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: March 13, 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 [4–8]:

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

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]. 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 [12, 13], and their products may exert feedback regulations [14, 15]. 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 [16, 17]. Similar enhancer-sharing mechanisms also occur between genes on the same chromosome or between promoters of the same gene [18, 19]. 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, these effects may be non-negligible [11, 19]. While some indirect effects, such as feedback regulations, can be separated using short-lived reporters (e.g., nascent RNA reporters) or time-correlation analysis [11, 20], 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 in a pair of intrinsically coupled identical genes [Fig. 1(a), see Supplemental Material] [21–24, 31]. 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 with rate k_{INI}, followed by a constant-speed (V_{EL}) RNA synthesis (of length L) and a post-synthesis residence duration T_{S}

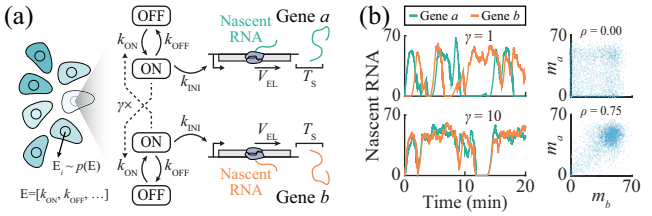

FIG. 1. A stochastic model of coupled gene transcription. (a) Model schematic. (b) Simulated results for the model with and without intrinsic coupling under a fixed environmental condition. Left: Time traces of gene transcription. Right: Scatter plot showing nascent RNA levels of both gene copies across all time points. $\gamma \neq 1$ leads to correlated fluctuation. See Supplemental Material for detailed simulation settings [31].

before the nascent RNA is released.

To account for extrinsic variability, we assume that all kinetic parameters $\mathrm{E} = [k_{\mathrm{ON}}, k_{\mathrm{OFF}}, \cdots]$ may vary across cells and over time following a probability distribution $p(\mathrm{E})$ [Fig. 1(a), see Supplemental Material] [31]. 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 [25, 26]. Among the parameters in E, k_{ON} and k_{OFF} are likely the primary targets of extrinsic fluctuations [4, 26]. 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 [21, 27]. 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 $T_{\mathrm{RES}} = L/V_{\mathrm{EL}} + T_{\mathrm{S}}$ prior to t, i.e., $m_i(t) = \sum_{t-T_{\mathrm{RES}} \leq t_j \leq t} g(t_j - t)$ (Supplemental Material) [21, 31]. Here, $g(\tau)$ represents the contribution of a nascent RNA molecule initiated at time τ (relative to t, $-T_{\mathrm{RES}} \leq \tau \leq 0$) to the total nascent RNA signal at t. By defining $\mathcal{M}_i(\tau, t) \equiv \sum_{t-T_{\mathrm{RES}} \leq t_j \leq t+\tau} g(t_j - t)$ as the partial accumulation of nascent RNA signal over the history from $t - T_{\mathrm{RES}}$ to $t + \tau$, we write the following master equation:

$$\begin{aligned}\frac{d\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b)}{d\tau} &= (\mathbf{K} - \mathbf{K}_{\mathrm{INI},a} - \mathbf{K}_{\mathrm{INI},b})\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b) \\ &\quad + \mathbf{K}_{\mathrm{INI},a}\mathbf{P}(\mathcal{M}_a - g(\tau), \mathcal{M}_b) \\ &\quad + \mathbf{K}_{\mathrm{INI},b}\mathbf{P}(\mathcal{M}_a, \mathcal{M}_b - g(\tau)), \qquad (2)\end{aligned}$$

where $\mathbf{K} = \begin{bmatrix} -2k_{\mathrm{ON}} & k_{\mathrm{OFF}} & k_{\mathrm{OFF}} & 0 \\ k_{\mathrm{ON}} & -\gamma k_{\mathrm{ON}} - k_{\mathrm{OFF}} & 0 & k_{\mathrm{OFF}} \\ k_{\mathrm{ON}} & 0 & -\gamma k_{\mathrm{ON}} - k_{\mathrm{OFF}} & k_{\mathrm{OFF}} \\ 0 & \gamma k_{\mathrm{ON}} & \gamma k_{\mathrm{ON}} & -2k_{\mathrm{OFF}} \end{bmatrix}$, $\mathbf{K}_{\mathrm{INI},a} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & k_{\mathrm{INI}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{\mathrm{INI}} \end{bmatrix}$, $\mathbf{K}_{\mathrm{INI},b} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & k_{\mathrm{INI}} & 0 \\ 0 & 0 & 0 & k_{\mathrm{INI}} \end{bmatrix}$, 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. (2) 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 [21]. Note that Eq. (2) can be nondimensionalized by redefining τ as τ/T_{RES} and multiplying all kinetic rates by T_{RES}. Without loss of generality, we set $T_{\mathrm{RES}} = 1$ for the rest of the paper.

Under a fixed environmental condition, the steady-state marginal distribution of n_a and n_b is given by (Supplemental Material) [31]:

$$\tilde{\mathbf{P}}_0 = \frac{1}{1 + 2\left(\frac{k_{\mathrm{ON}}}{k_{\mathrm{OFF}}}\right) + \gamma\left(\frac{k_{\mathrm{ON}}}{k_{\mathrm{OFF}}}\right)^2} \begin{bmatrix} 1 \\ k_{\mathrm{ON}}/k_{\mathrm{OFF}} \\ k_{\mathrm{ON}}/k_{\mathrm{OFF}} \\ \gamma(k_{\mathrm{ON}}/k_{\mathrm{OFF}})^2 \end{bmatrix}, \qquad (3)$$

which yields the covariance between gene states [Fig. 2(a)]:

$$C_{n_a,n_b|\mathrm{E}} = (\gamma - 1)\left(\frac{\left(\frac{k_{\mathrm{ON}}}{k_{\mathrm{OFF}}}\right)}{1 + 2\left(\frac{k_{\mathrm{ON}}}{k_{\mathrm{OFF}}}\right) + \gamma\left(\frac{k_{\mathrm{ON}}}{k_{\mathrm{OFF}}}\right)^2}\right)^2. \qquad (4)$$

As expected, $\gamma > 1$ and $0 < \gamma < 1$ lead to positive and negative covariance, respectively. For each $\gamma \neq 1$, Eq. (4) 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 (Sup-

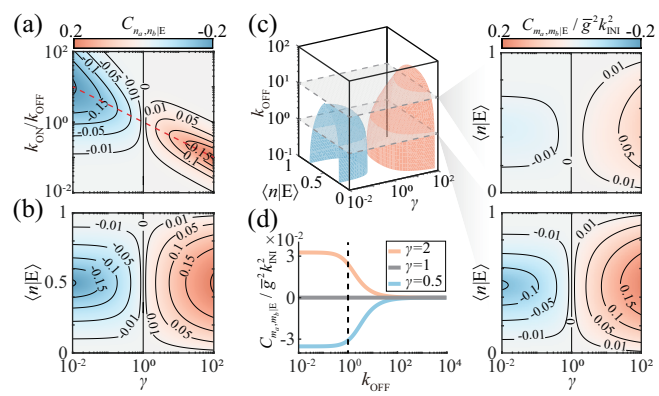

FIG. 2. Intrinsic covariance in coupled gene transcription. (a-b) Covariance of gene state as a function of γ and $k_{\mathrm{ON}}/k_{\mathrm{OFF}}$ (a) or γ and $\langle n|\mathrm{E}\rangle$ (b) under a fixed environmental condition. (c) Covariance of nascent RNA as a function of γ, $\langle n|\mathrm{E}\rangle$,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

plemental Material) [31]:

$$f(n,\gamma) = (\gamma-1)\left[\frac{(1-n)\left(2n-1+\sqrt{4(\gamma-1)n(1-n)+1}\right)}{2\gamma(1-n)+2n-1+\sqrt{4(\gamma-1)n(1-n)+1}}\right]^2. \tag{5}$$

Because $\langle n|\mathrm{E}\rangle$ is experimentally more accessible than k_{ON} and k_{OFF}, this reformulation helps connect our theory to experimental data in subsequent derivations.

To further solve the fluctuations of m_a and m_b, we apply characteristic function method to obtain (Supplemental Material) [19, 31]:

$$\langle m|\mathrm{E}\rangle = k_{\mathrm{INI}}\bar{g}\,\langle n|\mathrm{E}\rangle\,, \tag{6}$$

$$\langle m_i m_j|\mathrm{E}\rangle = 2\mathbf{u}\mathbf{K}_{\mathrm{INI},i}\left[\int_{-1}^{0} d\tau_1 g(\tau_1)\int_{-1}^{\tau_1} d\tau_2 g(\tau_2)\, e^{\mathbf{K}(\tau_1-\tau_2)}\right]\mathbf{K}_{\mathrm{INI},j}\tilde{\mathbf{P}}_0 + \overline{g^2}k_{\mathrm{INI}}\,\langle n|\mathrm{E}\rangle\,\delta_{ij}, \tag{7}$$

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_{ON} and k_{OFF}. Therefore, $\langle m_i m_j|\mathrm{E}\rangle$ decays with the scale of gene-state transition rates.

Derived from these equations, the covariance at nascent RNA level follows a similar trend to that at the gene-state level [Fig. 2(c)]. Specifically, in the limit of slow gene-state transitions (k_{ON} & $k_{\mathrm{OFF}} \to 0$), the covariances at these two levels are proportional, i.e.:

$$C_{m_a,m_b|\mathrm{E}} = \bar{g}^2 k_{\mathrm{INI}}^2 C_{n_a,n_b|\mathrm{E}}. \tag{8}$$

Outside this limiting case, the covariance can be expressed as a series expansion of $e^{\mathbf{K}(\tau_1-\tau_2)}$ (Supplemental Material) [31]:

$$C_{m_a,m_b|\mathrm{E}} = \bar{g}^2 k_{\mathrm{INI}}^2 C_{n_a,n_b|\mathrm{E}} + \sum_{s=2}^{\infty}\frac{2\overline{\Delta\tau^s}}{s!}\mathbf{u}\mathbf{K}_{\mathrm{INI},a}\mathbf{K}^s\mathbf{K}_{\mathrm{INI},b}\tilde{\mathbf{P}}_0, \tag{9}$$

where $\overline{\Delta\tau^s} = \int_{-1}^{0} d\tau_1 \int_{-1}^{\tau_1} d\tau_2 g(\tau_1)g(\tau_2)(\tau_1-\tau_2)^s$. Notably, since the first-order term ($s=1$) in Eq. (9) cancels out and $\overline{\Delta\tau^2} \ll 1$ (Supplemental Material) [31], a significant deviation of $C_{m_a,m_b|\mathrm{E}}$ from Eq. (8) arises only when k_{ON} & $k_{\mathrm{OFF}} > 1$ [Fig. 2(d)]. Therefore, for experiments measuring short-lived nascent RNA signals, whose lifetime satisfies k_{ON} & $k_{\mathrm{OFF}} \lesssim 1$ in many biological systems [28], Eq. (8) remains a good approximation.

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 ($\sigma^2_{\mathrm{int,ind}}$) and coupled ($\sigma^2_{\mathrm{ind,cp}}$) components of intrinsic fluctuations:

$$\sigma_m^2 = \underbrace{\left\langle \sigma^2_{m|\mathrm{E}} - C_{m_a,m_b|\mathrm{E}}\right\rangle}_{\sigma^2_{\mathrm{int,ind}}} + \underbrace{\left\langle C_{m_a,m_b|\mathrm{E}}\right\rangle}_{\sigma^2_{\mathrm{int,cp}}} + \underbrace{\sigma^2_{\langle m|\mathrm{E}\rangle}}_{\sigma^2_{\mathrm{ext}}}, \tag{10}$$

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

$$C_{m_a,m_b} = \left\langle C_{m_a,m_b|\mathrm{E}}\right\rangle + \sigma^2_{\langle m|\mathrm{E}\rangle}. \tag{11}$$

Hypothetically, $\sigma^2_{\mathrm{ind,cp}}$ can be determined from Eq. (11) if the average expression level under every environmental condition ($\langle m|\mathrm{E}\rangle$) can be precisely measured.

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

$$\sigma^2_{\langle m|\mathrm{E}\rangle} \approx \bar{g}^2 k_{\mathrm{INI}}^2 \sigma^2_{\langle n|\mathrm{E}\rangle}, \tag{12}$$

i.e., the extrinsic fluctuation is simply proportional to the variance of ON-state probabilities over fluctuating environments ($\sigma^2_{\langle n|\mathrm{E}\rangle}$), independent of their mean ($\langle n\rangle$). In contrast, for k_{ON} & $k_{\mathrm{OFF}} \lesssim 1$, the coupled intrinsic fluctuation is a function of both $\langle n\rangle$ and $\sigma^2_{\langle n|\mathrm{E}\rangle}$ (Supplemental Material) [31]:

$$\left\langle C_{m_a,m_b|\mathrm{E}}\right\rangle \approx \bar{g}^2 k_{\mathrm{INI}}^2\left[f(\langle n\rangle,\gamma) + \frac{1}{2}\frac{\partial^2 f}{\partial n^2}\bigg|_{\langle n\rangle}\sigma^2_{\langle n|\mathrm{E}\rangle}\right]. \tag{13}$$

Therefore, the dependence of total covariance on $\langle n\rangle$ and $\sigma^2_{\langle n|\mathrm{E}\rangle}$ varies with the relative contributions of $\sigma^2_{\mathrm{ind,cp}}$ and σ^2_{ext} [Fig. 3(a)]. Specifically, for an independent gene pair ($\gamma = 1$), C_{m_a,m_b} is solely a function of $\sigma^2_{\langle n|\mathrm{E}\rangle}$. 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) [31]. In all other cases, C_{m_a,m_b} depends on both $\langle n\rangle$ and $\sigma^2_{\langle n|\mathrm{E}\rangle}$, with its gradient offering insights into γ.

Based on these results, we propose a quantitative framework for estimating γ, $\sigma^2_{\mathrm{ind,cp}}$, and σ^2_{ext} from experimental data (assuming k_{ON} & $k_{\mathrm{OFF}} \lesssim 1$) through the following steps:

1. Determining $\langle n\rangle$: Since $k_{\mathrm{INI}} \gg 1$ in most biological systems [28], we can use the total noise of nascent RNA (quantified as the squared coefficient of variation) to estimate $\langle n\rangle$ for any given experimental dataset, i.e., $\langle n\rangle \approx 1/(\eta_m^2+1)$ (Supplemental Material) [31].

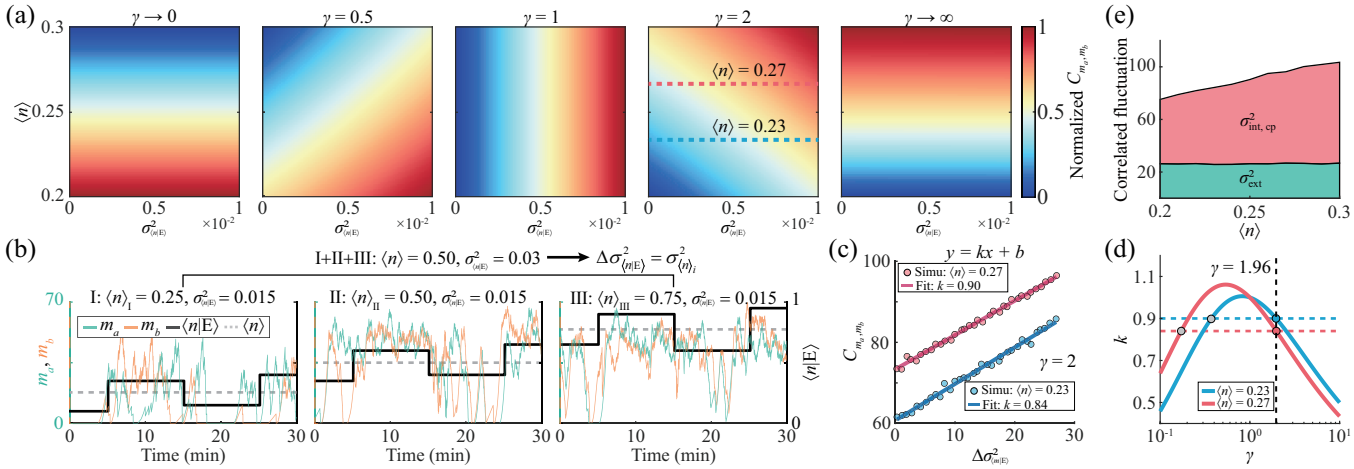

FIG. 3. Distinguishing extrinsic and intrinsic components in correlated transcriptional fluctuations. (a) Total covariance of nascent RNA between gene copies as a function of $\langle n\rangle$ and $\sigma^2_{\langle n|\mathrm{E}\rangle}$ for different γ values. (b) Simulated time traces of coupled gene transcription ($\gamma = 2$) under three noisy environments with different $\langle n\rangle$. Combining the three datasets increases $\sigma^2_{\langle n|\mathrm{E}\rangle}$ (See Supplemental Material for derivation of $\Delta\sigma^2_{\langle n|\mathrm{E}\rangle}$) [31]. (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) See Supplemental Material for detailed simulation settings [31].

2. Measuring the partial slope of C_{m_a,m_b}: Even without knowing the absolute level of $\sigma^2_{\langle n|\mathrm{E}\rangle}$, we can tune certain external factors to gather multiple datasets with distinct $\langle n\rangle$ (and similar $\sigma^2_{\langle n|\mathrm{E}\rangle}$). By combining these datasets, we can quantitatively increase $\sigma^2_{\langle m|\mathrm{E}\rangle}$ (and $\sigma^2_{\langle n|\mathrm{E}\rangle}$) while keeping $\langle m\rangle$ (and $\langle n\rangle$) fixed [Fig. 3(b), see Supplemental Material] [31]. This allows us to measure the partial slope of C_{m_a,m_b} with respect to $\sigma^2_{\langle n|\mathrm{E}\rangle}$ [Fig. 3(c)], i.e.:

$$\left.\frac{\partial C_{m_a,m_b}}{\partial \sigma^2_{\langle m|\mathrm{E}\rangle}}\right|_{\langle n\rangle} \approx \frac{1}{2}\left.\frac{\partial^2 f}{\partial n^2}\right|_{\langle n\rangle} + 1. \tag{14}$$

Evaluating this expression at a given $\langle n\rangle$ typically yields two possible solutions for γ.

3. Determining γ: By repeating step 2 for different values of $\langle n\rangle$, we can identify the correct γ as the common solution across all cases [Fig. 3(d), see Supplemental Material] [31].

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. (12)-(13) [Figs. 3(e) and S1, see Supplemental Material] [31].

Applying this framework to simulated data confirms its effectiveness and accuracy.

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 (Supplemental Material) [29, 31]. 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 $k_{\mathrm{ON}}T_{\mathrm{RES}}$ & $k_{\mathrm{OFF}}T_{\mathrm{RES}} \lesssim 1$ [22, 27, 30] [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 [22, 27, 30].

To apply our framework, we use the AP position differences between nuclei as a proxy and tuning factor for extrinsic fluctuations. By measuring the average *hb* expression level at each AP position ($\langle m|x\rangle$), we resample single-nucleus data to construct datasets with varying $\langle m\rangle$ and $\sigma^2_{\langle m|x\rangle}$ [Fig. 4(b), Supplemental Material][31]. We find that the covariance between non-sister alleles primarily varies with $\sigma^2_{\langle m|x\rangle}$, indicating independent transcription. 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 S2(a), see Supplemental Material] [31]. This finding uncovers a previously unidentified anti-coupling, which may result from competition between adjacent sister alleles for transcription factor recruitment [12, 13].

We then apply this strategy to investigate whether the two alternative promoters, P1 and P2, of the *hb* gene are

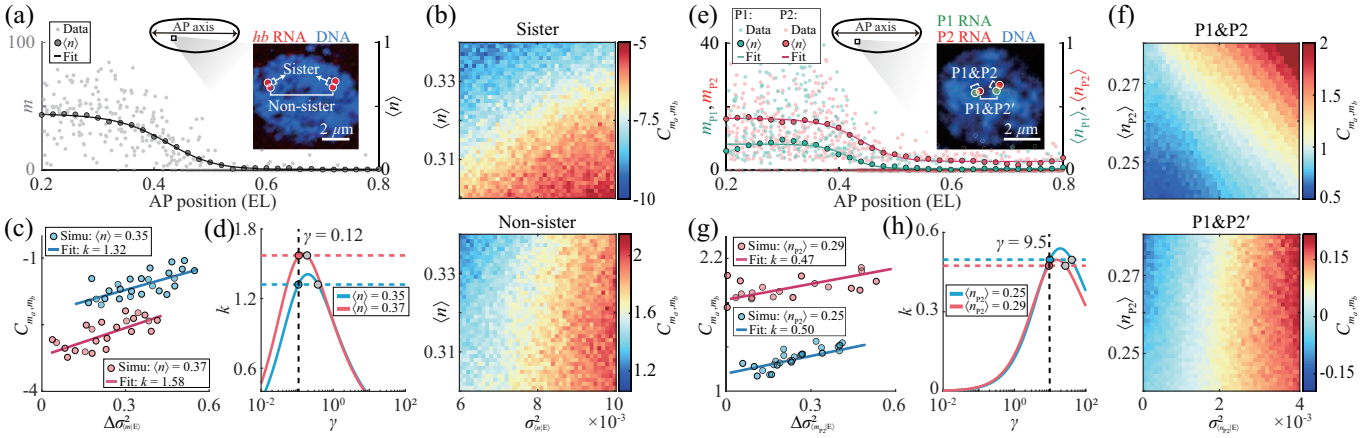

FIG. 4. Decomposing correlated transcriptional fluctuations from experimental data. (a) Nascent RNA signal at individual *hb* gene in a single embryo as a function of AP position (in units of total embryo length (EL)). Data were binned along the AP axis to estimate $\langle n\rangle$ for fitting to a logistic function. Inset: confocal image of a nucleus labeled for 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 $\sigma^2_{\langle n|\mathrm{E}\rangle}$. (c) Linear relationships between total covariance of nascent RNA and $\sigma^2_{\langle n|\mathrm{E}\rangle}$ 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. Data were 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 labeled for 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 $\sigma^2_{\langle m_{\mathrm{P2}}|\mathrm{E}\rangle}$. (g) Linear relationships between total covariance of nascent RNA and $\sigma^2_{\langle m_{\mathrm{P2}}|\mathrm{E}\rangle}$ 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).

coupled, a question unresolved in previous research due to the existence of extrinsic noise [29] [Fig. 4(e), Supplemental Material] [31]. These promoters are regulated by the same set of enhancers and therefore exhibit similar expression patterns (with different amplitudes) during early development [29]. By quantifying nascent RNA production from each promoter, we compare correlated 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 (see Supplemental Material and Figs. S2(b-c)) [31], 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 S2(d)]. This previously unrecognized positive coupling is consistent with enhancer-shared promoter coupling observed in synthetic systems [18].

Summary and outlook.—Direct gene-gene coupling and extrinsic noise are both key contributors to correlated fluctuations in gene transcription, yet a precise theoretical characterization and dissection of their contributions have been lacking. 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 for nascent RNA fluctuations, we show that intrinsic coupling alone can generate correlated fluctuation, whose magnitude depends on both the coupling strength and the ON-state probability of the gene. Based on these results, we extend the traditional noise decomposition scheme to distinguish correlated transcriptional fluctuations caused by intrinsic coupling from those induced by extrinsic variability. Since these two components exhibit distinct relationships with the mean and variance of the gene's ON-state probabilities across fluctuating environments, we develop a quantitative framework to disentangle them and estimate gene-pair coupling strength from experimental data. Applying this framework 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 primarily focuses on couplings occurring at gene-state transitions, our approach can be readily extended to explore similar effects at other stages of gene expression and 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.

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] M. Sun and J. Zhang, Nucleic Acids Res. **48**, 533 (2020).
[10] A. Rhee, R. Cheong, and A. Levchenko, Proc. Natl. Acad. Sci. USA **111**, 17330 (2014).
[11] J. Rodriguez, G. Ren, C. R. Day, K. Zhao, C. C. Chow, and D. R. Larson, Cell **176**, 213 (2019).
[12] R. C. Brewster, F. M. Weinert, H. G. Garcia, D. Song, M. Rydenfelt, and R. Phillips, Cell **156**, 1312 (2014).
[13] Y. Zhang, T. D. Ho, N. E. Buchler, and R. Gordân, Genome Res. **31**, 1216 (2021).
[14] 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).
[15] U. Alon, *An introduction to systems biology: design principles of biological circuits* (Chapman and Hall/CRC, 2019).
[16] I. W. Duncan, Annu. Rev. Genet. **36**, 521 (2002).
[17] B. Lim, T. Heist, M. Levine, and T. Fukaya, Mol. Cell **70**, 287 (2018).
[18] T. Fukaya, B. Lim, and M. Levine, Cell **166**, 358 (2016).
[19] M. Levo, J. Raimundo, X. Y. Bing, Z. Sisco, P. J. Batut, S. Ryabichko, T. Gregor, and M. S. Levine, Nature **605**, 754 (2022).
[20] B. Munsky, G. Neuert, and A. Van Oudenaarden, Science **336**, 183 (2012).
[21] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Phys. Rev. Lett. **117**, 128101 (2016).
[22] B. Zoller, S. C. Little, and T. Gregor, Cell **175**, 835 (2018).
[23] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).
[24] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[25] L. Ham, R. D. Brackston, and M. P. Stumpf, Phys. Rev. Lett. **124**, 108101 (2020).
[26] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Science **307**, 1962 (2005).
[27] H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding, Nat. Methods **12**, 739 (2015).
[28] J. V. Meeussen and T. L. Lenstra, Trends Genet. **40**, 160 (2024).
[29] J. Wang, S. Zhang, H. Lu, and H. Xu, Nat. Commun. **13**, 2714 (2022).
[30] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).
[31] See Supplemental Material at [URL] for detailed derivations, simulation and experimental settings, data analysis methods, and discussions, which includes Refs. [32-38].
[32] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).
[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] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[37] 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).
[38] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, Nat. Genet. **43**, 554 (2011).