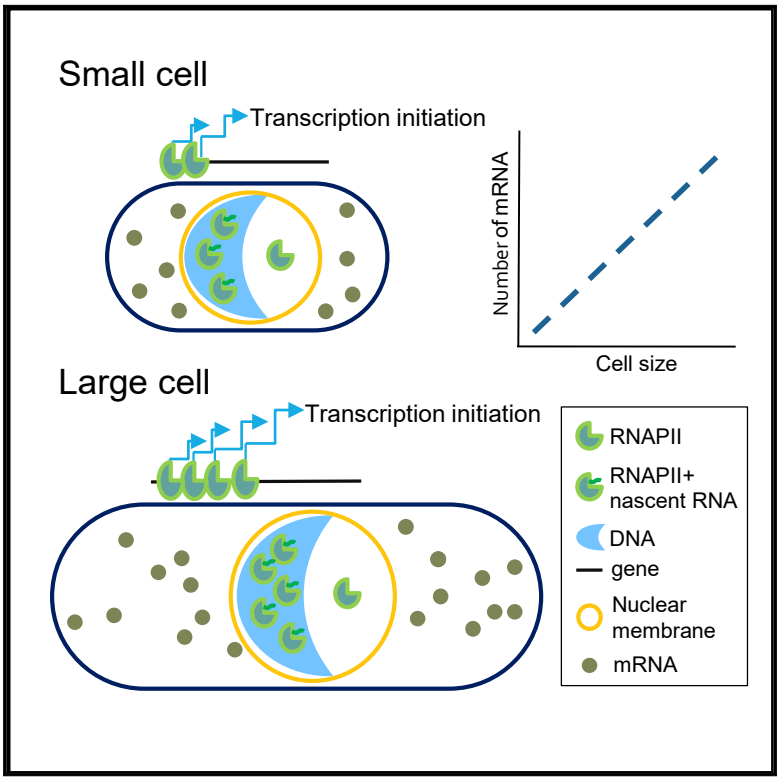
Article

Current Biology

Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size

Graphical Abstract



Highlights

- Numbers of constitutive and inducible mRNAs scale with cell size
- Coordination of RNAPII initiation rates with cell size underpins scaling
- Amounts of DNA-bound RNAPII increase with cell size and are limiting
- Transcription of constitutive and periodic mRNAs is a non-bursty Poisson process

Authors

Xi-Ming Sun, Anthony Bowman, Miles Priestman, ..., Dirk Dormann, Vahid Shahrezaei, Samuel Marguerat

Correspondence

v.shahrezaei@imperial.ac.uk (V.S.), samuel.marguerat@imperial.ac.uk (S.M.)

In Brief

Messenger RNA numbers scale with cell size to maintain constant concentrations. Using experimental approaches and mathematical models, Sun et al. show that scaling results from the coordination of RNAPII initiation rates with cell size. They propose a model where scaling is the consequence of competition between genes for limiting RNAPII.

Sun et al., 2020, Current Biology *30*, 1–14
April 6, 2020
https://doi.org/10.1016/j.cub.2020.01.053

CellPress

Please cite this article in press as: Sun et al., Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.01.053

Current Biology
Article

Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size

Xi-Ming Sun,[1,2,4] Anthony Bowman,[3,4] Miles Priestman,[1,2,3] Francois Bertaux,[1,2,3,5,6] Amalia Martinez-Segura,[1,2] Wenhao Tang,[3] Chad Whilding,[1,2] Dirk Dormann,[1,2] Vahid Shahrezaei,[3,*] and Samuel Marguerat[1,2,7,*]
[1]MRC London Institute of Medical Sciences (LMS), Du Cane Road, London W12 0NN, UK
[2]Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK
[3]Department of Mathematics, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, UK
[4]These authors contributed equally
[5]Present address: Institut Pasteur, USR 3756 IP CNRS, 28 rue du Docteur-Roux, 75015 Paris, France
[6]Present address: Inria Paris, 2 rue Simone Iff, 75012 Paris, France
[7]Lead Contact
*Correspondence: v.shahrezaei@imperial.ac.uk (V.S.), samuel.marguerat@imperial.ac.uk (S.M.)
https://doi.org/10.1016/j.cub.2020.01.053

SUMMARY

Cell size varies during the cell cycle and in response to external stimuli. This requires the tight coordination, or "scaling," of mRNA and protein quantities with the cell volume in order to maintain biomolecule concentrations and cell density. Evidence in cell populations and single cells indicates that scaling relies on the coordination of mRNA transcription rates with cell size. Here, we use a combination of single-molecule fluorescence *in situ* hybridization (smFISH), time-lapse microscopy, and mathematical modeling in single fission yeast cells to uncover the precise molecular mechanisms that control transcription rates scaling with cell size. Linear scaling of mRNA quantities is apparent in single fission yeast cells during a normal cell cycle. Transcription of both constitutive and periodic genes is a Poisson process with transcription rates scaling with cell size and without evidence for transcriptional off states. Modeling and experimental data indicate that scaling relies on the coordination of RNA polymerase II (RNAPII) transcription initiation rates with cell size and that RNAPII is a limiting factor. We show using real-time quantitative imaging that size increase is accompanied by a rapid concentration-independent recruitment of RNAPII onto chromatin. Finally, we find that, in multinucleated cells, scaling is set at the level of single nuclei and not the entire cell, making the nucleus a determinant of scaling. Integrating our observations in a mechanistic model of RNAPII-mediated transcription, we propose that scaling of gene expression with cell size is the consequence of competition between genes for limiting RNAPII.

INTRODUCTION

Gene expression is coordinated with cell size in order to maintain biomolecule concentrations. Understanding the mechanisms that mediate this coordination, hereafter called "scaling," is a fundamental and intriguing problem in cell biology [1, 2]. Messenger RNAs (mRNA) and proteins are synthesized from the cell DNA genome, which is one of the few cellular components that do not scale with size. Because cell volume increases during the cell cycle and mRNA half-lives are typically short [3], constant rates of mRNA or protein production cannot lead to gene expression scaling. Recent work in yeast [4], animal [5, 6], and plant cells [7] has shown that mRNA synthesis rates instead are coordinated globally with cell size and are a major mechanism of scaling. Conversely, mRNA degradation seems to be mostly unconnected to scaling [4, 5, 7], although evidence suggests that degradation rates are adjusted early after budding yeast asymmetric division [8] and when growth rate changes [9, 10]. Scaling is pervasive, and only few mRNAs deviate from its regulation [11–13]. Interestingly, two of them participate in the control of size homeostasis [11, 12].

What could be the molecular mechanism behind transcription scaling? For a gene with an active promoter, mRNA numbers follow a Poisson distribution [14]. Transcription is however often discontinuous, and periods of RNA synthesis or "bursts" alternate with periods of promoter inactivity (off state) [15]. Work in single mammalian cells has shown that scaling of mRNA numbers results from a coordination of the size of the transcription bursts with cell volume and not from their frequency [5]. Therefore, the mechanism behind scaling may not be related to activation of transcription but rather to properties of active promoters. Critically, transcription is a complex process and is regulated at many levels, including RNA polymerase II (RNAPII) initiation, pause/release, elongation, and termination [16–19]. Which of these processes is coordinated with cell size to mediate scaling remains unclear.

How could a complex set of molecular reactions, such as transcription, become more efficient as cell size increases? In an elegant experiment, Padovan-Merhar and colleagues fused cells

of small and large size and found that the number of mRNAs produced from a gene encoded in the small cell genome increased in response to the large cell environment. However, it only reached about half the concentration observed in the small cell. This suggests that scaling responds to both cell volume and DNA content [5]. This is consistent with the observation that the cell synthesis capacity is split between genome copies in diploid budding yeast cells [11]. Importantly, changes in gene dosage, in the case of increased ploidy, are associated with overall cell size increase in many organisms [1]. This indicates that the number of genome copies present in a cell is linked to its volume and number of macromolecules. This also suggests that the cell's overall synthetic capacity could be limiting and have a determining role in setting its size [20]. It is therefore likely that scaling depends on a limiting factor involved in transcription, but its identity and regulation with the cell volume is not known [20, 21].

In this study, we measured gene expression in over 20,000 single cells of the fission yeast *Schizosaccharomyces pombe* by single-molecule fluorescence *in situ* hybridization (smFISH) (Tables S1, S2, and S3) [22]. We combined these data with agent-based models of growing and dividing cells [23–25], stochastic models of gene expression [14, 26, 27], and Bayesian inference [28–34] to investigate the quantitative parameters of gene expression that mediate scaling. This integrative approach enabled us to determine which step of the transcription process is scaling with cell size and which molecular event connects transcription with the cell volume.

RESULTS

Gene Expression Scaling Is an Attribute of Constitutive and Inducible Gene Expression

We first tested whether scaling of gene expression with cell size is an attribute of single fission yeast cells during rapid proliferation. To do this, we measured mRNA levels of 7 constitutively expressed genes in wild-type haploids (WT) and in conditional mutants of the Wee1 (*wee1-50*) and Cdc25 (*cdc25-22*) cell cycle regulators by smFISH (Figure 1A; Table S1). At semi-permissive temperature, mutant cells divide at smaller and larger sizes than WT, respectively (Figure S1A) [4]. Consistent with population microarray data, mean numbers of the 7 mRNAs measured in single cells scale with the average size of the three strains (Figure 1B) [4].

We then wondered whether scaling could be detected as cells elongate during a normal cell cycle. The fission yeast elongation phase is restricted to G2. Confounding effects arising from changes in genome content during DNA replication are therefore unlikely. Using cell length as a measure of size (STAR Methods), we analyzed *rpb1* mRNA expression relative to cell size in WT, mutant, and diploid cells and observed linear scaling within each genotype across a wide range of sizes (Figures 1C and 1F). Linear scaling during the cell cycle was apparent for all the constitutive genes analyzed in this study (Figures S1B and S1C). This indicates that gene expression scaling is not an artifact of mutations in the *wee1* and *cdc25* cell-cycle regulators. Moreover, the mutant data demonstrate that gene expression scales over a dynamic range of sizes larger than that observed in WT cells. Together with earlier studies in cell populations, this analysis establishes fission yeast as a powerful model for studying scaling [4].

We next analyzed the relation of scaling with ploidy using diploid cells. Diploids follow scaling parameters (slope and intercept) similar to smaller haploids (Figures 1C and 1F). Interestingly, for a given size, mRNA numbers in diploids and in *cdc25-22* haploids were similar (Figures S1D–S1F). Moreover, when analyzing 3 diploid strains bearing heterozygous deletions of single genes, we observed significant linear scaling of the corresponding mRNA but decreased concentrations compared to both diploid and *cdc25-22* cells (Figures S1D–S1F). This suggests that scaling of single gene copies in fission yeast is coordinated with the cell genome content and that the machinery behind scaling may be limiting.

At the level of single genes, is scaling a property of constitutive expression or do genomes of larger cells have a globally higher gene expression capacity? To answer this question, we analyzed three mRNAs induced during specific phases of the cell cycle (Figure S1G). All of them showed stronger induction levels in larger cells when comparing WT, *wee1-50*, and *cdc25-22* mutants (Figures 1D, 1G, S1I, and S1J). This indicates that scaling is not restricted to constitutive genes and extends to inducible transcripts, which were not expressed in smaller cells at the beginning of the cycle.

We next analyzed the induction of two genes that respond to acute changes in external conditions. The *urg1* mRNA responds to changes in uracil concentration, while *sib1* is induced by 2,2-dipyridyl (DIP) [35, 36]. Induction of *urg1* was heterogeneous and scaled across size mutants (Figures 1E, 1H, and S1H). The *sib1* mRNA showed a much more homogeneous response scaled within and across cell types (Figure S1K). This is evidence for scaling occurring in response to unexpected changes in external conditions. Taken together, these data indicate that scaling is universal and does not depend on the mode of gene regulation. Scaling does not result from a passive accumulation of mRNA during the cell cycle but requires a change in the cell gene expression capacity that is coordinated with cell size.

Coupling of mRNA Decay Rates with Cell Size Is Not a Mechanism of Scaling

mRNA quantities are regulated at the level of transcription but also post-transcriptionally through modulation of degradation rates [37]. Three studies have reported that mRNA degradation rates are not regulated with cell size in fission yeast, plant, and mammals [4, 5, 7]. To test these observations in single fission yeast cells, we analyzed expression of 3 genes by smFISH in WT, *wee1-50*, and *cdc25-22* cells treated with thiolutin, which inhibits *S. pombe* transcription efficiently (Figure S2A) [12]. We derived mRNA half-lives within each strain using cells binned by size (Figure S2B). We observed mRNA half-lives of around 30–40 min for the *rpb1* and *rpb2* mRNAs consistent with previous observations [3]. In *wee1-50* and *cdc25-22* mutants, both mRNAs showed half-lives similar to WT, consistent with an absence of scaling of mRNA degradations rates. In addition, within strains, this analysis did not show consistent positive or negative coordination of degradation rates with cell size, although one dataset showed a significant decrease in stability in *cdc25-22* cells (Figure S2B, left). The absence of consistent scaling of mRNA degradation rates was further confirmed using

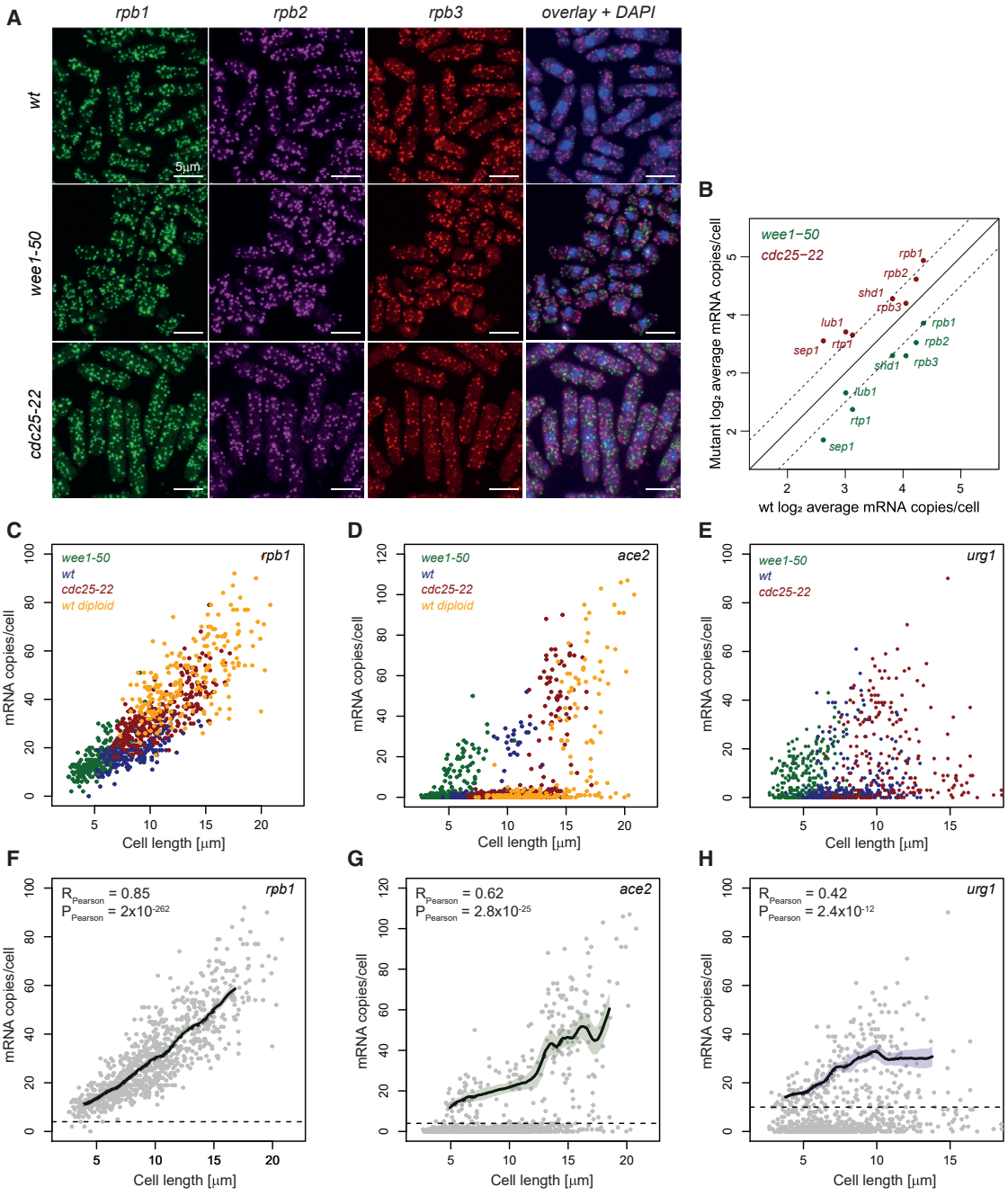


Figure 1. Gene Expression Scaling Is an Attribute of Constitutive and Inducible Gene Expression

(A) smFISH images of *rpb1*, *rpb2*, and *rpb3* mRNA in *wee1-50*, *cdc25-22*, and WT cells and of all channels overlaid with DAPI (last column). White scale bars, 5 μm.

(B) Mean expression of 7 mRNAs in *wee1-50* (green) and *cdc25-22* (red) cells plotted against their expression in WT cells. Lines show equality (plain) and a 2× up or down interval (dotted).

(C–E) Expression of the *rpb1* (C), *ace2* (D), and *urg1* (E) genes plotted against cell length for *wee1-50* (green), WT (blue), *cdc25-22* (red), and WT diploid (orange) cells. Cells in (E) were analyzed 3 h after uracil addition.

(F–H) Same as (C)–(E), respectively, with all cells in gray. Median counts in running windows sampled from 100 experimental data bootstrap samples (line) and 95% confidence intervals (shaded area) are shown. Cells expressing each mRNA above the dotted line cutoff are included in the running window. Pearson correlation coefficients and p values are shown.

See also Figure S1 and Tables S1, S2, and S3.

an orthogonal promoter switch-off approach for the *rpb1*, *rpb2*, and *rtp1* genes, where only *rpb2* showed a slight significant change in half-life (Figure S2B, right). Finally, as discussed in the next section, mathematical modeling and inference do not support scaling of degradation rate. Overall, in agreement with previous studies, our analysis indicates that mRNA degradation is not a major mediator of scaling.

Coupling of Transcription Rates to Cell Size and Not Burst Frequency Mediates Scaling

We next explored the contribution of transcription rates to scaling using smFISH and mathematical modeling. This combined approach allowed us to study dynamic transcription rates based on static smFISH measurements in single cells. We developed agent-based models that incorporate the two-state model of gene expression inside growing and dividing cells, which are themselves described by phenomenological models of cell growth and size control (Figure 2A; STAR Methods). We used an approximate Bayesian computation (ABC) inference approach on cell size and smFISH measurements to infer mechanism of scaling by competing models with different assumptions. This inference approach determines the simplest model that captures the statistics of the experimental measurements and returns posteriors of the parameters and model probabilities for all models (STAR Methods). We used 2 classes of gene expression models that are the limiting cases of the two-state model (Figure 2A). The first class describes the switching of transcription between off and on states explicitly (called hereafter bursting) and comprises three models: one assuming constant transcription rates (v) (burst size constant); one assuming transcription rates (v) coupled to cell size (burst size scaling); and one assuming burst frequency (k_{on}) coupled to cell size (burst frequency scaling). The second class assumes no transcriptional off state, resulting in a simple birth-death process that produces Poisson distributions. It comprises two models, the first assuming constant transcription rates (v) (Poisson constant) and the second assuming transcription rates (v) linearly coupled to cell size (Poisson scaling; Figure 2A; Tables 1 and S4; STAR Methods).

We performed model selection between these five models on the smFISH WT data for 7 constitutively expressed genes. This analysis generated two clear conclusions. First, models assuming bursty transcription were strongly penalized and supported by small-model probabilities (Figure 2B, compare proportion of blue and red colors). This was the case even when burst frequency parameters (k_{on}) were coupled to cell size. Second, among the Poisson models, the one assuming transcription scaling with size was preferred for all but one mRNA (*rpb3* in Figure 2B, compare orange and dark red colors). From this, we conclude that transcription of constitutively expressed genes in fission yeast is Poissonian (non-bursty) and scales with cell size.

We then asked whether transcription regulated during the cell cycle or by external cues followed the same paradigm. We found that, as for constitutively expressed genes, the mRNA distributions of cell-cycle-regulated genes can be explained by cell-cycle-dependent Poissonian transcription (Figure 2B, middle). This indicates that transcription, even when regulated in defined sections of the cell cycle falls into the Poisson regime (non-bursty). In terms of scaling, model selection is in overall support of scaling of transcription rate with cell size as it is for constitutive genes (Figure 2B, middle). mRNAs that respond to external stimuli showed a different picture. Model selection favored the bursting models but was inconclusive on whether burst size or burst frequency was coupled to cell size (Figure 2B, right). This could point to bursty transcription of inducible genes, to the presence of strong extrinsic noise, or to a scenario where cells respond heterogeneously to external signals.

In order to validate the model selection analysis, we generated simulations for all constitutive and cell-cycle-regulated genes from Figure 2B, using models where transcription rates scale with cell size and follow a Poisson regime. This analysis reproduced quantitatively the mean and coefficient of variation (CV) of all experimental measurements (Figure 2C). Notably, our experimental data showed noise levels identical to those produced by a simple birth-death process of mRNA in growing and dividing cells (compare red dots and gray Poisson line). This indicates that their low noise levels do not show the deviation from the Poisson distribution typical of bursty transcription dynamics, supporting our model selection analysis. The cell-cycle-regulated mRNA (*ace2*, *fkh2*, and *mid2*) deviate from the Poisson line as their expression is restricted to specific phases of the cell cycle (Figures 1D, 1G, S1I, and S1J). Importantly, this additional source of noise had been incorporated in the model selection analysis (STAR Methods).

We then used the inferred parameters from the Poisson models to simulate expression of multiple genes in single cells and compared them to smFISH data where 3 different mRNAs were measured per cell (STAR Methods). This analysis suggests that most gene-to-gene expression correlations are explained by scaling of transcription rates with cell size and may not require extensive additional regulation (Figure 2D). However, our data also suggest that, for some gene pairs (e.g., *rpb1-rpb2*), correlations can be explained better by including additional extrinsic variability.

We finally asked whether our modeling approach was in support of a negligible role of mRNA degradation in scaling. We performed model selection and simulation analyses on the transcription shutoff experiment from Figure S2 using models that consider either scaling of transcription rates or degradation rates (STAR Methods; Figure 2E). This analysis shows that the model assuming transcriptional scaling with constant rates of mRNA degradation with cell size is overwhelmingly chosen (Figure 2E), in line with our analysis in Figure S2B.

In summary, we show that scaling in a rapidly growing unicellular organism occurs through coordination of transcriptional rates with cell size as in metazoans. However, we find that transcription during the fission yeast rapid cell cycle is mainly Poissonian except during acute response to external changes, where we detect signs of bursty expression. This is consistent with previous observation in the budding yeast *Saccharomyces cerevisiae* [38]. Finally, we find that scaling of transcription rates explains most of the expression correlation of multiple mRNAs.

Coordination of RNAPII Initiation Rates with Cell Size as the Main Mechanism of Scaling

We then wondered which specific aspect of the transcription process is coordinated with cell size to mediate scaling. To

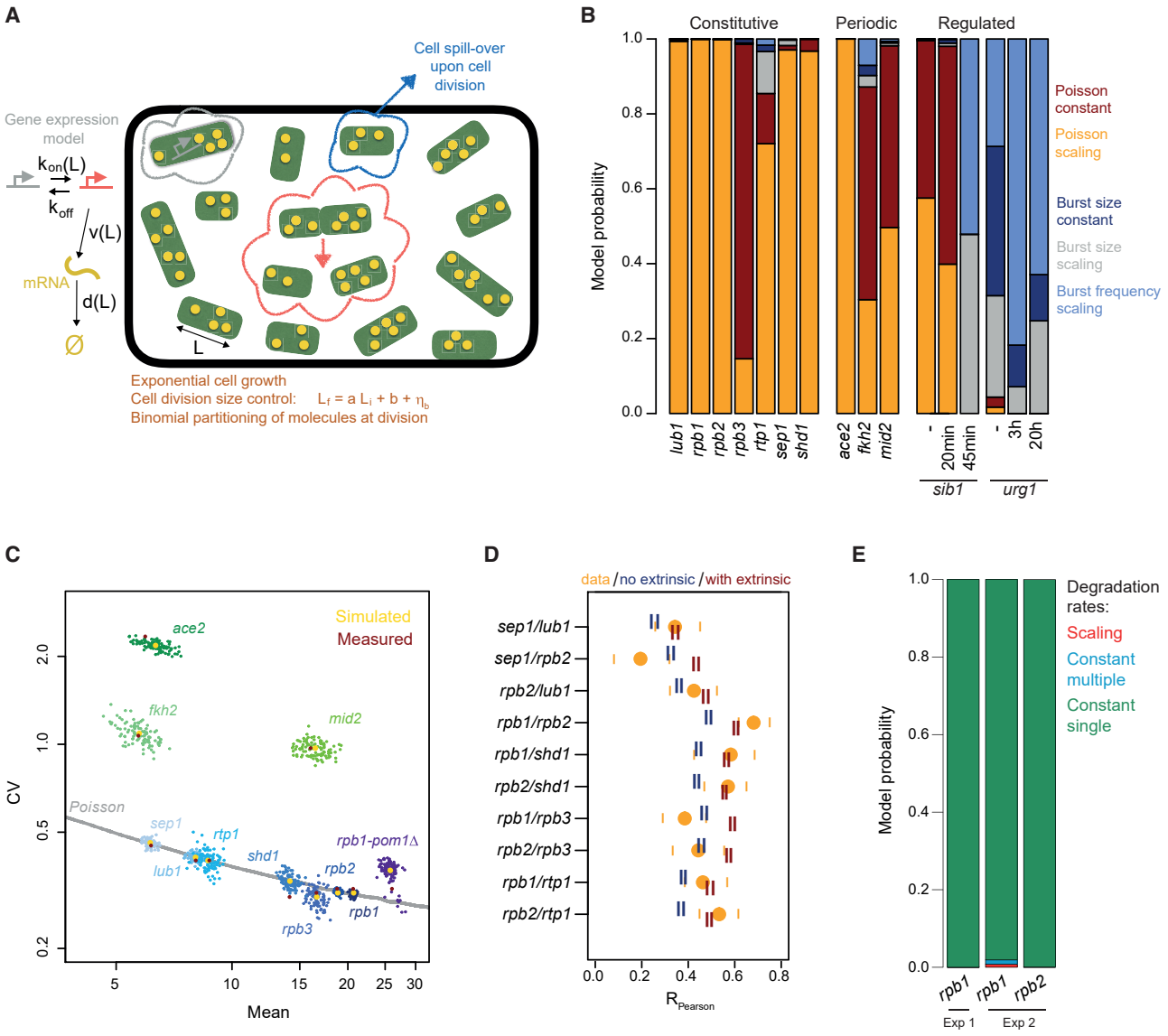


Figure 2. Coordination of Transcription Rates to Cell Size and Not Burst Frequency or mRNA Decay Rates Mediates Scaling

(A) Cartoon of the combined modeling strategy. Left: two-state model of gene expression is shown. The transcription rate (v), burst frequency (k_{on}), or degradation rate (d) can be coupled to cell length (L). Right: cell cycle modeling with noisy linear maps is shown. a/b, length parameters; Lf, division length; Li, birth length; η_b, noise term. Rectangles: cells/agents; yellow dots: mRNAs. Population size is kept constant by removing a random cell (blue arrow) after every division event (red arrow).

(B) Model probabilities for five models of transcription scaling after ABC model selection (key on the right). Results for genes constitutively expressed (constitutive), cell cycle regulated (periodic), or regulated by external stimuli (regulated) are shown. For *sib1* and *urg1*, times after addition of 2,2-dipyridyl or uracil, respectively, are indicated.

(C) Simulation of mRNA numbers during the cell cycle using a model where transcription rates scale under a Poisson regime. Simulated mRNA numbers' coefficients of variation (CVs) and means are plotted for each particle. Yellow circles: median of simulated data for all particles; red circles: median of experimental data used for parameter inference; gray line (Poisson): mean-CV relationship of a gene expressed with parameters scaling and at the Poisson limit. rpb1-pom1Δ, *rpb1* expression in *pom1*Δ cells.

(D) Expression correlation between mRNA pairs in single cells. Orange dots: experimental measurements; orange bars: 95% confidence intervals for experimental measurements; blue bars: 95% confidence intervals for correlations obtained from model simulations as in (C); red bars: 95% confidence interval for simulations including 20% extrinsic noise.

(E) Model probabilities for three models of mRNA degradation after ABC model selection (two experiments). Decay rates are scaling ("scaling") or kept constant ("constant single" and "constant multiple") as cell size varies. See STAR Methods for details.

See also Figure S2 and Tables S1, S2, S3, and S4.

Table 1. Maximum *A Posteriori* Estimates of Cell Cycle and Size Control Parameters Obtained by ABC Inference on the Snap-Shot Cell Size Data

Strain	a	b (μm)	η_b (μm)	η_d	m_g (1/Cell Cycle)	σ_g (1/Cell Cycle)
WT	0.195	11.03	0.253	0.045	1	0.111
pom1Δ	0.276	8.86	1.70	0.110	1	0.478

Used in Figure 2.

investigate this and to test the model predictions from Figure 2, we measured transcription rates experimentally. Single cells' transcription rates can be estimated from smFISH images by measuring intensities of nuclear transcription sites [38]. As a model, we used probes directed against the 5′ region of the *rpb1* mRNA. This probe design provides strong sensitivity for detection of nascent transcripts (Figures S3A and S3B). Transcription site intensities and cell size were weakly but significantly correlated when comparing WT, *wee1-50*, and *cdc25-22* cells, confirming modeling results and previous observation in a metazoan cell line (Figures 3A, 3B, and S3C) [5]. We then investigated the impact of increased ploidy on transcription rates scaling by comparing nascent site intensities of four mRNAs in haploid and diploid cells. We used probes against the 5′ end of *rpb1* as above and against three cell-cycle-regulated mRNAs, because inducible expression increases the sensitivity of nascent site detection. Intensities of individual transcription sites were similar between haploid and diploid cells (Figures 3C and S3D). However, when considering the total intensity per cell, diploids showed increased rates and scaling was apparent again (Figures 3C and S3D). These data support a model of scaling where the cell transcriptional capacity is limiting and distributed between the two gene copies of diploids. This is further strengthened by the analysis of the nascent intensity of the *rpb1* mRNA in a heterozygous deletion strain, which showed no increased rates in the remaining copy (Figures 3C and S3D).

To investigate further the mechanism behind scaling of transcription rates, we designed an orthogonal modeling approach where transcription is modeled as RNAPII particles hopping on a gene represented by a lattice (Figure 3D). This approach, which is based on a totally asymmetric simple exclusion process (TASEP), has been used successfully to study transcription and translation [7, 39–42]. In our model, the lattice is of length L_G and non-bursty transcription is modeled using three rates: (1) the transcription initiation rate α is the rate at which RNAPII molecules enter the first site of the lattice and (2) the elongation rate β is the rate at which RNAPII molecules hop one site forward on the lattice. If this site is already occupied by another RNAPII molecule, the rate of hopping is set to zero (Figure 3D, orange arrow); we note that sources of pausing, such as RNAPII sequence-specific stop or backtracking, are not modeled, as they are not supported by our chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) data (see below) [41]. (3) The termination rate γ is the rate at which RNAPII molecules leave the last site of the lattice and produce full-length mRNAs (Figure 3D). We incorporated this model in the agent-based framework from Figure 2, assuming each rate could be linearly coupled to cell size.

By sampling the rates α,β, and γ over physiological timescales estimated from previous studies (STAR Methods), we found that coupling of initiation rates with size produced the most robust linear scaling (Figure 3E) and the strongest positive correlation of nascent intensities with size (Figure S3E). Although a model coupling transcription elongation rates with cell size could also generate linear scaling in some parameter regions (Figures 3E and S3E), these required elongation rates to be much slower than the ~2 kb/min observed experimentally in either yeast or metazoans (Figure S3F) [38, 43–47]. Importantly, linear scaling could only be observed in regimes with slow initiation rates relative to elongation and termination, indicating that initiation is rate limiting (Figure S3G). Interestingly, this also suggests that fast, non-limiting initiation rates could be a mechanism by which some genes escape scaling (e.g., *rpb3*; Figure 2B). Scaling of termination rates does not lead to linear transcription scaling (Figure 3E) because limiting termination rates generate high traffic and stalling of polymerases. Finally, these results were strengthened by ABC inference analysis using the same TASEP model and nascent sites intensities for 3 genes in different strains, which showed clear preference for the initiation model (Figure S3H). This *in silico* analysis suggests that scaling of initiation rates with cell size could be the mechanism of scaling.

The initiation model generated two important predictions. First, even for non-bursty genes, cells in a population should not be transcribing actively at all times and the frequency of transcribing cells should increase with cell size (Figure S3I). To test this prediction, we compared the fraction of cells with a nascent transcription site ("active cell fraction") in *wee1-50*, WT, *cdc15-22*, and diploid cells. As predicted by the model, a clear increase in the fraction of transcriptionally active cells with size could be observed (Figure 3F). Moreover, a strong positive correlation of the active cells fraction with cell length was apparent when calculated in sliding windows of increasing cell numbers during the normal cell cycle (Figure S3J). The second prediction of the initiation scaling model is a positive correlation between the number of transcribing RNAPII and cell size (Figure S3E). To test this, we analyzed RNAPII occupancy across the genomes of WT, *wee1-50*, and *cdc25-22* cells by ChIP-seq. We used three antibodies against total RNAPII and serine 5 phosphorylation of its carboxy-terminal domain (CTD) (STAR Methods). Using data normalized to occupancy at histone genes (STAR Methods), we could observe a significant increase in overall RNAPII occupancy with size consistent with previous observation and supporting model predictions (Figure 3G; Table S6) [4]. Importantly, we did not find evidence for a redistribution of RNAPII from the 5′ to the 3′ of genes in larger cells, which makes regulation of scaling at the level of RNAPII pause/release unlikely (Figures 3H and S3K). Similar distributions of RNAPII molecules could be reproduced quantitatively by our minimal TASEP model in the case of RNAPII initiation rates scaling, but not elongation rates, supporting this conclusion further (Figures 3D and S3L). In summary, our *in silico* and experimental data indicate that transcriptional scaling is mediated by an increase in RNAPII initiation rates coordinated with cell size. In addition, our modeling data together with the increase in RNAPII occupancy observed in larger cells suggest that RNAPII could be a limiting factor for transcription as cell size increases.

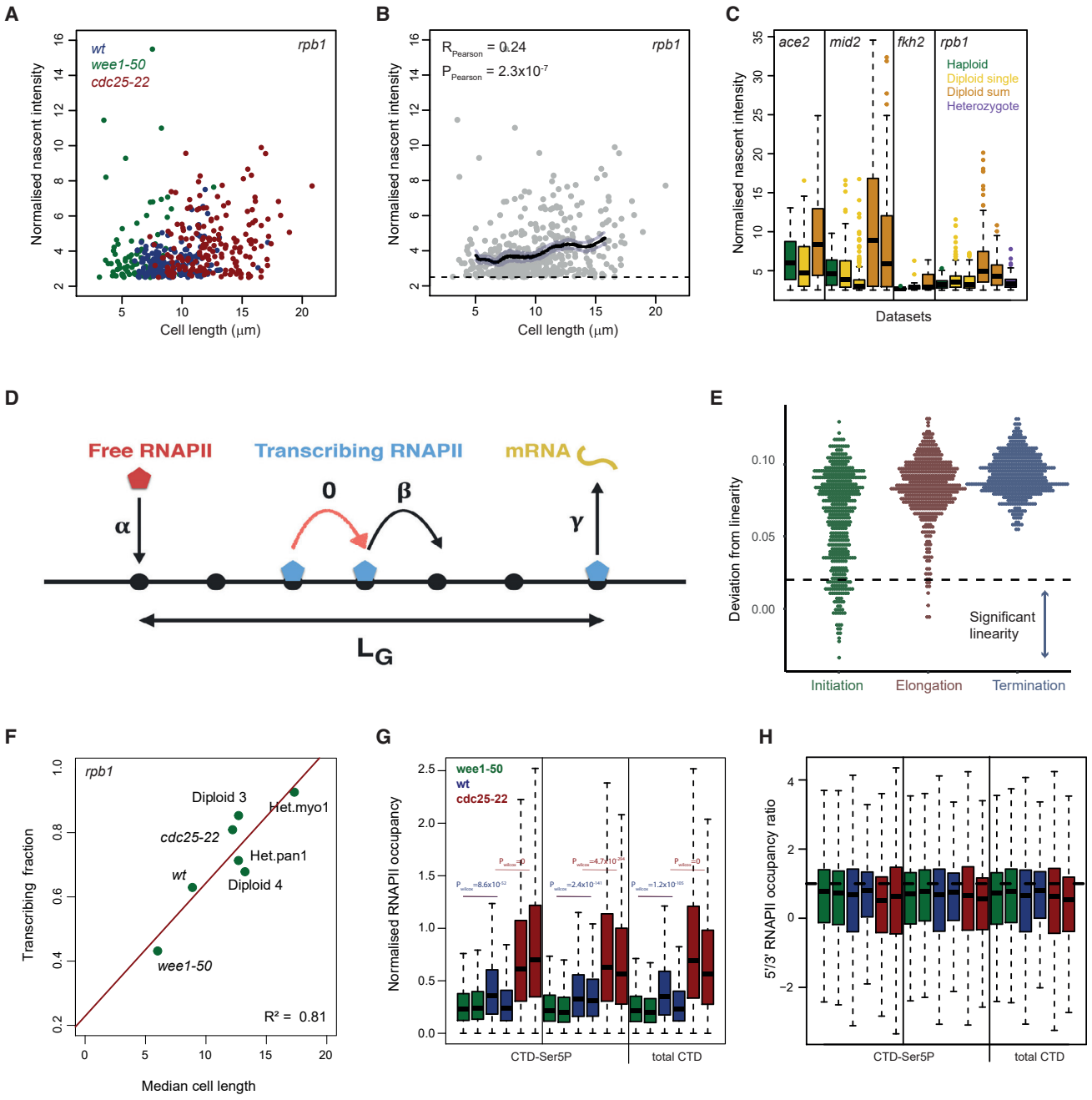


Figure 3. Coordination of RNAPII Initiation Rates with Cell Size as the Main Mechanism of Scaling

(A) *rpb1* normalized nascent sites intensities plotted against cell length for *wee1-50* (green), WT (blue), and *cdc25-22* (red).

(B) Same as (A) with all the cells in gray. Median intensities in running windows sampled from 100 experimental data bootstrap samples (line) and 95% confidence interval (shaded area) are shown. Cells lacking nascent sites are excluded. Pearson correlation coefficient and p value are indicated.

(C) Normalized nascent intensities for *rpb1* and the cell-cycle-regulated mRNAs *ace2*, *mid2*, and *fkh2* in haploid (green), diploid (yellow and orange), and *rpb1* heterozygote deletion cells (purple). Yellow: intensities of single nascent sites; orange: sum of nascent intensities per cell. Same color boxplots for a given mRNA denote multiple experiments.

(D) Cartoon of the RNAPII transcription TASEP model. Genes are represented by a lattice of length L_G. α, initiation rate; β, elongation rate (set to 0 if next site is occupied); γ, termination rate.

(E) Linear scaling of transcription rates captured by the TASEP model when α, β, or γ are coupled to cell length. Deviation from linearity between linear fits on the mRNA numbers versus cell length data for the smaller and larger half of simulated cells is plotted. Values close to zero indicate a linear scaling, and larger values indicate saturation of mRNA numbers in large cells (STAR Methods). The dashed line shows the error in estimating the slope of the linear fits.

(F) Fraction of cells with a nascent site ("transcribing fraction") plotted as a function of mean cell length for the *rpb1* mRNA in strains of varying average length. Linear fit (red line) and R^2 are indicated.

(legend continued on next page)

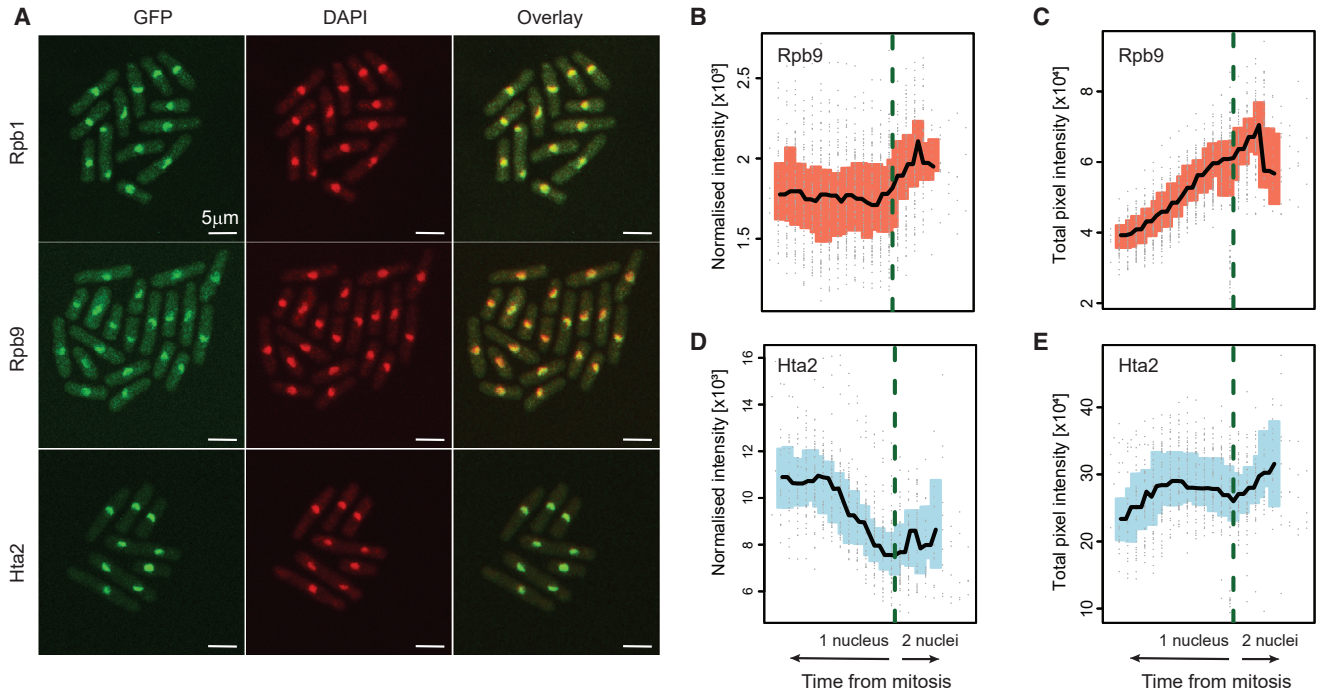


Figure 4. Nuclear RNAPII Concentration Increases with Cell Size

(A) Confocal images of Rpb1, Rpb9, and Hta2 proteins tagged with GFP in WT cells. DAPI staining for DNA and overlay of both channels are shown. White scale bars, 5 μm.

(B) Wide-field fluorescence data from live-cell imaging of the RNAPII subunit Rpb9. Normalized intensity/cell as a proxy for total cellular protein concentration is plotted along time relative to the mitotic phase of the cell cycle.

(C) As in (B) but showing the total pixel intensity in the area of strong fluorescence signal as a proxy for chromatin-bound amounts.

(D and E) As in (B) and (C) but for the histone Hta2.

See also Figure S4 and Tables S1 and S4.

RNAPII Amounts on Chromatin Increase with Cell Size

If scaling of initiation rates is the mechanism behind scaling and RNAPII is limiting, the amount of RNAPII complexes on the genome of single cells should increase during the cell cycle. To test this hypothesis, we measured localization of RNAPII in single cells by live-cell imaging. To do this, we tagged components of the RNAPII complex with green fluorescent protein (GFP) and imaged them during the cell cycle (Figures 4A and S4A). We observed that the cellular concentration of the RNAPII subunits Rpb1 and Rpb9 remained constant during *S. pombe* growth phase in G2 (Figures 4B and S4B). This indicates that scaling of transcription initiation is not controlled by regulation of the cellular concentration RNAPII. Our image data show that a very large fraction of the tagged RNAPII subunits localizes in the nucleus in the DAPI-positive area similar to the DNA-bound histone protein Hta2 (Figures 4A, S4F, and S4G). We therefore asked whether the amount of RNAPII on chromatin changes with cell size during the cell cycle. To assess this, we measured fluorescence intensities of the nuclear region occupied by RNAPII subunits (Figures 4C and S4C). Strikingly, the signal for Rpb1 and Rpb9 increased steadily during G2 when most cell elongation occurs (Figures 4C and S4C). Confocal analysis of Rpb1 intensities also showed a clear increase in signal in the DAPI-positive area of the nucleus that was correlated with cell size in line with our live-cell imaging data (STAR Methods; Figure S4D). Importantly, this was not apparent for another DNA-bound protein, the histone Hta2 (H2Aβ; Figures 4D, 4E, and S4E). These data are consistent with the predictions of the initiation scaling model and indicate that RNAPII quantities are likely to be limiting for transcription. In addition, these experiments indicate that RNAPII is efficiently imported in the nucleus and rapidly recruited onto chromatin.

A Role of the Nucleus in Scaling

In fission yeast and other organisms, nuclear and cytoplasmic volumes are intimately connected [48]. As scaling of initiation rates is coordinated with increased RNAPII levels in the nucleus, we wondered whether nuclear size rather than cell size itself could be the quantitative determinant of scaling.

(G) ChIP-seq analysis of RNAPII occupancy in *wee1-50* (green), WT (blue), and *cdc25-22* (red) cells. Data for antibodies against serine 5 CTD phosphorylation (left, middle) and total Rbp1 (right) are shown. p values for one-sided Wilcoxon tests are shown.

(H) As in (G) for RNAPII occupancy 5′/3′ ratios. Note that ratios are not changing with size and are consistently lower than 1.

See also Figure S3 and Tables S1, S2, S3, S5, and S6.

To test this idea, we analyzed nascent site intensities of *cdc11-119* mutant cells cultivated at non-permissive temperature. Under these conditions, cells elongate and undergo mitosis and nuclear division but do not divide (Figure 5A) [49]. Strikingly, scaling of nascent site intensities from individual nuclei with cell size was not apparent in this system (R^2 = 0.001; Figure 5B, right). Scaling was only restored when all nascent sites present in a cell were added together (R^2 = 0.2; Figure 5B, left). This mirrors data from multinucleated cells showing that, although the overall nuclear volume scales with cell volume, the volume of individual nuclei is proportional to their surrounding cytoplasmic volume [48]. Consistent with this, the ratios of nascent site intensities between nuclei of *cdc11-119* cells were weakly correlated with their adjacent cytoplasmic volume delimited by either the cytoplasmic membrane or half the distance to the next nucleus (STAR Methods; Figure S5A). This indicates that scaling could be dependent on physical properties of the nucleus correlated with cell size and not directly on the cell volume.

We next analyzed scaling in conditions where the correlation between cell and nuclear size is compromised. Pom1 is a regulator of cell polarity and division whose deletion leads to increased size variability at birth due to cell partitioning errors [50]. We analyzed expression of three mRNAs by smFISH in *pom1Δ* mutant cells expressing a marker of the nuclear envelope to allow measurement of nuclear size (Figure 5C). As expected, cell and nuclear size show smaller correlation in the *pom1Δ* mutants compared to WT cells (Figure S5B). Scaling in *pom1Δ* cells was linear as in WT cells, but mRNA numbers showed higher y axis intercept when plotted as a function of cell size (Figure S5C). This indicates that mRNA concentrations in *pom1Δ* cells are the highest in smaller cells after birth and decrease as cells elongate in G2 (Figure 5D). Although such a deviation from perfect concentration homeostasis is also observed in WT cells and has been reported in mammals, it is exacerbated in the mutant (Figure 5D) [5]. The modeling approach described in Figure 2, which is based on transcription scaling with cell size and binomial partitioning of mRNAs based on daughter cell sizes, failed to capture the increased deviation from concentration homeostasis of *pom1Δ* cells (Figure 5E, magenta line). However, a modified model coupling transcription rates to nuclear size instead captured it well (Figure 5E, green line). Overall, this analysis indicates that nuclear size may be a better predictor of scaling than cytoplasmic volume in multinucleated cells and in cells where the nucleo-cytoplasmic ratio is transiently perturbed. This raises the possibility that properties of the nucleus are direct and quantitative determinants of scaling.

A Mechanistic Model of Scaling

We used the results from this study to develop a mechanistic model of scaling centered around RNAPII-mediated transcription to integrate and explain the respective role of the cell volume and nuclear features in predicting scaling (Figure 5F; STAR Methods; see also [20, 21]). In this model, the rates of RNAPII complex synthesis and maturation scale with cell size. This is consistent with our smFISH and live-cell imaging data showing that RNAPII subunits have a constant cellular concentration during the cell cycle (Figures 1 and 4). RNAPII is then transported to the nucleus with a rapid rate, which is not limiting and is proportional to nuclear size. This results in depletion of RNAPII from the cytosol (as observed in live-cell imaging data; Figure 4). Once in the nucleus, RNAPII binds to DNA with a constant high affinity and transcription rates are proportional to the numbers of DNA-RNAPII complexes present on each gene at any given time to reflect scaling of initiation rates (Figure 3). Finally, RNAPII levels are set to be limiting in line with the initiation scaling model (Figure 3), the behavior of diploid and heterozygous mutants (Figure 3) [11], and the fact that the cell synthetic capacity is titrated against the number of genes in heterokaryons [5]. Moreover, this assumption fits the observation that many RNAPII subunits are limiting for growth in fission yeast [51–53]. The high affinity and the limiting amount of RNAPII ensures that the majority of RNAPII molecules are bound to DNA and increase with cell size consistent with our imaging data (Figure 4) and with biochemical evidence from mammalian cells [5]. Finally, DNA replication occurs close to cell division and each daughter cell inherits about half of the DNA-bound RNAPII, mostly independent of its size, as observed in our live-cell experiments (Figures 4C and S4C). This simple model captures the different features of the data presented in this study. First, it retrieves the scaling of DNA-bound RNAPII with cell size, while keeping its overall cellular concentration constant, as observed in Figure 4 (Figure S5D). Second, it explains the scaling of mRNA numbers with cell size, including the higher mRNA concentration observed in smaller cells (Figure 5E, blue line). Interestingly, based on this model, the total rate of transcription in a multinucleated cell is proportional to total RNAPII quantities, which is proportional to cell size (Figure 5B). However, consistent with our data, the amounts of RNAPII imported to each nucleus are proportional to the nucleus size and determine transcription rates (Figures 5B and S5A). Overall, this simple mechanistic model of RNAPII transcription, by integrating the experimental findings from this study, explains the origin of transcriptional scaling and the role taken by the nucleus. In summary, it identifies the competition between genes for the limited pool of transcriptional machinery as central to the phenomenon of transcriptional scaling.

DISCUSSION

We performed an extensive experimental and modeling study of gene expression scaling with cell size in single fission yeast cells. We found that scaling is a pervasive feature that impinges on constitutive and regulated expression. We then showed that scaling relies on an increase in RNAPII initiation rates with cell size and a concentration-independent recruitment of a limiting RNAPII on the genome. Finally, we propose that nuclear size may participate in setting scaling levels.

Our work supports a simple and robust model for the scaling of gene expression with cell size, in which the competition between promoters for a limited pool of RNAPII determines their relative strength. Because RNAPII maintains a constant concentration as cell size increases (as proteins do in general), the number of RNAPII complexes increase linearly with cell size. Cell size increase will not affect the relative strength of promoters but will cause their absolute rate of transcriptional initiation to scale linearly with cell size, exactly as required to produce the observed gene expression scaling.

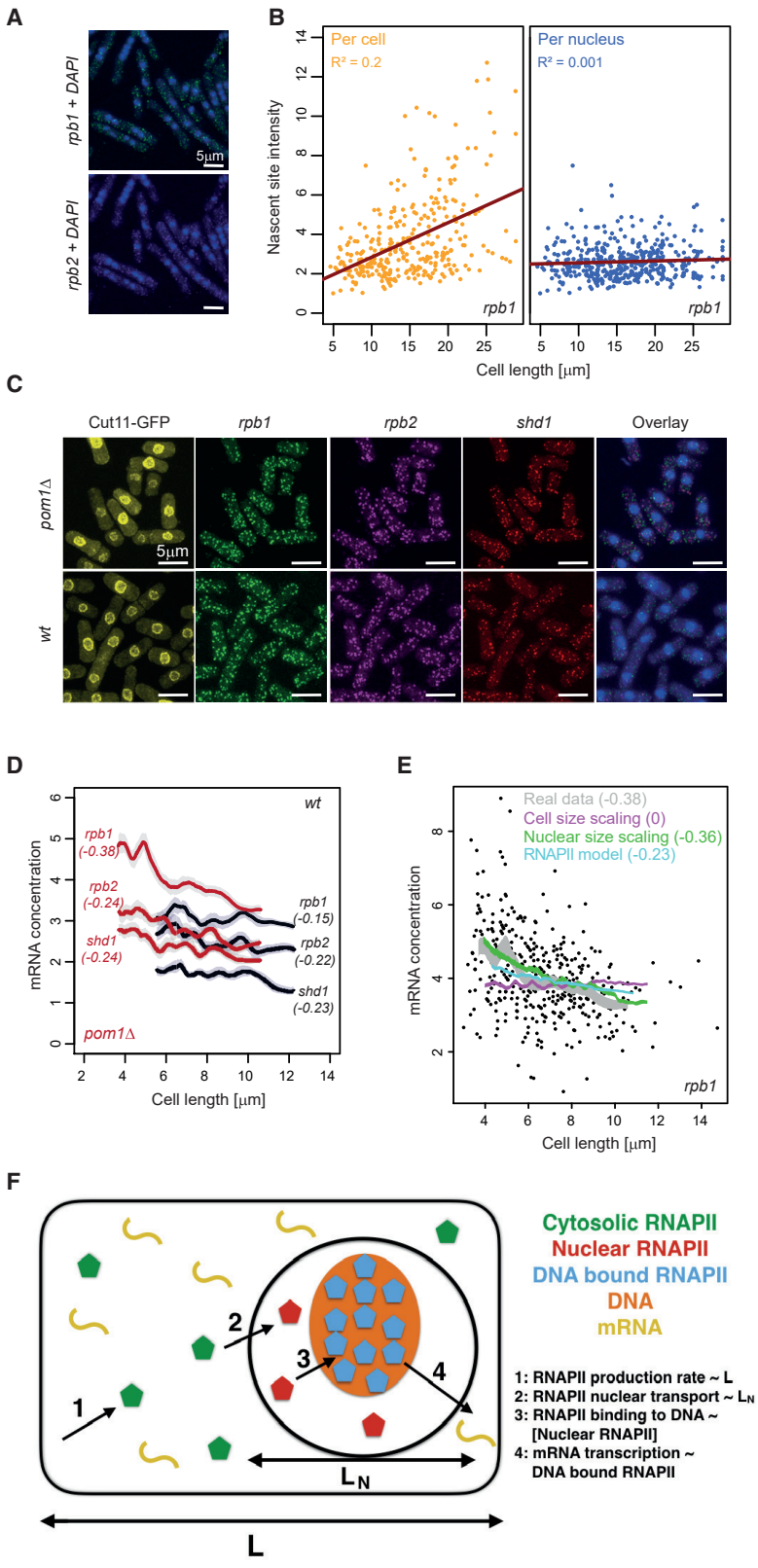


Figure 5. A Role for the Nucleus in Scaling

(A) smFISH images for the *rpb1* and *rpb2* mRNA in *cdc11-119* cells at restrictive temperature. White scale bars, 5 μm.

(B) *rpb1* normalized nascent sites intensities per nucleus (blue) or per cell (orange) plotted as a function of cell size. Linear fits (red lines) and adjusted R^2 are indicated.

(C) smFISH images for the *rpb1*, *rpb2*, and *shd1* mRNA in WT and *pom1Δ* cells carrying the Cut11-GFP nuclear marker (left column). Overlay is shown on the right. White scale bars, 5 μm.

(D) mRNA concentration (mRNA/cell length) for the *rpb1*, *rpb2*, and *shd1* mRNA in WT (black) and *pom1Δ* (red) cells. Pearson correlations are indicated between parentheses (all $P_{Pearson} < 0.05$). The solid line shows median counts in running windows sampled from a count distribution identical to the experimental data. Shaded area represents 95% confidence intervals.

(E) *rpb1* concentrations plotted as a function of cell length. Running average for the experimental data (gray) or derived from model predictions assuming transcription rates scaling with cell length (magenta), scaling with nuclear area (green), or predicted from a mechanistic model of RNAPII prediction (light blue) are shown. Pearson correlation coefficients are indicated between parentheses.

(F) Cartoon of a mechanistic model of RNAPII transcription (see STAR Methods for details and parameters).

See also Figure S5 and Tables S1, S2, and S3.

Our model assumes that the general and specific transcription factors that regulate relative promoter strength are in excess even in small cells and thus are not affected by cell size. To gain a deeper mechanistic understanding of scaling, it will be important to determine whether this assumption is met for most regulators or whether some escape the rule. It could also inform about mechanisms through which some mRNA escape regulation by scaling, as our modeling results suggest that non-limiting initiation rates can produce this behavior [11–13]. Another interesting question will be to determine the role of chromatin remodelers in facilitating transcription initiation in larger cells. It is possible that a more permissive chromatin environment in large cells synergizes with increased RNAPII local concentrations to support higher transcription rates of inducible genes.

Recent evidence suggests that, in mammals, burst size is regulated at the level of the proximal promoter sequence, although distal enhancers are involved in setting burst frequency [54]. Moreover, burst initiation and RNAPII pause/release, but not RNAPII recruitment, have been shown to be regulated in response to biological perturbations [55]. Our model of scaling through initiation of RNAPII transcription fits well with these data, as this process is independent of both gene activation and control of burst frequency. It is also consistent with the observation that promoters maintain their relative expression levels in changing growth conditions [56].

We find that scaling is regulated at the level of single nuclei in multinucleated cells and may be linked to nuclear size. RNA synthesis levels have been connected to nuclear size in other systems, such as multinucleated muscle cells [57], or for the HTLV-1 mRNA [58]. It could reflect a higher availability of RNA polymerases around larger nuclei, as these tend to be surrounded by larger cytoplasmic volume (Figure S5A) [48, 57]. Interestingly, the nucleus was also found to be an independent transcriptional unit in mature osteoclasts [59] and in multinucleated fungi, where nuclei retain local control of cell cycle periodic transcription [60, 61]. This suggests that more complex feedback and molecular mechanisms may also be at play.

An important result from our study, which is not directly related to scaling, is that transcription rates of most fission yeast genes tested are constant and show no evidence for off states or bursty transcription. This is in line with previous observations in budding yeast and plants [30, 38, 62]. Further experiments in live cells will be important to define precisely the molecular mechanisms behind this observation. Transcriptional bursts result in high gene expression noise and are associated with Fano factors ((σ^2/μ) of mRNA numbers) greater than 1, whereas Poissonian birth-death processes have a Fano factor = 1. Our finding that most transcription followed a non-bursty regime relied on our modeling, taking cell size and the cell cycle into account explicitly. Without doing so, all genes in this study would have been called bursty, as Fano factors calculated on smFISH raw counts were greater than 1 (not shown). This reiterates the importance of studying gene expression considering potential confounding effects of morphological features, such as cell size and the cell cycle [24, 25, 30, 63, 64].

Finally, in addition to progressing our understanding of the mechanisms behind scaling, this study provides a large quantitative dataset of gene expression and cell size measurement in over 20,000 cells in various conditions. This will support future modeling efforts aimed at understanding regulation of gene expression.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABLITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Strains and culture conditions
- METHOD DETAILS
  - RNA single molecule fluorescence *in situ* hybridization (smFISH)
  - ChIP-seq
  - ChIP-seq analysis
  - Live-cell microscopy and analysis
  - Cell size measurements
  - Mathematical modeling
  - ABC inference and model selection
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.01.053.

ACKNOWLEDGMENTS

We thank Nick Rhind for stimulating discussion and his help with formulating our model of scaling. We are grateful to Kurt Schmoller, Philipp Thomas, Marc Sturrock, and Nick Rhind for critical reading of the manuscript. We thank Snezhana Oliferenko, Paul Nurse, and Jürg Bähler for sharing strains. This research was supported by the UK Medical Research Council and a Leverhulme Research Project Grant (RPG-2014-408). A.B. is supported by an EPSRC PhD studentship and V.S. is supported by the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1). We used the computing resources of the UK Medical Bioinformatics partnership (UK MED-BIO), which is supported by the UK Medical Research Council (grant MR/L01632X/1) and the Imperial College High Performance Computing Service.

AUTHOR CONTRIBUTIONS

X.-M.S., A.B., V.S., and S.M. designed the study. X.-M.S. performed most experimental work. A.B. and V.S. performed the mathematical modeling work. M.P. and F.B. performed live-cell imaging experiments and image analysis on Figures 4 and S4A–S4C with help from X.-M.S. M.P. developed the PombeTrack image analysis package. A.M.-S., W.T., and S.M. performed additional statistical and computational analysis. C.W. and D.D. helped with setting up microscopy and image analysis protocols. V.S. and S.M. supervised the study and wrote the manuscript with help from X.-M.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 11, 2019
Revised: December 1, 2019
Accepted: January 16, 2020
Published: February 13, 2020

REFERENCES

1. Marguerat, S., and Bähler, J. (2012). Coordinating genome expression with cell size. Trends Genet. *28*, 560–565.
2. Schmoller, K.M., and Skotheim, J.M. (2015). The biosynthetic basis of cell size control. Trends Cell Biol. *25*, 793–802.
3. Amorim, M.J., Cotobal, C., Duncan, C., and Mata, J. (2010). Global coordination of transcriptional control and mRNA decay during cellular differentiation. Mol. Syst. Biol. *6*, 380.
4. Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J., and Nurse, P. (2010). A coordinated global control over cellular transcription. Curr. Biol. *20*, 2010–2015.
5. Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell *58*, 339–352.
6. Kempe, H., Schwabe, A., Crémazy, F., Verschure, P.J., and Bruggeman, F.J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol. Biol. Cell *26*, 797–804.
7. Ietswaart, R., Rosa, S., Wu, Z., Dean, C., and Howard, M. (2017). Cell-size-dependent transcription of FLC and its antisense long non-coding RNA COOLAIR explain cell-to-cell expression variation. Cell Syst. *4*, 622–635.e9.
8. Mena, A., Medina, D.A., García-Martínez, J., Begley, V., Singh, A., Chávez, S., Muñoz-Centeno, M.C., and Pérez-Ortín, J.E. (2017). Asymmetric cell division requires specific mechanisms for adjusting global transcription. Nucleic Acids Res. *45*, 12401–12412.
9. García-Martínez, J., Delgado-Ramos, L., Ayala, G., Pelechano, V., Medina, D.A., Carrasco, F., González, R., Andrés-León, E., Steinmetz, L., Warringer, J., et al. (2016). The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons. Nucleic Acids Res. *44*, 3643–3658.
10. García-Martínez, J., Troulé, K., Chávez, S., and Pérez-Ortín, J.E. (2016). Growth rate controls mRNA turnover in steady and non-steady states. RNA Biol. *13*, 1175–1181.
11. Schmoller, K.M., Turner, J.J., Kõivomägi, M., and Skotheim, J.M. (2015). Dilution of the cell cycle inhibitor Whi5 controls budding-yeast cell size. Nature *526*, 268–272.
12. Keifenheim, D., Sun, X.-M., D'Souza, E., Ohira, M.J., Magner, M., Mayhew, M.B., Marguerat, S., and Rhind, N. (2017). Size-dependent expression of the mitotic activator Cdc25 suggests a mechanism of size control in fission yeast. Curr. Biol. *27*, 1491–1497.e4.
13. Saint, M., Bertaux, F., Tang, W., Sun, X.-M., Game, L., Köferle, A., Bähler, J., Shahrezaei, V., and Marguerat, S. (2019). Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. Nat. Microbiol. *4*, 480–491.
14. Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. USA *99*, 12795–12800.
15. Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in noisy gene expression. Science *342*, 1188–1193.
16. Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. *14*, 572–584.
17. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.
18. Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478.
19. Proudfoot, N.J. (2016). Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. Science *352*, aad9926.
20. Heldt, F.S., Lunstone, R., Tyson, J.J., and Novák, B. (2018). Dilution and titration of cell-cycle regulators may control cell size in budding yeast. PLoS Comput. Biol. *14*, e1006548.
21. Lin, J., and Amir, A. (2018). Homeostasis of protein and mRNA concentrations in growing cells. Nat. Commun. *9*, 4496.
22. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.
23. Ciechonska, M., Sturrock, M., Grob, A., Larrouy-Maumus, G., Shahrezaei, V., and Isalan, M. (2019). Ohm's Law for emergent gene expression under fitness pressure. bioRxiv. https://doi.org/10.1101/693234.
24. Thomas, P. (2019). Intrinsic and extrinsic noise of gene expression in lineage trees. Sci. Rep. *9*, 474.
25. Bertaux, F., Marguerat, S., and Shahrezaei, V. (2018). Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits. R. Soc. Open Sci. *5*, 172234.
26. McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA *94*, 814–819.
27. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. USA *98*, 8614–8619.
28. Munsky, B., Trinh, B., and Khammash, M. (2009). Listening to the noise: random fluctuations reveal gene network parameters. Mol. Syst. Biol. *5*, 318.
29. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., and Koeppl, H. (2012). Moment-based inference predicts bimodality in transient gene expression. Proc. Natl. Acad. Sci. USA *109*, 8340–8345.
30. Zechner, C., Unger, M., Pelet, S., Peter, M., and Koeppl, H. (2014). Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. Nat. Methods *11*, 197–202.
31. Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G., and Legewie, S. (2018). Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. Mol. Syst. Biol. *14*, e7678.
32. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472–474.
33. Komorowski, M., Costa, M.J., Rand, D.A., and Stumpf, M.P.H. (2011). Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. Proc. Natl. Acad. Sci. USA *108*, 8645–8650.
34. Tiberi, S., Walsh, M., Cavallaro, M., Hebenstreit, D., and Finkenstädt, B. (2018). Bayesian inference on stochastic gene transcription from flow cytometry data. Bioinformatics *34*, i647–i655.
35. Watt, S., Mata, J., López-Maury, L., Marguerat, S., Burns, G., and Bähler, J. (2008). urg1: a uracil-regulatable promoter system for fission yeast with short induction and repression times. PLoS ONE *3*, e1428.
36. Mercier, A., and Labbé, S. (2010). Iron-dependent remodeling of fungal metabolic pathways associated with ferrichrome biosynthesis. Appl. Environ. Microbiol. *76*, 3806–3817.
37. Wilusz, C.J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet. *20*, 491–497.
38. Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.
39. Zhao, Y.-B., and Krishnan, J. (2014). mRNA translation and protein synthesis: an analysis of different modelling methodologies and a new PBN based approach. BMC Syst. Biol. *8*, 25.
40. Sahoo, M., and Klumpp, S. (2013). Backtracking dynamics of RNA polymerase: pausing and error correction. J. Phys. Condens. Matter *25*, 374104.
41. Voliotis, M., Cohen, N., Molina-París, C., and Liverpool, T.B. (2008). Fluctuations, pauses, and backtracking in DNA transcription. Biophys. J. *94*, 334–348.

42. Dobrzynski, M., and Bruggeman, F.J. (2009). Elongation dynamics shape bursty transcription and translation. Proc. Natl. Acad. Sci. USA *106*, 2583–2588.
43. Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat. Struct. Mol. Biol. *14*, 796–806.
44. Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J. Biol. Chem. *266*, 71–75.
45. Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831–840.
46. Pelechano, V., Chávez, S., and Pérez-Ortín, J.E. (2010). A complete set of nascent transcription rates for yeast genes. PLoS ONE *5*, e15442.
47. Swinburne, I.A., and Silver, P.A. (2008). Intron delays and transcriptional timing during development. Dev. Cell *14*, 324–330.
48. Neumann, F.R., and Nurse, P. (2007). Nuclear size control in fission yeast. J. Cell Biol. *179*, 593–600.
49. Verde, F., Mata, J., and Nurse, P. (1995). Fission yeast cell morphogenesis: identification of new genes and analysis of their role during the cell cycle. J. Cell Biol. *131*, 1529–1538.
50. Bähler, J., and Pringle, J.R. (1998). Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. Genes Dev. *12*, 1356–1370.
51. Hayles, J., Wood, V., Jeffery, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Salas-Pino, S., Heichinger, C., and Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. Open Biol. *3*, 130053.
52. Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. *28*, 617–623.
53. Weston, L., Greenwood, J., and Nurse, P. (2017). Genome-wide screen for cell growth regulators in fission yeast. J. Cell Sci. *130*, 2049–2055.
54. Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.
55. Bartman, C.R., Hamagami, N., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A., and Raj, A. (2019). Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. Mol. Cell *73*, 519–532.e4.
56. Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. Mol. Syst. Biol. *9*, 701.
57. Windner, S.E., Manhart, A., Brown, A., Mogilner, A., and Baylies, M.K. (2019). Nuclear scaling is coordinated among individual nuclei in multinucleated muscle fibers. Dev. Cell *49*, 48–62.e3.
58. Billman, M.R., Rueda, D., and Bangham, C.R.M. (2017). Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. *2*, 87.
59. Youn, M.Y., Takada, I., Imai, Y., Yasuda, H., and Kato, S. (2010). Transcriptionally active nuclei are selective in mature multinucleated osteoclasts. Genes Cells *15*, 1025–1035.
60. Roberts, S.E., and Gladfelter, A.S. (2015). Nuclear autonomy in multinucleate fungi. Curr. Opin. Microbiol. *28*, 60–65.
61. Dundon, S.E.R., Chang, S.-S., Kumar, A., Occhipinti, P., Shroff, H., Roper, M., and Gladfelter, A.S. (2016). Clustered nuclei maintain autonomy and nucleocytoplasmic ratio control in a syncytium. Mol. Biol. Cell *27*, 2000–2007.
62. Wu, Z., Ietswaart, R., Liu, F., Yang, H., Howard, M., and Dean, C. (2016). Quantitative regulation of FLC via coordinated transcriptional initiation and elongation. Proc. Natl. Acad. Sci. USA *113*, 218–223.
63. Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. Cell *163*, 1596–1610.
64. Soltani, M., Vargas-Garcia, C.A., Antunes, D., and Singh, A. (2016). Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. PLoS Comput. Biol. *12*, e1004972.
65. Takahashi, C.N., Miller, A.W., Ekness, F., Dunham, M.J., and Klavins, E. (2015). A low cost, customizable turbidostat for use in synthetic circuit characterization. ACS Synth. Biol. *4*, 32–38.
66. Heinrich, S., Geissen, E.-M., Kamenz, J., Trautmann, S., Widmer, C., Drewe, P., Knop, M., Radde, N., Hasenauer, J., and Hauf, S. (2013). Determinants of robustness in spindle assembly checkpoint signalling. Nat. Cell Biol. *15*, 1328–1339.
67. Trcek, T., Chao, J.A., Larson, D.R., Park, H.Y., Zenklusen, D., Shenoy, S.M., and Singer, R.H. (2012). Single-mRNA counting using fluorescent in situ hybridization in budding yeast. Nat. Protoc. *7*, 408–419.
68. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. Nat. Methods *10*, 277–278.
69. Rodríguez-López, M., Gonzalez, S., Hillson, O., Tunnacliffe, E., Codlin, S., Tallada, V.A., Bähler, J., and Rallis, C. (2019). The GATA transcription factor Gaf1 represses tRNA genes, inhibits growth, and extends chronological lifespan downstream of fission yeast TORC1. bioRxiv. https://doi.org/10.1101/700286.
70. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
71. Lock, A., Rutherford, K., Harris, M.A., Hayles, J., Oliver, S.G., Bähler, J., and Wood, V. (2019). PomBase 2018: user-driven reimplementation of the fission yeast database provides rapid and intuitive access to diverse, interconnected information. Nucleic Acids Res. *47* (D1), D821–D827.
72. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
73. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
74. Wiśniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol. Cell. Proteomics *13*, 3497–3506.
75. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.
76. Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1*, 10.
77. Amir, A. (2014). Cell size regulation in bacteria. Phys. Rev. Lett. *112*, 208102.
78. Tanouchi, Y., Pai, A., Park, H., Huang, S., Stamatov, R., Buchler, N.E., and You, L. (2015). A noisy linear map underlies oscillations in cell size and gene expression in bacteria. Nature *523*, 357–360.
79. Jun, S., and Taheri-Araghi, S. (2015). Cell-size maintenance: universal strategy revealed. Trends Microbiol. *23*, 4–6.
80. Wang, P., Robert, L., Pelletier, J., Dang, W.L., Taddei, F., Wright, A., and Jun, S. (2010). Robust growth of Escherichia coli. Curr. Biol. *20*, 1099–1103.
81. Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.
82. Shahrezaei, V., Ollivier, J.F., and Swain, P.S. (2008). Colored extrinsic fluctuations and stochastic gene expression. Mol. Syst. Biol. *4*, 196.
83. Peccoud, J., and Ycart, B. (1995). Markovian modeling of gene-product synthesis. Theor. Popul. Biol. *48*, 222–234.

84. Shahrezaei, V., and Swain, P.S. (2008). Analytical distributions for stochastic gene expression. Proc. Natl. Acad. Sci. USA *105*, 17256–17261.

85. Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. Ann. Stat. *12*, 1151–1172.

86. Marin, J.-M., Pudlo, P., Robert, C.P., and Ryder, R.J. (2012). Approximate Bayesian computational methods. Stat. Comput. *22*, 1167–1180.

87. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M.P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface *6*, 187–202.

88. Lenormand, M., Jabot, F., and Deffuant, G. (2013). Adaptive approximate bayesian computation for complex models. Comput. Stat. *28*, 2777–2796.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-RNA polymerase II CTD repeat YSPTSPS antibody [8WG16]	abcam	Cat# ab817; RRID: AB_306327
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody [4H8]	abcam	Cat# ab5408; RRID: AB_304868
Anti-RNA polymerase II subunit B1 (phospho-CTD Ser-5) Antibody, clone 3E8	Sigma Aldrich	Cat# 04-1572-I; RRID: AB_2801296
Dynabeads Pan Mouse IgG	Invitrogene	11041
Dynabeads Sheep anti-Rat IgG	Invitrogene	11035
Dynabeads M-280 Sheep anti-Mouse IgG	Invitrogene	11201D
Chemicals, Peptides, and Recombinant Proteins		
Dynabeads Protein A	Invitrogene	10001D
Dynabeads Protein G	Invitrogene	10003D
Critical Commercial Assays		
NEBNext ChIP-seq Library Prep Master Mix Set for Illumina	NEW ENGLAND BioLabs	E6240S
NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)	NEW ENGLAND BioLabs	E7335S
NEBNext Multiplex Oligos for Illumina (Index Primers Set 2)	NEW ENGLAND BioLabs	E7500S
Experimental Models: Organisms/Strains		
S. pombe: h-972wt	Nurse lab collection	ySBM2
S. pombe: h+ cut11-GFP::ura4 leu1-32 ura4-D18	Nurse lab	ySBM3
S. pombe: h-cdc25-22	Bähler lab	ySBM14
S. pombe: h-wee1-50	Bähler lab	ySBM15
S. pombe: h- cdc11-119	Bähler lab	ySBM42
S. pombe: h- rpb1-GFP ade6-M216 leu1-32	NBRP-Yeast	ySBM45
S. pombe: h+ pom1Δ::KanMX4 ade6-216 leu1-32 Ura4-D18	Bioneer v5 deletion library	ySBM59
S. pombe: h- pom1Δ::KanMX4, cut11-GFP	this paper	ySBM61
S. pombe: h90hta2-GFP-HA::KanMX ade6-216 leu1-32 lys1-131 ura4-D18	Bioneer	ySBM69
S. pombe: h-hta2-GFP-HA::KanMX	this paper	ySBM134
S. pombe: h90 rpb9-GFP-HA::KanMX ade6-216 leu1-32 lys1-131 ura4-D18	Bioneer	ySBM182
S. pombe: h+uch2-GFP	Oliferenko lab	ySBM186
S. pombe: h+uch2-mcherry	Oliferenko lab	ySBM188
S. pombe: h-rpb1-GFP, uch2-mcherry	this paper	ySBM189
S. pombe: h+ rpb9-GFP-HA::KanMX, uch2-mcherry	this paper	ySBM193
S. pombe: h90 rho3-GFP-HA::KanMX, uch2-mcherry	this paper	ySBM194
S. pombe: h-hta2-GFP-HA::KanMX, his7:SV40pr-mCherry-NLS-leu1	this paper	ySBM262
S. pombe: h+/h+leu1-32/leu32 ade6-M216ura4-D18/ura4-D18Δ rpb1/WTΔ	Bioneer	ySBM98
S. pombe: h+/h+ leu1-32/leu32 ade6-M216ura4-D18/ura4-D18Δ pan1/WTΔ	Bioneer	ySBM105

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. pombe: h+/h+ leu1-32/leu32 ade6-M216ura4-D18/ura4-D18Δ myo1/WTΔ	Bioneer	ySBM106
S. pombe: h+/h+ ade6-M210/ade6-M216 ura4-D18 leu1-32/leu1-32	Bioneer	ySBM125
S. pombe: h-,nmt(41)-rpb1::KanMX	this paper	ySBM295
S. pombe: h-,nmt(41)-rpb2::KanMX	this paper	ySBM296
S. pombe: h+,nmt(41)-rtp1::KanMX	this paper	ySBM297
Oligonucleotides		
See Table S7	N/A	N/A
Recombinant DNA		
pFA6a-kanMX6-P41nmt1	Addgene	Cat#39281
Software and Algorithms		
FIJI (ImageJ)	NIH	https://fiji.sc/
MATLAB	Mathworks	R2012a, R2015b
FISH-quant v2c	Florian Mueller, muellerf.research@gmail.com	http://www.app.asso.fr/en/
Models and simulation data	This paper	https://github.com/vshahrez/Sun-Bowman-et-al
ABC implementation in Julia	This paper	https://github.com/vshahrez/AdaptiveABC.jl
Live-cell microscopy analysis (widefiled timelapse image analysis)	This paper	https://github.com/ImperialCollegeLondon/PombeTrack/releases/tag/v0.1
ChIP-seq data	This paper	E-MTAB-8522
Other		
μ-Slide IV 0.4 Uncoated	ibidi	cat#81601
CellASIC ONIX Microfluidic Plates	Millipore	cat no Y04c-02-5PK

LEAD CONTACT AND MATERIALS AVAILABLITY

Further information and requests for resources and reagents (including *S. pombe* strains) should be directed to and will be fulfilled by the Lead Contact, Samuel Marguerat (samuel.marguerat@imperial.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains and culture conditions

The strains used in this study are listed in Table S1. Genetic crossing confirmed by polymerase chain reaction was used for strains generated unless otherwise specified. Strains were revived from glycerol frozen stocks on solid yeast extract agar (YE agar), or YE agar supplemented with 25 mg l^{-1} adenine, L-histidine, L-leucine, uracil, L-lysine, and L-arginine (Sigma), and with or without appropriate antibiotics for selection. YE agar plates were incubated for approximately 48 h at 32°C in a static incubator until visible large colonies could be observed. Single colonies were transferred into liquid yeast extract medium (YE), in YE supplemented as above (YES), Edinburgh minimal medium (EMM), or EMM supplemented as above (EMMS), unless otherwise indicated in figure legends, and incubated at 170 rpm in a shaking incubator. Temperature sensitive strains were grown at 32°C and shifted to 36.5°C for the time indicated in figure legends. For the induction of *sib1* expression, the strains were grown at 25°C to an optical density at 600 nm (OD_{600}) of ≈0.4 and treated with 2,2-dipyridyl (DIP; ACROS) at a final concentration of 250 μM for the time indicated in figure legends, or left untreated. For measuring mRNA decay rates, cells were grown in YE at 25°C to OD_{600} ≈0.4; cells were treated with thiolutin (AXXORA) for the time indicated at a final concentration of 15 μg/ml, or left untreated. For *urg1* induction, cells were grown in EMM supplemented with or without 0.25 mg/ml uracil for the time indicated. For transcription inhibition, log phase cultures (OD600∼0.5) were treated with thiolutin (15 ug/ml) and same volume of DMSO (used for dissolving thiolutin) was added to thiolutin untreated culture. Samples were taken at 0, 25, 35, 45 mins and processed as for smFISH. For live-cell experiments, cells were grown in EMMS in syphonstats – chemostat-like devices (http://klavinslab.org/hardware.html) which maintain the turbidity of liquid cultures by diluting with fresh medium appropriately – and maintained at OD_{600} 0.4 at 32°C by frequent dilution [65].

METHOD DETAILS

RNA single molecule fluorescence *in situ* hybridization (smFISH)

All smFISH datasets are described in Table S2. All the mRNA counts, nascent site intensities and cell size measurements are available in Table S3. smFISH samples were prepared according to a method modified from published protocols [66, 67]. Briefly, cells were fixed in 4% formaldehyde and the cell wall partially digested using zymolyase. Cells were permeabilised in 70% ethanol, pre-blocked with bovine serum albumin and salmon sperm DNA, and incubated overnight with custom Stellaris oligonucleotide sets (Biosearch Technologies) labeled with CAL Fluor Red 610, Quasar 670, or Quasar 570 (probe sequences are listed in Table S7). Cells were mounted in ProLong Gold antifade mountant with DAPI (Molecular Probes), and imaged on a Leica TCS Sp8 confocal microscope, using a 63x oil objective (NA 1.40). Optical z sections were acquired (0.3 μm step size) for each scan to cover the entire depth of cells. Cell boundaries were outlined manually and single mRNA molecules were identified and counted using the FISH-quant MATLAB package [68]. Cell area, length, and width were quantified using custom ImageJ macros. The technical error in FISH-quant detection was estimated at 6%–7% by quantifying *rpb1* mRNA foci with two sets of probes labeled with different dyes. Nascent intensities were normalized by the mode of the intensities of that gene in a given cell (unless there were not enough intensities in the cell, where the mode across the whole scan was substituted instead). The nascent mRNA foci were identified and using two approaches. In the first approach nascent site were identified using an in-built function of the FISH-quant package that relies on DAPI images to identify nuclear dots. In the second approach, we defined single mRNA dots with intensities 2.5 to 3-fold above the modal intensity within the same cell as nascent sites ("threshold method). The quality of the identification of nascent sites was validated manually by visualizing high intensity foci in the nucleus, with an accuracy of over 90% in all three strains (wild-type, *cdc25-22*, and *wee1-50*).

ChIP-seq

Chromatin immunoprecipitation (ChIP) assays was carried out essentially according to published methods [69]. In brief, cells were grown in YES to an OD600 of ~0.8 and fixed with formaldehyde solution (1% final) and then quenched with glycine. After washing twice with cold PBS (phosphate-buffed saline), cells were re-suspended in lysis buffer containing proteases inhibitors and disrupted vigorously with acid-washed glass beads 8-11 times for 20 s in a FastPrep instrument. Samples were then sonicated in Bioruptor (at High setting and 6 times 5 mins with 30 s ON/30 s OFF). Chromatin were immunoprecipitated with antibody against Rpb1 (ab817) or Rpb1 CTD-ser 5 (ab5408 or sigma 04-1572), which were coupled to Dynabead protein-G and protein-A and Dynabead sheep anti-mouse or rat IgG (Invitrogene). DNA was purified from immunoprecipitated samples using MinElute QIAGEN kit. Quantification of the DNA was done using Qubit dsDNA HS assay kit and quality was verified using Bioanalyzer.

For sequencing DNA from immunoprecipitated samples, the libraries were made using the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (E6240S) with the indexes provided in NEBBext Multiplex Oligos for Illumina (Index Primers Set1,2 and 3). Negative control DNA are those from the same chromatin extracts without going through immunoprecipitation steps. Pools of libraries were sequenced on an Illumina HiSeq 2500 instrument at the MRC LMS genomics facility. Paired-end reads (100 nt) were generated from two pools of 12 or 18 samples per sequencing lane. Data were processed using RTA 1.18.64, with default filter and quality settings. The reads were de-multiplexed with bcl2fastq-1.8.4 (CASAVA, allowing 0 mismatches).

ChIP-seq analysis

A description of ChIP-seq libraries can be found in Table S5. Sequencing reads were aligned to the fission yeast genome as available in PomBase in July 2019 using BWA [70, 71]. For Figure 3G, RNAPII occupancy counts were extracted for each transcript using HTseq [72] and the fission yeast annotation available in PomBase in July 2019 [71]. Data were normalized using DESeq2 (Table S6) and scaled using the mean counts of fission yeast histone genes as a scaling factor to allow comparison of global RNAPII occupancy between size mutants [73]. Amounts of histone proteins are thought to scale with DNA content rather than cell volume and are commonly used as normalization factors for absolute proteomics measurements [74]. Moreover, average synthesis rates of histones were found to remain constant across a wide range of sizes in budding yeast (K. Schmoller, personal communication). For Figures 3H and S3K, RNAPII immunoprecipitation data were normalized with their respective input and average gene analysis was performed using the deeptools analysis suite [75]. For Figure 3H, RNAPII occupancy values across the length of each gene were divided into 100 bins using the deeptools package. Ratios between mean occupancy in bins 1-50 (5′) compared to bins 51-100 (3′) are shown.

Live-cell microscopy and analysis

Strains of interest and wild-type ySBM2 were grown from single colonies in 5 mL YES before they were transferred into syphonstats and maintained at OD_{600} 0.4 overnight in EMMS. Prior to microscopy, ySBM2 cells were mixed at a 1:10 ratio with each strain of interest and diluted to a final OD_{600} of 0.3 in fresh EMMS. Cells were loaded directly into a CellASIC ONIX Y04C-02 microfluidic plate (EMD Millipore) according to the manufacturer instructions. Fresh EMMS was continually perfused through the growth chamber with a constant pressure of 6.9 kPa (approximately 3 μl/h). Cells were imaged on an Olympus IX70 inverted widefield fluorescence microscope with an environmental chamber maintained at 32°C, with a high precision motorised XYZ stage (ASI), controlled with μManager version 1.4.22 [76]. Cells were continually imaged at a 10 min interval with a 40 × objective (NA 0.95, UPlanSApo; Olympus) with brightfield (30 ms exposure), GFP (250 ms exposure, emission filter Semrock 514/30 nm), and dsRed (500 ms exposure, emission filter Semrock 617/73 nm) channels captured by a Hamamatsu Orca Flash 4.0 V2 sCMOS camera, with illumination provided by a Lumencor Spectra X LED light source set to 20% power. For each of the four growth chambers within the microfluidic plate, three

positions were defined, each of which was focused using the software autofocus using the brightfield channel; since this feature was generally inaccurate, a 3 μm Z stack was used around the autofocus position to ensure that at least one Z-position was in focus for each position.

Initial analysis was performed with Fiji – a derivative of ImageJ – with time-lapses assembled if necessary, and in-focus slices for each time-point selected using a custom macro (written by Stephen Rothery, Facility for Imaging by Light Microscopy, Imperial College London), resulting in a 4-dimensional OME-TIFF file for each field-of-view (XYCT). Each file was subsequently analyzed using a series of custom Python scripts utilizing the scikit-image, SciPy, NumPy, and pandas packages extensively among others. The scripts have been assembled in a python package called PombeTrack available in github (https://github.com/ImperialCollegeLondon/PombeTrack/releases/tag/v0.1). The scripts permit the semi-automated definition of cell boundaries (segmentation) within the brightfield channel, followed by the quantification of fluorescence within cell boundaries, identification of nuclei, further quantification within nuclear boundaries, and assignment of cells into lineages. Cell segmentation was effected using a custom 'balloon-filling' algorithm, in which a connected series of nodes is 'inflated' from its center inducing an outward force on all nodes; this outward force is counteracted by an 'image force', which applies an opposing force inversely proportional to the intensity of pixels neighboring the node, this has the effect of preventing nodes from expanding through areas of low light intensity – which generally surrounds the cell boundary; finally each node is also affected by its direct neighboring nodes which pull each node sideways according to their position, ensuring smooth contours. Together, after multiple iterations, nodes migrate from a central location until the forces equilibrate at areas of low light intensity (generally the edge of the cell). This procedure can be performed in an automated manner in which iterations cease when the area contained within the nodes does not significantly change, or in a supervised manner, in which further iterations are prompted manually via a keyboard command, with progress displayed via a graphical interface. Initial centers of cells are defined either by manually clicking, or by the center of the cell in the previous frame. Nuclei are defined as areas within the cell boundary which have red fluorescence pixel intensity values greater than 1.1 standard deviations above the mean intensity within the whole cell boundary. Time from mitosis is defined as the number of hours from the point at which the number of detected nuclei increases from one to two. Fluorescence intensity is adjusted for uneven illumination according to a series of empty fields imaged using the same settings. Background and autofluorescence is determined from wild-type cells cultured within the same field-of-view, with the mean level of fluorescence within these cells subtracted from measurements. Fluorescence is normalized by cell or nuclear area by dividing total fluorescence of all pixels within their respective boundary by the area of that boundary. Scripts are available upon request.

For live cell imaging in Figure S4, cells were imaged in ibidi microfluidic channel slides (μslide VI 0.4, #80601) instead of a CellASIC ONIX microfluidic plate. 30 μL of cells sampled from syphonstat cultures maintained at OD600 = 0.5 for at least 15 generations were loaded into channels of a pre-warmed slide. 40 μL of pre-warmed EMMS was then added to each reservoir of channels, followed by 10 μL of mineral oil (Sigma, M5904) to prevent evaporation during imaging. Live-cell microscopy was performed as described above.

Image analysis was semi-automated and performed using custom interactive MATLAB scripts available upon request. Tracks corresponding to individual cell cycles are recorded by user clicking. Only cell cycles for which cells remained in in focus for at least the 2 hours preceding mitosis were collected. Local background subtraction was automatically applied to fluorescence images. Cell segmentation was performed automatically based on thresholding of the brighfield images. Nuclear segmentation was performed automatically based on thresholding of the uch2-mCherry fluorescence images.

For the live cell imaging in Figures S4D and S4E, cells were seeded in ibidi microfluidic channel slides as described for Figures S4A–S4C. Live-cell microscopy was performed on a Leica TCS Sp8 confocal microscope, using a 40x oil immersion objective (NA 1.40) with pinhole equal 1 and resonant scanning. An environmental chamber was used to maintained temperature at 32°C and GFP with brightfield were acquired with an optical z stacks of 5μm with 6 steps to ensure covering the whole range of nuclear thickness. The image analysis was semi-automated and performed using custom scripts in ImageJ.

Cell size measurements

We extracted both cell area and cell length measurements form the smFISH images as proxies for cell volume. We observed that both measurements support robustly data characteristics such as scaling of mRNA numbers and positive intercepts. We used cell length as a proxy for cell volume throughout the manuscript as it proved to be a simpler and more consistent measure. Importantly, as fission yeast has a cylindrical shape, its length is directly proportional to its volume. For the nucleus, we acquired area measurements only. As the nucleus is spherical, area and volume are not proportional. We have therefore derived volume estimates from area measurements assuming a perfect sphere using:

$$\text{volume} = 4 \Big/ 3^{*}\pi^{*}(\text{area}/\pi)^{3/2}$$

Mathematical modeling

We use agent-based simulations of stochastic gene expression coupled to cell size in growing and dividing cells (Figure 2A) [25]. We assume cells grow exponentially with a constant growth rate from birth to division that is sampled from a truncated Gaussian distribution with mean m_g and standard deviation σ_g. A cell that is born with birth length L_i grows until it reaches the division length L_f We

use a phenomenological model of cell size control that relates the final size to initial size through a noisy linear map, which captures experimentally observed variability and correlations in cell size [77–79].

$$L_f = aL_i + b + \eta_b, \qquad \text{(Equation 1)}$$

where a and b are size control parameters ($a = 0$ denotes a sizer mechanism and $a = 1$ an adder mechanism) and η_b is a truncated Gaussian with mean zero and standard deviation σ_b. The dividing cell of length L_f produces two daughter cells with birth sizes L_i' and $L_f - L_i'$, where $L_i' = L_f\eta_d$ and η_d is another truncated Gaussian with mean 0.5 and standard deviation σ_d. The biomolecules such as mRNA molecules (except for DNA) are binomially partitioned in the daughter cells with a probability proportional to the daughter cell size L_i'As shown in Figure 2A, we simulate a fixed number of cells, so upon cell division one of the existing cells (including one of the newly born daughter cells) is chosen randomly and taken out of the simulation. This ensures we are simulating a constant number of cells in time and can produce snap-shot data with steady-state cell age and size distribution as observed in the experimental data. This has been used in the simulations used in the ABC inference (Figure 2). For the modeling results shown in Figures 3 and 5, where the results are conditioned on cell size, we have used a simpler scheme [25], where upon cell division only one of the daughter cells is followed modeling a single lineage (similar to the data generated in a mother machine) [80]. The simulations are performed using a simple algorithm that uses discretized time steps to simulate exactly the Gillespie method [81] with time-dependent parameters [82]. The simulation code for inference is written in the Julia programming language (ABC code is available at https://github.com/vshahrez/AdaptiveABC.jl) and the rest of simulations are performed in R. The agent-based simulation codes (used for Figures 2, 3, and 5) are available on github (https://github.com/vshahrez/Sun-Bowman-et-al).

Our main gene expression model is the so-called telegraph model or the two-state model [83] where the gene can be in an 'off' or an 'on' state and transcription can only occur when the gene is on (Figure 2A). If the gene is always on then this model reduces to a simple non-bursty birth-death process with parameters transcription rate v and mRNA decay rate d. Here the mRNA counts per cell have a Poisson distribution (in the absence of cell cycle effects). In the limit where the duration of the promotor on-state is shorter than the mRNA life time we have the bursty limit characterized by 3 parameters of average burst frequency k_{on}, average burst size v/k_{off} and mRNA decay rate d. In this model, birth events are simulated as geometrically distributed increases in mRNA numbers [84]. For Figure 2B model selection, we used five variants of this model including two non-scaling models (Poisson or bursty) and three transcription scaling models (Poisson model with transcription scaling or bursty limit with either burst size scaling, or burst frequency scaling). Transcriptional scaling is modeled as linear dependence of transcription rate or burst size to cell size ($v(L) = v^*L$, where, v^* is a constant and L is the cell length) or burst frequency ($k_{on}(L) = k_{on}^*L$, where, k_{on}^* is a constant and L is the cell length). The models for the cell cycle regulated genes, assume that there is a time point in the cell cycle, where gene expression increases from a basal level to an active level and there is another time point in the cell cycle, where gene expression switches to basal level again. The inference in Figure 2 is performed in two steps. First the cell cycle and size control parameters are inferred by ABC on the snapshot size data for the different strains. The maximum *a posteriori* estimates are shown in Table 1, which support the idea that fission yeast is more of a sizer ($a \ll 1$) and *pom1Δ* strain has more variability in cell division. These estimates are then used on the ABC model selection of different models of gene expression scaling in Figures 2B and 2C (samples of the posterior from this ABC model selection inference is included as Supplementary Table S4). Poisson line in Figure 2C is created by fixing noisy linear map parameters of cell cycle using the ABC fits to the wt data and varying v^* over a range while keeping mRNA life-time constant in the Poisson scaling model to produce a range of different mean expressions. ABC model selection in Figure 2E is performed on the data from the transcription shut-off experiments of the 3 strains of *wt*, *wee1-50*, and *cdc25-22.* The 3 models included are all the Poisson limit with different scaling assumptions for transcription and decay rates. Model one assumes transcriptional scaling with a single constant decay rate across the 3 strains (''Constant single''), the second model assumes also transcriptional scaling but with 3 different constant decay rates within each strain (''Constant multiple''). And the third model assumes constant transcription and decay rate that is proportional to inverse cell size across the 3 strains ($d(L) = d^*L$, where d^* is a constant and$L$ is the cell length) (''Scaling''). The priors used in the model selection, are wide over a physiological range. The model selection results were not too sensitive to the choice of the priors.

In Figure 3D a more detailed model of transcription is illustrated, which is based on the totally asymmetric simple exclusion process (TASEP) [7, 39–41]. Here, the promoter is assumed to be always active, i.e., we are modeling a non-bursty gene. The gene is modeled as a lattice of size L_G. Transcription starts by initiation through binding of a RNAPII molecule to the first site on the gene with rate α if the site is empty. Elongation is modeled as hopping of RNAPII forward with rate β if the next site is available (otherwise, the hopping is blocked, as illustrated in Figure 3D). Termination is modeled with RNAPII leaving the last site on the gene with rate γ, which gives rise to a fully transcribed mRNA. In our model, we have ignored pausing, backtracking and incomplete termination. In Figure 3E, we compare 3 variants of this model, where size scaling is through linear coupling of initiation $(\alpha(L) = \alpha^*L)$, elongation $(\beta(L) = \beta^*L)$ or termination rate $(\gamma(L) = \gamma^*L)$ to cell length (L). We chose $L_G = 20$ for computational efficiency and as it is larger than the typically observed number of RNAPIIs on the genes, which is related to nascent site intensity (Figure 3A). In each model, we randomly picked the initiation timescale τ_I, elongation timescale through the whole gene τ_E, and termination timescale τ_T between 0.001-0.1 hours that produces average mRNA numbers of between 20-30 for a moderately expressed gene. The lower limit on the timescale is significantly shorter than the mRNA life time and the upper limit represents very slow steps to achieve moderate transcription, given the mRNA life time, and is also slower than the time-scales reported in the literature [3]. The rates are inversely proportional to the time-scales as $\alpha = 1/\tau_I$, $\beta = L_G/\tau_E$ and $\gamma = 1/\tau_I$. Note that as β is the rate of hopping per site, it is also proportional to the number of sites on the gene L_G. Simulations where run using 500 random sets of α, β and γ. For each parameter set 1500 cells are simulated and two

linear regressions are done on the mRNA numbers versus cell length data for the smaller and larger half of simulated cells. The deviation from linearity is estimated as the difference between the linear regression coefficients of small versus large cells. Values close to zero indicate a linear scaling and larger values indicate saturation of mRNA numbers in large cells. We also performed an ABC model selection using an implementation of our TASEP model in Julia on several datasets (Figure S3G). We used the same prior as discussed above.

The nuclear scaling model and the RNAPII model in Figure 5E rely on both cell size and nuclear size dynamics. It is known that nuclear size scales closely with cell size [48]. There has not been much modeling of nuclear size control in the literature. We introduce a phenomenological and passive model of cell and nuclear size control, extending the noisy linear map of cell size control (Equation 1). We assume cellular exponential growth, cell size control and division as before. We assume nuclear size also grows exponentially and follows its own noisy linear map:

$$L_{Nf} = a_N L_{Ni} + b_N + \eta_{Nb}$$

Cell division time is determined when cells reach their final size (L_f).For simplicity, we assume mitosis is taking place at cell division and the size of the newly divided daughter cells and their nucleus is determined by $L_i' = L_f \eta_d$ and $L_{Ni}^{'} = L_{Nf}\eta_{Nd}$, where η_d and η_{Nd} are truncated correlated Gaussian noise with mean equal 0.5, standard deviations σ_d and σ_{Nd} and correlation coefficient of ρ_d. We choose $\rho_d = 0.5$ based on analysis of our time-lapse imaging data (Figure 4) and the rest of the parameters of our dual noisy linear map model of cell and nuclear control were fitted on the static pom1 mutant size data using the ABC inference.

Given, our dual noisy linear map model discussed above, in the nuclear scaling model (Figure 5E), we assume transcription rate v to be linearly dependent on the nuclear size L_N but mRNAs are partitioned upon division based on the size of the daughter cells (not nuclear size of the daughter cells). In this model a small daughter cell is likely to inherit a nucleus of average size, with transcription rates higher than expected from cell size, resulting in an increase in mRNA concentration for small cells.

The RNAPII model (Figure 5F) provides a mechanistic RNAPII based model of transcriptional scaling. In this hybrid deterministic and stochastic model, transcription, translation, complex formation and maturation of RNAPII molecules are modeled as simple cell size dependent production steps. The RNAPII is then transported to the nucleus by a nuclear size dependent rate and it binds to DNA with high affinity with a rate that is dependent on the concentration of DNA in the nucleus (inversely proportional to nuclear size). In this model transcription rate of a gene is assumed to be proportional to the amount of DNA-bound RNAPII.

In this hybrid deterministic and stochastic model, RNAPII dynamics are modeled deterministically by a series of ODEs inside growing and dividing cells, while transcription of mRNA is modeled stochastically. Upon cell division, we assume mRNA are partitioned binomially according to the size of the daughter cells. The free cytosolic and nuclear RNAPII are portioned binomially according to the size of the daughter cells and their nucleus. The scaling in this model comes about from sequestration of RNAPII on the DNA. The model is very robust to different model parameters and assumptions as long as the level of free cytosolic and nuclear RNAPII is much smaller than the DNA bound RNAPII. The qualitative model results for Rpb1 shown in Figure 5E, are obtained by using the parameters of the dual noisy linear map model discussed above for the *pom1* mutant, tuning RNAPII parameters to obtain about 10% free RNAPII and linear scaling of DNA bound RNAPII, as well as setting the transcription rate to match expression levels of Rpb1. The model without any further tuning recovers deviation from concentration homeostasis observed at small cell sizes, which is observed for the different genes in the wt and *pom1* mutant strains.

ABC inference and model selection

In this study we have used Approximate Bayesian Computation (ABC) for inference. When the likelihood function is intractable, we require a tool for carrying out inference without it. ABC is precisely such a tool. The algorithm originated in the 1980s and 90 s (see e.g., [85]). For review of more recent developments, see [15]. ABC aims to carry out Bayesian inference without having to evaluate the likelihood function. Given data D and model with parameter set θ, this is done by approximating the posterior distribution:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \approx P(\theta|\rho(D, D_\theta) < \varepsilon),$$

where D_θ is a set of data generated from the model with parameters θ, sampled from the prior $P(\theta)$, $\rho(D, D_\theta)$ is a distance measure that is defined on the set of such datasets (or their summary statistics) and ε is a tolerance, representing the degree of approximation we are willing to accept. The simplest ABC algorithm, that is based on sampling θ repeatedly and rejecting the ones that produce data with larger distance than our tolerance (which is called ABC rejection sampling [86]), is too inefficient. Much work has been carried out over the past decade in this area, leading to a variety of different implementations with much more favorable scaling of computation time with the dimensionality of the parameter space [86]. For the purpose of this project we will use a Sequential Monte Carlo implementation, based on the implementations of Toni et al. [87] (ABC-SMC) and Lenormand et al. [88] (APMC). In the ABC-SMC, one fixes the size of the posterior sample, N, and a finite sequence of decreasing tolerances, $\{\varepsilon_t\}$, *a priori*. The primary differences between APMC and ABC-SMC are first that the sequence of epsilons is not determined *a priori*; it is dynamically determined from the previous iteration's distribution of errors until a stopping criterion $(pacc_{min})$ is fulfilled and second that simulations from earlier iterations are not discarded.

ABC lends itself very naturally to model selection [86, 87]. In essence, all we have to do is to extend our priors to one extra dimension, representing different models. Formally, we require a joint prior distribution over models and parameters, $P(M, \theta)$. We have combined the model selection aspects of ABC-SMC implementation and adaptive aspect of APMC to obtain our APMC with model selection algorithm.

In order to apply our APMC algorithm, we need to choose an appropriate distance $\rho(D,D_\theta)$. As the problem at hand is stochastic in nature, we have chosen to use sum of square differences of summary statistics of the data and simulated data in the distance measure:

$$\rho(D, D_\theta) = \sqrt{\sum_i \left(\frac{m_i(D) - m_i(D_\theta)}{sd(m_i(D))} \right)^2},$$

where we used central sample moments and cross moments $m_i(D)$ of our data up to order 3. With our data sample sizes, moments beyond the first three are usually too noisy to be useful. Also, each term in the distance measure are weighted by the bootstrap estimates of standard deviation of the central moments. This rescales the terms in the sum appropriately and downweighs the noisier moments, helping to prevent overfitting of the data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistical analysis are reported in each figure legend including center and precision measures. Wilcoxon tests on Figure 3 are one-sided and those on Figure S3 are two-sided. Boxplots represent median, interquartile range, and most extreme data points that are not more than 1.5 times the interquartile range. All statistical analysis has been performed in R and mathematical modeling was performed using code implemented in Julia and R. Detailed description of all the mathematical modeling approaches are available in the STAR Methods. Number of cells in each smFISH experiment are provided in Table S2.

DATA AND CODE AVAILABILITY

All smFISH, cell size and nuclear size data are available in Table S3. The accession number for the ChIP-seq sequencing data reported in this paper is ArrayExpress: E-MTAB-8522. Processed ChIP-seq data are available in Table S6. The model posteriors related to Figure 2 are available in Table S4. The ABC code for inference is available at https://github.com/vshahrez/AdaptiveABC.jl. The agent-based simulation codes and simulation results (related to Figures 2, 3, 5, and S3) are available at https://github.com/vshahrez/Sun-Bowman-et-al. The image analysis scripts use on Figures 4 and S4 have been assembled in a python package called PombeTrack available at: https://github.com/ImperialCollegeLondon/PombeTrack/releases/tag/v0.1.