Supplementary Materials for

A DNA-repair pathway can affect transcriptional noise to promote cell fate transitions

Ravi V. Desai, Xinyue Chen, Benjamin Martin, Sonali Chaturvedi, Dong Woo Hwang, Weihan Li, Chen Yu, Sheng Ding, Matt Thomson, Robert H. Singer, Robert A. Coleman, Maike M. K. Hansen, Leor S. Weinberger*

*Corresponding author. Email: leor.weinberger@gladstone.ucsf.edu

Published 22 July 2021 on *Science* First Release
DOI: 10.1126/science.abc6506

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/cgi/content/full/science.abc6506/DC1)

Tables S1 to S8 (Excel)
MDAR Reproducibility Checklist (PDF)

Materials and Methods

Cell Culture and Growth Conditions

Mouse E14 embryonic stem cells (male) were routinely cultured in feeder-free conditions on gelatin-coated plates with ESGRO-2i medium (Millipore, cat:SF016-200) at 37°C, 5% CO_2, in humidified conditions *(28)*. Jurkat T Lymphocytes (male) were cultured in RPMI-1640 medium (supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin), at 37°C, 5% CO_2, in humidified conditions at 0.1×10^6 to 1×10^6 cells/mL. Isoclonal Jurkat T Lymphocytes with lentivirally integrated EF1α-d_2GFP construct were previously described *(28)*. Human immortalized myelogenous leukemia (K652, female) cells were cultured in RPMI-1640 medium (supplemented with L-glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin), at 37°C, 5% CO_2, in humidified conditions at 2×10^5 to 2×10^6 cells/mL. Isoclonal K562 cells with lentivirally integrated EF1α-d_2GFP and UBC- d_2GFP constructs were previously described *(28)*. Human embryonic kidney (HEK293, female) cells were cultured in DMEM (supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin, 25mM HEPES, and 2mM L-glutamine) at 37°C, 5% CO_2, in humidified conditions at 30 to 90% confluency.

Noise Enhancer Testing on Isoclonal Jurkat and K562 Cells

Jurkat and K562 cells were seeded into 12-well plates at densities of 0.2×10^6 and 0.4×10^6 cells/mL respectively in media containing 20 μM IdU (Sigma, cat:I7125, dissolved in DMSO) or equivalent volume of DMSO for 24 hours. Flow cytometry was performed using a BD LSRII cytometer. Treated cells were run unfixed and live to avoid additional sources of variability from fixation. 50k live cells were collected from each sample for noise measurements. Conservative gating for a live subset of approximately 3k cells of similar size, volume, and state, was applied on the FSC vs. SSC to reduce extrinsic noise contributions as previously described *(22,24)*.

Single-Cell RNA Sequencing Preparation and Analysis

$1x10^6$ mESCs were seeded in a gelatin-coated, 10cm dish in 2i/LIF media. 24 hours following seeding, cultures were replenished with 2i/LIF media containing 10μM IdU or an equivalent volume of DMSO for 24 hours. After treatment, cells were trypsinized with TrypLE and spun down for 5 minutes at 200 x g. Single-cell suspensions were prepared in DPBS at a concentration of 83,000 cells/ml. Approximately 3000 cells from each sample were loaded into a chip and processed with the Chromium Single Cell Controller (10x Genomics). To generate single-cell gel beads in emulsion (GEMs), DMSO- and IdU- treated samples were assigned unique indexes using Single Cell 3′ Library and Gel Bead Kit V2 (10x Genomics, cat:120237). Sequencing was performed on an Illumina HiSeq4000 with a paired-end setup specific for 10x libraries.

Data were aligned to mm10 reference genome using 10x Cell Ranger v2. Quality control, normalization and analysis were carried out using two packages: Seurat and BASiCS. For analysis

in Seurat, gene-barcode matrices were filtered and normalized using the "LogNormalize" method, resulting in 812 and 744 transcriptomes from DMSO and IdU samples. Transcript variability was quantified using variance (σ^2), coefficient of variation ($\frac{\sigma^2}{\mu^2}$), and Fano factor ($\frac{\sigma^2}{\mu}$). During the normalization procedure in Seurat, counts for the i^{th} gene in the j^{th} cell (x_{ij}) are multiplied by the following scaling factor: $S = \frac{10000}{\sum_{i=1}^{n} x_{ij}}$, where n is the number of genes in the dataset. The scaling factor is therefore dependent on the number of UMIs detected per cell. The coefficient of variation is insensitive to this scaling factor as it is a dimensionless quantity (i.e, σ and μ are scaled by the same factor and thus cancel out when calculating coefficient of variation). However, the Fano factor, which has units, must be re-scaled to account for the differential effect that this normalization procedure has on σ^2 vs. μ (i.e., σ^2 gets scaled by S^2 while μ gets scaled by S). To negate the carryover of this scaling factor, calculated Fano factors from the Seurat-normalized dataset were multiplied by $\frac{1}{S}$ where S is a unique value for the DMSO and IdU samples: $\frac{10000}{\text{avg. number of UMIs per cell in sample}}$.

For analysis using BASiCS, quality control and filtering was performed using the BASiCS_Filter function resulting in an identical number of transcriptomes (812 and 744) as produced by Seurat. Posterior estimates of mean and over-dispersion for each gene were computed using a Markov Chain Monte Carlo (MCMC) simulation with 40,000 iterations and a log-normal prior. For differential mean testing, a threshold of fold change >2 with an FDR cutoff of 0.05 was used. Differential variability was tested with a threshold of fold change >1.5 with an FDR cutoff of 0.05. Only genes with no change in mean expression were considered for interpreting changes in variability. Code for scRNA-seq analysis is available at $https://github.com/weinbergerlab-ucsf/Code\_Desai\_et\_al$ and archived on Zenodo *(79)*.

Gene features and sequences from the GRCm38 reference were used for analysis of gene characteristics that potentiate noise enhancement. TAD boundary locations in mESCs were taken from Hi-C maps produced by Elphège et al *(80)*. DAVID v6.8 was used to test for gene ontology (GO) enrichment among highly variable genes. All tested genes from BASiCS were used as background. Bonferroni-corrected p-values (adjusted p-values) were used to visualize GO enrichment. Cell cycle determination was performed using *cyclone* as implemented in scran *(33)*. The default set of cell cycle marker genes for mESCs (mouse_cycle_markers.rds) was used. Cells were assigned to G1, S, and G2/M phases using their normalized genes counts produced by Seurat. Pseudotime analysis was conducted using *destiny (81)* with the Seurat-normalized cell-gene matrix as input. Gene-gene correlation matrices were assembled by first filtering out genes from the Seurat-normalized matrix whose mean abundance <1 in each treatment group to avoid spurious correlations that may emerge from low expression. 961 genes remained for downstream analysis. Pearson correlation for each gene pair was calculated. Clustering of gene-pairs based on similarity in correlation patterns was performed using the hierarchical clustering method within the *seriation* package. Change in correlation strength was calculated by subtracting absolute value of gene-pair correlation in DMSO condition from IdU condition.

Bulk RNA Sequencing Preparation and Analysis

$2x10^5$ mESCs were seeded in each well of a gelatin-coated, 6-well plate in 2i/LIF media. 24 hours following seeding, cultures were replenished with 2i/LIF media containing 10μM IdU, 5μM IdU or an equivalent volume of DMSO in triplicate for 24 hours. After treatment, cells were trypsinized with TrypLE and RNA was extracted using a RNeasy minikit (Qiagen) according to manufacturer's instructions. ERCC spike-in RNA (2μl diluted at 1:100) was added to each RNA extraction (Ambion, cat:4456740). A total of 9 cDNA libraries were prepared with an NEBNext Ultra II RNA Library Prep kit (NEB, cat:E7770S) and sequenced with an Illumina HiSeq4000. Sequencing yielded a median of ∼40 million single-end reads per library. Read quality was checked via FASTQC. Reads were aligned to an edited version of the mm10 reference genome containing the ERCC spike-in sequences using TopHat with default parameters. Transcript level quantification was performed using Cufflinks with default parameters. The quantification matrix was then imported into R and analyzed via DESeq2. Samples were normalized using ERCC transcripts as controls for size factor estimation. Differential mean testing was conducted with a threshold of fold change >2 and an FDR cutoff of 0.05.

Single Molecule RNA FISH

Probes for detection of nascent and mature Nanog transcripts were developed using the designer tool from Stellaris (LGC Biosearch Technologies) (Table S1). 30 probes (TAMRA conjugated) for mature Nanog mRNA were targeted towards the 3′ GFP segment of transcripts. 48 probes (Quasar 670 conjugated) for nascent Nanog mRNA were targeted towards the first intronic sequence as taken from the mm10 genome reference. Probes were designed using a masking level of 5, and at least 2 base pair spacing between single probes.

$1x10^5$ Nanog-GFP mESCs were seeded into each well of a gelatin-coated, 35mm Ibidi dish (quad-chambered, cat:80416) in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF containing 10μM 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 μg/mL glucose oxidase (Sigma Aldrich, cat: G7141), 520 μg/mL catalase (Sigma Aldrich, cat: C3515), and 0.5 mg/mL Trolox (Sigma Aldrich, cat: 238813). Images were taken on a Zeiss Axio Observer Z1 microscope equipped with a Yokogawa CSU-X1 spinning disk unit and 100x/1.4 oil objective. Approximately 20 xy locations were randomly selected for each condition. For each xy location, Nyquist sampling was performed by taking 30, 0.4μM steps along the z-plane.

Image analysis and spot counting was performed using FISH-quant *(82)*. Cells were manually segmented and analysis was conducted on cells of a similar size to minimize extrinsic noise. Tran-

scriptional centers (TCs) were identified by signal overlap in exon, intron and DAPI channels. The amount of nascent mRNA at TCs was quantified through a weighted superposition of point spread functions.

Rate calculations for random-telegraph model

From smRNA-FISH data for Nanog, the kinetic parameters of the random-telegraph model were inferred using the empirically derived values of mRNA mean (μ), mRNA Fano factor ($Fano$), transcriptional center frequency (f_{ON}) and transcriptional center intensity (TC_{mRNA}) *(28)*. The transcription rate (k_{tx}) is calculated as:

$$k_{tx} = TC_{mRNA}\frac{k_{elongation}}{L} \tag{1}$$

where $k_{elongation}$ is the elongation rate of RNAPII (1.9 kb/min) *(83)* and L is the length of the transcribed region of Nanog. The degradation rate (k_{deg}) is calculated as:

$$k_{deg} = \frac{f_{ON}\cdot k_{tx}}{\mu} \tag{2}$$

The rate of promoter activation (k_{ON}) is given by:

$$k_{ON} = k_{deg}\left(-\frac{\mu(f_{ON}-1)+f_{ON}(Fano-1)}{Fano-1}\right) \tag{3}$$

The rate of promoter inactivation (k_{OFF}) is given by:

$$k_{OFF} = -k_{deg}\left(-\frac{\mu(f_{ON}-1)+f_{ON}(Fano-1)}{Fano-1}\right)\cdot\left(\frac{1}{f_{ON}}-1\right) \tag{4}$$

Live-cell imaging of transcription with p21-MS2 reporter cell line

A p21-MS2 reporter U2OS cell line *(38)* was cultured at 37°C and 5% CO_2 in the Dulbecco's Modified Eagle Medium (DMEM) with high glucose (Corning), containing 10% Fetal Bovine Serum (FBS - Atlanta Biologicals) and 1% penicillin–streptomycin (Gibco). The reporter cells were pre-treated with either IdU (10μM) or an equivalent amount of DMSO for 48 hours before live-cell imaging. To induce expression of the p21 gene, the reporter cells were treated with 10μM Nutlin-3 (Sigma, cat: N6287) for 2 hours prior to live-cell imaging.

During the entire imaging sessions, p21-MS2 reporter cells were kept in Leibovitz's L-15 Medium (Gibco), containing 10% FBS at 37 °C and supplemented with Nutlin-3 and IdU or DMSO. Time-lapse z-series images were acquired on a wide-field microscope (Olympus IX-81

stand), equipped with an electron multiplying CCD camera (iXon3 DU-897E-CS0-# BV; Andor) and an objective with a 1.4 NA and 60X magnification, which yielded a pixel-size of 266.8nm. The microscope also was equipped with an automated XY stage (MS2000-XY with an extra-fine lead-screw pitch of 0.635 mm and a 10-nm linear encoder resolution; Applied Scientific Instrumentation) and a piezo-Z stage (Applied Scientific Instrumentation). Microscope control and image acquisition were achieved using MetaMorph software (Molecular Devices, Sunnyvale, CA). A 491-nm laser (Calypso-25; Cobolt) was used for illumination, and cells were imaged for 118 minutes with 11 stacks at 500 nm intervals (1 picture every minute).

Time-lapse images were acquired by maximum intensity projection of the z-series in ImageJ (http://rsb.info.nih.gov/ij/). The fluorescence of each transcription site (TS) in time-lapse images was quantified with a custom MATLAB software. The TS of each cell was detected semi-automatically, and the fluorescence signals were measured for each time point. Fast fluctuations in the fluorescence signal of the TS were removed using a rolling average. Transcriptional activity was normalized by dividing the fluorescence of the TS by the background signal. The TS was defined as active when transcriptional activity was higher than 2.5-fold the background.

The cumulative transcription occurring in each cell was calculated based on the normalized transcriptional activity over 118 minutes of imaging. The K_{OFF} (hr^{-1}) is calculated by taking the inverse of the average of the burst duration. The K_{ON} (hr^{-1}) is calculated by taking the inverse of the average time between bursts. Errors represent the standard error of the mean. Significance was determined by a two-sample t-test performed in MATLAB.

Extrinsic Noise Filtering on Flow Cytometry Data

All flow cytometry data were collected on BD FACSCalibur, LSRII or LSRFortessa X-20 with 488-nm laser used to detect GFP. For all measurements of Nanog-GFP mean and variability, >50k cells are collected per sample. Gating of cytometry data was performed with FlowJo. Prior to quantification of Nanog-GFP mean and variability, the smallest possible forward- and side-scatter region containing at least 3k cells was used to isolate cells of similar size and shape. This filters out gene expression variability arising from cell-size heterogeneity as previously established *(22,24,28)*.

Cell-cycle Analysis by Propidium-Iodide Staining

$2x10^5$ Nanog-GFP mESCs were seeded in each well of a gelatin-coated, 6-well plate in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF media containing 10μM IdU or an equivalent volume of DMSO in triplicate for 24 hours. After treatment, cells were washed with DPBS, dissociated with TrypLE, pelleted, washed with DPBS, and resuspended in ice-cold 70% ethanol. Samples were stored overnight at -20°C and pelleted the following day at 200g for 5 minutes at 4°C. Cells were washed twice with DPBS supplemented with 0.5% BSA to prevent cell loss. Pellets were resuspended in 150μL of DPBS supplemented with 0.1mg/ml RNAse A

(Thermo, cat: EN0531) and 30 μg/ml Propidium Iodide (Thermo, cat:P3566). After overnight incubation at 4°C, cells were directly analyzed on a BD LSRII cytometer.

Actinomycin D assay for post-transcriptional noise

1x10^5 Nanog-GFP mESCs were seeded in each well of gelatin-coated, 12-well plates in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF supplemented with 10μM IdU, 6ug/ml Actinomycin D (ActD) (Sigma cat:A1410), 10μM IdU + 6ug/ml ActD or equivalent volume DMSO. After 24 hours of incubation, cells were run unfixed and live on BD Fortessa.

Noise Enhancer Testing in Serum/LIF culture

Serum/LIF media was prepared with 85% DMEM (supplemented with 2mM of L-glutamine), 15% FBS, 0.1mM 2-mercaptoethanol, and 1000U/ml of LIF (Sigma Aldrich, cat: LIF1010). Nanog-GFP mESCs grown feeder-free in 2i/LIF were passaged and seeded onto gelatin-coated 10cm dishes in serum/LIF media. Cells were passaged twice in serum/LIF media prior to noise enhancer testing. 4x10^5 Nanog-GFP mESCs were seeded into each well of a gelatin-coated 6-well plate in serum/LIF media. 24 hours following seeding, media was replaced with serum/LIF supplemented with either 10μM IdU or equivalent volume DMSO in triplicate. After 24 hours of treatment, cells were run unfixed and live on BD LSRII flow cytometer.

Live-cell time-lapse microscopy of Nanog-GFP expression

1x10^5 Nanog-GFP mESCs were seeded into each well of a gelatin-coated, 35mm Ibidi dish (quad-chambered, cat:80416) in 2i/LIF media. 24 hours following seeding, media was replenished with 2i/LIF containing 10μM IdU or equivalent volume DMSO in replicate. Time-lapse imaging commenced immediately after addition of compounds with IdU- and DMSO- treated cells imaged in the same experiment (neighboring wells). Imaging was performed on a Zeiss Axio Observer Z1 microscope equipped with Yokogawa CSU-X1 spinning disk unit and a Cool-SNAP HQ2 14-bit camera (PhotoMetrics). 488nM laser line (50% laser power, 500-ms excitation) was used for GFP imaging. Samples were kept in an enclosed stage that maintained humidified conditions at 37°C and 5% CO_2. Images were captured every 20 minutes for 24 hours. For each xy location, three z-planes were sampled at 4-μm intervals. The objective used was 40x oil, 1.3 N.A.

Cell segmentation, tracking and GFP quantification were carried out using CellProfiler *(84)*. Tracking of cells was manually verified. Segmented cells tracked for less than 4 hours were discarded. Cell division triggered the start of 2 new trajectories. After illumination correction and background subtraction, the mean GFP fluorescence intensity of a segmented cell was taken from each z-plane and averaged over the entire z-stack. For each trajectory, noise autocorrelation ($\tau_{1/2}$) and noise magnitude (intrinsic-CV^2) were calculated as previously described *(41)*. Fluorescence trajectories were first detrended (normalized) by subtracting the population time-dependent average fluorescence to isolate intrinsic noise. Distributions of noise frequency ranges

(F_N) were extracted from normalized autocorrelation functions (ACFs) of individual trajectories, where $F_N = \frac{1}{\tau_{1/2}}.\tau_{1/2}$ is the value of τ (lag time) where the normalized ACF reaches a value of 0.5.

Sox2 two-color reporter assay

The endogenous alleles of Sox2 are tagged with P2A-mClover and P2A-tdTomato. Both fluorophores have a PEST tag, thus shortening their half-lives to approximately 2.5 hours. 2x10^5 Sox2-dual-tag mESCs were seeded in each well of a gelatin-coated, 6-well plate in 2i/LIF media. 24 hours following seeding, cultures were replenished with 2i/LIF media containing 10μM IdU or an equivalent volume of DMSO in triplicate for 24 hours. Cells were run unfixed and live on BD LSRII flow cytometer. Data from all three replicates were pooled together. No cell-size gating was performed as assay allows for separation of extrinsic noise. To align fluorescence values of mClover and tdTomato on the same scale, each cell's fluorescence intensity was normalized to the mean expression level of that fluorophore for the population. Since Sox2 expression spans several orders of magnitude, cells were binned according to their total Sox2 expression (mClover + tdTomato). Bins with fewer than 100 cells were discarded. Intrinsic noise (CV^2) of Sox2 expression for each bin was calculated using the following formula:

$$\eta^2_{intrinsic} = \frac{\langle (tdTomato_i - mClover_i)^2 \rangle}{2\langle tdTomato \rangle \langle mClover \rangle} \tag{5}$$

This value was then multiplied by the mean Sox2 expression for each bin to obtain the Fano factor. Given that the number of cells in each bin differs and variance estimates are affected by sample size, we calculated 95% confidence intervals around the Fano factor for each bin through bootstrapping. Bin populations were resampled 10,000 times with replacement.

Nucleoside analog screening

14 nucleoside analogs (compound names and sources listed in Table S3) were resuspended in DMSO. 1x10^5 Nanog-GFP mESCs were seeded in gelatin-coated 12-well plates in 2i/LIF media. 24 hours after seeding, media was swapped with 2i/LIF containing 10μM of nucleoside analog or equivalent volume DMSO in replicate. After 24 hours of treatment, cells were run unfixed and live on BD LSRII cytometer. Extrinsic noise filtering via cell-size gating was performed prior to calculation of Nanog Fano factor. Fano factor for Nanog-GFP expression for each treatment was normalized to DMSO control.

Generation of stable CRISPRi Nanog-GFP mESC line

To stably integrate the CRISPRi machinery into the ROSA26 locus of Nanog-GFP mESCs, AAVS1 homology arms of the CRISPRi knockin construct (krab-dCas9-p2a-mCherry, Addgene:73497) were swapped with ROSA26 homology arms. The dox-inducible promoter of this construct was replaced with a constitutive CAGGS promoter and the kanamycin resistance cassette was replaced with puromycin resistance. Two million Nanog-GFP mESCs were nucleofected with the CRISPRi

knockin construct and left to recover for 48 hours. Puromycin (1μg/ml) selection was run until single colonies could be picked. Clonal CRISPRi Nanog-GFP mESC lines were assessed for mCherry expression and ability to knockdown Nanog. We selected the clone with the highest percentage of mCherry-positive cells.

CRISPRi gRNA design and cloning

gRNA sequences were were taken from the mCRISPRi-v2 library *(85)*. gRNA oligos were annealed and cloned into the pU6-sgRNA EF1Alpha-puro-T2A-BFP lentiviral vector (Addgene:60955) using the BstXI/BlpI ligation strategy *(85)*.

CRISPRi screening for genetic dependencies of noise enhancer

25 genes involved in nucleotide metabolism, DNA repair, and chromatin remodeling were screened for their potential role in noise enhancement from IdU. Three gRNAs were designed per gene (gene names and gRNA sequences listed in Table S4). Three non-targeting controls (scrambled gRNAs) were taken from the mCRISPRi-v2 library *(85)*. Each gRNA expression plasmid was separately packaged into lentivirus in HEK293T cells as previously described *(85)*. For each gRNA lentivirus, $1.5x10^5$ CRISPRi Nanog-GFP mESCs were spinoculated with filtered viral supernatant for 90 minutes at 200 x g in replicate. Following spinoculation, infected cells were seeded into gelatin-coated, 6-well plates in 2i/LIF media. 48 hours following seeding, media was swapped with 2i/LIF supplemented with either 10μM IdU or equivalent volume DMSO. Consequently, for every knockdown there is a DMSO and IdU treatment group. After 24 hours of treatment, cells were run unfixed and live on a BD LSRII flow cytometer. To minimize technical variability, analysis was restricted to cells with homogeneous levels of dCas9-KRAB and gRNA expression through stringent gating on mCherry/BFP double-positive cells. Extrinsic noise filtering through cell-size gating was then applied. For each gRNA, Nanog Fano factor for the DMSO and IdU treatments were normalized to the Nanog Fano factor of the non-targeting controls treated with DMSO.

qPCR verification of CRISPRi knockdown

To verify CRISPRi knockdown of Apex1 and Tk1, each of the six gRNA-expression plasmids targeting these two genes along with a non-targeting control and empty vector were packaged into lentivirus. $1.5x10^5$ CRISPRi Nanog-GFP mESCs were spinoculated with filtered viral supernatant for 90 minutes at 200 x g in replicate. Following spinoculation, infected cells were seeded into gelatin-coated, 6-well plates in 2i/LIF media. 72 hours following seeding, $1x10^6$ mCherry/BFP double-positive cells from each infected cell population were sorted on a FACSAria II. Total RNA was extracted using an RNeasy Mini Kit (QIAGEN cat:74104) and reverse-transcribed using a QuantiTect Reverse Transcription Kit (QIAGEN cat:205311). cDNA from each independent biological replicate was plated in triplicate and run on a 7900HT Fast Real-Time PCR System (Thermo) using designed primers (Table S5) and Fast SYBR Green Master Mix (Applied Biosystems, cat:4385612). Expression of GAPDH was used for normalization. Relative mRNA levels

of Apex1 and Tk1 were calculated by the $\Delta\Delta C_t$ method using the empty-vector populations as the control. All reported levels of repression are relative to the non-targeting control.

Tk1 competition assay

$1x10^5$ Nanog-GFP mESCs were seeded in each well of gelatin-coated, 12-well plates in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF supplemented with 10μM IdU in combination with thymidine (Sigma cat:T1895) or uridine (Sigma cat:U3003) at concentrations ranging from 0 to 100μM. Concentration combinations were done in triplicate. After 24 hours of treatment, cells were run unfixed and live on BD FACS Calibur cytometer. Extrinsic noise filtering via cell-size gating was performed prior to calculation of Nanog Fano factor.

ChIP-qPCR assay of Apex1 binding to Nanog promoter

$1x10^6$ Nanog-GFP mESCs were seeded in gelatin-coated 10cm dish in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF media containing 10μM IdU, 100μM CRT0044876, 10μM IdU + 100μM CRT0044876 or an equivalent volume of DMSO in triplicate for 24 hours.

After treatment, cells were washed with DPBS, dissociated with TrypLE, and subsequently washed three times with 10ml PBS followed by fixation with 1% formaldehyde at 37°C for 10 minutes. Fixation was quenched with 0.125M glycine (Sigma-Aldrich, cat:50046) for 5 minutes at room temperature. Cells were washed three more times with 10ml of ice-cold PBS (pH 7.0) containing 1X protease inhibitor cocktail (P8340; Sigma-Aldrich) followed by lysing in lysis buffer (1.0% SDS, 10mM EDTA, 50mM Tris-HCl pH7.8) for 10 minutes on ice. The lysate was subjected to sonication in Covaris S2 sonicator (Covaris inc.) The sonicated lysate was centrifuged at 14,000 rpm for 15 minutes at 4°C and the supernatant was subjected to immunoprecipitation with 5ug of Apex-1 antibody (Thermofisher Scientific, cat: 501726108) in buffer containing 0.01% SDS, 1% Triton x-100, 1.2mM EDTA, 16.7mM Tris-HCl (pH 8.0) and 167mM NaCl at 4°C for 4 hours with constant shaking. Next, 25ul of Protein A/G Dynabeads (Thermofisher Scientific, cat: 88802) was added and incubated at 4°C for 2 hours with constant shaking. The beads were subjected to three washes with wash buffer-A (50mM Hepes, pH 8.0, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 0.5mM PMSF and protease inhibitor cocktail) followed by three washes with wash buffer-B (20mM Tris, pH 8.0, 1mM EDTA, 250mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 0.5 mM PMSF and protease inhibitor cocktail) followed by two washes with TE buffer (10mM Tris HCl and 1.0mM EDTA). Protein-DNA complex was eluted in ChIP elution buffer (50mM Tris, pH 8.0, 1mM EDTA, 1% SDS and 50mM NaHCO3) and incubated at 65°C for 10 minutes, followed by centrifugation at 14,000 rpm for 1 minute. Supernatant was transferred to a new tube and the beads were eluted again with 200ul of elution buffer, followed by reverse crosslinking by incubating the eluate in 200mM NaCl overnight at 65°C, and subjected to digestion with RNAse and proteinase K. DNA was precipitated with phenol-chloroform extraction, followed by ethanol precipitation. Purified ChIP DNA was resuspended in TE buffer and subjected

to qPCR using 7900HT Fast Real-Time PCR system (Thermo-Fisher Scientific, cat:4329003) with Fast SYBR Green Master Mix. qPCR primers were directed against Nanog promoter (sequences in Table S7).

H_2O_2 oxidative stress assay

Nanog-GFP mESCs were seeded in gelatin-coated 12-well plate at a density of 20,000 cells/cm^2 in 2i/LIF media. 24 hours after seeding, media was replaced with 2i/LIF containing the indicated concentrations of H_2O_2 (Sigma cat: H1009) in triplicate. After 1 hour of treatment, H_2O_2 was washed off, and cells were cultured in fresh 2i/LIF for 24 hours. Cells were run unfixed and live on BD Fortessa.

MMS alkylation damage assay

1x10^5 Nanog-GFP mESCs were seeded in each well of gelatin-coated 12-well plate in 2i/LIF media. 24 hours after seeding, media was replaced with 2i/LIF containing 2mM MMS (Sigma cat: 129925), 2mM MMS + 100μM CRT0044876 or equivalent volume DMSO in triplicate. After 1 hour of treatment, media was washed off all wells and replaced with fresh 2i/LIF for 24 hours. Cells were run unfixed and live on BD Fortessa.

UV stress assay

1x10^5 Nanog-GFP mESCs were seeded in each well of gelatin-coated, 12-well plates in 2i/LIF media. 24 hours following seeding, cultures were exposed to 3kJ of 365nM light (Fotodyne UV Transilluminator 3-3000 with 15W bulbs) for 15, 30 or 60 minutes at room temperature in the dark. Control plates were left at room temperature in the dark for equivalent periods of time. Cells from UV-exposed and control plates were run unfixed and live on BD FACS Calibur cytometer 1,2,4,8, and 12 hours post-exposure in replicate. Extrinsic noise filtering via cell-size gating was performed prior to calculation of Nanog Fano factor.

Mpg and Ung overexpression assay

cDNA clones for mouse Methylpurine glycosylase (Mpg) (Origene MR204931) and mouse Uracil DNA glycosylase (Ung) (Origene MR204314) were PCR amplified, ligated with IRES-mCherry and cloned into MluI/NotI sites of pHR-TREp-Tat-Dendra lentiviral vectors *(86)*.

Mpg, Ung, and rtTA (pHR-SFFVp-rtTA) vectors were packaged into lentivirus in HEK293Ts. Nanog-GFP mESCs were spinoculated with filtered viral supernatant containing rtTA lentivirus in combination with Mpg or Ung lentivirus for 90 minutes at 200 x g. After expansion of infected cells, 1x10^5 Mpg- and Ung-overexpressing Nanog-GFP mESCs were seeded in gelatin-coated 12-well plates in 2i/LIF media. 24 hours after seeding, media was replaced with 2i/LIF containing 2ug/ml doxycycline (Sigma-Aldrich, cat: D9891) or equivalent volume DMSO in triplicate. After

24 hours of doxycycline-induction, cells were run unfixed and live on BD Fortessa. To minimize technical variability, analysis was restricted to cells with homogeneous levels of Mpg/Ung expression through stringent gating on mCherry positive cells. Extrinsic noise filtering through cell-size gating was then applied.

Rescue of noise-enhancement with catalytically inactive Apex1

Coding sequences of wild-type and catalytically inactive (CI) (G283C, G625A) mouse Apex1 (NM009687.2) were synthesized from IDT gBlocks, ligated with IRES-mOrange, and cloned into BamHI/NotI sites of pHR-TREp-Tat-Dendra *(86)*.

Apex1-WT, Apex-CI, mOrange control, and rtTA (pHR-SFFVp-rtTA) vectors were packaged into lentivirus in HEK293T cells. CRISPRi Nanog-GFP mESCs were spinoculated with filtered viral supernatant containing rtTA lentivirus in combination with Apex1-WT, Apex1-CI or mOrange control lentivirus for 90 minutes at 200 x g. Following spinoculation, infected cells were expanded and grown in 2i/LIF supplemented with 2ug/ml doxycycline. After 2 days of doxycycline induction, mOrange positive cells from each infected populated were bulk sorted on a FACSAria II to obtain successfully transduced cells.

Apex1 gRNA-expression (4th sequence in Table S4) plasmid (contains BFP reporter) was packaged into lentivirus in HEK293T cells. Importantly, Apex1 gRNA was targeted towards the endogenous promoter sequence and thus does not interfere with expression of transduced vectors. Apex1-WT, Apex-CI, mOrange control CRISPRi cell lines were spinoculated with filtered viral supernatant containing Apex1 gRNA lentivirus for 90 minutes at 200 x g. Following spinoculation, infected cells were seeded into gelatin-coated, 12-well plates in 2i/LIF media at a density of 25,000 cells/cm^2. 24 hours after infection, media was replaced with 2i/LIF supplemented with 2ug/ml doxycycline. After 24 hours of doxycycline induction, media was replaced with 2i/LIF supplemented with 2ug/ml doxycycline and either 10μM IdU or equivalent volume DMSO in triplicate. After 24 hours of treatment, cells were run unfixed and live on BD Fortessa. To minimize technical variability, analysis was restricted to cells with homogeneous levels of Apex1-WT/CI, dCas9-KRAB and gRNA expression through stringent gating on mOrange/mCherry/BFP triple-positive cells. Extrinsic noise filtering through cell-size gating was then applied. Nanog Fano factor for all treatment conditions were normalized to the Nanog Fano factor of the mOrange control cell line treated with DMSO.

Biotinylated-trimethylpsoralen (bTMP) supercoiling assay

1x10^5 Nanog-GFP mESCs were seeded into each well of a gelatin-coated, 35mm Ibidi dish (quad-chambered, cat:80416) in 2i/LIF media. 24 hours following seeding, media was replaced with 2i/LIF supplemented with 10μM IdU, 10μM IdU + 100μM CRT0044876, or equivalent volume DMSO in replicate. After 24 hours of treatment, media was replaced with 2i/LIF supplemented with 1μM aphidicolin (Sigma-Aldrich, cat:A0781) for two hours. For control experiments,

Nanog-GFP mESCs were cultured with or without 10μM IdU for 24 hours followed by treatment with 100μM bleomycin for one hour. Cells were then washed 1xDPBS and then permeabilized with 0.1% Tween-20 in DPBS for 15 minutes. Cells were then incubated with 0.3mg/ml EZ-Link Psoarlen-PEG3-Biotin (Thermo, cat:29986) for 15 minutes. Cultures were then exposed to 365nM light (AlphaImager HP with 15W bulbs, ProteinSimple) for 15 minutes at room temperature. Cells were then washed 2xDPBS, fixed with cold 70% ethanol for 30 minutes at 4°C, and then washed 2xDPBS. Cells were then incubated with Alexa Fluor 594 Streptavidin (Thermo cat:S32356) for one hour at room temperature in the dark, washed 2xDPBS, and stained with DAPI for 10 minutes at room temperature in the dark. Cells were imaged in a buffer containing 50% glycerol (Thermo, cat: 17904), 75 μg/mL glucose oxidase (Sigma Aldrich, cat: G7141), 520 μg/mL catalase (Sigma Aldrich, cat: C3515), and 0.5 mg/mL Trolox (Sigma Aldrich, cat: 238813). Images were taken on a Zeiss Axio Observer Z1 microscope equipped with a Yokogawa CSU-X1 spinning disk unit and 63x/1.4 oil objective. Approximately 20 xy locations were randomly selected for each condition. For each xy location, three z-planes were sampled at 4-μm intervals. Nuclear segmentation using DAPI signal and quantification of psoralen staining intensity were carried out using CellProfiler. After illumination correction and background subtraction, the mean psoralen fluorescence intensity of a segmented nucleus was taken from each z-plane and averaged over the entire z-stack.

Topoisomerase 1 overexpression assay

Topoisomerase 1 (NM 009408) expression vector (Origene, MR218547L3) with puromycin selection marker was packaged into lentivirus using HEK293T cells. Nanog-GFP mESCs were spinoculated with lentiviral supernatant. Three days following transduction, infected cells were subjected to seven days of puromycin (1ug/ml) selection to isolate cells stably overexpressing Topoisomerase 1.

$1x10^5$ wildtype and Topoisomerase 1 overexpressing Nanog-GFP mESCs were seeded in gelatin-coated 12-well plates in 2i/LIF media. 24 hours after seeding, media was swapped with 2i/LIF containing 10μM of IdU or equivalent volume DMSO in triplicate. After 24 hours of treatment, cells were run unfixed and live on BD LSRII cytometer. Extrinsic noise filtering via cell-size gating was performed prior to calculation of Nanog Fano factor. Fano factor for Nanog-GFP expression for each treatment was normalized to the DMSO control for wildtype Nanog-GFP mESCs.

96-dose combination for testing of noise phase space

Compound plates containing 96 concentration combinations of IdU, BrdU (Sigma Aldrich, cat: B5002), or HmU (Santa Cruz Biotech, cat: SC-221028) with CRT0044876 were prepared by the Gladstone Assay Development and Drug Discovery Core using an Agilent Bravo liquid handling system. All wells contained equivalent volumes of DMSO. Compound mixtures were suspended in 200μL of 2i/LIF media. $1x10^4$ Nanog-GFP mESCs were seeded into each well of a gelatin-coated, 96-well dish in 200μL of 2i/LIF media. 24 hours after seeding, 100μL of media was removed from each well and 100μL of compound-containing 2i/LIF was added in replicate. Layout and final

concentrations of treatments are listed in Table S6. IdU and BrdU concentrations ranged from 0 to 50μM while HmU concentrations ranged from 0 to 10μM. CRT0044876 ranged from 0 to 150μM. After 24 hours of treatment, cells were detached using TrypLE and plates were run on BD LSR-Fortessa high-throughput system. After extrinsic noise filtering via cell-size gating, Nanog mean and Fano factor for each treatment were normalized to DMSO control well. Reported fold changes in mean and Fano factor are the average of two replicates.

Estimation of promoter toggling kinetics from scRNA-seq data

Gene expression data from the scRNA-seq dataset were fit to the 2-state model using the D3E algorithm, allowing for estimation of $k_{ON}, k_{OFF}, and k_{tx}$ in proportion to the rate of mRNA degradation which is the lone parameter that is not estimable from this dataset alone *(70)*. Parameter estimation was conducted using the methods of moments approach with the normalise and removeZeros options. Analysis was run for the 945 genes classified as highly variable according to the BASiCS algorithm. The Cramer-von Mises test was used for goodness-of-fit testing. Values of k_{decay} were then retrieved from an existing dataset of mRNA degradation rates in mESCs *(87)*, with the assumption that degradation rates are unchanged between DMSO and IdU conditions. Parameter estimates were then verified against experimental values of mean mRNA counts using the following relationship: $\langle RNA \rangle = \frac{k_{ON}}{k_{ON}+k_{OFF}} \cdot \frac{k_{tx}}{k_{decay}}$. Genes whose predicted mean was within 10% of experimental value were used for downstream analysis. 314 genes passed this filtering process based on availability of mRNA degradation rates and alignment of parameter estimates with expected mean mRNA counts.

Mathematical modeling and simulations

Detailed descriptions of model development, parameterization, evaluation, and sensitivity analysis can be found in the supplementary text along with details of Gillespie simulations.

Differentiation of mESCs into neural ectodermal lineage

Nanog-GFP mESCs were seeded at a density of 20000 cells/cm^2 in gelatin-coated 12-well plate in 2i/LIF media. 24 hours after seeding, cells were washed with PBS, and differentiation was carried out in N2B27 medium *(88)* supplemented with 4uM IdU or equivalent volume DMSO in 6 replicates. Spent medium was exchanged with fresh medium after 48 hours, and IdU was only kept in culture for the first 48 hours of differentiation.

After 48 hours and 96 hours of differentiation, cells were washed once with PBS and then incubated in basal (N2B27) medium with Alexa Fluor700 Rat Anti-Mouse CD24 (BioLegend cat: 101836, clone M1/69) at a dilution of 1:1000 for 30 min at 37 °C, in the dark. Subsequently, cells were washed once with PBS, dissociated with Accutase (Life Technologies cat: 00455556), and analyzed on BD Fortessa.

Cellular reprogramming assays

Three cellular reprogramming systems were tested in this study: (1) Nanog-GFP secondary mouse embryonic fibroblasts (MEFs) harboring stably integrated, doxycycline-inducible cassettes for Oct4, Sox2, and Klf4. GFP is expressed from the endogenous Nanog locus. (2) Oct4-GFP MEFs that express GFP from the endogenous Oct4 locus. (3) Oct4-GFP MEFs that harbor knocked-in, doxycycline-inducible cassettes for Oct4, Sox2, C-Myc and Klf4 (OKSM) at Col1a1 locus.

In the first reprogramming assay, Nanog-GFP secondary MEFs were seeded onto gelatin-coated, 12-well plates at a density of 10,000 cells/cm^2 in MEF medium (DMEM supplemented with 10% FBS and 0.1mM non-essential amino acid, and 2mM Glutamax). 24 hours after seeding, wells were washed with DPBS and media was switched to ESC media (knockout DMEM, 10% FBS, 10% KSR, 2mM Glutamax, 0.1mM non-essential amino acid, 0.1mM 2-mercaptoethanol, 10^3 units/ml leukemia inhibitory factor) supplemented with 1μg/ml doxycycline. Additionally, IdU (1uM or 4uM) or equivalent volume DMSO (Day 0) were added to media. 48 hours after the start of IdU treatment, wells were washed with DPBS and media was replaced with ESC media supplemented with 1μg/ml doxycycline alone. Media was refreshed every other day until day 10 of reprogramming. Alkaline phosphatase staining was performed according to manufacturer's instructions using the Alkaline Phosphatase Detection Kit (Millipore, cat: SCR004). For flow cytometric analysis of Nanog-GFP expression, cells were dissociated with TrypLE and run unfixed on BD FACS Calibur cytometer.

In the second reprogramming assay, Oct4-GFP primary MEFs were transduced with lentiviral vectors encoding Oct4, Sox2, Klf4, and c-Myc. Lentiviruses encoding these factors were individually packaged in PLAT-E cells (ATCC) using pMX-based vectors. 48 hours after transfection of lentiviral vectors, viral supernatant was collected and filtered. For infection, Oct4-GFP primary MEFs were seeded on gelatin-coated, 6-well plates at a density of 10,000 cells/cm^2 in MEF medium 24 hours prior to transduction (Day -2). Oct4, Sox2, Klf4, and c-Myc viruses were mixed in equal volume along with 5μg/ml polybrene and incubated with primary MEFs for 24 hours in MEF medium (Day -1). Following infection, wells were washed with ESC media and cells were incubated with ESC media supplemented with 10μM Forskolin, 1mM Valproic Acid, 4μM IdU or equivalent volume DMSO (Day 0). ESC media was refreshed every other day. IdU supplementation was discontinued after 48 hours while Forskolin and Valproic Acid were kept in media continuously. Oct4-GFP(+) colonies were counted on days 8, 10 and 12.

In the third reprogramming assay, Apex1 was knocked down via shRNAs (sequences in Table S8) to assess its necessity for IdU-mediated enhancement of reprogramming. shRNA sequences were inserted into AgeI/EcoRI sites of pLKO.1 (Addgene # 10878) and puroR was replaced with mCherry. Oct4-GFP MEFs with doxycycline-inducible cassettes for OKSM factors were seeded on gelatin-coated, 12-well plates at a density of 6000 cells/cm^2 in MEF medium 24 hours prior to transduction (Day -2). shRNA viruses along with 5μg/ml polybrene were incubated with MEFs

for 24 hours in MEF medium (Day -1). Following infection, wells were washed with ESC media and cells were incubated with ESC media supplemented with 2μg/ml doxycycline and 4uM IdU or equivalent volume DMSO (Day 0). IdU supplementation was discontinued after 48 hours. ESC media with 2μg/ml doxycycline was refreshed every other day, and from Day 6, medium was refreshed every day. On day 10, cells were dissociated with TrypLE and run unfixed on BD Fortessa. Oct4-GFP percentages were analyzed among mCherry positive (shRNA expressing) cells. qPCR validation of Apex1 knockdown was carried out as described in above section (qPCR verification of CRISPRi knockdown). qPCR primer sequences are listed in Table S8.

Bulk RNA-seq of secondary MEFs undergoing reprogramming

Secondary MEFs were seeded onto gelatin-coated, 6-well plates at a density of 10,000 cells/cm^2 in MEF medium. For each timepoint (2- and 5-day), 4 wells were seeded (2 replicates for standard reprogramming and 2 replicates for IdU-assisted reprogramming). 24 hours after seeding, wells were washed with DPBS and media was switched to ESC media supplemented with 1μg/ml doxycycline. Additionally, 4μM IdU or equivalent volume DMSO (Day 0) were added to media. 48 hours after the start of reprogramming, cells for the 2-day timepoint in DMSO and IdU conditions were dissociated with TrypLE, pelleted, and snap frozen with liquid nitrogen. Media in the remaining wells was refreshed with ESC media supplemented with 1μg/ml doxycycline alone at 48 hour timepoint. This was repeated on day 4. On day 5, remaining cells were dissociated and frozen identically to that of the 2-day timepoint.

RNA was extracted from each cell pellet using a RNeasy minikit (Qiagen) according to manufacturer's instructions. A total of 8 cDNA libraries were prepared with an NEBNext Ultra II RNA Library Prep kit (NEB, cat:E7770S) and sequenced with an Illumina HiSeq4000. Sequencing yielded a median of ∼50 million single-end reads per library. Read quality was checked via FASTQC. Reads were aligned to the mm10 reference genome using TopHat with default parameters. Transcript level quantification was performed using Cufflinks with default parameters.

Supplementary Text

Contents

Overview

The overall goal of our computational modeling initiative was to develop a mechanistic model of transcription that was capable of quantitatively fitting empirical data of the Apex1-generated increase in transcriptional noise without changing mean. We refer to this phenomenon as **D**iscordant **T**ranscription t**h**rough **R**epair (DiThR) and herein develop a series of models of increasing complexity (summarized in Supp. Fig. 22) to identify a minimal model capable of capturing how Apex1 recruitment to DNA (mediated by IdU) affects transcription, thus recapitulating DiThR. We refer to the model that most accurately recapitulates the experimental data as the DiThR model.

All tested models (Supp. Fig. 22) build from the canonical two-state random-telegraph model (Model 0) and are described in Sections 2–3 (below). Experimental data are used to infer parameter estimates for the models (Section 4) and Monte-Carlo simulations of the models were used for model negation and identification of the best-fit model (Sections 5 and 6). Parameter sensitivity analysis was performed (Section 7), and the best-fit model was tested for its ability to recapitulate observed DiThR for other genes (section 8).

1 General modeling assumptions

Based upon existing literature of nucleotide incorporation into DNA *(89)* and the observed linear increase in Fano factor with increasing IdU concentration (Fig. S22D), all models assume first-order IdU incorporation. To prevent overparameterization, IdU incorporation is treated as a single kinetic step (i.e. Michaelis Menten kinetics with $k_{incorpo} = [IdU] \cdot k_{0,incorpo}$).

Based on the data that IdU does not increase noise when Apex1 is knocked down (Fig. 3C), Apex1 recruitment to DNA—rather than IdU incorporation—was modeled as the key transcription-altering step. Based on the Apex1 catalytic inhibitor data, models assume that the steric and topological actions of Apex1 on DNA generates a transcriptionally non-productive state and that RNA polymerase II cannot read through the Apex1-IdU DNA complex. Notably, two types of off states were modeled: the *canonical* off state of the promoter (OFF) and an *effective* off state. The *effective* off state can be seen as all the promoter states that are incapable of active transcription, irrespective of the underlying mechanism inhibiting transcription. In other words, this non-productive, *effective* off state is equivalent to the measured off state (see section 4 for further details).

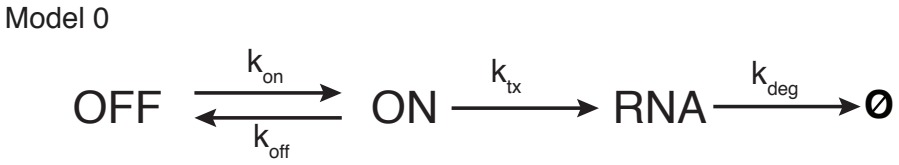
2 Detailed mathematics and derivation of parameter constraints for various models

Five models of increasing complexity are considered. Relationships between kinetic rates at equilibrium are derived and these relationships are then used to constrain the parameter phase space in

section 4.

Model 0

Model 0, used essentially as a null hypothesis, is the well-known canonical two-state *random telegraph* model *(17)* and is described by the following schematic and rate scheme:



$$
\begin{aligned}
OFF &\xrightarrow{k_{ON}} ON \\
ON &\xrightarrow{k_{OFF}} OFF \\
ON &\xrightarrow{k_{tx}} RNA \\
RNA &\xrightarrow{k_{deg}} \varnothing
\end{aligned}
$$

Model 0 is used for the Maximum Likelihood Estimation and Akaike Information Criterion (AIC)-based model selection (see section 5.1) and for analysis of smFISH results to determine reaction rates (see main methods section) and parametric relationships as follows.

Model 0 can be described by the following set of coupled linear Ordinary Differential Equations (ODEs):

$$
\begin{cases}
\frac{dOFF}{dt} = -k_{ON}OFF + k_{OFF}ON \\
\frac{dON}{dt} = k_{ON}OFF - k_{OFF}ON \\
\frac{dRNA}{dt} = k_{tx}ON - k_{deg}RNA
\end{cases}
\tag{6}
$$

At steady state:

$$
\frac{dOFF}{dt} = \frac{dON}{dt} = \frac{dRNA}{dt} = 0
$$

we obtain:

$$\begin{cases} K \doteq \frac{k_{ON}}{k_{OFF}} = \frac{ON_{eq}}{OFF_{eq}} \\ \\ RNA_{eq} = \frac{k_{tx}}{k_{deg}} \times ON_{eq} \end{cases} \tag{7}$$

The fraction of time spent in $ON + OFF = 1$ (since a single promoter is being considered). So:

$$ON_{eq} = \frac{k_{ON}}{k_{OFF}}(1 - ON_{eq}) = \frac{k_{ON}}{k_{ON} + k_{OFF}}$$

Thus, we arrive at the fundamental relationships used herein:

$$\langle ON \rangle \doteq ON_{eq} = \frac{k_{ON}}{k_{ON} + k_{OFF}} \quad ; \quad \langle OFF \rangle \doteq OFF_{eq} = \frac{k_{OFF}}{k_{ON} + k_{OFF}}$$

$$K \doteq \frac{k_{ON}}{k_{OFF}} = \frac{ON_{eq}}{OFF_{eq}} \quad ; \quad \langle RNA \rangle \doteq RNA_{eq} = \frac{k_{tx}}{k_{deg}} \cdot \frac{k_{ON}}{k_{ON} + k_{OFF}}$$

Importantly, at equilibrium only ratios between kinetic parameters are important.

Intuitively, we can write a ratio involving the promoter kinetic constants as:

$$\frac{k_{ON}}{k_{ON} + k_{OFF}} = \frac{k_{ON}}{k_{OFF}} \cdot \frac{1}{\frac{k_{ON}}{k_{OFF}} + 1}$$

Thus, from the previously derived relations, it follows that:

$$\frac{k_{ON}}{k_{ON} + k_{OFF}} = \frac{ON_{eq}}{ON_{eq} + OFF_{eq}}$$

Lumped kinetic parameters

Often the true promoter states are more complex than the simple two-state model *(90)*. Since, with smFISH we are restricted to measuring two promoter states, i.e., a measured ON and measured OFF state (Figure S22B), the kinetic rates that were measured by smFISH need to be treated as *lumped kinetic parameters*. We therefore adopt capital notations for the lumped or *effective* kinetic parameters that are measured through smFISH ($K_{ON}, K_{OFF}, K_{TX}, K_{deg}$). In other words, all following rates that have a capital notation (i.e. K) are lumped parameters that can be measured by smFISH.

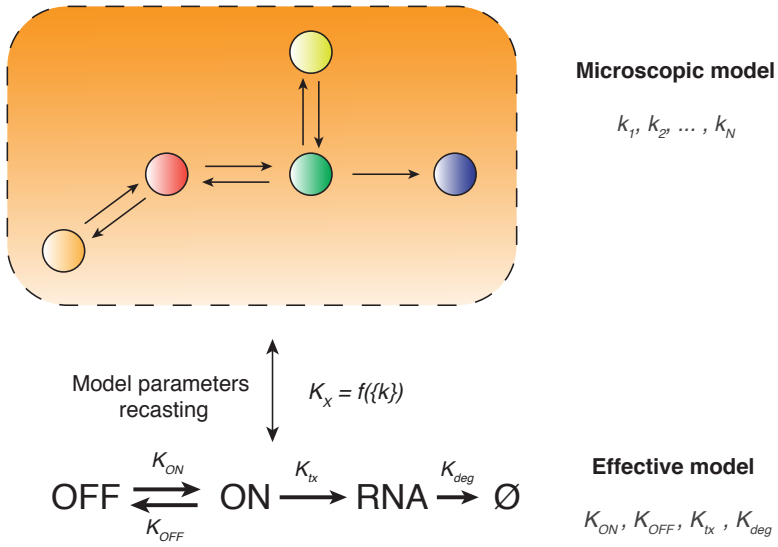
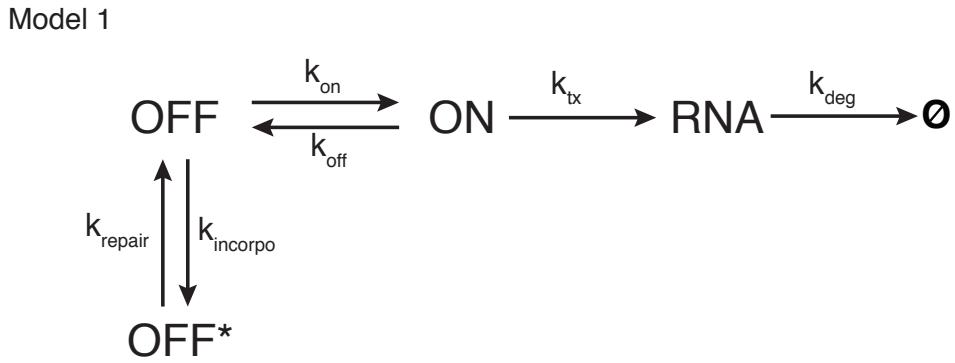


Figure 1: **Model parameters recasting procedure.** From kinetic parameters inferred using the effective model at equilibrium, *via* smFISH experiments, it is possible to derive constraints for microscopic kinetic parameters.

Model 1

This model assumes that IdU incorporation and subsequent interaction of Apex1 with DNA occurs only in the OFF state of the promoter:



Biologically, this would occur if certain additional mechanisms were at play that would inhibit DNA repair during active transcription.

This model is described by the following set of ODEs:

$$\begin{cases} \frac{dOFF}{dt} = k_{ON} \cdot ON + k_{repair} \cdot OFF^* - (k_{incorpo} + k_{ON}) \cdot OFF \\ \frac{dON}{dt} = k_{ON} \cdot OFF - k_{OFF} \cdot ON \\ \frac{dOFF^*}{dt} = k_{incorpo} \cdot OFF - k_{repair} \cdot OFF^* \\ \frac{dRNA}{dt} = k_{tx} \cdot ON - k_{deg} \cdot RNA \end{cases}$$

In accordance with the derivations made for Model 0, the time fraction the system remains in the ON state is defined as follows:

$$\frac{K_{ON}}{K_{ON} + K_{OFF}} \equiv \frac{ON}{ON + OFF + OFF^*} \tag{8}$$

K_{OFF} represents the transition rate to the *effective* OFF state (i.e., comprising all states that are incapable of transcribing mRNA). In this case the *effective* OFF state consists of both OFF and OFF* and the transition into these states is governed by k_{off} and $k_{incorpo}$ respectively.

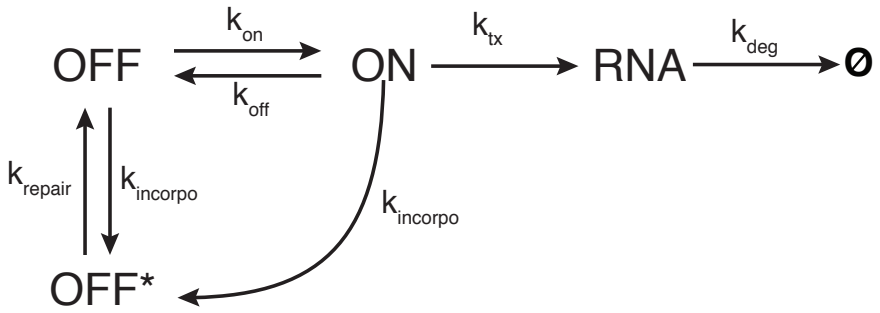
We can thus derive the following constraint between k_{repair} and $k_{incorpo}$:

$$k_{repair} = \frac{K_{ON} + K_{OFF}}{K_{ON}} \cdot \frac{k_{ON}}{k_{ON} + k_{OFF} + k_{incorpo}} - k_{incorpo} \tag{9}$$

Model 2

In this model, Apex1 can interact with DNA in both the ON and OFF states of the promoter. If Apex1 interacts with the DNA in the ON state, this leads to an effective turning *off* of the system (i.e., the system enters the *effective* off state when in the OFF* state).

Model 2



As noted in Model 1, this *effective* off state encompasses all states that are not transcriptionally active due to mechanisms distinct from the typical switching off of a promoter and thus differs from the *canonical* off state (OFF). Once repair is completed, the system enters the *canonical* off state (OFF), which could happen if, for instance, the interaction of Apex1 with DNA may recruit chromatin modifiers (e.g. histone deacetylases, histone methyltransferases) that in turn silence gene expression. As a consequence, stalled polymerases, at both the promoter proximal region or further downstream would unbind DNA.

This model can be described by the following set of ODEs:

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot ON + k_{repair} \cdot OFF^* - (k_{incorpo} + k_{ON}) \cdot OFF \\ \frac{dON}{dt} = k_{ON} \cdot OFF - (k_{OFF} + k_{incorpo}) \cdot ON \\ \frac{dOFF^*}{dt} = k_{incorpo} \cdot OFF + k_{incorpo} \cdot ON - k_{repair} \cdot OFF^* \\ \frac{dRNA}{dt} = k_{tx} \cdot ON - k_{deg} \cdot RNA \end{cases}$$

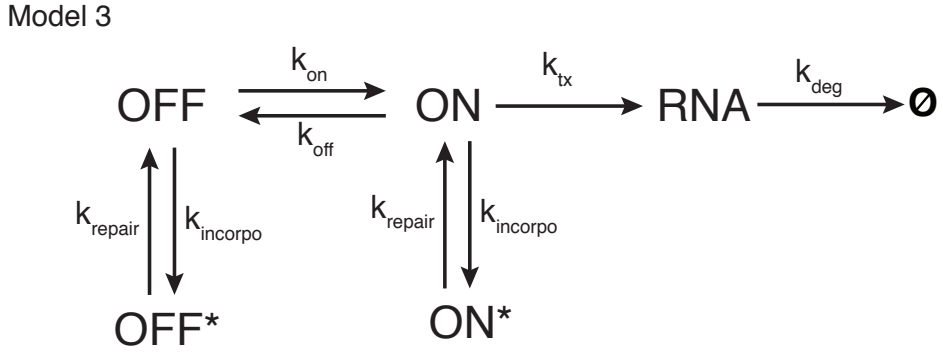
Using equation (8) we derive the following constraint between k_{repair} and $k_{incorpo}$:

$$\frac{k_{incorpo}}{k_{repair}} = \frac{K_{OFF}}{K_{ON}} \cdot \frac{k_{ON}}{k_{OFF}} - 1 \tag{10}$$

where K_{OFF} represents the transition rate to the *effective* OFF state, consisting of both OFF and OFF* and the transition into these states is governed by k_{off} and $k_{incorpo}$ respectively.

Model 3

This model assumes that Apex1 can still interact with IdU in the ON state but that this interaction does not alter the "primed" characteristic of the gene-expression system. In this transcriptionally non-productive ON* state, when Apex1 is recruited to the DNA, active transcription cannot occur (i.e., cannot be measured) but the transcriptional permissiveness of the state is not altered:



Physiologically, this could be achieved if primed polymerases remain paused but do detach from the DNA, or if transcription-enhancing epigenetic marks are not erased. As a result, exiting into this "primed" ON* state would allow for continuation of transcription once repair is completed. Notably, while this ON* state is still primed for continuation of transcription, it contributes to the *effective* off state, since active transcription does not occur and cannot be measured in this state.

This model can be described by the following set of ODEs:

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot ON + k_{repair} \cdot OFF^* - (k_{incorpo} + k_{ON}) \cdot OFF \\ \frac{dON}{dt} = k_{ON} \cdot OFF + k_{repair} \cdot ON^* - (k_{OFF} + k_{incorpo}) \cdot ON \\ \frac{dOFF^*}{dt} = k_{incorpo} \cdot OFF - k_{repair} \cdot OFF^* \\ \frac{dON^*}{dt} = k_{incorpo} \cdot ON - k_{repair} \cdot ON^* \\ \frac{dRNA}{dt} = k_{tx} \cdot ON - k_{deg} \cdot RNA \end{cases}$$

In accordance with the derivations made for Model 0, the time fraction the system remains in the ON state is defined as follows:

$$\frac{K_{ON}}{K_{ON}+K_{OFF}} \equiv \frac{ON}{ON+ON^*+OFF+OFF^*} \tag{11}$$

K_{OFF} represents the transition rate into the effective OFF state, which in this case consists of the OFF, OFF*, and ON* states. The transition into these states is governed by k_{off} and $k_{incorpo}$ respectively. We can thus derive the following constraint:

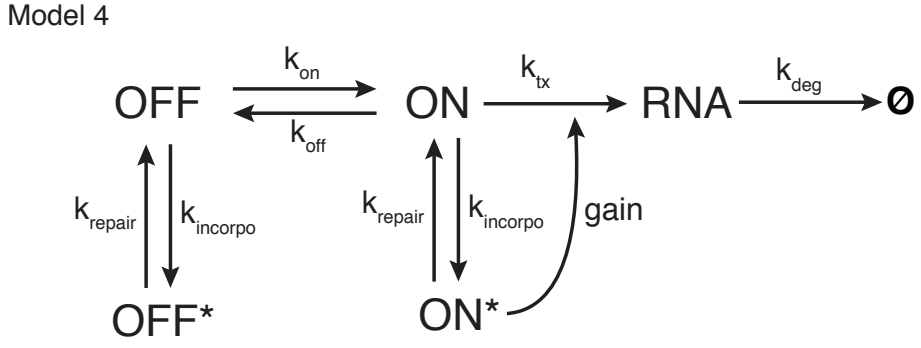
$$\frac{k_{incorpo}}{k_{repair}} = \frac{1}{k_{ON}+k_{OFF}} \cdot [k_{ON} \cdot \frac{K_{OFF}}{K_{ON}} - k_{OFF}] \tag{12}$$

The details of this derivation are described below for Model 4.

Model 4

Model 4 is an extension of Model 3, where a positive transcriptional response to the repair (a coherent feedforward loop) is included. When the system transitions from ON* to ON, the basal transcription rate (k_{tx1}) increases by a multiplicative factor (*gain*), which establishes the feedforward circuit:

$$k_{tx2} = gain \cdot k_{tx1}. \tag{13}$$



Transcription rate (k_{tx}) is a product of the polymerase occupancy and the elongation rate of each polymerase ($k_{tx} = Pol \cdot k_{elong}$) *(91)*. The positive activation (*gain*) on transcription rate (k_{tx}) may be caused by a number of underlying mechanisms impacting either polymerase occupancy and/or elongation rate of polymerases: (i) modification of supercoiling which may facilitate polymerase loading or increase transcriptional elongation upon relaxation of supercoiling (Chong et. al., *Cell* 2014, see Fig. S4D); (ii) by forcing polymerases to pause while the system is in the ON* state, chromatin remains in the accessible state *(92)* causing accumulation of polymerase 'convoys' and increased total polymerase occupancy; or (iii) the accumulation of polymerases in turn can enhance the elongation rate of all bound polymerases *(93)*.

Notably, while the primed promoter state (ON*) is considered part of the effective off state because transcriptional elongation does not occur, when transitioning back into the transcriptionally active (ON) state, the system will retain its primed state resulting in a feedforward (*gain*) on transcription.

This model can be described by the following set of ODEs :

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot ON + k_{repair} \cdot OFF^* - (k_{incorpo} + k_{ON}) \cdot OFF \\ \frac{dON}{dt} = k_{ON} \cdot OFF + k_{repair} \cdot ON^* - (k_{incorpo} + k_{OFF}) \cdot ON \\ \frac{dOFF^*}{dt} = k_{incorpo} \cdot OFF - k_{repair} \cdot OFF^* \\ \frac{dON^*}{dt} = k_{incorpo} \cdot ON - k_{repair} \cdot ON^* \\ \frac{dRNA}{dt} = \langle k_{tx} \rangle \cdot ON - k_{deg} \cdot RNA \end{cases}$$

Where $k_{incorpo} = [IdU] \cdot k_{0,incorpo}$. Here, [IdU] is the concentration of IdU in the cell, which is assumed to remain constant.

At equilibrium :

$$\frac{dOFF}{dt} = \frac{dON}{dt} = \frac{dOFF^*}{dt} = \frac{dON^*}{dt} = \frac{dRNA}{dt} = 0$$

$$\Longrightarrow \begin{cases} OFF_{eq} = \frac{k_{OFF}}{k_{ON}} \cdot ON_{eq} \\ ON_{eq} = \frac{k_{ON}}{k_{OFF}} \cdot OFF_{eq} \\ OFF^*_{eq} = \frac{k_{incorpo}}{k_{repair}} \cdot OFF_{eq} \\ ON^*_{eq} = \frac{k_{incorpo}}{k_{repair}} \cdot ON_{eq} \\ RNA_{eq} = \frac{\langle k_{tx} \rangle}{k_{deg}} \cdot ON_{eq} \end{cases} \tag{14}$$

The relationship between the *effective* ON and *effective* OFF states is defined as:

$$\frac{K_{ON}}{K_{ON} + K_{OFF}} \equiv \frac{ON}{ON + OFF + ON^* + OFF^*} \tag{15}$$

where K_{OFF} represents the transition rate to the *effective* OFF state. For this model the *effective* off state consists of the OFF, OFF*, and ON* states, and is governed by both k_{off} and $k_{incorpo}$. Using the set of equations in (14) and (15) we obtain:

$$\frac{K_{ON}}{K_{ON} + K_{OFF}} = \frac{k_{repair}}{k_{repair} + k_{incorpo}} \cdot \frac{k_{ON}}{k_{ON} + k_{OFF}} \tag{16}$$

Equation (16) can be seen as :

$$P(ON_{effective}) = P(\text{"}Repaired\ state\text{"}) \cdot P(ON)$$

Because $P(ON_{effective}) \equiv P(\text{" Repaired state "} \cap ON)$ we can deduce that "Repaired state" and ON are independent probabilistic events.

Thus, equation (16) can be rewritten as:

$$\frac{K_{ON}}{K_{OFF}} = \frac{k_{repair} \cdot k_{ON}}{k_{OFF} \cdot (k_{repair} + k_{incorpo}) + k_{ON} \cdot k_{incorpo}} \tag{17}$$

Which yields:

$$\frac{k_{incorpo}}{k_{repair}} = \frac{1}{k_{ON}+k_{OFF}} \cdot [k_{ON} \cdot \frac{K_{OFF}}{K_{ON}} - k_{OFF}] \tag{18}$$

Next, we define $\langle k_{tx} \rangle \equiv$ mean transcription rate in the presence of IdU and $k_{tx1} \equiv$ the basal transcription rate in the control condition (DMSO). This yields:

$$\langle k_{tx} \rangle = P(ON \mid OFF) \cdot k_{tx1} + P(ON \mid ON^*) \cdot gain \cdot k_{tx1} \tag{19}$$

where the *gain* term represents the amplification of the transcription rate following completion of repair. $P(ON \mid OFF)$ and $P(ON \mid ON^*)$ represent the probability that the gene transitioned to the ON state from the OFF and ON* states, respectively.

These probabilities read:

$$\begin{cases} P(ON \mid OFF) = \frac{k_{ON} \cdot OFF}{k_{ON} \cdot OFF + k_{repair} \cdot ON^*} \\ \\ P(ON \mid ON^*) = \frac{k_{repair} \cdot ON^*}{k_{ON} \cdot OFF + k_{repair} \cdot ON^*} \end{cases} \tag{20}$$

Next, combining equations (18) and (20) yields:

$$\begin{cases} P(ON \mid OFF) = \frac{k_{OFF}}{k_{OFF} + k_{incorpo}} \\ \\ P(ON \mid ON^*) = \frac{k_{incorpo}}{k_{OFF} + k_{incorpo}} \end{cases} \tag{21}$$

As shown in equations (19) and (21), the effect of the feedforward circuit depends on the frequency of time spent in the ON* state, which in turn depends on $k_{incorpo}$.

Rewriting (19) using (21) we obtain:

$$\frac{\langle k_{tx} \rangle}{k_{tx1}} = \frac{1}{k_{OFF} + k_{incorpo}} \cdot (k_{OFF} + gain \cdot k_{incorpo}) \tag{22}$$

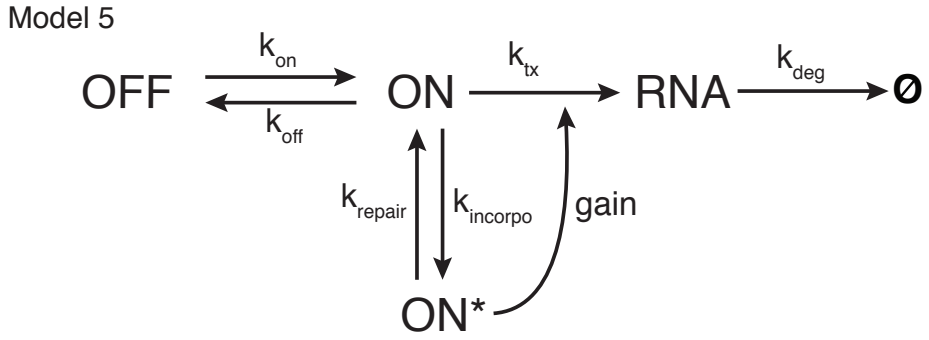
$$\implies gain = \frac{k_{OFF} + k_{incorpo}}{k_{incorpo}} \cdot \frac{\langle k_{tx} \rangle}{k_{tx1}} - \frac{k_{OFF}}{k_{incorpo}} \tag{23}$$

Equation (23) can be rewritten to explicitly consider $[IdU]$:

$$gain = \frac{k_{OFF} + k_{0,incorpo} \cdot [IdU]}{k_{0,incorpo} \cdot [IdU]} \cdot \frac{\langle k_{tx} \rangle}{k_{tx1}} - \frac{k_{OFF}}{k_{0,incorpo} \cdot [IdU]} \tag{24}$$

Model 5

Model 5 makes the same assumptions as model 4 except that the dominant Apex1 interaction with DNA occurs in the ON state. (i.e., the interaction of Apex1 in the OFF state is negligible compared to that in the ON state):



In this model, base excision and repair are coupled to transcription. The ODEs describing this model are as follows:

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot ON - k_{ON} \cdot OFF \\ \frac{dON}{dt} = k_{ON} \cdot OFF + k_{repair} \cdot ON^* - (k_{incorpo} + k_{OFF}) \cdot ON \\ \frac{dON^*}{dt} = k_{incorpo} \cdot ON - k_{repair} \cdot ON^* \\ \frac{dRNA}{dt} = \langle k_{tx} \rangle \cdot ON - k_{deg} \cdot RNA \end{cases}$$

The *gain* expression remains unchanged from Model 4 but the ratio between k_{repair} and $k_{incorpo}$ changes as follows:

$$\frac{K_{ON}}{K_{OFF}} \equiv \frac{ON}{OFF + ON^*} \tag{25}$$

Thus we obtain:

$$\frac{K_{ON}}{K_{OFF}} = \frac{k_{repair} \cdot k_{ON}}{k_{OFF} \cdot k_{repair} + k_{incorpo} \cdot k_{ON}} \tag{26}$$

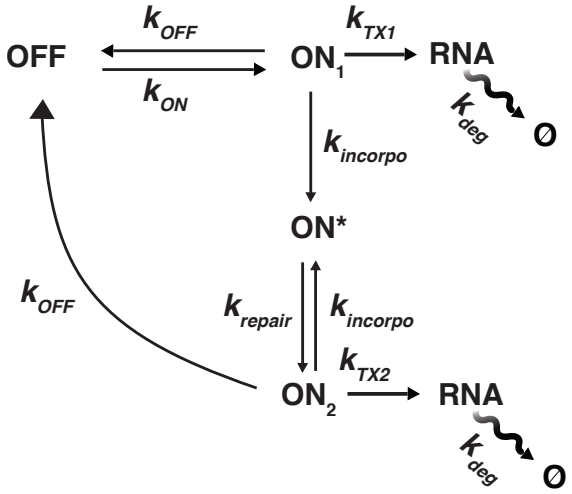
And:

$$\frac{k_{incorpo}}{k_{repair}} = \frac{K_{OFF}}{K_{ON}} - \frac{k_{OFF}}{k_{ON}} \tag{27}$$

3 Chemical Master Equation

For all models we constructed an associated stochastic reaction scheme. As an example, Model 5 can be rewritten using the following scheme in which ON_2 was introduced for computational expediency (see Box 1) and represents the repaired and "primed" ON state of the gene.

$$
\begin{aligned}
OFF &\xrightarrow{k_{ON}} ON_1 \\
ON_1 &\xrightarrow{k_{OFF}} OFF \\
ON_1 &\xrightarrow{k_{tx1}} RNA \\
ON_1 &\xrightarrow{k_{incorpo}} ON^* \\
ON^* &\xrightarrow{k_{repair}} ON_2 \\
ON_2 &\xrightarrow{k_{incorpo}} ON^* \\
ON_2 &\xrightarrow{k_{OFF}} OFF \\
ON_2 &\xrightarrow{k_{tx2}} RNA \\
RNA &\xrightarrow{k_{deg}} \varnothing
\end{aligned}
$$



Box 1
Starting from the stochastic reaction scheme, we can write the corresponding set of ODEs:

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot (ON_1 + ON_2) - k_{ON} \cdot OFF \\ \frac{dON_1}{dt} = k_{ON} \cdot OFF - (k_{incorpo} + k_{OFF}) \cdot ON_1 \\ \frac{dON_2}{dt} = ON^* \cdot k_{repair} - (k_{incorpo} + k_{OFF}) \cdot ON_2 \\ \frac{dON^*}{dt} = k_{incorpo} \cdot (ON_1 + ON_2) - k_{repair} \cdot ON^* \\ \frac{dRNA}{dt} = k_{tx1} \cdot ON_1 + k_{tx2} \cdot ON_2 - k_{deg} \cdot RNA \end{cases}$$

Here we define $ON \equiv ON_1 + ON_2$. As defined earlier, $k_{tx2} \equiv gain \cdot k_{tx1}$. It comes that:

$$\begin{cases} \frac{dOFF}{dt} = k_{OFF} \cdot ON - k_{ON} \cdot OFF \\ \frac{dON}{dt} \equiv \frac{dON_1}{dt} + \frac{dON_2}{dt} = k_{ON} \cdot OFF + k_{repair} \cdot ON^* - (k_{incorpo} + k_{OFF}) \cdot ON \\ \frac{dON^*}{dt} = k_{incorpo} \cdot ON - k_{repair} \cdot ON^* \\ \frac{dRNA}{dt} = k_{tx1} \cdot ON_1 + gain \cdot k_{tx1} \cdot ON_2 - k_{deg} \cdot RNA \end{cases}$$

That is:

$$\begin{cases} \vdots \\ \frac{dRNA}{dt} = \left(k_{tx1} \cdot \frac{ON_1}{ON} + gain \cdot k_{tx1} \cdot \frac{ON_2}{ON} \right) \cdot ON - k_{deg} \cdot RNA \end{cases}$$

By identification, we recover the set of ODEs presented in section 2.

This stochastic scheme represents the ODEs from section 2. ON_2 was introduced for computational expediency (see Box 1) and represents the repaired and "primed" ON state of the gene. This repaired and primed ON state has a higher transcription rate (k_{tx2}), due to the positive feedforward loop (*gain*), where $k_{tx2} = gain \cdot k_{tx1}$. Notably, as shown in equation (19), the effect of this feedforward gain depends on the frequency of time spent in the ON* state, which in turn depends on $k_{incorpo}$ (see Figure S26C).

Using this scheme, we can construct the chemical master equation (CME) describing the time dependent distributions of mRNA copy number *(94)*:

$$\frac{d\mathbf{P}(m,t)}{dt} = \mathbf{A}\cdot\mathbf{P}(m,t) + \delta(\mathbf{E}-\mathbf{I})[m\mathbf{P}(m,t)] + \Delta(\mathbf{E}^{-1}-\mathbf{I})[\mathbf{P}(m,t)] \tag{28}$$

Where $\mathbf{A}$, Δ, and δ are the transition, transcription, and degradation matrices respectively:

$$\mathbf{A} = \begin{bmatrix} -k_{ON} & 0 & -k_{OFF} & -k_{OFF} \\ 0 & -k_{repair} & k_{incorpo} & k_{incorpo} \\ k_{ON} & 0 & -(k_{OFF}+k_{incorpo}) & 0 \\ 0 & k_{repair} & 0 & -(k_{incorpo}+k_{OFF}) \end{bmatrix}$$

$$\Delta = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & k_{tx1} & 0 \\ 0 & 0 & 0 & k_{tx2} \end{bmatrix}$$

$$\delta = \begin{bmatrix} k_{deg} & 0 & 0 & 0 \\ 0 & k_{deg} & 0 & 0 \\ 0 & 0 & k_{deg} & 0 \\ 0 & 0 & 0 & k_{deg} \end{bmatrix}$$

$\mathbf{P}(m,t)$ is a four-element column vector consisting of the time-dependent mRNA probability distributions while in the $OFF, ON^*, ON_1, and ON_2$ states respectively. $\mathbf{E}$ and $\mathbf{E}^{-1}$ are the forward and backward shift operators while $\mathbf{I}$ is the identity matrix.

At steady-state, the mRNA probability distribution can be reconstructed as a sum of the binomial moments *(94)*

$$P(m) = \sum_{k\geq m} (-1)^{m-k} \binom{k}{m} b_k, m = 0,1,2,\ldots \tag{29}$$

Where b_k is the k^{th} binomial moment of the distribution given by:

$$b_k = \frac{1}{\prod_{i=1}^{k} det(i\delta - \mathbf{A})} \cdot \prod_{i=k}^{1} [\mathbf{u}_N (i\delta - \mathbf{A})^* \Delta] \cdot \mathbf{b}_0, k = 1, 2, \ldots \tag{30}$$

where $\mathbf{u}_N$ = [1,1,1,1]. $(i\delta - \mathbf{A})^*$ and $det(i\delta - \mathbf{A})$ are the adjugate and the determinant of matrix $(i\delta - \mathbf{A})$ respectively. b_1 is equivalent to the mean mRNA abundance of the system at equilibrium. $\mathbf{b}_0$ is the corresponding eigenvector for the zero eigenvalue of $\mathbf{A}$. The 4 elements of $\mathbf{b}_0$ therefore represent the fraction of time spent in the $OFF, ON^*, ON_1, and\ ON_2$ states at equilibrium. The n^{th} component of $\mathbf{b}_0$ is given by:

$$b_0^{(n)} = \prod_{i=1}^{3} \frac{\beta_i^{(n)}}{\alpha_i}, 1 \leq n \leq 4 \tag{31}$$

where $\alpha_1, \alpha_2, \alpha_3$ are the three non-zero eigenvalues of $\mathbf{A}$. $\beta_1^{(n)}, \beta_2^{(n)}, \beta_3^{(n)}$ are the three eigenvalues of the sub-matrix $\mathbf{M}_n$ which is constructed by removing the n^{th} row and n^{th} column of $\mathbf{A}$.

The Fano factor for mRNA counts at equilibrium is then given by:

$$FF = \frac{2b_2 + b_1 - b_1^2}{b_1} \tag{32}$$

An exact numerical simulation of such a stochastic process can be generated by Gillespie's method (a.k.a., Doob-Gillespie algorithm) *(95)*. The algorithm was implemented in-house using a Julia 1.1.1 script. For each model 1500 simulations were run for a virtual duration of 200 h. Since the time is not discrete, a 'parsing' algorithm, based on recursive binary search, was used to align all the trajectories on a common time scale of 2000 intervals. For each interval of the discretized time, the average and variance of the number of mRNAs, taking into account all trajectories, was computed. The analytic relationships described above are used to verify the inferred kinetic rates from stochastic simulations.

4 Estimation of model parameters from experimental data

Using smRNA FISH data, we can infer the *effective* kinetic rates (Table S2). The kinetic rates computed in the control condition (DMSO in Table S2) are the starting point for all simulations and were considered as constant for all the models and associated results.

Using these measured *effective* rates (K_{ON}, K_{OFF}, K_{tx}, K_{deg}) along with the relationships developed in section 2, we are left with one remaining degree of freedom for our models: k_{repair} (or equivalently $k_{incorpo}$).

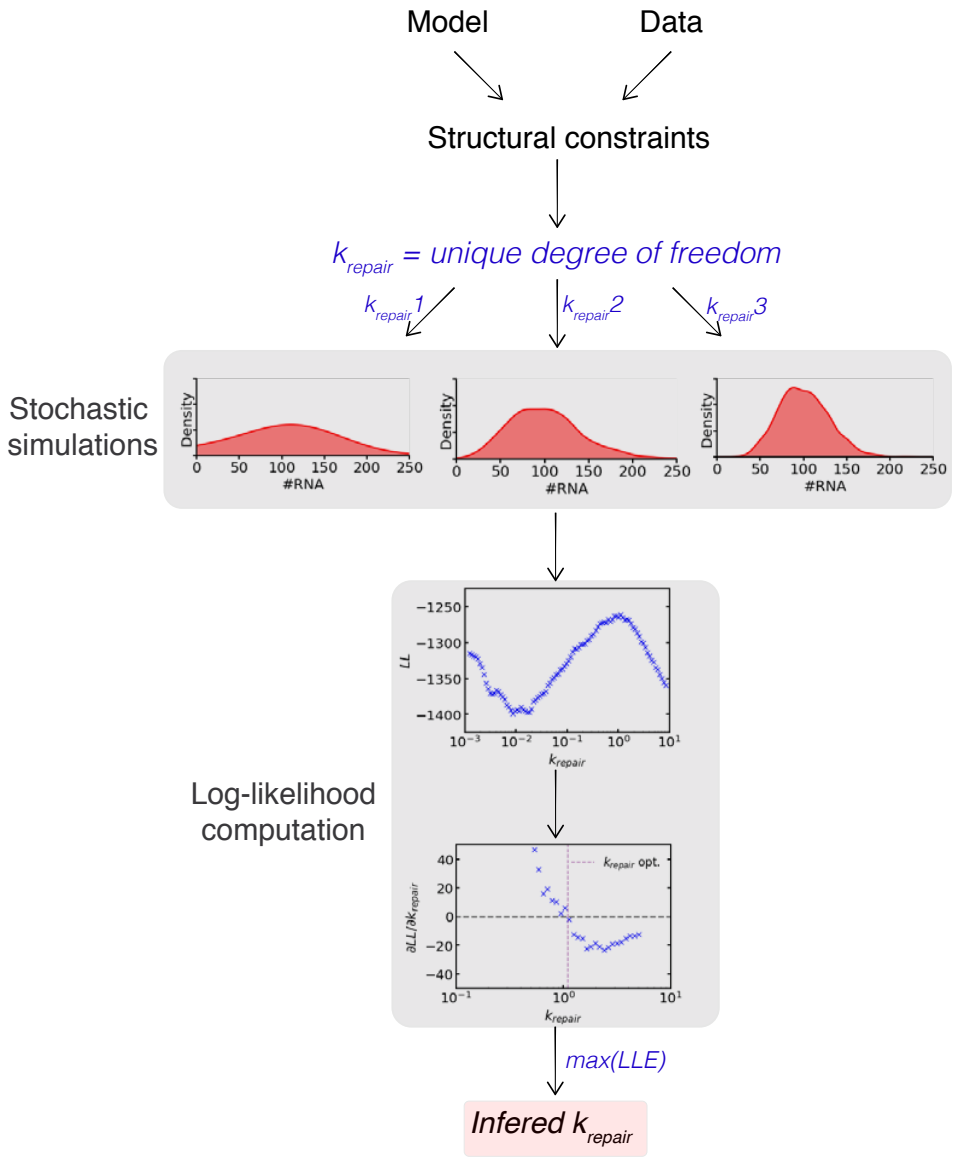
Comparison between experimental data and the results of stochastic simulations for each of the models described in section 2 allows determination of which models most accurately fit the experimental data.

5 Model selection: Comparison of simulation results to experimental data

5.1 Information theory-based approach: Maximum Likelihood Estimation (MLE) and Akaike information criterion (AIC)

5.1.1 Workflow

In this approach, we compared the effective behavior of simulation results to the experimental data. After deriving constraints on the phase space spanned by k_{repair}, each of the models was simulated over a range of k_{repair} values. For each of the 500 chosen values (logarithmically spaced) of our unique degree of freedom (k_{repair}), 50,000 simulations for a total duration of 200h were run. Only the last point of each simulation was stored. Then the steady-state distribution of RNA was computed using the same binning as in the experimental data. From these stochastic simulations, the effective behavior of the system can be extracted for a given model. K_{OFF} and K_{ON} were computed using a non-linear curve fitting assuming exponentially distributed residence times (Poisson process). Both ν_{on} (fraction of time that promoter is active) and the burst size BS were computed using their basic definitions *(96)*. The density of mRNA population was computed using the last points of the simulations. A classic kernel density was used to represent the data.



5.1.2 Basis for Maximum likelihood estimation and AIC computation

The computation of the log-likelihood is as follows:

Let $\mathbf{X}$ the vector of n empirical observations and let $\mathscr{X}_\Delta : \{x_i^\Delta : x \in [x_{i-1}, x_i] \iff x \to x_i^\Delta \quad \forall x \in \mathbf{X}\}$ the set of bins[1] derived from $\mathbf{X}$, where $x_i = x_{i-1} + \Delta$ for $i = 1, \ldots, N$, $x_0 = 0$ and Δ is the size of a bin. We define $\hat{p}(x^\Delta)$ as the binned empirical probability (in fact the observed binned frequency), derived from $\mathbf{X}$, and $P(x^\Delta)$ as the probability of a given bin x^Δ for a given model. In addition in

[1] "$\to$" means here "is associated to"

the following we denoted by $P(x_j)$ the model probability of an observation x_j. The likelihood $\mathscr{L}$ is defined as :

$$\mathscr{L} \doteq \prod_{x_j \in \mathbf{X}} P(x_j) \tag{33}$$

Following the binning procedure described before, we can derive a binned likelihood $\mathscr{L}^{\Delta}$ given by:

$$\mathscr{L}^{\Delta} \doteq \prod_{\substack{x_j \in \mathbf{X} \\ x_j \to x^{\Delta} \in \mathscr{X}_{\Delta}}} P(x^{\Delta}) = \prod_{x^{\Delta} \in \mathscr{X}_{\Delta}} P(x^{\Delta})^{n \cdot \hat{p}(x^{\Delta})} \tag{34}$$

In the following we drop the Δ superscript in $\mathscr{L}^{\Delta}$ and denote the binned likelihood as $\mathscr{L}$. The (binned) log-likelihood is then given as:

$$\log(\mathscr{L}) = \sum_{x^{\Delta} \in \mathscr{X}_{\Delta}} n\hat{p}(x^{\Delta}) \cdot \log(P(x^{\Delta})) \tag{35}$$

In this case, $\mathscr{L}$ is a function of k_{repair}. We then try to find the value of k_{repair} that maximizes $\log(\mathscr{L})$:

$$\hat{k}_{repair,MLE} = \underset{k_{repair}}{\arg\max} \log(\mathscr{L}^{(k_{repair})}) \tag{36}$$

With $k_{repair} \in [10^{-4}, 10]$, by assumption. After computing $\log(\mathscr{L}^{(k_{repair})})$ for each of the 500 values of k_{repair}, we apply a smoothing (moving average) to the data, take the derivative, smooth the derivative, find the two points on each side of the abscissa, and then interpolate the point for which the derivative is equal to zero using a linear interpolation. The maximum $\log(\mathscr{L})$ is computed after the first smoothing. Then, we compute the effective behavior of the system using the inferred value of k_{repair}, namely $\hat{k}_{repair,MLE}$.

The model selection is based on the Akaike information criterion (*AIC*) and the resulting measures $\Delta_i AIC$ and Akaike weights w_i *(97)*. Because $\hat{k}_{repair,MLE}$ is dependent on the empirical distribution, we can assume that it is only an estimate of the *true* value, which we call k^0_{repair}. Thus we want to reduce as much as possible the distance between $\hat{k}_{repair,MLE}$ and k^0_{repair}. This optimization problem allows us to derive the so-called Akaike information criterion (AIC) as a measure to compare models *(97)*. *AIC* is an estimate of the expected relative distance between the fitted model and the unknown true mechanism that actually generated the observed data:

$$AIC = -2\log(\mathscr{L}^{(\hat{k}_{repair,LLE})}) + 2K \tag{37}$$

with K being the number of degrees of freedom of the system.

5.1.3 Results

As shown in Figures S23A-B, Model 5 is selected on the basis of minimized AIC. Model 4 is second-best. Model 5 qualitatively and quantitatively matches experimental data with its inferred k_{repair}.

5.1.4 Parameter identifiability

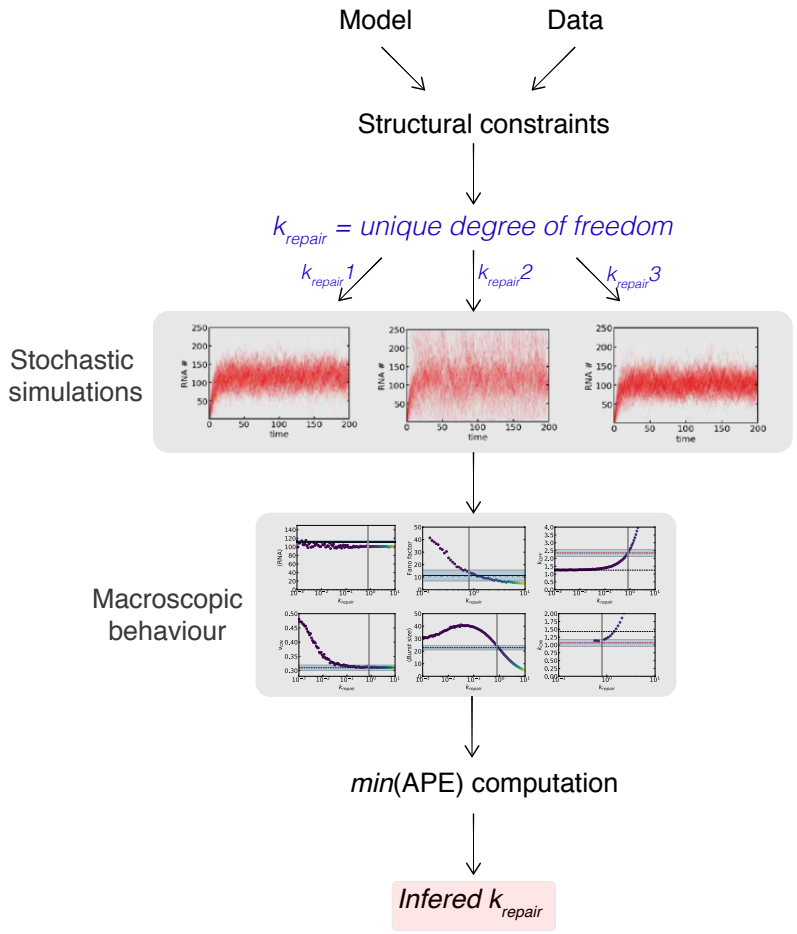
Bootstrapping was used to assess the quality of the inference of k_{repair} for Model 5 using MLE. This method allows the computation of the confidence interval (CI) for k_{repair}, and, in the framework of Bayesianism, a posterior distribution $P(Data|k_{repair})$ using a flat, minimally-informative prior *(97)*. The distribution of the MLE is peaked around a particular value, suggesting parameter identifiability (Figure S23C).

5.2 APE-based approach

5.2.1 Workflow

As in the first approach for model selection, after deriving constraints on the phase space spanned by k_{repair}, each of the models was simulated over a range of k_{repair} values. From these stochastic simulations, the effective behavior of the system can be extracted for a given model.

For each model, inference of the best value of k_{repair} is based on the minimization of a loss function (absolute percentage error, a.k.a. APE). This quantitative approach is coupled to a visual comparison of the model behavior to experimental data. The model whose inferred k_{repair} value minimizes divergence between model and data behavior is thus chosen. A graphical representation of such a process is given below:



5.2.2 Method of APE calculation

To discriminate between models based on their effective behavior, we utilized a measure that quantifies the discrepancy between model-derived results and experimental data. Because the effective observations (Fano factor, K_{ON}, burst size BS, etc.) derived from smFISH experiments are of different orders of magnitude, we utilized a relative measure to avoid the largest parameters carrying the highest weight on the error. The absolute percentage error (APE) satisfied these constraints. The procedure is as follows:

Consider the vector $\mathbf{M} \equiv (\langle RNA \rangle, \nu_{ON}, BS, FF, K_{ON}, K_{OFF})$. Thus, $\mathbf{M}_{model}$ and $\mathbf{M}_{exp}$ contain all the effective observations from modeling and experiment respectively. $\mathbf{M}_{model}$ is a function of k_{repair} or $k_{incorpo}$ equivalently. $\hat{k}_{repair,APE}$ is the best inferred value of the degree of freedom for a particular model. It is given by:

$$\hat{k}_{repair,APE} = \text{argmin}_k \left| \frac{\mathbf{M}_{exp} - \mathbf{M}_{model}(k)}{\mathbf{M}_{exp}} \right| \tag{38}$$

With $k \in [10^{-4}, 10]$, by assumption. This notation implies minimization of the $\ell 1$ norm (sum of the vector components). In an operative manner, we simulated each model for 250 logarithmic distributed values of k_{repair} using the previously described Doob-Gillespie algorithm *(95)*.

To validate our approach, we devised an alternative loss function where the $\ell 1$ norm is computed using a non-biased (i.e symmetric) measure of relative prediction accuracy: the absolute log accuracy (ALA). The procedure is as follows:

$$\hat{k}_{repair,ALA} = \text{argmin}_k \left| \log \left(\frac{\mathbf{M}_{model}(k)}{\mathbf{M}_{exp}} \right) \right| \tag{39}$$

With $k \in [10^{-4}, 10]$, by assumption.

5.2.3 Results

Both the APE and ALA approaches yielded *exactly* the same results for parameter inference (except for model 1 for which $\hat{k}_{incorpo,ALA} = 0.39$) and model selection. According to the APE-based approach for parameter inference and model selection, model 5 best recapitulates the observed data of increased expression noise without change in mean expression (Figure S24) and is referred to as the **D**iscordant **T**ranscription t**h**rough **R**epair (DiThR) model.

5.3 Selection of Model 5 (DiThR Model)

Based on both the MLE- and APE-based approaches, model 5 (DiThR model) best matches experimentally measured mRNA distrubtions for both FISH data and scRNA-seq data for Nanog (Fig. S23-24). The DiThR model utilizes a repair mechanism that is dependent on a transcriptionally active state, in which repair events only occur while the gene is transcriptionally permissive (i.e., in the ON state). The key adjustment we make is the introduction of a coherent feedforward loop caused by the system entering a primed yet transcriptionally inactive state (ON*). Completion of this repair results in a higher transcription rate due to either: (i) modification of supercoiling *(57)* which in turn has a feedforward effect on transcriptional elongation once continuation of transcription is enabled; (ii) by forcing polymerases to pause while the system is in the ON* state, chromatin remains in the accessible state *(92)* causing accumulation of polymerases and increased total polymerase occupancy; (iii) the accumulation of polymerases in turn can enhance the elongation rate of all bound polymerases *(93)*.

This feedforward loop is supported by the model-selection process, as the second-best model (model 4) also has an amplified transcription rate following Apex1 interaction and subsequent

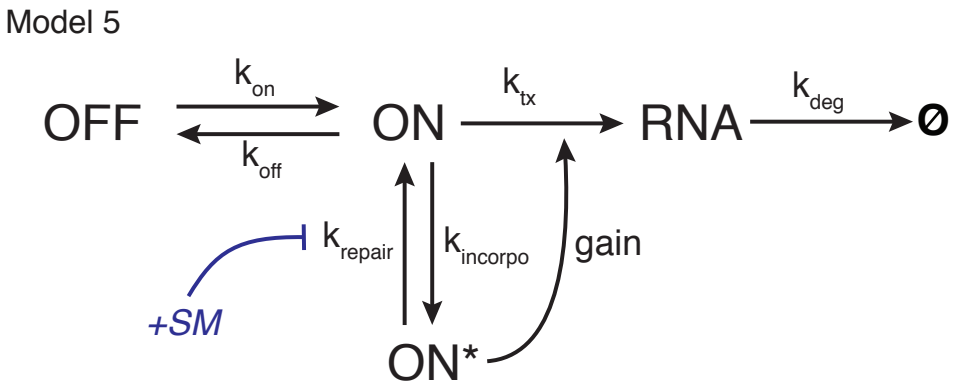
DNA repair. Interestingly, for the same value of k_{repair}, this transcription-coupled base excision mechanism (model 5) leads to a less significant increase in noise and better maintenance of mean as compared to model 4 in which repair can also take place in the OFF state (model 4). This implies that coupling of base excision to transcription is the most efficient method of DNA repair in terms of minimizing excess transcriptional variability.

It is important to note that the dynamic binding and unbinding of Apex1 triggers noise enhancement more than the repair per-se. One implication is that other protein-DNA dynamic interactions may lead to unavoidable noise modulation through structural constraints like supercoiling. The strength of such modulation will depend on the kinetic rates of interaction.

6 Validation of Model 5 (DiThR Model)

To validate that it is indeed the feedforward loop (i.e., *gain*) and not simply an increase in transcription rate (i.e., higher k_{tx}), that causes the experimentally observed increase in Fano with no effect on the mean, we modified Model 5 to include two additional scenarios: i) no positive feedforward (Figure S25A); and ii) no positive feedforward together with a manually increased k_{TX} (Figure S25B). These two scenarios were compared with Model 5 (Figure S25C) for ability to fit the observed data. Model 5 was again found to be the most accurate model at fitting the experimental data.

A perturbative approach was followed to test whether the DiThR model can recapitulate experimental data of the Nanog gene expression system in the presence of topoisomerase inhibitors. We computationally tuned k_{repair} to predict the behavior of the gene expression system. First, k_{repair} was decreased by a factor n (akin to addition of a small molecule '+SM' topoisomerase inhibitor), which increases the average time spent in the ON* state by the same factor. Decreasing k_{repair} is meant to simulate the effect of topoisomerase inhibition (+SM) which is thought to trap the system in the ON* state for a longer time due to reduced rate of supercoiling relaxation, represented in the following schematic.



When the system is in the ON* state, it exhibits Poisson dynamics, characterized by exponentially distributed escape time, τ, with rate k_{repair}/n. It follows that $\langle\tau\rangle = n/k_{repair} = n \cdot \langle\tau\rangle_{n=1}$. The subscript $n = 1$ in the previous expression indicates the average value in the absence of perturbation (SM = 0). Densities were computed using the final point of 10000 Gillespie's simulations.

As a first approximation, we expect that an increase in residence time in the ON* state should be translated into an equivalent increase in feedforward strength. In other words, the longer the system stays in the ON* state, the higher the feedforward gain, and consequent enhancement of transcription rate (k_{tx}). Interestingly, we observe (Figure S24C) that significant perturbation ($n = 2$) of k_{repair} slightly changes average gene expression ($\times 1.41$) while leading to a striking increase in

Fano factor ($\times 3.12$).

The modeling results indicate that noise-without-mean amplification is positively correlated with time spent in the ON* state. If Apex1-induced supercoiling is a mediator of both transcriptional repression (time in ON* state) and feedforward gain, it stands that modulation of baseline supercoiling levels should influence transcriptional noise in the presence and absence of BER activity *(58)*. We experimentally verify this through overexpression of Topoisomerase I (Fig. S21C) and inhibition of topoisomerase catalytic activity using small molecules such as topotecan and etoposide (Figure S21B). Inhibition of topoisomerase catalytic activity yields a stark increase in Fano factor ($\times 3.16$ for topotecan and $\times 5.03$ for etoposide), while the mean remains comparable ($\times 0.98$ for topotecan and $\times 1.16$ for etoposide). Conversely, overexpression of Topoisomerase I reduced IdU-mediated noise enhancement indicating that increased rates of supercoiling relaxation reduces time spent in ON* state and thus attenuates the noise modulatory properties of BER. Model 5 therefore captures the effects of topoisomerase —and thus supercoiling —perturbation.

Importantly, modeling of topoisomerase inhibition through reduction of k_{repair} can be generalized to a broad range of BER activity (i.e., range of $k_{incorpo}$ rates). To constrain the modeling by experimental data, we modeled AP-site formation using the kinetic rates derived from IdU experimental data: in the absence of IdU, the baseline noise level should change (reduced BER activity), but not the qualitative modulation of the Fano factor in response to topoisomerase inhibition. This is shown quantitatively in the following sensitivity analyses (Section 7.1).

7 Sensitivity analysis of Model 5 (DiThR Model)

7.1 Modulation of k_{repair} and $k_{incorpo}$ with fixed feedforward strength

The model incorporating Apex1 interaction with chromatin, and an associated transcriptional amplification (i.e., feedforward loop), can recapitulate the noise-without-mean amplification of Nanog expression observed with IdU treatment. We next asked how this behavior – increase in noise without substantial modification of mean expression – is related to the dynamics of Apex1 binding and unbinding (k_{repair} and $k_{incorpo}$ describe this interaction).

We conducted a phase-plane analysis of the mean and Fano factor for both k_{repair} and $k_{incorpo}$ (Fig S26A-B). We assumed that the feedforward strength is fixed and equal to the deduced feedforward strength from the previous analysis using Nanog RNA FISH data for $10\mu M$ IdU. As expected, the Fano factor increases as $k_{incorpo}$ increases (for $k_{incorpo}$ lower than ≈ 1). This demonstrates a positive dose-dependent relationship between IdU and noise (Fig S26B).

We observe an inverse relation for k_{repair}, where noise increases as k_{repair} decreases. Experimentally, the small-molecule inhibitor of the Apex1 endonuclease domain (CRT0044876) was used to decrease k_{repair}. We note that when $k_{incorpo}$ is higher than ≈ 1 the Fano factor starts to decrease (Fig. S26B). These observations can be understood examining equation (40) for the Fano

factor:

$$FF = 1 + \frac{(1-\nu_{on}) \cdot k_{tx}}{K_{ON} + K_{OFF} + k_{decay}} \tag{40}$$

For $k_{incorpo} > 1$, ν_{on} decreases slowly and the effective $\langle k_{tx} \rangle$ starts to increase slowly as compared to when $k_{incorpo} \in [0.1, 1]$ (Fig S26C-D). These changes are counteracted by a larger increase in K_{OFF}. The behavior for the mean number of RNA produced with increasing $k_{incorpo} > 1$ can also be understood using the previous considerations: the decrease in mean corresponds to a decrease of the frequency in the ON state that is not counteracted by a strong enough feedforward gain.

7.2 Modulation of k_{ON} and k_{OFF} with fixed feedforward gain

We next wanted to define the parameter regime for k_{ON} and k_{OFF} in which homeostatic maintenance of mean expression is possible with the DiThR model.

For this analysis, simulations were run with values of $k_{ON}, k_{OFF} \in [10^{-3}, 10]$ for both the null model (model 0, DMSO condition) and DiThR model. For the same values of k_{ON}, k_{OFF}, the fold change in mean of mRNA counts was calculated by comparing results of DiThR model to the null model. This provides insight into how IdU treatment may impact expression of genes with different bursting kinetics. When $k_{OFF} \gg k_{ON}$, the addition of IdU in DiThR model increases the average number of mRNA produced as compared to the null model (Fig S26E). This can be explained by a competition between the OFF and ON^* states and by the fact that in this portion of the phase space $k_{ON} < k_{OFF} < k_{incorpo}$ (and $k_{incorpo} < k_{repair}$). Therefore, the probability of presence in the ON_2 state (transcriptionally more productive state), increases. This implies that IdU treatment would increase the mean of very lowly expressed genes. This was seen experimentally in bulk RNA-seq measurements of transcript abundance in mESCs as the ~ 100 genes that showed an increase in mean with IdU treatment were from the lowest expression regime (Fig S3). The exact inverse effect is observed in the upper left corner of the heatmap, where $k_{ON} > k_{OFF} > k_{incorpo}$. Thus for highly expressed genes, the effect of IdU on mean expression is minimal, as seen experimentally.

7.3 Feedforward strength and noise enhancement

When all the kinetics rates of the system are fixed, increasing the feedforward gain, and thus the effective transcription rate, leads both to an increase in mean and Fano factor (Fig S26F).

8 Testing of DiThR Model for other noise-enhanced genes

The analysis above indicated that IdU mediates the recruitment of Apex1 (and perhaps other DNA base-excision repair machinery) to DNA while the gene is transcriptionally permissive (DiThR model). By postulating a second ON state characterized by increased k_{tx} the model is sufficient to recapitulate the experimental observations of increased noise while mean is relatively unchanged.

However, this model was parameterized and validated based on the Nanog expression data. Consequently, we next asked whether the DiThR model could explain the experimental data collected for other noise-enhanced genes within the scRNA-seq dataset (Fig. 1C).

To simulate DiThR model for additional genes, estimates for the following parameters were needed: k_{ON}, k_{OFF}, k_{tx1}(basal transcription rate in DMSO), $k_{deg}, and \langle k_{tx} \rangle$ (effective transcription rate in IdU). To derive these parameter estimates, we used an established moments-matching technique described in *(70)*, where the first, second, and third exponential moments of the mRNA distributions in the DMSO and IdU conditions are used to calculate the parameters of a Poisson-beta distribution (that describes the two-state model) which best fits experimental count data. The parameter estimates were derived in proportion to k_{deg}. Values of k_{deg} were retrieved from an existing dataset of mRNA degradation rates in mESCs *(87)*.

Of the 945 genes classified as highly variable with IdU treatment, 314 genes remained for downstream analysis based on availability of mRNA degradation rates and high confidence parameter estimates (Fig. S28A). Using the above parameter estimates, we computed the constraints on the feedforward (*gain*) term and $k_{incorpo}$ as a function of k_{repair} using the relationships derived in Section 2 for Model 5. As above, there was again one remaining degree of freedom in our model system: k_{repair}. Using the MLE-based approach (Section 5.1), simulation results for a range of k_{repair} values were compared against scRNA-seq data to identify $\hat{k}_{repair,LLE}$ for each of the 314 genes.

Once $\hat{k}_{repair,LLE}$ was identified, the effective values of K_{ON} and K_{OFF} from simulation results were compared to experimentally derived estimates of K_{ON} and K_{OFF} from scRNA-seq data for each gene (Fig. S29). Overall, simulated values for effective rates of promoter toggling in the IdU condition align with experimental results, indicating that the DiThR model holds explanatory power for noise-enhanced genes beyond just Nanog. This suggests that the use of a second, transcriptionally-enhanced ON state may be a unified mechanism for maintenance of transcriptional homeostasis during base excision repair across a broad range of genes with different bursting kinetics.

Supplementary Figures S1-S32

Supplemental Figure 1

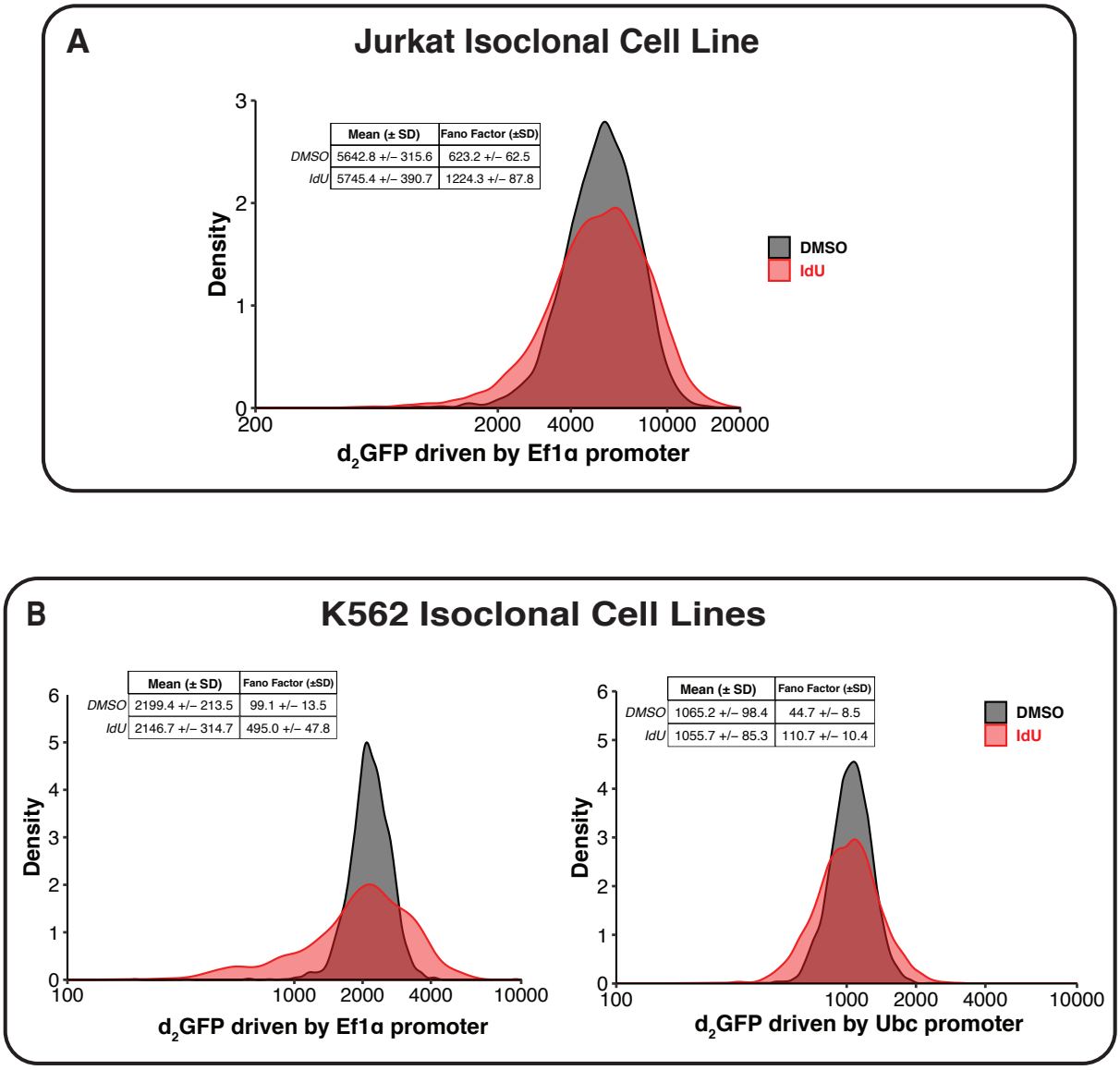


Figure S1: Nucleoside analog increases expression variability of housekeeping promoters in Jurkat and K562 cells.
(A) Representative flow cytomtery distributions of d_2GFP expression in an isoclonal population of Jurkat cells treated with either 20μM IdU or equivalent volume DMSO for 24 hours. Mean and SD are derived from 2 biological replicates. **(B)** Representative flow cytomtery distributions of d_2GFP expression in isoclonal populations of K562 cells treated with either 20μM IdU or equivalent volume DMSO for 24 hours. Mean and SD are derived from 2 biological replicates.

Supplemental Figure 2

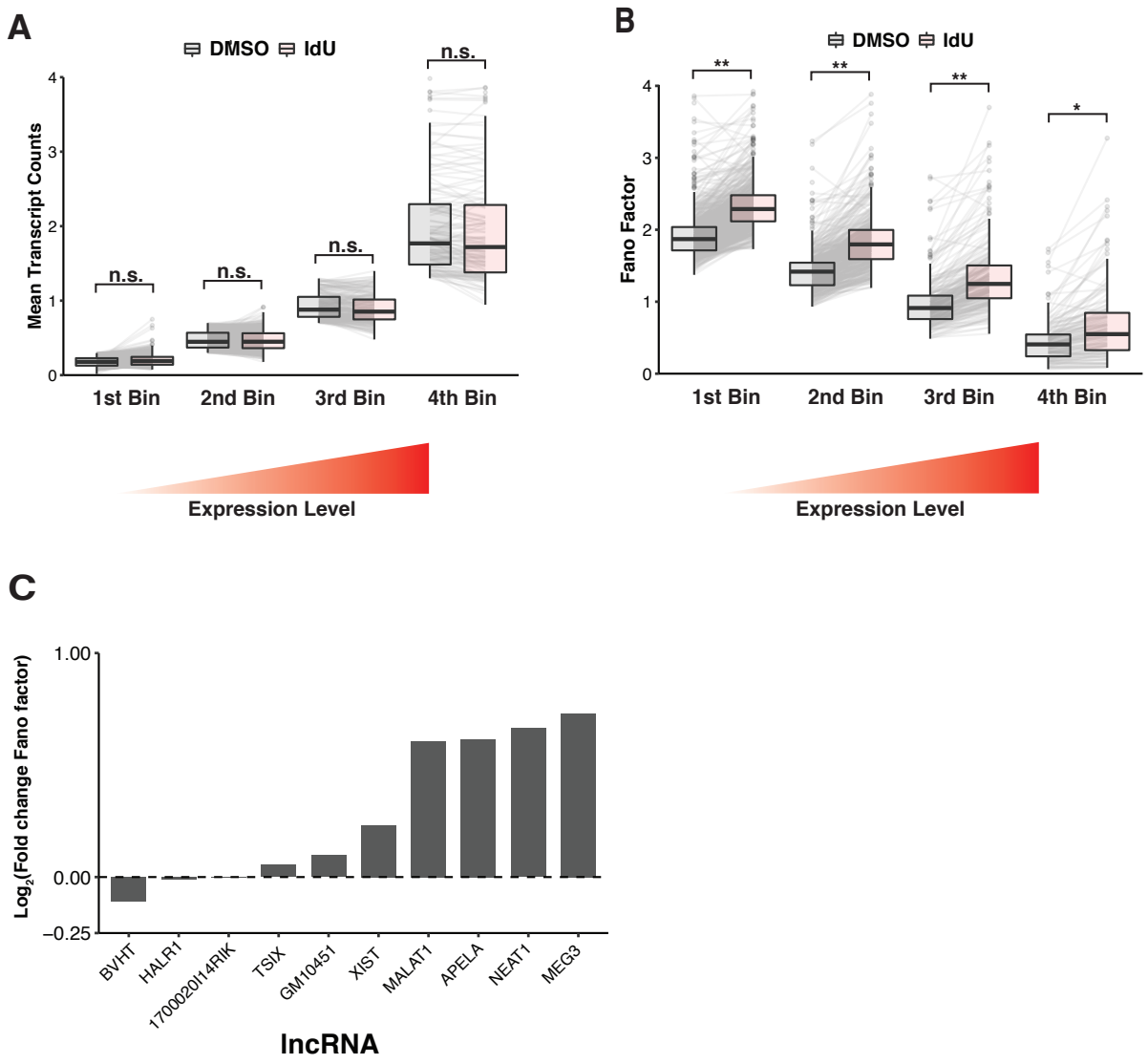


Figure S2: Noise enhancement occurs for coding and non-coding genetic elements across all expression levels.
(A) 4,578 genes from scRNA-seq dataset were binned into one of four groups (quartiles) based on mean expression level in DMSO condition. Comparison of mean expression level for each gene in IdU and DMSO treatment groups. Boxplots show median ±interquartile range of mean values for genes within each bin. Solid lines connect the same gene in the DMSO and IdU boxplots. P values were calculated using a two-tailed, paired Student's t test. **(B)** Comparison of Fano factor for each gene in IdU and DMSO treatment groups. Boxplots show median ±interquartile range of Fano factors for genes within each bin. Solid lines connect the same gene in the DMSO and IdU boxplots. P values were calculated using a two-tailed, paired Student's t test. **$p < 0.001$, *$p = 0.0016$ **(C)** Fold change (Log_2) in Fano factor for expression of 10 long non-coding RNAs (lncRNAs). Comparison is between 10μM IdU and DMSO treatments. lncRNAs with greater than 100 total counts per sample in scRNA-seq dataset were used for analysis. Majority of lncRNAs show enhancement of expression variability with IdU treatment.

Supplemental Figure 3

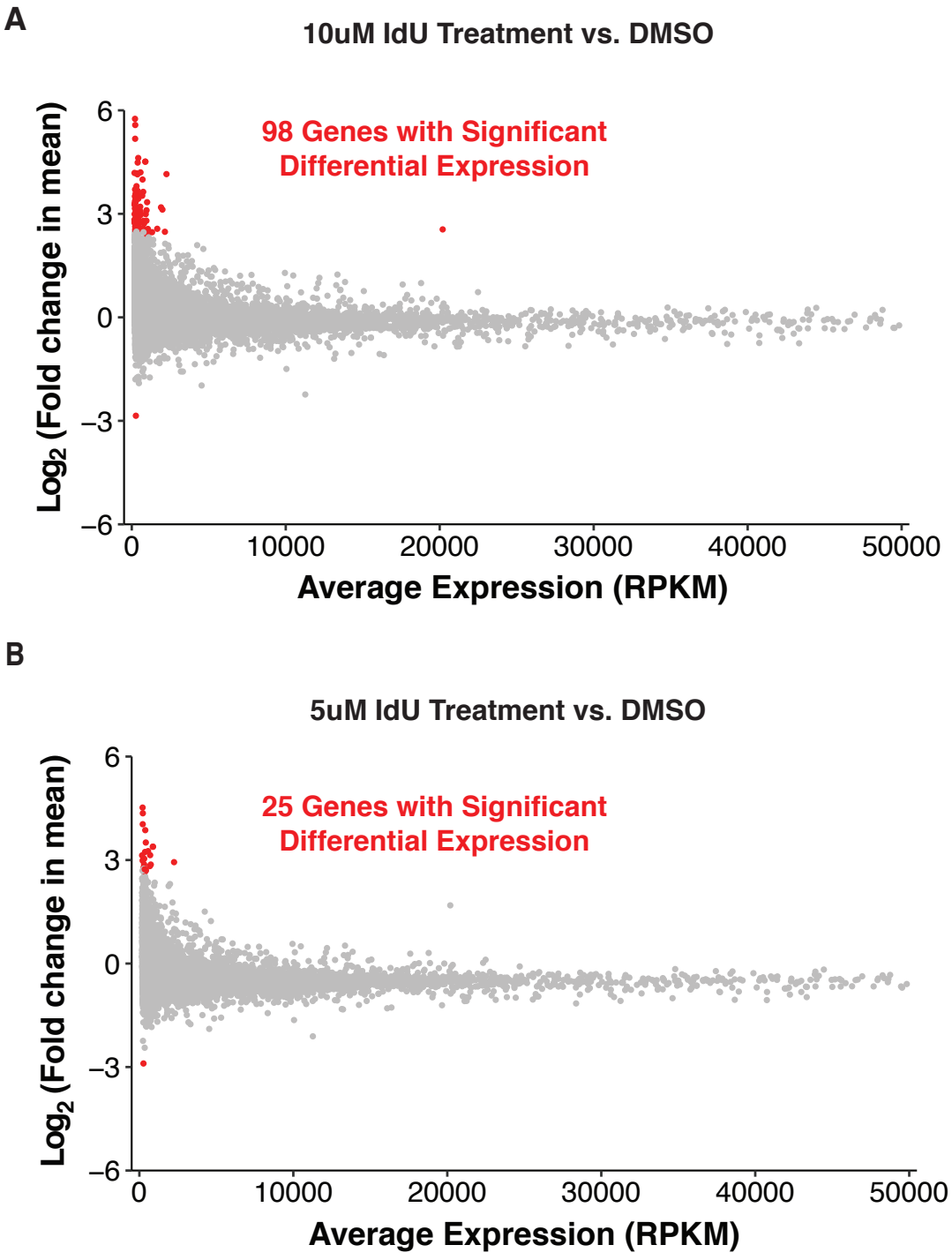


Figure S3: IdU causes minimal change in mean gene expression levels as measured by bulk RNA-seq.
Transcript abundances were normalized using ERCC spike-in counts. Differential mean testing was conducted with a threshold of fold change > 2 and an FDR cutoff of 0.05. Genes considered differentially expressed are highlighted in red. **(A)** Mean transcript abundance vs. fold change (Log_2) in mean for 12,502 genes. Comparison is between 10μM IdU and DMSO treatments. 98 genes were identifed as differentially expressed. **(B)** Mean transcript abundance vs. fold change (Log_2) in mean for 12,054 genes. Comparison is between 5μM IdU and DMSO treatments.

Supplemental Figure 4

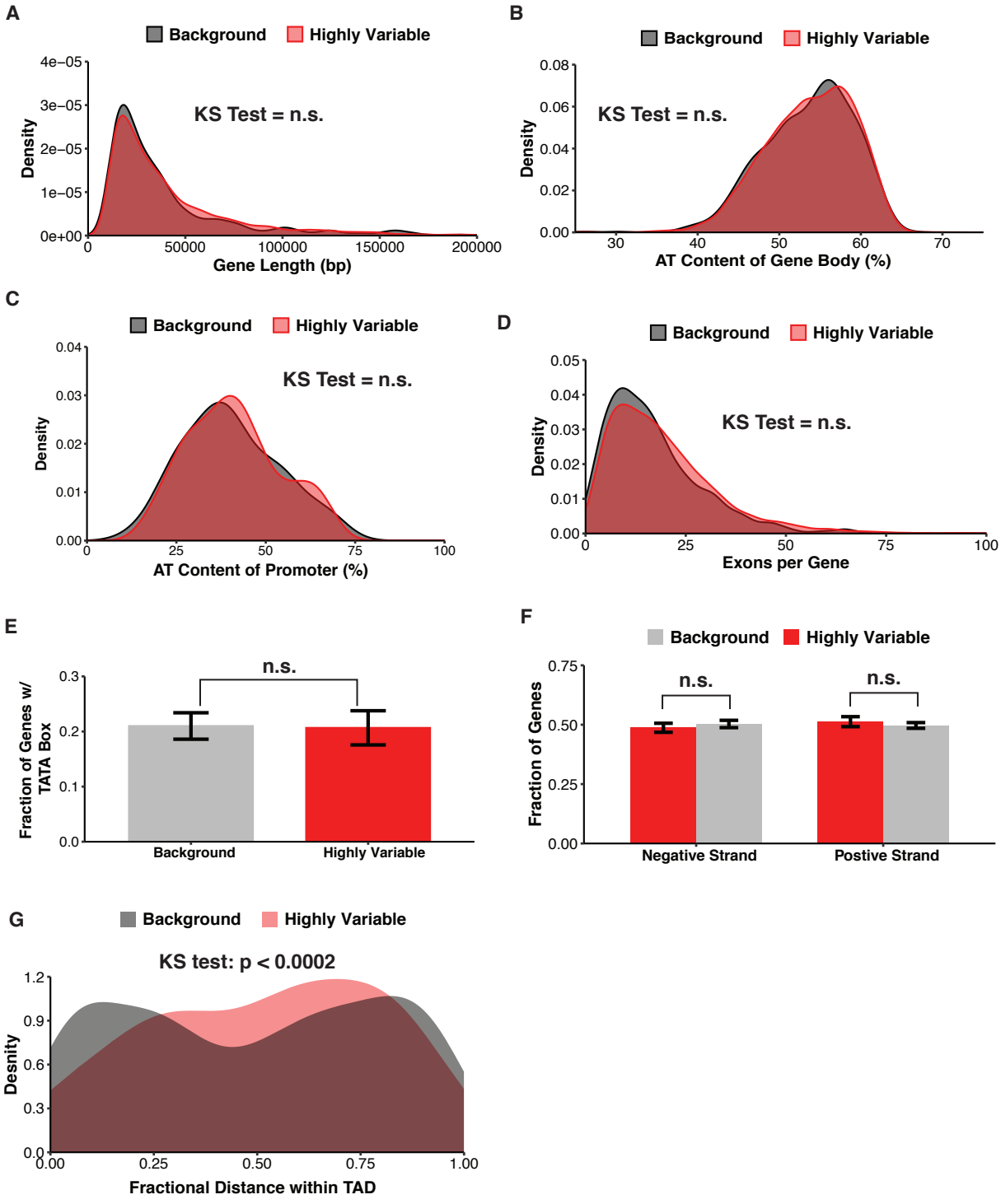


Figure S4: Noise-enhanced genes tend to be centrally located within topologically associating domains.
Comparisons are between 945 genes classified as highly variable and 3,513 genes classified as non-variable (background) according to BASiCS algorithm. Gene characteristics and sequences were taken from Ensembl GRCm38 reference genome. **(A)** Distributions of gene lengths for highly variable and background genes. Length was calculated as distance between Ensembl gene start and end coordinates which correspond to outermost transcript start and end coordinates. **(B)** Percentage of base-pairs in gene body (based on gene start and end coordinates) that are A:T. **(C)** Percentage of base-pairs in 200bp region upstream of gene start that are A:T. **(D)** The number of exons was averaged over all transcripts associated with a gene. Distributions of average exon quantity for genes in the highly variable and background group were then plotted. **(E)** Fraction of genes with TATA sequence in 200bp region upstream of gene start. Data represent mean and SD from bootstrapping procedure with 10,000 resamplings of 100 genes from each group with replacement. **(F)** Fraction of genes whose coding sequence is located on negative and positive strands. Data represent mean and SD from bootstrapping procedure with 10,000 resamplings of 100 genes from each group with replacement. **(G)** Fractional distance of gene within TAD was calculated as (gene start coordinate - TAD start coordinate)/(TAD end coordinate - TAD start coordinate). Coordinates of TAD boundaries in mESCs were taken from Elphege et al. *(80)*.

Supplemental Figure 5

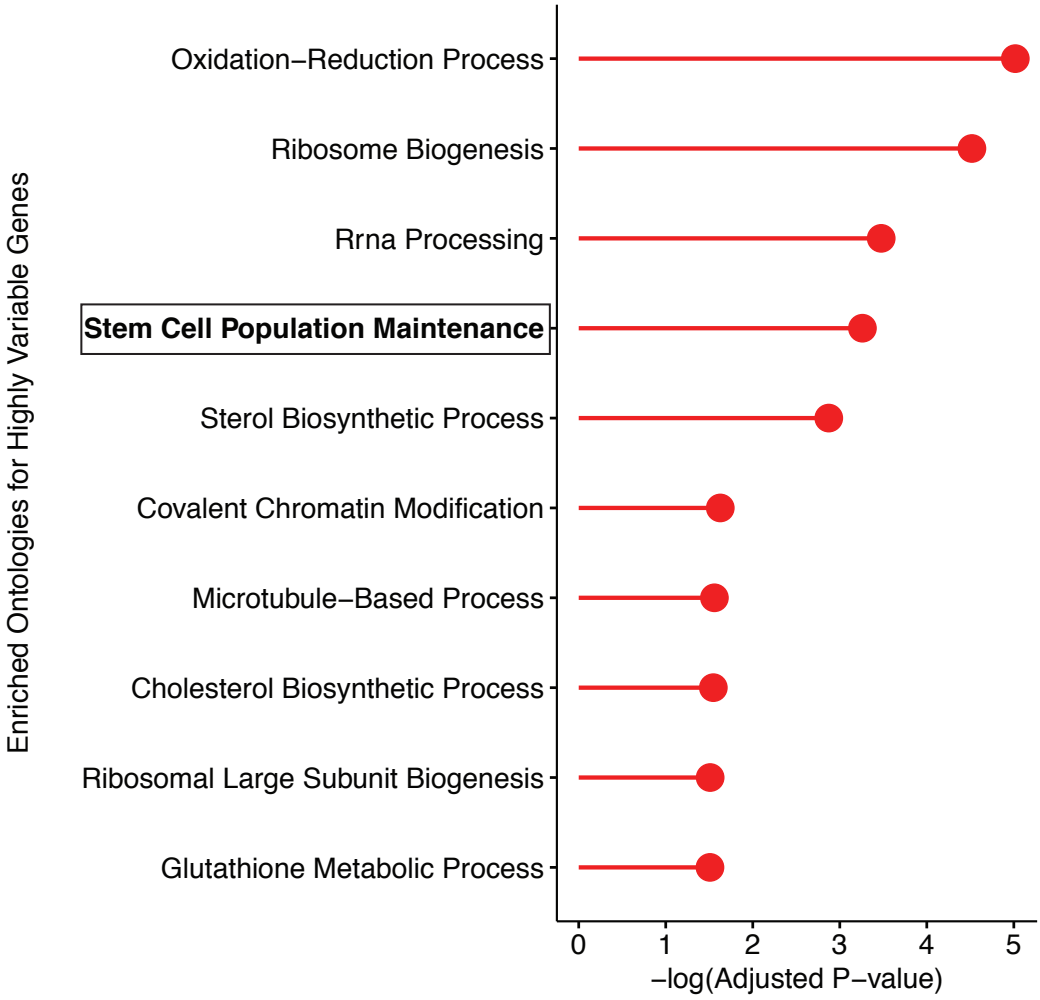


Figure S5: Ontology analysis of variably expressed genes shows enrichment for housekeeping and pluripotency maintenance pathways.
DAVID v6.8 was used to identify enriched ontologies among the 945 genes classified as highly variable according to BASiCS algorithm.

Supplemental Figure 6

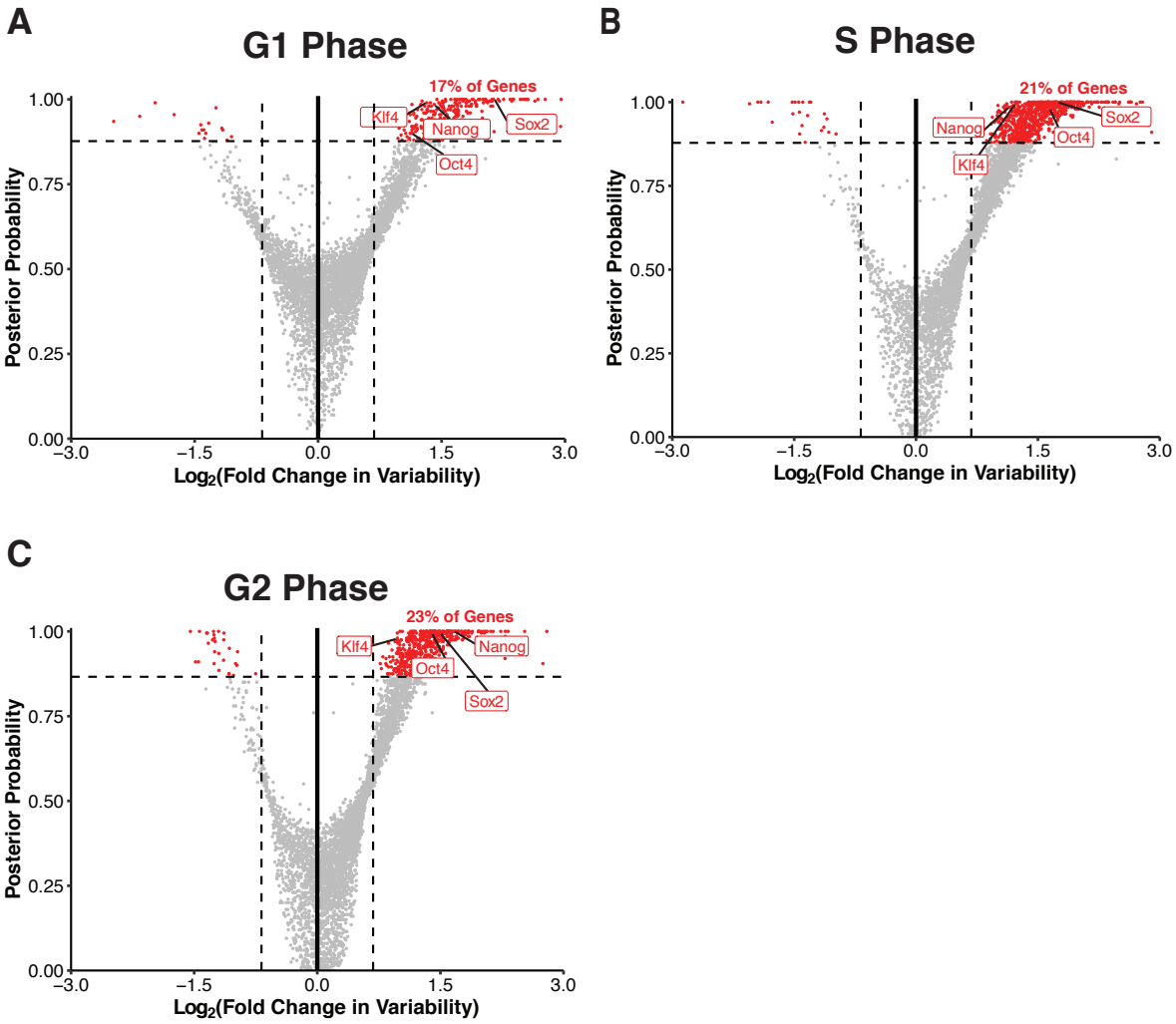


Figure S6: Noise-enhancement of pluripotency factors occurs in all three phases of the cell cycle.
A total of 1556 cells in the scRNA-seq dataset were classified into one of three cell-cycle phases. Differential variability testing was then conducted between cells in the DMSO and IdU treatment groups with the same cycle classification. 17%, 21% and 23% of genes are classified as highly variable in G1, S, and G2 phases respectively.

Supplemental Figure 7

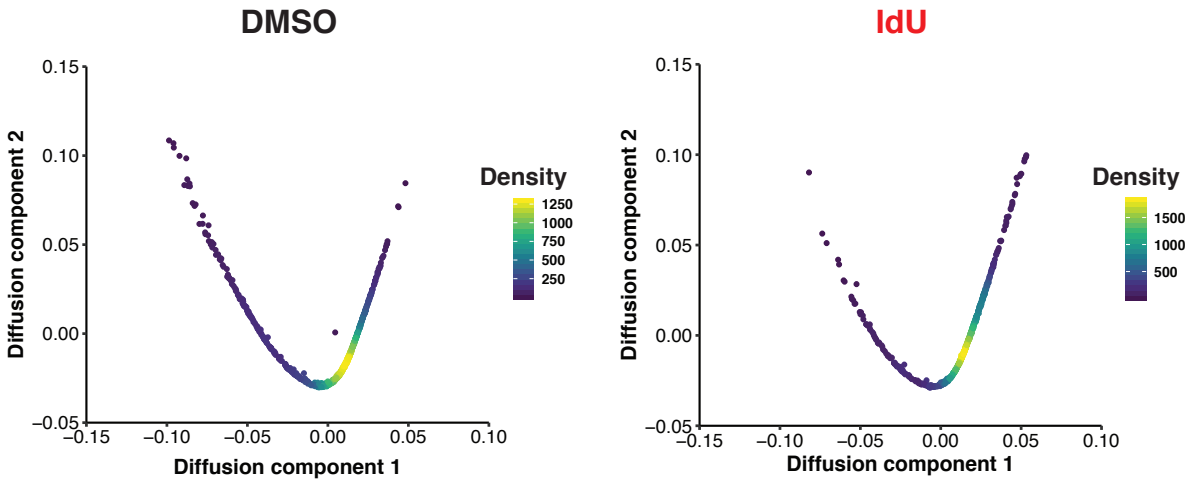


Figure S7: Transcript variability is not caused by bifurcation of mESCs into separate developmental lineages.
Pseudotime analysis of IdU-treated cells shows no differentiation of mESCs into separate developmental lineages.

Supplemental Figure 8

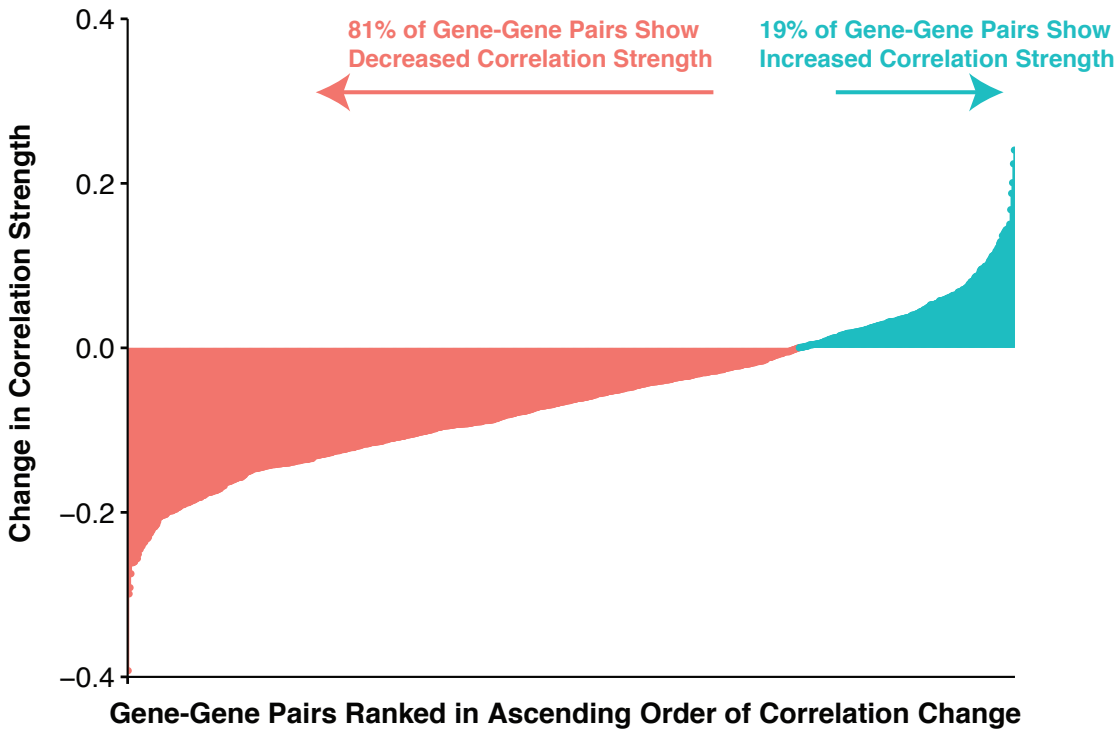


Figure S8: Majority of gene-gene pairs show a decrease in correlation strength.
The Pearson correlation of expression for 923,521 (961 x 961) gene-gene pairs were compared between DMSO and IdU treatment groups. For each gene-gene pair, the absolute value of the correlation strength in DMSO was subtracted from the absolute value of the correlation strength in IdU. Negative values indicate loss of correlation in expression.

Supplemental Figure 9

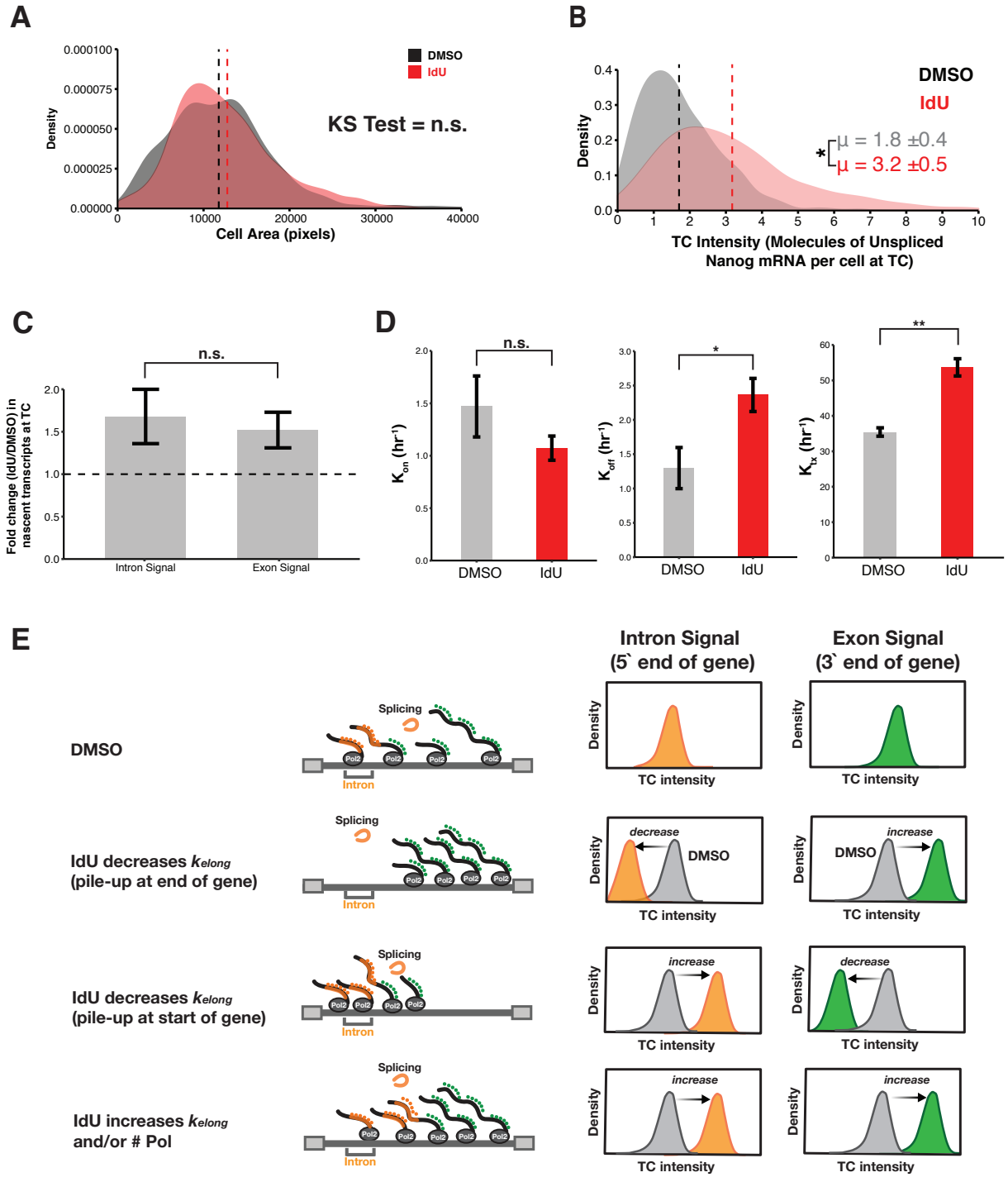


Figure S9: Shortened burst duration and increased transcription rate causes enhanced cell-to-cell variability in Nanog mRNA counts.
(A) Distribution of cell sizes for analyzed cells in DMSO and IdU conditions. Cell size was calculated as number of pixels within segmented cell boundary. Dashed lines represent means of each distribution. Data represent pooling of cells from all four biological replicates for each condition. KS test shows no significant difference between cell size distributions. **(B)** Distributions of unspliced/nascent Nanog mRNA per cell as quantified by probes for first intron of Nanog. Data represent pooling of cells from all four biological replicates for each condition. With IdU, active transcriptional centers (TCs) have more unspliced/nascent mRNA, *p = 0.0029 by a two-tailed, unpaired Student's t test. **(C)** Fold change in number of nascent Nanog transcripts detected at TCs by intron (5‘ end) and exon (3‘ end) probes (number of nascent transcripts in IdU condition / number of nascent transcripts in DMSO condition). IdU treatment causes an equivalent increase in nascent transcripts at 5‘ and 3‘ ends of gene body, indicating that . P values were calculated using a two-tailed, unpaired Student's t test. **(D)** Inference of parameters for 2-state model of transcription. P values were calculated using a two-tailed, unpaired Student's t test. *p = 0.0017, **p = 0.0001. **(E)** Schematic describing expected changes in TC intensity (as quantified by intron and exon probe set) based on changes in polymerase elongation through gene body. As described in the second row, slowed elongation with a pileup of polymerases at the end of the gene would cause an increase in exon signal accompanied by a decrease in intron signal. As described in the third row, slowed elongation with a pileup of polymerases at the start of the gene would cause a decrease in exon signal and increase in intron signal. As described in the fourth row, an increased transcription rate would result in an increase of both the intron and exon signal. Experimental data (Fig. 2D and fig. S9, B& C) is most consistent with the scenario described in row four.

Supplemental Figure 10

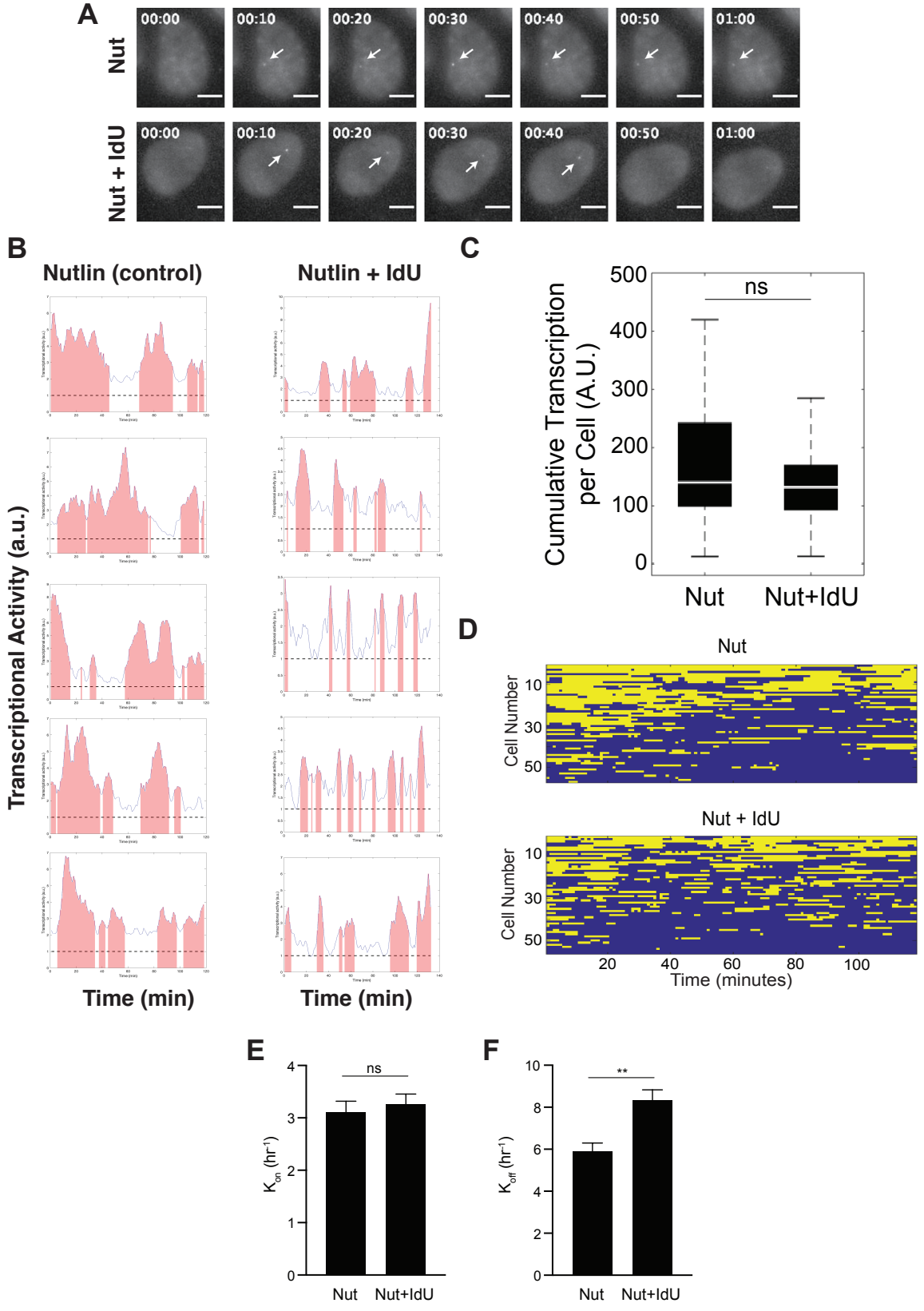


Figure S10: Live-cell imaging of transcription demonstrates shortened burst duration with maintenance of transcriptional output.
(A) Images of U2OS cells containing p21-MS2 reporter (maximum intensity projections of z-dimension stacks). Scale bar, 5 μm. Images were taken 10 minutes apart. White arrow points to transcription site (TS). Top panel: U2OS p21-MS2 reporter cells treated with Nutlin (activator of p21 transcription). Bottom panel: U2OS p21-MS2 reporter cells treated with Nutlin and 20μM IdU. **(B)** Quantification of the fluorescence intensity at the TS for 5 representative cells in the control and IdU conditions. Shadings represent bursts of transcription where fluorescence signal (a.u.) exceeds background signal by 2.5-fold. **(C)** Cumulative transcriptional output per cell in control and IdU conditions. Data represent analysis of 56 cells for each condition over 118 minutes of imaging. Boxplots show median ±interquartile range of cumulative transcriptional output, which is calculated based on the normalized transcriptional activity from fluorescence tracings exemplified in panel B. **(D)** Transcriptional activity (yellow bars) of each tracked cell over time in control and IdU conditions. **(E)** Rate of activation for p21 promoter (K_{ON}). Data represent mean (± SEM) of 56 cells for each condition. **(F)** Rate of inactivation for p21 promoter (K_{OFF}). Data represent mean (± SEM) of 56 cells for each condition. **p < 0.001 by a two-sample t-test.

Supplemental Figure 11

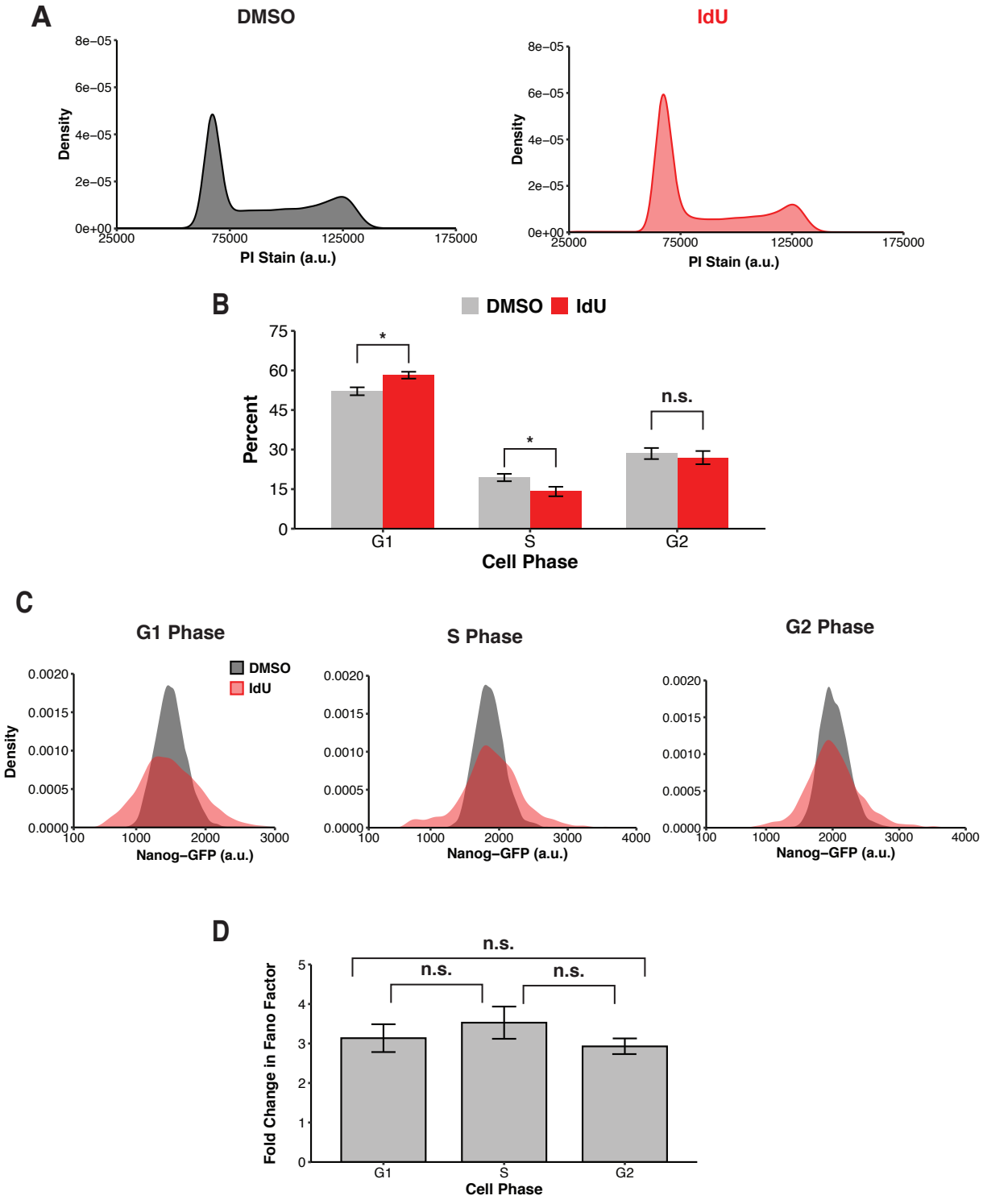


Figure S11: Noise-enhancement of Nanog protein expression is independent of cell-cycle state.
(A) Representative flow cytometry distributions of propidium iodide staining for Nanog-GFP mESCs treated with either DMSO or 10μM IdU for 24 hours. No signs of aneuploidy are visible, indicating transcriptional variability is not due to cell-to-cell variability in gene copy numbers. **(B)** Percent of cells in each phase of the cell cycle for DMSO and IdU treatments based on propidium iodide staining. IdU treatment slightly slows entry into S phase. Data represent mean and SD of three biological replicates. P values were calculated using a two-tailed, unpaired Student's t test. *$p < 0.01$ **(C)** Representative flow cytometry distributions of Nanog-GFP for mESCs within the G1, S and G2 phases of the cell cycle. mESCs were treated with 10μM IdU or equivalent volume DMSO for 24 hours followed by propidium iodide staining. **(D)** IdU-induced noise-enhancement of Nanog-GFP protein levels is unchanged across all three phases of the cell cycle. Nanog-GFP Fano factor with IdU treatment was normalized to DMSO control for calculation of fold change. Data represent mean and SD of three biological replicates. P values were calculated using a two-tailed, unpaired Student's t test.

Supplemental Figure 12

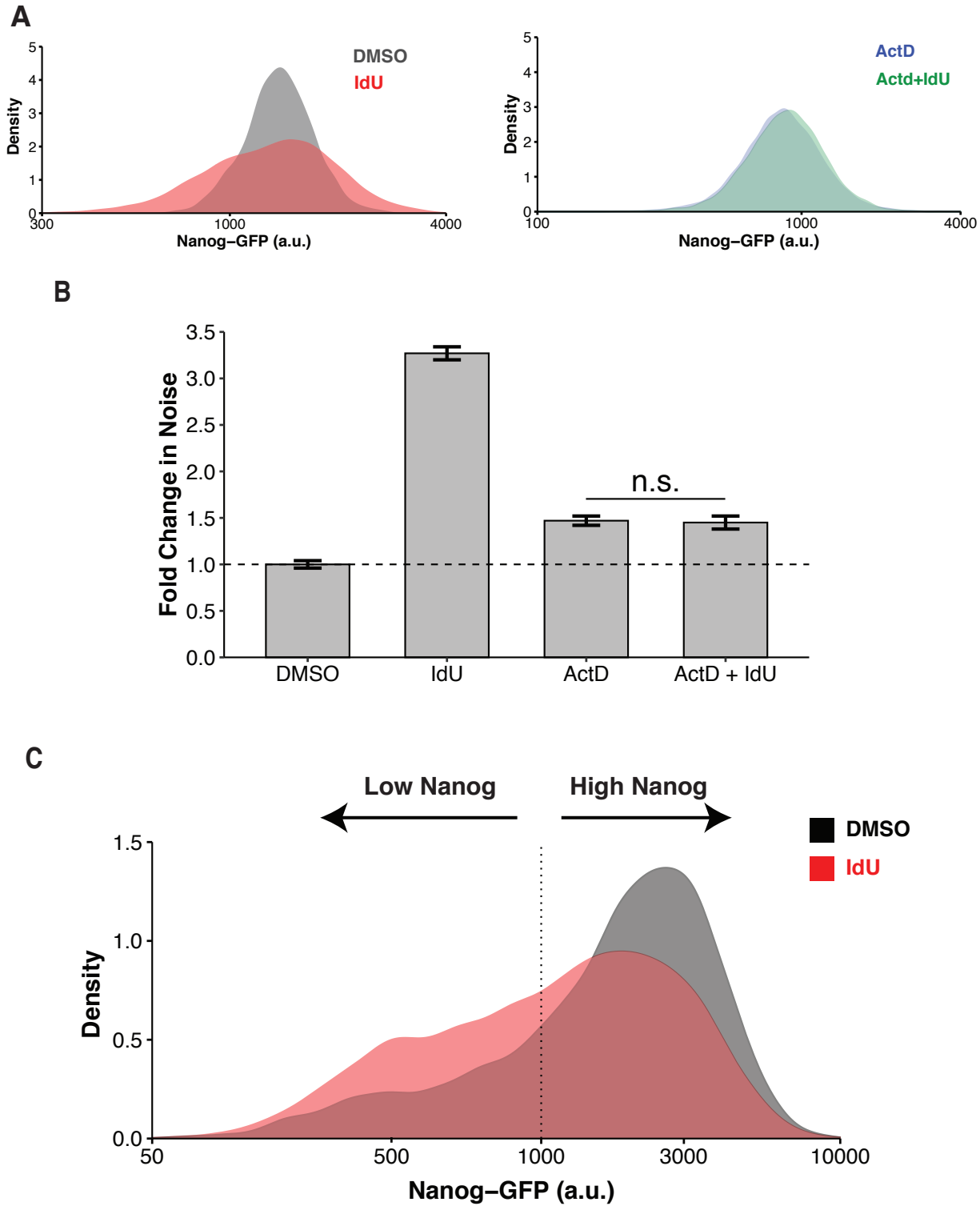


Figure S12: Increased expression variability of Nanog protein originates from transcriptional sources of noise which drives a greater number of mESCs into the low-Nanog state while cultured in serum/LIF.
(A) Representative flow cytometry distributions of Nanog-GFP expression for mESCs cultured in 2i/LIF and treated with either 10μM IdU or equivalent volume DMSO for 24 hours (left histogram) or 6ug/ml Actinomycin D in the presence or absence of 10μM IdU for 24 hours (right histogram). **(B)** Inhibition of transcription ablates IdU-induced noise-enhancement of Nanog-GFP protein levels. Nanog-GFP Fano factor was normalized to DMSO control for calculation of fold change. Data represent mean and SD of three biological replicates. P values were calculated using a two-tailed, unpaired Student's t test. **(C)** Flow cytometry distribution of Nanog-GFP expression for mESCs cultured in serum/LIF and treated with 10μM IdU or equivalent volume DMSO for 24 hours. Data is pooled from three biological replicates.

Supplemental Figure 13

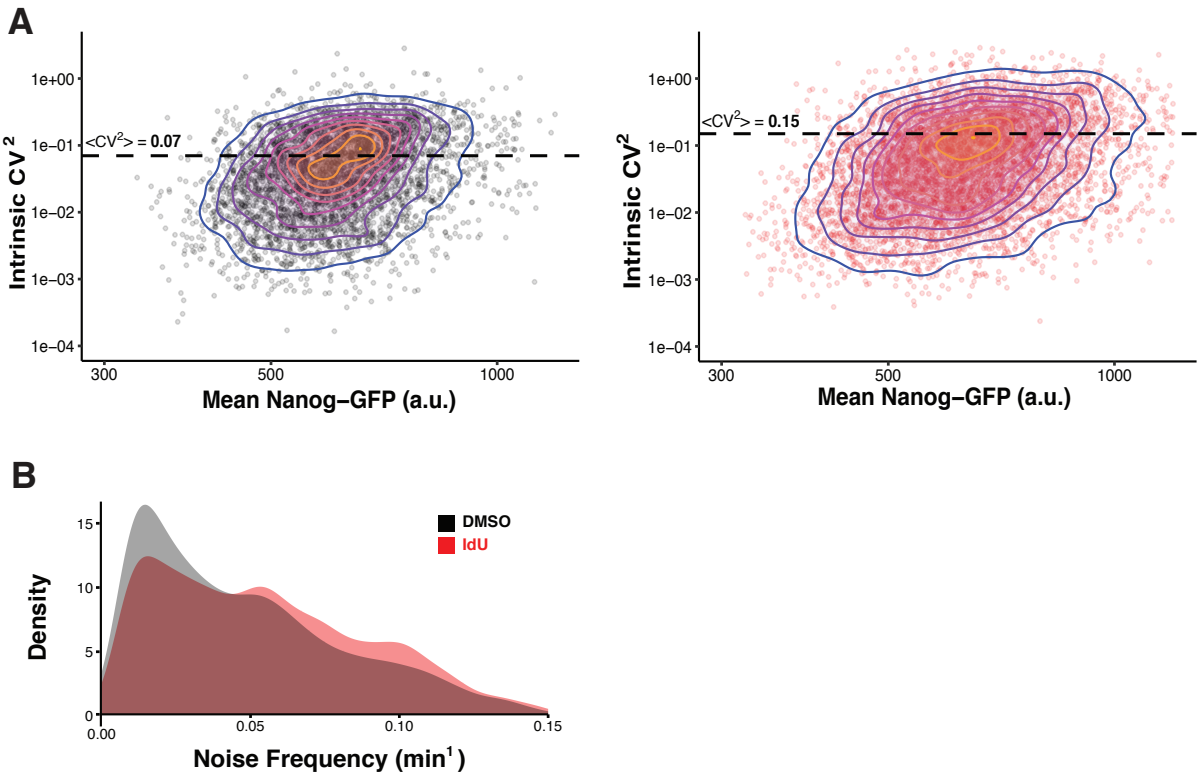


Figure S13: Time-lapse imaging demonstrates that altered kinetics of promoter toggling cause individual cells to experience larger fluctuations in Nanog protein expression.
(A) Each point represents a single-cell fluorescence trajectory (DMSO on left, n = 1513; IdU on right, n = 1414). Single-cell fluorescence trajectories were detrended by subtracting time-dependent population average for Nanog-GFP fluorescence. The mean Nanog-GFP fluorescence for each raw trajectory is then plotted versus the CV^2 of the detrended version of the trajectory to isolate intrinsic noise. The dashed lines represent the average intrinsic CV^2 of all trajectories for each treatment group. Time-lapse imaging shows that for individual cells the magnitude of Nanog protein fluctuations increases with IdU treatment. **(B)** Distributions of noise frequencies from autocorrelation functions of each detrended trajectory. Noise frequency is calculated as the inverse of the autocorrelation time ($\tau_{1/2}$). Shorter but more productive transcriptional bursts with IdU treatment pushes the frequency content of Nanog-GFP fluctuations to higher spectra.

Supplemental Figure 14

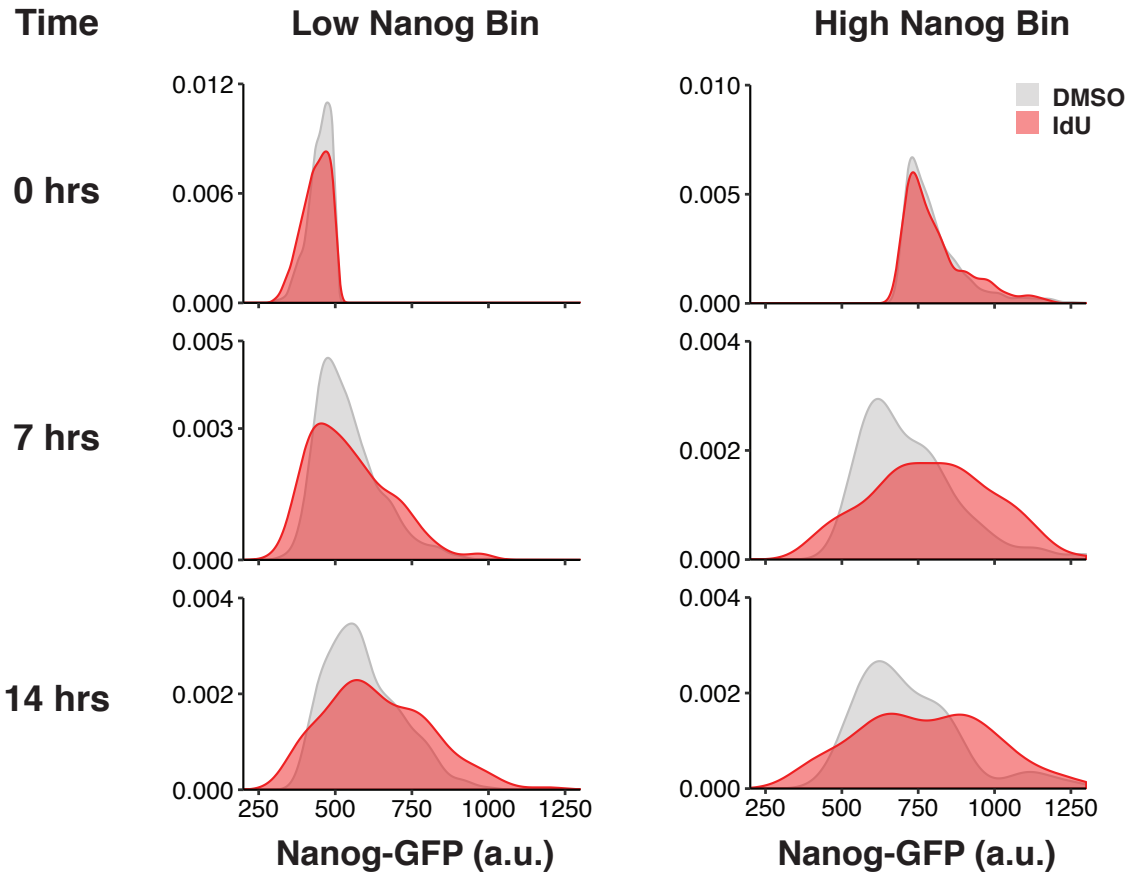
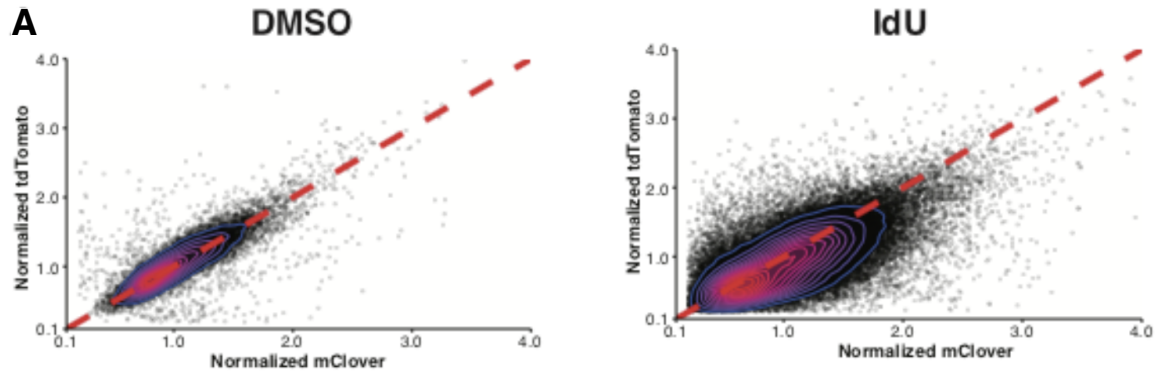


Figure S14: Amplification of expression fluctuations occurs independently of starting Nanog level.
Single-cell trajectories whose starting fluorescence value was below 500 a.u. or above 700 a.u. were binned into low (left column) and high (right column) groups respectively. Only trajectories whose starting point coincided with addition of DMSO or IdU at time zero were used. Distributions of trajectory fluorescence values at zero, seven, and 14 hours into treatment conditions are shown. By 14 hours into IdU treatment, there is visible interconversion of cells between the low and high Nanog states, indicating that memory of initial Nanog expression level is erased. This precludes the possibility that noise enhancement is due to promoter mutations that create sub-populations of cells with stable expression of Nanog at low and high levels.

Supplemental Figure 15



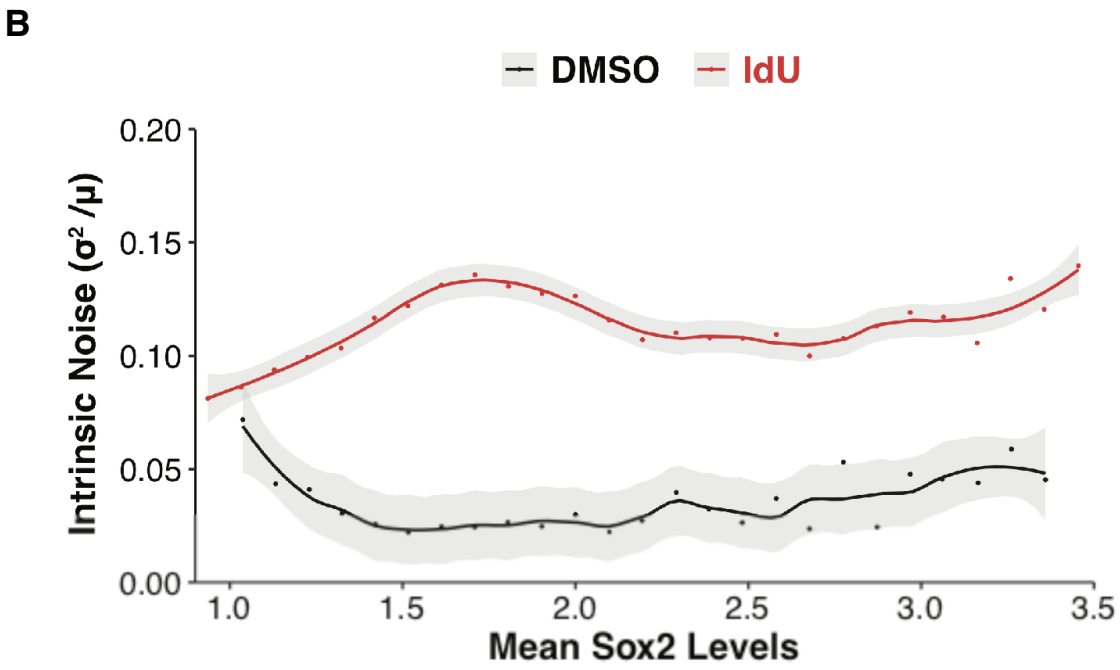


Figure S15: IdU treatment increases intrinsic noise of Sox2 expression.
(A) Flow cytometry dot-plot of mESCs with Sox2 dual color tags (P2A-mClover fused to one allele and P2A-tdTomato fused to other allele). Dashed red line has slope of one. mClover and tdTomato fluorescence values were normalized to population average. Data shown is pooled from three biological replicates. **(B)** Cells were binned according to total Sox2 expression from both alleles. Each point represents the intrinsic noise (Fano factor) of Sox2 expression for cells within a particular bin. Grey shadings represent 95% confidence intervals as determined by bootstrapping. Smooth lines are produced from loess regression. IdU increases Sox2 intrinsic noise across all expression levels.

Supplemental Figure 16

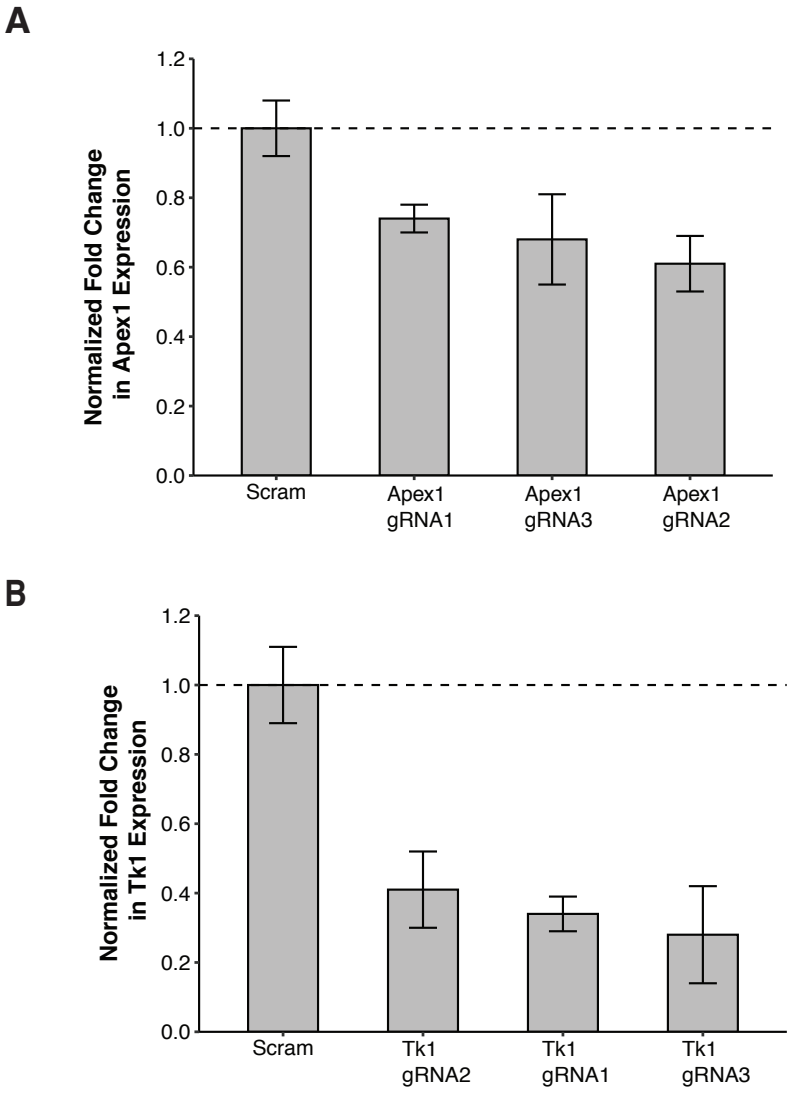
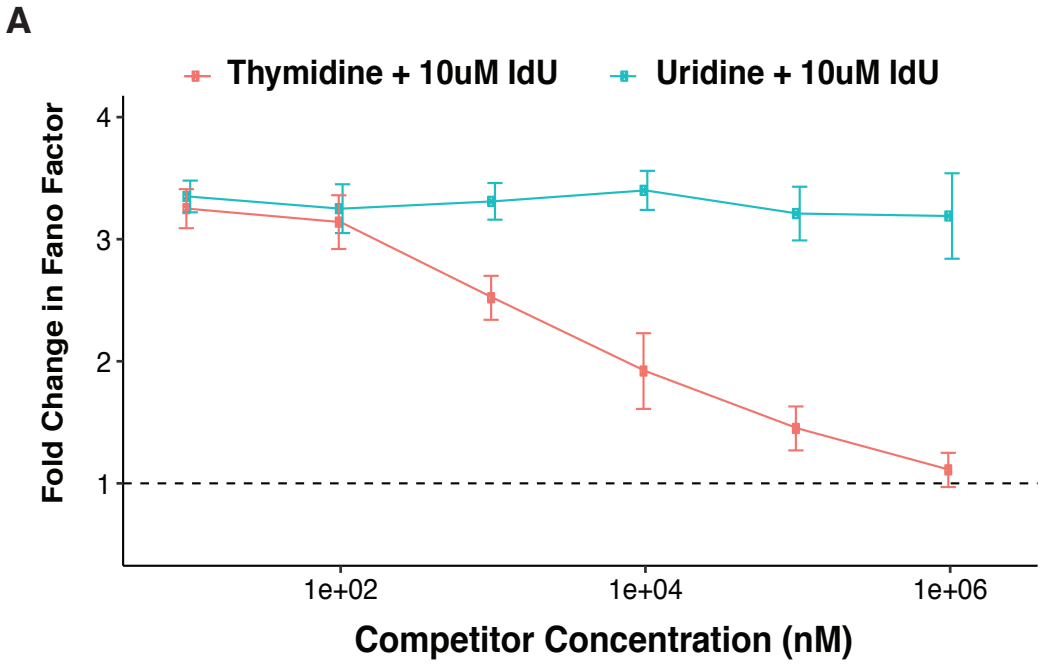


Figure S16: Validation of CRISPRi knockdown of Apex1 and Tk1 via qPCR measurements. $\Delta\Delta C_t$ method was used with the empty-vector cell population as the control. Levels of Apex1 and Tk1 repression are relative to the non-targeting (scrambled) population. Data represent mean and SD of two biological replicates.

Supplemental Figure 17



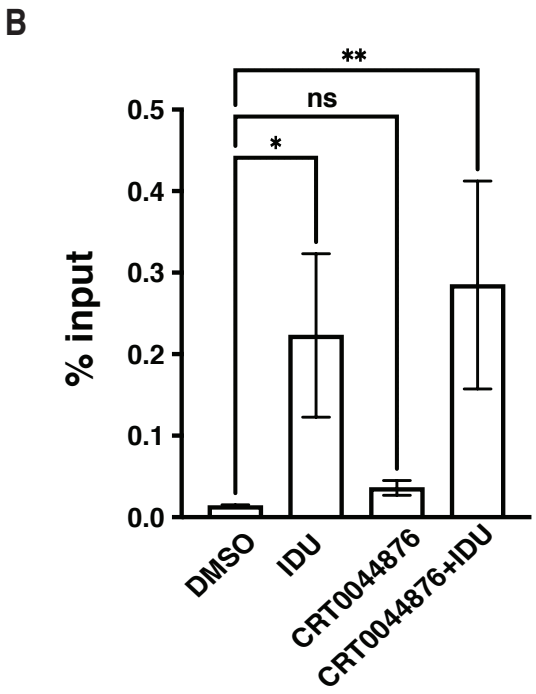
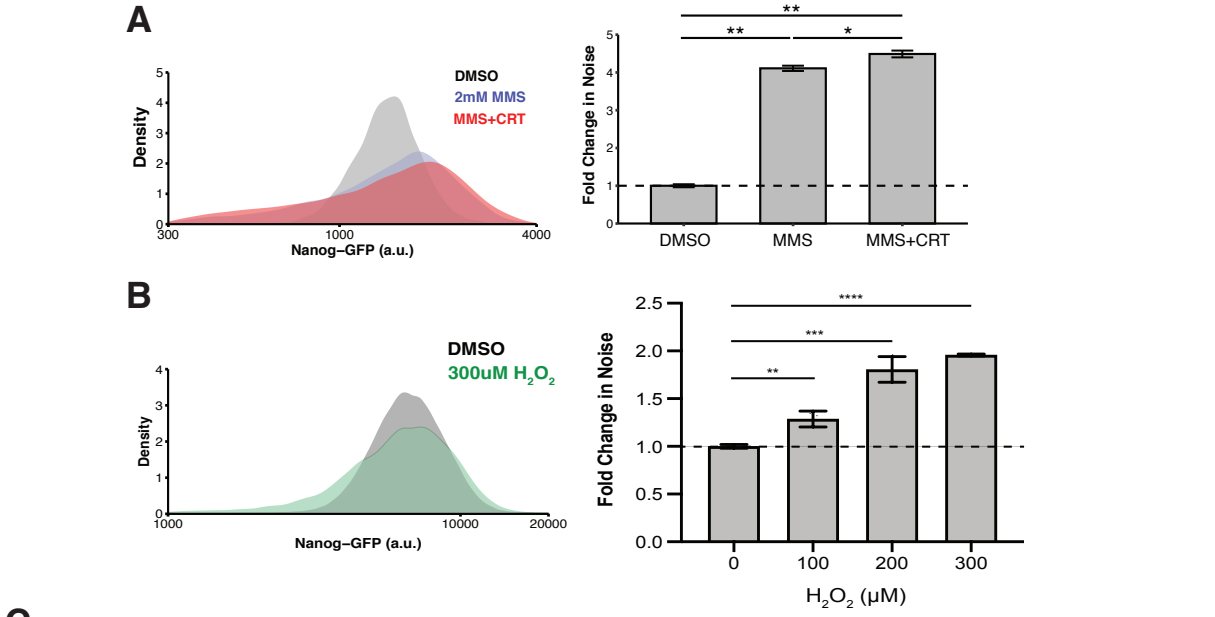
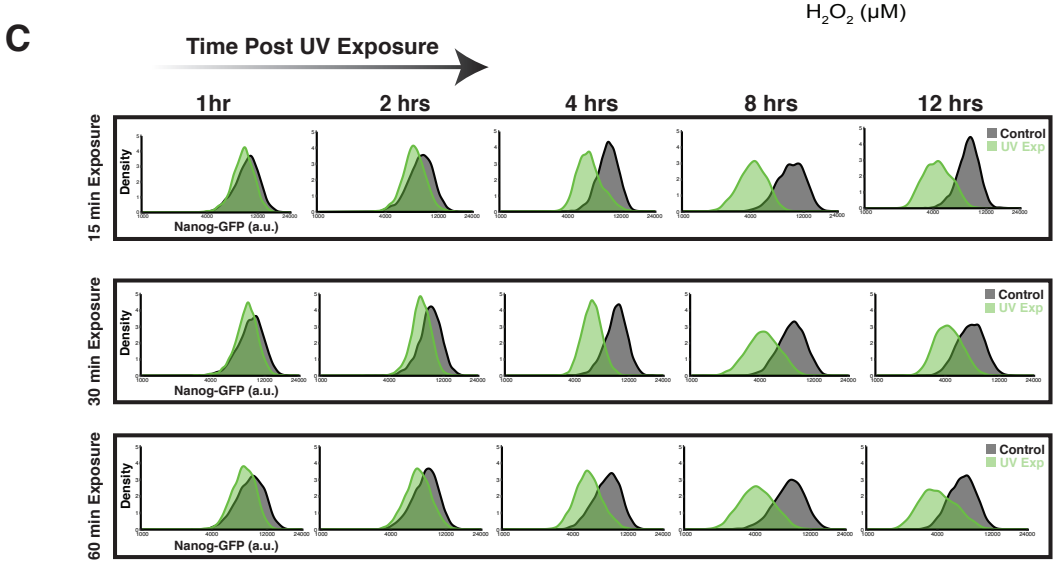


Figure S17: Thymidine competition ablates Nanog noise-enhancement from IdU and ChIP-qPCR demonstrates increased recruitment of Apex1 to Nanog promoter with IdU treatment. **(A)** The Fano factor of Nanog for each concentration combination is normalized to DMSO control. For all treatment combinations, IdU concentration is held constant at 10μM. Concentration of thymidine (red) and uridine (blue) is reported on the x-axis. Combination of 100μM thymidine and 10μM IdU returns Nanog Fano factor to baseline level (DMSO control). Uridine, which is not a substrate of Tk1, fails to ablate IdU-induced noise-enhancement. Data points represent mean and SD of three biological replicates. **(B)** ChIP-qPCR analysis of Apex1 binding to Nanog promoter. IdU treatment enhances recruitment of Apex1 to the Nanog promoter; *p = 0.0334, **p = 0.0089 by Dunnett's multiple comparisons test.

Supplemental Figure 18





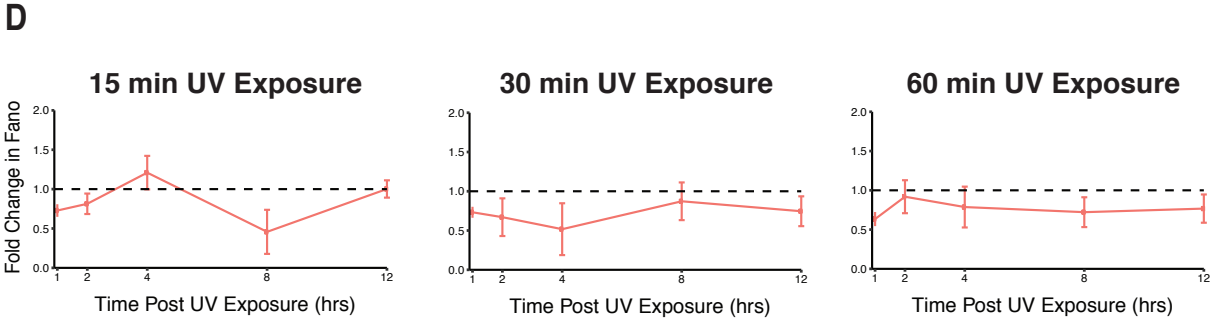


Figure S18: Activation of BER with MMS and H_2O_2 increases gene expression variability while activation of NER with UV exposure fails to increase noise.
(A) (Left) Representative flow cytometry distributions of Nanog-GFP expression for mESCs treated with 2mM MMS (blue), 2mM MMS with 100uM CRT0044876 (red) and DMSO (grey). Compounds were washed off after 1 hour of treatment. Flow-cytometry analysis performed 24 hours after wash. (Right) Nanog-GFP Fano factor was normalized to DMSO control for calculation of fold change. Data represent mean ($\pm$SD) of three biological replicates. **p $<$ 0.00086, *p = 0.014 by a two-tailed, unpaired Student's t test. **(B)** (Left) Representative flow cytometry distributions of Nanog-GFP expression for mESCs treated with 300uM H_2O_2 (green) and DMSO (grey). Compounds were washed off after 1 hour of treatment. Flow-cytometry analysis performed 24 hours after wash. (Right) Nanog-GFP Fano factor was normalized to DMSO control for calculation of fold change. Data represent mean and SD of three biological replicates. **p = 0.0045, ***p = 0.0005, ****p $<$ 0.0001 by a two-tailed, unpaired Student's t test. **(C)** Representative flow cytometry distributions of Nanog-GFP expression from UV-exposed (green) and control (grey) cell populations. Cells were analyzed one, two, four, eight, and 12 hours post exposure. **(D)** For each exposure group (15, 30, and 60 minutes), the fold change in Fano factor is calculated as the Fano factor for Nanog in the UV-exposed population normalized to the Fano factor of its respective control population. Data points represent mean and SD of two biological replicates. Across all time points (except 4 hour point in 15 minute exposure group) UV stress reduces the Fano factor of Nanog.

Supplemental Figure 19

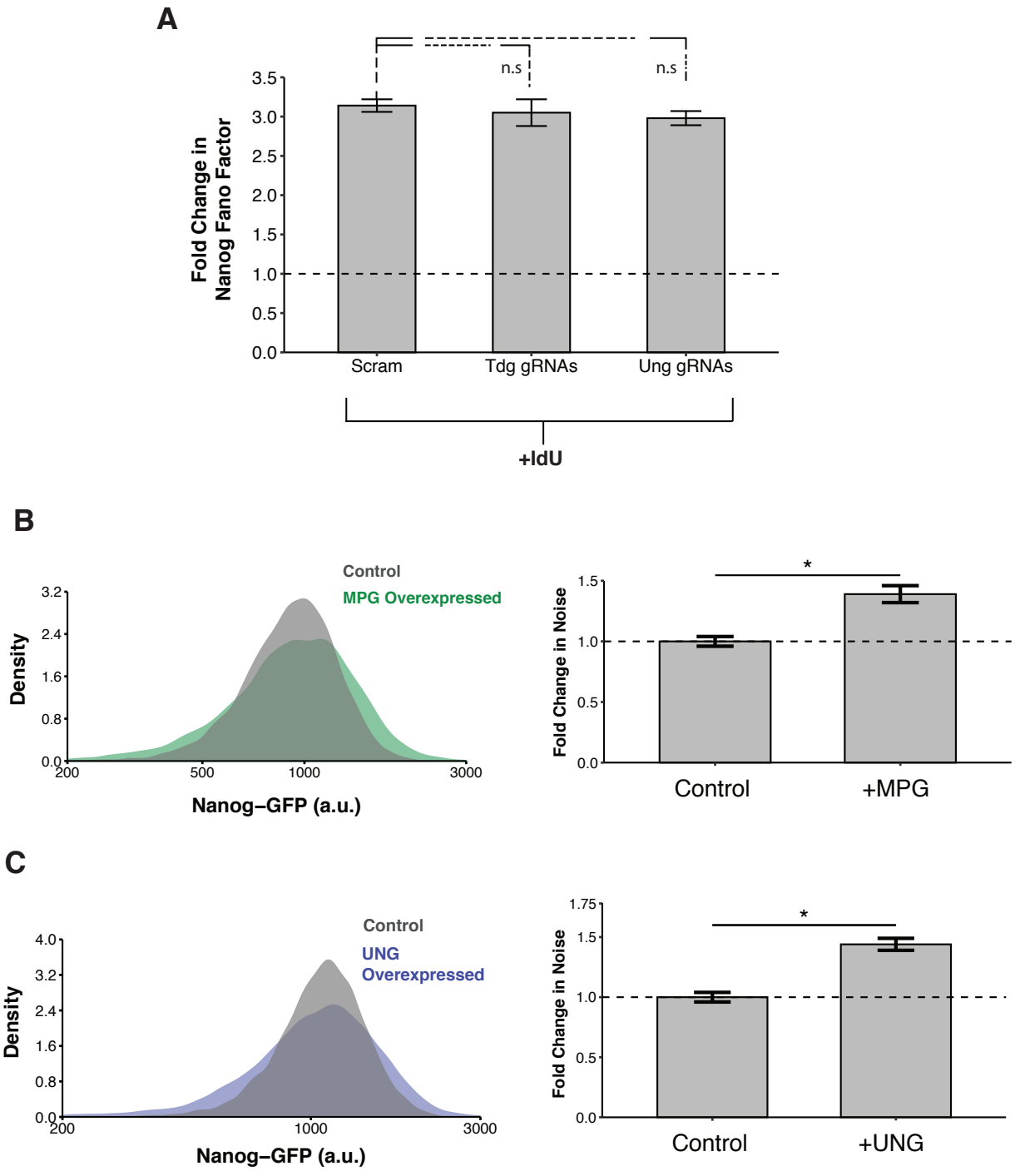


Figure S19: Activation of BER through overexpression of Mpg or Ung increases gene expression noise.
(A) Knockdown of Thymine-DNA glycosylase (Tdg) or Uracil-DNA glycosylase (Ung) failed to ablate IdU-mediated noise enhancement of Nanog-GFP expression. Fold change in Fano factor is normalized to Fano factor of Nanog-GFP expression in CRISPRi mESCs expressing scrambled gRNA. Three gRNAs for Tdg and three gRNAs for Ung were each tested in triplicate. Data represent mean and SD of all nine samples for each knockdown. P values were calculated using a two-tailed, unpaired Student's t test. **(B)** (Left) Representative flow cytometry distributions of Nanog-GFP expression from mESCs overexpressing methylpurine glycosylase (Mpg, green) and control (grey) cell populations. (Right) Nanog-GFP Fano factor was normalized to DMSO control for calculation of fold change. Data represent mean and SD of three biological replicates. *p = 0.0246, by a two-tailed, unpaired Student's t test. **(C)** (Left) Representative flow cytometry distributions of Nanog-GFP expression from mESCs overexpressing Uracil-DNA glycosylase (Ung, purple) and control (grey) cell populations. (Right) Nanog-GFP Fano factor was normalized to DMSO control for calculation of fold change. Data represent mean and SD of three biological replicates. *p = 0.0138, by a two-tailed, unpaired Student's t test.

Supplemental Figure 20

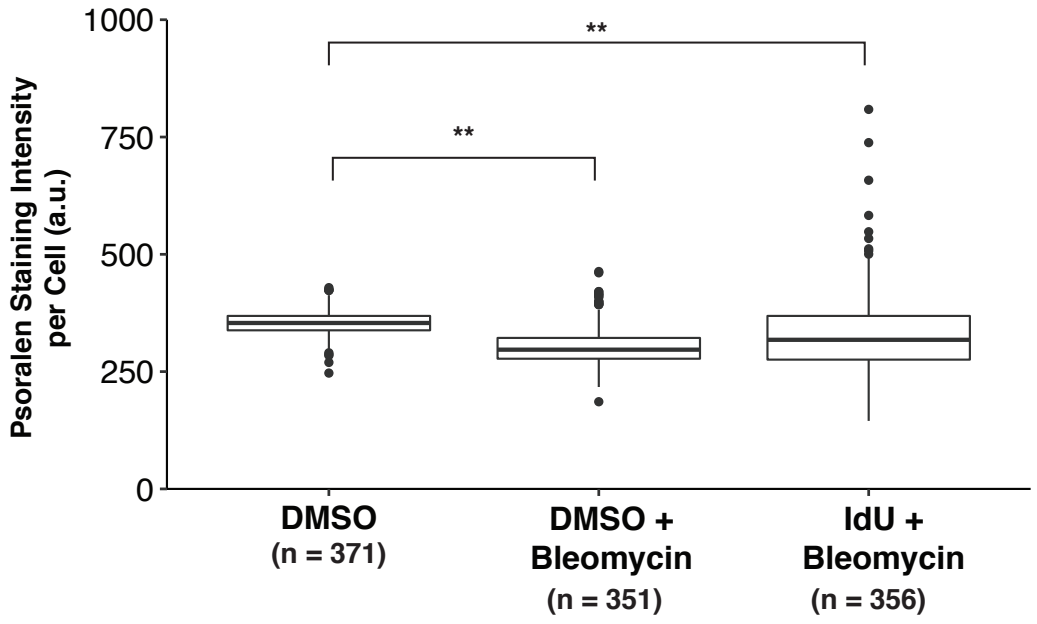


Figure S20: Bleomycin treatment reduces bTMP intercalation into DNA, validating assay sensitivity for negative supercoiling levels.
Boxplots show median ±interquartile range of single-cell bTMP staining intensities. Treatment of mESCs with 100μM bleomycin was performed for 1 hour just prior to bTMP incubation. Bleomycin reduces the mean bTMP staining intensity for cells treated with DMSO or 10μM IdU as compared to DMSO control with no bleomycin treatment (**$p < 0.0001$). The reduction in bTMP staining when IdU is coupled with bleomycin indicates that IdU alone in uncoiled DNA does not increase bTMP intercalation. Data shown are pooled from two biological replicates. P values were calculated using Kruskal-Wallis test followed by Tukey's multiple comparison test.

Supplemental Figure 21

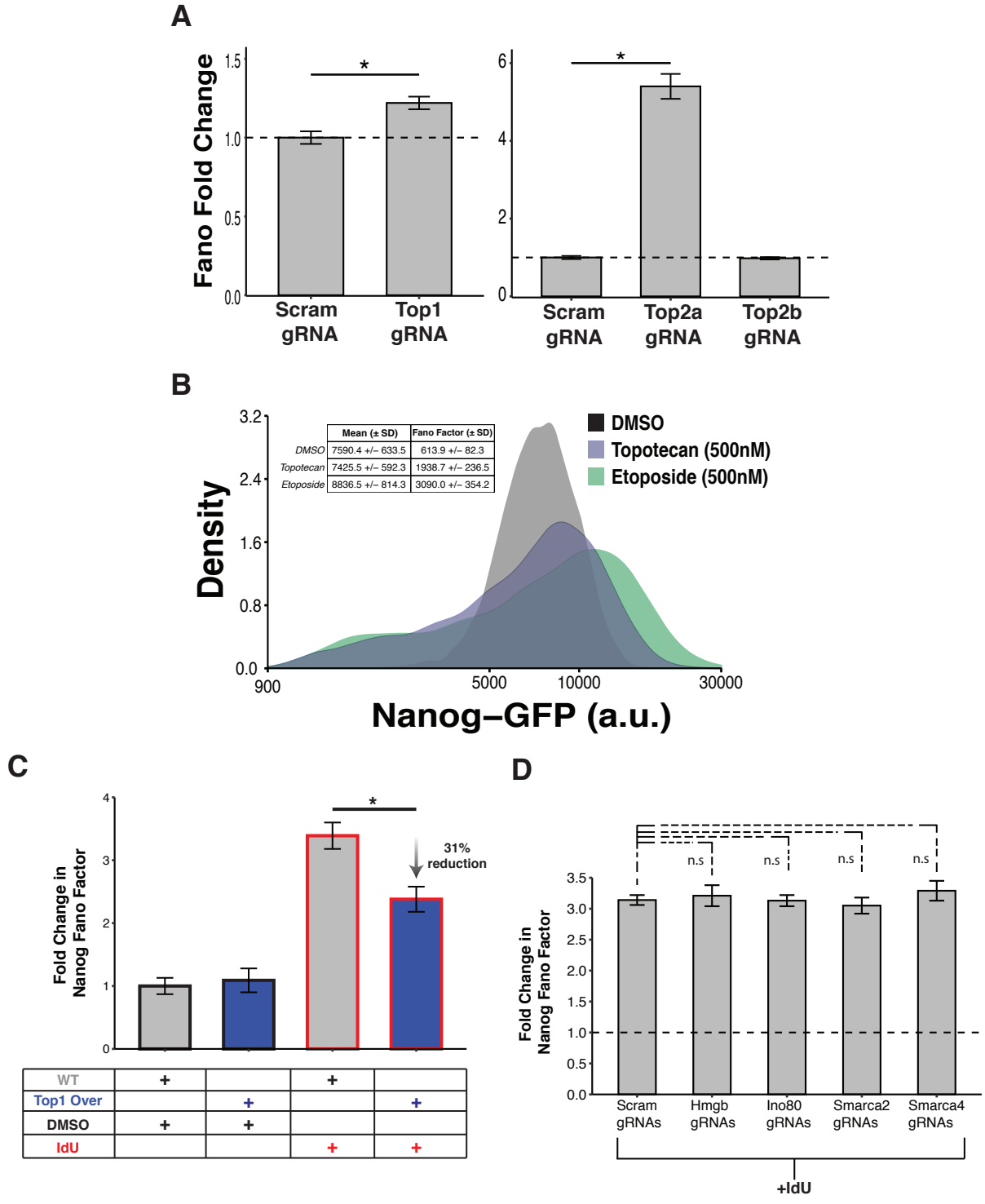


Figure S21: Loss of Topoisomerase activity increases Nanog expression variability while overexpression of Topoisomerase 1 partially ablates IdU-mediated noise-enhancement of Nanog expression.
(A) CRISPRi knockdown of Topoisomerases involved in relaxation of DNA supercoiling. Nanog Fano factor was normalized to scrambled gRNA population. Data represent mean (± SD) of three biological replicates. Knockdown of Top1 (*p = 0.002) and Top2a (*p = 0.003) increases Nanog expression variability. P values were calculated by two-tailed, unpaired Student's t test. **(B)** Representative flow cytometry distributions of Nanog-GFP expression in mESCs treated with DMSO, 500nM topotecan or 500nM etoposide for 24 hours in 2i/Lif media. Extrinsic noise filtering via cell-size gating was performed prior to calculation of Nanog Fano factor. Table inset shows mean and Fano factor (±SD) of Nanog expression averaged over three biological replicates of each treatment. **(C)** Overexpression of Topoisomerase 1 partially ablates IdU-mediated noise-enhancement of Nanog expression. Wildtype (WT) and topoisomerase 1 (Top1) overexpressing Nanog-GFP mESCs were treated with DMSO or 10μM IdU for 24h in 2i/LIF media. Nanog Fano factor was normalized to wildtype population of Nanog-GFP mESCs treated with DMSO. Data represent mean (± SD) of three biological replicates. Overexpression of Top1 (*p = 0.0036) reduces IdU-mediated enhancement of Nanog variability by 31%. P values were calculated by two-tailed, unpaired Student's t test. **(D)** Knockdown of histone remodelling factors known to interact with BER pathway failed to ablate IdU-mediated noise enhancement of Nanog-GFP expression. Fold change in Fano factor is normalized to Fano factor of Nanog-GFP expression in CRISPRi mESCs expressing scrambled gRNA. Three gRNAs for each histone remodelling factor were tested in triplicate. Data represent mean and SD of all nine samples for each knockdown. P values were calculated using a two-tailed, unpaired Student's t test.

Supplementary Figure 22

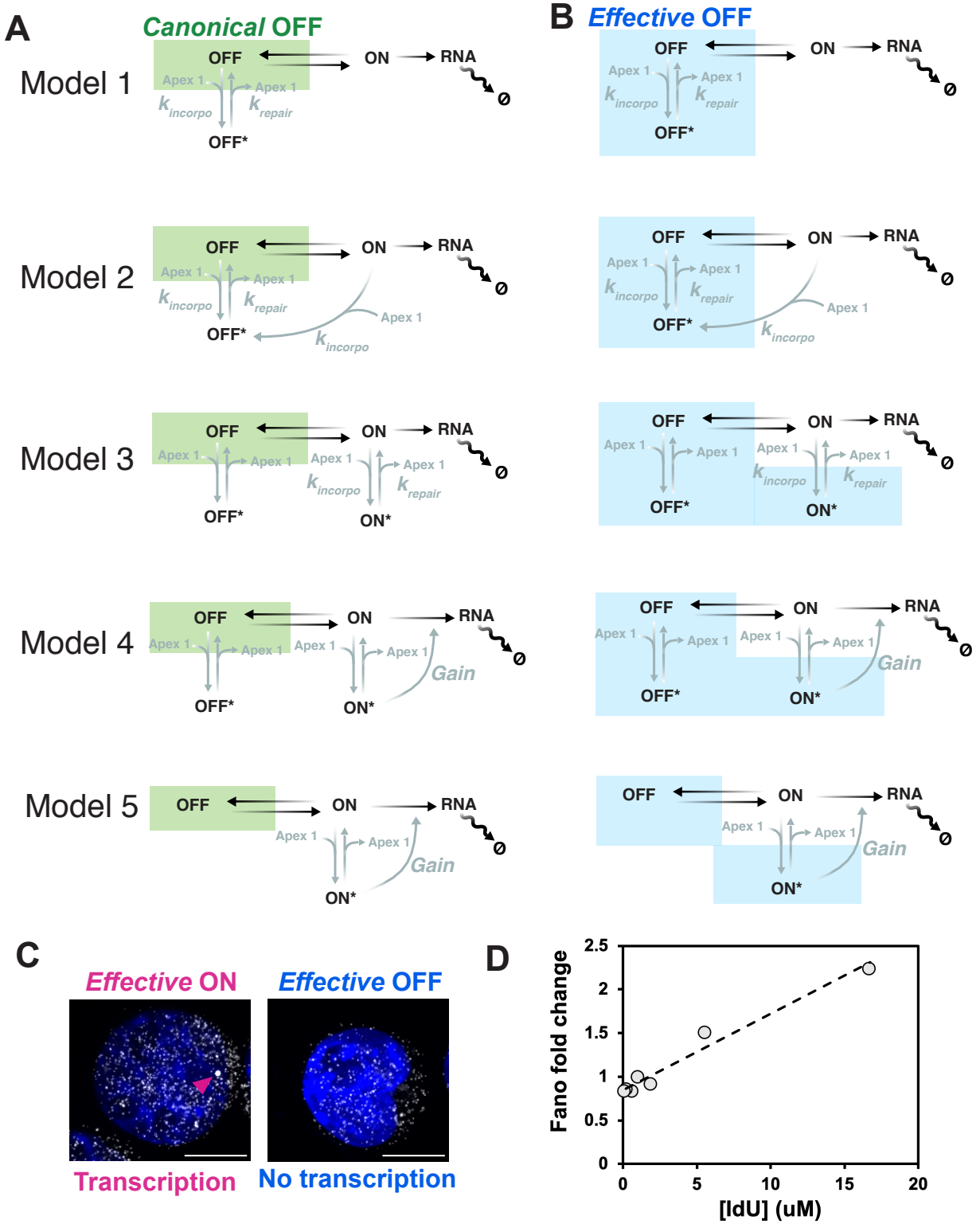
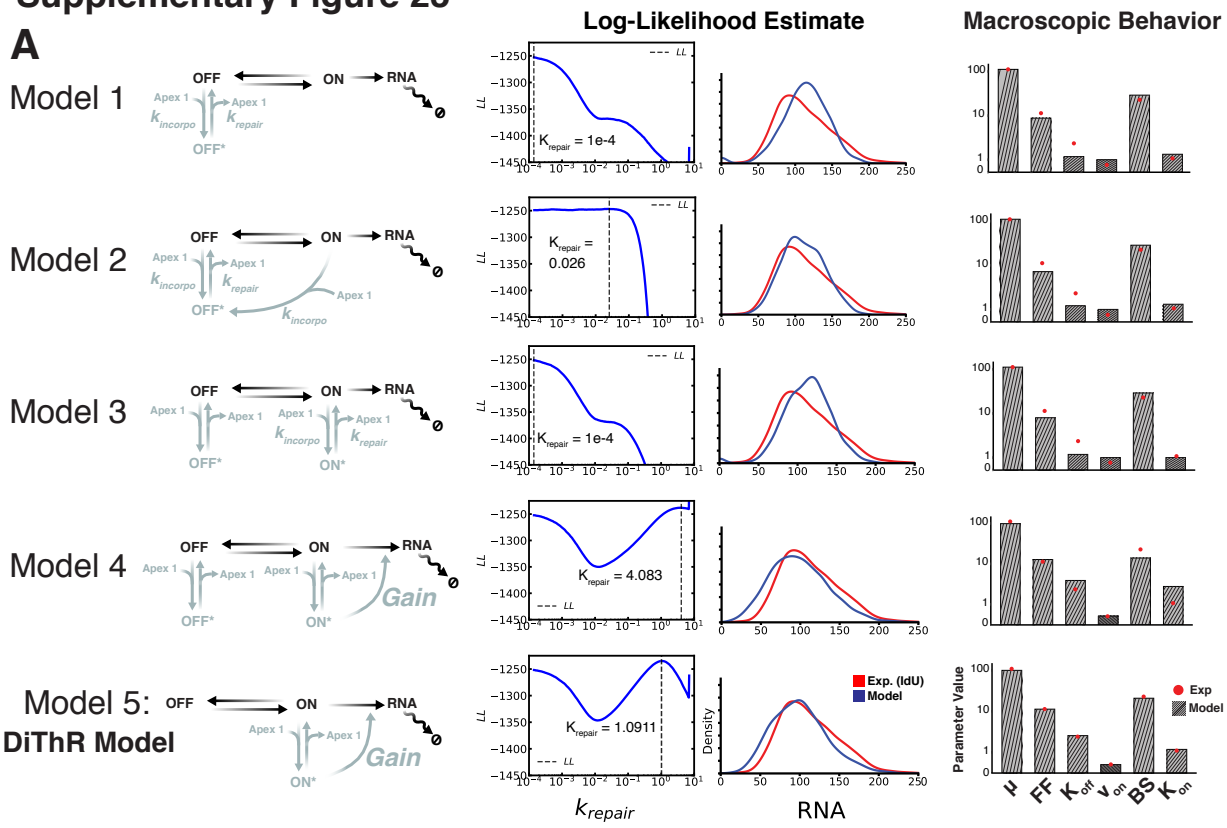


Figure S22: Schematic of tested models with canonical and effective OFF states highlighted. **(A)** Schematic of simulated models incorporating Apex1 into standard two-state model of transcription. Canonical OFF state of each model is highlighted in green. **(B)** Identical schematic of simulated models as in panel A. Effective OFF state (all states incapable of producing mRNA) of each model is highlighted in blue. **(C)** Representative micrographs from smRNA FISH experiment with DAPI staining (maximum intensity projection). Nanog-eGFP transcripts are labelled with probe-set for eGFP. Scale bar is 5μm. (Left) Purple arrow indicates active transcriptional center as verified by intron probe set, which represents the effective ON state. (Right) Lack of active transcriptional center represents effective OFF state. **(D)** Fold change of Nanog-GFP Fano factor as a function of IdU concentration (adapted from data in Figure 4C). Incorporation of IdU (a thymidine nucleotide analogue) is a result of DNA polymerase insertion during replication and so incorporation of IdU is likely in pseudo-first order regime of Michaelis Menten kinetics. This is consistent with experimental evidence demonstrating that IdU-mediated noise-enhancement of Nanog expression is dose-dependent.

Supplementary Figure 23



B

Model	max (LL)	Δ_iAIC	w_i
Control	-1247	24	0.000
1	-1253	36	0.000
2	-1247	24	0.000
3	-1252	34	0.000
4	-1238	6	0.05
5	-1235	0	0.95

C

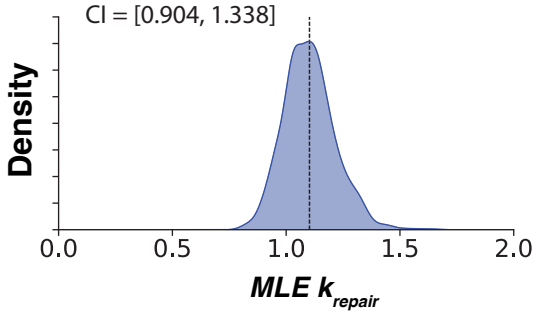
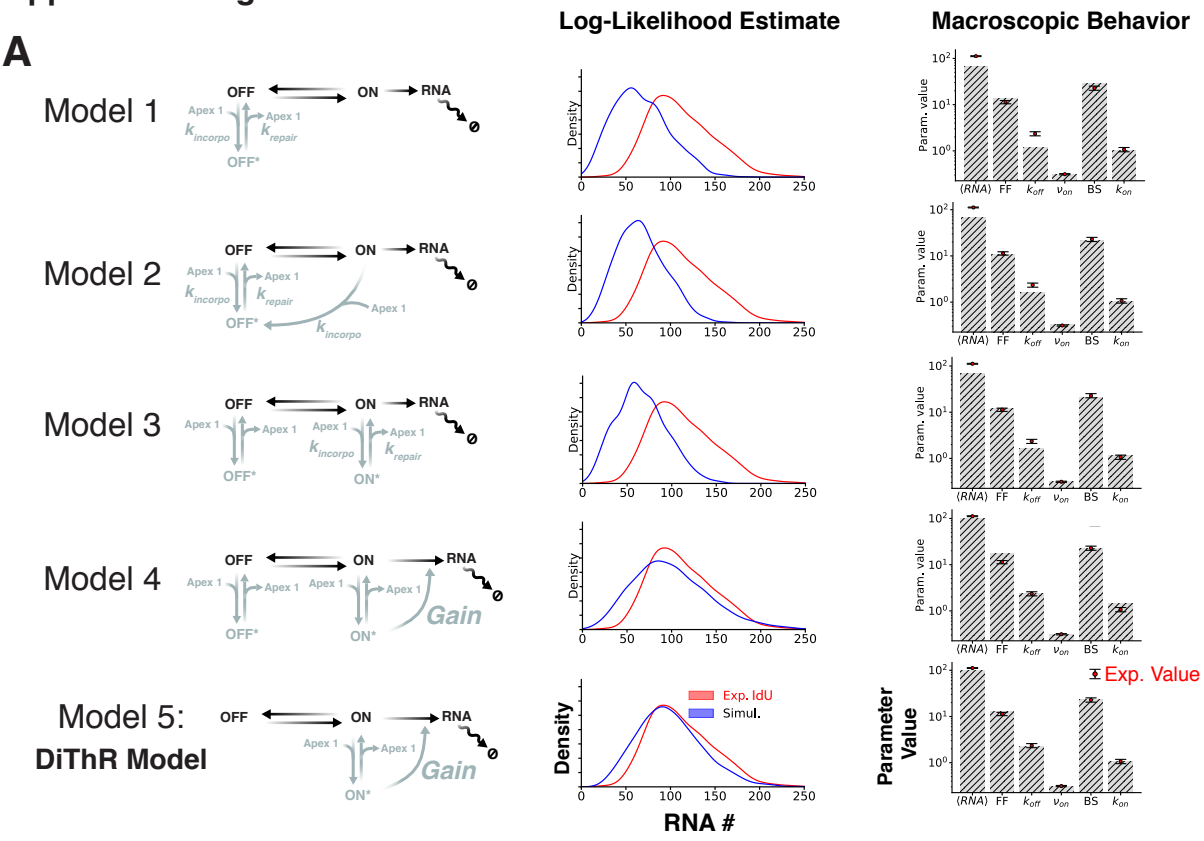


Figure S23: MLE-based approach for model selection reveals transcription-coupled base-excision mechanism best recapitulates experimental data.
(A) (First Column) Schematic of simulated models incorporating Apex1 into standard 2-state model of transcription. (Second Column) For each model, 500 logarithmically-spaced values of $k_{repair} \in [10^{-4}, 10]$ were simulated. For each simulated value of k_{repair}, log-likelihood is calculated as described in supplementary text 5.1.2 and plotted. Dashed vertical line in each plot denotes value of k_{repair} that maximizes log-likelihood estimate. (Third Column) Comparison of experimental Nanog mRNA distribution (red) to simulated distributions of Nanog mRNA (blue) for each model using value of k_{repair} that maximizes log-likelihood. (Fourth Column) The effective behavior (mean Nanog mRNA [μ], Fano factor [FF], K_{OFF}, fraction of time active [ν_{on}], burst size [BS], K_{ON}) of each model using value of k_{repair} that maximizes log-likelihood estimate are compared to experimental data (supplementary text 5.1). Bars represent simulation values of Nanog gene expression system while red points with vertical line represent experimental data on Nanog expression from smRNA-FISH of mESCs treated with 10μM IdU. **(B)** For each tested model, the maximum log-likelihood value is listed along with the associated $\Delta_i AIC$. Model 5 (DiThR model) best describes experimental data based on these metrics. **(C)** Distribution and confidence interval (CI) of inferred k_{repair} values (based on MLE) for DiThR model using bootstrapping method in which the empirical distribution of Nanog mRNA counts from smRNA-FISH data was re-sampled 1000 times with replacement (supplementary text 5.1.4). Bootstrapping results show a well peaked distribution indicating practical parameter identifiability for k_{repair}.

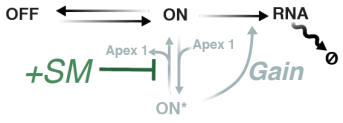
Supplemental Figure 24



B

Model	APE	$\hat{k}_{repair,APE}$	log(L)
1	1.43	3.27	-1650.7
2	0.89	0.43	-1766.7
3	0.92	0.74	-1752.5
4	0.99	1.71	-1315.5
5	0.27	0.89	-1256.3

C



	Mean	FF
- *SM*	100.23	11.58
+SM	142.31	36.20

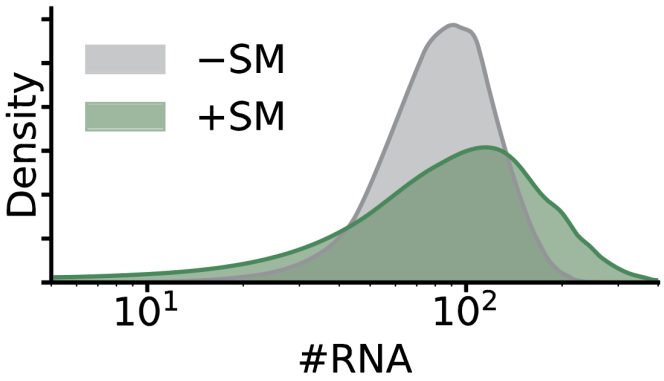


Figure S24: APE-based approach for model selection concurs with MLE-based approach, identifying DiThR model as best match to experimental data.
(A) (First Column) Schematic of simulated models incorporating Apex1 into standard 2-state model of transcription. (Second Column) Comparison of experimental Nanog mRNA distribution (red) to simulated distributions of Nanog mRNA (blue) for each model using value of k_{repair} that minimizes absolute percentage error. (Third Column) Effective behavior of simulation results (using value of k_{repair} that minimizes absolute percentage error) are compared to experimental data (supplementary text 5.2). Bars represent simulation values of Nanog gene expression system while red points with vertical line represent experimental data on Nanog expression from smRNA-FISH of mESCs treated with 10μM IdU. **(B)** Values of k_{repair} that minimize the absolute percentage error for each model are listed. DiThR model yields the smallest APE and the largest log-likelihood. **(C)** Predicted effect of small-molecule (SM) topoisomerase inhibitors with DiThR model (supplementary text section 6). $(-)$SM condition was simulated using the inferred kinetic parameters for Nanog gene expression system with IdU treatment (10μM). Inhibition of topoisomerase activity causes k_{repair} to decrease by a factor n (+SM), which increases the average time spent in the ON* state by the same factor. The longer the residence time in the ON* state, the stronger the enhancement of transcription rate (k_{TX}) through a positive feedforward loop. Densities were computed using final point of 10000 Gillespie simulations. Significant perturbation (n = 2) of k_{repair} only slightly changed average gene expression (×1.41) while leading to a stronger increase in Fano factor (×3.12), which matched experimental data (Fig. S21B).

Supplemental Figure 25

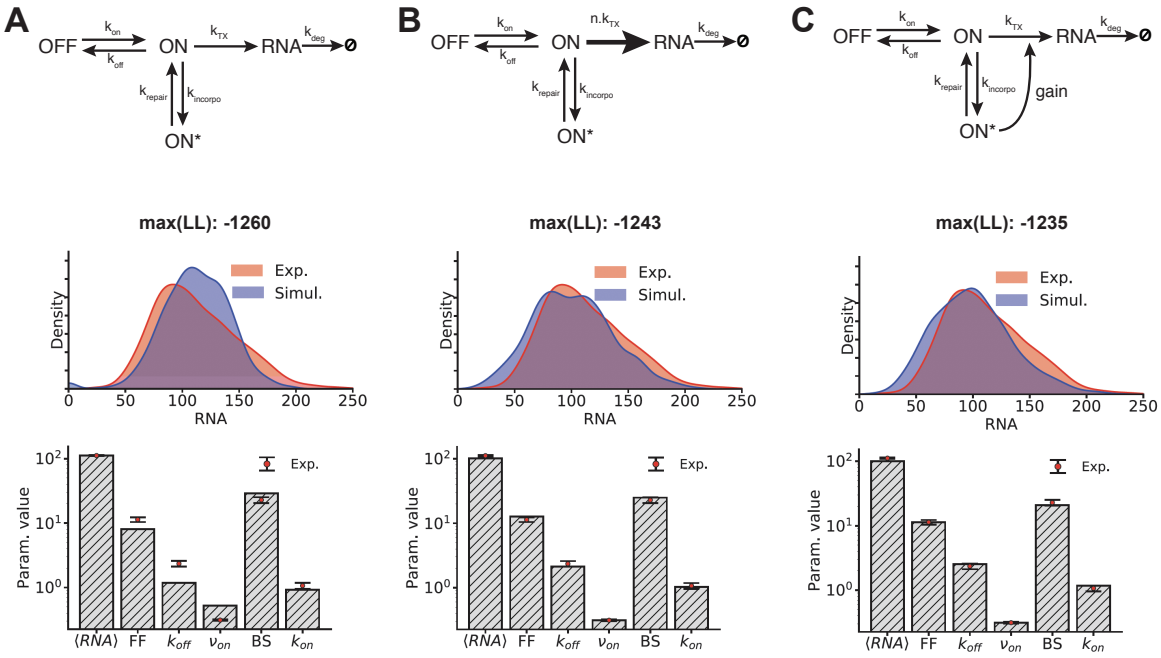


Figure S25: Control of transcription rate through feedforward mechanism in DiThR model allows for best fit of experimental data.
(A) Schematic and macroscopic behavior of DiThR model with no feedforward mechanism (i.e., no boost of transcription rate after repair completion). Simulation results are compared to experimental data of Nanog expression from smRNA-FISH of mESCs treated with 10μM IdU (supplementary text 6). **(B)** Schematic and macroscopic behavior of DiThR model with a manually increased k_{TX} (i.e., k_{TX} is increased by a factor n following completion of repair). **(C)** Schematic and macroscopic behavior of DiThR model with a feedforward mechanism (*gain*) where the proposed speedup of transcription is dependent on the frequency of time spent in the ON* state.

Supplemental Figure 26

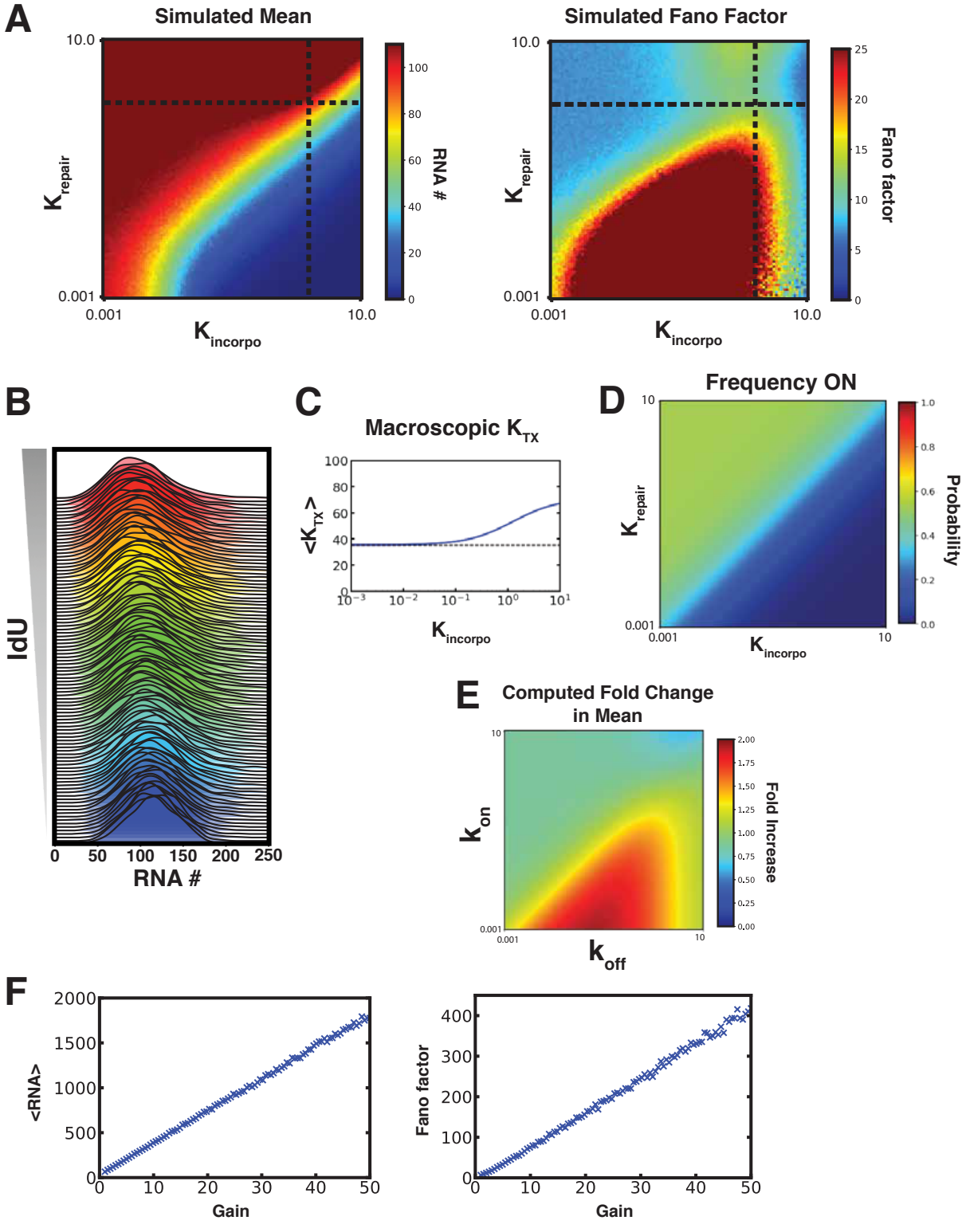


Figure S26: Sensitivity analysis of model parameters reveals phase-space for modulation of Nanog variability independently of mean.
(A) Heatmaps displaying mean (left) and Fano factor (right) of Nanog mRNA from simulation results of DiThR model as a function of k_{repair} and $k_{incorpo}$ values spanning four orders of magnitude (supplementary text 7.1). Dashed horizontal and vertical lines represent inferred values of k_{repair} and $k_{incorpo}$ that match experimental Nanog gene expression system in the presence of 10μM IdU. Multiple regions of the parameter phase-space exhibit constant mean output with unique levels of variability (Fano factor) demonstrating how mean and variability are tuned independently. **(B)** Simulated distributions of Nanog mRNA with increasing concentration of IdU which increases $k_{incorpo}$. Simulation results demonstrate how DiThR model allows for maintenance of mean output with increasing variability as concentration of IdU is increased. **(C)** Effective transcription rate of Nanog gene expression system as a function of $k_{incorpo}$. As IdU incorporation and subsequent Apex1 recruitment increases, the effective transcription rate increases as well. This represents the compensatory, feedforward mechanism of the DiThR model allowing for maintenance of mean output with increasing incorporation of IdU. **(D)** Heatmap displaying fraction of time that the Nanog gene expression is in the effective ON state as a function of k_{repair} and $k_{incorpo}$ values. **(E)** Heatmap displaying fold change in mean as a function of canonical k_{off} and k_{on} values (supplementary text 7.2). Fold change is calculated as the output of DiThR model relative to model 0 (canonical 2-state model) for the same set of k_{off} and k_{on} values. For a gene whose $k_{off} >> k_{on}$, addition of IdU to the system increases the mean output. **(F)** Mean mRNA and Fano factor of DiThR model output as a function of the *gain* term which describes how strongly the transcription rate is amplified following completion of repair (supplementary text 7.3).

Supplementary Figure 27

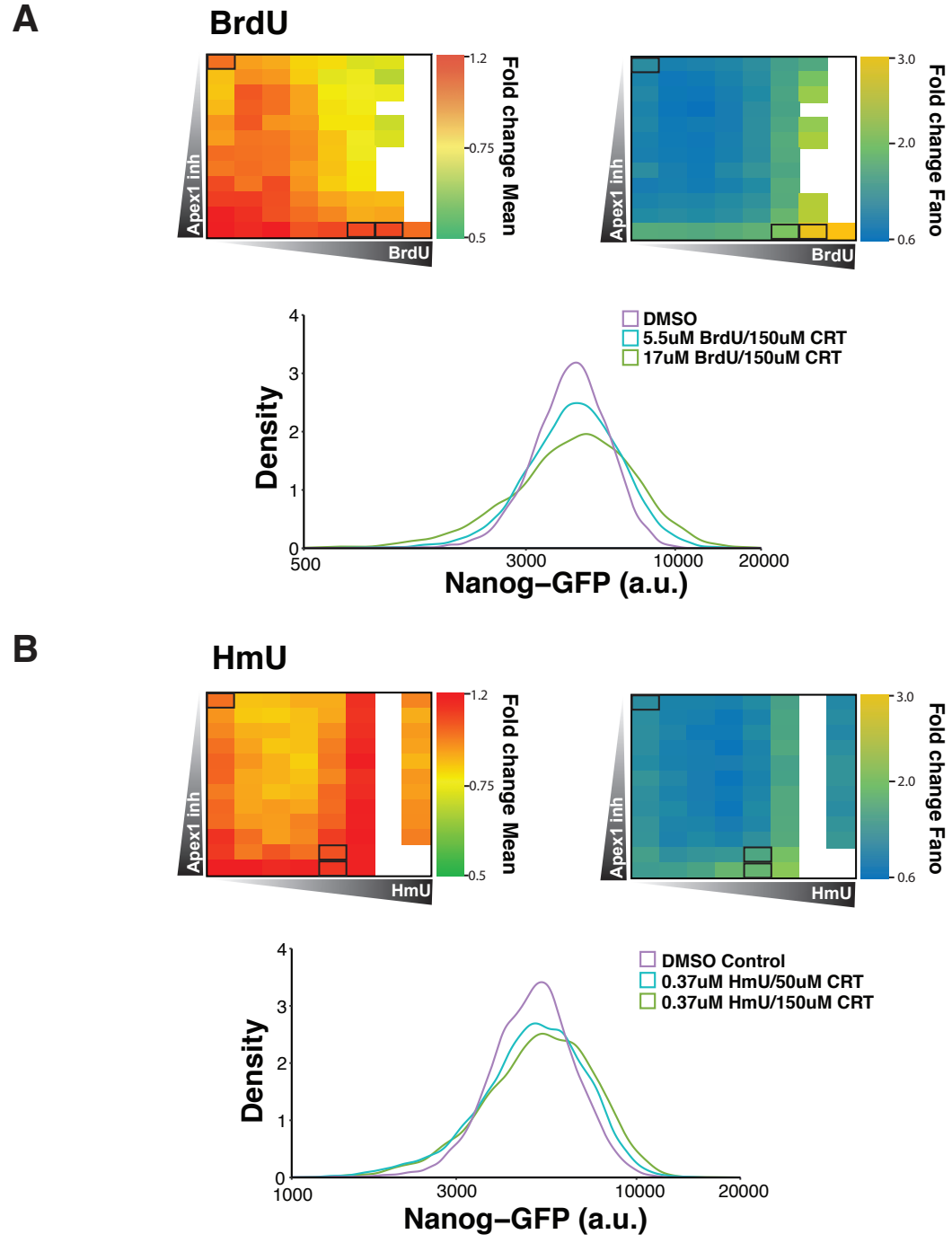


Figure S27: Treatment with BrdU or hmU in combination with CRT0044876 allows for tuning of Nanog variability independently of the mean.
(A) Testing of 96 concentration combinations of BrdU and CRT0044876 (Apex1 endonuclease domain inhibitor) to validate tunability of Nanog variability. BrdU and CRT0044876 were used to increase binding and decrease unbinding of Apex1 respectively. Nanog-GFP mESCs grown in 96-well plates were treated with 12 concentrations of CRT0044876 ranging from 0 to 150uM in combination with 8 concentrations of BrdU ranging from 0 to 50uM. Data represent average of two biological replicates. (Top left and top right panels) 96-well heatmaps displaying fold change in Nanog mean and Fano factor for each drug combination as compared to DMSO (top-leftmost well). Insufficient number of cells (<50,000) for extrinsic noise filtering were recorded from white wells. (Bottom Panel) Representative flow cytometry distributions from highlighted wells (black rectangles). Nanog variability increases independently of the mean. **(B)** Testing of 96 concentration combinations of HmU and CRT0044876 (Apex1 endonuclease domain inhibitor) to validate tunability of Nanog variability. hmU is a naturally found, Tet-induced oxidation product of thymine. Nanog-GFP mESCs grown in 96-well plates were treated with 12 concentrations of CRT0044876 ranging from 0 to 150uM in combination with 8 concentrations of hmU ranging from 0 to 10uM. Data represent average of two biological replicates. (Top left and top right panels) 96-well heatmaps displaying fold change in Nanog mean and Fano factor for each drug combination as compared to DMSO (top-leftmost well). Insufficient number of cells (<50,000) for extrinsic noise filtering were recorded from white wells. (Bottom Panel) Representative flow cytometry distributions from highlighted wells (black rectangles). As with IdU and BrdU, Nanog variability increases independently of the mean.

Supplemental Figure 28

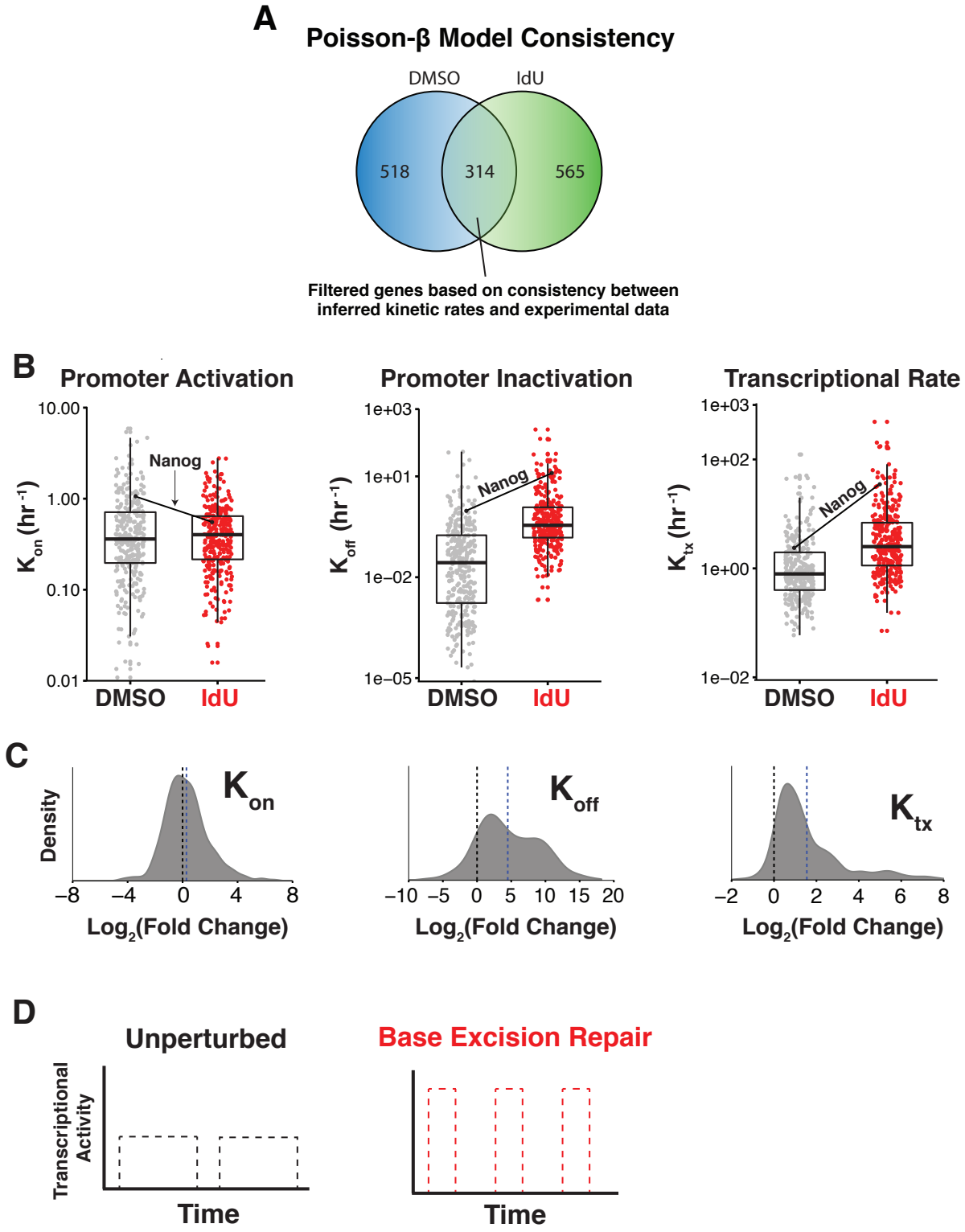


Figure S28: Highly variable genes exhibit shorter but more intense transcriptional bursts. **(A)** Of the 945 genes classified as highly variable with IdU treatment, we were able to estimate parameters of the 2-state model for 314 of these genes (supplementary text 8). **(B)** Boxplots show median ±interquartile range of parameter estimates with each point representing a gene. **(C)** Distributions of fold change in bursting kinetics between IdU and DMSO conditions for 314 highly variable genes. Dashed blue line signifies mean of distribution. Majority of highly variable genes exhibit increased K_{OFF} and K_{TX}, which is consistent with DiThR model and changes in bursting kinetics observed for Nanog. **(D)** Base-excision repair orchestrates shorter but more intense transcriptional bursts to maintain mean expression for genes with diverse bursting kinetics.

Supplementary Figure 29

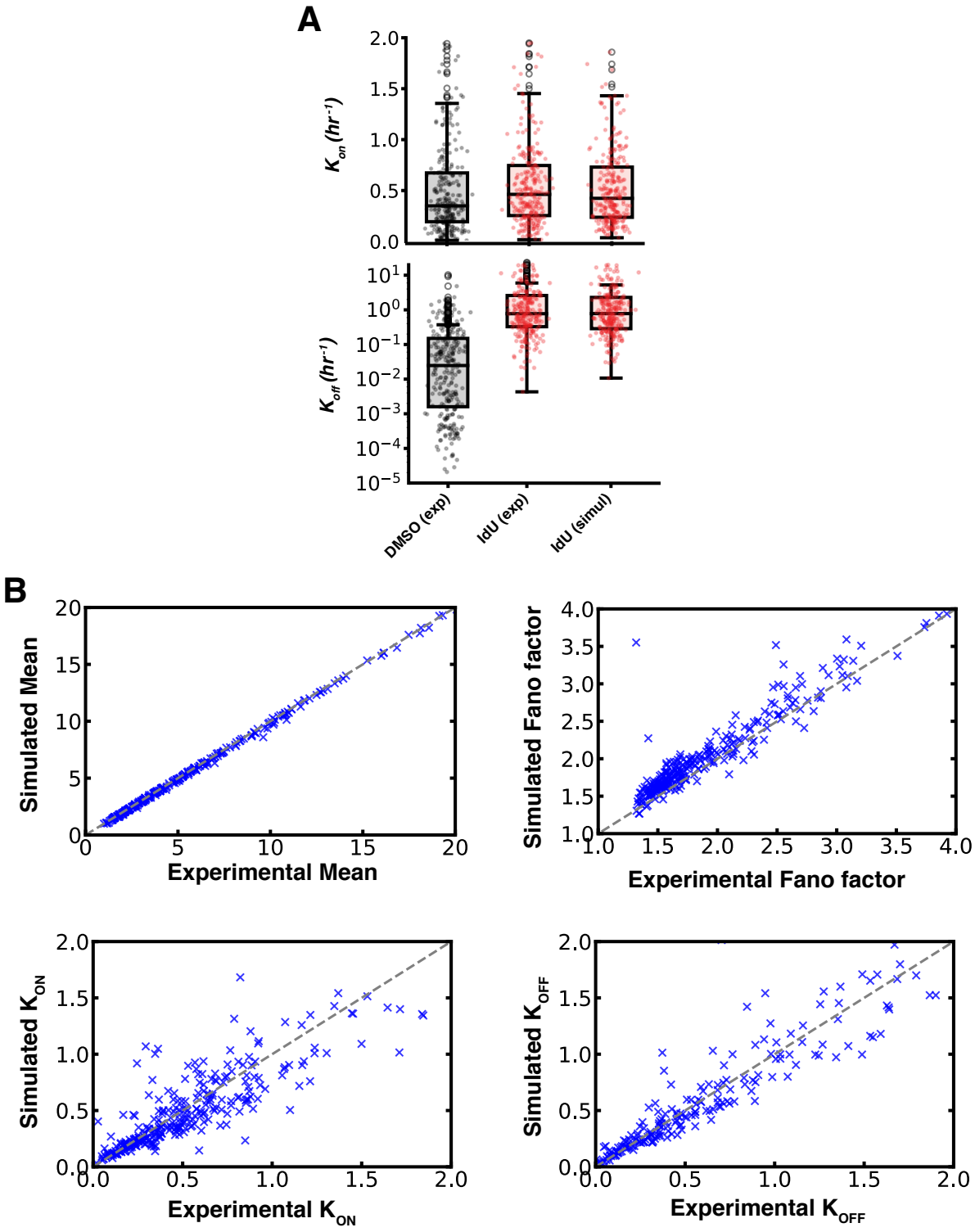


Figure S29: DiThR model provides unifying mechanism for noise-enhancement of genes with different bursting kinetics.
(A) Experimental values (exp) of effective K_{ON} and K_{OFF} in IdU condition, as derived from a moments-matching technique applied to scRNA-seq data, are compared to predicted values (simul) derived from simulations of DiThR model in the IdU condition. Each point represents a gene. Boxplots show median $\pm$interquartile range of parameter values. **(B)** Experimental values of mean, Fano factor, K_{ON}, and K_{OFF} (based on scRNA-seq data) in the IdU condition are compared to simulated values derived from DiThR model.

Supplemental Figure 30

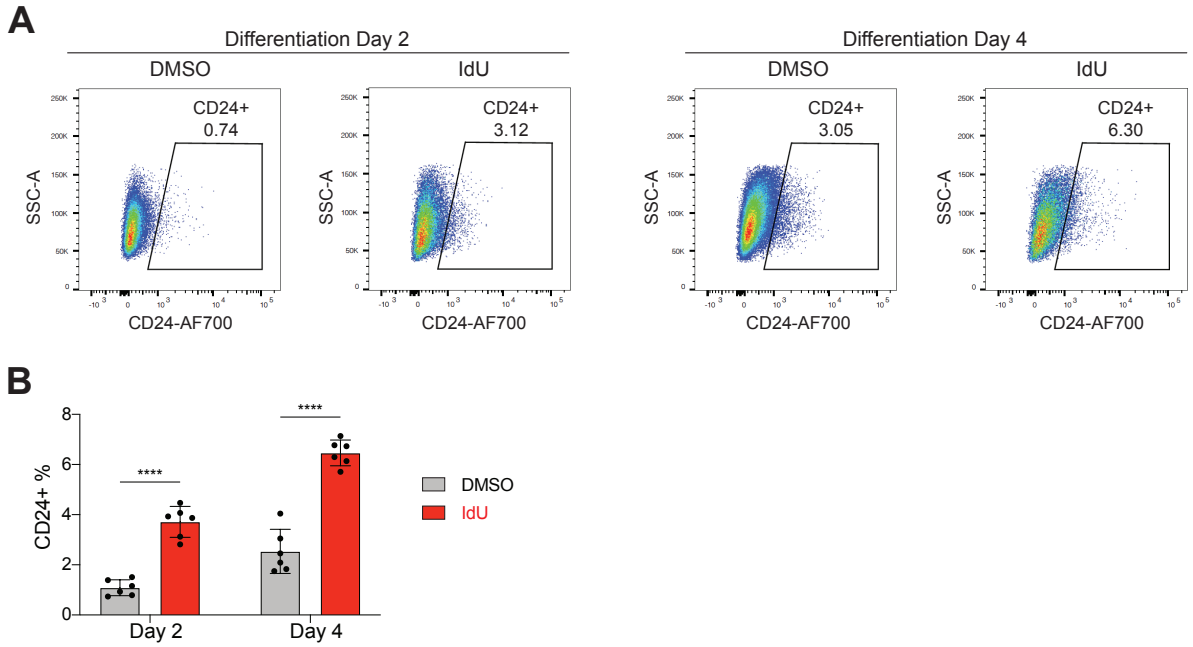


Figure S30: Noise-enhancement potentiates differentiation of mESCs into neural ectoderm lineage.
(A) Representative flow cytometry dot plots of CD24 staining at days 2 and 4 of differentiation with 4uM IdU or equivalent volume DMSO added for first 48 hours. **(B)** Percent of CD24(+) cells at days 2 and 4 of differentiation as measured by flow cytometry. Data represent mean and SD of 6 replicates for each condition. ****$p < 0.0001$, by a two-tailed, unpaired Student's t test.

Supplemental Figure 31

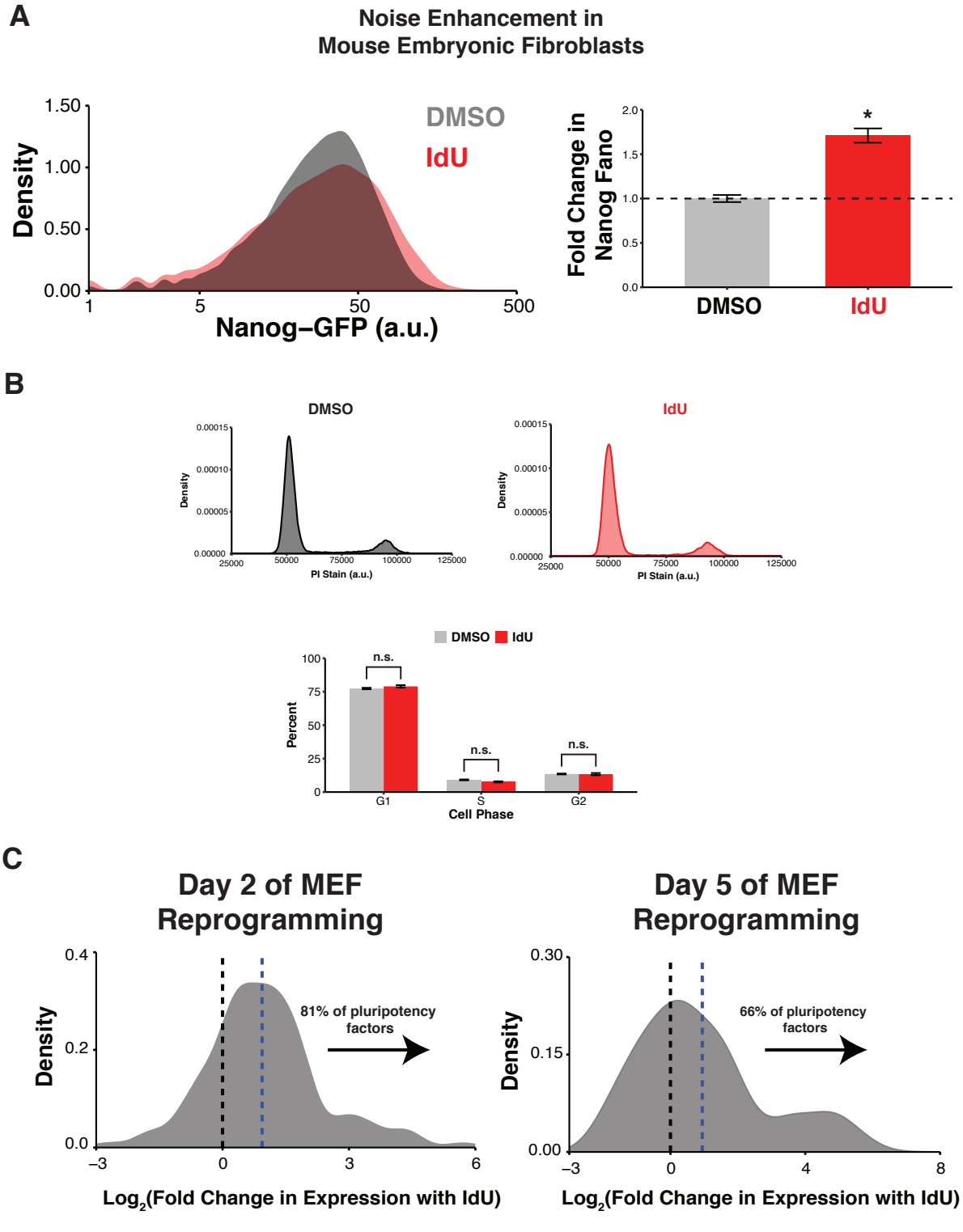


Figure S31: IdU treatment enhances conversion of mouse embryonic fibroblasts (MEFs) into induced pluripotent stem cells (iPSCs).
(A) Secondary MEFs with the endogenous Nanog locus tagged with GFP were treated with 4μM IdU or equivalent volume DMSO for 48 hours in MEF media. (Right) Representative flow cytometry distributions of Nanog-GFP expression in secondary MEFs after 48 hour treatment with IdU or DMSO. (Left) Quantification of Nanog Fano factor demonstrates that IdU treatment increases expression variability as compared to DMSO control (*p = 0.003, by a two-tailed, unpaired Student's t test). Data represent mean and SD of three biological replicates. **(B)** (Top) Representative flow cytometry distributions of propidium iodide staining for Nanog-GFP secondary MEFs treated with either DMSO or 4μM IdU for 48 hours in MEF media. No signs of aneuploidy are visible, indicating Nanog expression variability is not due to cell-to-cell variability in gene copy numbers. (Bottom) Percent of cells in each phase of the cell cycle for DMSO and IdU treatments based on propidium iodide staining. IdU treatment does not alter cell-cycle progression, indicating enhanced reprogramming is not due to accelerated cellular division. Data represent mean and SD of three biological replicates. P values were calculated using a two-tailed, unpaired Student's t test. **(C)** Bulk RNA-seq was conducted on days 2 and 5 of doxycycline-induced reprogramming of secondary MEFs supplemented with 4μM IdU or equivalent volume DMSO for the first 48 hours. Distributions of fold change in expression for 129 pluripotency genes (taken from Mouse Genome Informatics, gene ontology term: 0019827) in the IdU condition as compared to the DMSO control are shown. Dashed blue line represents mean of distribution. 81% and 66% of the pluripotency factors show increased expression with the addition of IdU as compared to DMSO control at days 2 and 5 of reprogramming, respectively. Noise amplification during early stages of reprogramming accelerates activation of pluripotency network.

Supplemental Figure 32

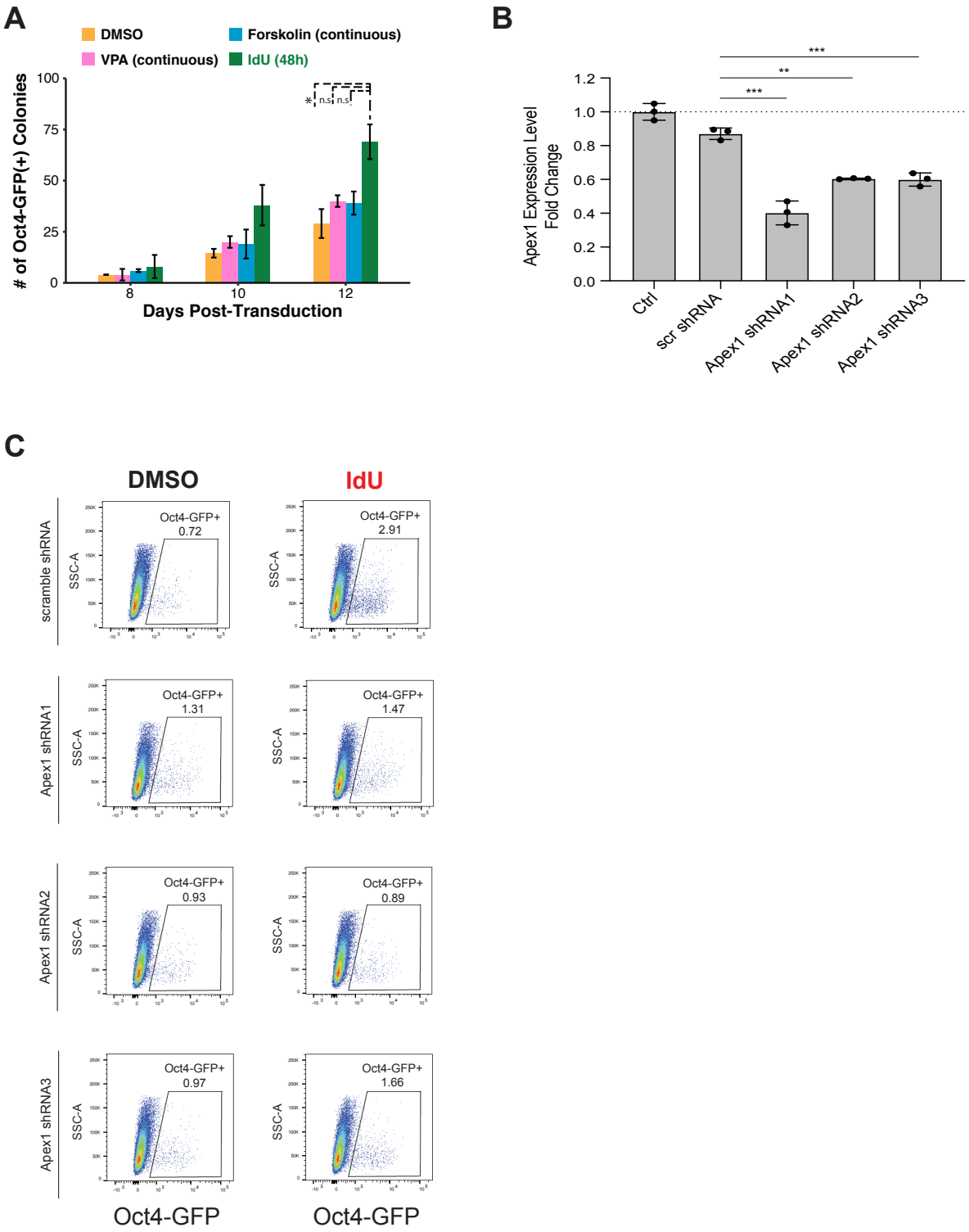


Figure S32: Knockdown of Apex1 in MEFs ablates IdU-mediated enhancement of reprogramming efficiency.
(A) Oct4-GFP primary MEFs (seeded at 10,000 cells/cm2) were retrovirally transduced with cDNAs encoding Oct4, Sox2, Klf4, and c-Myc. 24 hours after transduction, infected cells were treated with DMSO (continuously), 1mM valproic acid (VPA, continuously), 10μM forskolin (continuously), or 4μM IdU (first 48 hours). VPA and forskolin are established enhancers of cellular reprogramming. The number of Oct4-GFP(+) stem cell colonies were counted 8, 10, and 12 days from the start of drug treatment. Data represent mean and SD of 2 biological replicates. Treatment of transduced MEFs with IdU during early stages of reprogramming increases the number of Oct4-GFP(+) colonies that form as compared to DMSO control, *p =0.039 by one-way ANOVA with Bonferroni post hoc test. **(B)** Validation of shRNA knockdown of Apex1 in Oct4-GFP secondary MEFs via qPCR measurements. Relative mRNA levels of Apex1 were quantified using Gapdh as a reference gene. $\Delta\Delta C_t$ method was used with the empty-vector cell population as the control. Data represent mean and SD of biological triplicates. **(C)** Representative flow cytometry dot plots of Oct4-GFP expression at day 10 of reprogramming for secondary MEFs with Apex1 depleted. Oct4-GFP secondary MEFs were treated with either 4μM IdU or equivalent volume DMSO for first 48 hours of reprogramming in ESC media.

Captions for supplementary tables S1-S8

Table S1 (attached separately)
Sequences of smRNA-FISH oligonucleotide probes for first intron of Nanog and GFP.

Table S2 (attached separately)
Inferred effective kinetic rates of 2-state random telegraph model for Nanog in DMSO and IdU conditions.

Table S3 (attached separately)
List of nucleoside analogs that were screened for ability to increase to Nanog protein variability.

Table S4 (attached separately)
Gene targets and sequences of gRNAs used in CRISPRi screen.

Table S5 (attached separately)
Sequences of primers used for RT-qPCR verification of Apex1 and Tk1 knockdown.

Table S6 (attached separately)
Concentrations and layout of compound plates used for testing of IdU, BrdU, or hmU in combination with CRT0044876.

Table S7 (attached separately)
Sequences of primers used for ChIP-qPCR measurement of Apex1 interaction with Nanog promoter.

Table S8 (attached separately)
Sequences of Apex1 shRNAs and RT-qPCR primers used in Figure S32.

References and Notes

1. L. Boltzmann, Weitere Studien uber das Wärmegleichgewicht unter Gasmolekulen, Sitzungsber. Kais. Akad. Wiss. Wien Math. *Naturwiss* **66**, 275–370 (1872).

2. S. Arrhenius, Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Z. Phys. Chem.* **4**, 226–248 (1889). doi:10.1515/zpch-1889-0416

3. L. Roberts, Picture coding using pseudo-random noise. *IEEE Trans. Inf. Theory* **8**, 145–154 (1962). doi:10.1109/TIT.1962.1057702

4. P. Fatt, B. Katz, Some observations on biological noise. *Nature* **166**, 597–598 (1950). doi:10.1038/166597a0 Medline

5. A. A. Priplata, J. B. Niemi, J. D. Harry, L. A. Lipsitz, J. J. Collins, Vibrating insoles and balance control in elderly people. *Lancet* **362**, 1123–1124 (2003). doi:10.1016/S0140-6736(03)14470-4 Medline

6. D. Cohen, Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**, 119–129 (1966). doi:10.1016/0022-5193(66)90188-3 Medline

7. A. Arkin, J. Ross, H. H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics* **149**, 1633–1648 (1998). doi:10.1093/genetics/149.4.1633 Medline

8. J. L. Spudich, D. E. Koshland Jr., Non-genetic individuality: Chance in the single cell. *Nature* **262**, 467–471 (1976). doi:10.1038/262467a0 Medline

9. P. B. Gupta, C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, E. S. Lander, Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011). doi:10.1016/j.cell.2011.07.026 Medline

10. L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, D. V. Schaffer, Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182 (2005). doi:10.1016/j.cell.2005.06.006 Medline

11. A. Raj, A. van Oudenaarden, Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008). doi:10.1016/j.cell.2008.09.050 Medline

12. J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006). doi:10.1016/j.cub.2006.03.092 Medline

13. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005). doi:10.1016/j.cell.2005.09.031 Medline

14. T. B. Kepler, T. C. Elston, Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001). doi:10.1016/S0006-3495(01)75949-8 Medline

15. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLOS Biol.* **4**, e309 (2006). doi:10.1371/journal.pbio.0040309 Medline

16. W. J. Blake, M. KAErn, C. R. Cantor, J. J. Collins, Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003). doi:10.1038/nature01546 Medline

17. J. Peccoud, B. Ycart, Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995). doi:10.1006/tpbi.1995.1027

18. J. Rodriguez, D. R. Larson, Transcription in living cells: Molecular mechanisms of bursting. *Annu. Rev. Biochem.* **89**, 189–212 (2020). doi:10.1146/annurev-biochem-011520-105250 Medline

19. D. Nicolas, B. Zoller, D. M. Suter, F. Naef, Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7153–7158 (2018). doi:10.1073/pnas.1722330115 Medline

20. A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, N. Barkai, Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006). doi:10.1038/ng1807 Medline

21. R. D. Dar, S. M. Shaffer, A. Singh, B. S. Razooky, M. L. Simpson, A. Raj, L. S. Weinberger, Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. *PLOS ONE* **11**, e0158298 (2016). doi:10.1371/journal.pone.0158298 Medline

22. J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, J. S. Weissman, Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006). doi:10.1038/nature04785 Medline

23. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (Chapman & Hall/CRC, 2007).

24. R. D. Dar, N. N. Hosmane, M. R. Arkin, R. F. Siliciano, L. S. Weinberger, Screening for noise in gene expression identifies drug synergies. *Science* **344**, 1392–1396 (2014). doi:10.1126/science.1250220 Medline

25. M. M. K. Hansen, W. Y. Wen, E. Ingerman, B. S. Razooky, C. E. Thompson, R. D. Dar, C. W. Chin, M. L. Simpson, L. S. Weinberger, A post-transcriptional feedback mechanism for noise suppression and fate stabilization. *Cell* **173**, 1609–1621.e15 (2018). doi:10.1016/j.cell.2018.04.005 Medline

26. Y. Li, Y. Shan, R. V. Desai, K. H. Cox, L. S. Weinberger, J. S. Takahashi, Noise-driven cellular heterogeneity in circadian periodicity. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10350–10356 (2020). doi:10.1073/pnas.1922388117 Medline

27. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi:10.1038/nbt.4096 Medline

28. M. M. K. Hansen, R. V. Desai, M. L. Simpson, L. S. Weinberger, Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability. *Cell Syst.* **7**, 384–397.e6 (2018). doi:10.1016/j.cels.2018.08.002 Medline

29. B. Munsky, G. Neuert, A. van Oudenaarden, Using gene expression noise to understand gene regulation. *Science* **336**, 183–187 (2012). doi:10.1126/science.1216379 Medline

30. D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474 (2011). doi:10.1126/science.1198817 Medline

31. N. Eling, A. C. Richard, S. Richardson, J. C. Marioni, C. A. Vallejos, Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data. *Cell Syst.* **7**, 284–294.e12 (2018). doi:10.1016/j.cels.2018.06.011 Medline

32. M. Kaern, T. C. Elston, W. J. Blake, J. J. Collins, Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005). doi:10.1038/nrg1615 Medline

33. A. Scialdone, K. N. Natarajan, L. R. Saraiva, V. Proserpio, S. A. Teichmann, O. Stegle, J. C. Marioni, F. Buettner, Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015). doi:10.1016/j.ymeth.2015.06.021 Medline

34. J. M. Pedraza, A. van Oudenaarden, Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005). doi:10.1126/science.1109090 Medline

35. R. Bargaje, K. Trachana, M. N. Shelton, C. S. McGinnis, J. X. Zhou, C. Chadick, S. Cook, C. Cavanaugh, S. Huang, L. Hood, Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2271–2276 (2017). doi:10.1073/pnas.1621412114 Medline

36. M. Mojtahedi, A. Skupin, J. Zhou, I. G. Castaño, R. Y. Y. Leong-Quong, H. Chang, K. Trachana, A. Giuliani, S. Huang, Cell fate decision as high-dimensional critical state transition. *PLOS Biol.* **14**, e2000640 (2016). doi:10.1371/journal.pbio.2000640 Medline

37. C. Sokolik, Y. Liu, D. Bauer, J. McPherson, M. Broeker, G. Heimberg, L. S. Qi, D. A. Sivak, M. Thomson, Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst.* **1**, 117–129 (2015). doi:10.1016/j.cels.2015.08.001 Medline

38. L. A. Carvajal, D. B. Neriah, A. Senecal, L. Benard, V. Thiruthuvanathan, T. Yatsenko, S.-R. Narayanagari, J. C. Wheat, T. I. Todorova, K. Mitchell, C. Kenworthy, V. Guerlavais, D. A. Annis, B. Bartholdy, B. Will, J. D. Anampa, I. Mantzaris, M. Aivado, R. H. Singer, R. A. Coleman, A. Verma, U. Steidl, Dual inhibition of MDMX and MDM2 as a therapeutic strategy in leukemia. *Sci. Transl. Med.* **10**, eaao3003 (2018). doi:10.1126/scitranslmed.aao3003 Medline

39. E. Abranches, A. M. V. Guedes, M. Moravec, H. Maamar, P. Svoboda, A. Raj, D. Henrique, Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* **141**, 2770–2779 (2014). doi:10.1242/dev.108910 Medline

40. T. Kalmar, C. Lim, P. Hayward, S. Muñoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, A. Martinez Arias, Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLOS Biol.* **7**, e1000149 (2009). doi:10.1371/journal.pbio.1000149 Medline

41. D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, M. L. Simpson, Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006). doi:10.1038/nature04194 Medline

42. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005). doi:10.1126/science.1106914 Medline

43. A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, U. Alon, Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006). doi:10.1038/nature05316 Medline

44. S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010). doi:10.1038/nature09303 Medline

45. T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014). doi:10.1038/nchembio.1532 Medline

46. Y. F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011). doi:10.1126/science.1210944 Medline

47. L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013). doi:10.1016/j.cell.2013.04.002 Medline

48. E. S. J. Arnér, S. Eriksson, Mammalian deoxyribonucleoside kinases. *Pharmacol. Ther.* **67**, 155–186 (1995). doi:10.1016/0163-7258(95)00015-9 Medline

49. B. Demple, T. Herman, D. S. Chen, Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: Definition of a family of DNA repair enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11450–11454 (1991). doi:10.1073/pnas.88.24.11450 Medline

50. T. Lindahl, R. D. Wood, Quality control by DNA repair. *Science* **286**, 1897–1905 (1999). doi:10.1126/science.286.5446.1897 Medline

51. U. Müller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* **42**, 8592–8604 (2014). doi:10.1093/nar/gku552 Medline

52. M. L. Hegde, T. K. Hazra, S. Mitra, Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* **18**, 27–47 (2008). doi:10.1038/cr.2008.8 Medline

53. S. Madhusudan, F. Smart, P. Shrimpton, J. L. Parsons, L. Gardiner, S. Houlbrook, D. C. Talbot, T. Hammonds, P. A. Freemont, M. J. E. Sternberg, G. L. Dianov, I. D. Hickson, Isolation of a small molecule inhibitor of DNA base excision repair. *Nucleic Acids Res.* **33**, 4711–4724 (2005). doi:10.1093/nar/gki781 Medline

54. O. A. Kladova, M. Bazlekowa-Karaban, S. Baconnais, O. Piétrement, A. A. Ishchenko, B. T. Matkarimov, D. A. Iakovlev, A. Vasenko, O. S. Fedorova, E. Le Cam, B. Tudek, N. A. Kuznetsov, M. Saparbaev, The role of the N-terminal domain of human apurinic/apyrimidinic endonuclease 1, APE1, in DNA glycosylase stimulation. *DNA Repair (Amst.)* **64**, 10–25 (2018). doi:10.1016/j.dnarep.2018.02.001 Medline

55. C. D. Mol, T. Izumi, S. Mitra, J. A. Tainer, DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**, 451–456 (2000). doi:10.1038/35000249 Medline

56. D. R. McNeill, D. M. Wilson 3rd, A dominant-negative form of the major human abasic endonuclease enhances cellular sensitivity to laboratory and clinical DNA-damaging agents. *Mol. Cancer Res.* **5**, 61–70 (2007). doi:10.1158/1541-7786.MCR-06-0329 Medline

57. S. Chong, C. Chen, H. Ge, X. S. Xie, Mechanism of transcriptional bursting in bacteria. *Cell* **158**, 314–326 (2014). doi:10.1016/j.cell.2014.05.038 Medline

58. S. A. Sevier, D. A. Kessler, H. Levine, Mechanical bounds to transcriptional noise. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13983–13988 (2016). doi:10.1073/pnas.1612651113 Medline

59. S. Corless, N. Gilbert, Investigating DNA supercoiling in eukaryotic genomes. *Brief. Funct. Genomics* **16**, 379–389 (2017). doi:10.1093/bfgp/elx007 Medline

60. K. N. Babos, K. E. Galloway, K. Kisler, M. Zitting, Y. Li, Y. Shi, B. Quintino, R. H. Chow, B. V. Zlokovic, J. K. Ichida, Mitigating antagonism between transcription and proliferation allows near-deterministic cellular reprogramming. *Cell Stem Cell* **25**, 486–500.e9 (2019). doi:10.1016/j.stem.2019.08.005 Medline

61. P. Guptasarma, Cooperative relaxation of supercoils and periodic transcriptional initiation within polymerase batteries. *BioEssays* **18**, 325–332 (1996). doi:10.1002/bies.950180411 Medline

62. S. Kim, B. Beltran, I. Irnov, C. Jacobs-Wagner, Long-distance cooperative and antagonistic RNA polymerase dynamics via DNA supercoiling. *Cell* **179**, 106–119.e16 (2019). doi:10.1016/j.cell.2019.08.033 Medline

63. F. Kouzine, S. Sanford, Z. Elisha-Feil, D. Levens, The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.* **15**, 146–154 (2008). doi:10.1038/nsmb.1372 Medline

64. J. Liu, F. Kouzine, Z. Nie, H.-J. Chung, Z. Elisha-Feil, A. Weber, K. Zhao, D. Levens, The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression. *EMBO J.* **25**, 2119–2130 (2006). doi:10.1038/sj.emboj.7601101 Medline

65. M. Mizutani, T. Ohta, H. Watanabe, H. Handa, S. Hirose, Negative supercoiling of DNA facilitates an interaction between transcription factor IID and the fibroin gene promoter. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 718–722 (1991). doi:10.1073/pnas.88.3.718 Medline

66. S. S. Teves, S. Henikoff, Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* **21**, 88–94 (2014). doi:10.1038/nsmb.2723 Medline

67. M. Bazlekowa-Karaban, P. Prorok, S. Baconnais, S. Taipakova, Z. Akishev, D. Zembrzuska, A. V. Popov, A. V. Endutkin, R. Groisman, A. A. Ishchenko, B. T. Matkarimov, A. Bissenbaev, E. Le Cam, D. O. Zharkov, B. Tudek, M. Saparbaev, Mechanism of stimulation of DNA binding of the transcription factors by human apurinic/apyrimidinic endonuclease 1, APE1. *DNA Repair* **82**, 102698 (2019). doi:10.1016/j.dnarep.2019.102698 Medline

68. J. F. Breit, K. Ault-Ziel, A.-B. Al-Mehdi, M. N. Gillespie, Nuclear protein-induced bending and flexing of the hypoxic response element of the rat vascular endothelial growth factor promoter. *FASEB J.* **22**, 19–29 (2008). doi:10.1096/fj.07-8102com Medline

69. B. Leuner, E. R. Glasper, E. Gould, Thymidine analog methods for studies of adult neurogenesis are not equally sensitive. *J. Comp. Neurol.* **517**, 123–133 (2009). doi:10.1002/cne.22107 Medline

70. M. Delmans, M. Hemberg, Discrete distributional differential expression (D3E)—A tool for gene expression analysis of single-cell RNA-seq data. *BMC Bioinformatics* **17**, 110 (2016). doi:10.1186/s12859-016-0944-6 Medline

71. W. J. Blake, G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, J. J. Collins, Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006). doi:10.1016/j.molcel.2006.11.003 Medline

72. H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, S. Huang, Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008). doi:10.1038/nature06965 Medline

73. G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, M. B. Elowitz, Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719 (2007). doi:10.1126/science.1137455 Medline

74. J. Pruszak, W. Ludwig, A. Blak, K. Alavian, O. Isacson, CD15, CD24, and CD29 define a surface biomarker code for neural lineage differentiation of stem cells. *Stem Cells* **27**, 2928–2940 (2009). doi:10.1002/stem.211 Medline

75. S. Semrau, J. E. Goldmann, M. Soumillon, T. S. Mikkelsen, R. Jaenisch, A. van Oudenaarden, Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat. Commun.* **8**, 1096 (2017). doi:10.1038/s41467-017-01076-4 Medline

76. C. Naughton, N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra, N. Gilbert, Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20**, 387–395 (2013). doi:10.1038/nsmb.2509 Medline

77. D. Racko, F. Benedetti, J. Dorier, A. Stasiak, Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Res.* **46**, 1648–1660 (2018). doi:10.1093/nar/gkx1123 Medline

78. A. Singh, Transient changes in intercellular protein variability identify sources of noise in gene expression. *Biophys. J.* **107**, 2214–2220 (2014). doi:10.1016/j.bpj.2014.09.017 Medline

79. R. Desai, B. Martin, scRNA-seq and modeling code for: A DNA-repair pathway can affect transcriptional noise to promote cell fate transitions, Zenodo (2021); https://doi.org/10.5281/zenodo.4891977.

80. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted degradation of CTCF decouples local

insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017). doi:10.1016/j.cell.2017.05.004 Medline

81. P. Angerer, L. Haghverdi, M. Büttner, F. J. Theis, C. Marr, F. Buettner, destiny: Diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016). doi:10.1093/bioinformatics/btv715 Medline

82. F. Mueller, A. Senecal, K. Tantale, H. Marie-Nelly, N. Ly, O. Collin, E. Basyuk, E. Bertrand, X. Darzacq, C. Zimmer, FISH-quant: Automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013). doi:10.1038/nmeth.2406 Medline

83. H. Ochiai, T. Sugawara, T. Sakuma, T. Yamamoto, Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci. Rep.* **4**, 7125 (2014). doi:10.1038/srep07125 Medline

84. C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo, A. E. Carpenter, CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* **16**, e2005970 (2018). doi:10.1371/journal.pbio.2005970 Medline

85. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016). doi:10.7554/eLife.19760 Medline

86. B. S. Razooky, A. Pai, K. Aull, I. M. Rouzine, L. S. Weinberger, A hardwired HIV latency program. *Cell* **160**, 990–1001 (2015). doi:10.1016/j.cell.2015.02.009 Medline

87. L. V. Sharova, A. A. Sharov, T. Nedorezov, Y. Piao, N. Shaik, M. S. H. Ko, Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* **16**, 45–58 (2009). doi:10.1093/dnares/dsn030 Medline

88. W. Wongpaiboonwattana, M. P. Stavridis, Neural differentiation of mouse embryonic stem cells in serum-free monolayer culture. *J. Vis. Exp.* (99): e52823 (2015). doi:10.3791/52823 Medline

89. M. T. Walsh, X. Huang, Measurement of incorporation kinetics of non-fluorescent native nucleotides by DNA polymerases using fluorescence microscopy. *Nucleic Acids Res.* **45**, e175–e175 (2017). doi:10.1093/nar/gkx833 Medline

90. B. Zoller, D. Nicolas, N. Molina, F. Naef, Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11**, 823 (2015). doi:10.15252/msb.20156257 Medline

91. K. Bahar Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, S. Itzkovitz, Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* **13**, 2653–2662 (2015). doi:10.1016/j.celrep.2015.11.036 Medline

92. D. A. Gilchrist, G. Dos Santos, D. C. Fargo, B. Xie, Y. Gao, L. Li, K. Adelman, Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010). doi:10.1016/j.cell.2010.10.004 Medline

93. K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, E. Bertrand, A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016). doi:10.1038/ncomms12248 Medline

94. J. Zhang, T. Zhou, Promoter-mediated transcriptional dynamics. *Biophys. J.* **106**, 479–488 (2014). doi:10.1016/j.bpj.2013.12.011 Medline

95. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977). doi:10.1021/j100540a008

96. R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, L. S. Weinberger, Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17454–17459 (2012). doi:10.1073/pnas.1213530109 Medline

97. K. P. Burnham, D. R. Anderson, K. P. Burnham, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, ed. 2, 2002).