

RESEARCH

Open Access



3'-5' crosstalk contributes to transcriptional bursting

Massimo Cavallaro^{1,2,3*} , Mark D. Walsh¹, Matt Jones¹, James Teahan⁴, Simone Tiberi⁵, Bärbel Finkenstädt³ and Daniel Hebenstreit^{1*}

*Correspondence:

m.cavallaro@warwick.ac.uk;

d.hebenstreit@warwick.ac.uk

¹School of Life Sciences, University of Warwick, Coventry, UK

²Mathematics Institute and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK

³Department of Statistics, University of Warwick, Coventry, UK

Full list of author information is available at the end of the article

Abstract

Background: Transcription in mammalian cells is a complex stochastic process involving shuttling of polymerase between genes and phase-separated liquid condensates. It occurs in bursts, which results in vastly different numbers of an mRNA species in isogenic cell populations. Several factors contributing to transcriptional bursting have been identified, usually classified as intrinsic, in other words local to single genes, or extrinsic, relating to the macroscopic state of the cell. However, some possible contributors have not been explored yet. Here, we focus on processes at the 3' and 5' ends of a gene that enable reinitiation of transcription upon termination.

Results: Using Bayesian methodology, we measure the transcriptional bursting in inducible transgenes, showing that perturbation of polymerase shuttling typically reduces burst size, increases burst frequency, and thus limits transcriptional noise. Analysis based on paired-end tag sequencing (PoIII ChIA-PET) suggests that this effect is genome wide. The observed noise patterns are also reproduced by a generative model that captures major characteristics of the polymerase flux between the ends of a gene and a phase-separated compartment.

Conclusions: Interactions between the 3' and 5' ends of a gene, which facilitate polymerase recycling, are major contributors to transcriptional noise.

Keywords: Gene expression, Parameter inference, Mathematical modelling, Gene looping, Biological noise, Liquid-liquid phase separation

Introduction

In many cellular systems, mRNAs appear to be produced in burst-like fashion. This is directly observed in real-time experimental studies [1–3] and also agrees with theoretical analyses of steady-state mRNA distributions among single cells [4, 5]. Such bursty dynamics are thought to be the signature of gene regulation and are often described in terms of transcriptional “noise” [5, 6]. Due to the central role of transcription in cellular functions, it is important to understand the mechanisms from which the bursting originates [7].



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The microscopic dynamics underlying transcription are not yet well understood. Various factors have been found to influence transcriptional dynamics, mostly by modulating bursting parameters such as the size or frequency of bursts [3, 5]. These factors are often classified as either intrinsic or extrinsic, although this distinction is blurred in many cases. This classification originally derives from the observation that fluctuations in expression levels are partially correlated across multiple genes [8], thus suggesting common, extrinsic causes, while the remaining, independent fluctuations are intrinsic to each gene. Typical major extrinsic noise sources are the cell cycle [9–11] and cell-size fluctuations [12], the latter partially due to the former. Numerous additional factors such as neighbouring cells, cell morphology, and others have been found to affect transcription to varying degrees [13]. Intrinsic factors include non-linear transcription factor interactions [4, 5, 8], changing chromatin status [14, 15], promoter architecture [3], transcription factor diffusion [16], and several others [17–19].

It is unclear how these phenomena relate to the local environment at transcribing genes. These are associated to clusters of RNA polymerase II (PolII), which have been interpreted as “transcription factories” [20] and suggested to modulate the temporal patterns of transcription [21, 22]. More recently, it has been found that, in proximity to active genes, the PolIIs are incorporated in membrane-less droplets, maintained by liquid-liquid phase separation (LLPS) from the rest of the nucleus, with the net effect of locally increasing the population of factors involved in initiation; when PolII is liberated from this domain, transcription can be initiated [23–28]. LLPS also provides an explanation for the hitherto enigmatic action-at-a-distance type of gene regulation by distal enhancers, as the nuclear condensates are indeed able to restructure the genome, albeit results on LLPS are relatively preliminary at this stage [29, 30].

While a comprehensive description of the interactions between PolIIs, other factors, and the chromatin within these niches is missing, several observations suggest that termination is linked to reinitiation; these include the presence of the same factor species at both ends of a gene, the reduction of initiation upon perturbation of 3' processes, and protein interactions that have been suggested to juxtapose the promoter and the terminator DNA, forming a structure that has been referred to as a “gene loop” [31, 32]. Importantly, it has been demonstrated that 3'-end processing favours transcription initiation; the presence of such 3'-5' crosstalk in a gene increases its mean expression level [33]. The concept of LLPS appears highly important in this regard, as PolII undergoes a sequence of post-translational modifications on its C-terminal domain during transcription, while integration into phase-separated domains and reinitiation requires it to be unmodified [24]. In line with this, recent studies suggest that LLPS is also involved in 3'-end transcriptional processes [34]. We generically refer to the shuttling of PolIIs from 5' to 3', potentially passing through the LLPS compartment, as the recycling. It has been suggested that a repetitive cycle of reinitiation and termination due to these mechanisms is likely to produce a rapid succession of mRNA creation events, thus potentially contributing to the transcriptional bursts [35], but to the best of our knowledge, an experimental verification is as yet lacking.

In this paper, we investigate the interplay between bursty expression and 3'-5' interactions using an interdisciplinary approach. We first consider two integrated genes that permit studying transcription upon perturbation of their 3'-5' processes at different induction levels; we demonstrate that these interactions strikingly influence the

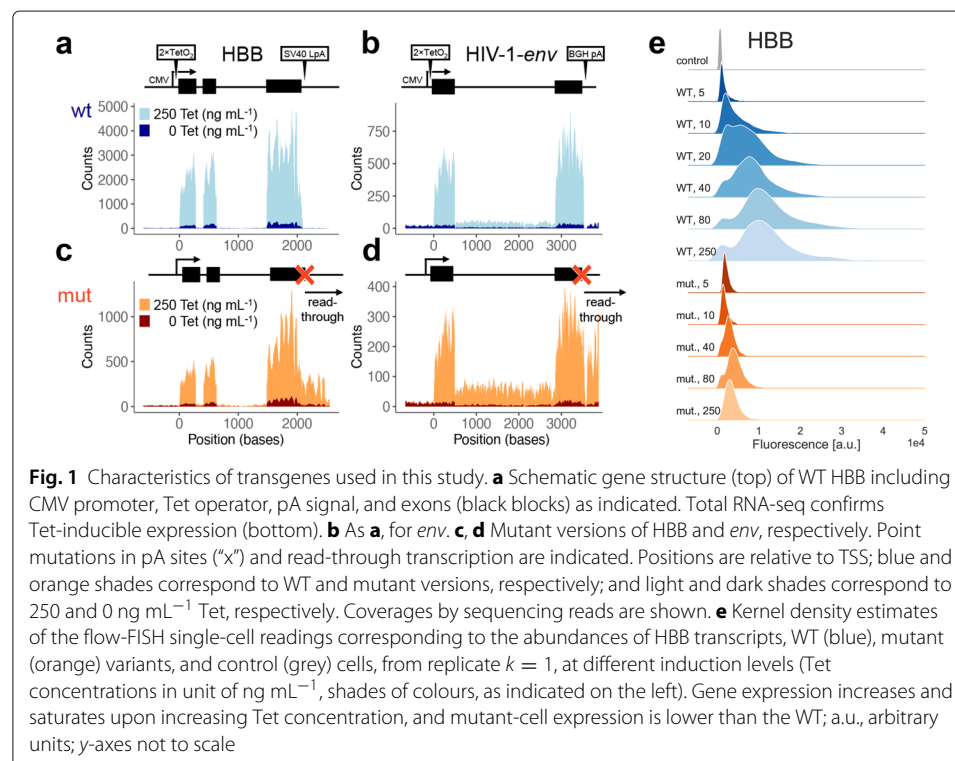
transcription kinetics and typically elicit the transcriptional noise, by decreasing burst frequency and increasing burst size. We then focus on genome-wide 3'-5' interactions involved in transcription by means of PolII ChIA-PET sequencing data, showing that they are related to the gene-expression parameters similarly to the transgenes' results. This scenario is well described by a microscopic stochastic model of gene expression, where tuning a single parameter—corresponding to the probability of local polymerase recycling—naturally yields the observed expression patterns, without involving extrinsic-noise contributors or alternative intrinsic mechanisms.

Results

Cell lines as model systems for PolII recycling

We utilised two HEK293 cell lines which contain on their genomes copies of the genes *β -globin* (HBB) [33] and a modified version of HIV-1-*env* [36], respectively, driven by inducible CMV promoters (Fig. 1a, b).

This transgene approach allowed us to exploit very well-characterised model systems for recycling perturbation, which achieve mono-allelic expression and, most importantly, allow precise control of expression levels with inducers [33]. The first gene, HBB, is an example for long-range chromosomal interactions in its native genomic neighbourhood. Its expression involves spatial proximity between the promoter and a locus control region (LCR) over 50 Kb away [37]. The LCR has been studied extensively in murine and human cells (see, for example, references [38, 39]) and jointly regulates expression of several *β -globin*-like genes at the locus, likely involving LLPS [40]. A recent study demonstrates burst-like expression of murine HBB and suggests that interactions between the LCR and the HBB promoter modulate the bursting parameters [9]. Our cell line features an ectopic



insertion of human HBB under control of a tetracycline (Tet) responsive promoter. A previous study of this system has provided a substantial number of results suggesting that 3' mRNA processing contributes to reinitiation of transcription [33]. This notion is based on several findings relating to the introduction of a single point mutation in the SV40 late poly-adenylation (pA) site (Fig. 1c). This includes decreased average mRNA expression levels, while “read-through” transcription downstream of the pA site is increased. Furthermore, the mutation leads to a decrease of PolII, TBP, and TFIIB levels at the promoter shortly after gene induction, and to an accumulation at the “read-through” region instead. Reduced transcription initiation compared to wild-type (WT) cells was also supported by nuclear run on assays and by a changed profile of post-translational modifications of PolII. Noticeably, TFIIB has been demonstrated to be functionally involved in linking 3' and 5' transcriptional activities [41], while post-translational modifications of PolII are in part carried out by Ssu72, which is associated with gene-loop formation in yeast [42] and appears to have similar roles in vertebrates [43]. A further recent study that utilised the ectopic HBB system reports direct detection of gene loops based on a 3C assay in the WT cell line, but not the mutant [44].

The second cell line, containing a Tet-inducible version of HIV-1-*env*, was previously studied in similar fashion to the HBB constructs. Results using a mutated version of the pA site (Fig. 1d) mirrored those obtained with HBB, suggesting extensive 3'-5' crosstalk and recycling of factors including polymerase [33, 45]. The *env* construct uses a BGH, not an SV40 pA site, which suggests that the findings are independent of the type of pA site. Notably, expression of the HIV-1 gene using its native long terminal repeat (LTR) promoter exhibits bursting dynamics [6].

We used these cell lines and their mutant versions as a model system for mammalian gene expression in the presence and absence of 3'-5' crosstalk. We confirmed by total RNA-seq that HBB and *env* mRNAs are expressed inducibly in all cell lines (Fig. 1a–d). At high Tet concentration (250 ng mL⁻¹), the fold changes over the un-induced samples were ≈ 16 and ≈ 26 for HBB and *env*, respectively. The mutants were expressed at lower levels and featured read-through transcription as described, with intact transcript sequences, i.e. not subject to splicing defects (Fig. 1c, d) [46–54]. This indicated specificity of the pA site mutations.

In order to detect transcripts at the single molecule level, we designed probes for single molecule RNA-FISH (smFISH) and confirmed detection of large transcript numbers upon Tet stimulation of the cells, while the expression of a control gene, AKT1, remained constant (Additional file 1: Section S1 and Figure S3). Microscopy-based smFISH is not ideal for HEK293 cells, since they tend to overlap and form aggregates when growing. We therefore decided to record the smFISH signal by adapting a flow-FISH technique based on flow cytometry [55]; this also resolves extrinsic-noise contributors such as cell size, morphology, and cycle, and, thanks to its high throughput, permits recording vast numbers of cells to analyse overall population structures (Additional file 1: Sections S1 and S7).

While the flow-cytometer fluorescence signal from stained cells serves as a proxy for the mRNA abundance, it is returned in arbitrary units (a.u.) rather than in absolute counts. We thus used microscope imaging and nCounter[®] data to calibrate the flow-FISH fluorescence readings of HBB and *env* cells, respectively. Applying the clustering algorithm

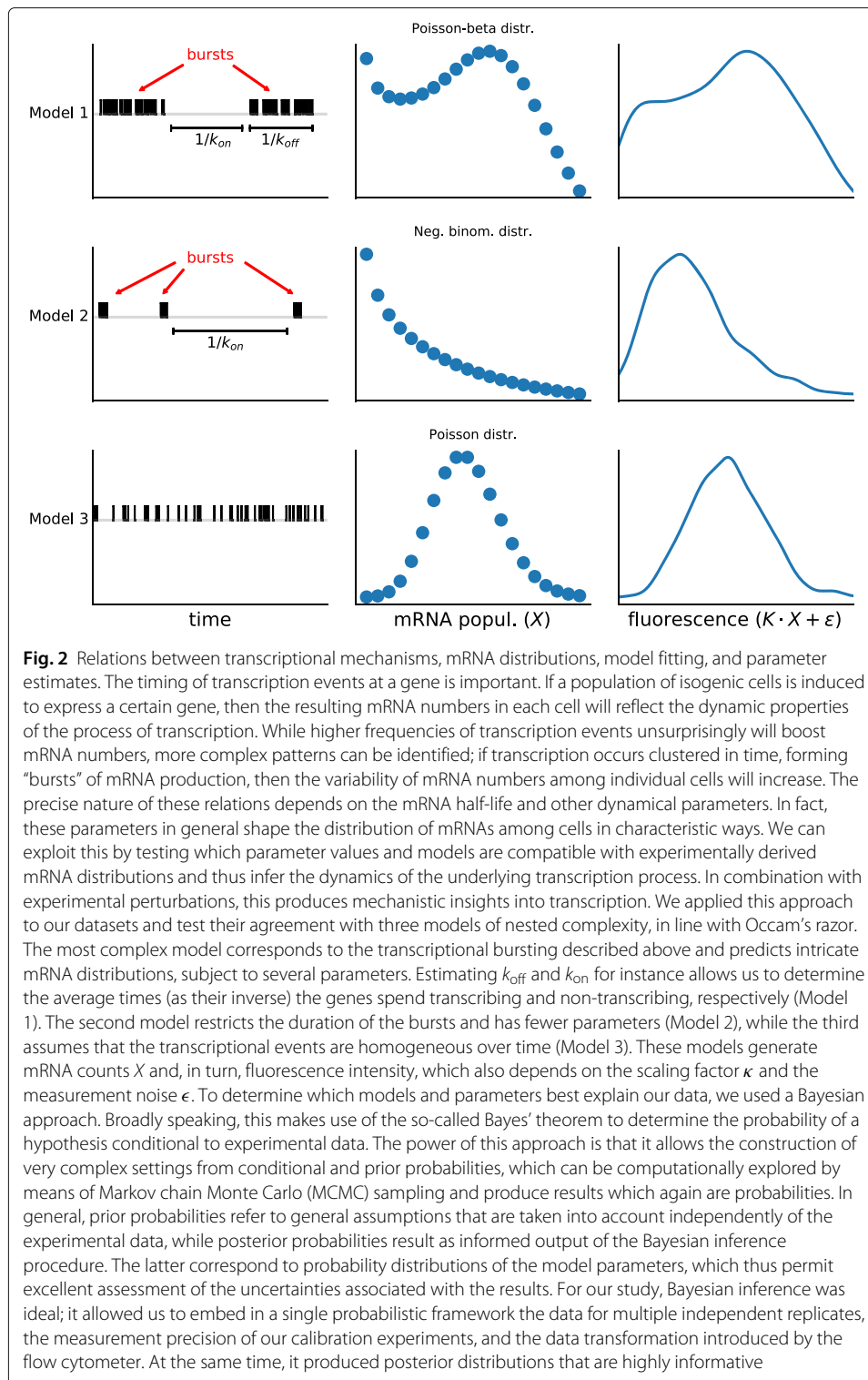
of [56–58] to the flow-FISH recordings allowed us to select single-cell readings against those from cell clumps, doublets, and debris (Additional file 1: Section S1 and Figure S1).

Flow-FISH data demonstrate Tet-dose dependent expression of HBB and *env*, indicating specific detection of transcripts above background noise. The stationary expression levels appeared to reach saturation at 80 ng mL⁻¹ Tet (Fig. 1e and Additional file 1: Section S1 and Figure S2). Staining for the DNA content demonstrates a mild increase of HBB and *env* expression with increasing cell cycle stage. We found that the contribution to the total variability, measured as the squared coefficient of variation (CV²) of the mRNA population, due to the cell cycle and size was minor (Additional file 1: Section S6) and therefore focused on local genic mechanisms to investigate the observed noise pattern. The measured signal includes a background of unspecific staining and auto-fluorescence of the cells, which is subtracted from the total signal [59]. To gauge this background, we deleted the *env* gene from its host cell line with Cas9 [60] and performed the staining procedure as before. The resulting control cells had low fluorescence intensity that remained virtually unchanged upon maximal Tet stimulation, thus confirming specificity of our system and validating the use of this control to estimate the background (Additional file 1: Section S1 Table S1). Nuclear RNA export was largely unaltered by the mutations (Wilcoxon rank sum test on nuclear/cytoplasmic ratios from 83 HBB cells at 250ng mL⁻¹ Tet, $P = 0.85$; for 203 *env* cells $P = 6 \cdot 10^{-9}$, but the ratios differed only by 10%). Note that flow-FISH and its analysis/interpretation are unaffected by nuclear export issues.

Increased transcriptional bursting upon 3'-5' crosstalk

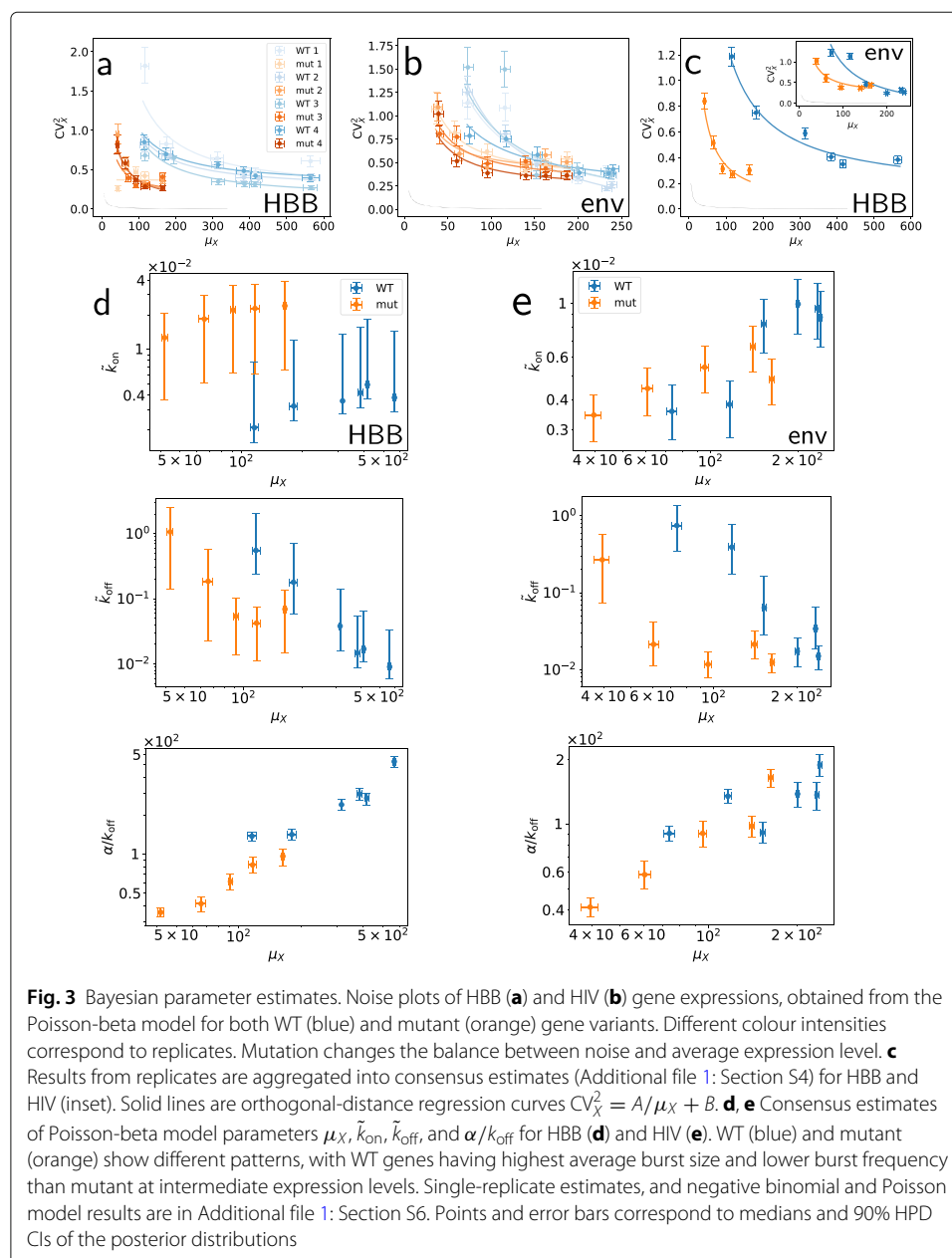
In order to gain insights into the transcriptional dynamics driving WT and mutant expression of HBB and *env*, we employed a Markov chain Monte Carlo (MCMC) sampling approach to fit statistical models to the flow-FISH data (Fig. 2). Importantly, Bayesian modelling permitted using microscope and nCounter[®] data to estimate informative prior distributions that calibrate the absolute mRNA quantification, while retaining flexibility in this respect. We further incorporated the background signal in the Bayesian framework based on the estimates from the Tet-stimulated control cells (“Materials and methods” section and Additional file 1: Sections S2-S3).

Our strategy requires flexible models to represent the absolute mRNA abundance. We considered three stochastic models of gene expression to capture the phenomenology of the transcription process (Fig. 2 and “Materials and methods” section). According to the first model, the gene can stay in an “on” state, in which transcription occurs at rate $\tilde{\alpha}$, or in an “off” state, in which no transcription occurs. The gene switches from “off” to “on” and “on” to “off” at rates \tilde{k}_{on} and \tilde{k}_{off} , respectively. Assuming that the mRNA degrades at constant rate \tilde{d} , this model corresponds to a Poisson-beta mixture distribution for the stationary per-cell mRNA population, which can be expressed in terms of the dimensionless rates α , k_{on} , and k_{off} (Additional file 1: Section S2) [4, 61]. The second model is a simplified version of the former two-state model, where α and k_{off} approach infinity, while the ratio α/k_{off} , which is referred to as the average burst size [62] and incorporated as a single parameter, is held finite; this model gives rise to a negative-binomial stationary mRNA distribution and allows much more efficient MCMC sampling than the Poisson-beta model (Additional file 1: Sections S3-S4). The third model is the most naïve as it assumes that transcription events of individual mRNAs occur independently at constant rate $\mu_X \cdot \tilde{d}$, where μ_X is the mean mRNA population, thus yielding a Poisson distributed



mRNA population at equilibrium which is thought to characterise genes with unregulated expression [5]. Noise levels consistent with the Poisson model [63, 64] or higher [4, 13] have both been reported in the literature. Estimates of the degradation rates \tilde{d} for both mutant and WT transgenes are listed in Additional file 1: Section S5 [65].

We obtained better fits for the Poisson-beta and the negative-binomial models than the Poisson model (Additional file 1: Sections S4 and S6) for all the replicates. In the Poisson-beta case, the MCMC traces of the rates k_{off} and α had a strong correlation; this revealed that most of the information about these two parameters is encoded in the ratio α/k_{off} (Additional file 1: Section S6 and Figure S10), which is more straightforwardly inferred by means of the negative-binomial model. In fact, for our data, these two models give consistent results in terms of CV^2 , average burst size α/k_{off} , and burst frequency \tilde{k}_{on} . To study the transcriptional noise, we obtained the CV^2 of the mRNA abundance (which we refer to as CV_X^2) from the estimated parameters (Additional file 1: Sections S2 and S6), and plotted it against the estimated mean expression levels μ_X (Fig. 3a–c). These reveal a trend observed before in other systems [6, 66–68], i.e. the transcriptional noise decreases



as μ_X increases, with the data of each experiment well fitted by a curve of the form $CV_X^2 = A/\mu_X + B$, and seems to approach a lower limit beyond which it does not further decrease. Such a limit is known as the noise floor [69–73]. Strikingly, the presence of the mutation alters the noise trends, thus suggesting that PolII recycling indeed contributes to the noise. The transcriptional noise at intermediate expression levels is significantly higher in WT than mutant cells. For the HBB gene, this pattern extends throughout the range of all induction levels. *Env* shows less pronounced differences between WT and mutant cells for the highest expression levels but resembles HBB otherwise. In all these cases, the noise clearly appears higher than postulated by the Poisson prediction curve $CV_X^2 = 1/\mu_X$ (solid lines in Fig. 3a–c).

Using the DNA content and the forward scatter signal (FSC-A) as proxies of the cell cycle progression and the cell size, respectively, we heuristically selected populations corresponding to G1, S, and G2 phases of three different sizes each from 40 ng mL⁻¹ Tet-induced cells (Fig. 4a–c); we fitted the negative-binomial model to their mRNA-expression reads, and estimated kinetic parameters and noise for each population, separately. Based on this, we found that the cell cycle and size, which typically are major extrinsic-noise contributors, only account for less than 20% of total mRNA variability for the transgenes (Fig. 4d–f), in contrast with [9, 10]; for further details, see Additional file 1: Section S7.

Modulation of rates

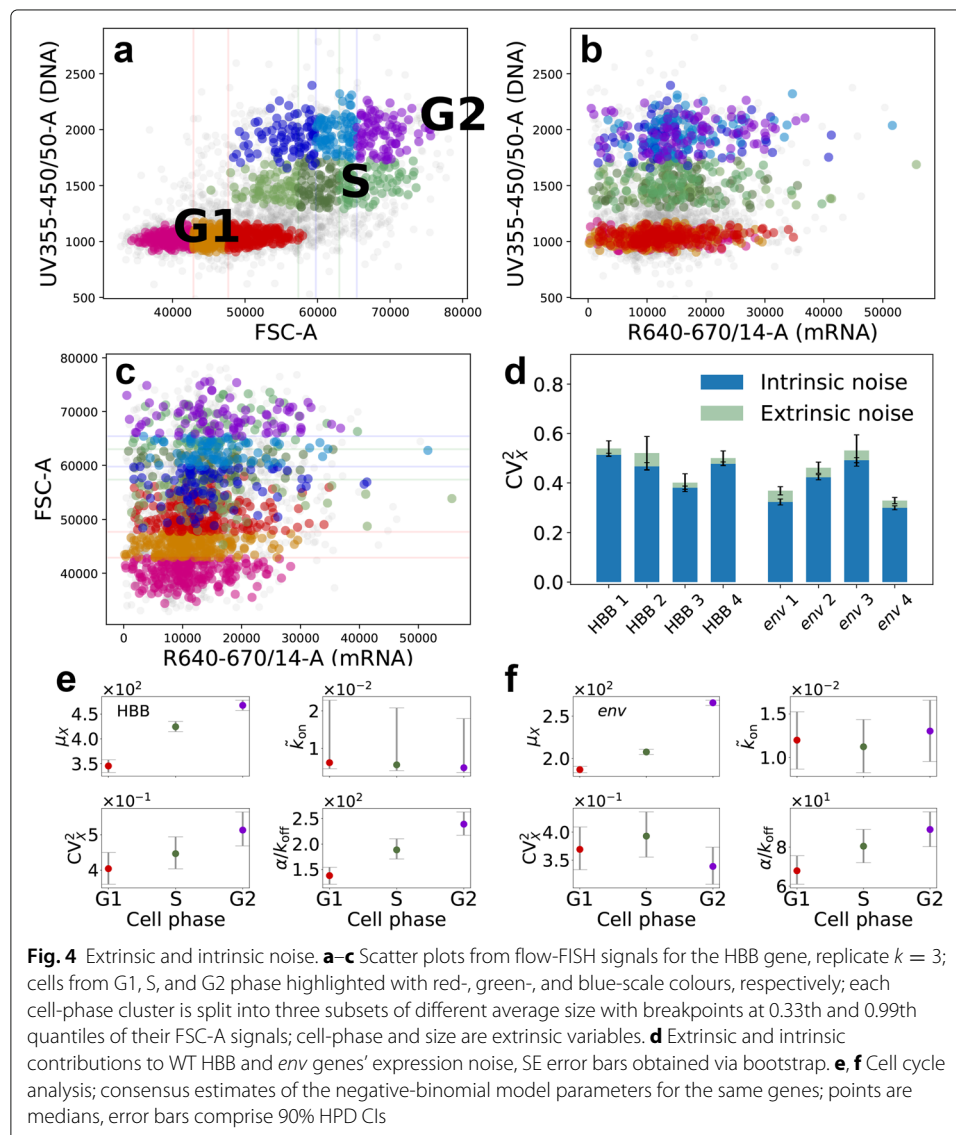
The overall rate estimates obtained from our fits are largely in agreement with previous findings from similar systems [3]. In fact, estimated values of \tilde{k}_{off} ranged up to ≈ 2.5 events per minute, with \tilde{k}_{on} roughly an order of magnitude lower. Increasing the Tet concentration boosts transcription by increasing the average burst size and the frequency \tilde{k}_{on} (Fig. 3d), thus shortening the average “off” state duration ($1/\tilde{k}_{\text{on}}$). Intriguingly, for the HBB gene, \tilde{k}_{on} is higher in mutant than WT cells in all cases, while the average burst size is lower in mutant cells in all cases. These patterns are less definite for the *env* gene but appear to support the conclusions from the HBB gene (Fig. 3e and Additional file 1: Section S6). In other words, the 3′-5′ crosstalk imposes a constraint on the transcriptional dynamics whose removal can cause bursts to be more frequent and smaller than in the WT gene.

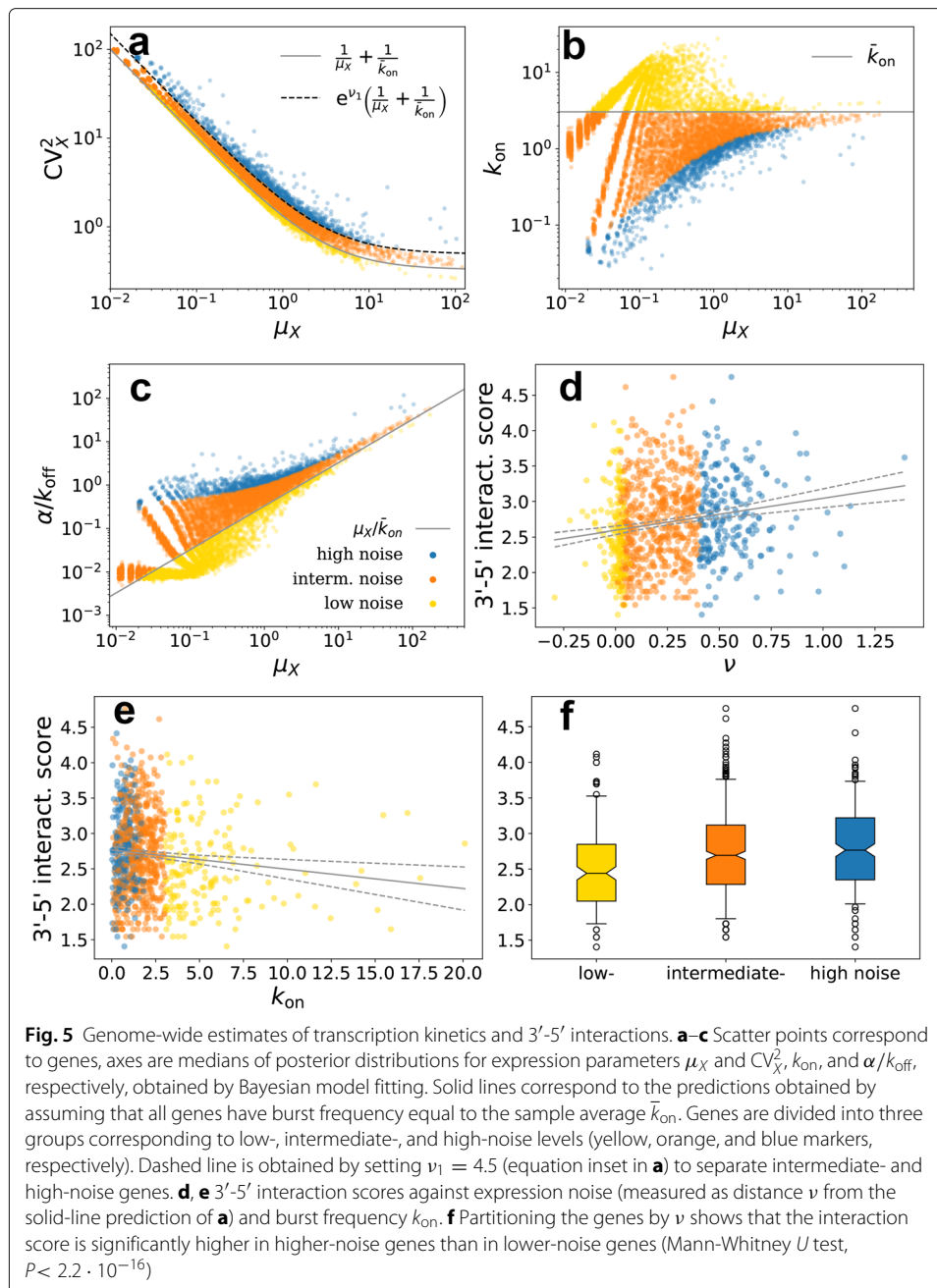
PolII-mediated 3′-5′ interactions by ChIA-PET

To jointly study the expression of a gene and its 3′-5′ interactions, we analysed publicly available datasets for the human cell line K562, obtained from chromatin-interaction analysis by paired-end tag sequencing (ChIA-PET) [74] and single-cell RNA-seq data (scRNAseq) [64]. We chose to use ChIA-PET against PolII to target chromatin interactions that are involved in transcription. We generated HiC-style interaction matrices (whose entries correspond to 2-Kb regions) from the ChIA-PET data using CHIA-PET2 [75]. We filtered the list of genes from the RefGene database with the hg19 reference genome to only contain those with unique gene symbols on chromosomes 1–22 and X, thus excluding alternatively spliced genes. As a proxy of the 3′-5′ interaction of a gene, we first aggregated the reads corresponding to the interaction between the bins that include its transcription start site (TSS) and transcription end site (TES). The resulting metrics depend on the gene length, which we addressed by dividing the

number of reads for each gene by the average read number from 10^4 genomic intervals of the same length as the gene, randomly sampled across the chromosome. We then applied the $\text{arcsinh}\sqrt{x+0.5}$ transformation to obtain a variance-stable interaction score [76]. Note that 5' to 3' interaction scores correlate with those for 5' to gene body interactions; this appears unsurprising, given that spatial proximity at one location will favour interaction signals at neighbouring regions, and is tangential to our analyses. We also discarded genes that are shorter than the resolution of our interaction matrices.

Fitting a negative-binomial distribution to the scRNA-seq UMI counts data of [64] allows us to conveniently classify expressed genes (sample UMI mean > 0.05) based on the estimated noise CV_X^2 , the burst frequency k_{on} , and the average burst size α/k_{off} (“Materials and methods” section, see also [7, 77, 78]). These are plotted against the mean expression μ_X in Fig. 5a–c. It is worth noting that burst frequency averaged over all the genes, \bar{k}_{on} , seems to determine the average trends of CV_X^2 and α/k_{off} . The noise trend



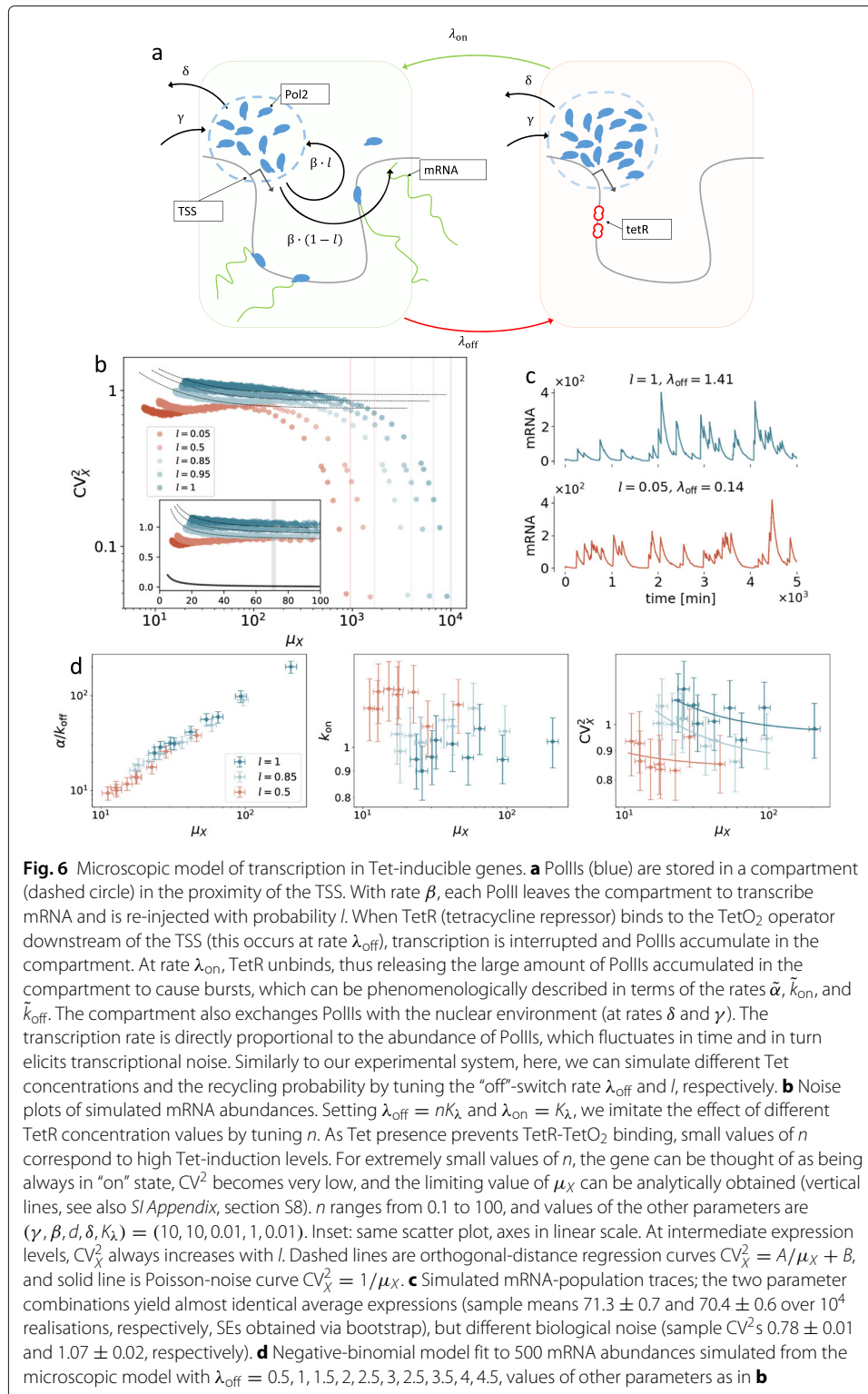


appears to be explained by the curve $CV_X^2 = 1/\mu_X + 1/\bar{k}_{on}$ (derived under the negative-binomial assumption, see Additional file 1: Section S2), which in fact separates the genes whose noise levels are higher than the mean predicts (blue and orange markers in Fig. 5) from those whose noise is lower than the prediction (yellow markers). As a measure of the deviation from this prediction, for each gene, we calculated the vertical distance ν of its expression noise to the curve $CV_X^2 = 1/\mu_X + 1/\bar{k}_{on}$ in logarithmic scale, further separating noisy genes for which $\nu > \nu_1$ (blue markers in Fig. 5) from those for which $0 < \nu < \nu_1$ (orange markers). The interaction score of the high-noise genes is significantly higher than the score of the intermediate group, which in turn is higher than the low-noise genes' (Mann-Whitney U test, $P < 2.2 \cdot 10^{-16}$).

There is a significant positive correlation between the distance ν and the interaction score ($P < 2.2 \cdot 10^{-16}$, lm), thus showing that the noise level of genes with high interaction score is typically higher than the mean predicts; we also observe a significant negative correlation between the interaction score and the burst frequency k_{on} ($P < 2.2 \cdot 10^{-16}$, lm) and a significant positive correlation between the interaction score and the burst size ($P < 2.2 \cdot 10^{-16}$, lm), consistent with the results on the transgenes. Filtering out zero-count genes, for which there is little statistical information, increases the P values above to $2.0 \cdot 10^{-7}$, $2.0 \cdot 10^{-3}$, and $1.68 \cdot 10^{-5}$, respectively, due to smaller sample sizes, and yields the scatter plots of Fig. 5d, e and the boxplot of Fig. 5f for the three groups. These results agree with those obtained from different ChIA-PET biological repeats and different bin resolutions (1 Kb and 7 Kb; Additional file 1: Section S8 and Figure S16).

Microscopic model

To shed further light on the biological mechanisms involved and test whether PolII shuttling can a priori alter the transcriptional noise as seen in the previous section, we constructed and simulated a more complex stochastic model that captures the most important features of our expression system, i.e. induction, polymerase flux between the LLPS droplet (or, more generically, a cluster of PolII [20]) and the gene, transcription, and decay, while stripping away non-essential details (Fig. 6a). Its precise formulation, along with additional details, is illustrated in the “Materials and methods” section and Additional file 1: Section S9. The model is designed around the idea that each PolII waits in a compartment until the transcription occurs [22], where the compartment represents an LLPS droplet (Fig. 6a). This is immersed in its nuclear environment, which adds and removes PolIIs at rates γ and δ , respectively. In addition to this, by transcribing at rate β , the PolIIs leave the compartment with probability $1 - l$ or are re-injected otherwise. This latter reaction represents the crosstalk between the 3'-end processing and the transcription initiation and helps to sustain the compartment population despite the presence of initiation, which on average contributes to depleting it. Consistently with the two genes integrated in our cell lines, the model encodes a Tet-repressor binding site downstream of the TSS which binds to the TetR factor, present at concentration n . Such a binding event interrupts the transcription; therefore, tuning n allows us to control the blocking rate λ_{off} . The model parameters l and n are akin to the pA mutation and the Tet concentration, respectively, in the experimental settings. We assume that the pA mutation hinders but does not completely block PolII flux back to the compartment (which can also be facilitated by diffusion, see for instance [16, 24]); therefore, the parameter l is assumed to be small but still strictly positive even in the presence of pA mutation. During a TetR blockade, PolIIs cannot transcribe and accumulate in the compartment. When the blockade is released, the transcription occurs at a rate directly proportional to the available PolII (consistently with the law of mass action and experimental observations [22, 73]); therefore, at the end of the TetR blockade, the compartment is highly populated and the transcription occurs repeatedly while the PolII population quickly drops. As the simulation results demonstrate, the model is able to reproduce an increase of transcriptional bursting upon increasing the recycling probability l (Fig. 6). This behaviour is conserved under a broad range of different parameter settings, demonstrating that this is a



generic result of our model. Fitting a negative-binomial distribution with vague prior distributions to an ensemble of mRNA abundances, simulated from this microscopic model, shows patterns consistent with those obtained from the experimental data (Fig. 6c and Additional file 1: Section S9).

While actual transcriptional mechanisms are more complex than our idealised model, the latter provides a significant step towards a mechanistic explanation of our observations. In fact, it captures the essential features of the two gene constructs, and naturally reproduces the observed pattern by tuning only the shuttling probability l and the factor abundance n . Notably, our results demonstrate a minor role for extrinsic contributions to noise (Fig. 6b); in fact, intrinsic factors suffice to yield the noise floor for a wide range of λ_{off} and μ_X , which contrasts with several other studies [69–73].

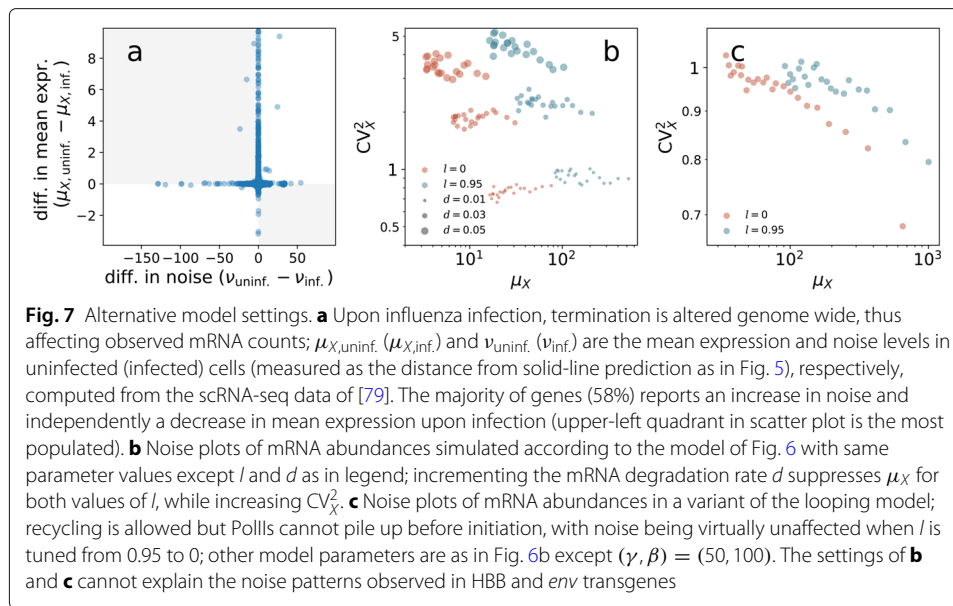
Alternative model settings

The pA site mutations in HBB and *env* transgenes cause termination defects which in turn affect the mRNA degradation rate (Additional file 1: Section S5, and [33]). To establish whether the observed noise patterns are ascribable to this, we considered both single-cell expression data and numerical simulations. We analysed human genes in the publicly available dataset of [79], which includes scRNA-seq UMI count data from both influenza-infected and uninfected human A594 cells. Influenza infection causes termination defects in human genes, where transcription can continue for tens of kilobases after the pA site [80, 81]. Native elongation transcript sequencing (NET-seq) also shows that infected cells do not have a difference in initiation of transcription [81]. As suggested in [79], we assumed that a cell is infected if it has at least 0.02% of transcripts coming from influenza genes after 6 h from virus inoculation; otherwise, it is assumed to be uninfected. We then computed the mean expression levels $\mu_{X,\text{inf.}}$ and $\mu_{X,\text{uninf.}}$ and the noise levels $\nu_{\text{inf.}}$ and $\nu_{\text{uninf.}}$ for all human genes (where the subscripts “inf.” and “uninf.” indicate infected and uninfected conditions, respectively). The presence of the termination defect increases the transcript degradation rate, which lowers the UMI counts; we found indeed that $\mu_{X,\text{inf.}} < \mu_{X,\text{uninf.}}$ for the overwhelming majority of genes. We also found an overall increase in noise with infection, i.e. $\nu_{\text{inf.}} > \nu_{\text{uninf.}}$ for many genes, as illustrated in Fig. 7a. A similar scenario is obtained simulating our model with increasing values of mRNA degradation rate d and with recycling rate l held fixed (Fig. 7b): increasing d lowers the average amount of in silico mRNA and increments its CV^2 . This scenario does not fit the experimental transgene observations, where pA mutation equally lowers μ_X but decreases CV_X^2 , and therefore, it is not a plausible representation of their true biological mechanisms.

Further, we considered a variant of our model where PolIIs are not allowed to condensate in a compartment before the transcription begins. The importance of particle condensation to fluctuations in the presence of an on-off switch has been mathematically described [82]. In the modified in silico model, indeed, increasing the recycling rate does not increase the noise (Fig. 7c), thus suggesting that a reservoir of PolIIs may be a crucial component of gene regulation.

Discussion

The wealth of existing results strongly suggests the occurrence of 3′-5′ crosstalk in the WT variants of our transgene systems, involving physical interaction between factors at either gene end and recycling of polymerases, which can be disrupted or strongly reduced upon a point mutation. Similarly, information of the interactions between the ends of genes involved in transcription can be accessed genome wide by means of PolII ChIA-PET sequencing.



Based on both an in-depth analysis of the transgene systems (which provide a controlled experimental setting) and an observational study of ChIA-PET sequencing data (which provide a genome-wide view of chromatin interactions involved in transcription), we present results to suggest that PolII-mediated 3'-5' interactions are major contributors to transcriptional noise.

Building on standard phenomenological models, transcription parameters, such as average burst size and frequency, are consistently inferred across the different conditions using a Bayesian methodology, to demonstrate the presence of association between 3'-5' interactions and transcription kinetics. Modelling transcription requires abstraction and simplification due to the complexity of the molecular processes involved and the inadequacy of current experimental methodologies to dynamically resolve structural interactions at individual loci. Furthermore, the Bayesian estimates of the kinetic parameters reflect the incomplete quantitative information available on the experimental device. Also note that our transgenes might not exactly represent the average endogenous gene. Nevertheless, our setting is sufficient to resolve specific patterns, which can be reproduced by an ab initio mechanistic model, thus supporting our conclusions.

The analysis suggests that recycling of the polymerase typically increases noise at a given expression level, while an alternative symmetric interpretation is possible, viz., that recycling permits higher expression at a given noise level. These relations either are a byproduct of the construction of the transcriptional machinery or were selected for. It will be interesting to further explore our findings from an evolutionary perspective. In particular, many studies show how selection of noisy expression can be critical by contributing to cell fate diversity [83, 84] and by favouring their long-term survival in adverse environments [85]. This could also have implications in synthetic biology, where the optimisation of gene expression and the control of its noise are desirable features [86, 87]. Our work provides an important contribution to the field of systems biology by identifying a single base, and thus a genetic determinant, that modulates the balance between the average expression level and its variation.

Materials and methods

Measurement equation and Monte Carlo estimation

We assume that the measured fluorescence Y_i of cell i is proportional to the true mRNA abundance X_i and therefore can be expressed as $Y_i^{(k)} = \epsilon_i^{(k)} + \kappa^{(k)} X_i^{(k)}$ where (k) indexes the replicate, $\kappa^{(k)}$ can be thought of as a scale, and $\epsilon_i^{(k)}$ is the zero of such a scale, also corresponding to the background of unspecific staining and auto-fluorescence of the i th cell [59]. The background noise is measured, for each replicate k , by means of control cells whose gene of interest has been deleted. These are used to define informative priors for $\epsilon_i^{(k)}$. Our choice is $\epsilon_i^{(k)} \sim \text{SN}(a^{(k)}, \mu_\epsilon^{(k)}, \sigma_\epsilon^{(k)})$, i.e. the control-cell fluorescence y is supposed to have Azzalini's skew-normal distribution

$$f_\epsilon(y|a^{(k)}, \mu_\epsilon^{(k)}, \sigma_\epsilon^{(k)}) = 2\Phi\left(\frac{(y - \mu_\epsilon^{(k)})}{\sigma_\epsilon^{(k)}} a^{(k)}\right) \phi\left(\frac{y - \mu_\epsilon^{(k)}}{\sigma_\epsilon^{(k)}}\right),$$

where Φ and ϕ are the standard normal CDF and normal PDF, respectively, while the mean $\mu_\epsilon^{(k)}$, the standard deviation $\sigma_\epsilon^{(k)}$, and the skewness parameter $a^{(k)}$ are point estimates from the control datasets. Prior distributions for $\kappa^{(k)}$ are chosen based on the regression coefficients of gamma generalised linear model fits with identity link. For the remaining parameters, we assume vague gamma priors with mean 1 and variance 10^3 . Adaptive Metropolis-Hastings samplers for model fitting were implemented (Additional file 1: Section S4) [88].

Phenomenological two-state gene-expression models

The transcriptional bursting is fully characterised by the rates $\tilde{\alpha}$, \tilde{k}_{on} , and \tilde{k}_{off} in units of min^{-1} . It is convenient to express the rates in units of the inverse of the mean mRNA lifetime \tilde{d} , i.e. $\tilde{k}_{\text{off}} = k_{\text{off}} \tilde{d}$, $\tilde{k}_{\text{on}} = k_{\text{on}} \tilde{d}$, $\tilde{\alpha} = \alpha \tilde{d}$. It can be shown that the stationary mRNA abundance X for this model is Poisson beta with probability density function (PDF)

$$f_X(x|\alpha, k_{\text{on}}, k_{\text{off}}) = \int_0^1 f_{\text{Poi}}(x|\alpha p) f_{\text{Be}}(p|k_{\text{on}}, k_{\text{off}}) dp,$$

where $f_{\text{Poi}}(x|\alpha) = \alpha^x e^{-\alpha}/x!$ and $f_{\text{Be}}(p|k_{\text{on}}, k_{\text{off}}) = p^{k_{\text{on}}-1} (1-p)^{k_{\text{off}}-1} \Gamma(k_{\text{on}} + k_{\text{off}}) / (\Gamma(k_{\text{off}})\Gamma(k_{\text{on}}))^{-1}$ are PDFs of Poisson and beta random variables (RVs), respectively. This expresses the hierarchy

$$X|\alpha, P \sim \text{Poi}(\alpha P), \quad P|k_{\text{on}}, k_{\text{off}} \sim \text{Beta}(k_{\text{on}}, k_{\text{off}}).$$

It is convenient to reparametrise the Poisson-beta PDF in terms of its mean $\mu_X = \alpha k_{\text{on}} / (k_{\text{off}} + k_{\text{on}})$, to get

$$X|\mu_X, k_{\text{on}}, k_{\text{off}}, P \sim \text{Poi}(\mu_X P (k_{\text{off}} + k_{\text{on}}) / k_{\text{on}}), \\ f_X(x|\alpha, k_{\text{on}}, k_{\text{off}}) =: f_X'(x|\mu_X, k_{\text{on}}, k_{\text{off}}).$$

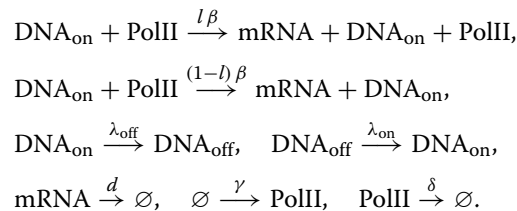
In fact, this allows us to exploit knowledge on μ_X in the form of informative priors and infer the dimensionless rates α , k_{off} , and k_{on} . These are converted to min^{-1} by using \tilde{d} estimated from data (Additional file 1: Section S5). In the limit as $k_{\text{off}} \rightarrow \infty$, $\alpha \rightarrow \infty$, with their ratio α/k_{off} held finite, the population mean satisfies $\mu_X = k_{\text{on}}\alpha/k_{\text{off}}$, while the PDF of X approaches the negative-binomial distribution

$$f_X''(x|k_{\text{on}}, k_{\text{off}}/\alpha) = \int_0^\infty f_{\text{Poi}}(x|\lambda) f_{\text{Gamma}}(\lambda|k_{\text{on}}, k_{\text{off}}/\alpha) d\lambda,$$

where $f_{\text{Gamma}}(x|k_{\text{on}}, k_{\text{off}}/\alpha)$ is the density of a Gamma RV with mean μ_X and variance $\mu_X k_{\text{off}}/\alpha$; when this RV concentrates near the mean as $k_{\text{on}} \rightarrow \infty$ and $k_{\text{off}}/\alpha \rightarrow 0$, X is Poisson with PDF $f_{\text{Poi}}(x|\mu_X)$.

Microscopic model

The microscopic model is defined by means of the following chemical reaction scheme:



By the law of mass action, $\lambda_{\text{off}} = nK_\lambda$, $\lambda_{\text{on}} = K_\lambda$, where K_λ and n represent the chemical affinity and concentration of TetR homodimers that bind to the TetO₂ operators downstream of the TSS, respectively. When such a binding event occurs, the transcription is inhibited as elongation is impeded and the resulting locked DNA configuration is represented by the species DNA_{off}. The switch to DNA_{on} corresponds to the release of the lock. A variant of this model that does not allow PolIII to accumulate before transcription is obtained with $\gamma > 0$ when the PolIII compartment is empty and $\gamma = 0$ otherwise.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-020-02227-5>.

Additional file 1: "Supporting Information to: "3'-5' crosstalk contributes to transcriptional bursting".

Additional file 2: Review history.

Acknowledgements

We thank Louise Dyson, Matt Moores, Lucy Ternent, and Jonathan Keith for valuable discussions, and Sharon Collier and Charlotte Petersen for minor contributions. pSpCas9(BB)-2A-GFP (PX458) was a gift from Feng Zhang (Addgene plasmid #48138; <http://n2t.net/addgene:48138>; RRID:Addgene 48138). We also thank Søren Lykke-Andersen and Torben Heick Jensen for providing cell lines.

Peer review information

Andrew Cosgrove was the primary editor of this manuscript and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 2.

Authors' contributions

M.C., S.T., B.F., and D.H. designed the research; M.C., M.D.W., M.J., J.T., and D.H. performed the research; M.C., M.D.W., M.J., and D.H. analysed the data; M.C., M.D.W., and D.H. wrote the paper with input from all authors. All authors read and approved the final manuscript.

Funding

The research was supported by BBSRC grant BB/L006340/1 and utilised WISB computational and experimental facilities (grant ref: BB/M017982/1) funded under the UK Research Councils' Synthetic Biology for Growth programme.

Availability of data and materials

Custom scripts have been made available at <https://github.com/mcavallaro/gLoop> under GNU-GPLv3.0 licence [89]. Data that support the findings of this study have been deposited in Zenodo [90] and in the National Center for Biotechnology Information Gene Expression Omnibus with accession number GSE124682 [91].

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Life Sciences, University of Warwick, Coventry, UK. ²Mathematics Institute and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK. ³Department of Statistics, University of Warwick, Coventry, UK. ⁴Department of Chemistry, University of Warwick, Coventry, UK. ⁵Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland.

Received: 4 April 2020 Accepted: 8 December 2020

Published online: 04 February 2021

References

1. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell*. 2005;123(6):1025–36. <https://doi.org/10.1016/j.cell.2005.09.031>.
2. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. *Curr Biol*. 2006;16(10):1018–25. <https://doi.org/10.1016/J.CUB.2006.03.092>.
3. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011;332(6028):472–4. <https://doi.org/10.1126/science.1198817>.
4. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006;4(10):309. <https://doi.org/10.1371/journal.pbio.0040309>.
5. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science*. 2012;336(6078):183–7. <https://doi.org/10.1126/science.1216379>.
6. Singh A, Razoooky B, Cox CD, Simpson ML, Weinberger LS. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in hiv-1 gene expression. *Biophys J*. 2010;98(8):32–34. <https://doi.org/10.1016/j.bpj.2010.03.001>.
7. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, Segerstolpe Å, Rivera CM, Ren B, Sandberg R. Genomic encoding of transcriptional burst kinetics. *Nature*. 2019;565(7738):251–4. <https://doi.org/10.1038/s41586-018-0836-1>.
8. Elovitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;297(5584):1183–6. <https://doi.org/10.1126/science.1070919>.
9. Zopf CJ, Quinn K, Zeidman J, Maheshri N. Cell-cycle dependence of transcription dominates noise in gene expression. *PLoS Comput Biol*. 2013;9(7):1003161. <https://doi.org/10.1371/journal.pcbi.1003161>.
10. Sherman MS, Lorenz K, Lanier MH, Cohen BA. Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. *Cell Syst*. 2015;1(5):315–25. <https://doi.org/10.1016/j.cels.2015.10.011>.
11. Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. Single-cell analysis of transcription kinetics across the cell cycle. *eLife*. 2016;5:e12175. <https://doi.org/10.7554/eLife.12175>.
12. Padovan-Merhar O, Nair G, Biaesch A, Mayer A, Scarfone S, Foley S, Wu A, Churchman LS, Singh A, Raj A. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell*. 2015;58(2):339–52. <https://doi.org/10.1016/J.MOLCEL.2015.03.005>.
13. Battich N, Stoeger T, Pelkmans L. Control of transcript variability in single mammalian cells. *Cell*. 2015;163(7):1596–610. <https://doi.org/10.1016/j.cell.2015.11.018>.
14. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. *Science*. 2004;304(5678):1811–4. <https://doi.org/10.1126/science.1098641>.
15. Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I, Barkai N. Expression noise and acetylation profiles distinguish HDAC functions. *Mol Cell*. 2012;47(2):193–202. <https://doi.org/10.1016/J.MOLCEL.2012.05.008>.
16. van Zon JS, Morelli MJ, Tanase-Nicola S, ten Wolde PR. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys J*. 2006;91(12):4350–67. <https://doi.org/10.1529/BIOPHYSJ.106.086157>.
17. Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell*. 2014;158(2):314–26. <https://doi.org/10.1016/j.cell.2014.05.038>.
18. Fukaya T, Lim B, Levine M. Enhancer control of transcriptional bursting. *Cell*. 2016;166(2):358–68. <https://doi.org/10.1016/j.cell.2016.05.025>.
19. Bartman C, Hsu S, Hsiung C-S, Raj A, Blobel G. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol Cell*. 2016;62(2):237–47. <https://doi.org/10.1016/J.MOLCEL.2016.03.007>.
20. Papanonis A, Cook PR. Transcription factories: Genome organization and gene regulation. *Chem Rev*. 2013;113(11):8683–705. <https://doi.org/10.1021/cr300513p>.
21. Cisse II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, Dugast-Darzacq C, Hajj B, Dahan M, Darzacq X. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science*. 2013;341(6146):664–7. <https://doi.org/10.1126/science.1239053>.
22. Cho WK, Jayanth N, English BP, Inoue T, Andrews JO, Conway W, Grimm JB, Spille JH, Lavis LD, Lionnet T, Cisse II. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife*. 2016;5:e13617. <https://doi.org/10.7554/eLife.13617>.
23. Plys AJ, Kingston RE. Dynamic condensates activate transcription. *Science*. 2018;361(6400):329–30. <https://doi.org/10.1126/science.aau4795>.
24. Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, McSwiggen DT, Kokic G, Dailey GM, Cramer P, Darzacq X, Zweckstetter M. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol*. 2018;25(9):833–40. <https://doi.org/10.1038/s41594-018-0112-y>.
25. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, Tjian R. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*. 2018;361(6400):2555. <https://doi.org/10.1126/science.aar2555>.
26. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, Sharp PA,

- Chakraborty AK, Young RA. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 2018;361(6400):3958. <https://doi.org/10.1126/science.aap9195>.
27. Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, Cisse II. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. 2018;361(6400):412–5. <https://doi.org/10.1126/science.aar4199>.
 28. Cramer P. Organization and regulation of gene transcription. *Nature*. 2019;573(7772):45–54. <https://doi.org/10.1038/s41586-019-1517-4>.
 29. Shin Y, Chang YC, Lee DSW, Berry J, Sanders DW, Ronceray P, Wingreen NS, Haataja M, Brangwynne CP. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell*. 2018;175(6):1481–9113. <https://doi.org/10.1016/j.cell.2018.10.057>.
 30. Wei MT, Chang YC, Shimobayashi SF, Shin Y, Strom AR, Brangwynne CP. Nucleated transcriptional condensates amplify gene expression. *Nat Cell Biol*. 2020;22(10):1187–96. <https://doi.org/10.1038/s41556-020-00578-6>.
 31. Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*. 2009;136(4):688–700. <https://doi.org/10.1016/J.CELL.2009.02.001>.
 32. Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol*. 2011;12(5):283–94. <https://doi.org/10.1038/nrm3098>.
 33. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell*. 2010;40(3):410–22. <https://doi.org/10.1016/j.molcel.2010.10.012>.
 34. Fang X, Wang L, Ishikawa R, Li Y, Fiedler M, Liu F, Calder G, Rowan B, Weigel D, Li P, Dean C. Arabidopsis FLL2 promotes liquid–liquid phase separation of polyadenylation complexes. *Nature*. 2019;569(7755):265–9. <https://doi.org/10.1038/s41586-019-1165-8>.
 35. Hebenstreit D. Are gene loops the cause of transcriptional noise? *Trends Genet*. 2013;29(6):333–8. <https://doi.org/10.1016/j.tig.2013.04.001>.
 36. Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J. A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*. 2008;29(2):271–8. <https://doi.org/10.1016/J.MOLCEL.2007.11.035>.
 37. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. *Nat Genet*. 2002;32(4):623–6. <https://doi.org/10.1038/ng1051>.
 38. Reik A, Telling A, Zitnik G, Cimbara D, Epner E, Groudine M. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol*. 1998;18(10):5992–6000. <https://doi.org/10.1128/mcb.18.10.5992>.
 39. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*. 2002;10(6):1453–65. [https://doi.org/10.1016/s1097-2765\(02\)00781-5](https://doi.org/10.1016/s1097-2765(02)00781-5).
 40. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A phase separation model for transcriptional control. *Cell*. 2017;169(1):13–23. <https://doi.org/10.1016/j.cell.2017.02.007>.
 41. Wang Y, Fairley JA, Roberts SGE. Phosphorylation of TFIIB links transcription initiation and termination. *Curr Biol CB*. 2010;20(6):548–53. <https://doi.org/10.1016/j.cub.2010.01.052>.
 42. Ansari A, Hampsey M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*. 2005;19(24):2969–78. <https://doi.org/10.1101/gad.1362305>.
 43. Wani S, Yuda M, Fujiwara Y, Yamamoto M, Harada F, Ohkuma Y, Hirose Y. Vertebrate Ssu72 regulates and coordinates 3'-end formation of RNAs transcribed by RNA polymerase II. *PLoS ONE*. 2014;9(8):106040. <https://doi.org/10.1371/journal.pone.0106040>.
 44. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. Gene loops enhance transcriptional directionality. *Science*. 2012;338(6107):671–5. <https://doi.org/10.1126/science.1224350>.
 45. Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ. Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Mol Cell*. 2008;29(1):56–68. <https://doi.org/10.1016/j.molcel.2007.11.030>.
 46. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>.
 47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
 48. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
 49. Dyer NP, Shahrezaei V, Hebenstreit D. LiBiNorm: an htseq-count analogue with improved normalisation of Smart-seq2 data and library preparation diagnostics. *PeerJ*. 2019;7:6222. <https://doi.org/10.7717/peerj.6222>.
 50. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
 51. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):187–91. <https://doi.org/10.1093/nar/gku365>.
 52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
 53. Goldstein LD, Cao Y, Pau G, Lawrence M, Wu TD, Seshagiri S, Gentleman R. Prediction and quantification of splice events from RNA-Seq data. *PLOS ONE*. 2016;11(5):0156132. <https://doi.org/10.1371/journal.pone.0156132>.
 54. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res*. 2019;47(8):47. <https://doi.org/10.1093/nar/gkz114>.
 55. Shapiro HM. Practical flow cytometry. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2003. <https://doi.org/10.1002/0471722731>.
 56. Lo K, Brinkman RR, Gottardo R. Automated gating of flow cytometry data via robust model-based clustering. *Cytom Part A*. 2008;73A(4):321–32. <https://doi.org/10.1002/cyto.a.20531>.

57. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics*. 2009;10:106. <https://doi.org/10.1186/1471-2105-10-106>.
58. Lo K, Hahne F, Brinkman RR, Gottardo R. flowClust: a Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics*. 2009;10(1):145. <https://doi.org/10.1186/1471-2105-10-145>.
59. Tiberi S, Walsh M, Cavallaro M, Hebenstreit D, Finkenzstädt B. Bayesian inference on stochastic gene transcription from flow cytometry data. *Bioinformatics*. 2018;34(17):647–55. <https://doi.org/10.1093/bioinformatics/bty568>.
60. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281–308. <https://doi.org/10.1038/nprot.2013.143>.
61. Kim J, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol*. 2013;14(1):7. <https://doi.org/10.1186/gb-2013-14-1-r7>.
62. Dobrzynski M, Bruggeman FJ. Elongation dynamics shape bursty transcription and translation. *Proc Natl Acad Sci*. 2009;106(8):2583–8. <https://doi.org/10.1073/pnas.0803507106>.
63. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*. 2008;15(12):1263–71. <https://doi.org/10.1038/nsmb.1514>.
64. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161(5):1187–201. <https://doi.org/10.1016/j.cell.2015.04.044>.
65. Spiess A. qpcR: modelling and analysis of real-time PCR data. 2013. <https://CRAN.R-project.org/package=qpcR>. Accessed 2018.
66. Dar RD, Shaffer SM, Singh A, Razooky BS, Simpson ML, Raj A, Weinberger LS. Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. *PLoS ONE*. 2016;11(7):0158298. <https://doi.org/10.1371/journal.pone.0158298>.
67. Dar RD, Razooky BS, Weinberger LS, Cox CD, Simpson ML. The low noise limit in gene expression. *PLoS ONE*. 2015;10(10):0140969. <https://doi.org/10.1371/journal.pone.0140969>.
68. Soltani M, Vargas-Garcia CA, Antunes D, Singh A. Inter-cellular variability in protein levels from stochastic expression and noisy cell cycle processes. *PLoS Comput Biol*. 2016;12(8):1004972. <https://doi.org/10.1371/journal.pcbi.1004972>.
69. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O’Shea E, Pilpel Y, Barkai N. Noise in protein expression scales with natural protein abundance. *Nat Genet*. 2006;38(6):636–43. <https://doi.org/10.1038/ng1807>.
70. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006;441(7095):840–6. <https://doi.org/10.1038/nature04785>.
71. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 2010;329(5991):533–8. <https://doi.org/10.1126/science.1188308>.
72. Stewart-Ornstein J, Weissman JS, El-Samad H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol Cell*. 2012;45(4):483–93. <https://doi.org/10.1016/j.molcel.2011.11.035>.
73. Yang S, Kim S, Rim Lim Y, Kim C, An HJ, Kim J-H, Sung J, Lee NK. Contribution of RNA polymerase concentration variation to protein expression noise. *Nat Commun*. 2014;5(1):4761. <https://doi.org/10.1038/ncomms5761>.
74. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei C-L, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung W-K, Snyder M, Ruan Y. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 2012;148(1–2):84–98. <https://doi.org/10.1016/j.cell.2011.12.014>.
75. Li G, Chen Y, Snyder MP, Zhang MQ. ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res*. 2017;45(1):4. <https://doi.org/10.1093/nar/gkw809>.
76. Bartlett MS. The use of transformations. *Biometrics*. 1947;3(1):39–52. <https://doi.org/10.2307/3001536>.
77. Svensson V. Droplet scRNA-seq is not zero-inflated. *Nat Biotechnol*. 2020;38(2):147–50. <https://doi.org/10.1038/s41587-019-0379-5>.
78. Townes FW, Hicks SC, Aryee MJ, Irizarry RA. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol*. 2019;20(1):295. <https://doi.org/10.1186/s13059-019-1861-6>.
79. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. *eLife*. 2018;7:e32303. <https://doi.org/10.7554/eLife.32303>.
80. Zhao N, Sebastiano V, Moshkina N, Mena N, Hultquist J, Jimenez-Morales D, Ma Y, Rialdi A, Albrecht R, Fenouil R, Sánchez-Aparicio MT, Ayllon J, Ravisankar S, Haddad B, Ho JSY, Low D, Jin J, Yurchenko V, Prinjha RK, Tarakhovskiy A, Squatrito M, Pinto D, Allette K, Byun M, Smith ML, Sebra R, Guccione E, Tumpey T, Krogan N, Greenbaum B, van Bakel H, García-Sastre A, Marazzi I. Influenza virus infection causes global RNAPII termination defects. *Nat Struct Mol Biol*. 2018;25(9):885–93. <https://doi.org/10.1038/s41594-018-0124-7>.
81. Bauer DLV, Tellier M, Martínez-Alonso M, Nojima T, Proudfoot NJ, Murphy S, Fodor E. Influenza virus mounts a two-pronged attack on host RNA polymerase II transcription. *Cell Rep*. 2018;23(7):2119–29. <https://doi.org/10.1016/j.celrep.2018.04.047>.
82. Cavallaro M, Mondragón RJ, Harris RJ. Temporally correlated zero-range process with open boundaries: steady state and fluctuations. *Phys Rev E*. 2015;92(2):022137. <https://doi.org/10.1103/PhysRevE.92.022137>.
83. Kussell E, Leibler S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. 2005;309(5743):2075–8. <https://doi.org/10.1126/science.1114383>.
84. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010;467(7312):167–73. <https://doi.org/10.1038/nature09326>.
85. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. *Nature*. 2009;462(7269):90–93. <https://doi.org/10.1038/nature08504>.
86. Murphy KF, Adams RM, Wang X, Balázs G, Collins JJ. Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res*. 2010;38(8):2712–26. <https://doi.org/10.1093/nar/gkq091>.
87. Bandiera L, Furini S, Giordano E. Phenotypic variability in synthetic biology applications: dealing with noise in microbial gene expression. *Front Microbiol*. 2016;7:479. <https://doi.org/10.3389/fmicb.2016.00479>.
88. Patil A, Huard D, Fannesbeck C. PyMC: Bayesian stochastic modelling in Python. *J Stat Softw*. 2010;35(4):1–81. <https://doi.org/10.18637/jss.v035.i04>.

89. Cavallaro M, Walsh MD, Jones M, Teahan J, Tiberi S, Finkenstädt B, Hebenstreit D. Supporting software to "3'-5' crosstalk contributes to transcriptional bursting". GitHub. 2020. <https://doi.org/10.5281/zenodo.4127058>.
90. Cavallaro M, Walsh MD, Jones M, Teahan J, Tiberi S, Finkenstädt B, Hebenstreit D. Supporting data to "3'-5' crosstalk contributes to transcriptional bursting". Zenodo. 2020. <https://doi.org/10.5281/zenodo.4304833>.
91. Cavallaro M, Walsh MD, Jones M, Teahan J, Tiberi S, Finkenstädt B, Hebenstreit D. Supporting RNA-seq data to "3'-5' crosstalk contributes to transcriptional bursting". Gene Expression Omnibus. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124682>. Accessed 2019.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

