

Allele-specific single-cell RNA sequencing reveals different architectures of intrinsic and extrinsic gene expression noises

Mengyi Sun and Jianzhi Zhang*

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Received August 28, 2019; Revised October 19, 2019; Editorial Decision November 18, 2019; Accepted November 20, 2019

ABSTRACT

Gene expression noise refers to the variation of the expression level of a gene among isogenic cells in the same environment, and has two sources: extrinsic noise arising from the disparity of the cell state and intrinsic noise arising from the stochastic process of gene expression in the same cell state. Due to the low throughput of the existing method for measuring the two noise components, the architectures of intrinsic and extrinsic expression noises remain elusive. Using allele-specific single-cell RNA sequencing, we here estimate the two noise components of 3975 genes in mouse fibroblast cells. Our analyses verify predicted influences of several factors such as the TATA-box and microRNA targeting on intrinsic or extrinsic noises and reveal gene function-associated noise trends implicating the action of natural selection. These findings unravel differential regulations, optimizations, and biological consequences of intrinsic and extrinsic noises and can aid the construction of desired synthetic circuits.

INTRODUCTION

Gene expression noise refers to the variation in gene expression level among genetically identical cells in the same environment (1). Gene expression noise is often deleterious, because it leads to imprecise cellular behaviors. For example, it may ruin the stoichiometric relationship among functionally related proteins, which may further disrupt cellular homeostasis (2–7). However, under certain circumstances, gene expression noise can be beneficial. Prominent examples include bet-hedging strategies of microbes in fluctuating environments (8,9) and stochastic mechanisms for initiating cellular differentiation in multicellular organisms (10–12).

Gene expression noise has extrinsic and intrinsic components. The extrinsic noise arises from the among-cell variation in cell state such as the cell cycle stage or the concentra-

tions of various transcription factors (TFs), while the intrinsic noise is due to the stochastic process of gene expression even under a given cell state such as the stochastic binding of a promoter to RNA polymerase (13–15). Note that our definitions of intrinsic and extrinsic noises are based on the source of the noise. Under these definitions, both intrinsic and extrinsic noises can vary among genes. For instance, the intrinsic expression noise of a gene is predicted to be negatively correlated with the mean expression level of the gene (16), whereas the extrinsic noise can be different for genes belonging to different biological pathways (1). Dissecting gene expression noise into the two components provides insights into its mechanistic basis (17). Furthermore, the two noise components can have different biological consequences. For instance, genes regulating the cell cycle should ideally have high extrinsic noise but low intrinsic noise, because their expression levels should be variable among different cell states but stable under the same state. Dissecting the expression noise of a gene into intrinsic and extrinsic components requires a dual reporter assay typically performed in haploid cells by placing two copies of the same gene into the genome, each fused with a distinct reporter gene such as the yellow fluorescent protein (YFP) gene or cyan fluorescent protein (CFP) gene (18). This way, the intrinsic noise in protein concentration can be assessed by the difference between YFP and CFP concentrations within cells while the extrinsic noise can be measured by the covariation between YFP and CFP concentrations among cells. However, such experiments are laborious in strain construction and expression quantification, hindering the examination of many genes. Consequently, past genome-wide studies of gene expression noise measured only the total noise (19–22). Some authors attempted to focus on the intrinsic noise by limiting the analysis to cells of similar morphologies (20,21). But because the extrinsic noise is not completely eliminated in the above experiments, the estimated intrinsic noise is inaccurate. Furthermore, these experiments could not study the extrinsic noise. As a result, accurate knowledge about intrinsic and extrinsic noise is limited to only a few genes (18,23), and a general under-

*To whom correspondence should be addressed. Tel: +1 734 763 0527; Fax: +1 734 763 0544; Email: jianzhi@umich.edu

standing of the pattern, regulation, and evolution of these two noise components is lacking.

Here, we propose to use allele-specific single-cell RNA sequencing (scRNA-seq) to estimate the intrinsic and extrinsic expression noises at the mRNA level. When the two alleles of a gene are distinguished by their DNA sequences, the distinct sequences serve as dual reporters of mRNA concentrations in scRNA-seq. Our method is thus in principle similar to the classical dual reporter assay except that we study the intrinsic and extrinsic expression noises at the mRNA level whereas the classical assay studies them at the protein level. Because the protein noise is widely believed to arise primarily from the mRNA noise (16,24), findings about the latter will not only inform us the mRNA noise but also largely the protein noise. Because the dual reporters exist naturally at any heterozygous locus of the genotype investigated and because single-cell expression levels of all genes in the genome are measured simultaneously by scRNA-seq, our method can estimate the intrinsic and extrinsic expression noises at the genomic scale from one scRNA-seq experiment of a highly heterozygous genotype. Although scRNA-seq has been used to estimate mRNA expression noise (25–28), it has never been used to estimate the two noise components separately. Using publicly available allele-specific scRNA-seq data from mouse fibroblast cells (29), we estimate the intrinsic and extrinsic noises of 3975 genes, allowing depicting the architectures of the two noise components in mouse cells.

MATERIALS AND METHODS

Intrinsic and extrinsic noise in diploid cells

Let Y be the expression level of a gene in a cell and let X describe the cell state. Y is a random variable that is a function of the random variable X . Gene expression noise is commonly measured by noise strength $\eta_{tot}^2 = Var(Y)/E^2(Y)$, where Var stands for variance and E stands for expectation. According to the law of total variance, $\frac{Var(Y)}{E^2(Y)} = \frac{E(Var(Y|X))}{E^2(Y)} + \frac{Var(E(Y|X))}{E^2(Y)}$, where the first term on the right-hand side of the equation describes the variation of Y given X , or intrinsic noise strength η_{int}^2 , and the second term describes the variation of Y due to the variation of X , or extrinsic noise strength η_{ext}^2 .

Most past studies of intrinsic and extrinsic expression noises of a gene were conducted in haploid cells by placing two copies of the gene (under the control of two identical, independent promoters) in the genome, each carrying a unique marker. Let the expression levels of the two gene copies be Y_1 and Y_2 , respectively. It was found that the intrinsic noise of each gene copy can be expressed by $\eta_{int,H}^2 = \frac{E(Y_1 - Y_2)^2}{2E(Y_1)E(Y_2)}$ and the extrinsic noise of each gene copy can be expressed by $\eta_{ext,H}^2 = \frac{Cov(Y_1, Y_2)}{E(Y_1)E(Y_2)}$, where the subscript H indicates haploid and Cov indicates covariance (14).

Now let us consider a diploid cell in which the two alleles of the focal gene are controlled by two identical, independent promoters and have unique markers. We are interested in the noise of the total expression level of the two

alleles. Because the expression levels of the two alleles are independent given the cell state, by definition, the intrinsic expression noise strength in diploid cells is $\eta_{int,D}^2 = \frac{E(Var((Y_1 + Y_2)|X))}{E^2(Y_1 + Y_2)} = \frac{E((Var(Y_1) + Var(Y_2))|X)}{4E^2(Y_1)} = \frac{2E(Var(Y)|X)}{4E^2(Y_1)} = \eta_{int,H}^2/2$. Similarly, by definition, the extrinsic expression noise strength in diploid cells is $\eta_{ext,D}^2 = \frac{Var(E(Y_1 + Y_2)|X)}{E^2(Y_1 + Y_2)} = \frac{Var(2E(Y)|X)}{4E^2(Y_1)} = \frac{Var(E(Y)|X)}{E^2(Y_1)} = \eta_{ext,H}^2$. Thus, we can adapt previously obtained formulas of intrinsic and extrinsic noise in haploid cells for the study of diploid cells.

Allele-specific single-cell RNA-seq data and data preprocessing

The raw read counts of allele-specific scRNA-seq data (29) were downloaded from <https://github.com/RickardSandberg/Reinius.et.al.Nature.Genetics.2016?files=1> (mouse.c57.counts.rds and mouse.cast.counts.rds). We preprocessed the dataset by requiring that (1) all cells have the same genotype and (2) there are spike-in standards in each cell. Two groups of cells satisfied our criteria: 60 cells from clone 7 and 75 cells from different clones or different individuals (IDs in the raw read-count dataset are 24–26, 28, 29, 31–35, 37–44, 46, 48–51, 53, 55, 58–60 and 124–170). Note that the latter group of cells are non-clonal and were isolated in different experiments; so they likely have larger variations in expression. Our analysis thus focused primarily on clone 7, although most results were also reproduced in the non-clonal cells. Because of the dual reporter design of our analysis, sex-linked genes were removed. For clone 7, we further removed genes on Chromosomes 3 and 4 due to aneuploidy. To ensure the relative reliability of our noise estimates, we limited the analysis to genes that have ≥ 5 reads mapped to each allele per cell averaged across cells. We then corrected the read counts mapped to each allele in each cell using spike-ins according to the following procedure. First, we obtained the number of reads mapped to spike-in molecules in each cell, yielding an array of 60 numbers, each specifying the number of reads mapped to spike-in molecules in one cell. Second, we divided each entry in the array by the largest number in the array, creating an array of 60 normalization factors that are all between 0 and 1. Third, we calibrated the number of reads mapped to each allele in each cell by dividing the original read number by the corresponding normalization factor in the array.

The noise decomposition requires the two reporters to have the same among-cell expression distribution. However, due to imprinting and polymorphisms in regulatory sequences, not all genes have the two identically regulated alleles. We thus performed a Kolmogorov–Smirnov test on the single-cell expression levels of the two alleles of each gene, and removed genes with $P < 0.05$ after multiple-testing correction (Benjamini–Hochberg correction). The data from the non-clonal cells were processed similarly. Some authors suggested normalizing single-cell expression levels of each reporter by its mean expression level to deal with unequal regulations between alleles (30,31). While this processing should allow analyzing more genes, the statistical properties

of the normalization are not well understood. To be conservative, we chose to remove genes that do not satisfy the assumption of the dual reporter experiment.

Estimation of intrinsic and extrinsic noises

We estimated the intrinsic and extrinsic expression noises of haploids using an existing program (30) and then converted them to the corresponding values in diploids using the formulas described above. We then derived noise estimates that are independent of both the mean expression level and the mean read number, the latter of which is inversely correlated with the amount of technical noise (32). Because the exact forms of the above dependencies are unknown, we used a rank-based measure. Specifically, we performed robust linear regression of the rank of intrinsic (or extrinsic) noise on the rank of expression level and the rank of read number using the 'rlm' function of the 'MASS' package with default options in R; the residual from the regression, D_{int} (or D_{ext}), is the measurement of intrinsic (or extrinsic) noise. To obtain the intrinsic noise estimate of a gene that is also independent of its extrinsic noise, we regressed the rank of intrinsic noise on the rank of mean expression level, the rank of mean read number, and the rank of extrinsic noise simultaneously. The obtained residual is referred to as D'_{int} . We similarly obtained D'_{ext} . The procedure used to process the data and estimate the two noise components is summarized in Supplementary Figure S1.

Assessment of technical extrinsic noise using spike-in molecules

We assessed the extrinsic technical noise using spike-in molecules from clone 7 and non-clonal cells. First, we estimated the mean read number of each spike-in species from the corrected read number of each spike-in molecule in each cell. The correction procedure was the same as used for correcting allele-specific reads mapped to each gene. Second, we ordered the spike-in molecules by their mean read numbers and paired neighboring spike-in molecules whose mean read numbers are similar. For each pair of spike-in molecules, we used binomial sampling to down-sample in each cell the raw reads of the spike-in molecule whose mean read number is larger, according to the ratio between the mean read numbers of the two spike-in molecules. Finally, each pair of spike-in molecules was treated as two alleles of the same spike-in transcripts for estimating extrinsic noise. As in the analysis of actual genes, we filtered out spike-in molecules whose mean (raw) read numbers are smaller than 5.

Factors influencing intrinsic and extrinsic noise

Mouse genes with a TATA-box were downloaded from the Eukaryotic Promoter Database (EPD) (33). Information of mouse miRNAs and their targets was downloaded from the RegNetwork database (34). Information about mouse *trans*-regulators and their target genes was also downloaded from RegNetwork (34). Note that miRNAs were considered *trans*-regulators in the database; so were they in our analysis. Some transcription factors target themselves. Because

the total noise of a gene by definition correlates with the intrinsic and extrinsic noises of the gene, we removed the self-targeting pairs in the analysis of *trans*-regulators. This problem does not involve miRNAs because we have no miRNA noise measures.

To test the hypothesis that genes targeted by the same *trans*-regulator tend to have similar D_{ext} , we grouped genes that share a *trans*-regulator and computed the standard deviation (SD) of their D_{ext} within the group. We then computed the median SD across all groups. Because SD is undefined for groups containing only one gene, such groups were discarded. We also removed *trans*-regulators that have noise measures and are target genes, such that the regulators and targets have no overlaps.

To analyze the relationship between histone modifications and expression noise, we downloaded the computed modification peak position data from Cistrome (35). We focused on four types of histone modifications in mouse wild-type fibroblast cells: H3K4Me1 (36), H3K4Me2 (36), H3K4Me3 (37) and H3K27AC (37). All four datasets used were of high quality and passed quality criteria of Cistrome. For each modification, we computed Spearman's correlation between the number of peaks overlapping core promoters, which are from 200 nucleotides upstream to 100 nucleotides downstream of transcription start sites (19), and D'_{int} or D'_{ext} .

Noise comparison among genes of different functions

GO terms of mouse genes were downloaded from Ensembl BioMart (GRC38m.p5) (38). Genes functioning in the mitochondrion are associated with the GO cellular component term of 'mitochondria', whereas cell cycle genes are associated with the GO biological process term of 'cell cycle'. Mouse protein complex data were downloaded from the CORUM database (<http://mips.helmholtz-muenchen.de/corum/>) (39).

To evaluate if a group of genes with a certain function (i.e., focal genes) are enriched/depleted with the TATA-box or miRNA targeting, we compared the group with other genes (i.e., non-focal genes) after controlling mean expression levels across 13 mouse tissues (40). Specifically, we ranked the focal genes by the mean expression level and divided them into 50 equal-size bins. We then obtained non-focal genes falling into each of these expression bins and identified the smallest number (m) of non-focal genes of all bins. We randomly picked m non-focal genes per bin and used this set of non-focal genes to compare with the focal genes. As expected, the non-focal genes showed similar expression levels as the corresponding focal genes ($P = 0.28$ for genes functioning in the mitochondrion, $P = 0.37$ for genes encoding protein complex members, and $P = 0.45$ for cell cycle genes; Mann–Whitney U test). The non-focal genes are referred to as the 'expression stratified control genes'.

DAVID GO web server with default options was used to perform the GO term enrichment analysis (41), in which all genes with estimated D_{int} and D_{ext} were used as the background. The web server returned the P -value after Benjamini–Hochberg correction for multiple testing. We ranked the GO terms by the significance level and reported

the three most significant GO terms for each group of genes with specific noise properties, if more than three GO terms were significantly enriched.

RESULTS

High-throughput estimation of intrinsic and extrinsic expression noises

The expression noise of a gene is commonly measured by the noise strength η^2 , which is the among-cell variance in expression level divided by the squared mean expression level. On the basis of previously derived formulas of intrinsic and extrinsic noises in haploids (14), we derived formulas for estimating intrinsic (η_{int}^2) and extrinsic (η_{ext}^2) noises in diploids (see Materials and Methods). Let the expression levels of the two alleles of a gene in a diploid cell be Y_1 and Y_2 , respectively. If the two alleles are controlled by two independent, identical promoters, $\eta_{int}^2 = \frac{E[(Y_1 - Y_2)^2]}{4E(Y_1)E(Y_2)}$ and $\eta_{ext}^2 = \frac{Cov(Y_1, Y_2)}{E(Y_1)E(Y_2)}$, where E and Cov respectively stand for expectation and covariance. Graphically, when the expression levels of the two alleles in each cell are respectively plotted on the x -axis and y -axis of a dot plot, extrinsic noise is represented by the spread of dots along the diagonal line of $y = x$, whereas the intrinsic noise is represented by the spread of dots along the direction perpendicular to the diagonal (left panel in Figure 1A). As an example, single-cell expression levels of the gene *Tcofl* are plotted (right panel in Figure 1A).

To estimate intrinsic and extrinsic gene expression noises, we used the scRNA-seq data of mouse fibroblast cells from an F1 hybrid of two mouse strains (29). Note that scRNA-seq data are subject to large technical noises, which may also be decomposed into intrinsic and extrinsic technical noises (32). The intrinsic technical noise is primarily caused by the low capturing efficiency of cellular transcripts and can result in a high variance and high dropout rate in estimating the mRNA expression level. The intrinsic technical noise artificially increases the level of the estimated intrinsic expression noise. The extrinsic technical noise is mainly due to tube-to-tube variability in capturing efficiency and artificially increases the level of the estimated extrinsic expression noise. Imputation, which substitutes the observed expression level of a gene in a cell by its expected expression level, is often used to deal with technical noises in scRNA-seq-based cell classification (42). But, imputation cannot be used in our study because it leads to underestimation of gene expression noise. Therefore, we only used spike-in control molecules to normalize expression levels in individual cells (see Materials and Methods).

Our analysis focused on clone 7 (derived from the hybrid of CAST/EiJ male \times C57BL/6J female) in the data, because (1) the number of sequenced cells ($n = 60$) is the largest in this clone, and (2) all sequenced cells from this clone have spike-in control molecules, permitting accurate read count estimation. Upon the removal of genes whose two alleles show significantly different among-cell expression distributions and other steps of data processing (Supplementary Figure S1; see Materials and Methods), we obtained the intrinsic and extrinsic expression noises of 3975

genes. To assess the precision of our noise estimates, we randomly separated the cells of clone 7 into two 30-cell groups. We found that the estimates of the intrinsic noise of a gene from the two subsamples are highly correlated (Pearson's $r = 0.79$, $P < 1 \times 10^{-300}$; Spearman's $\rho = 0.79$, $P < 1 \times 10^{-300}$; Figure 1B), while those of extrinsic noise are moderately correlated ($r = 0.42$, $P = 2.3 \times 10^{-151}$; $\rho = 0.44$, $P = 3.8 \times 10^{-185}$; Figure 1C). Note that the above correlations demonstrate the precision rather than the accuracy of our measurements. The accuracy of our measurements depends on technical noises, which can in principle be estimated using spike-in molecules, because they have no biological variation among cells. However, two factors render the technical noises of spike-in molecules not directly comparable with those of natural transcripts. First, spike-in molecules provide information of the technical noise in sample preparation steps after the addition of spike-in molecules, so the technical noises associated with earlier steps are unknown (42). Second, spike-in molecules have much lower capturing efficiencies (43) than natural transcripts. Nonetheless, it can be shown that, after normalization by spike-in molecules (see Materials and Methods), extrinsic noises disappear for spike-in molecules (red dots in Supplementary Figure S2), whereas extrinsic noises for natural transcripts remain substantial (black dots in Supplementary Figure S2), indicating that the tube-to-tube variation in sample preparation steps after the addition of spike-in molecules has been corrected. Because the magnitudes of technical noises cannot be estimated in our dataset and because the measurements of intrinsic and extrinsic noises are subject to different technical noises, it is not possible to directly compare the contributions of intrinsic noise and extrinsic noise to the total noise in the data analyzed. Nevertheless, with proper statistical processing, we can compare extrinsic or intrinsic noise among genes.

In addition to clone 7, there is another group of cells with $n = 75$ that fulfill the above two criteria (see Materials and Methods), but this group of cells are non-clonal and were isolated in different experiments, so may be more heterogeneous in cell state and subject to larger technical variabilities. Our analysis thus focused primarily on clone 7, although most results were also reproduced in the non-clonal cells. While the precision of the intrinsic noise estimates is similarly high in the non-clonal cells ($r = 0.80$, $P < 1 \times 10^{-300}$; $\rho = 0.79$, $P < 1 \times 10^{-300}$; Supplementary Figure S3A) when compared with that in the clonal cells (Figure 1B), the estimates of the extrinsic noise are much less precise in the non-clonal cells ($r = 0.31$, $P = 1.25 \times 10^{-102}$; $\rho = 0.24$, $P = 6.9 \times 10^{-65}$; Supplementary Figure S3B) than in the clonal cells (Figure 1C), probably for the aforementioned reasons. The assessment of technical noise in non-clonal cells (Supplementary Figure S3C) yielded similar results as in clone 7 cells (Supplementary Figure S2).

In theory, the intrinsic expression noise of a gene should decrease with the mean expression level of the gene (16,44), whereas no such relationship is expected for the extrinsic noise. We confirmed that our estimate of the intrinsic noise is indeed strongly negatively correlated with the mean expression level (Spearman's $\rho = -0.81$, $P < 1.0 \times 10^{-300}$; Figure 1D). A similar trend was observed from the non-clonal cells (Supplementary Figure S3D). Intriguingly, we

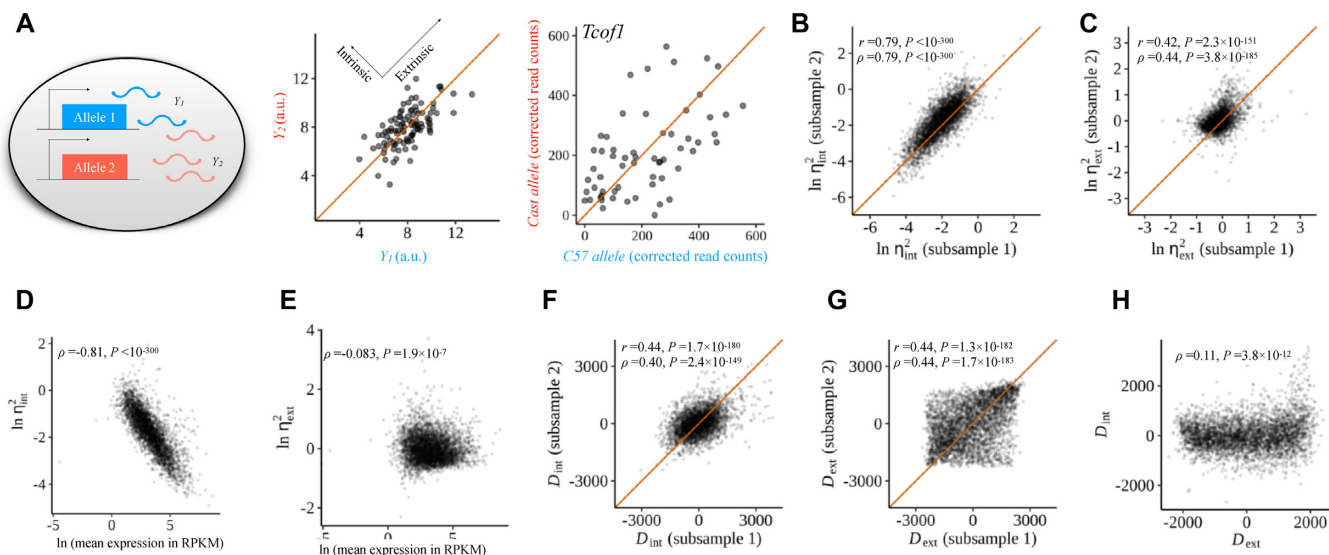


Figure 1. Decomposition of gene expression noise into intrinsic and extrinsic noise. (A) Gene expression noise can be decomposed to its intrinsic and extrinsic components by the dual reporter assay, where two reporters represented respectively by the blue and orange boxes are controlled by independent, identical promoters. When plotting the expression level of one reporter against that of the other in each cell, the spread along the diagonal represents extrinsic noise, whereas the spread orthogonal to the diagonal represents intrinsic noise. Y_1 and Y_2 are the expression levels of the two reporters, respectively. The left plot shows hypothetical data from a gene, whereas the right plot presents the spike-in adjusted read-counts of the two alleles of *Tcofl* from individual cells. (B) Intrinsic noises (η_{int}^2) estimated from two sub-samples of clone 7 are highly correlated with each other. Ln-transformed η_{int}^2 is shown. Each dot is a gene. The orange line shows the diagonal. (C) Extrinsic noises (η_{ext}^2) estimated from two sub-samples of clone 7 are moderately correlated with each other. Ln-transformed η_{ext}^2 is shown. Each dot is a gene. The orange line shows the diagonal. (D) The intrinsic expression noise of a gene is strongly negatively correlated with the mean expression level of the gene. Expression level is measured by Reads Per Kilobase of transcript per Million mapped reads (RPKM). (E) The extrinsic expression noise of a gene is weakly negatively correlated with the mean expression level of the gene. Because the extrinsic noise could be negative (see Materials and Methods), we added a small value, which is 0.1 minus the minimum extrinsic noise, to all η_{ext}^2 values before taking the natural log. (F) Intrinsic noise estimates adjusted for mean expression level and technical noise (D_{int}) are significantly correlated between two sub-samples of clone 7. The orange line shows the diagonal. (G) Extrinsic noise estimates adjusted for mean expression level and technical noise (D_{ext}) are significantly correlated between two sub-samples of clone 7. The orange line shows the diagonal. (H) D_{int} and D_{ext} are positively correlated.

also found a weak, but significant negative correlation between the extrinsic noise and mean expression level ($\rho = -0.083, P = 1.9 \times 10^{-7}$; Figure 1E). Because the extrinsic noise is the normalized covariance between Y_1 and Y_2 , and because the normalized covariance tends to be underestimated for lowly expressed genes due to larger sampling errors, the estimated extrinsic noise is expected to be positively correlated with the mean expression level for technical reasons. To assess the impact of the technical noise on extrinsic expression noise, we correlated across genes the extrinsic noise with the mean allele-specific read number, because the mean read number is not normalized by gene length so contains more information about the technical variation when compared with the mean expression level. Indeed, a positive correlation is observed between the estimated extrinsic noise and mean allele-specific read number instead of expression level ($\rho = 0.06, P = 3.4 \times 10^{-5}$). Thus, the observed negative correlation between extrinsic noise and expression level is likely biological. The trend observed in the non-clonal cells is similar to that in the clonal cells (Supplementary Figure S3E).

It is preferable to remove the correlation between a noise measure and the mean expression level in order to identify factors that impact intrinsic or extrinsic noise not simply due to their influences on the mean expression level. In addition, because technical noise in scRNA-seq decreases with mean read number (32), it would be important to fur-

ther remove the impact of the mean read number on our expression noise measures. To this end, we used robust linear regressions to remove the covariations with the mean expression level and mean read number in our measures of intrinsic and extrinsic noise (see Materials and Methods), which are referred to as D_{int} and D_{ext} , respectively. Note that D_{int} and D_{ext} are residuals in the regressions of expression noise ranks so have values potentially from -3975 to 3975 . We used ranks instead of raw noise estimates because (1) we do not know the exact relationship between the noises and the mean expression level or read number, (2) the expression noise estimates contain contributions from technical noises, and (3) rank statistics are robust to outliers. As expected, D_{int} is correlated with neither the mean expression level ($\rho = -0.003, P = 0.85$) nor the mean read number ($\rho = -0.004, P = 0.82$). Similarly, D_{ext} is correlated with neither the mean expression level ($\rho = -0.002, P = 0.89$) nor the mean read number ($\rho = -0.0005, P = 0.98$). To assess the precision of these new noise measures, we plotted the correlation between the estimates from two subsamples of clone 7 for D_{int} (Figure 1F) and D_{ext} (Figure 1G), respectively. We found the correlation of D_{int} from the two subsamples ($r = 0.44, P = 1.7 \times 10^{-180}$; $\rho = 0.40, P = 2.4 \times 10^{-149}$) similar to that of D_{ext} from the two subsamples ($r = 0.44, P = 1.3 \times 10^{-182}$; $\rho = 0.44, P = 1.7 \times 10^{-183}$). Because our subsequent statistical analyses of D_{int} and D_{ext} are all rank-based, the measurement precision of D_{int} and D_{ext} can be treated

as comparable. Compared with those in the clonal cells, the precision of D_{int} is similar ($r = 0.48$, $P = 6.1 \times 10^{-272}$; $\rho = 0.40$, $P = 1.3 \times 10^{-188}$; Supplementary Figure S3F) but that of D_{ext} is lower ($r = 0.24$, $P = 8.8 \times 10^{-66}$; $\rho = 0.23$, $P = 2.7 \times 10^{-64}$; Supplementary Figure S3G) in the non-clonal cells.

Interestingly, we observed a weak, but significant positive correlation between D_{int} and D_{ext} ($\rho = 0.11$, $P = 3.8 \times 10^{-12}$; Figure 1H). Similar results were obtained from the non-clonal cells ($\rho = 0.047$, $P = 0.0008$; Supplementary Figure S3G). Although previous theoretical studies predicted a dependency of intrinsic noise on extrinsic noise, the direction of the correlation was unpredicted (13,24,45). Because of this observed correlation, we further acquired an intrinsic noise estimate that is independent of the extrinsic noise by regressing the rank of intrinsic noise on the rank of mean expression level, the rank of mean read number, and the rank of extrinsic noise simultaneously. The obtained rank residual, referred to as D'_{int} , is correlated with none of the mean expression level ($\rho = -0.002$, $P = 0.88$), mean read number ($\rho = -0.002$, $P = 0.90$), and extrinsic noise ($\rho = -0.003$, $P = 0.85$). We similarly obtained D'_{ext} , which is correlated with none of the mean expression level ($\rho = -0.005$, $P = 0.76$), mean read number ($\rho = -0.002$, $P = 0.91$), and intrinsic noise ($\rho = 0.005$, $P = 0.72$). Finally, we used the 'scran' package to divide the cells from clone 7 into G1 and G2-M cell cycle stages based on the total reads of each gene in each cell (46). We then computed D'_{int} and D'_{ext} of each gene in each stage. We found that both D'_{int} and D'_{ext} are similar between the stages (Supplementary Figure S3I and J, which can be compared with Figure 1F and G, respectively), indicating that the adjusted noise is a robust property of a gene across cell cycle stages.

The TATA-box is associated with elevated intrinsic and extrinsic noises

Our estimates of D_{int} and D_{ext} for thousands of mouse genes allow testing the potential impacts of several factors on the two noise components. We focused on four factors with prior predictions of their effects. The first factor is the presence/absence of the TATA-box in the promoter region. The TATA-box has been predicted to increase the intrinsic noise because it enlarges the burst size in bursty gene expression through interacting with nucleosomes (44,47). In addition, the TATA-box can increase intrinsic noise by reducing the number of states in promoter cycles (22). Indeed, D_{int} is significantly higher for genes with the TATA-box in the promoter than those without (Figure 2A). The same is true for D'_{int} , which is independent of D_{ext} (Figure 2A). Similar results were obtained from the non-clonal cells (Supplementary Figure S4A).

The presence of the TATA-box sensitizes the promoter to *trans*-regulation (48,49) so should also increase the susceptibility of the promoter to cell state changes (50,51). Hence, we predict that the TATA-box also raises the extrinsic noise. Supporting this prediction, genes with the TATA-box show significantly higher D_{ext} and D'_{ext} than those without (Figure 2B). Similar patterns were observed in the non-clonal cells (Supplementary Figure S4B).

Because the above analyses of the TATA-box are based on correlations, they do not prove causality. Nevertheless, the only other known property of the TATA-box on gene expression is to increase the mean expression level (52), which has already been controlled in our D_{int} and D_{ext} estimates. Our observations, coupled with manipulative experiments showing increased (total) expression noise conferred by the TATA-box (17,47,53), suggests that the influences of the TATA-box on both intrinsic and extrinsic noise revealed here is causal.

Opposing effects of microRNAs on the intrinsic and extrinsic noise of target genes

A microRNA (miRNA) regulates the expressions of its target genes by degrading their mRNAs and/or suppressing their translations (54). Combining mathematical modeling and experimental validation, Schmiedel *et al.* showed that a gene would have an elevated extrinsic protein expression noise if it is targeted by a miRNA than when it is not, because the miRNA concentration varies among cells (55). For the same reason, we expect that miRNA targeting increases the extrinsic mRNA expression noise. Schmiedel *et al.* also showed that the protein intrinsic noise of a gene is reduced when it is targeted by a miRNA than when it is not (56). This is because, under the assumption that the mean mRNA concentration is unaltered, being targeted by a miRNA means a reduction in mRNA half-life and a compensatory increase in transcription. Even though the magnitude of the fluctuation of the mRNA concentration in a cell may be unaltered (see below), the frequency of the fluctuation is higher, which leads to a lower protein intrinsic noise. However, the impact of miRNA targeting on the mRNA intrinsic noise depends on the mechanism underlying the compensatory increase in transcription. If the increased transcription is caused by a higher burst frequency in transcriptional initiation, mRNA intrinsic noise will be reduced. Alternatively, if it is caused by a greater burst size, mRNA intrinsic noise will be increased. It is also possible that the increased transcription is due to a combination of the two mechanisms. We thus explore the following three questions. First, do genes targeted by miRNAs have lower or higher D_{int} and D'_{int} than those not targeted by miRNAs? Second, do genes targeted by more miRNA species have lower or higher D_{int} and D'_{int} ? Third, do genes targeted by miRNAs have higher D_{ext} and D'_{ext} than those not targeted by miRNAs? We obtained relationships between miRNAs and their targets from the RegNetwork database (34) (see Materials and Methods). We found that genes targeted by miRNAs have significantly lower D_{int} and D'_{int} than genes not targeted by miRNAs (Figure 2C). Furthermore, D_{int} (Figure 2D) and D'_{int} (Figure 2E) of a gene are significantly negatively correlated with the number of miRNA species targeting the gene. Regarding the extrinsic noise, D_{ext} and D'_{ext} are significantly higher for genes targeted by miRNAs than those not targeted by miRNAs (Figure 2F). Similar results were obtained from the non-clonal cells (Supplementary Figure S4C–F), except that the results on D_{ext} and D'_{ext} are statistically non-significant (Supplementary Figure S4F), probably due to the aforementioned lower precision of extrinsic noise estimates in the non-clonal cells. Because

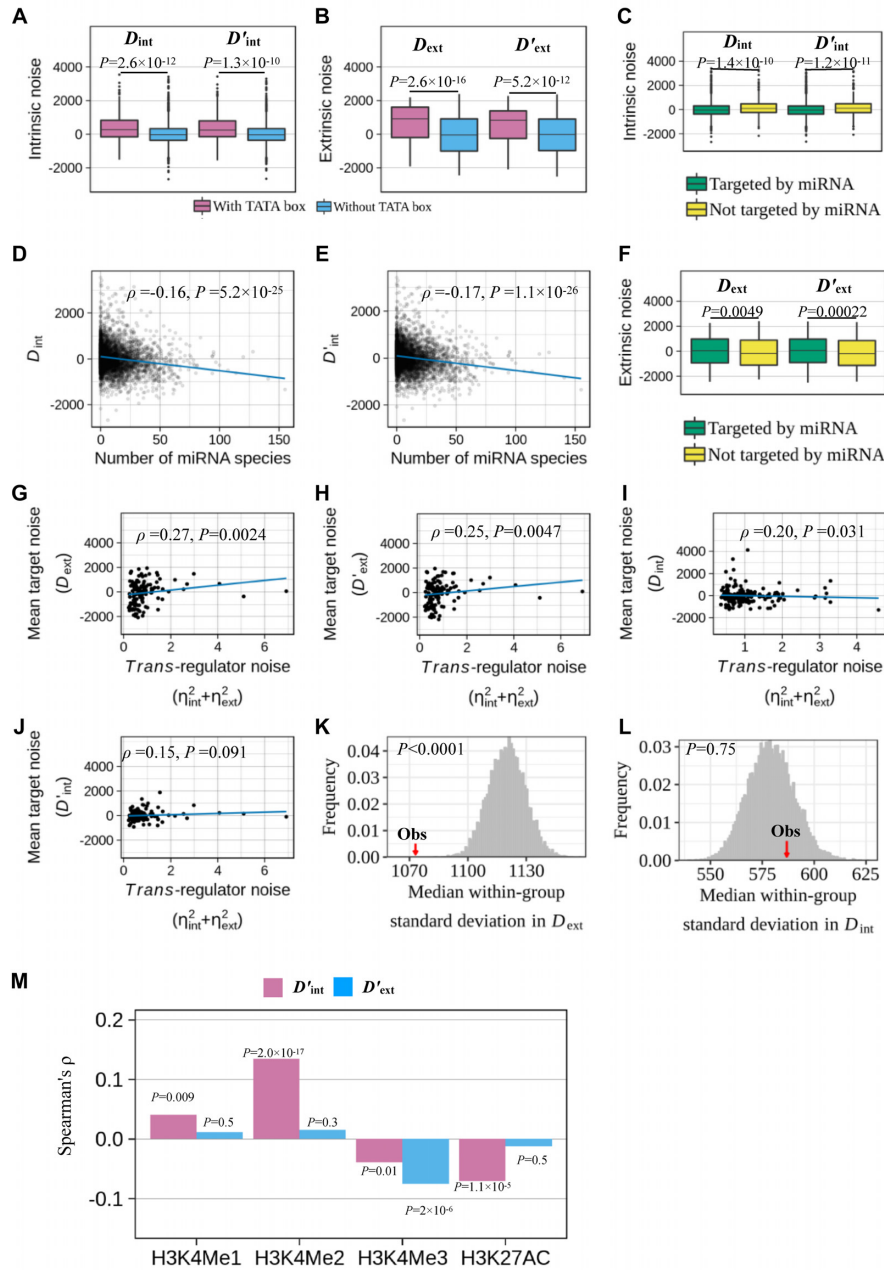


Figure 2. Factors influencing intrinsic and/or extrinsic gene expression noise. **(A)** Genes with a TATA-box in the promoter (pink) have significantly higher intrinsic noise (D_{int}) than genes without a TATA-box (blue). The same is true when intrinsic noise is measured by D'_{int} , which is uncorrelated with extrinsic noise. The lower and upper edges of a box represent the first (qu_1) and third (qu_3) quartiles, respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3 - qu_1)$, and the dots represent values outside the inner fences (outliers). **(B)** Genes with a TATA-box in the promoter (pink) have significantly higher extrinsic noise (D_{ext}) than genes without a TATA-box (blue). The same is true when extrinsic noise is measured by D'_{ext} , which is uncorrelated with intrinsic noise. **(C)** Genes targeted by miRNA (green) have significantly lower intrinsic noise (D_{int} and D'_{int}) than genes not targeted by miRNA (yellow). **(D)** Genes targeted by more miRNA species have lower D_{int} . The blue line displays the linear regression of D_{int} of a target gene on the number of miRNA species targeting it. **(E)** Genes targeted by more miRNA species have lower D'_{int} . The blue line displays the linear regression of D'_{int} of a target gene on the number of miRNA species targeting it. **(F)** Genes targeted by miRNA (green) have significantly higher extrinsic noise (D_{ext} and D'_{ext}) than genes not targeted by miRNA (yellow). **(G)** The mean extrinsic noise (D_{ext}) of genes targeted by the same *trans*-regulator is significantly positively correlated with the total noise ($\eta_{int}^2 + \eta_{ext}^2$) of the *trans*-regulators. **(H)** The mean extrinsic noise (upon the control for intrinsic noise) (D'_{ext}) of genes targeted by the same *trans*-regulator is significantly positively correlated with the total noise ($\eta_{int}^2 + \eta_{ext}^2$) of the *trans*-regulators. **(I)** The mean intrinsic noise (D_{int}) of genes targeted by the same *trans*-regulator is significantly positively correlated with the total noise ($\eta_{int}^2 + \eta_{ext}^2$) of the *trans*-regulator. **(J)** The mean intrinsic noise (upon the control for extrinsic noise) (D'_{int}) of genes targeted by the same *trans*-regulator is not significantly positively correlated with the total noise ($\eta_{int}^2 + \eta_{ext}^2$) of the *trans*-regulator. **(K)** The observed median standard deviation of D_{ext} among genes regulated by the same *trans*-regulator (red arrow) is significantly smaller than the random expectation (histograms). **(L)** The observed median standard deviation of D_{int} among genes regulated by the same *trans*-regulator is not significantly different from the random expectation (histograms). **(M)** Spearman's correlation between the number of histone modification peaks that overlap the core promoter and intrinsic or extrinsic noise.

the only other known function of miRNAs is to regulate the mean expression levels of their targets (54), which are uncorrelated with our noise measures, it is likely that the effects observed here are causal.

Similar extrinsic noises of genes regulated by the same *trans*-regulator

According to the definitions of intrinsic and extrinsic noises, we predict that, if gene A *trans*-regulates gene B, the extrinsic but not intrinsic noise of gene B should rise with the expression noise of gene A. To test this prediction, we obtained the relationship between *trans*-regulators and their target genes from RegNetwork (34). Because both η_{int}^2 and η_{ext}^2 of the *trans*-regulator affect the extrinsic noise of the target genes, we need a measure of the *trans*-regulator noise that takes into account both η_{int}^2 and η_{ext}^2 . For each *trans*-regulator that has estimated η_{int}^2 and η_{ext}^2 , we computed its $\eta_{tot}^2 = \eta_{int}^2 + \eta_{ext}^2$. Here, we gave equal weights to the measured η_{int}^2 and η_{ext}^2 , because of the lack of knowledge of the relative measurement accuracy of η_{int}^2 and η_{ext}^2 . We then computed the average D_{int} and average D_{ext} of all the targets of the *trans*-regulator, respectively, after excluding the *trans*-regulator itself if it self-regulates, because the extrinsic noise of a gene is by definition correlated with its total noise irrespective of the validity of our hypothesis. In support of our hypothesis, we found a positive correlation between the mean target D_{ext} and η_{tot}^2 of their *trans*-regulator ($\rho = 0.27$, $P = 0.0024$; Figure 2G). The same is true for D'_{ext} ($\rho = 0.25$, $P = 0.0047$; Figure 2H). By contrast, although the mean D_{int} of the targets and η_{tot}^2 of their *trans*-regulator are correlated ($\rho = 0.20$, $P = 0.031$; Figure 2I), the correlation becomes non-significant for D'_{int} ($\rho = 0.15$, $P = 0.091$; Figure 2J). In the above, we considered η_{tot}^2 because it is the total noise of the regulator regardless of its source that influences the target extrinsic noise.

It can be further predicted that genes regulated by the same *trans*-regulator should have more similar D_{ext} values but not necessarily more similar D_{int} values, when compared with genes that are not co-regulated by a *trans*-regulator. To test this prediction, we grouped all target genes of each *trans*-regulator, followed by calculation of the standard deviation (SD) of D_{int} and that of D_{ext} within the group. We then computed the median SD of D_{int} and median SD of D_{ext} across all *trans*-regulators. As a comparison, we randomized the targets of each regulator, requiring only that the number of targets of each regulator remained unaltered (see Materials and Methods). We then similarly computed the median SD of D_{int} and median SD of D_{ext} across all *trans*-regulators. This randomization was repeated 10,000 times. We found that the observed median SD of D_{ext} is significantly lower than that from each of the 10 000 randomizations (i.e. $P < 0.0001$; Figure 2K). By contrast, the observed median SD of D_{int} is smaller than that in only 25% of the 10 000 randomizations (i.e. $P = 0.75$; Figure 2L). Together, our results confirm the theoretical prediction that the expression noise of *trans*-regulators primarily affects the extrinsic but not intrinsic expression noise of their targeted genes. We also performed the same analyses in the non-clonal cells. Although the trends exist, they are not sta-

tistically significant (Supplementary Figure S4G-J), likely due to the less precise estimation of expression noise in the non-clonal cells.

Differential effects of histone modification on intrinsic and extrinsic noises

Prompted by recent reports of correlations between several histone modifications and gene expression noise (25,57), we respectively correlated histone modification with intrinsic and extrinsic expression noises. Specifically, we collected histone modification peak data from Cistrome (35), and computed the correlation between histone modification strength in the core promoter and D'_{int} or D'_{ext} . We found H3K4Me1 modification to be significantly positively correlated with D'_{int} but not significantly correlated with D'_{ext} (Figure 2M). The same can be said for H3K4Me2 (Figure 2M). By contrast, H3K4Me3 modification is significantly negatively correlated with both D'_{int} and D'_{ext} , but the correlation with D'_{ext} is much stronger than that with D'_{int} (Figure 2M). H3K27Ac modification is significantly negatively correlated with D'_{int} but not significantly correlated with D'_{ext} (Figure 2M). These observations suggest that histone modification often differentially impacts intrinsic and extrinsic expression noises.

The genome-wide finding that (i) the TATA-box increases both D_{int} and D_{ext} , (ii) miRNAs decrease the D_{int} but increase the D_{ext} of its targets, (iii) the D_{ext} but not D_{int} of a gene is impacted by the expression noise of its *trans*-regulator and (iv) histone modification is differentially correlated with D'_{int} and D'_{ext} not only reveals mechanisms responsible for the variations of intrinsic and extrinsic expression noises among genes, but also demonstrates that our high-throughput estimation of intrinsic and expression noises is reliable. Note, however, because the above analyses were based on rank statistics, the absolute effect sizes are unknown. Consequently, it is unclear whether the above findings are biologically important. In the following section, we address this question by asking if the two noise components have been subject to differential natural selection for groups of genes of different functions and whether the above studied mechanisms have been used in the optimizations of the two noise components.

Genes with mitochondrial functions show lowered extrinsic expression noise

Previous studies found that the variation in mitochondrial function among cells is a primary source of global extrinsic noise of gene expression, because protein synthesis requires ATP, which is largely produced by the mitochondrion (58,59). We thus predict that natural selection should have minimized the expression noise of (nuclear) genes that function in the mitochondrion in order to reduce the gene expression noise globally. Indeed, one source of the protein level noise of proteins localized to the mitochondrion is the partition of mitochondria during the cell division, and recent work showed that this partition is tightly regulated presumably to ensure equal partitions (60). To achieve a low expression noise at the mRNA level for nuclear genes with mitochondrial functions, selection could have reduced the

intrinsic noise, extrinsic noise, or both. However, for highly expressed genes, the extrinsic noise is the main contributor to expression noise, because the intrinsic noise is naturally low when the mean expression is high (21,55). We noticed in our data that nuclear genes of mitochondrial functions are highly expressed relative to other nuclear genes ($P = 1.9 \times 10^{-15}$, Mann–Whitney U test). Because D_{int} and D_{ext} are independent of the mean expression level, we predict that genes functioning in the mitochondrion should have reduced D_{ext} but not necessarily reduced D_{int} . Indeed, D_{ext} is significantly lower for nuclear genes functioning in the mitochondrion when compared with other nuclear genes (Figure 3A), and this disparity remains for D'_{ext} (Figure 3A). By contrast, D_{int} is not significantly different between the two groups of genes (Figure 3B), whereas D'_{int} is even slightly larger for genes functioning in the mitochondrion than other genes (Figure 3B). Similar results were obtained from the non-clonal cells (Supplementary Figure S5).

What are the underlying molecular mechanisms responsible for the reduction of D_{ext} of genes functioning in the mitochondrion? Based on the earlier results (Figure 2), possible mechanisms include the underrepresentation of the TATA-box in genes functioning in the mitochondrion, underrepresentation of miRNA targeting, preferential regulation by quiet *trans*-regulators, and certain histone modifications. Because our noise data do not include many *trans*-regulators and because the causal relationship between histone modification and noise is less clear, we focused on the first two mechanisms. Indeed, compared with other genes, those functioning in the mitochondrion are depleted of the TATA-box ($P = 4.6 \times 10^{-5}$, Fisher's exact test; Figure 3C) and are less targeted by miRNAs ($P = 0.036$, Fisher's exact test; Figure 3D). To explore whether the depletion of TATA-box and miRNA targeting can fully account for the reduction in extrinsic noise of nuclear genes functioning in the mitochondrion, we regressed D_{ext} as a linear function of the presence/absence of TATA-box and miRNA targeting. The residual of the above regression provided an extrinsic noise measure upon the control for TATA-box and miRNA targeting. We found that the difference in extrinsic noise between nuclear genes that function in the mitochondrion and other genes remains significant (D_{ext} : $P = 0.001$, Mann–Whitney U test; D'_{ext} : $P = 0.00065$, Mann–Whitney U test). Thus, depletions of the TATA-box and miRNA targeting are only part of the mechanisms responsible for the selective reduction of the D_{ext} of genes functioning in the mitochondrion.

Genes encoding protein complex members have lowered intrinsic expression noise

Because dosage balance is important for protein complex members (61–63) and because as long as members of the same protein complex are co-regulated in expression, extrinsic noise does not create dosage imbalance (23), we predict that protein complex members have reduced intrinsic noise but not necessarily reduced extrinsic noise. An early yeast study showed that, compared with other proteins, protein complex members have lowered protein level noises measured in morphologically similar cells, suggesting that they have reduced intrinsic noise (5). In our data

where intrinsic and extrinsic noises are explicitly separated, we found D_{int} significantly lower for genes encoding protein complex members than other genes (Figure 4A). The same is true for D'_{int} (Figure 4A). By contrast, although D_{ext} is significantly lower for genes encoding protein complex members than other genes (Figure 4B), this disparity becomes non-significant for D'_{ext} (Figure 4B). Similar patterns were observed in the non-clonal cells (Supplementary Figure S6).

Potential mechanisms underlying the D_{ext} difference between genes encoding protein complex members and other genes can include a depletion of the TATA-box and an enrichment of miRNA targeting in the former group. Indeed, compared with other genes, those encoding protein complex members tend not to use the TATA-box (Figure 4C), tend to be targeted by miRNAs (Figure 4D), and tend to be targeted by more miRNA species (Figure 4E). The difference between genes encoding protein complex members and other genes in intrinsic noise after adjusting the presence/absence of TATA-box and the number of miRNA species targeting the gene by linear regression remains significant for both D_{int} ($P = 0.017$, Mann–Whitney U test) and D'_{int} ($P = 0.031$, Mann–Whitney U test), suggesting that other mechanisms also contribute to the lowered intrinsic noise of protein complex members.

Cell cycle genes have low intrinsic but high extrinsic noise

Cell cycle genes are those that control the cell cycle and hence should express differently at different cell cycle stages (64). However, within a cell that is at a cellular stage, cell cycle genes should preferably show consistent expressions. Thus, we predict that cell cycle genes have been selected to have low D_{int} but high D_{ext} . Indeed, compared with other genes, cell cycle genes show significantly lower D_{int} and D'_{int} (Figure 5A), but significantly higher D_{ext} and D'_{ext} (Figure 5B). This finding echoes the recent report that the genetic circuit underlying the biological clock often has an architecture to buffer the harmful internal fluctuation of signals while responding to the variation of the functional external stimuli (65). The analysis of the non-clonal cells yielded similar results (Supplementary Figure S7).

Given the noise features of the cell cycle genes, we predict that they should be preferentially targeted by miRNAs, because miRNA targeting lowers the intrinsic noise but raises the extrinsic noise. In addition, we know that the impact of miRNAs on the intrinsic noise (but not necessarily the extrinsic noise) of a target rises with the number of miRNA species targeting the gene (Figure 2C). We found that the fraction of genes targeted by miRNAs is not significantly higher for cell cycle genes than other genes ($P = 0.30$, Fisher's exact test; Figure 5C), but the median number of miRNA species targeting a gene is significantly higher for cell cycle genes than other genes ($P = 0.0071$, Mann–Whitney U test; Figure 5D). These observations suggest that miRNA targeting is not responsible for cell cycle genes' high D_{ext} but is responsible for their low D_{int} . Notwithstanding, we cannot rule out the possibility that the non-significant result in Figure 5C is due to the relatively small sample size of cell cycle genes ($n = 570$, as opposed to 935 for genes encoding protein complex members and 1603 for

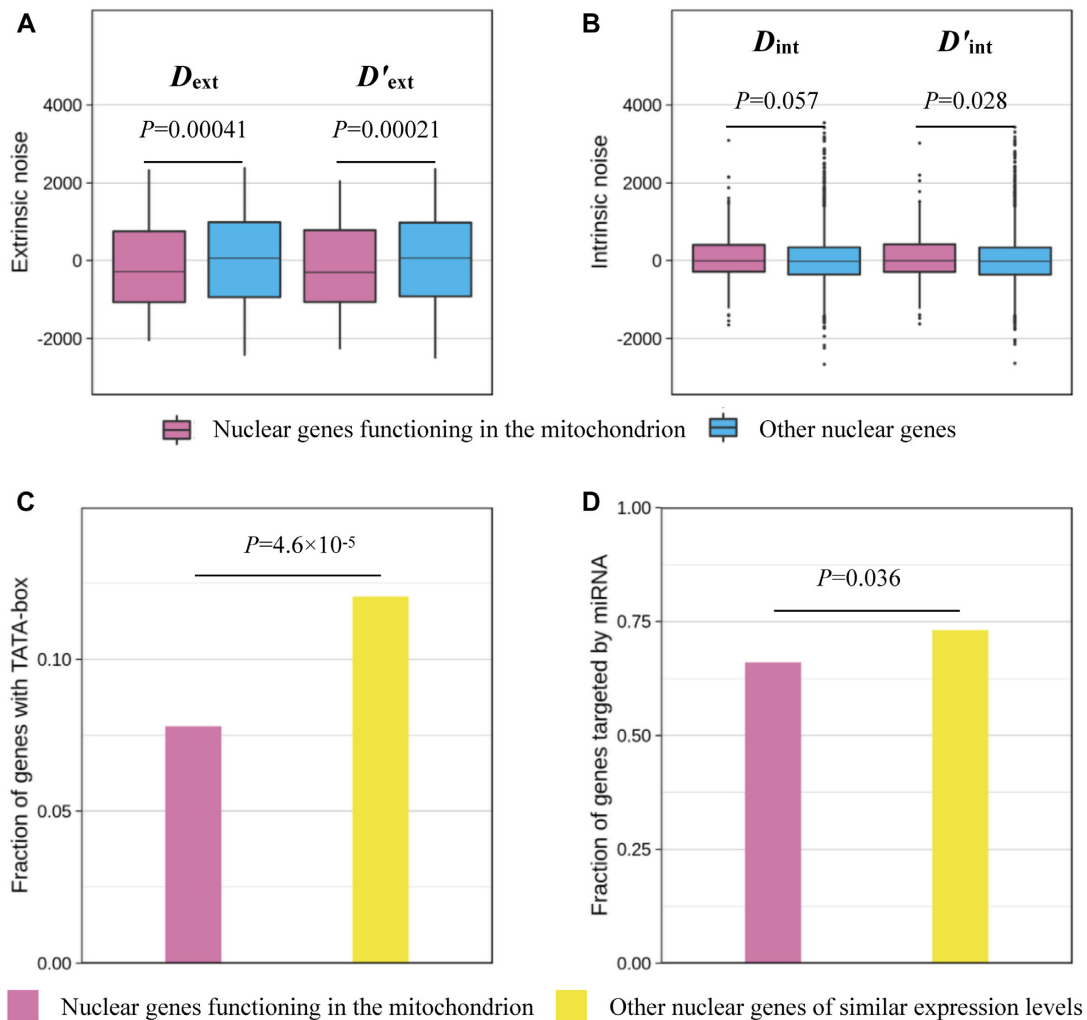


Figure 3. Nuclear genes functioning in the mitochondrion have lower extrinsic noise but not lower intrinsic noise when compared with other genes. (A) Nuclear genes functioning in the mitochondrion (pink) have significantly lower extrinsic noise (D_{ext} and D'_{ext}) than other genes (blue). The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3 - qu_1)$, and the dots represent values outside the inner fences (outliers). (B) Nuclear genes functioning in the mitochondrion (pink) do not have significantly lower intrinsic noise D_{int} and even have significantly higher D'_{int} than other genes (blue). (C) TATA-box is underrepresented in the promoters of nuclear genes functioning in the mitochondrion (pink) when compared with other genes of similar expression levels (yellow). (D) Nuclear genes functioning in the mitochondrion (pink) are less targeted by miRNAs than other genes with similar expression levels (yellow).

genes functioning in the mitochondrion). After adjusting the number of miRNA species targeting a gene, we found that cell cycle genes still have lower D_{int} ($P=0.0057$, Mann–Whitney U test) and D'_{int} ($P=0.0013$, Mann–Whitney U test) than other genes, suggesting the existence of other factors contributing to the low intrinsic noise of cell cycle genes.

Other genes with exceptionally high or low extrinsic or intrinsic noise

To learn more about the biological implications of intrinsic and extrinsic noise, we performed gene ontology (GO) analysis on genes with extreme D_{ext} and/or D_{int} values. We first defined high D_{ext} genes as those genes whose D_{ext} values are in the highest 10% of all 3975 genes and low D_{ext} genes as

those whose D_{ext} values are in the lowest 10% of all 3975 genes. We similarly defined high D_{int} genes and low D_{int} genes. These genes show enrichments of various functional categories (Table 1). For instance, both the high D_{ext} group and high D_{int} group are enriched with genes encoding secreted proteins and extracellular proteins. Secreted and extracellular proteins synthesized from many individual cells are mixed together and function outside the cells, so there is no need to reduce their expression noise at the mRNA level. Thus, their high noise likely reflects a lack of selection minimizing their noise. By contrast, the low D_{ext} group are enriched with genes whose products interact with RNAs, whereas the low D_{int} group are enriched with genes encoding phosphoproteins and proteins with coiled coil structure, again indicating that the biological implications of extrinsic noise and intrinsic noise can be different. Similar results

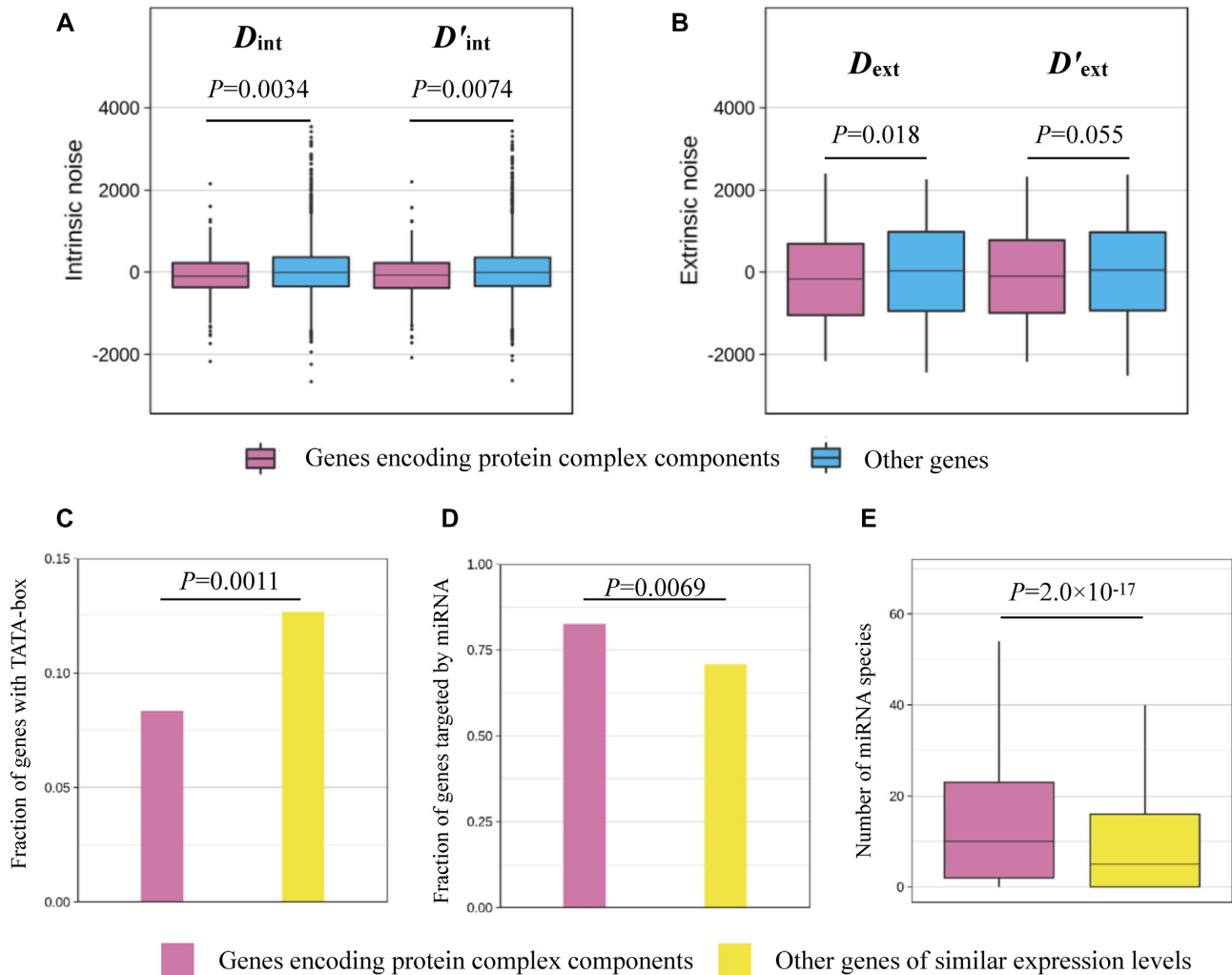


Figure 4. Genes encoding protein complex components have lower intrinsic noise but not lower extrinsic noise than other genes. (A) Genes encoding protein complex components (pink) have significantly lower intrinsic noise (D_{int} and D'_{int}) than other genes (blue). The lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3 - qu_1)$, and the dots represent values outside the inner fences (outliers). (B) Genes encoding protein complex components (pink) have significantly lower D_{ext} but not significantly lower D'_{ext} than other genes (blue). (C) TATA-box is underrepresented in the promoters of genes encoding protein complex components (pink) when compared with other genes of similar expression levels (yellow). (D) Genes encoding protein complex components (pink) are more likely to be targeted by miRNAs when compared with other genes of similar expression levels (yellow). (E) Genes encoding protein complex components (pink) tend to be targeted by more miRNA species when compared with other genes of similar expression levels (yellow).

were found for the non-clonal cells (Supplementary Table S1).

We further examined genes with different combinations of extreme extrinsic and intrinsic noises (Table 1 and Supplementary Table S1). Specifically, we identified genes with both high D_{ext} and high D_{int} , high D_{ext} but low D_{int} , low D_{ext} but high D_{int} , and both low D_{ext} and low D_{int} , respectively. Here, a gene is considered to have high (or low) noise if its noise is ranked in the top (or bottom) 25% among the 3975 genes. As expected, the group with both high D_{ext} and high D_{int} is enriched with genes encoding secreted and extracellular proteins, while the group with high D_{ext} but low D_{int} is enriched with cell cycle genes. The group with low D_{ext} but high D_{int} is not enriched with any GO category. Finally, the group with both low D_{ext} and low D_{int} is enriched with genes encoding RNA-interacting proteins and phosphopro-

teins. The identification of genes with extreme noise values can help further understand the biological significance and constraints of intrinsic and extrinsic gene expression noises.

DISCUSSION

Using allele-specific scRNA-seq, we performed the first genomic estimation of intrinsic and extrinsic expression noises of any species. The mRNA noise estimates obtained allowed us to evaluate the predicted effects of various factors. In particular, we found that (i) the presence of the TATA-box in the promoter of a gene increases both the intrinsic and extrinsic expression noise of the gene, (ii) miRNAs lower the intrinsic noise but increase the extrinsic noise of their target genes, (iii) the extrinsic noise of a gene increases with the

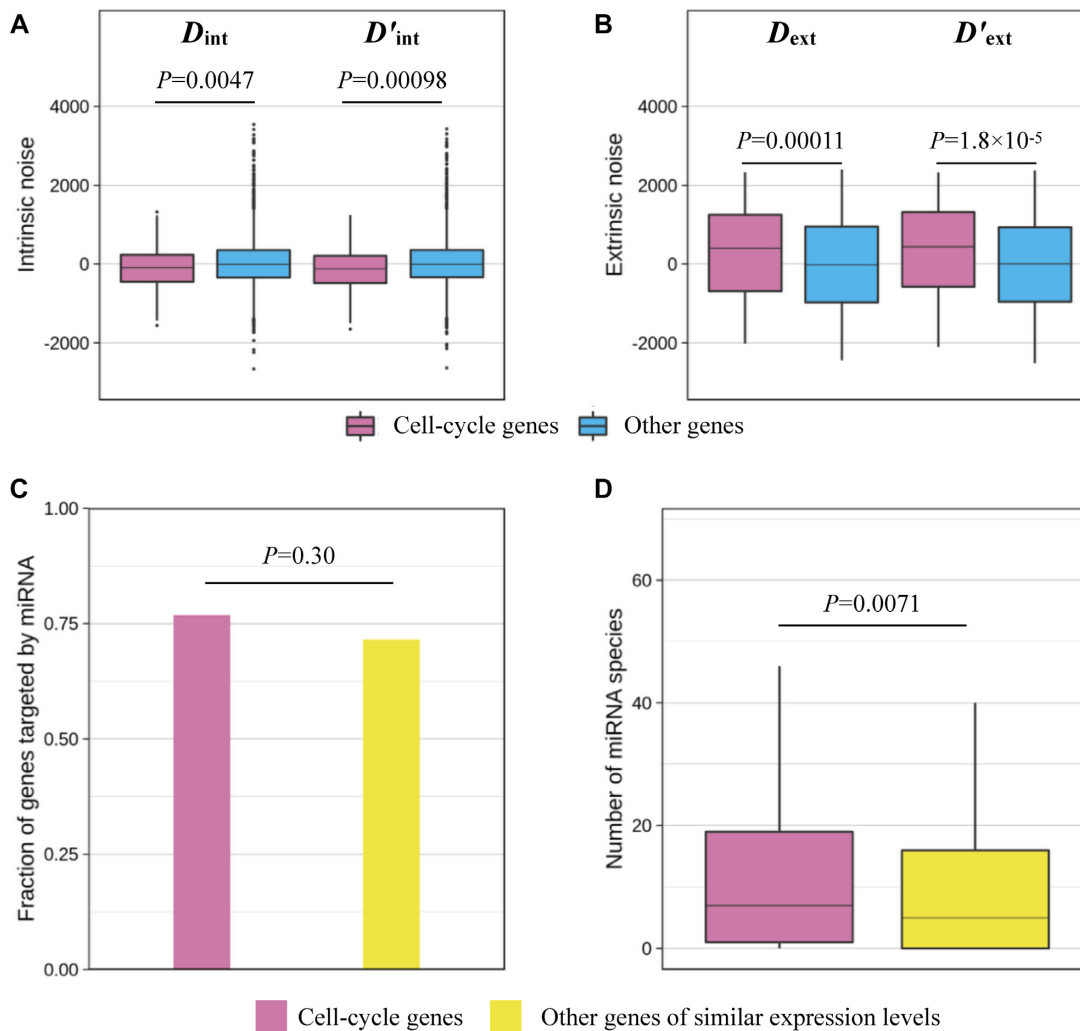


Figure 5. Cell cycle genes have lower intrinsic noise but higher extrinsic noise than other genes. (A) Cell cycle genes (pink) have significantly lower intrinsic noise (D_{int} and D'_{int}) when compared with other genes (blue). The lower and upper edges of a box represent the first (q_1) and third quartiles (q_3), respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(q_3 - q_1)$, and the dots represent values outside the inner fences (outliers). (B) Cell cycle genes (pink) have significantly higher extrinsic noise (D_{ext} and D'_{ext}) when compared with other genes. (C) Fraction of genes targeted by miRNAs is not significantly different between cell cycle genes (pink) and other genes of similar expression levels (yellow). (D) Cell cycle genes (pink) tend to be targeted by more miRNA species than other genes of similar expression levels (yellow).

total expression noise of its *trans*-regulator, (iv) genes regulated by the same *trans*-regulator have more similar extrinsic expression noises than genes not co-regulated and (v) histone modification is often differentially correlated with intrinsic and extrinsic noises. Considering gene functions, we formulated hypotheses on natural selection for lowered or elevated intrinsic and/or extrinsic noise of groups of genes, and were able to find evidence supporting these hypotheses. Specifically, we predicted and then demonstrated that (nuclear) genes functioning in the mitochondrion have reduced extrinsic noise, genes encoding protein complex members have decreased intrinsic noise, and cell cycle genes have lowered intrinsic noise but elevated extrinsic noise.

It is valuable to compare our results with previous genome-wide studies of total protein expression noise. For example, a study in yeast showed that nuclear genes functioning in the mitochondrion have unusually high protein

noise, presumably due to the random partition of mitochondria during cell division (20). Multiple studies reported that expression noise of nuclear genes functioning in the mitochondrion can result in large, presumably harmful among-cell variation in global gene expression (58,59,66). It was thus unclear whether the gene expression noise of nuclear genes functioning in mitochondrion has been subject to selective minimization. Our results on the mRNA expression noise of nuclear genes functioning in the mitochondrion provide clear evidence for the minimization. Our ability to detect this signal is likely because mRNAs are located in the cytoplasm so are not subject to the problem of block partition of mitochondrial proteins. Regarding genes encoding protein complex members, a previous study (5) suggested that their low noise may be explained by one or more of the following reasons. First, protein complex members are enriched for essential genes and essential genes tend to

Table 1. Significantly enriched GO terms among genes with extreme intrinsic and/or extrinsic expression noise in clone 7. The three most significant terms are presented if more than three terms are significantly enriched

GO terms	Corrected <i>P</i> -values
High extrinsic noise	
Secreted	3.5×10^{-12}
Extracellular region	1.5×10^{-11}
Signal peptide	7.3×10^{-9}
Low extrinsic noise	
Poly (A) RNA binding	6.7×10^{-7}
RNA binding	1.2×10^{-6}
rRNA processing	1.8×10^{-6}
High intrinsic noise	
Extracellular region	1.4×10^{-8}
Signal peptide	3.1×10^{-8}
Disulfide bond	1.0×10^{-7}
Low intrinsic noise	
Phosphoprotein	1.9×10^{-6}
Coiled coil	4.6×10^{-6}
High extrinsic noise and high intrinsic noise	
Signal peptide	1.2×10^{-13}
Secreted	4.9×10^{-13}
Extracellular region	2.1×10^{-12}
High extrinsic noise and low intrinsic noise	
Cell cycle	0.01
Low extrinsic noise and low intrinsic noise	
Poly (A) RNA binding	1.7×10^{-24}
Nucleus	1.1×10^{-8}
Nucleolus	1.9×10^{-8}

have low noise. Second, protein complex members are more dosage-sensitive due to the requirement for dosage balance among members of the same complex. Third, the low noise of protein complex members is a by-product of their short protein half-lives. Our results do not support the first or third reason, because the first reason would predict both low extrinsic noise and low intrinsic noise, contrasting our observation of reduction in D_{int} but not D_{ext} , while the third reason would predict no reduction in the mRNA expression noise, contradictory to our observation of lowed D_{int} . With respect to cell cycle genes, no previous research has ever found them to have low expression noise despite the suggestion that cell cycle should be robust to biochemical noise (67,68). This is possibly because previous studies did not separate intrinsic from extrinsic noise, while cell cycle genes are expected to and indeed have low D_{int} but high D_{ext} .

Our analyses have several caveats that are worth discussion. First, although many of our statistical results are highly significant, the effect sizes of some factors appear small. This may be due to the high technical noises of scRNA-seq-based expression level measures (69), which is further exacerbated in allele-specific scRNA-seq, because only reads containing information of the allele of origin, which constitute a small fraction of all reads, are useful to our analysis. The high technical noise introduces both random errors and systematic errors in our estimation of expression noise. Random errors are not expected to create spurious results in large samples (70). By contrast, systematic errors may create spurious results. In our analysis, we removed known systematic errors from technical noises (32) by controlling for the number of reads per gene. Thus, the

remaining errors in our estimation of intrinsic noise and extrinsic noise should be largely random, and these random errors have likely caused underestimation of effect sizes in our study. Furthermore, whether an effect is evolutionarily important depends on whether it is detectable by natural selection. Our observation of differential uses of various molecular mechanisms such as the TATA-box and miRNA targeting in the optimization of intrinsic and extrinsic noise levels demonstrates that the detected effects are important. Second, previous theoretical studies showed that noise decomposition using the dual reporter system is accurate under static environments but may not be accurate under dynamic environments; in the latter case, noise decomposition may not reveal the underlying mechanism (13,24,45). Notwithstanding, we found that the intrinsic and extrinsic noises estimated in this study largely follow expectations. More importantly, intrinsic and extrinsic noises do have different biological meanings and hence are differentially tuned evolutionarily. Hence, the noise decomposition appears biologically meaningful and useful. Third, a central topic about noise decomposition is the absolute magnitudes of intrinsic and extrinsic noises (16–18). As mentioned, because of the relatively large size of the technical noise from allele-specific scRNA-seq and different impacts of the technical noise on measures of intrinsic and extrinsic noises, it is impossible to compute and compare the absolute magnitudes of intrinsic and extrinsic noises. This limitation forced us to use rank-based statistics, which made it difficult to estimate absolute effect sizes of various factors. Fourth, our study focused on mRNA expression noise, but one might argue that mRNA noise does not directly correspond to protein noise. We believe that this should not be an issue, because of substantial evidence that mRNA noise is the major source of protein noise (3,16,24,71,72). Finally, to obtain reliable noise estimates, we filtered out genes with low average read counts. Therefore, our conclusions mainly apply to genes with moderate to high expressions. Because lowly expressed genes are impacted most by noise (16), it will be important to study intrinsic and extrinsic noises of lowly expressed genes in the future.

In sum, our study performed the first genome-scale estimation of intrinsic and extrinsic gene expression noise at the mRNA level. We demonstrated the general reliability of our noise estimates and illustrated the utility of these estimates for understanding the mechanisms controlling and selections on the two noise components. Our findings may have implications for synthetic biology, where one often needs to design genetic circuits that have robust yet dynamic behaviors. For example, the detailed mechanisms that cells employ to allow cell cycle genes to have high extrinsic noise but low intrinsic noise may provide insights for designing oscillators that are sensitive to different cell states yet are robust to intrinsic noise (73–75).

DATA AVAILABILITY

The original single-cell RNA-seq data analyzed here have been published (29). Results and computer code are available at GitHub (<https://github.com/mengysun/Dissecting-noise-project>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank members of the Zhang lab for valuable comments.

FUNDING

U.S. National Institutes of Health [GM120093 to J.Z.]. Funding for open access charge: U.S. National Institutes of Health [GM120093].

Conflict of interest statement. None declared.

REFERENCES

- Raser, J.M. and O'shea, E.K. (2005) Noise in gene expression: origins, consequences, and control. *Science*, **309**, 2010–2013.
- Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dollé, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H. and Klein, C.A. (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**, 1011–1014.
- Batada, N.N. and Hurst, L.D. (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.*, **39**, 945–947.
- Kemkemer, R., Schrank, S., Vogel, W., Gruler, H. and Kaufmann, D. (2002) Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 13783–13788.
- Lehner, B. (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.*, **4**, 170.
- Wang, Z. and Zhang, J. (2011) Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, E67–E76.
- Xu, H., Liu, J.-J., Liu, Z., Li, Y., Jin, Y.-S. and Zhang, J. (2019) Synchronization of stochastic expressions drives the clustering of functionally related genes. *Sci. Adv.*, **5**, eaax6525.
- Veening, J.-W., Smits, W.K. and Kuipers, O.P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, **62**, 193–210.
- Zhang, Z., Qian, W. and Zhang, J. (2009) Positive selection for elevated gene expression noise in yeast. *Mol. Syst. Biol.*, **5**, 299.
- Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. and Huang, S. (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, **453**, 544–547.
- Huang, S. (2009) Non-genetic heterogeneity of cells in development: more than just noise. *Development*, **136**, 3853–3862.
- Turing, A.M. (1952) The chemical basis of morphogenesis. *Philos. Trans. R. Soc. London*, **237**, 37–72.
- Hilfinger, A. and Paulsson, J. (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 12167–12172.
- Swain, P.S., Elowitz, M.B. and Siggia, E.D. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12795–12800.
- Sharon, E., van Dijk, D., Kalma, Y., Keren, L., Manor, O., Yakhini, Z. and Segal, E. (2014) Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.*, **24**, 1698–1706.
- Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y. and Barkai, N. (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.*, **38**, 636–643.
- Raser, J.M. and O'shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science*, **304**, 1811–1814.
- Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
- Faure, A.J., Schmiedel, J.M. and Lehner, B. (2017) Systematic analysis of the determinants of gene expression noise in embryonic stem cells. *Cell Syst.*, **5**, 471–484.
- Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L. and Weissman, J.S. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, **441**, 840–846.
- Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X.S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533–538.
- Zoller, B., Nicolas, D., Molina, N. and Naef, F. (2015) Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.*, **11**, 823.
- Stewart-Ornstein, J., Weissman, J.S. and El-Samad, H. (2012) Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell*, **45**, 483–493.
- Sherman, M.S., Lorenz, K., Lanier, M.H. and Cohen, B.A. (2015) Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. *Cell Syst.*, **1**, 315–325.
- Wu, S., Li, K., Li, Y., Zhao, T., Li, T., Yang, Y.-F. and Qian, W. (2017) Independent regulation of gene expression level and noise by histone modifications. *PLoS Comput. Biol.*, **13**, e1005585.
- Morgan, M.D. and Marioni, J.C. (2018) CpG island composition differences are a source of gene expression noise indicative of promoter responsiveness. *Genome Biol.*, **19**, 81.
- Kim, J.K., Kolodziejczyk, A.A., Illic, T., Teichmann, S.A. and Marioni, J.C. (2015) Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.*, **6**, 8687.
- Jiang, Y., Zhang, N.R. and Li, M. (2017) SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol.*, **18**, 74.
- Reinius, B., Mold, J.E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J. and Sandberg, R. (2016) Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.*, **48**, 1430–1435.
- Fu, A.Q. and Pachter, L. (2016) Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. *Stat. Appl. Genet. Mol. Biol.*, **15**, 447–471.
- Rhee, A., Cheong, R. and Levchenko, A. (2014) Noise decomposition of intracellular biochemical signaling networks using nonequivalent reporters. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 17330–17335.
- Grün, D., Kester, L. and Van Oudenaarden, A. (2014) Validation of noise models for single-cell transcriptomics. *Nat. Methods*, **11**, 637–640.
- Dreos, R., Ambrosini, G., Groux, R., Cavin Périer, R. and Bucher, P. (2016) The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res.*, **45**, D51–D55.
- Liu, Z.-P., Wu, C., Miao, H. and Wu, H. (2015) RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*, **2015**, bav095.
- Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J. and Lei, Y. (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.*, **12**, R83.
- Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J. and Plath, K. (2017) Cooperative binding of transcription factors orchestrates reprogramming. *Cell*, **168**, 442–459.
- Xie, W., Nagarajan, S., Baumgart, S.J., Kosinsky, R.L., Najafova, Z., Kari, V., Hennion, M., Indenbirken, D., Bonn, S. and Grundhoff, A. (2017) RNF40 regulates gene expression in an epigenetic context-dependent manner. *Genome Biol.*, **18**, 32.
- Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C. and Clapham, P. (2016) Ensembl 2017. *Nucleic Acids Res.*, **45**, D635–D642.
- Ruepp, A., Waegel, B., Lechner, M., Brauner, B., Dungen-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C. and Mewes, H.-W. (2009) CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.*, **38**, D497–D501.
- Söllner, J.F., Lepar, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E. and Simon, E. (2017) An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci. Data*, **4**, 170185.
- Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2008) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.

42. Wagner, A., Regev, A. and Yosef, N. (2016) Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.*, **34**, 1145–1160.
43. Svensson, V., Natarajan, K.N., Ly, L.-H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A. and Teichmann, S.A. (2017) Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods*, **14**, 381–387.
44. Hornung, G., Bar-Ziv, R., Rosin, D., Tokuriki, N., Tawfik, D.S., Oren, M. and Barkai, N. (2012) Noise–mean relationship in mutated promoters. *Genome Res.*, **22**, 2409–2417
45. Shahrezaei, V., Ollivier, J.F. and Swain, P.S. (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, **4**, 196.
46. Lun, A.T., McCarthy, D.J. and Marioni, J.C. (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*, **5**, 2122.
47. Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R. and Collins, J.J. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell*, **24**, 853–865.
48. Hornung, G., Oren, M. and Barkai, N. (2012) Nucleosome organization affects the sensitivity of gene expression to promoter mutations. *Mol. Cell*, **46**, 362–368.
49. Tirosh, I. and Barkai, N. (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res.*, **18**, 1084–1091.
50. Paulsson, J. (2004) Summing up the noise in gene networks. *Nature*, **427**, 415–418.
51. Pedraza, J.M. and van Oudenaarden, A. (2005) Noise propagation in gene networks. *Science*, **307**, 1965–1969.
52. Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–510.
53. Murphy, K.F., Adams, R.M., Wang, X., Balazsi, G. and Collins, J.J. (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res.*, **38**, 2712–2726.
54. Bartel, D.P. (2018) Metazoan MicroRNAs. *Cell*, **173**, 20–51.
55. Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Blüthgen, N., Marks, D.S. and van Oudenaarden, A. (2015) MicroRNA control of protein expression noise. *Science*, **348**, 128–132.
56. Schmiedel, J., Marks, D.S., Lehner, B. and Blüthgen, N. (2017) Noise control is a primary function of microRNAs and post-transcriptional regulation. bioRxiv doi: <https://doi.org/10.1101/168641>, 26 July 2017, preprint: not peer reviewed.
57. Chen, X. and Zhang, J. (2016) The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.*, **2**, 347–354.
58. Das Neves, R.P., Jones, N.S., Andreu, L., Gupta, R., Enver, T. and Iborra, F.J. (2010) Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.*, **8**, e1000560.
59. Johnston, I.G., Gaal, B., das Neves, R.P., Enver, T., Iborra, F.J. and Jones, N.S. (2012) Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput. Biol.*, **8**, e1002416.
60. Jajoo, R., Jung, Y., Huh, D., Viana, M.P., Rafelski, S.M., Springer, M. and Paulsson, J. (2016) Accurate concentration control of mitochondria and nucleoids. *Science*, **351**, 169–172.
61. Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14746–14753.
62. Papp, B., Pal, C. and Hurst, L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, **424**, 194–197.
63. Sun, M. and Zhang, J. (2019) Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLoS Genet.*, **15**, e1008389.
64. Cho, R.J., Campbell, M.J., Winzler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D. and Lockhart, D.J. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
65. Pittayakanchit, W., Lu, Z., Chew, J., Rust, M.J. and Murugan, A. (2018) Biophysical clocks face a trade-off between internal and external noise resistance. *eLife*, **7**, e37624.
66. Dhar, R., Missarova, A.M., Lehner, B. and Carey, L.B. (2019) Single cell functional genomics reveals the importance of mitochondria in cell-to-cell phenotypic variation. *eLife*, **8**, e38904.
67. Li, F., Long, T., Lu, Y., Ouyang, Q. and Tang, C. (2004) The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4781–4786.
68. Vilar, J.M., Kueh, H.Y., Barkai, N. and Leibler, S. (2002) Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5988–5992.
69. Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M. and Wold, B.J. (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.*, **24**, 496–510.
70. Hedge, C., Powell, G. and Sumner, P. (2018) The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods*, **50**, 1166–1186.
71. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. and Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.*, **4**, e309.
72. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J. and Eisen, M.B. (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol.*, **2**, e137.
73. Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
74. Fung, E., Wong, W.W., Suen, J.K., Bulter, T., Lee, S.-g. and Liao, J.C. (2005) A synthetic gene–metabolic oscillator. *Nature*, **435**, 118–122.
75. Potvin-Trottier, L., Lord, N.D., Vinnicombe, G. and Paulsson, J. (2016) Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, **538**, 514–517.