

**Comparative analysis between single-cell RNA-seq and single-molecule RNA FISH  
indicates that the pyrimidine nucleobase idoxuridine (IdU) globally amplifies  
transcriptional noise**

Giuliana P. Calia<sup>1</sup>, Xinyue Chen<sup>1</sup>, Binyamin Zuckerman<sup>1</sup>, Leor S. Weinberger<sup>1,2,3\*</sup>

<sup>1</sup> Gladstone|UCSF Center for Cell Circuitry, <sup>2</sup> Department of Biochemistry and Biophysics, and

<sup>3</sup> Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158

\*to whom correspondence should be addressed: [leor.weinberger@gladstone.ucsf.edu](mailto:leor.weinberger@gladstone.ucsf.edu)

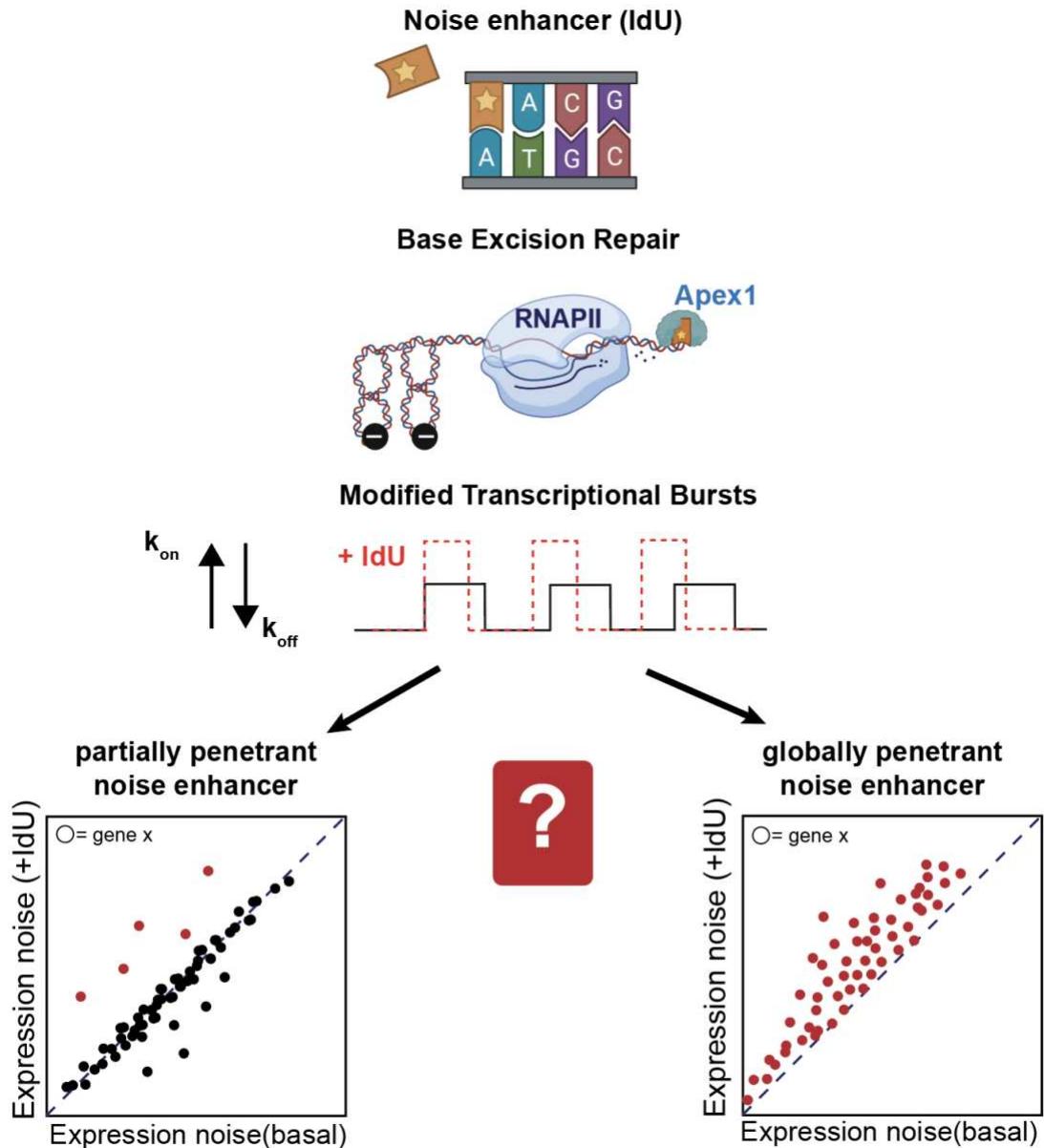
Running Title: Global penetrance of IdU-induced noise amplification

## **ABSTRACT**

**Stochastic fluctuations (noise) in transcription generate substantial cell-to-cell variability, but the physiological roles of noise have remained difficult to determine in the absence of generalized noise-modulation approaches.** Previous single-cell RNA-sequencing (scRNA-seq) suggested that the pyrimidine-base analog (5'-iodo-2'-deoxyuridine, IdU) could generally amplify noise without substantially altering mean-expression levels but scRNA-seq technical drawbacks potentially obscured the *penetrance* of IdU-induced transcriptional noise amplification. Here we quantify global-vs.-partial penetrance of IdU-induced noise amplification by assessing scRNA-seq data using numerous normalization algorithms and directly quantifying noise using single-molecule RNA FISH (smFISH) for a panel of genes from across the transcriptome. Alternate scRNA-seq analyses indicate IdU-induced noise amplification for ~90% of genes, and smFISH data verified noise amplification for ~90% of tested genes. Collectively, this analysis indicates which scRNA-seq algorithms are appropriate for quantifying noise and argues that IdU is a globally penetrant noise-enhancer molecule that could enable investigations of the physiological impacts of transcriptional noise.

**Keywords:** noise enhancer molecule/ transcriptional noise/ single-cell RNA sequencing

## Graphical Abstract



## **INTRODUCTION**

Cell-to-cell variability is an unavoidable consequence of the biochemical processes occurring in individual cells <sup>1</sup>. While some portion of cell-to-cell variability arises from extrinsic factors (e.g., cell size, cycle phase, or microenvironment), a substantial body of literature has demonstrated that, in isogenic populations of cells—particularly mammalian cells—a large fraction of the variability originates from intrinsic sources, such as stochastic fluctuations (noise) in transcription <sup>2,3</sup>. These intrinsic stochastic fluctuations can be quantitatively accounted for by the ‘toggling’ of genes between active and inactive expression periods leading to episodic bursts of transcription—commonly referred to as the two-state random-telegraph model of gene expression <sup>4-7</sup>. The resulting transcriptional bursts are amplified upon nuclear export and cytoplasmic mRNA processing <sup>8</sup>. Ultimately, these stochastic transcriptional fluctuations can generate substantial cellular variability, which has been implicated in cell-fate specification decisions ranging from HIV latency to cancer <sup>9-11</sup>.

Although a few specific roles of transcriptional noise have been elucidated, it remains unclear how broadly this phenomenon impacts generalized physiological processes, particularly in eukaryotic and mammalian systems. Isolating the physiological roles of expression noise would typically require one to modulate noise independent of the mean level of expression <sup>12</sup>. While this orthogonal perturbation of noise has been achieved in some notable cases <sup>13-15</sup>, in practice, it is non-trivial to implement. The complication stems in part from the fundamental linkage between the variance ( $\sigma^2$ ) and the mean ( $\mu$ ) for most physical processes. For example, the common metric for quantifying expression noise, the coefficient of variation (CV), defined as  $\sigma/\mu$ , typically scales inversely with mean, such that CV decreases as the mean increases <sup>16</sup>. Consequently, the normalized variance, or Fano factor ( $\sigma^2/\mu$ ) is often used to compare noise for processes with different mean values as it does not scale with the mean <sup>17,18</sup>. Regardless of the metric, orthogonal noise modulation remains challenging and generalized approaches to modulate noise would be useful physiological probes.

To break the  $1/\mu$  dependence of CV and orthogonally modulate noise, specific autoregulatory architectures (e.g., feedback and feedforward) are typically required <sup>19,20</sup>. However, small molecules called “noise enhancers” can also generate increased expression noise *without* altering the mean-expression level <sup>21</sup> by a process known as homeostatic noise amplification <sup>22</sup>; a notable

contrast to transcriptional activators, which increase mean-expression levels. We reported the molecular mechanism for one particular class of noise-enhancer molecule, pyrimidine-base analogs such as 5'-ido-2'-deoxyuridine (IdU), and used scRNA-seq to show that IdU increase noise. However, scRNA-seq, despite its utility in measuring genome-wide expression<sup>23</sup>, suffers from well-established issues of technical noise due to small inputs of RNA, amplification bias, and differences in capture ability<sup>24</sup> that could obscure the *penetrance* of IdU-mediated noise enhancement.

To better understand the genome-wide effect of IdU, here we reanalyzed scRNA-seq data using different normalization algorithms that minimize extrinsic and technical noise<sup>25</sup>. Each algorithm identified a different proportion of the genes exhibiting amplified noise and—in the fraction of genes with amplified noise—substantial differences in the magnitude of noise amplification. To validate scRNA-seq measurements, we employed single molecule-RNA FISH (smFISH)—the gold-standard for mRNA quantification due to its high sensitivity for mRNA detection<sup>26</sup>—to probe a panel of genes—from across the transcriptome—that displayed the greatest differences in noise between different scRNA-seq algorithms. Collectively, these analyses indicate that IdU is a globally penetrant noise-enhancer molecule and could, in principle, be a candidate to probe the physiological roles of expression noise for diverse genes of interest.

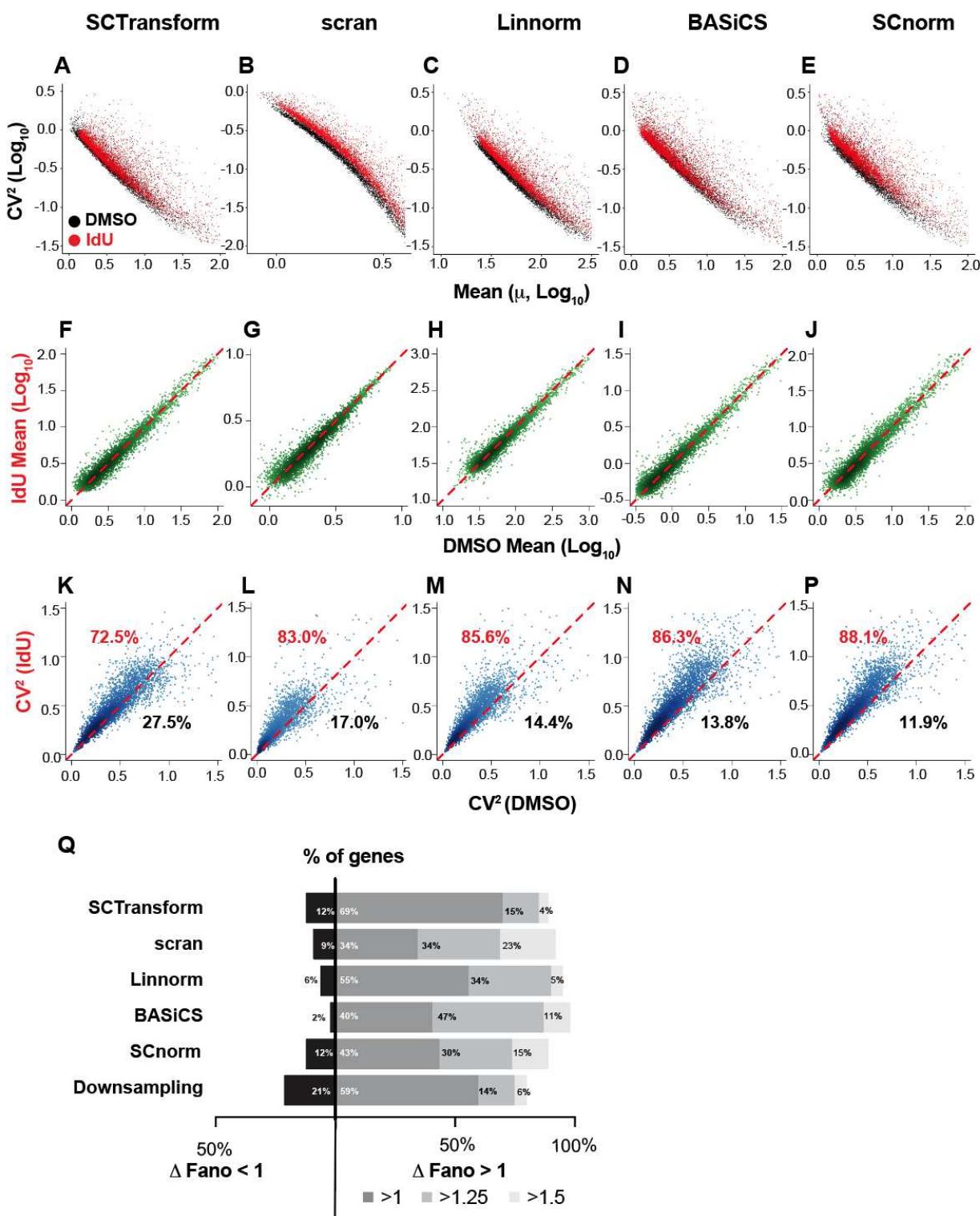
## **RESULTS**

### **Alternate scRNA-seq normalization algorithms generate differing profiles of expression noise indicating noise amplification ( $\Delta Fano > 1$ ) for ~90% of expressed genes**

To examine how different scRNA-seq normalization methods influence the quantification of transcriptional noise, we employed six commonly-used normalization algorithms to analyze scRNA-seq data from IdU-treated versus DMSO-treated (control) mouse embryonic stem cells (mESCs)<sup>22</sup>. In this dataset, cells were treated with the established noise-enhancer molecule IdU<sup>21,27</sup>, noise was quantified by examining  $CV^2$  relative to the mean, and quality control was performed using Seurat<sup>28</sup>, prior to all normalizations, to filter out low-expressing cells.

Initially, we attempted to account for differences in sequencing depth between IdU and control samples using a rudimentary random subsampling (or downsampling) algorithm<sup>29</sup> to normalize control and IdU-treated samples to the equal read depths. The rationale for random subsampling is based upon reports that lower sequencing depth can be associated with increased variance<sup>30</sup>, and subsampling reads to normalize the sequencing depth between samples can, in principle, correct for this. Surprisingly, downsampling suggested that IdU-mediated noise amplification was only partially penetrant, with ~78% of transcripts exhibiting amplified noise measured by Fano factor—this is compared to over 96% of transcripts exhibiting amplified noise when a standard log normalization without downsampling is used (**Fig. S1A-B**).

To determine if these subsampling results were consistent with other algorithms, we compared the results to five other established scRNA-seq normalization algorithms (**Fig 1**): *SCTransform*<sup>31</sup>, *scran*<sup>32</sup>, *Linnorm*<sup>33</sup>, *BASiCS*<sup>34</sup>, and *SCnorm*<sup>35</sup>. *SCTransform* is a commonly used normalization algorithm that accounts for changes in depth and includes a variance-stabilization transformation step. *scran* and *BASiCS* are widely used normalization methods that use “size factors” to eliminate technical noise while maintaining biological heterogeneity. *Linnorm* utilizes homogenously expressed genes and scales the reads accordingly. *SCnorm*, similar to *SCTransform*, groups genes based on count-depth relationships and uses quantile regression to generate normalization factors.



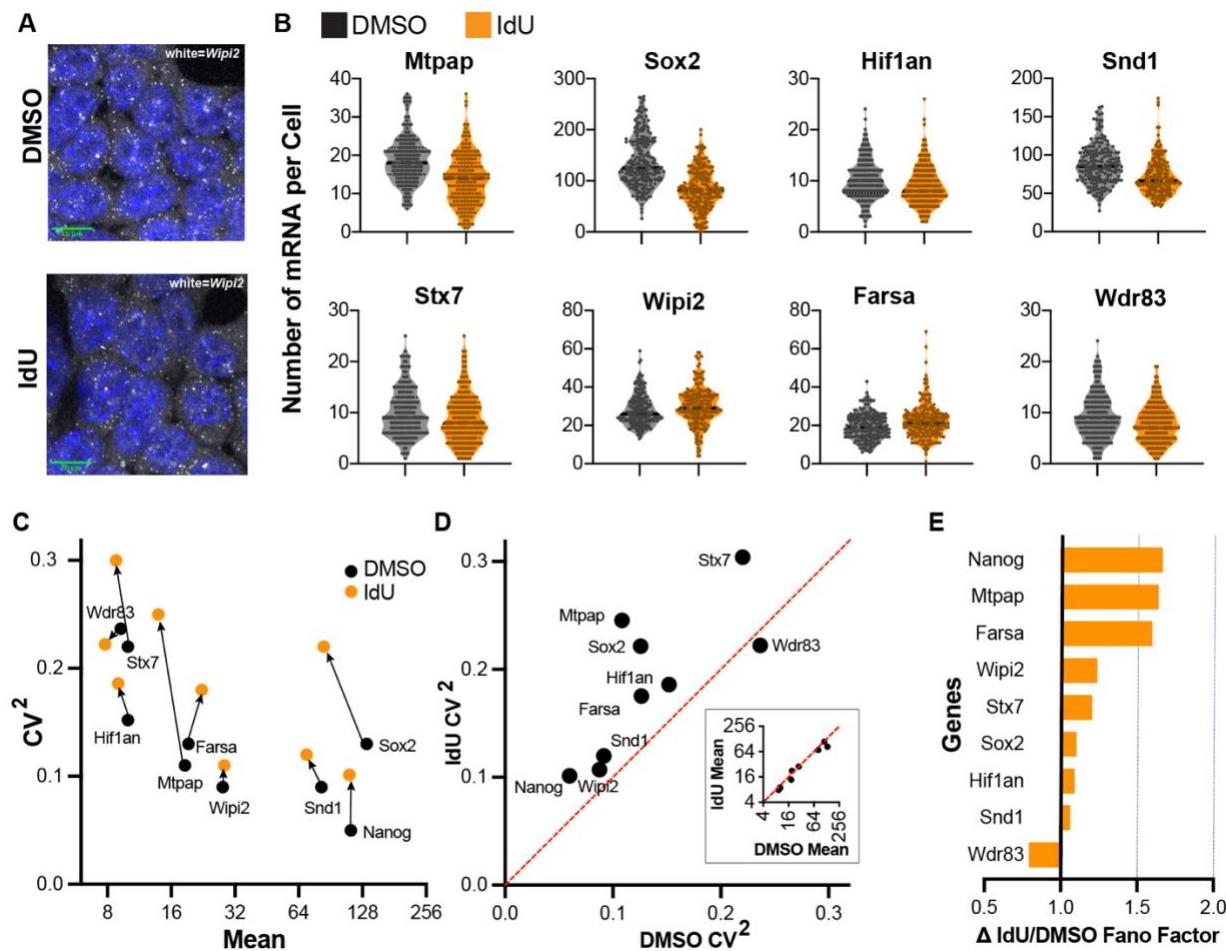
**Figure 1: Common scRNA-seq normalization algorithms generate different quantifications of mRNA noise.** scRNA-seq analysis of  $CV^2$ -vs-mean for ~5,000 transcripts in mESCs treated with IdU (red) or DMSO control (black) as analyzed by commonly used normalization algorithms: (A) SCTransform, (B) scran, (C) Linnorm, (D) BASiCS, or (E) SCnorm. (F–J) Mean expression for each of the ~5,000 transcripts in presence and absence of IdU using each normalization algorithm; none of the normalization algorithms generate substantial changes in mean expression for IdU-treated cells. (K–P)  $CV^2$  for each transcript in presence and absence of IdU using each normalization algorithm; different algorithms generate substantially different fractions of transcripts with amplified noise ranging from ~70% of transcripts with amplified noise (SCTransform) to ~88% of transcripts with amplified noise (SCnorm). (Q) Quantification showing percentages of transcripts with Fano factor fold change ( $\Delta$ ) <1, >1, >1.25 or >1.5 with IdU treatment. Calculations from all scRNA-seq normalization algorithm are shown (see Fig. S1 for Downsampling analysis).

Despite their substantially different technical schemes, analysis with each of these normalizations indicated that IdU induces a substantial amplification of noise ( $CV^2$ ) for most expressed genes (**Fig. 1A–E**). The IdU-induced noise amplification appeared to be homeostatic (**Fig. 1F–J**), with mean-expression levels largely unchanged by IdU under all algorithms. However, each algorithm calculated a somewhat different percentage of expressed genes with increased  $CV^2$  ranging from 72% to 88% of genes exhibiting increased noise (**Fig. 1K–P**). Quantification of noise by analysis of the normalized variance (i.e., Fano factor), which in principle eliminates the scaling dependence on mean-expression level, showed a similar profile (**Fig. 1Q**). Whilst every algorithm showed a majority of expressed genes exhibit amplified noise by Fano factor, each calculated a different penetrance (i.e., percentage of noise-amplified genes) as well as substantial differences in the magnitude of noise amplification among those transcripts with amplified noise.

Overall, scRNA-seq analysis by all normalization algorithms showed that IdU homeostatically amplifies noise for the majority of expressed genes without observable changes in the mean-expression level. However, each scRNA-seq normalization algorithm calculated a different transcriptional *penetrance* for IdU noise amplification, indicating that additional analysis of transcriptional noise via an independent (i.e., non-scRNA-seq approach) was warranted.

### **RNA quantification by smFISH indicates widespread penetrance of IdU-induced noise amplification**

To directly quantify RNA levels in individual cells using a non-sequencing-based approach, we employed smFISH, a well-established imaging method to assess noise and the gold standard for quantitative assessment of mRNA expression <sup>36</sup>. While smFISH enables quantification of mRNA abundance in individual cells, it is a relatively low-throughput method requiring distinct fluorescent probes and image quantification for each transcript species of interest. Consequently, to generate a pseudo-global view of transcription using smFISH, we selected a panel of eight genes (**Fig. S1**) that satisfied two criteria: (i) they displayed the greatest difference in noise amplification as calculated by the different RNA-seq algorithms (**Fig. S1C–D**), and (ii) the genes were compatible with smFISH probe design (i.e., a minimum of 30 probes were predicted to hybridize to the transcript). The selected genes spanned a wide range of gene-expression levels (i.e., across 2-Logs and genome locations) (**Fig. S1E**) <sup>37</sup>. We also included previously reported smFISH data <sup>22</sup> from a ninth gene, *Nanog*, to



**Figure 2:** smFISH analysis of IdU-induced noise amplification for a subset of genes from across the transcriptome. **(A)** Representative smFISH images of *Wipi2* transcripts (white dots) in mESCs treated with DMSO (top) or IdU (bottom). Nuclei are stained with DAPI (blue). **(B)** Distribution of per-cell mRNA transcripts for eight representative genes +/- IdU. **(C)** CV<sup>2</sup>-vs.-mean analysis as calculated from smFISH +/- IdU for each of the genes in the representative panel. **(D)** Direct comparison of CV<sup>2</sup> +/- IdU as calculated from smFISH data. **Inset:** direct comparison of IdU-induced change in mean mRNA from smFISH data. **(E)** Fold change in Fano Factor for IdU-treated samples compared to DMSO control from smFISH data.

benchmark the smFISH analysis. smFISH probe sets for each gene in the panel were generated, cells were imaged in the presence/absence of IdU, and images (Fig. 2A, S2A) segmented and analyzed using FishQuant<sup>38</sup>. Extrinsic noise was filtered by analyzing cell size (i.e., cell-area distribution in pixel number) by Kolmogorov-Smirnov test and excluding cells in the upper and lower tails of cell-size distribution (Fig. S2B). For each gene, the per-cell mRNA abundance was quantified for ~200 cells in order to generate the distribution of each mRNA per cell, in the presence and absence of IdU (Fig. 2B). This analysis revealed greater variance in per-cell mRNA levels in IdU-treated samples than controls, and the calculated CV<sup>2</sup> values verified that noise increased for most genes (Fig. 2C). Direct comparison of mRNA CV<sup>2</sup> values showed significant noise amplification for 8 of the 9 genes in the representative panel, without substantial changes in

mean (**Fig. 2D, S3A**), in agreement with the scRNA-seq analysis. Notably, *Sox2* appeared to be an exception that exhibited a reduction in mean expression (see Discussion).

To ensure that noise amplification, as reported by  $CV^2$ , could not be explained by changes in mean expression, we also analyzed the Fano factor calculated from the smFISH data (**Fig. 2E, S3B**). Fano factor analysis verified that all but one gene in the panel (*Wdr83*) exhibited IdU-induced amplification of noise ( $\Delta$ Fano factor  $> 1$ ). The greatest increased in noise was for *Nanog*, *Mtpap*, and *Farsa* ( $\Delta$ Fano factor  $> 1.5$ ) whereas *Sox2*, *Snd1*, *Stx7*, *Hiflan*, and *Wipi2* exhibited smaller noise amplifications.

To further verify IdU-mediated noise amplification, we tested if the mechanistic underpinnings<sup>21</sup> of homeostatic noise amplification were satisfied. Theory predicts that homeostatic noise amplification requires reciprocal changes in transcriptional bursting parameters<sup>21</sup>; i.e., a decrease in transcriptional burst frequency with a corresponding but reciprocal increase in transcriptional burst size. Transcriptional burst frequency can be calculated by scoring the percentage of cells with active transcriptional centers (TCs) in smFISH images, and the burst size can be similarly calculated if other parameters, such as the decay rate of the specific mRNA, are also known<sup>39</sup>. Consequently, we calculated burst frequency and burst size for two genes, *Sox2* and *Mtpap*, for which mRNA decay rates in mouse embryonic stem cells were known<sup>40</sup>. In agreement with the model of homeostatic transcriptional noise amplification, IdU substantially decreased the burst frequency (**Fig. S4A**) and correspondingly increased the burst size (**Fig. S4B**) for both genes.

Overall, these smFISH data indicate that IdU generates a homeostatic amplification of transcriptional noise for 8 out of 9 genes in the representative panel taken from diverse expression profiles and locations in the genome.

## **DISCUSSION**

This study set out to determine if the pyrimidine base analog IdU, which is incorporated into cellular DNA and removed by the cellular base-excision repair (BER) surveillance machinery<sup>41</sup>, is a globally penetrant noise-enhancer molecule. Analysis of scRNA-seq data using six different normalization algorithms indicated that the amplification of noise by IdU is globally penetrant and

likely not an artifact of a particular scRNA-seq analysis algorithm. However, the scRNA-seq analysis also indicated that scRNA-seq algorithms generate variable genome-wide noise profiles (**Fig. 1**). To validate IdU's genome-wide effect, we curated a panel of genes that exhibited high sensitivity to individual scRNA-seq normalization algorithms and represented both high and low expressing genes for downstream orthogonal analysis of transcriptional noise by smFISH. This approach revealed that IdU increased transcriptional noise for 8 out of 9 genes (**Fig. 2**) and that most published scRNA-seq algorithms (with the exception of rudimentary downsampling) were fairly accurate compared to the smFISH direct measurement (compare **Fig. S1C** to **Fig. 2E**).

One technical limitation of this study is that smFISH analysis is necessarily low throughput and limited to the subset of genes for which good probe sets can be designed, which limits the spectrum of measurements and the number of genes that can be analyzed. We attempted to mitigate this by exploring genes from across the expression spectrum (**Fig. 2C**), and by analyzing a substantial number of cells per treatment (>170 cells per treatment **after** filtering extrinsic noise). Thus, our data indicate that IdU likely acts as a noise-enhancer molecule for a large fraction of genes irrespective of their mean-expression level.

Notably, our smFISH analysis did not detect an increase in noise for *Wdr83*. Considering that IdU increases noise via BER, a genome-wide surveillance pathway, it may be surprising that any gene fails to exhibit an increase in noise. It is possible that the low-throughput nature of smFISH may have obscured IdU-induced noise amplification. However, a potential explanation for the selective response lies in the putative mechanism by which BER increases noise via DNA topology changes. Specifically, the BER enzyme AP endonuclease 1 (Apex1) generates DNA supercoiling which leads to an accumulation of RNA Polymerase II; when released, this amplifies transcriptional burst size; indeed, the IdU noise effect can be phenocopied by topoisomerase inhibitors, which increase DNA supercoiling <sup>22</sup>. Consequently, regions which naturally have increased supercoiling and corresponding high levels of topoisomerase may be less sensitive to changes in topology caused by IdU and BER- induced supercoiling. It has been previously reported that Topologically Associated Domain (TAD) boundaries are regions that exhibit substantial supercoiling and are enriched in insulator binding protein CTCF <sup>42,43</sup>, and that Topoisomerase IIB is prevalent at CTCF sites <sup>44</sup>. Intriguingly, *Wdr83* has two relatively unique features among the panel of genes tested: (i) it contains a relatively large CTCF binding region, and (ii) it is found at a TAD boundary <sup>42</sup>.

Together these two features may be consistent with IdU acting as a topology-dependent global noise-enhancer molecule, as shown by scRNA-seq (**Fig. 1**) and smFISH (**Fig. 2**) analysis.

A second point of interest is the case of *Sox2*. Whilst most genes analyzed exhibited a homeostatic increase in noise—without a substantial change in mean—smFISH revealed that IdU induced a decrease in mean *Sox2* mRNA, though we note the change in mRNA numbers was less than two-fold. Notably, IdU did not alter single-cell *Sox2* protein levels in our previous analysis <sup>22</sup>. It is possible that the reported negative-feedback regulation of *Sox2* <sup>45</sup> acts to buffer changes at the protein level.

From the practical perspective of quantifying expression noise, this study reveals that common analyses can fail to resolve quantitative changes in intrinsic expression noise, particularly for individual genes. A number of approaches have been proposed to overcome these types of scRNA-seq limitations, including elegant solutions using mathematical modeling methods to address technical variability without compromising quantification of biological noise <sup>46</sup>, though these can be cumbersome to implement. The analyses herein indicate that it may be advisable to combine high-throughput analyses (e.g., scRNA-seq) with lower-throughput direct quantification (e.g., smFISH) to quantify changes in transcriptional noise for individual genes as the “ground-truth” may lie somewhere in between the results of each analysis. Regardless, both the scRNA-seq and smFISH data argue that IdU appears to be a global noise enhancer which could be leveraged to modulate noise without altering mean expression for the majority of genes.

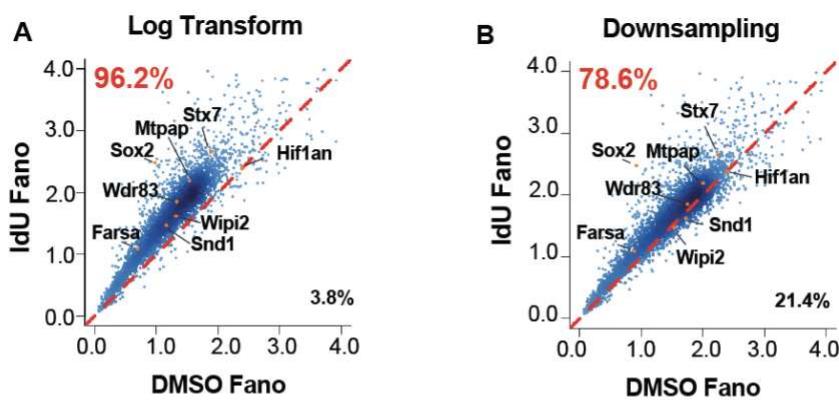
**Acknowledgements:** We thank L. Miranda for administrative support; R. Desai, G. Vasen, K. Delucas, and D. Lewis for technical guidance. Harrington and S. Kim for imaging guidance; K. Claiborn for editing and O. Rechavi for subsampling guidance. Data for this study were acquired at the Nikon Imaging Center at UCSF using NIH S10 Shared Instrumentation grant (1S10OD017993-01A1).

**Funding:** This work was supported by NIH award R37AI109593 (to LSW), BZ acknowledges support from the EMBO fellowship (ALTF 388-2021) and XC acknowledges support from CIRM training grant EDUC4-12766.

**Author contributions:** GC, XC, and LSW conceived and designed the study. GC and XC designed the smFISH experiments. GC performed, analyzed, and curated the data. GC and BZ performed single-cell Seq normalizations and GC curated the data. LSW provided reagents and resources. GC and LSW wrote the paper.

**Declaration of Interests:** No conflicting interests to declare

## SUPPORTING FIGURES



**C Fold Change in Fano**

	Log Transform	Downsampling	BASiCS	SCTransform	Linnorm	SCnorm	scran
<i>Wdr83</i>	1.41	1.05	1.11	1.13	1.32	1.31	1.41
<i>Mtpap</i>	1.43	1.10	1.10	1.28	1.39	1.12	1.42
<i>Farsa</i>	1.57	1.23	1.13	1.23	1.32	1.44	1.45
<i>Hif1an</i>	1.26	1.02	1.04	1.08	1.28	0.92	1.09
<i>Wipi2</i>	1.28	0.96	1.04	1.13	1.23	1.23	1.35
<i>Stx7</i>	1.39	1.18	1.21	1.19	1.42	1.47	1.38
<i>Snd1</i>	1.23	0.94	1.01	1.01	1.17	0.98	1.27
<i>Sox2</i>	2.81	2.69	1.42	1.50	1.76	1.81	3.87
<i>Nanog</i>	2.40	2.27	1.78	1.88	2.03	2.05	2.81

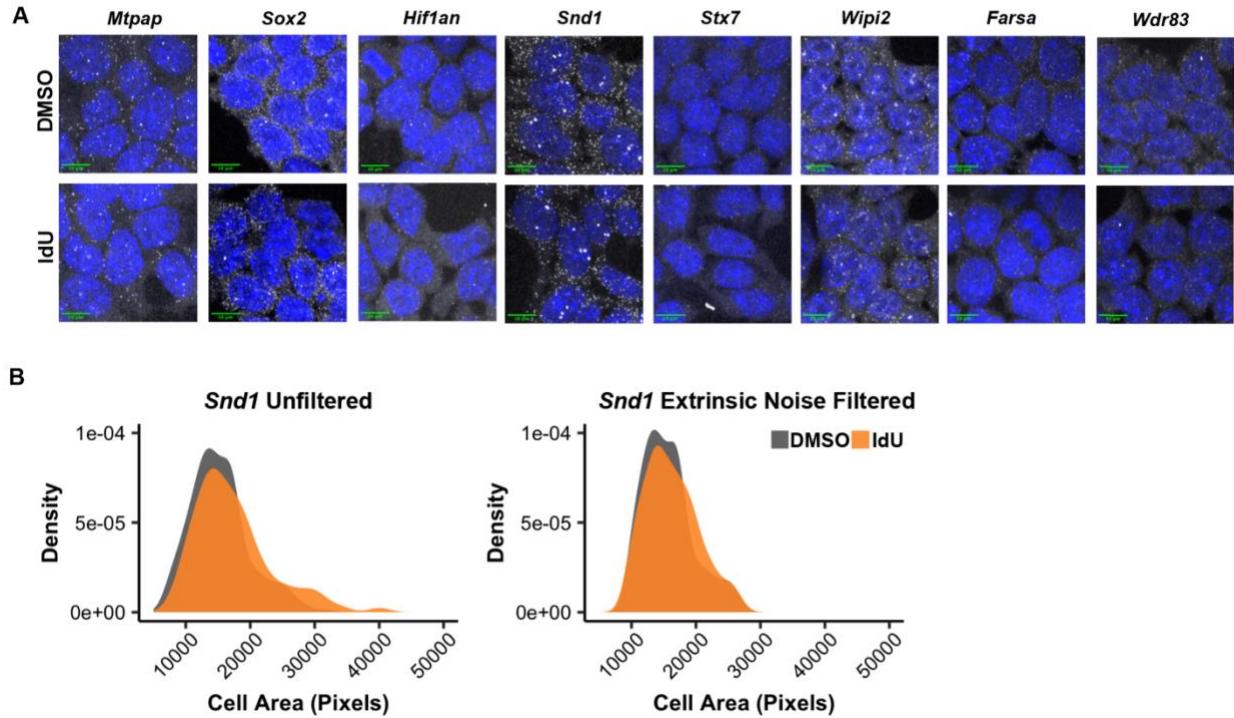
**D Fold Change in CV<sup>2</sup>**

	Log Transform	Downsampling	BASiCS	SCTransform	Linnorm	scNorm	scran
<i>Wdr83</i>	1.52	1.11	1.45	1.26	1.41	1.46	1.55
<i>Mtpap</i>	1.51	1.09	1.33	1.29	1.40	1.44	1.49
<i>Farsa</i>	1.62	1.26	1.38	1.22	1.39	1.38	1.47
<i>Hif1an</i>	1.41	1.06	1.33	1.21	1.41	1.19	1.44
<i>Wipi2</i>	1.30	0.96	1.24	1.12	1.26	1.26	1.37
<i>Stx7</i>	1.45	1.07	1.31	1.27	1.39	1.35	1.36
<i>Snd1</i>	1.33	0.98	1.29	1.07	1.23	1.26	1.34
<i>Sox2</i>	4.06	3.89	2.92	2.56	2.54	2.84	5.11
<i>Nanog</i>	3.00	2.81	2.79	2.43	2.59	2.67	3.32

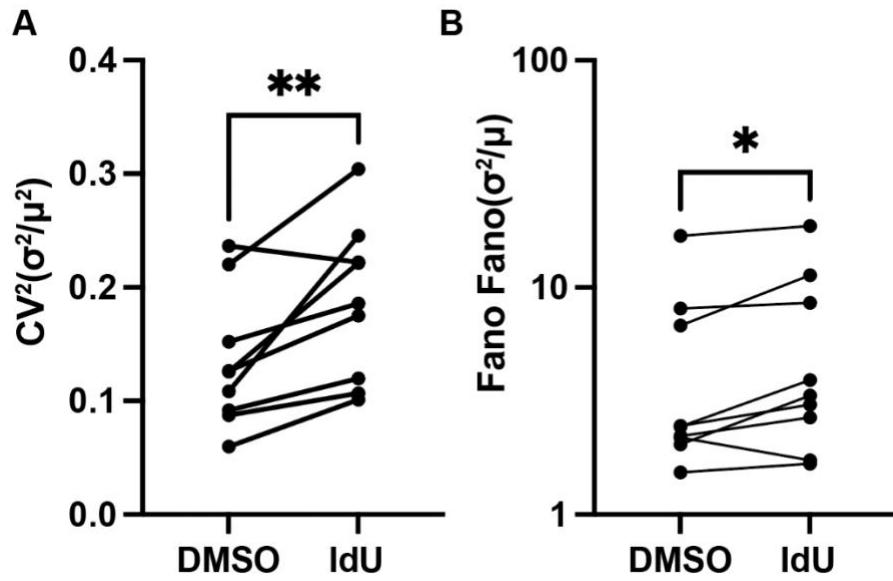
**E Genomic Locations**

<i>Wdr83</i>	<i>Mtpap</i>	<i>Farsa</i>	<i>Hif1an</i>	<i>Wipi2</i>	<i>Stx7</i>	<i>Snd1</i>	<i>Sox2</i>	<i>Nanog</i>
chr8:85,075,035	chr18:4,375,592	chr8:84,856,989	chr19:44,562,850	chr5:142,629	chr10:24,149,327	chr6:28,480,359	chr3:34,650,405	chr6:122,707,568

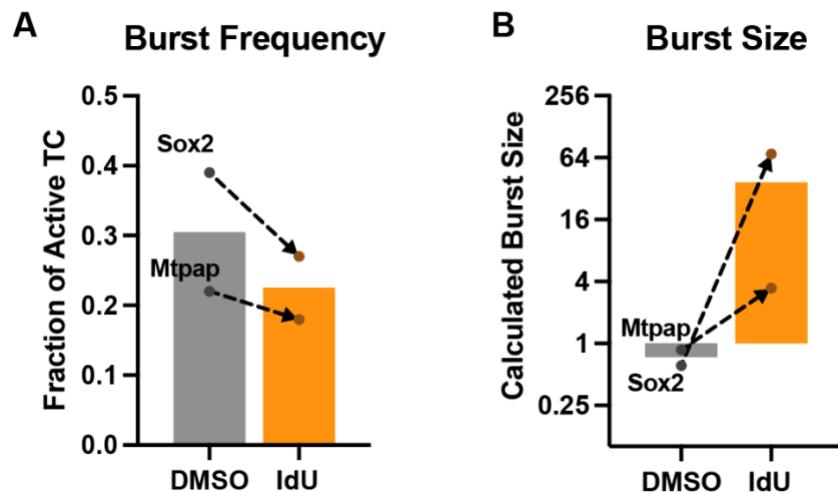
**Figure S1: Single Cell RNA Sequencing algorithms show varying penetrance of IdU-induced amplification of expression noise.** Related to Figure 1. Direct comparison of Fano value for different single cell sequencing normalizations. The 8 genes chosen for smFISH are labeled in (A) Data used from Desai et al., 2021. (B) Downsampling algorithm that matches DMSO sequencing depth with IdU sequencing depth using random subsampling. Table of values showing the fold change of IdU vs control for Fano (C) and CV<sup>2</sup> (D) for the 8 genes selected for smFISH (together with *Nanog* from Desai et al. 2021) across the seven normalization algorithms. (E) Genomic locations of genes selected for smFISH.



**Figure S2: Representative images and extrinsic-noise filtering (by cell-size) for smRNA FISH.** Related to Figure 2. (A) Representative smFISH images of mESCs treated with DMSO (top) or IdU (bottom) for 24 hours with DAPI stain (blue) and mRNA transcripts (gray) fluorescent labeled with TAMRA probe-set. Max intensity projections are shown. (B) Left: The cell size distribution of 194 DMSO treated cells and 196 IdU treated cells are significantly different (KS test  $p = 0.007$ ). Right: the 5th and 95th percentile of the whole population were removed. The cell size distribution for the remaining 178 DMSO Treated cells and 174 IdU treated cells are not significantly different (KS test  $p=0.21$ )



**Figure S3: smRNA FISH shows significantly increased Fano Factor and  $CV^2$  with IdU Treatment.** (A) Fano Factor comparison of samples +/- IdU compared using Wilcoxon matched-pairs signed rank test  $p = 0.0117$ . (B)  $CV^2$  comparison of samples +/- IdU compared using Wilcoxon matched-pairs signed rank test  $p = .0078$



**Figure S4: Changes in burst frequency and burst size with IdU treatment.** (A) Transcriptional burst frequency for *Mtpap* and *Sox2* as calculated from the fraction of active transcription centers (TC) in smFISH images. (B) Calculated transcriptional burst size for *Mtpap* and *Sox2* (see Methods).

# Captions for supplementary table S1

## **Table S1** (attached separately)

Sequences of smRNA-FISH oligonucleotide probe for eight genes

## Data Availability

This study includes no new data deposited in external repositories.

## Star Methods

### Single-Cell RNA Sequencing Analysis

The scRNA-seq count matrix available at GSE176044\_mesc\_bulk\_rnaseq\_gene\_counts was used (Data ref: Desai *et al*, 2021). Before applying each normalization method, both DMSO and IdU datasets were subjected to a quality control process. First, Seurat <sup>28</sup> was used to filter for high-quality cells using a minimum of 4000 detected genes, 10000 UMI counts, and <10% reads mapping to mitochondrial genes per cell.. The resulting count matrix was then further filtered by the BASiCS\_Filter function from the BASiCS <sup>34</sup> R package with default parameters, which limited the analysis to genes with sufficient sequencing coverage for reliable noise quantification. The output count matrix consisted of 811 cells for DMSO and 732 cells for IdU across 4456 genes. This output was then used to run 5 normalization pipelines according to their protocols: *SCTtransform* <sup>31</sup>, *scran* <sup>32</sup>, *Linnorm* <sup>33</sup>, *BASiCS* <sup>34</sup>, and *SCnorm* <sup>35</sup>. BASiCS was run with default parameters and recommended settings (N=20000, Thin=20 and Burn=10000) using the horizontal integration strategy (no-spikes). The other packages were run with default parameters. To perform rudimentary downsampling (**Fig. S1B**) the python random.seed function was used to simulate an identical sequencing depth for both DMSO and IdU. 81% of DMSO reads were sampled, and 100% of IdU reads were subsampled. From this subsampled matrix, the same 4445 genes and 1543 cells included previously were used for further quantification steps.

### Single Molecule RNA FISH

Probes for the detection of transcripts were developed using the designer tool from Stellaris (LGC Biosearch Technologies) (Table S1) setting the minimum number of probes to 30 (TAMRA conjugated) for gene transcripts.  $1 \times 10^5$  mouse embryonic stem cells were seeded into each well of

a gelatin-coated, 8-well Ibidi dish (cat: 80826) in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF containing 10mM IdU or equivalent volume DMSO. After 24 hours of treatment, cells were then fixed with DPBS in 4% paraformaldehyde for 10 minutes. Fixed cells were washed with DPBS and stored in 70% EtOH at 4°C for one hour to permeabilize the cell membranes. Probes were diluted 200-fold and allowed to hybridize at 37°C overnight. Wash steps and DAPI (Thermo, cat: D1306) staining were performed as described (<https://www.biosearchtech.com/support/resources/stellaris-protocols>). To minimize photo-bleaching, cells were imaged in a buffer containing 50% glycerol (Thermo, cat: 17904), 75 mg/mL glucose oxidase (Sigma Aldrich, cat: G7141), 520 mg/mL catalase (Sigma Aldrich, cat: C3515), and 0.5 mg/mL Trolox (Sigma Aldrich, cat: 238813). Images were collected on an inverted Nikon TiE microscope (Nikon) run using Micromanager 2.0<sup>47</sup> equipped with a CSU-W1 Spinning Disk with Borealis Upgrade (Yokogawa, Andor), ILE Laser launch with 4 laser lines (450/488/561/646nm, Andor), quad-band dichroic ZT405/488/561/647 (Chroma), emission filters for DAPI (ET447/60), GFP (ET525/50), RFP (ET607/36), and Cy5 (ET685/40) (Chroma), piezo XYZ stage (ASI), and Zyla 4.2 CMOS camera (Andor), using a Plan Apo VC 60x/1.4 Oil objective (Nikon). Approximately 10 XY regions of interest were randomly selected for each condition. For each image, XY pixel size was 108nm/px, and a Z-step size of 250nm was used with over 60 image planes to fully cover the tissue. Image analysis and spot counting was performed using FISH-quant<sup>38</sup>. Cells were manually segmented and analysis was conducted on cells of a similar size to minimize extrinsic noise. Active transcription sites were measured according to<sup>39</sup>. Burst size was calculated as previously reported<sup>8</sup> using the following equation:

$$\text{Burst Size} = (\lambda_{\text{mRNA}} * \mu_{\text{mRNA}}) / k_{\text{off}} \text{ where } k_{\text{off}} = 1 - (\% \text{ of active TC}).$$

## REFERENCES

1. Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* *135*, 216–226. 10.1016/j.cell.2008.09.050.
2. Blake, W.J., KÆrn, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature* *422*, 633–637. 10.1038/nature01546.
3. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol* *4*, e309. 10.1371/journal.pbio.0040309.
4. 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. *Molecular Cell* *24*, 853–865. 10.1016/j.molcel.2006.11.003.
5. Harper, C.V., Finkenstädt, B., Woodcock, D.J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D.G., Mullins, J.J., Rand, D.A., Davis, J.R.E., et al. (2011). Dynamic Analysis of Stochastic Transcription Cycles. *PLoS Biol* *9*, e1000607. 10.1371/journal.pbio.1000607.
6. Kepler, T.B., and Elston, T.C. (2001). Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophysical Journal* *81*, 3116–3136. 10.1016/S0006-3495(01)75949-8.
7. Peccoud, J., and Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology* *48*, 222–234. 10.1006/tpbi.1995.1027.
8. Hansen, M.M.K., Desai, R.V., Simpson, M.L., and Weinberger, L.S. (2018). Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Systems* *7*, 384-397.e6. 10.1016/j.cels.2018.08.002.
9. Balázsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* *144*, 910–925. 10.1016/j.cell.2011.01.030.
10. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* *546*, 431–435. 10.1038/nature22794.
11. Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* *40*, 466–470. 10.1038/ng.116.
12. Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using Gene Expression Noise to Understand Gene Regulation. *Science* *336*, 183–187. 10.1126/science.1216379.
13. Bonny, A.R., Fonseca, J.P., Park, J.E., and El-Samad, H. (2021). Orthogonal control of mean and variability of endogenous genes in a human cell line. *Nat Commun* *12*, 292. 10.1038/s41467-020-20467-8.
14. Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* *463*, 913–918. 10.1038/nature08781.

15. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat Genet* *31*, 69–73. 10.1038/ng869.
16. Sanchez, A., and Golding, I. (2013). Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science* *342*, 1188–1193. 10.1126/science.1242975.
17. Kærn, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* *6*, 451–464. 10.1038/nrg1615.
18. Olsman, N., Xiao, F., and Doyle, J. (2018). Evaluation of Hansen et al.: Nuance Is Crucial in Comparisons of Noise. *Cell Systems* *7*, 352–355. 10.1016/j.cels.2018.10.003.
19. Hansen, M.M.K., Wen, W.Y., Ingberman, E., Razooky, B.S., Thompson, C.E., Dar, R.D., Chin, C.W., Simpson, M.L., and Weinberger, L.S. (2018). A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. *Cell* *173*, 1609–1621.e15. 10.1016/j.cell.2018.04.005.
20. Zhang, H., Chen, Y., and Chen, Y. (2012). Noise Propagation in Gene Regulation Networks Involving Interlinked Positive and Negative Feedback Loops. *PLoS ONE* *7*, e51840. 10.1371/journal.pone.0051840.
21. Dar, R.D., Hosmane, N.N., Arkin, M.R., Siliciano, R.F., and Weinberger, L.S. (2014). Screening for noise in gene expression identifies drug synergies. *Science* *344*, 1392–1396. 10.1126/science.1250220.
22. Desai, R.V., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H., et al. (2021). A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* *373*, eabc6506. 10.1126/science.abc6506.
23. Kim, J.K., Kolodziejczyk, A.A., Ilicic, 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. 10.1038/ncomms9687.
24. Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* *10*, 1093–1095. 10.1038/nmeth.2645.
25. Chu, S.-K., Zhao, S., Shyr, Y., and Liu, Q. (2022). Comprehensive evaluation of noise reduction methods for single-cell RNA sequencing data. *Briefings in Bioinformatics* *23*, bbab565. 10.1093/bib/bbab565.
26. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* *5*, 877–879. 10.1038/nmeth.1253.
27. Li, Y., Shan, Y., Desai, R.V., Cox, K.H., Weinberger, L.S., and Takahashi, J.S. (2020). Noise-driven cellular heterogeneity in circadian periodicity. *Proc. Natl. Acad. Sci. U.S.A.* *117*, 10350–10356. 10.1073/pnas.1922388117.

28. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* *184*, 3573-3587.e29. 10.1016/j.cell.2021.04.048.
29. Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* *15*. 10.15252/msb.20188746.
30. Rizzetto, S., Eltahla, A.A., Lin, P., Bull, R., Lloyd, A.R., Ho, J.W.K., Venturi, V., and Luciani, F. (2017). Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. *Sci Rep* *7*, 12781. 10.1038/s41598-017-12989-x.
31. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* *20*, 296. 10.1186/s13059-019-1874-1.
32. L. Lun, A.T., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* *17*, 75. 10.1186/s13059-016-0947-7.
33. Sengupta, D., Rayan, N.A., Lim, M., Lim, B., and Prabhakar, S. (2016). Fast, scalable and accurate differential expression analysis for single cells (Genomics) 10.1101/049734.
34. Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLoS Comput Biol* *11*, e1004333. 10.1371/journal.pcbi.1004333.
35. Bacher, R., Chu, L.-F., Leng, N., Gasch, A.P., Thomson, J.A., Stewart, R.M., Newton, M., and Kendziora, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. *Nat Methods* *14*, 584–586. 10.1038/nmeth.4263.
36. Torre, E., Dueck, H., Shaffer, S., Gospocic, J., Gupte, R., Bonasio, R., Kim, J., Murray, J., and Raj, A. (2018). Rare Cell Detection by Single-Cell RNA Sequencing as Guided by Single-Molecule RNA FISH. *Cell Systems* *6*, 171-179.e5. 10.1016/j.cels.2018.01.014.
37. Nassar, L.R., Barber, G.P., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., Lee, C.M., Muthuraman, P., et al. The UCSC Genome Browser database: 2023 update. *8*.
38. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat Methods* *10*, 277–278. 10.1038/nmeth.2406.
39. Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., and Darzacq, X. (2014). Transcription Factors Modulate c-Fos Transcriptional Bursts. *Cell Reports* *8*, 75–83. 10.1016/j.celrep.2014.05.053.
40. Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H. (2009). Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. *DNA Research* *16*, 45–58. 10.1093/dnare/dsn030.

41. Lindahl, T., and Wood, R.D. (1999). Quality Control by DNA Repair. *Science* *286*, 1897–1905. 10.1126/science.286.5446.1897.
42. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* *485*, 376–380. 10.1038/nature11082.
43. Nanni, L., Ceri, S., and Logie, C. (2020). Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol* *21*, 197. 10.1186/s13059-020-02108-x.
44. Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M.V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E.J., et al. (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol* *17*, 182. 10.1186/s13059-016-1043-8.
45. Ormsbee Golden, B.D., Wuebben, E.L., and Rizzino, A. (2013). Sox2 Expression Is Regulated by a Negative Feedback Loop in Embryonic Stem Cells That Involves AKT Signaling and FoxO1. *PLoS ONE* *8*, e76345. 10.1371/journal.pone.0076345.
46. Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat Methods* *11*, 637–640. 10.1038/nmeth.2930.
47. Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using  $\mu$ Manager software. *J Biol Methods* *1*, e10. 10.14440/jbm.2014.36.