



Heterogeneity is not always a source of noise: Stochastic gene expression in regulatory heterozygoteJuneil Jang ¹, François Amblard^{2,3} and C.-M. Ghim ^{1,3,*}¹*Department of Biomedical Engineering, Ulsan National Institute of Science & Technology, Ulsan 44919, Republic of Korea*²*Center for Soft and Living Matter, Institute for Basic Science, Ulsan 44919, Republic of Korea*³*Department of Physics, Ulsan National Institute of Science & Technology, Ulsan 44919, Republic of Korea*

(Received 26 October 2020; revised 8 August 2021; accepted 16 September 2021; published 1 October 2021)

Zygoty of diploid genome (i.e., degree to which two parental alleles of a gene have varied genetic sequences) adds another dimension to stochastic gene expression. The allelic imbalance in chromatin accessibility or divergence in regulatory sequences leads to fitness effects but the quantitative aspects thereof are largely left unexplored. We investigate diploid gene expression systems with homozygous (the same) and heterozygous (varied) combination of alleles in *cis*-regulatory sequences, not in structural gene loci, and characterize the zygoty-associated stochastic fluctuations in protein abundance. An emerging feat of heterozygoty is its counterintuitive capacity for genetic noise control. Especially when the noise is dominantly contributed to by the fluctuations in duty cycle (“reliability”) rather than in process speed (“productivity”) of gene expression machinery, its interallelic discrepancy acts to reduce the gene expression noise. These findings offer a novel insight into the rich repertoire of balancing selection enriched in the regulatory elements of immune response genes.

DOI: [10.1103/PhysRevE.104.044401](https://doi.org/10.1103/PhysRevE.104.044401)**I. INTRODUCTION**

Origins and consequences of cell-to-cell variability are essential to the understanding of diverse biological processes underlying the development, aging, immune response, and tumorigenesis, just to name a few [1–4]. As a micrometer-sized chemical reactor, living cells call for analytical frameworks that respect the stochasticity of biochemical reactions and discrete nature of macromolecules. In particular, gene expression involves a series of macromolecular binding and unbinding events orchestrated by enzyme complexes and transcription factors often occurring in low copy numbers. Together with genetic and environmental factors, this fundamental randomness may play a crucial role in determining the phenotype, giving rise to cell-to-cell variations in a clonal population subject to the same environmental conditions. Recent advances in molecular imaging and microfluidics-based cytometric technology led to an unprecedented opportunity in single-cell biology [5,6]. “Noisy” gene expression [7–9], or random fluctuations in the gene product levels, turned out to be an essential notion for deepening our quantitative understanding of central dogma and intracellular metabolism alike. Efforts to date, however, have been predominantly devoted to the natural and synthetic gene expression systems in prokaryotic cells. Despite the recent surge in single-cell genomics and single-molecule biophysics, full-fledged stochastic treatment of eukaryotic gene expression has been limited with a few notable exceptions [10,11].

Although the majority of genes in diploid cells (i.e., cells with paired chromosomes, one from each parent) express both

the maternal and paternal alleles, genes are often subject to epigenetic modifications that allow only one allele to be expressed, while leaving the other transcriptionally silent. This random monoallelic expression (MAE) accounts for up to 24% of the human genome according to a recent study [12]. What is the rationale, if any, behind this “idling” of an allele? Yet, despite its prevalence in eukaryotic gene expression system, the roles of MAE remain by and large to be elucidated. Recent transcriptomic studies [13,14] suggest that MAE increases the variability of the gene expression level in some specific tissues by generating different expression patterns based on allelic compositions even in the genetically identical cells. Another novel hypothesis put forward is that MAE is a mechanism for exercising tight control over expression level by avoiding the “overshoot” [13,14]. In the same vein, it was proposed that the heterozygous combination of alleles could be an important regulatory mechanism to achieve fine-tuning of gene expression levels [15,16] especially when robust titration of gene products is required for accurate decision making, e.g., during cell division and differentiation.

Motivated by the putative rationale behind MAE, here we address the general question of the gene expression noise in a regulatory heterozygote and compare it with its mean-expression-preserving homozygous counterparts. To be specific, contrary to the previous model [16], where the promoter dynamics is only implicit under the “adiabatic” approximation and the source of the “extrinsic noise” is limited to the cell-to-cell variability in the transcription factor concentration, we consider a stochastic model of diploid gene expression in a more general nonequilibrium setting, which takes the previous model as a limiting case of fast promoter switching.

*cmghim@unist.ac.kr

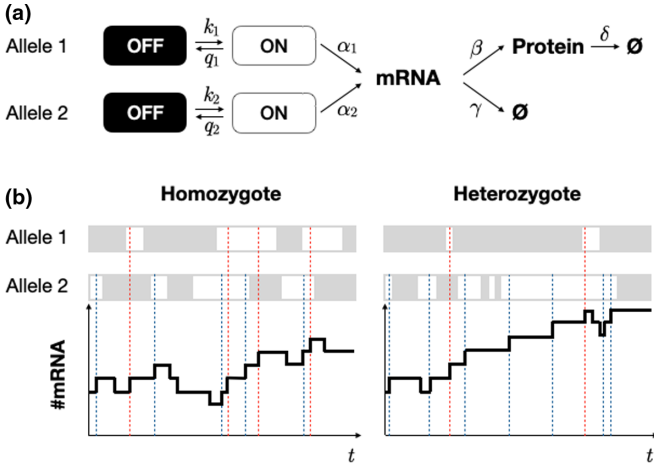


FIG. 1. Schematic of diploid gene expression system. (a) In general two alleles have divergent regulatory sequences or distinct epigenetic status, differing either in kinetic rates of activation and deactivation (k_i, q_i) or the rate of lumped transcription process α_i ($i = 1, 2$). (b) A realization of stochastic gene expression at the level of the transcript in homo- and heterozygous cell. The white and gray bands respectively denote the durations of “ON” and “OFF” state and the dotted lines in red or blue represent the events of mRNA synthesis from the allele 1 or 2.

II. MODEL OF DIPLOID GENE EXPRESSION

We employ the chemical master equation (CME) to describe a stochastic model of gene expression, where the distinct stages of transcription and translation are “coarse-grained” as discrete birth and death events of single mRNA and protein molecules (Fig. 1). The synthesis and degradation (or dilution) of the transcript is respectively governed by the propensity α and γ , but the synthesis occurs only when the allele is in an active state [17–21]. Unlike the transcription and other upstream events, translation and protein decay is not affected by the activity of an allele but governed by a constant propensity β and δ . The promoter switches back and forth between active and inactive promoter states at a rate k (activation) and q (deactivation), reflecting random binding and unbinding of the transcription factor(s) to the control element of the allele. When the transcription factors are non-competitively shared by the two alleles, the mean activity of an allele is controlled by a dimensionless equilibrium constant $K = q/k$ with $1/(1 + K) \equiv \theta$ being the duty cycle, i.e., probability that the allele is in an active state.

The model of diploid gene expression under consideration is composed of two alleles independently expressing the same gene products. The allelic differences lying only in the regulatory sequences, not in the coding region, a heterozygous gene with no interallelic interaction is modeled as a pair of alleles with distinct expression kinetics. The kinetic rates associated with each allele are independently modulated and the copy numbers of the gene products as a phenotype are thus additive. To be specific, the divergence in the regulatory sequences or varied epigenetic status may lead to a different set of propensity values (k_i, q_i), $i = 1, 2$, which are related to the “reliability” in a sense that the ratio q_i/k_i determines the probability of the failure-free operation of the gene expression

machinery. Another manifestation of heterozygosity, related to the “productivity” of the gene expression, is the interallelic discrepancy in the overall rate α_i of the open complex formation and the ensuing processive downstream events down to the synthesis of the nascent transcript. Since we consider identical coding regions, the other rates β and δ associated with the translation and protein decay are not affected by the zygosity.

For quantitative analysis, let us now define the probability $P_s(m, p, t)$ that a diploid cell has m and p copies of the mRNA and protein associated with a gene when its two promoters are in a state s at time t . The promoter state s can assume one of the four possible combinations of the active/inactive allele pairs. Then the chemical master equation is given by

$$\frac{\partial}{\partial t} P_s(m, p, t) = - \sum_{s'} (v_{ss'} + w_{ss'}) P_{s'}(m, p, t), \quad (1)$$

where the transition rate matrix $V = (v_{ss'})$ includes the rates involved in the change of promoter states, while $W = (w_{ss'})$ denotes the rates of the transcription, translation, and degradation events of the gene products. Further details of V and W can be found in Appendix A.

III. NOISE PROFILES

Because of the linearity of the underlying CME, the first and second moment of the steady-state probability distribution $|P(m, p, \infty)\rangle$ can be obtained exactly by the method of probability generating function. The mean and variance of the protein level thus obtained is

$$\langle p \rangle = \frac{\beta}{\delta} \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i} \equiv \sum_{i=1}^2 \langle p_i \rangle, \quad (2a)$$

$$\sigma_p^2 = (1 + b) \langle p \rangle + \frac{\delta}{1 + \delta} \sum_{i=1}^2 \langle p_i \rangle^2 g(k_i, q_i), \quad (2b)$$

where $b \equiv \beta/(1 + \delta)$, $\omega_i \equiv k_i + q_i$, and

$$g(k_i, q_i) \equiv \frac{K_i}{1 + \omega_i} \left(1 + \frac{1}{\delta + \omega_i} \right). \quad (3)$$

All the rates are now rescaled in units of γ . The noise, defined as the coefficient of variation squared (CV^2), is given by

$$\eta \equiv \frac{\sigma_p^2}{\langle p \rangle^2} = \underbrace{\frac{1 + b}{\langle p \rangle}}_{\equiv \eta_0} + \underbrace{\frac{\delta}{1 + \delta} \frac{1}{\langle p \rangle^2} \sum_{i=1}^2 \langle p_i \rangle^2 g(k_i, q_i)}_{\equiv \eta_1}. \quad (4)$$

The first term on the right-hand side, which we will denote as η_0 , originates from the fluctuations inherent to the birth-death process of the gene products. Being inversely proportional to mean copy number, η_0 is likely to matter only when the mean expression level is sufficiently low. The second term, η_1 , reflects the propagation of the “upstream” noise due to the ON-OFF switching of the promoter down to the protein level. As shown explicitly in Appendix B, the copy-number statistics of mRNA can be exactly described, up to its second moment, an intermittent Poisson process gated by a telegraph noise arising from the random ON/OFF switching of the promoter state. The factor $K_i = q_i/k_i = (1 - \theta_i)/\theta_i$ is, in fact,

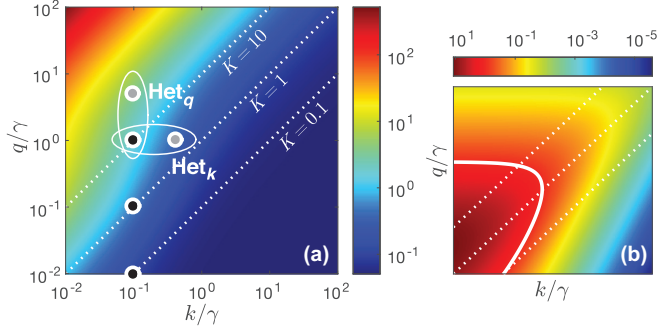


FIG. 2. Noise landscape of the monoallelic gene expression. (a) The color code represents CV^2 of the protein copy number produced by a single allele as a function of the activation and deactivation rates (k, q). The other rates are $\alpha = 4.2$, $\beta = 1.8$, and $\delta = 0.15$, which are all rescaled in units of γ . The broken lines along the diagonal are, from top to bottom, the contours of the average copy number $\langle p \rangle = 4.6, 25.2$, and 45.8 , each of which is the y -intercept of Figs. 3(a), 3(d), and 3(g). The three black dots indicate the reference alleles used in Figs. 3 and 4. Het_k/Het_q refers to the allele pairs composed of these reference alleles and their k -/ q -modified partner alleles. (b) The heatmap encodes the ratio η_1/η_0 . The white solid trajectory signifies $\eta_1 = \eta_0$, below which η_1 is the dominant contribution to the total protein noise. The axis ranges, left out for brevity, are the same as in panel (a).

CV^2 of the number of active alleles, which follows Bernoulli distribution. The remaining factors $(1 + \omega_i)^{-1}$ and $\delta/(1 + \delta)$ have an intuitive interpretation as a damping factor reflecting the time-averaging of the promoter noise propagated to the transcript level and of the mRNA noise to the protein level, respectively [22].

For notational simplicity and analytical utility as well, we introduce a parameter λ that captures the interallelic discrepancy in the rate parameters, and designate a general genotype by the allelic rate pairs $(k, \lambda k)$, $(q, \lambda q)$, and/or $(\alpha, \lambda \alpha)$, where $\lambda = 1$ represents a homozygote. Note that with λ being defined as the ratio of the higher rate to the lower one, we always choose the reference allele to have (k, q, α) and set $\lambda \geq 1$ without losing generality. To investigate the impact of the individual processes on the genetic noise profile [6,23,24], we employ three types of heterozygotes, k -, q -, and α -heterozygote, which we denote as Het_k , Het_q , and Het_α , respectively. Then, for a meaningful comparison of the noise profile between homo- and heterozygotes, we first determine the rates \bar{k} , \bar{q} , and $\bar{\alpha}$ of the mean-expression-preserving homozygote. See Fig. 2 and Eq. (C1) in Appendix C.

IV. PROMOTER NOISE UNDER NONEQUILIBRIUM BINDING DYNAMICS

For a single allele in the limit of fast promoter dynamics, i.e., for $\omega \gg 1$ with a physiologically sound assumption that $\delta \ll 1$ and a finite ratio $q/k \equiv K$, the contribution of the promoter switching, contained in the g function in Eq. (4), is only $O(K/\omega)$ to both the mRNA- and protein-level noise, and thus becomes negligible. In this limit, the slow birth-death dynamics of the gene products is shielded from the promoter noise, and the overall gene expression is accurately

reproduced by a Poisson point process with the rate parameter $\alpha \cdot (1 + K)^{-1} = \theta \alpha$.

However, as shown in Fig. 2(b), when either of the rates (k, q) decrease down to the mRNA degradation rate γ , η_1 starts to dominate. This tendency is prominent near the main diagonal characterized by a 50% duty cycle of the promoter ($K = 1$), but weakens when either of the rates k or q exceeds γ . Notice that, even with a constant duty cycle, η_1/η_0 changes its value by more than three orders of magnitude. In general, as the duration of the ON or OFF state of the promoter becomes longer, the noise propagated from the ON-OFF switching becomes more significant.

Since, at the mean protein level, a c -fold change in k is indistinguishable from a reciprocal change in q , and since $g(ck, q)$ from Eq. (3) is always less than $g(k, q/c)$ for $c > 1$, the unilateral increase in k is the solution to increase the expression level by adjusting k or q with minimal noise. Likewise, to decrease the expression, a unilateral increase in q is the solution for $c < 1$. Moreover, tuning α entails even higher noise as is clearly seen in Figs. 3(c), 3(f), and 3(i), where the dotted lines fall deep below the corresponding solid lines at a given mean.

The biallelic gene expression as the sum of the contributions from each independent allele shows nontrivial noise behavior. As the binding affinity of the reference allele enhances when K shifts from 10 to 1 to 0.1 [Figs. 3(b), 3(e), and 3(h)], the noise level shows an overall trend of decrease. However, at higher duty cycle, due to the earlier onset of the saturation in the mean expression level with growing discrepancy [Figs. 3(d) and 3(g)], the noise level rebounds, rendering it a nonmonotonic function of K . The increased duty cycle in the reference allele from 9% ($K = 10$) to 50% ($K = 1$) to 91% ($K = 0.1$) results in $\sim 1500\%$ and $\sim 30\%$ relative increase in the total expression for a hundred-fold discrepancy in K , leaving little room for improvement in noise reduction at higher duty cycles.

V. HETEROZYGOSITY IN PRODUCTIVITY AND RELIABILITY

From Eq. (4), the reduced noise ν of a heterozygous gene with respect to its mean-expression-preserving homozygote can be expressed as

$$\nu \equiv \frac{\eta_{\text{het}} - \eta_{\text{hom}}}{\eta_{\text{hom}}} = \frac{\sum_{i=1}^2 \langle p_i \rangle^2 g(k_i, q_i) - \frac{1}{2} \langle p \rangle^2 \bar{g}}{\frac{\beta(1+b^{-1})\langle p \rangle}{\delta} + \frac{1}{2} \langle p \rangle^2 \bar{g}}, \quad (5)$$

where \bar{g} is either $g(\bar{k}, \bar{q})$ or $g(k, \bar{q})$ depending on whether the heterozygote in question is Het_k or Het_q . Notice that, from Eq. (2a), $\langle p_i \rangle$ contains the allele-specific variation of both α_i and K_i .

Specifically, when the interallelic heterogeneity is introduced into the productivity measure α_i with homozygous (k_i, q_i) , the numerator in Eq. (5), up to the positive common factor \bar{g} , becomes $\frac{1}{2}(\langle p_1 \rangle - \langle p_2 \rangle)^2 \propto (1 - \lambda)^2 > 0$, showing that the heterozygous gene expression is noisier than the homozygote's. This is a conspicuous feature of Het_α as shown by the solid lines in Figs. 4(c), 4(f), and 4(i). The reduced noise ν for Het_α monotonically grows with the interallelic discrepancy in the transcriptional processing rate α .

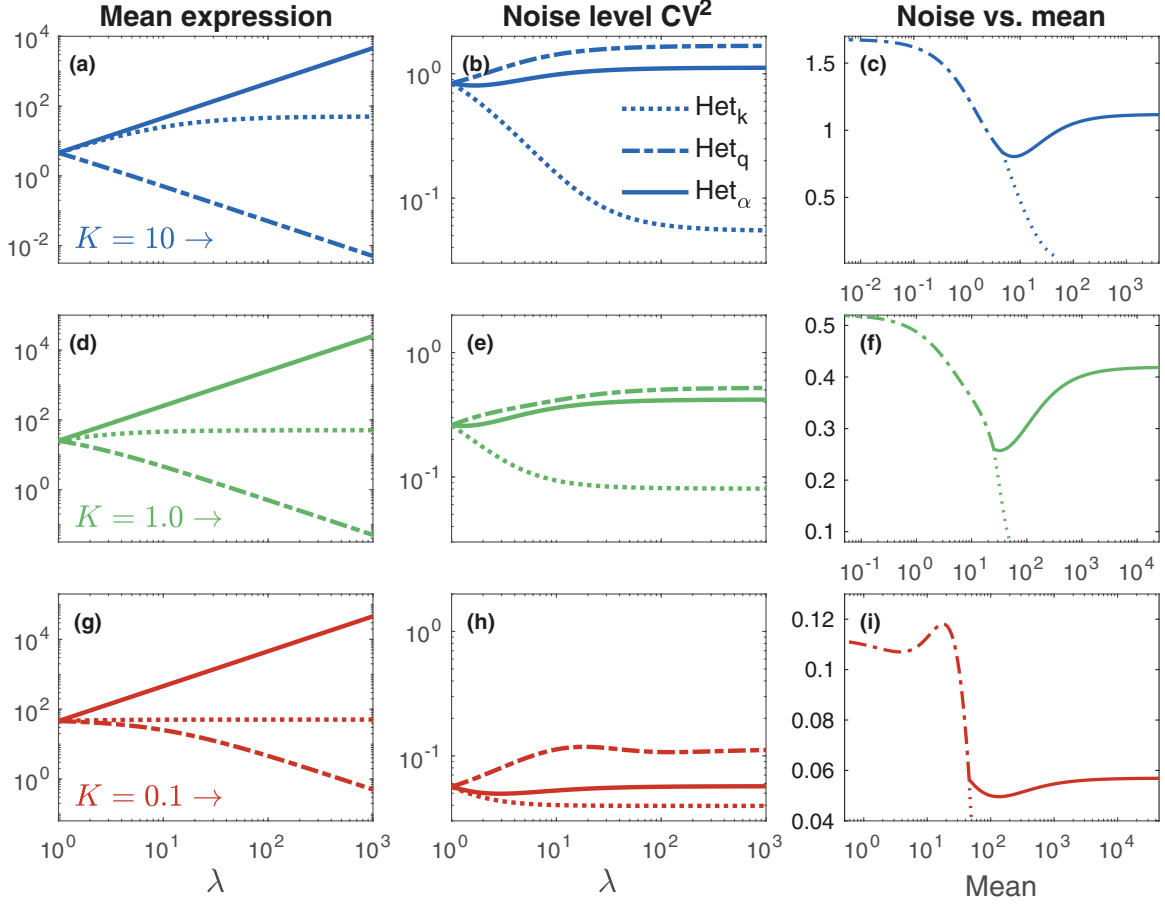


FIG. 3. Mean and noise level of biallelic gene expression. The first two columns are the mean expression and noise level as a function of discrepancy parameter λ , and the last column is the resultant 2D map between the mean and the noise. The rows are arranged according to the values of the equilibrium constant: $K = 10$ (blue, a–c), $K = 1$ (green, d–f), and $K = 0.1$ (red, g–i). The dotted, dash-dotted, and solid lines represent Het_k , Het_q , and Het_α , respectively. The intersection point of the three distinct lines in each of the mean–noise relation denotes the mean and noise level of the homozygote at each equilibrium constant K . All the zygosity-independent rates are chosen the same as in Fig. 2.

However, when the heterozygosity manifests as a difference in the reliability rather than the productivity of the gene expression machine, the rates affected are k_i and/or q_i . In the fast switching limit, i.e., $k_i \gg 1$ and $q_i \gg 1$,

$$\langle p_i \rangle^2 g(k_i, q_i) \sim \frac{k_i q_i}{(k_i + q_i)^3}. \quad (6)$$

Hence, from Eq. (5), for Het_k ,

$$v(\lambda) \sim \frac{kq}{(k+q)^3} \left[1 + \lambda \left(\frac{1+K}{\lambda+K} \right)^3 - 2r_k \left(\frac{1+K}{r_k+K} \right)^3 \right], \quad (7)$$

where $r_k \equiv \bar{k}/k$. The same form of approximation is obtained for Het_q with the substitutions $k \leftrightarrow q$ and $K \leftrightarrow K^{-1}$. It follows that, as long as $K > 2$ for Het_k and $K < 1/2$ for Het_q , v remains negative, i.e., Het_k and Het_q are less noisy than their mean-preserving homozygotes regardless of λ . This is consistent with the results in Figs. 4(a) and 4(h), where the Het_k 's noise buffering capacity is clearly enhanced in the weak-promoter limit, while Het_q is beneficial as the binding affinity becomes stronger. Even with moderate values of k and q —actually the k value of the reference allele here is set to 0.1, the noise reduction reaches as high as 40% (Het_k) to 60%

(Het_q) for the reference allele with not too “balanced” duty cycle ($K \sim 1$).

In the large-discrepancy limit ($\lambda \gg 1$), Het_k with a weak reference promoter [blue solid line in Fig. 4(a)] and Het_q with a strong promoter [red solid line in Fig. 4(h)] corresponds to MAE, where both the mean and noise levels are contributed to almost exclusively by a single allele, i.e., the reference (non-reference) allele for the former (latter) case. In this regime, the heterozygotes show lower noise level than their mean-preserving homozygotes, which is consistent with a recent report [14] on the noise reduction as a potential rationale of MAE.

VI. DISCUSSION

The phenotypic diversity stemming from stochastic gene expression and its evolutionary implications have been the target of recent research [25–28]. Matsumoto *et al.* [29] investigated an individual-based diploid model to assess the impact of stochastic gene expression in shaping the gene pool through the interplay among the selection strength, magnitude of genetic noise, and frequency of environmental changes. Here, as an effort to understand a seemingly counterintuitive

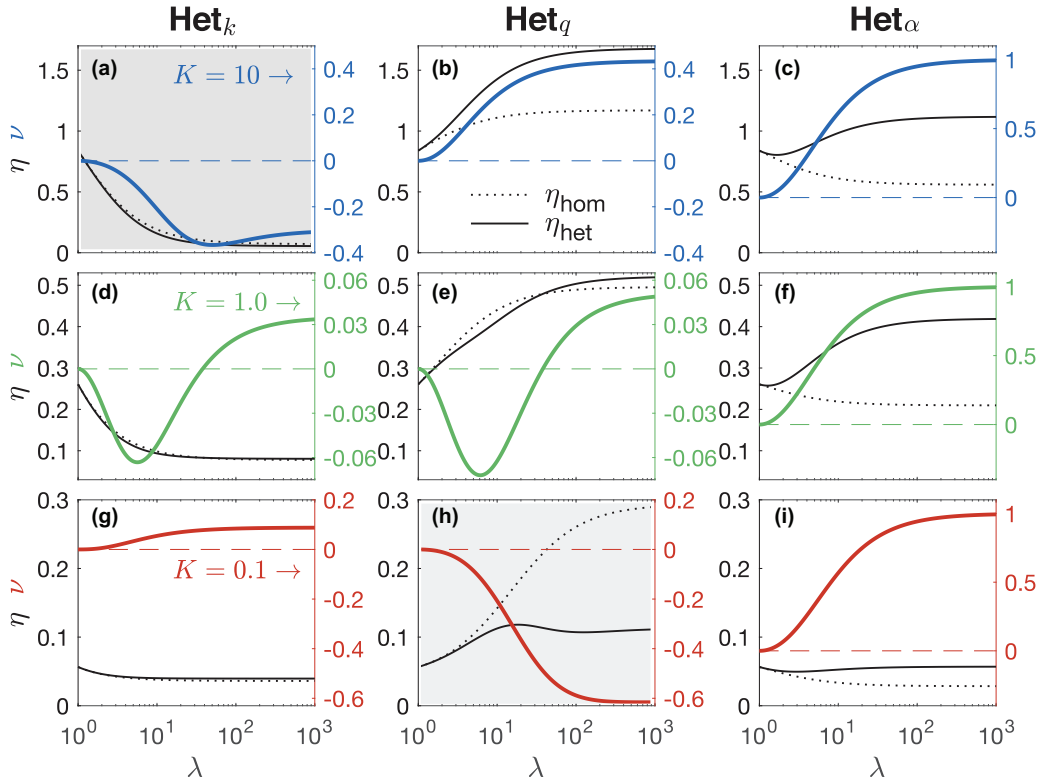


FIG. 4. The noise and reduced noise of the heterozygotes in comparison with their mean-preserving homozygotes. The solid (dotted) black lines represent the heterozygotes (homozygotes) and should refer to the y-axis on the left. The solid lines in color represent the reduced noise $(\eta_{\text{het}} - \eta_{\text{hom}})/\eta_{\text{hom}}$ and refer to the y-axis on the right. The rows are arranged according to the equilibrium constant of the reference allele: $K = 10$ (a)–(c), $K = 1$ (d)–(f), and $K = 0.1$ (g)–(i). As clearly seen in the shaded subfigures (a) and (h), the Het_k 's noise buffering capacity is prominent in the low duty-cycle regime (a), while Het_q plays the same role in the opposite regime of the duty-cycle (h).

role of regulatory heterozygote as a noise buffer, we differentiate the “productivity” of the gene expression machinery from its “reliability.” Unlike the productivity measure α_i for the processive events that are hardly an engineering target for obvious physiological reasons, the reliability measures (k_i, q_i) controlling the intermittency of the promoter activity provide a novel heterozygote advantage in diploid genome through its noise reduction capacity.

The analytical origin of this effect could be attributed to the convexity of the noise with respect to k_i and q_i . From Eq. (4) and the discussion in Appendix E, we see that $\partial^2\eta/\partial k_i^2 < 0$ and $\partial^2\eta/\partial q_i^2 < 0$ for certain conditions while $\partial^2\eta/\partial \alpha^2$ is always positive. That is, under the modulation of k_i and/or q_i , the birth-death process gated by a telegraph noise develops concave regions, where the opposite of Jensen’s inequality holds. This observation is in line with the recent study [15], where the heterozygosity mitigates the gene expression noise under the *trans*-acting noise due to the fluctuating transcription factor availability. In the context of the balancing selection emerging in the *cis*-regulatory regions of the human genome, our results hint at yet another type of heterozygote advantage. That is, the heterozygosity may arise as a selectable trait in the sequence evolution through its effects on noise reduction.

The results presented in this article also shed a new light on the double-sided roles of monoallelic expression as an

extreme of Het_q with a high degree of interallelic discrepancy ($\lambda \gg 1$). The intuitive rationale of MAE is to enhance the cellular “mosaicism” by differential expression of structural alleles [14,30,31]. At the same time, a gene ontology-based annotation reveals that the MAE pattern is found at disproportionately high frequencies in the cell fate decisions such as cell division, quiescence, and differentiation [13,14], suggesting a means to tighten the control of gene expression. Our analytical and numerical results show that, compared with the allele-independent regulations mediated by transcription factors or microRNA molecules, the allele-specific modification of the gene dosage provides a more controlled way to buffer the expression noise— Het_q at high duty cycles ($K \ll 1$) and Het_k at low duty cycles ($K \gg 1$) [11]. Experimental studies on the genetic and epigenetic control of k and q , instead of their ratio K , will be an important step toward fruitful ramifications in biological and engineering domains.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea grants funded by the Korea government (MSIT): Grants No. NRF-2018K1A4A3A01063890 (Global R&D Center Program) and No. NRF-2020R1F1A1075942. We gratefully acknowledge the UNIST Supercomputing Center for the support of computing resources.

APPENDIX A: CHEMICAL MASTER EQUATION AND ITS EXACT SOLUTION

The matrix V appearing in the CME, Eq. (1), writes

$$V = \begin{pmatrix} k_1 + k_2 & -q_1 & -q_2 & 0 \\ -k_1 & k_2 + q_1 & 0 & -q_2 \\ -k_2 & 0 & k_1 + q_2 & -q_1 \\ 0 & -k_2 & -k_1 & q_1 + q_2 \end{pmatrix}, \quad (\text{A1})$$

where the subindices 1 and 2 denote the allele identity. The matrix W capturing the birth and death process is diagonal in the promoter-state representation, i.e.,

$$w_{ss'} = \delta_{ss'} [n_s \alpha (1 - \hat{L}_m^-) + \gamma (1 - \hat{L}_m^+) m + \beta (1 - \hat{L}_p^-) + \delta (1 - \hat{L}_p^+) p], \quad (\text{A2})$$

where $\delta_{ss'} = 1$ only if $s = s'$ and $\delta_{ss'} = 0$ otherwise. Note that $n_s \in \{0, 1, 2\}$ is the number of active alleles in state s . The birth and death operators are defined as $\hat{L}_m^\pm f(m, p) = f(m \pm 1, p)$ and $\hat{L}_p^\pm f(m, p) = f(m, p \pm 1)$ for an arbitrary function $f(m, p)$ defined on the domain of nonnegative integers (m, p) .

To work around the difficulty of dealing with discrete variables, (m, p) , we introduce the probability generating function

$$\Pi_s(x, y; t) \equiv \sum_{p=0}^{\infty} \sum_{m=0}^{\infty} x^m y^p P_s(m, p; t), \quad (\text{A3})$$

$$\Pr(n|\Theta) = \frac{\Gamma(\frac{\omega}{\delta}) \Gamma(n + c_+) \Gamma(n + c_-)}{\Gamma(c_+) \Gamma(c_-) \Gamma(n + \frac{\omega}{\delta}) \Gamma(n + 1)} \left(\frac{1}{1 + \beta} \right)^{c_+} \left(\frac{\beta}{1 + \beta} \right)^n {}_2F_1 \left(n + c_+, \frac{\omega}{\delta} - c_-, n + \frac{\omega}{\delta}; \frac{\beta}{1 + \beta} \right), \quad (\text{A6})$$

where ${}_2F_1(\cdot)$ is a Gaussian hypergeometric function and

$$c_{\pm} = \frac{\alpha + \omega}{2\delta} \left(1 \pm \sqrt{1 - \frac{4k\alpha}{(\alpha + \omega)^2}} \right).$$

Thus, obtained distributions are shown in Fig. 5.

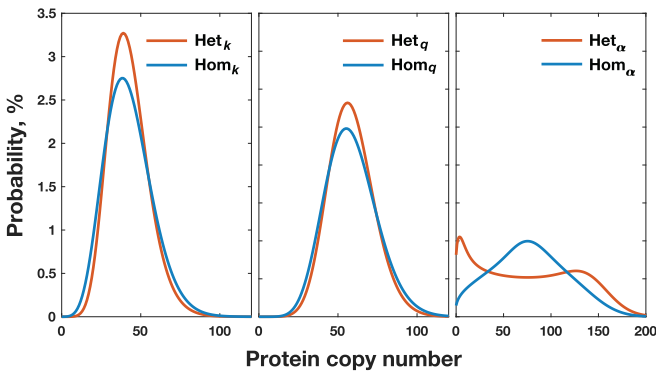


FIG. 5. The exact copy number distribution of protein. Here the discrepancy parameter is set to $\lambda = 30$ and the rate parameters for the reference allele are chosen as $\beta = 1.8$, $\delta = 0.15$, and $k = 0.1$. The promoter deactivation rate is chosen as $q = 1.0, 0.01$, and 0.1 , representing the cases of Figs. 4(a), 4(h), and 4(f). The reduced noise ν defined in Eq. (5) is $-0.33, -0.45$, and 0.82 from left to right.

the equivalent CME with respect to the continuous variables, (x, y) , becomes

$$\frac{\partial}{\partial t} \Pi_s(x, y; t) = - \sum_{s'} (v_{ss'} + \tilde{w}_{ss'}) \Pi_{s'}(x, y; t), \quad (\text{A4})$$

where

$$\tilde{w}_{ss'} = \delta_{ss'} [(1 - x)(n_s \alpha - \gamma \partial_x) + (1 - y)(\beta x \partial_x - \delta \partial_y)]. \quad (\text{A5})$$

The steady-state solution of Eq. (A4) can be obtained analytically by the method of characteristics [32]. In practice, once the steady-state copy number distribution, $\Pr(n|\Theta)$, for a haploid gene with a given set of rate parameters $\Theta \equiv \{k, q, \alpha, \beta, \gamma, \delta\}$ is given, the distribution for the diploid gene can be expressed as their convolution, i.e.,

$$\Pr(n|\Theta, \Theta') = \sum_{m=0}^n \Pr(m|\Theta) \Pr(n - m|\Theta').$$

And the haploid distribution is given by [9]

APPENDIX B: EXACT FORM OF MOMENTS

Thanks to the linearity of the CME, the exact form of the first and second moment of the steady-state probability distribution $\sum_s P_s(m, p, \infty)$ can be obtained by the method of probability generating function [22]. Thus, obtained mean and variance reads

$$\langle m \rangle = \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i}, \quad (\text{B1a})$$

$$\langle p \rangle = \frac{\beta}{\delta} \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i}, \quad (\text{B1b})$$

$$\begin{aligned} \sigma_m^2 &= \langle m \rangle + \sum_{i=1}^2 \langle m_i \rangle^2 \frac{K_i}{1 + \omega_i} \\ &= \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i} \left(1 + \frac{\alpha_i}{1 + K_i} \frac{K_i}{1 + \omega_i} \right), \end{aligned} \quad (\text{B1c})$$

$$\begin{aligned} \sigma_p^2 &= \langle p \rangle (1 + b) + \frac{\delta}{1 + \delta} \sum_{i=1}^2 \langle p_i \rangle^2 \frac{K_i}{1 + \omega_i} \frac{1 + \delta + \omega_i}{\delta + \omega_i} \\ &= \frac{\beta}{\delta} \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i} \left[1 + \frac{\beta}{1 + \delta} \left(1 + \frac{\alpha_i}{1 + K_i} g_i \right) \right], \end{aligned} \quad (\text{B1d})$$

where the parameters are defined the same as in the main text and g_i is short for $g(k_i, q_i)$ from Eq. (3). Note that all the rates

are rescaled in units of γ . The stationary cross-correlation between the mRNA and protein copy numbers captures how tightly the level of the two gene products are correlated and provides a clue to the gene expression dynamics responsible for the seeming absence of correlation [5]. From the expression

$$\text{Cov}(m, p) = \frac{\beta}{1 + \delta} \sum_{i=1}^2 \frac{\alpha_i}{1 + K_i} \left(1 + \frac{\alpha_i}{1 + K_i} g_i \right), \quad (\text{B2})$$

the normalized cross-correlation can be computed as

$$\frac{\text{Cov}(m, p)}{\sigma_m \sigma_p} = \sqrt{\frac{\beta \delta}{(1 + \delta)(1 + \delta + \beta)}} \chi(\bar{k}, \bar{q}, \bar{\alpha}), \quad (\text{B3})$$

where the vector notation is used to indicate the collective contribution from both of the alleles and $\chi(\bar{k}, \bar{q}, \bar{\alpha})$ approaches its minimum value of 1 in the limit of the constitutive promoter, $\bar{q} \rightarrow 0$.

From the exact results obtained above, the protein noise for an arbitrary pair of alleles is now given by Eq. (4). In particular, for a homozygote with $k_1 = k_2 = \bar{k}$ and $q_1 = q_2 = \bar{q}$, and thus $\langle m_i \rangle = \langle m \rangle / 2$ and $\langle p_i \rangle = \langle p \rangle / 2$, the noise level becomes

$$\eta_{\text{hom}} = \frac{1 + b}{\langle p \rangle} + \frac{1}{2} \frac{\delta}{1 + \delta} g(\bar{k}, \bar{q}). \quad (\text{B4})$$

This constitutes the reference level of noise to be compared with the heterozygotes that have the same mean level of gene expression.

APPENDIX C: CALCULATION OF THE AVERAGE RATE VALUES OF THE MEAN-EXPRESSION-PRESERVING HOMOZYGOTE

In computing the differential noise between hetero- and homozygotes for a given mean expression level, we require, from Eqs. (B1a) and (B1b), the following conditions:

$$\begin{aligned} (k, \lambda k) \text{ vs. } (\bar{k}, \bar{k}) &: \frac{1}{1 + K} + \frac{1}{1 + K/\lambda} = \frac{2}{1 + K\bar{k}/k} \\ \rightarrow \frac{\bar{k}}{k} &= a_\lambda \cdot \frac{K + h_\lambda}{K + a_\lambda} \leq a_\lambda, \end{aligned} \quad (\text{C1a})$$

$$\begin{aligned} (q, \lambda q) \text{ vs. } (\bar{q}, \bar{q}) &: \frac{1}{1 + K} + \frac{1}{1 + \lambda K} = \frac{2}{1 + K\bar{q}/q} \\ \rightarrow \frac{\bar{q}}{q} &= a_\lambda \cdot \frac{K^{-1} + h_\lambda}{K^{-1} + a_\lambda} \leq a_\lambda, \end{aligned} \quad (\text{C1b})$$

$$\begin{aligned} (\alpha, \lambda \alpha) \text{ vs. } (\bar{\alpha}, \bar{\alpha}) &: \frac{\alpha}{1 + K} + \frac{\lambda \alpha}{1 + K} = \frac{2\bar{\alpha}}{1 + K} \\ \rightarrow \frac{\bar{\alpha}}{\alpha} &= a_\lambda, \end{aligned} \quad (\text{C1c})$$

where $a_\lambda = (1 + \lambda)/2$ and $h_\lambda = 2\lambda/(1 + \lambda)$, respectively, denotes the arithmetic and harmonic mean of 1 and λ . With this preparation, the reduced noise of the heterozygote with reference to the mean-preserving homozygote is

$$\sigma_{p,\text{het}}^2 - \sigma_{p,\text{hom}}^2 = \frac{\delta}{1 + \delta} \left[\sum_{i=1}^2 \langle p_i \rangle^2 g(k_i, q_i) - \frac{1}{2} \langle p \rangle^2 \bar{g} \right]. \quad (\text{C2})$$

Here the terms $\langle p_i \rangle$, $\langle p \rangle$, and \bar{g} contain λ dependence, and the variables with an overbar can be substituted by using Eq. (C1). Hence, the reduced noise level is given by Eq. (5).

APPENDIX D: NOISE IN FAST SWITCHING LIMIT

In the fast switching limit, i.e., $k_i \gg 1$ and $q_i \gg 1$ with a physiological requirement $\delta \ll 1$ (upper right section of Fig. 2 in the main text, where the noise profile produces contours near parallel to the diagonal, reflecting the accuracy of ‘‘adiabatic’’ approximation),

$$\begin{aligned} \langle p_i \rangle^2 g(k_i, q_i) &= \left(\frac{\alpha_i \beta}{\delta} \right)^2 \frac{K_i}{(1 + K_i)^2} \frac{1 + \delta + \omega_i}{(\lambda + \omega_i)(\beta + \omega_i)} \\ &\sim \frac{1}{k_i} \frac{K_i}{(1 + K_i)^3} = \frac{k_i q_i}{(k_i + q_i)^3}. \end{aligned} \quad (\text{D1})$$

Hence, from Eq. (5),

$$\nu_{\text{Het}_k} \propto \frac{kq}{(k + q)^3} \left[1 + \lambda \left(\frac{1 + K}{\lambda + K} \right)^3 - 2r_k \left(\frac{1 + K}{r_k + K} \right)^3 \right], \quad (\text{D2a})$$

$$\nu_{\text{Het}_q} \propto \frac{kq}{(k + q)^3} \left[1 + \lambda \left(\frac{1 + K^{-1}}{\lambda + K^{-1}} \right)^3 - 2r_q \left(\frac{1 + K^{-1}}{r_q + K^{-1}} \right)^3 \right], \quad (\text{D2b})$$

where $r_k \equiv \bar{k}/k$, and $r_q \equiv \bar{q}/q$ whose explicit form is given by Eqs. ((C1a),(C1b)). If we define the function $f_K(x) \equiv x \left(\frac{1+K}{x+K} \right)^3$, then

$$\nu_{\text{Het}_k}(\lambda) \propto \frac{1}{2} [f_K(1) + f_K(\lambda)] - f_K(r_k), \quad (\text{D3a})$$

$$\nu_{\text{Het}_q}(\lambda) \propto \frac{1}{2} [f_{K^{-1}}(1) + f_{K^{-1}}(\lambda)] - f_{K^{-1}}(r_q). \quad (\text{D3b})$$

We find that ν is negative, i.e., the heterozygote is less noisy than the mean-preserving homozygote regardless of λ for $K > 2$ (Het_k) and for $K < 1/2$ (Het_q). This is consistent with the results presented in Fig. 4 of the main text, where the Het_k's noise buffering capacity is clearly enhanced in the weak-promoter limit, whereas Het_q exhibits noise reduction as the duty cycle becomes higher.

APPENDIX E: CONVEXITY OR CONCAVITY OF THE NOISE PROFILE

The observations presented here clearly show a possible reduction of the expression noise from a gene that harbors regulatory heterozygosity, i.e., when the same transcript is produced by two alleles that only differ by their promoter activity. This result seems to defeat our naive intuition that the ‘‘mixture’’ of two alleles with different statistical behaviors should produce more noise simply because of the statistical heterogeneity. The analytical origin of this effect can be attributed to the concavity of the noise with respect to k_i and q_i . Notice that the overall protein noise can be expressed as a linear combination of the allelic

contributions:

$$\begin{aligned} \eta_p &= \frac{\sigma_1^2 + \sigma_2^2}{(\langle p_1 \rangle + \langle p_2 \rangle)^2} \\ &= \left(\frac{\alpha_1 \theta_1}{\alpha_1 \theta_1 + \alpha_2 \theta_2} \right)^2 \eta_{p_1} + \left(\frac{\alpha_2 \theta_2}{\alpha_1 \theta_1 + \alpha_2 \theta_2} \right)^2 \eta_{p_2}, \end{aligned} \quad (E1)$$

where $\theta_i = 1/(1 + K_i)$ is the duty cycle of each allele as defined in the main text. From Eq. (4), since $\alpha^2 \theta^2 \eta(k, q)$ yields a quadratic function of α with a positive quadratic coefficient $g(k, q)/(1 + \delta^{-1})$, it is trivial to see the convexity of the overall noise with respect to α . Thus, the heterozygosity in productivity measure leads to noisier expression.

To check the convexity of the function with respect to k and q , we consider

$$\begin{aligned} \theta^2 \eta(k, q) &= A(1 + K)^{-1} + Bg(k, q)(1 + K)^{-2} \\ &= A\theta + B\theta(1 - \theta) \underbrace{\frac{1 + \delta + \omega}{(1 + \omega)(\delta + \omega)}}_{\equiv G(\omega)} \equiv F(\theta, \omega), \end{aligned} \quad (E2)$$

where $A = \delta(1 + b)/\alpha\beta$ and $B = \delta/(1 + \delta)$. Using the shorthand notation $\partial/\partial\Box \equiv \partial_\Box$, the chain rule reads:

$$\partial_k = \frac{1 - \theta}{\omega} \partial_\theta + \partial_\omega, \quad (E3a)$$

$$\partial_q = -\frac{\theta}{\omega} \partial_\theta + \partial_\omega, \quad (E3b)$$

$$\partial_k^2 = \frac{1 - \theta}{\omega} \left(\frac{1 - \theta}{\omega} \partial_\theta^2 + 2\partial_\theta \partial_\omega - \frac{2}{\omega} \partial_\theta \right) + \partial_\omega^2, \quad (E3c)$$

$$\partial_q^2 = \frac{\theta}{\omega} \left(\frac{\theta}{\omega} \partial_\theta^2 - 2\partial_\theta \partial_\omega + \frac{2}{\omega} \partial_\theta \right) + \partial_\omega^2. \quad (E3d)$$

Thus, the second-order derivatives are rearranged as

$$\begin{aligned} \partial_k^2 F &= -\frac{2(1 - \theta)}{\omega^2} [A + B(2 - 3\theta)G(\omega) + B(1 - 2\theta)\omega G'(\omega)] \\ &\quad + B(1 - 2\theta)G''(\omega), \end{aligned} \quad (E4a)$$

$$\begin{aligned} \partial_q^2 F &= \frac{2\theta}{\omega^2} [A + B(1 - 3\theta)G(\omega) - B(1 - 2\theta)\omega G'(\omega)] \\ &\quad + B(1 - 2\theta)G''(\omega), \end{aligned} \quad (E4b)$$

which can assume negative values. That is, under the modulation of k_i and/or q_i , the birth-death process gated by a telegraph noise develops concave regions, where the opposite of Jensen's inequality holds.

-
- [1] C. P. Martinez-Jimenez, N. Eling, H.-C. Chen, C. A. Vallejos, A. A. Kolodziejczyk, F. Connor, L. Stojic, T. F. Rayner, M. J. Stubbington, S. A. Teichmann *et al.*, Aging increases cell-to-cell transcriptional variability upon immune stimulation, *Science* **355**, 1433 (2017).
 - [2] T. Hagai, X. Chen, R. J. Miragaia, R. Rostom, T. Gomes, N. Kunowska, J. Henriksson, J.-E. Park, V. Proserpio, G. Donati *et al.*, Gene expression variability across cells and species shapes innate immunity, *Nature (London)* **563**, 197 (2018).
 - [3] Y. Baran, M. Subramaniam, A. Biton, T. Tukiainen, E. K. Tsang, M. A. Rivas, M. Pirinen, M. Gutierrez-Arcelus, K. S. Smith, K. R. Kukurba *et al.*, The landscape of genomic imprinting across diverse adult human tissues, *Genome Res.* **25**, 927 (2015).
 - [4] A. M. Hulse and J. J. Cai, Genetic variants contribute to gene expression variability in humans, *Genetics* **193**, 95 (2013).
 - [5] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells, *Science* **329**, 533 (2010).
 - [6] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, General properties of transcriptional time series in *Escherichia coli*, *Nat. Genet.* **43**, 554 (2011).
 - [7] J. M. Raser and E. K. O'shea, Noise in gene expression: origins, consequences, and control, *Science* **309**, 2010 (2005).
 - [8] B. Munsky, G. Neuert, and A. Van Oudenaarden, Using gene expression noise to understand gene regulation, *Science* **336**, 183 (2012).
 - [9] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. USA* **105**, 17256 (2008).
 - [10] B. Zoller, D. Nicolas, N. Molina, and F. Naef, Structure of silent transcription intervals and noise characteristics of mammalian genes, *Mol. Syst. Biol.* **11**, 823 (2015).
 - [11] Z. Cao and R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, *Proc. Natl. Acad. Sci. USA* **117**, 4682 (2020).
 - [12] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells, *Science* **343**, 193 (2014).
 - [13] M. A. Eckersley-Maslin, D. Thybert, J. H. Bergmann, J. C. Marioni, P. Flicek, and D. L. Spector, Random monoallelic gene expression increases upon embryonic stem cell differentiation, *Dev. Cell* **28**, 351 (2014).
 - [14] A.-V. Gendrel, M. Attia, C.-J. Chen, P. Diabangouaya, N. Servant, E. Barillot, and E. Heard, Developmental dynamics and disease potential of random monoallelic gene expression, *Dev. Cell* **28**, 366 (2014).
 - [15] M. K. Sung, J. Jang, K. S. Lee, C.-M. Ghim, and J. K. Choi, Selected heterozygosity at cis-regulatory sequences increases the expression homogeneity of a cell population in humans, *Genome Biol.* **17**, 164 (2016).
 - [16] H. Yang, J. Jang, and C.-M. Ghim, Effects of zygosity on stochastic gene expression of diploid cells, *J. Kor. Phys. Soc.* **74**, 312 (2019).
 - [17] D. L. Cook, A. N. Gerber, and S. J. Tapscott, Modeling stochastic gene expression: Implications for haploinsufficiency, *Proc. Natl. Acad. Sci. USA* **95**, 15641 (1998).
 - [18] B. Hat, P. Paszek, M. Kimmel, K. Piechor, and T. Lipniacki, How the number of alleles influences gene expression, *J. Stat. Phys.* **128**, 511 (2007).

- [19] Y. Jiang, N. R. Zhang, and M. Li, Scale: modeling allele-specific gene expression by single-cell RNA sequencing, *Genome Biol.* **18**, 74 (2017).
- [20] E. Roberts, A. Magis, J. O. Ortiz, W. Baumeister, and Z. Luthey-Schulten, Noise contributions in an inducible genetic switch: A whole-cell simulation study, *PLoS Comput. Biol.* **7**, e1002010 (2011).
- [21] M. Assaf, E. Roberts, and Z. Luthey-Schulten, Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise, *Phys. Rev. Lett.* **106**, 248102 (2011).
- [22] J. Paulsson, Models of stochastic gene expression, *Phys. Life Rev.* **2**, 157 (2005).
- [23] L. B. Carey, D. Van Dijk, P. M. Sloom, J. A. Kaandorp, and E. Segal, Promoter sequence determines the relationship between expression level and noise, *PLoS Biol.* **11**, e1001528 (2013).
- [24] A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq, Transcription factors modulate c-Fos transcriptional bursts, *Cell Rep.* **8**, 75 (2014).
- [25] K. Mineta, T. Matsumoto, N. Osada, and H. Araki, Population genetics of nongenetic traits: Evolutionary roles of stochasticity in gene expression, *Gene* **562**, 16 (2015).
- [26] Y. Ito, H. Toyota, K. Kaneko, and T. Yomo, How selection affects phenotypic fluctuation, *Mol. Sys. Biol.* **5**, 264 (2009).
- [27] K. Sato, Y. Ito, T. Yomo, and K. Kaneko, On the relation between fluctuation and response in biological systems, *Proc. Natl. Acad. Sci. USA* **100**, 14086 (2003).
- [28] Z. Wang and J. Zhang, Impact of gene expression noise on organismal fitness and the efficacy of natural selection, *Proc. Natl. Acad. Sci. USA* **108**, E67 (2011).
- [29] T. Matsumoto, K. Mineta, N. Osada, and H. Araki, An individual-based diploid model predicts limited conditions under which stochastic gene expression becomes advantageous, *Front. Genet.* **6**, 336 (2015).
- [30] S. Branciamore, Z. Valo, M. Li, J. Wang, A. D. Riggs, and J. Singer-Sam, Frequent monoallelic or skewed expression for developmental genes in cns-derived cells and evidence for balancing selection, *Proc. Natl. Acad. Sci. USA* **115**, E10379 (2018).
- [31] B. Reinius and R. Sandberg, Random monoallelic expression of autosomal genes: Stochastic transcription and allele-level regulation, *Nat. Rev. Genet.* **16**, 653 (2015).
- [32] W. A. Strauss, *Partial Differential Equations: An introduction*, 2nd ed., Chap. 3 (John Wiley & Sons, New York, 2007).