^2_{\langle m|\rm E\rangle}$ is simply

$$\frac{\partial C_{m_a,m_b}}{\partial \sigma^2_{\langle m|\rm E\rangle}}\bigg|_{\langle n\rangle} \approx \frac{1}{2}\frac{\partial^2 f}{\partial n^2}\bigg|_{\langle n\rangle} + 1. \tag{12}$$

To apply this equation to experimental data, we note that, based on Eqs. (1), (5), and (7), $\langle n\rangle$ can be estimated from the total noise of nascent RNA as

$$\eta^2_m \equiv \frac{\sigma^2_m}{\langle m\rangle^2} = \frac{1}{\langle n\rangle}\left(1 + \frac{\overline{g^2}}{\bar{g}^2 k_{\rm INI}}\right) - 1 \approx \frac{1}{\langle n\rangle} - 1. \tag{13}$$

Here, η^2_m, the squared coefficient of variation of nascent RNA, is experimentally measurable, and the approximation holds for $k_{\rm INI} \gg 1$, which is satisfied in most biological systems [13, 25, 51, 54–56]. Unlike $\langle n\rangle$, the absolute value of $\sigma^2_{\langle m|\rm E\rangle}$ (or $\sigma^2_{\langle n|\rm E\rangle}$), which may be influenced by multiple unknown or inaccessible external factors, is difficult to determine. However, if one of the $\langle n\rangle$-regulating factors, denoted as x, is accessible, we may perturb it from the original value x_0 to specific levels to collect datasets with distinct mean values ($\langle m|x\rangle$) but comparable $\sigma^2_{\langle m|\rm E\rangle}$ (Appendix B). By selecting a suitable set of perturbation levels $\{x_i\}$, we can construct a combined dataset, whose average nascent RNA level $\langle\langle m|x_i\rangle\rangle_i$ matches that of the unperturbed dataset at x_0 [Fig. 3(b)]. In contrast, $\sigma^2_{\langle m|\rm E\rangle}$ of the combined dataset increases with respect to the unperturbed one as

$$\Delta\sigma^2_{\langle m|\rm E\rangle} \approx \sigma^2_{\langle m|x_i\rangle}. \tag{14}$$

I.e., although the absolute value of $\sigma^2_{\langle m|\rm E\rangle}$ is unknown, we can quantitatively increase it while keeping $\langle m\rangle$ (and $\langle n\rangle$) fixed.

Building on these results, we propose a quantitative framework for estimating γ, $\sigma^2_{\rm ind,cp}$, and $\sigma^2_{\rm ext}$ from experimental data through the following steps:

1. Determining $\langle n\rangle$: For any given experimental dataset, we can compute η^2_m to estimate $\langle n\rangle$.

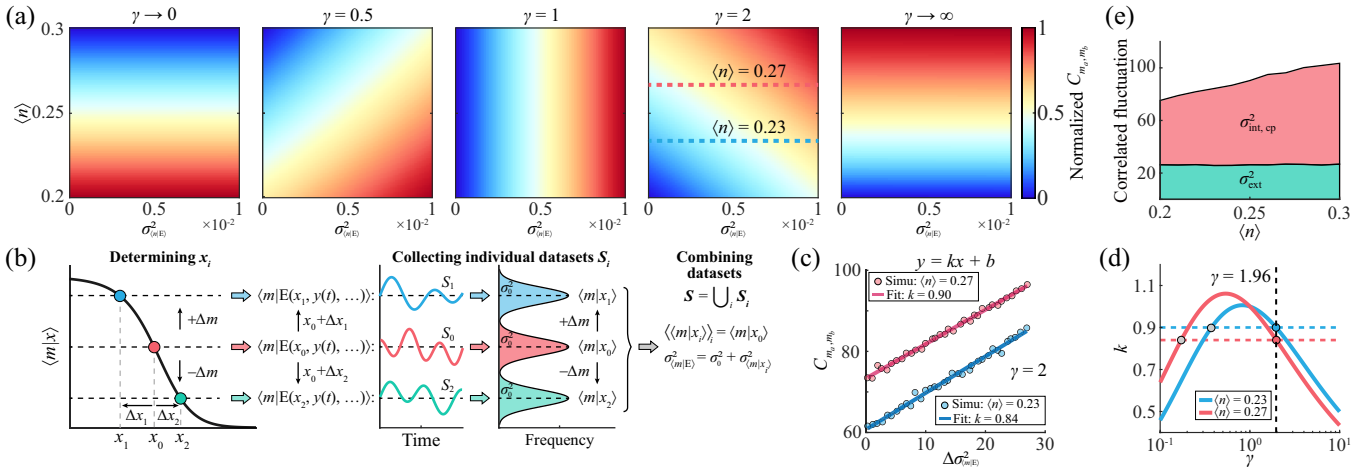

FIG. 3. Distinguishing extrinsic and intrinsic components in correlated transcriptional fluctuations. (a) Total covariance of nascent RNA between gene copies as a function of $\langle n\rangle$ and $\sigma^2_{\langle n|\mathrm{E}\rangle}$ for different γ values. (b) Schematic of perturbing an external factor x above and below its original value to construct a combined dataset with increased $\sigma^2_{\langle m|\mathrm{E}\rangle}$ at fixed $\langle m\rangle$ (and $\langle n\rangle$). (c) Linear relationships between total covariance of nascent RNA and $\sigma^2_{\langle m|\mathrm{E}\rangle}$ at different $\langle n\rangle$ values for the simulated data. (d) Estimation of γ from the slopes of the linear relationships in (c). (e) Noise decomposition of the simulated data in (c–d) as a function of $\langle n\rangle$ for fixed $\sigma^2_{\langle n|\mathrm{E}\rangle}$. (b–e) The kinetic parameters are set as: $k_{\mathrm{ON}} = 0.01$–0.2 min^{-1}, $k_{\mathrm{OFF}} = 0.05$ min^{-1}, $k_{\mathrm{INI}} = 50$ min^{-1}, $T_{\mathrm{RES}} = 2$ min, $g(\tau) = -\tau$, and $\gamma = 2$.

2. Measuring the partial slope of C_{m_a,m_b}: By carefully perturbing certain external factor, we can construct multiple datasets with varied $\sigma^2_{\langle m|\mathrm{E}\rangle}$ at fixed $\langle n\rangle$. Computing the differences between these datasets in C_{m_a,m_b} and $\sigma^2_{\langle m|\mathrm{E}\rangle}$ (using Eq. (14)) allows us to evaluate Eq. (12) at a given $\langle n\rangle$, which typically yields two possible solutions for γ [Fig. 3(c), see Supplemental Material [27]].

3. Determining γ: By repeating step 2 for different $\langle n\rangle$ values, we can identify the correct γ as the common solution across all cases [Fig. 3(d) and S2, see Supplemental Material [27]].

4. Decomposing C_{m_a,m_b}: Once γ is determined, we can separate total covariance into $\sigma^2_{\mathrm{ind,cp}}$ and σ^2_{ext} for any given $\langle n\rangle$ based on Eqs. (10)–(11) [Figs. 3(e), see Supplemental Material [27]].

Applying this framework to simulated data confirms its effectiveness and accuracy [Fig. 3(c–e)].

Uncovering intrinsic couplings in Drosophila gene expression—To demonstrate the application of the above framework on experimental data, we examine the transcription of the *hunchback* (*hb*) gene in *Drosophila melanogaster* embryos [55] (Supplemental Material [27]). During early development, *hb* exhibits a logistic expression pattern along the anterior-posterior (AP) axis of the embryo, with its activation probability varying significantly across AP positions and $k_{\mathrm{ON}}T_{\mathrm{RES}}$ & $k_{\mathrm{OFF}}T_{\mathrm{RES}} \lesssim 1$ [24,51,57] [Fig. 4(a)]. In the later mitotic interphase of each cell cycle, there are two pairs of sister *hb* alleles in each nucleus. By quantifying nascent RNA at each *hb* allele in individual nuclei, we separately analyze correlated fluctuations for adjacent sister alleles and distant non-sister alleles, both previously assumed to be independent [24,51,57].

To apply our framework, we use the nuclear AP position x as a tuning factor for dataset construction. Specifically, by measuring the average *hb* expression level at each AP position ($\langle m|x\rangle$), we uniformly sample single-nucleus data around each given position using varying x-ranges to construct datasets with distinct $\langle n\rangle$ and $\Delta\sigma^2_{\langle m|\mathrm{E}\rangle}$ [Fig. 4(a), see Appendix C]. We find that the covariance between non-sister alleles primarily varies with $\sigma^2_{\langle m|x\rangle}$, indicating independent transcription [Fig. 4(b)]. In contrast, the covariance between sister alleles depends on both $\langle n\rangle$ and $\sigma^2_{\langle m|x\rangle}$, suggesting potential (anti-)coupling between them. Moreover, consistent with our theory, datasets with fixed $\langle n\rangle$ exhibit a linear relationship between C_{m_a,m_b} and $\sigma^2_{\langle m|x\rangle}$ [Fig. 4(c)], whose slopes reveal a coupling strength of $\gamma \approx 0.12$ and a negative $\sigma^2_{\mathrm{ind,cp}}$ for sister alleles [Figs. 4(d) and S3(a), see Supplemental Material [27]]. This finding uncovers a previously unidentified anti-coupling, which may result from competition between adjacent sister alleles for transcription factor recruitment [14,15].

We then apply this strategy to investigate whether the two alternative promoters, P1 and P2, of the *hb* gene are coupled, a question unresolved in previous research due to the existence of extrinsic noise [55] [Fig. 4(e), Supplemental Material [27]]. These promoters are regulated by the same set of enhancers and therefore exhibit similar expression patterns (with different amplitudes) during early development [55]. By quantifying nascent RNA production from each promoter, we compare correlated

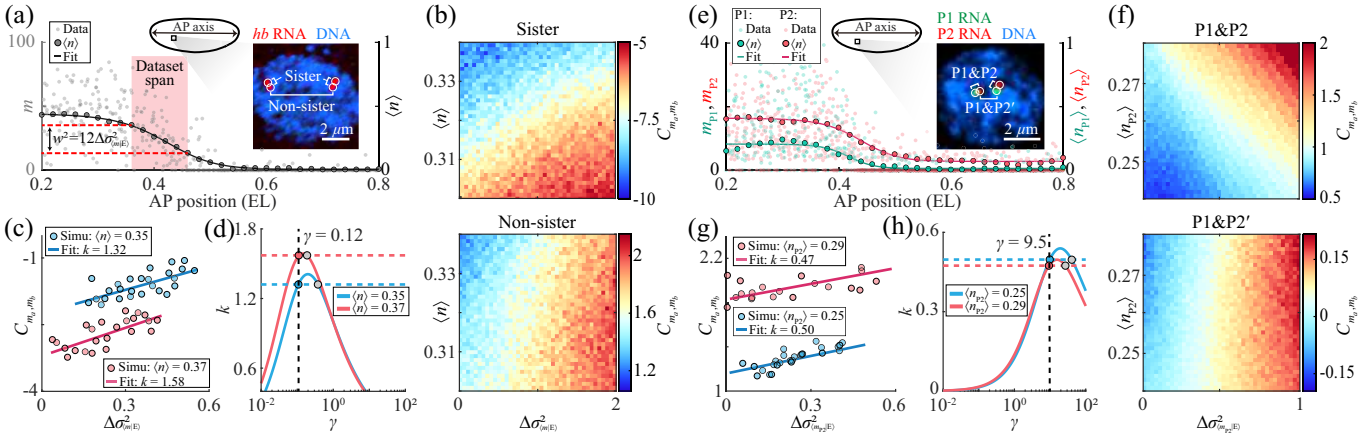

FIG. 4. Decomposing correlated transcriptional fluctuations from experimental data. (a) Nascent RNA signal at individual *hb* genes in a single embryo as a function of AP position. Single-nucleus data are binned along the AP axis to estimate $\langle n\rangle$ for fitting to a logistic function. Each dataset with given $\langle n\rangle$ and $\Delta\sigma^2_{\langle m|\mathrm{E}\rangle}$ is constructed by uniformly sampling nuclei within a specific AP window (red region). Inset: confocal image of a nucleus showing nascent RNAs from two sets of sister *hb* alleles. (b) Total covariance of nascent RNA between sister alleles (upper) and between non-sister alleles (lower) as functions of $\langle n\rangle$ and $\Delta\sigma^2_{\langle m|\mathrm{E}\rangle}$. (c) Linear relationships between total covariance of nascent RNA and $\Delta\sigma^2_{\langle m|\mathrm{E}\rangle}$ at different $\langle n\rangle$ values. (d) Estimation of γ between sister alleles from the slopes of the linear relationships in (c). (e) Nascent RNA signal at individual *hb* P1 and P2 promoters in a single embryo as functions of AP position. Single-nucleus data are binned along the AP axis and to estimate $\langle n_{\mathrm{P1}}\rangle$ and $\langle n_{\mathrm{P2}}\rangle$ for fitting to logistic functions. Inset: confocal image of a nucleus showing nascent RNAs from P1 and P2. (f) Total covariance of nascent RNA between P1 and P2 on the same (upper) or different (lower) homologous alleles as functions of $\langle n_{\mathrm{P2}}\rangle$ and $\Delta\sigma^2_{\langle m_{\mathrm{P2}}|\mathrm{E}\rangle}$. (g) Linear relationships between total covariance of nascent RNA and $\Delta\sigma^2_{\langle m_{\mathrm{P2}}|\mathrm{E}\rangle}$ at different $\langle n_{\mathrm{P2}}\rangle$ values. (h) Estimation of γ between P1 and P2 from the slopes of the linear relationships in (g).

fluctuations between P1 and P2 on either the same *hb* allele or different homologous alleles (denoted as P1 and P2'). We find that their C_{m_a,m_b} values exhibit distinct relationships with $\langle n\rangle$ and $\sigma^2_{\langle m|x\rangle}$ [Fig. 4(f)], suggesting potential coupling between P1 and P2 and independence between P1 and P2'. By extending our theory to describe non-identical gene pairs [Figs. S3(b–c), see Supplemental Material [27]], we estimate a coupling strength of $\gamma \approx 9.5$ and a positive $\sigma^2_{\mathrm{ind,cp}}$ for P1 and P2 [Figs. 4(g–h) and S3(d)]. This previously unrecognized positive coupling is consistent with enhancer-shared promoter coupling observed in synthetic systems [20].

Summary and outlook—In this Letter, we present a simple theoretical model to depict the stochastic nascent transcription of an intrinsically coupled gene pair under fluctuating environments. By analytically solving nascent RNA fluctuations, we develop a new noise decomposition framework to quantitatively distinguish correlated transcriptional fluctuations caused by intrinsic coupling from those induced by extrinsic variability, leveraging their distinct dependencies on the gene's ON-state statistics. Applying it to single-cell transcription data from *Drosophila* embryos, we uncover couplings between sister alleles and between alternative promoters of an endogenous gene, both of which were challenging to identify in previous research.

Our findings broaden the classical picture of gene expression noise and provide a quantitative framework for deciphering intrinsic gene-gene interactions in fluctuating environments. While this study focuses on a specific k_{ON}-coupling mechanism, our noise decomposition framework is broadly applicable to other coupling mechanisms (Supplemental Material [27]) and can be extended to investigate similar effects in other biochemical networks. Further generalizing this approach to analyze the temporal characteristics of correlated fluctuations among multiple genes will provide deeper insights into dynamic gene-gene interactions within complex gene regulatory networks.

Acknowledgments—We thank Jingyao Wang for providing previously published imaging data, and Michael Levine and Bomyi Lim for their valuable discussions and feedback. This work was supported by the National Key R&D Program of China (grant no. 2021YFA0910702), the National Natural Science Foundation of China (grant no. 12474194, 11774225), and the Natural Science Foundation of Shanghai (grant no. 22ZR1434000). We gratefully acknowledge the imaging and computing resources provided by the Student Innovation Center at Shanghai Jiao Tong University, and sincerely thank Liuyin Fan for the dedicated management and support of these resources.

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[2] A. Raj and A. Van Oudenaarden, Cell **135**, 216 (2008).

[3] J. M. Raser and E. K. O'Shea, Science **304**, 1811 (2004).
[4] M. S. Sherman, K. Lorenz, M. H. Lanier, and B. A. Cohen, Cell Syst. **1**, 315 (2015).
[5] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. USA **99**, 12795 (2002).
[6] A. Hilfinger and J. Paulsson, Proc. Natl. Acad. Sci. USA **108**, 12167 (2011).
[7] A. Hilfinger, M. Chen, and J. Paulsson, Phys. Rev. Lett. **109**, 248104 (2012).
[8] C. G. Bowsher and P. S. Swain, Proc. Natl. Acad. Sci. USA **109**, E1320 (2012).
[9] L. Ham, M. Jackson, and M. P. Stumpf, eLife **10**, e69324 (2021).
[10] J. Lin and A. Amir, Phys. Rev. Lett. **126**, 078101 (2021).
[11] M. Sun and J. Zhang, Nucleic Acids Res. **48**, 533 (2020).
[12] A. Rhee, R. Cheong, and A. Levchenko, Proc. Natl. Acad. Sci. USA **111**, 17330 (2014).
[13] J. Rodriguez, G. Ren, C. R. Day, K. Zhao, C. C. Chow, and D. R. Larson, Cell **176**, 213 (2019).
[14] R. C. Brewster, F. M. Weinert, H. G. Garcia, D. Song, M. Rydenfelt, and R. Phillips, Cell **156**, 1312 (2014).
[15] Y. Zhang, T. D. Ho, N. E. Buchler, and R. Gordân, Genome Res. **31**, 1216 (2021).
[16] J. E. Henninger, O. Oksuz, K. Shrinivas, I. Sagi, G. LeRoy, M. M. Zheng, J. O. Andrews, A. V. Zamudio, C. Lazaris, N. M. Hannett, *et al.*, Cell **184**, 207 (2021).
[17] U. Alon, *An introduction to systems biology: design principles of biological circuits* (Chapman and Hall/CRC, 2019).
[18] I. W. Duncan, Annu. Rev. Genet. **36**, 521 (2002).
[19] B. Lim, T. Heist, M. Levine, and T. Fukaya, Mol. Cell **70**, 287 (2018).
[20] T. Fukaya, B. Lim, and M. Levine, Cell **166**, 358 (2016).
[21] M. Levo, J. Raimundo, X. Y. Bing, Z. Sisco, P. J. Batut, S. Ryabichko, T. Gregor, and M. S. Levine, Nature **605**, 754 (2022).
[22] B. Munsky, G. Neuert, and A. Van Oudenaarden, Science **336**, 183 (2012).
[23] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Phys. Rev. Lett. **117**, 128101 (2016).
[24] B. Zoller, S. C. Little, and T. Gregor, Cell **175**, 835 (2018).
[25] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).
[26] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[27] See Supplemental Material at [URL] for detailed derivations, simulation and experimental settings, data analysis methods, and discussions, which includes Refs. [28-48].
[28] S. Komili and P. A. Silver, Nat. Rev. Genet. **9**, 38 (2008).
[29] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[30] H. G. Garcia, M. Tikhonov, A. Lin, and T. Gregor, Curr. Biol. **23**, 2140 (2013).
[31] N. C. Lammers, V. Galstyan, A. Reimer, S. A. Medin, C. H. Wiggins, and H. G. Garcia, Proc. Natl. Acad. Sci. USA **117**, 836 (2020).
[32] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, *et al.*, Nature **560**, 494 (2018).
[33] D. L. Bentley, Nat. Rev. Genet. **15**, 163 (2014).
[34] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Science **332**, 475 (2011).
[35] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[36] I. Golding and A. Amir, Rev. Mod. Phys. **96**, 041001 (2024).
[37] H. Chen, K. Shiroguchi, H. Ge, and X. S. Xie, Mol. Syst. Biol. **11**, 781 (2015).
[38] M. Wang, J. Zhang, H. Xu, and I. Golding, Nat. Microbiol. **4**, 2118 (2019).
[39] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[40] B. R. Cairns, Nature **461**, 193 (2009).
[41] A. Coulon, C. C. Chow, R. H. Singer, and D. R. Larson, Nat. Rev. Genet. **14**, 572 (2013).
[42] S. Patange, D. A. Ball, Y. Wan, T. S. Karpova, M. Girvan, D. Levens, and D. R. Larson, Cell Rep. **38** (2022).
[43] N. Yudkovsky, J. A. Ranish, and S. Hahn, Nature **408**, 225 (2000).
[44] S. J. Philips, M. Canalizo-Hernandez, I. Yildirim, G. C. Schatz, A. Mondragón, and T. V. O'Halloran, Science **349**, 877 (2015).
[45] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Science **332**, 472 (2011).
[46] B. Zoller, D. Nicolas, N. Molina, and F. Naef, Mol. Syst. Biol. **11**, 823 (2015).
[47] C. V. Harper, B. Finkenstädt, *et al.*, PLoS Biol. **9**, e1000607 (2011).
[48] H. Bremer and P. P. Dennis, EcoSal Plus **3**, 10 (2008).
[49] L. Ham, R. D. Brackston, and M. P. Stumpf, Phys. Rev. Lett. **124**, 108101 (2020).
[50] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Science **307**, 1962 (2005).
[51] H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding, Nat. Methods **12**, 739 (2015).
[52] J. Szavits-Nossan and R. Grima, Phys. Rev. Res. **5**, 013064 (2023).
[53] J. Szavits-Nossan and R. Grima, Biophys. J. **123**, 1034 (2024).
[54] J. V. Meeussen and T. L. Lenstra, Trends Genet. **40**, 160 (2024).
[55] J. Wang, S. Zhang, H. Lu, and H. Xu, Nat. Commun. **13**, 2714 (2022).
[56] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, Nat. Genet. **43**, 554 (2011).
[57] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).

End Matter

Appendix A: Master equation analysis of the model—The stochastic dynamics of the system can be solved using either master equation or queueing theory approaches [23, 52, 53]. In this work, we adopt the master equation framework, as it offers a straightforward derivation of the statistical moments of nascent RNA. Specifically, by defining $\mathcal{M}_i(\tau, t) \equiv \sum_{t-T_{\rm RES} \le t_j \le t+\tau} g(t_j - t)$ as the partial

accumulation of nascent RNA signal over the history from $t - T_{\rm RES}$ to $t + \tau$, we write the corresponding master equation as:

$$\frac{d\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b)}{d\tau} = (\mathbf{K} - \mathbf{K}_{{\rm INI},a} - \mathbf{K}_{{\rm INI},b})\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b) + \mathbf{K}_{{\rm INI},a}\mathbf{P}(\mathcal{M}_a - g(\tau), \mathcal{M}_b) + \mathbf{K}_{{\rm INI},b}\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b - g(\tau)), \quad \text{(A1)}$$

where $\mathbf{K} = \begin{bmatrix} -2k_{\rm ON} & k_{\rm OFF} & k_{\rm OFF} & 0 \\ k_{\rm ON} & -\gamma k_{\rm ON} - k_{\rm OFF} & 0 & k_{\rm OFF} \\ k_{\rm ON} & 0 & -\gamma k_{\rm ON} - k_{\rm OFF} & k_{\rm OFF} \\ 0 & \gamma k_{\rm ON} & \gamma k_{\rm ON} & -2k_{\rm OFF} \end{bmatrix}$, $\mathbf{K}_{{\rm INI},a} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & k_{\rm INI} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{\rm INI} \end{bmatrix}$, $\mathbf{K}_{{\rm INI},b} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & k_{\rm INI} & 0 \\ 0 & 0 & 0 & k_{\rm INI} \end{bmatrix}$, and $\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b) = \begin{bmatrix} P(n_a{=}0, n_b{=}0, \mathcal{M}_a, \mathcal{M}_b) \\ P(n_a{=}1, n_b{=}0, \mathcal{M}_a, \mathcal{M}_b) \\ P(n_a{=}0, n_b{=}1, \mathcal{M}_a, \mathcal{M}_b) \\ P(n_a{=}1, n_b{=}1, \mathcal{M}_a, \mathcal{M}_b) \end{bmatrix}$. The probability distribution of the system state at time t corresponds to the solution of Eq. (A1) at $\tau = 0$, i.e., $\mathbf{P}_t(m_a, m_b) = \mathbf{P}_{\tau=0}(\mathcal{M}_a, \mathcal{M}_b)$, given the initial marginal distribution of gene states [23]. Note that Eq. (A1) can be nondimensionalized by redefining τ as $\tau/T_{\rm RES}$ and multiplying all kinetic rates by $T_{\rm RES}$. Without loss of generality, we set $T_{\rm RES} = 1$ for this study.

Under a fixed environmental condition, the steady-state marginal distribution of gene state, $\tilde{\mathbf{P}}_0$, is easily solved from $\mathbf{K}\tilde{\mathbf{P}}_0 = 0$, yielding Eq. (2). To further determine the fluctuations of m_a and m_b, we apply characteristic function method to obtain Eq. (5) and (Supplemental Material [27]):

$$\langle m_i m_j | {\rm E}\rangle = 2\mathbf{u}\mathbf{K}_{{\rm INI},i}\left[\int_{-1}^{0} d\tau_1 g(\tau_1) \int_{-1}^{\tau_1} d\tau_2 g(\tau_2) e^{\mathbf{K}(\tau_1 - \tau_2)}\right]\mathbf{K}_{{\rm INI},j}\tilde{\mathbf{P}}_0 + \overline{g^2} k_{\rm INI}\langle n|{\rm E}\rangle\delta_{ij}, \quad \text{(A2)}$$

where $\mathbf{u} = [1, 1, 1, 1]$, $\bar{g} = \int_{-1}^{0} g(\tau)d\tau$, $\overline{g^2} = \int_{-1}^{0} g^2(\tau)d\tau$, and $i, j = a$ or b. Notably, $\mathbf{K}$ is negative semi-definite, whose eigenvalues are proportional to the magnitudes of $k_{\rm ON}$ and $k_{\rm OFF}$. Therefore, $\langle m_i m_j|{\rm E}\rangle$ decays with the scale of gene-state transition rates.

Through a Taylor expansion of $e^{\mathbf{K}(\tau_1-\tau_2)}$ in Eq. (A2), we derive expressions for the nascent-RNA covariance and total variance as:

$$C_{m_a,m_b|{\rm E}} = \bar{g}^2 k_{\rm INI}^2 C_{n_a,n_b|{\rm E}} + \sum_{s=2}^{\infty} \frac{2\overline{\Delta\tau^s}}{s!}\mathbf{u}\mathbf{K}_{{\rm INI},a}\mathbf{K}^s\mathbf{K}_{{\rm INI},b}\tilde{\mathbf{P}}_0, \quad \text{(A3)}$$

$$\sigma^2_{m|{\rm E}} = \bar{g}^2 k_{\rm INI}^2 \sigma^2_{n|{\rm E}} + \overline{g^2} k_{\rm INI}\langle n|{\rm E}\rangle + \sum_{s=1}^{\infty} \frac{2\overline{\Delta\tau^s}}{s!}\mathbf{u}\mathbf{K}_{{\rm INI},a}\mathbf{K}^s\mathbf{K}_{{\rm INI},a}\tilde{\mathbf{P}}_0, \quad \text{(A4)}$$

where $\overline{\Delta\tau^s} = \int_{-1}^{0} d\tau_1 \int_{-1}^{\tau_1} d\tau_2 g(\tau_1)g(\tau_2)(\tau_1 - \tau_2)^s$. Since $\overline{\Delta\tau^1}$ and $\overline{\Delta\tau^2} \ll 1$ under slow gene-state transitions (Supplemental Material [27]), Eqs. (A3) and (A4) reduce to (6) and (7) without significant deviations until $k_{\rm ON}$ & $k_{\rm OFF} > 1$ [Fig. 2(d) and S1(b)].

Appendix B: Tuning $\sigma^2_{\langle m|E\rangle}$ while keeping $\langle m\rangle$ fixed—Gene transcription in real biological systems is typically influenced by multiple external factors. While many of these factors remain unidentified or beyond experimental control, some may be sufficiently characterized to enable experimental modulation. Suppose we can manipulate a particular $k_{\rm ON}$- or $k_{\rm OFF}$-regulating factor, x, without influencing the fluctuations of other external factors (e.g., $y(t)$). By fixing x at a specific level, x_0, we can collect a dataset S_0 of m_a and m_b either from a time series of a single cell or from a snapshot of an ensemble of cells. The average gene transcription level in S_0, denoted as $\langle m|x_0\rangle \equiv \langle\langle m|{\rm E}(x_0, y(t), \cdots)\rangle\rangle_t$, is directly measurable. Due to the fluctuations of other external factors (and any trace fluctuation in x), the extrinsic variance of nascent RNA in S_0, denoted as $\sigma_0^2 \equiv \sigma^2_{\langle m|{\rm E}(x_0,\cdots)\rangle}$, is nonzero [Fig. 3(b)].

Now consider applying a constant perturbation to x (i.e., $x_1 = x_0 + \Delta x_1$) to generate a new dataset S_1 [Fig. 3(b)]. The average nascent RNA level will shift to $\langle m|x_1\rangle$ accordingly. More importantly, the extrinsic variance of nascent RNA, to a first-order approximation, remains unchanged for constant Δx_1, i.e.,

$$\sigma^2_{\langle m|{\rm E}(x_1,\cdots)\rangle} \approx \sigma_0^2 + \left(\partial_x\langle m|x\rangle|_{x_0}\right)^2\sigma^2_{\Delta x_1} = \sigma_0^2. \quad \text{(B1)}$$

Therefore, if we collect multiple datavsets at a series of x values ($x_i,\ i = 1, 2, \cdots$) [Fig. 3(b)], the average nascent RNA level in the combined dataset $S = \bigcup_i S_i$ becomes:

$$\langle m\rangle = \langle\langle m|x_i\rangle\rangle_i\,. \quad \text{(B2)}$$

By knowing how $\langle m|x\rangle$ varies with x, we can select a suitable set of x_i around x_0, so that S has the same $\langle m\rangle$ as S_0. Meanwhile, the extrinsic variance of m in S satisfies:

$$\sigma^2_{\langle m|{\rm E}\rangle} = \left\langle\sigma^2_{\langle m|{\rm E}(x_i),\cdots\rangle}\right\rangle + \sigma^2_{\langle m|x_i\rangle}. \quad \text{(B3)}$$

Since $\sigma^2_{\langle m|{\rm E}(x_i,\cdots)\rangle}$ is roughly unchanged across individual datasets, the increase in $\sigma^2_{\langle m|{\rm E}\rangle}$ due to dataset combination can be approximated by $\sigma^2_{\langle m|x_i\rangle}$ (Eq. 14). This strategy enables us to quantitatively tune $\sigma^2_{\langle m|{\rm E}\rangle}$ while keeping $\langle m\rangle$ (and $\langle n\rangle$) fixed.

Appendix C: Constructing datasets for Drosophila embryos—During early *Drosophila* development, the *hb* gene is primarily expressed in the anterior half of the embryo, forming a logistic expression pattern along the AP axis. Previous studies have reported that such spatial variation in transcription is primarily driven by an AP-position (x) dependent regulation of $k_{\rm ON}$ [51,55]. Therefore, by grouping single-nucleus data across different x

values, we can construct a series of datasets with systematically varied $\langle n \rangle$ and $\sigma^2_{\langle m|x \rangle}$ for covariance analysis. This proceeds as follows:

(1) Measuring $\langle n|x \rangle$ as a function of nuclear position: By binning the single-nucleus data along the AP axis (bin size: 0.05 EL), we compute, for each bin, the average levels of nascent RNA ($\langle m|x \rangle$) and ON-state probability ($\langle n|x \rangle$, using Eq. (13)). Since these two quantities are typically proportional, we concentrate on $\langle n|x \rangle$ in the following description for simplicity. Here, we focus primarily on the AP-position range 0.20–0.60 embryo length (EL), where the per-allele nascent RNA level decreases systematically from high to low. Within this region, we smooth the trend of $\langle n|x \rangle$ with respect to x by fitting it to a logistic function $L(x) = A\frac{e^{-(x-x_0)/\lambda}}{1+e^{-(x-x_0)/\lambda}} + B$, where x_0, λ, A, and B are fitting parameters.

(2) Selecting a set of x for dataset construction: To construct a combined dataset with desired values of $\langle n \rangle$ and $\Delta\sigma^2_{\langle n|\mathrm{E} \rangle}$, we first need to determine a set of x values for individual datasets. In principle, any x-set whose corresponding $\langle n|x \rangle$ values yield the target $\langle n \rangle$ and $\Delta\sigma^2_{\langle n|\mathrm{E} \rangle}$ is acceptable. For simplicity, we consider a uniform distribution of $\langle n|x \rangle$ values centered at $\langle n \rangle$, with a width satisfying $w^2 = 12\Delta\sigma^2_{\langle n|\mathrm{E} \rangle}$. By discretizing this distribution ($\Delta n = 0.005$), we obtain the corresponding x for each $\langle n|x \rangle$ via $x = L^{-1}(\langle n|x \rangle)$, yielding a set of x values.

(3) Sampling single-nucleus data from each x to construct a combined dataset: For each x in the above set, we collect all single-nucleus data around it within a bin width of $0.005/L'(x)$ EL. This x-sensitive bin width is chosen to maintain a fixed range of $\langle n|x \rangle$ within each bin. Since the number of nuclei in each bin (denoted as N_i) varies, we randomly sample an equal number of nuclei from each bin ($N = \min(N_i)$) to construct a combined dataset. This guarantees balanced contributions from each AP position in the final dataset.