Size-dependent increase in RNA Polymerase II initiation rates mediates gene expression scaling with cell size

Xi-Ming Sun[1,2]*, Anthony Bowman[3]*, Miles Priestman[1,2,3], Francois Bertaux[1,2,3,§], Amalia Martinez-Segura[1,2], Wenhao Tang[3], Dirk Dormann[1,2], Vahid Shahrezaei[3,4] & Samuel Marguerat[1,2,4]

[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] Correspondence to v.shahrezaei@imperial.ac.uk or samuel.marguerat@imperial.ac.uk

* These authors contributed equally to this work.

[§] Present address: Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France.

ABSTRACT

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 biomolecules 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* hybridisation (smFISH), time-lapse microscopy and mathematical modelling 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 rates of both constitutive and regulated genes scale with cell size without evidence for transcriptional bursting. Modelling and experimental data indicate that scaling relies on the coordination of 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 the transcriptional scaling unit. 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, called hereafter “scaling”, is a fundamental and intriguing problem in cell biology [1,2]. Messenger RNAs (mRNA) and proteins are synthesised from the cell DNA genome, which is one of the few cellular components that do not scale with size. Because cell volume increases exponentially and mRNA half-lives are typically short, constant rates of mRNA or protein production cannot lead to gene expression scaling. Recent work, in yeast [3], animal [4,5] and plant cells [6] 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 [3,4,6], although evidence suggests that degradation rates are adjusted early after budding yeast asymmetric division [7] and when growth rate changes [8,9]**.** Scaling is pervasive and only few mRNAs are able to deviate from its regulation [10–12]. Interestingly, two of them have been found to participate in the control of size homeostasis [10,11].

What could be the molecular mechanism behind transcription scaling with cell size? For a gene with an active promoter mRNA numbers follow a Poisson distribution [13]. Transcription is however often discontinuous and periods of RNA synthesis or ‘bursts” alternate with periods of promoter inactivity [14]. 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 [4]. This is compatible with the increased RNAPII occupancy observed genome-wide in large fission yeast cell cycle mutants [3]. This also indicates that the mechanism behind scaling may not be related to activation of transcription but rather to the efficiency of an active promoter. Critically, transcription is a complex process and is regulated at many levels including, RNAPII initiation, pause/release, elongation, and termination [15–18]. 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 from a gene encoded in the small cell genome would increase in response to the large cell environment, however its concentration was almost halved. This suggests that scaling responds to both cell volume and DNA content [4]. This is consistent with the observation that the cell synthesis capacity is split between genome copies in diploid budding yeast cells [10]. Importantly, changes in

74 gene numbers, in the case of increased ploidy for instance, are associated with overall cell
75 size increase in many organisms [1]. This indicates that the number of genes present in a
76 cell is linked to its volume and number of macromolecules. This also suggests that the cell's
77 overall synthetic capacity could be limiting and have a determining role in setting its size
78 [19]. It is therefore likely that scaling depends on a limiting factor involved in transcription but
79 its identity and regulation with the cell volume is not known [19,20].

80 In this study, we measured gene expression in over 20,000 single cells of the fission yeast
81 *Schizosaccharomyces pombe* by single molecule *in situ* hybridisation (smFISH, **Tables S1-**
82 **3**) [21]. We combined these data with agent-based models of growing and dividing cells [22–
83 24], stochastic models of gene expression [13,25,26] and Bayesian inference [27–33] to
84 investigate the quantitative parameters of gene expression that mediate scaling. This
85 integrative approach enabled us to determine which part of the transcription process is
86 scaling with cell size and which molecular event connects transcription with the cell volume.

87

88 **RESULTS**

89 **Gene expression scaling is a single cell attribute of constitutive and inducible gene**
90 **expression.**

91 We first confirmed that scaling of gene expression with cell size was an attribute of single
92 fission yeast cells during rapid proliferation. To do this, we measured mRNA levels of 7
93 constitutively expressed genes in wild-type haploids (*wt*), and in conditional mutants of the
94 Wee1 (*wee1-50*) and Cdc25 (*cdc25-22*) cell cycle regulators by smFISH (**Figure 1A, Table**
95 **S1**). At semi-permissive temperature, mutant cells divide at smaller and larger sizes than *wt*
96 respectively (**Figure S1A**) [3]. Consistent with population microarray data, mean numbers of
97 the 7 mRNAs measured in single cells scale with the average size of the three strains
98 (**Figure 1B**) [3].

99 We then wondered whether scaling could be detected as cells elongate and progress
100 through a normal cell cycle. The fission yeast elongation phase is restricted to G2.
101 Confounding effects arising from changes in genome content during DNA replication are
102 therefore unlikely. Using cell length as a measure of size (**methods**), we analysed
103 expression of the *rpb1* mRNA as a function of single cells' sizes in *wt*, mutant and diploid
104 cells and observed linear scaling within each genotype across a wide range of sizes (**Figure**
105 **1C and 1F**). Linear scaling during the cell cycle was apparent for all the constitutive genes
106 analysed in this study (**Figure S1B-C**). This indicates that gene expression scaling is not an
107 artefact 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 system for studying scaling [3].

We next analysed the relation of scaling with ploidy using diploid cells. Diploids follow scaling parameters (slope, intercept) similar to smaller haploids (**Figure 1C, 1F**). Interestingly, for a given size, mRNA numbers in diploids and in *cdc25-22* haploids were similar (**Figure S1D-F**). Moreover, when analysing 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. 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 (**Figure S1D-F**).

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 potential? To answer this question, we analysed three mRNA that are induced during specific phases of the cell cycle (**Figure S1G)**. All three mRNA showed stronger induction levels in larger cells when comparing *wt* with *wee1-50* and *cdc25-22* mutants (**Figure 1D, 1G**, **S1I and S1J**). This indicates that larger cells have an increased gene expression potential in order to support scaling of constitutive mRNA and inducible transcripts, which were not expressed in smaller cells at the beginning of the cycle.

To investigate this further, we analysed the induction of two genes, which respond to acute changes in external conditions. The uracil regulated gene *urg1* responds to changes in uracil concentration and *sib1* to addition of 2,2-dipyridyl (DIP) [34,35]. Induction of the *urg1* mRNA was heterogeneous and scaled with cell size across size mutants (**Figure 1E, 1H, and S1H**). The *sib1* mRNA showed a much more homogeneous response which also showed clear scaling within and across cell types (**Figure S1K**). This indicates that higher gene expression capacity in larger cells supports scaling 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 from 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

Messenger RNA quantities are regulated at the level of transcription but also post-transcriptionally through modulation of degradation rates [36]. Three studies have reported

142 that mRNA degradation rates are not regulated as a function of cell size in fission yeast,
143 plant and mammalian cells [3,4,6]. To confirm these observations and extend them to single
144 fission yeast cells, we analysed expression of 3 genes by smFISH after transcription
145 inhibition with Thiolutin in *wt*, *wee1-50*, and *cdc25-22* cells. Thiolutin has been shown to
146 inhibit transcription in *S. pombe* efficiently (**Figure S2A**) [11]. We observed mRNA half-lives
147 of around 30-40min for the *rpb1* and *rpb2* mRNAs consistent with previous observations
148 (**Figure S2B**) [37]. In *wee1-50* and *cdc25-22* mutants both mRNAs showed half-lives similar
149 to *wt*, consistent with an absence of scaling of mRNA degradations rates. We also measured
150 mRNA half-lives as a function of the cell cycle using cells binned by size. This analysis did
151 not show consistent positive or negative coordination of degradation kinetics with cell size
152 (**Figure S2B, left**). The absence of scaling of mRNA degradation rates was further
153 confirmed using an orthogonal promoter switch-off approach for the *rbp1*, *rbp2* and *rtp1*
154 genes (**Figure S2B, right**). Finally, as discussed in the next section mathematical modelling
155 and inference do not support scaling of degradation rate. Overall, in agreement with previous
156 studies, our analysis indicates that mRNA degradation is not a major mediator of scaling.

157

158 **Coupling of transcription rates to cell size and not burst frequency mediates scaling.**

159 We next explored the contribution of transcription rates to scaling using measurements of
160 mRNA quantities and mathematical modelling. Using this combined approach allowed us to
161 study dynamic transcription rates from static smFISH measurements of single cells. We
162 developed agent-based models that incorporate the two-state model of gene expression
163 inside growing and dividing cells, which are themselves described by phenomenological
164 models of cell growth and size control (**Figure 2A, methods**). We used an Approximate
165 Bayesian Computation (ABC) approach on cell size and smFISH measurements to infer
166 mechanism of scaling. This inference approach determines the simplest model that captures
167 the statistics of the experimental measurements and returns posteriors of the parameters
168 and model probabilities for all models (**methods**). We used 2 classes of gene expression
169 models that are the limiting cases of the two-state model shown in **Figure 2A**. The first class
170 describes transcriptional bursts explicitly, while the second assumes only non-bursty
171 transcription kinetics with a simple birth-death process that produces Poisson distributions
172 (**Figure 2A**). Each class is in turn composed of two models the first assuming constant
173 transcription rates during the cell cycle and the second assuming transcription rates that
174 increase linearly ("scale") with cell size (**methods**).

175 We performed model selection between these four models on the smFISH data for 7
176 constitutively expressed genes in the *wt*. This analysis generated two clear conclusions.

177 First, models assuming bursty transcription were strongly penalised and supported by small
178 model probabilities (**Figure 2B**, compare proportion of blue and red colours). Second,
179 among the Poisson models, the one assuming transcription scaling with size was preferred
180 for all but one mRNA (*rpb3* in **Figure 2B**, compare orange and dark red colours). From this
181 we conclude that transcription rates of constitutively expressed genes in fission yeast are not
182 bursty and scale with cell size.

183 We then asked whether transcription regulated by the cell cycle or by external cues followed
184 the same paradigm. We found that, as constitutively expressed genes, cell-cycle regulated
185 mRNAs do not follow a bursty transcription dynamics (**Figure 2B, middle**). This indicates
186 that transcription, even when regulated in defined sections of the cell cycle falls into the
187 Poisson regime. In terms of scaling, model selection is, as for the constitutive case, in overall
188 support of scaling of transcription rate with cell size (**Figure 2B, middle**). Messenger RNAs
189 that respond to external stimuli showed a different picture. Model selection favoured the
190 bursty model with support for burst size scaling. This could point to bursty transcription of
191 inducible genes, to the presence of strong extrinsic noise, or to a scenario where cells
192 respond heterogeneously to the external signals.

193 In order to validate the model selection analysis, we generated simulations for all constitutive
194 and cell-cycle regulated genes from **Figure 2B**, using models where transcription rates scale
195 with cell size and follow a Poisson regime. This analysis reproduced quantitatively the mean
196 and coefficient of variation (sd/mean, CV) of all experimental smFISH measurements, which
197 illustrated the non-bursty nature of transcription for these genes further (**Figure 2C**). We
198 then used the inferred parameters for the Poisson models to simulate gene expression of
199 multiple genes in single cells and compared them to experimental data where 3 different
200 mRNAs were measured in each cell (**methods**). This analysis suggests that most gene-to-
201 gene correlations in expression are explained by scaling of transcription rates with cell size
202 and may not require extensive additional regulation (**Figure 2D**). However, our data also
203 suggest that for some gene pairs (e.g. Rpb1-Rpb2), correlations can be explained better by
204 including some additional extrinsic variability.

205 We finally asked whether our modelling approach was also in support of a negligible role of
206 mRNA degradation in scaling. We performed further model selection and simulation
207 analyses on the transcription shut-off experiment from **Figure S2B** using models that
208 consider different possible scalings of transcription and degradation rates (**methods**). This
209 analysis shows that the model assuming transcriptional scaling with constant rates of mRNA
210 degradation with cell size is overwhelmingly chosen (**Figure 2E**) and could capture
211 experimental data quantitatively confirming our initial analysis (**Figure S2B**).

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

219

220 **Coordination of RNAPII initiation rates with cell size as the main mechanism of**
221 **scaling.**

222 We then wondered which specific aspect of the transcription process was directly
223 coordinated with cell size to mediate scaling. To investigate this and to confirm model
224 predictions from **Figure 2**, we measured transcription rates experimentally. Single cells'
225 transcription rates can be estimated from smFISH images by measuring intensities of
226 nuclear transcription sites [38]. As a model, we used probes directed against the 5' region of
227 the *rbp1* mRNA. This probe design provides strong sensitivity for detection of nascent
228 transcripts (**Figure S3A, S3B**). Transcription site intensities and cell size were overall
229 positively correlated when comparing *wt*, *wee1-50* and *cdc25-22* cells confirming modelling
230 results and previous observation in a metazoan cell line (**Figure 3A**, **3B**, **S3C**) [4]. We then
231 investigated the impact of increased ploidy on transcription rates scaling. For this we
232 compared nascent site intensities of 4 mRNA in haploid cells with those of diploid strains.
233 We used probes against the 5' end of *rpb1* as above and against three mRNA induced in
234 specific phases of the cell cycle, since inducible expression increases the sensitivity of
235 nascent site analysis. Intensities of individual transcription sites were similar between
236 haploid and diploid cells (**Figure 3C**, green and yellow boxes). However, when considering
237 the total intensity per cell, diploids showed increased rates and scaling was apparent again
238 (**Figure 3C**, orange boxes). This indicates that the cell transcriptional capacity is limiting and
239 distributed between the two gene copies of diploids. This was confirmed by analysis of the
240 nascent intensity of the *rbp1* mRNA in a heterozygous deletion strain which showed no
241 increased rates in the remaining copy (**Figure 3C**).

242 To investigate further the mechanism behind scaling of transcription rates, we designed an
243 orthogonal modelling approach where transcription is modelled as RNAPII particles hopping
244 on a gene represented by a lattice (**Figure 3D**). This approach, which is based on a totally
245 asymmetric simple exclusion process (TASEP), has been used successfully to study
246 transcription and translation [6,39–42]. In our model, a gene is represented by a lattice of

247 length L_G and non-bursty transcription is modelled using three rates: i) the transcription
248 initiation rate α is the rate at which RNAPII molecules enter the first site of the lattice (gene);
249 ii) the elongation rate β is the rate at which RNAPII molecules hop one site forward on the
250 lattice (gene); and iii) the termination rate γ is the rate at which the RNAPII molecule which
251 sits on the last site of the gene leaves the lattice and produces a full length mRNA (**Figure**
252 **3D**). We incorporated this model in the agent-based framework from **Figure 2** assuming
253 each rate could be linearly coupled to cell size.

254 By sampling the rates α, β and γ over physiological time scales estimated from previous
255 studies (**Methods**) we found that coupling of initiation rates with size produced the most
256 robust linear scaling (**Figure 3E**) and the strongest positive correlation of nascent intensities
257 with size (**Figure S3D**). Although a model coupling transcription elongation rates with cell
258 size could also generate linear scaling in some parameter regions (**Figure 3E, S3D**), these
259 required elongation rates to be much slower than experimental measurements either in
260 yeast or metazoans which report elongation rates of around 2kb/min (**Figure S3F**) [38,43–
261 47]. Importantly, linear scaling could only be observed in regimes with slow initiation rates
262 relative to elongation and termination indicating that initiation is rate limiting (**Figure S3E**).
263 Interestingly, this also suggests that fast, non-limiting initiation rates could be a mechanism
264 by which some genes escape scaling (e.g. *rpb3*; **Figure 2B**). Finally, these results were
265 confirmed by ABC inference analysis of nascent sites intensities data using the same
266 TASEP model for 3 genes in different strains which showed clear preference for the initiation
267 model (**Figure S3G**). This *in silico* analysis indicates that scaling of initiation rates with cell
268 size is the likely mechanism of scaling.

269 The initiation model generated two important predictions. First, the model predicts that, even
270 for non-bursty genes, cells in a population are not all actively transcribing at all times and the
271 frequency of active transcription should increase with cell size. To test this prediction, we
272 compared the fraction of cells with a nascent transcription site ("active cell fraction") in *wee1-*
273 *50*, *wt*, *cdc15-22* and diploid cells. As predicted by the model, a clear increase in the fraction
274 of transcriptionally active cells with size could be observed (**Figure 3F**). Moreover, a strong
275 positive correlation of the active cells fraction with cell length was apparent when calculated
276 in sliding windows of increasing cell numbers during the normal cell cycle (**Figure S3I**). The
277 second prediction of the initiation scaling model is a strong positive correlation between the
278 number of transcribing RNAPII and cell size (**Figure S3D**). To test this, we analysed RNAPII
279 occupancy across the genomes of *wt*, *wee1-50* and *cdc25-22* cells by chromatin
280 immunoprecipitation followed by next generation sequencing (ChIP-seq). We used three
281 antibodies against total RNAPII and Serine 5 phosphorylation of its carboxy terminal domain

282 (CTD, **methods**). Using data normalised to occupancy at histone genes (**methods**) we
283 could observe a significant increase in overall RNAPII occupancy with size consistent with
284 previous observation and confirming model predictions (**Figure 3G**) [3]. Importantly, we did
285 not find evidence for a redistribution of RNAPII from the 5' to the 3' of genes in larger cells
286 which makes regulation of scaling at the level of RNAPII pause/release unlikely (**Figure 3H,**
287 **S3J**). This profile was also reproduced quantitatively in our minimal TASEP model (**Figure**
288 **3D**, data not shown). In summary, our *in silico* and experimental data indicate that
289 transcriptional scaling is mediated by an increase in RNAPII initiation rates coordinated with
290 cell size. In addition, our modelling data together with the increase in RNAPII occupancy
291 observed in larger cells suggest that RNAPII could be a limiting factor for transcription as cell
292 size increases.

293

294 **RNAPII amounts on chromatin increases with cell size.**

295 If scaling of initiation rates is the mechanism behind scaling and RNAPII is limiting, the
296 amount of RNAPII complexes on the genome of single cells should increase during the cell
297 cycle. To test this hypothesis, we measured localisation of RNAPII in single cells by live-cell
298 imaging. To do this, we tagged components of the RNAPII complex with green fluorescent
299 protein (GFP) and imaged them during the cell cycle (**Figure 4A, S4A**). First, we observed
300 that the cellular concentration of Rpb1 and Rpb9, two RNAPII subunits, remained constant
301 during the fission yeast G2 growth phase (**Figure 4B, S4B**). This indicates that scaling of
302 transcription initiation is not controlled by regulation of the cellular concentration RNAPII. Our
303 image data show that a very large fraction of the tagged RNAPII subunits localises in the
304 nucleus in the DAPI-positive area (**Figure 4A**). We therefore asked whether the amount of
305 RNAPII on chromatin changes with cell size during the cell cycle. To assess this, we
306 measured fluorescence intensities of the nuclear region occupied by RNAPII subunits
307 (**Figure 4C, S4C**). Strikingly, the signal for both Rpb1 and Rpb9 increases steadily during
308 fission yeast G2 phase where most cell size increase occurs (**Figure 4C, S4C**). Importantly
309 this was not the case for another DNA bound protein the histone Hta2 (H2A beta), (**Figure.**
310 **4D-E**). This is consistent with the prediction of the initiation scaling model in the previous
311 section and indicates that RNAPII quantities are likely to be limiting for transcription. In
312 summary, these experiments indicate that increased initiation rates that mediate scaling are
313 sustained by efficient import of RNAPII in the nucleus together with rapid recruitment onto
314 chromatin.

315

316 **The nucleus is the scaling unit**

317 In fission yeast and other organisms, nuclear and cytoplasmic volume are intimately
318 connected [48]. As scaling of initiation rates depends on RNAPII levels in the nucleus we
319 wondered whether nuclear size rather than cell size itself could be the quantitative
320 determinant of scaling.

321 To test this idea, we analysed nascent site intensities of *cdc11-119* mutant cells cultivated at
322 non-permissive temperature. Under these conditions, cells elongate, undergo mitosis and
323 nuclear division but do not divide (**Figure 5A**) [49]. Strikingly, scaling of nascent site
324 intensities of individual nuclei with cell size was not apparent in this system ($R_{Pearson}$ = 0.06,
325 **Figure 5B, right**). Scaling was only restored when all nascent sites present in a cell were
326 added together ($R_{Pearson}$ = 0.46, **Figure 5B, left**). This mirrors data from multinucleated cells
327 showing that while the overall nuclear volume scales with cell volume, the volume of
328 individual nuclei is proportional to their surrounding cytoplasmic volume [48]. Consistent
329 with this, the ratios of nascent site intensities between nuclei of *cdc11-119* cells were weakly
330 correlated with their immediate cytoplasmic volumes (**Figure S5A**). This indicates that the
331 cell "scaling unit" could be the nucleus and not the whole cell.

332 We next analysed scaling in conditions where the correlation between cell and nuclear size
333 is compromised. Pom1 is a regulator of cell polarity and division which when deleted leads to
334 increased size variability at cell birth due to cell partitioning errors [50]. We analysed
335 expression of three mRNA by smFISH in *pom1Δ* mutant cells expressing a marker of the
336 nuclear envelope to allow measurement of nuclear size (**Figure 5C**). As expected cell and
337 nuclear size show smaller correlation in the *pom1Δ* mutants compared to *wt* cells (**Figure**
338 **S5B).** Scaling in this system was comparable to *wt* cells with the exception that *pom1Δ*
339 mutants showed a higher y-axis intercept when mRNA numbers are plotted as a function of
340 cell size (**Figure S5C**). This indicates that mRNA concentrations in *pom1Δ* cells are higher in
341 smaller cells after birth and negatively correlated with cell size (**Figure 5D**). This deviation
342 from perfect concentration homeostasis is also observed but to a lower degree in *wt* cells
343 (**Figure 5D**) as reported for mammalian cells [4]. Moreover, mRNA numbers divided by
344 nuclear volume show a negative correlation with nuclear volume (**Figure S5D)**. The
345 modelling approach described in **Figure 2** which is based on transcription scaling with cell
346 size and binomial partitioning of mRNAs based on daughter cell sizes failed to capture this
347 behaviour (**Figure 5E, magenta line**). However, a modified model coupling transcription
348 rates to nuclear size instead produced a good fit to the data (**Figure 5E, S5F green line**).
349 Together, this analysis suggests that beside the nucleus being the scaling unit, nuclear size
350 could be an important determinant of scaling.

351

352 **A mechanistic model of scaling**

353 We used the results from this study to develop a mechanistic model of scaling centred
354 around RNAPII mediated transcription (see also [19,20]). (**Figure 5F**, **methods**). In this
355 model the rates of RNAPII complex synthesis and maturation scale with cell size. This is
356 consistent with our smFISH and live-cell imaging data showing that RNAPII subunits have a
357 constant cellular concentration during the cell cycle (**Figure 1 and 4**). RNAPII is then
358 transported to the nucleus with a rapid rate that is depleting cytosolic RNAPII (as observed
359 in live-cell imaging data; **Figure 4)**. Once in the nucleus, RNAPII binds to DNA with a
360 constant high affinity and transcription rates are proportional to the numbers of DNA-RNAPII
361 complexes present on each gene at any given time consistent with scaling of initiation rates
362 (**Figure 3**). Finally, RNAPII levels is set to be limiting in line with the initiation scaling model
363 (**Figure 3**), the behaviour of diploid and heterozygous mutants (**Figure 3**) [10] and with the
364 fact that the cell synthetic capacity is titrated against the number of genes in heterokaryons
365 [4]. Moreover, this assumption fits the observation that many RNAPII subunits are limiting for
366 growth in fission yeast [51–53]. The high affinity and the limiting amount of RNAPII ensures
367 that the majority of RNAPII molecules are bound to DNA and increase with cell size
368 consistent with our imaging data (**Figure 4**) and with biochemical evidence from mammalian
369 cells [4]. Finally, DNA replication occurs close to cell division and each daughter cell inherits
370 about half of the DNA-bound RNAPII mostly independent of its size as observed in our live
371 cell imaging experiment (**Figure 4C, S4C**). This simple model is able to capture the different
372 features of our data. First, it retrieved the scaling of DNA-bound RNAPII with cell size while
373 keeping the overall cellular concentration constant as observed in **Figure 4** (**Figure S5E**).
374 Second, it explained the scaling of mRNA numbers with cell size including the higher mRNA
375 concentration observed in smaller cells (**Figure 5E**). As this mechanistic model relies on
376 both cell size mediated production of RNAPII and nuclear mediated partitioning of RNAPII it
377 captures better the negative correlation between mRNA numbers over nuclear area and cell
378 size than a purely nuclear size mediated transcriptional scaling (**Figure S5E**, compare blue
379 and green line). Overall, this simple mechanistic model of RNAPII transcription by
380 integrating the experimental findings from this study, explains the origin of transcriptional
381 scaling and its subtler features and identifies the competition between genes for the limited
382 pool of transcriptional machinery as central to the phenomenon of transcriptional scaling.

383

384 **CONCLUSIONS**

385 We performed an extensive experimental and modelling study of gene expression scaling
386 with cell size in single fission yeast cells. We found that scaling is a pervasive feature of

387 gene expression that impinges on constitutive and regulated expression. We then showed
388 that scaling relies on an increase in RNAPII initiation rates with cell size and a concentration
389 independent recruitment of a limiting RNAPII on the genome. Finally, we propose that the
390 nucleus is the scaling unit and that nuclear size may participate in setting scaling levels.

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

398 Our model assumes that the general and specific transcription factors that regulate relative
399 promoter strength are in excess even in small cells, and thus are not affected by cell size. To
400 gain a deeper mechanistic understanding of scaling, it will be important to determine whether
401 this assumption is met for most regulators or if some escape the rule. It could also inform
402 about mechanisms through which some mRNA escape regulation by scaling as our
403 modelling results suggest that non-limiting initiation rates can produce this behaviour [10–
404 12]. Another interesting question will be to determine the role of chromatin remodellers in
405 facilitating transcription initiation in larger cells. It is possible that a more permissive
406 chromatin environment in large cells synergises with increased RNAPII local concentrations
407 to support higher transcription rates of inducible genes.

408 Recent evidence suggests that, in mammals, burst size is regulated at the level of the
409 proximal promoter sequence, while distal enhancers are involved in setting burst frequency
410 [54]. Moreover, burst initiation and RNAPII pause/release but not RNAPII recruitment have
411 been shown to be regulated in response to biological perturbations [55]. Our model of
412 scaling through initiation of RNAPII transcription fits well with these data as this process is
413 independent of both gene activation and control of burst frequency. It is also consistent with
414 the observation that promoters maintain their relative expression levels in changing growth
415 conditions [56].

416 Our observation that scaling is regulated at the level of single nuclei in multinucleated cells
417 and may be linked to nuclear size is interesting. RNA synthesis levels have been connected
418 to nuclear size in other systems such as multinucleated muscle cells [57], or for the HTLV-1
419 mRNA [58]. It could reflect a higher availability of RNA polymerases around larger nuclei as
420 these tend to be surrounded by larger cytoplasmic volume (**Figure S5A**) [48,57].
421 Interestingly, the nucleus was also found to be an independent transcriptional unit in mature

422 osteoclasts [59] and in multinucleated fungi where nuclei retain local control of cell cycle
423 periodic transcription [60,61]. This suggests that more complex feedback and molecular
424 mechanisms may also be at play.

425 An important result from our study, which is not directly related to scaling, is the absence of
426 bursty transcription for most fission yeast genes tested. This is in line with previous
427 observations in budding yeast and plants [29,38,62]. Transcriptional bursts result in high
428 gene expression noise and are associated with Fano factors ($\frac{\sigma^2}{\mu}$ of mRNA numbers) greater
429 than 1, whereas Poissonian birth-death processes have a Fano factor = 1. Our finding that
430 most transcription followed a non-bursty regime relied on our modelling taking cell size and
431 the cell cycle into account explicitly. Without doing so, all genes in this study have been
432 thought to have bursty expression as Fano factors calculated on the raw counts were well
433 above 1 (not shown). This reiterates the importance of studying gene expression considering
434 potential confounding effects of morphological features such as cell size and the cell cycle
435 [23,24,29,63,64].

436 Finally, in addition to progressing our understanding of the mechanisms behind scaling, this
437 study provides a large quantitative dataset of gene expression and cell size measurement in
438 over 20000 cells in various conditions. This will support future modelling efforts aimed at
439 understanding regulation of gene expression.

440

441 **ACKNOWLEDGEMENTS**

442 We would like to thank Nick Rhind for stimulating discussion and his help with formulating
443 our model of scaling. We are grateful to Kurt Schmoller, Philipp Thomas, Marc Sturrock and
444 Nick Rhind for critical reading of the manuscript. We thank Snezhana Oliferenko and Jürg
445 Bähler for sharing strains. This research was supported by the UK Medical Research
446 Council, a Leverhulme Research Project Grant (RPG-2014-408). It used the computing
447 resources of the UK Medical Bioinformatics partnership (UK MED-BIO), which is supported
448 by the UK Medical Research Council (grant MR/L01632X/1) and the Imperial College High
449 Performance Computing Service. ChIP-seq data have been deposited in ArrayExpress,
450 accession number XXXXXX.

451

452

453

454 **METHODS**

455 **Strains and culture conditions**

456 The strains and their genotypes that were used in this study are listed in **Table S1.** Genetic
457 crossing confirmed by polymerase chain reaction was used for strains generated unless
458 otherwise specified. Strains were revived from glycerol frozen stocks on solid yeast extract
459 agar (YE agar), or YE agar supplemented with 25 mg l^{-1} adenine, L-histidine, L-leucine,
460 uracil, L-lysine, and L-arginine (Sigma), and with or without appropriate antibiotics for
461 selection. YE agar plates were incubated for approximately 48 h at 32°C in a static incubator
462 until visible large colonies could be observed. Single colonies were transferred into liquid
463 yeast extract medium (YE), in YE supplemented as above (YES), Edinburgh minimal
464 medium (EMM), or EMM supplemented as above (EMMS), unless otherwise indicated in
465 figure legends, and incubated at 170 rpm in a shaking incubator. Temperature sensitive
466 strains were grown at 32°C and shifted to 36.5°C for the time indicated in figure legends. For
467 the induction of *sib1* expression, the strains were grown at 25°C to an optical density at 600
468 nm (OD_{600}) of ≈0.4 and treated with 2,2-dipyridyl (DIP; ACROS) at a final concentration of
469 250 μM for the time indicated in figure legends, or left untreated. For measuring mRNA
470 decay rates, cells were grown in YE at 25°C to OD_{600} ≈0.4; cells were treated with thiolutin
471 (AXXORA) for the time indicated at a final concentration of 15 μg/ml, or left untreated. For
472 *urg1* induction, cells were grown in EMM supplemented with or without 0.25 mg/ml uracil for
473 the time indicated. For transcription inhibition, log phase cultures (OD600~0.5) were treated
474 with thiolutin (15 ug/ml) and same volume of DMSO (used for dissolve thiolutin) was added
475 to thiolutin untreated culture. Samples were taken at 0, 25, 35, 45 mins and processed as for
476 smFISH. For live-cell experiments, cells were grown in EMMS in syphonstats – chemostat-
477 like devices (http://klavinslab.org/hardware.html) which maintain the turbidity of liquid
478 cultures by diluting with fresh medium appropriately – and maintained at OD_{600} 0.4 at 32°C
479 by frequent dilution [65].

480 **RNA single molecule fluorescence in situ hybridisation (smFISH)**

481 All smFISH datasets are described in **Table S2**. All the mRNA counts, nascent site
482 intensities and cell size measurements are available in **Table S3**. smFISH samples were
483 prepared according to a method modified from published protocols [66,67]. Briefly, cells
484 were fixed in 4% formaldehyde and the cell wall partially digested using zymolyase. Cells
485 were permeabilised in 70% ethanol, pre-blocked with bovine serum albumin and salmon
486 sperm DNA, and incubated overnight with custom Stellaris oligonucleotide sets (Biosearch
487 Technologies) labelled with CAL Fluor Red 610, Quasar 670, or Quasar 570 (probe
488 sequences are listed in **Table S4**). Cells were mounted in ProLong™ Gold antifade

489 mountant with DAPI (Molecular Probes), and imaged on a Leica TCS Sp8 confocal
490 microscope, using a 63x oil objective (NA 1.40). Optical z sections were acquired (0.3 μm
491 step size) for each scan to cover the entire depth of cells. Cell boundaries were outlined
492 manually and single mRNA molecules were identified and counted using the FISH-quant
493 MATLAB package [68]. Cell area, length, and width were quantified using custom ImageJ
494 macros. The technical error in FISH-quant detection was estimated at 6–7% by quantifying
495 *rpb1* mRNA foci with two sets of probes labelled with different dyes. The nascent mRNA foci
496 were identified and quantified using intensity information from the above-mentioned FISH
497 quantification where an intensity 2.5 to 3-fold above the modal intensity within the same cell
498 was chosen as a threshold for nascent mRNA. The quality of the identification of nascent
499 sites was validated manually by visualising high intensity foci in the nucleus, with an
500 accuracy of over 90% in all three strains (wild-type, *cdc25-22*, and *wee1-50*).

501 **ChIP-seq**

502 Chromatin immunoprecipitation (ChIP) assays was carried out essentially according to
503 published methods [69]. In brief, cells were grown in YES to an OD600 of ~ 0.8 and fixed
504 with formaldehyde solution (1% final) and then quenched with glycine. After washing twice
505 with cold PBS (phosphate-buffed saline), cells were re-suspended in lysis buffer containing
506 proteases inhibitors and disrupted vigorously with acid-washed glass beads 8-11 times for
507 20 sec in a FastPrep instrument. Samples were then sonicated in Bioruptor (at High setting
508 and 6 times 5 mins with 30 secs ON/30 secs OFF). Chromatin were immunoprecipitated with
509 antibody against Rpb1 (ab817) or Rpb1 CTD-ser 5 (ab5408 or sigma 04-1572), which were
510 coupled to Dynabead protein-G and protein-A and Dynabead sheep anti-mouse or rat IgG
511 (Invitrogene). DNA was purified from immunoprecipitated samples using MinElute Qiagen
512 kit. Quantification of the DNA was done using QuDye dsDNA HS assay kit and quality was
513 verified using Bioanalyzer.

514 For sequencing DNA from immunoprecipitated samples, the libraries were made using the
515 NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (E6240S) with the indexes
516 provided in NEBBext Multiplex Oligos for Illumina (Index Primers Set1,2 and 3). Negative
517 control DNA are those from the same chromatin extracts without going through
518 immunoprecipitation steps. Pools of libraries were sequenced on an Illumina HiSeq 2500
519 instrument at the MRC LMS genomics facility. Paired-end reads (100 nt) were generated
520 from two pools of 12 or 18 samples per sequencing lane. Data were processed using RTA
521 1.18.64, with default filter and quality settings. The reads were de-multiplexed with bcl2fastq-
522 1.8.4 (CASAVA, allowing 0 mismatches).

523 **ChIP-seq analysis**

524 A description of ChIP-seq libraries can be found in **Table S5**. Sequencing reads were
525 aligned to the fission yeast genome as available in PomBase in July 2019 using BWA
526 [70,71]. For figure 3G, RNAPII occupancy counts were extracted for each transcript using
527 HTseq [72] and the fission yeast annotation available in PomBase in July 2019 [71]. Data
528 were normalised using DESeq2 and scaled using the mean counts of fission yeast histone
529 genes as a scaling factor to allow comparison of global RNAPII occupancy between size
530 mutants [73]. Amounts of histone proteins are thought to scale with DNA content rather than
531 cell volume and are commonly used as normalisation factors for absolute proteomics
532 measurements [74]. Moreover, average synthesis rates of histones were found to remain
533 constant across a wide range of sizes in budding yeast (Kurt Schmoller personal
534 communication). For figures 3H and S3I, RNAPII immunoprecipitation data were normalised
535 with their respective input and average gene analysis was performed using the deeptools
536 analysis suite [75].

537 **Live-cell microscopy and analysis**

538 Strains of interest and wild-type ySBM2 were grown from single colonies in 5 ml YES before
539 they were transferred into syphonstats and maintained at OD_{600} 0.4 overnight in EMMS.
540 Prior to microscopy, ySBM2 cells were mixed at a 1:10 ratio with each strain of interest and
541 diluted to a final OD_{600} of 0.3 in fresh EMMS. Cells were loaded directly into a CellASIC®
542 ONIX Y04C-02 microfluidic plate (EMD Millipore) according to the manufacturer instructions.
543 Fresh EMMS was continually perfused through the growth chamber with a constant pressure
544 of 6.9 kPa (approximately 3 µl/h). Cells were imaged on an Olympus IX70 inverted widefield
545 fluorescence microscope with an environmental chamber maintained at 32°C, with a high
546 precision motorised XYZ stage (ASI), controlled with µManager version 1.4.22 [76]. Cells
547 were continually imaged at a 10 minute interval with a 40× objective (NA 0.95, UPlanSApo;
548 Olympus) with brightfield (30 ms exposure), GFP (250 ms exposure, emission filter Semrock
549 514/30 nm), and dsRed (500 ms exposure, emission filter Semrock 617/73 nm) channels
550 captured by a Hamamatsu Orca Flash 4.0 V2 sCMOS camera, with illumination provided by
551 a Lumencor Spectra X LED light source set to 20% power. For each of the four growth
552 chambers within the microfluidic plate, three positions were defined, each of which was
553 focussed using the software autofocus using the brightfield channel; since this feature was
554 generally inaccurate, a 3 µm Z-stack was used around the autofocus position to ensure that
555 at least one Z-position was in focus for each position.

556 Initial analysis was performed with Fiji – a derivative of ImageJ – with time-lapses assembled
557 if necessary, and in-focus slices for each time-point selected using a custom macro (written
558 by Stephen Rothery, Facility for Imaging by Light Microscopy, Imperial College London),
559 resulting in a 4-dimensional OME-TIFF file for each field-of-view (XYCT). Each file was

560 subsequently analysed using a series of custom Python scripts utilising the scikit-image,
561 SciPy, NumPy, and pandas packages extensively among others. The scripts permit the
562 semi-automated definition of cell boundaries (segmentation) within the brightfield channel,
563 followed by the quantification of fluorescence within cell boundaries, identification of nuclei,
564 further quantification within nuclear boundaries, and assignment of cells into lineages. Cell
565 segmentation was effected using a custom 'balloon-filling' algorithm, in which a connected
566 series of nodes is 'inflated' from its centre inducing an outwards force on all nodes; this
567 outward force is counteracted by an 'image force', which applies an opposing force inversely
568 proportional to the intensity of pixels neighbouring the node, this has the effect of preventing
569 nodes from expanding through areas of low light intensity – which generally surrounds the
570 cell boundary; finally each node is also affected by its direct neighbouring nodes which pull
571 each node sideways according to their position, ensuring smooth contours. Together, after
572 multiple iterations, nodes migrate from a central location until the forces equilibrate at areas
573 of low light intensity (generally the edge of the cell). This procedure can be performed in an
574 automated manner in which iterations cease when the area contained within the nodes does
575 not significantly change, or in a supervised manner, in which further iterations are prompted
576 manually via a keyboard command, with progress displayed via a graphical interface. Initial
577 centres of cells are defined either by manually clicking, or by the centre of the cell in the
578 previous frame. Nuclei are defined as areas within the cell boundary which have red
579 fluorescence pixel intensity values greater than 1.1 standard deviations above the mean
580 intensity within the whole cell boundary. Time from mitosis is defined as the number of hours
581 from the point at which the number of detected nuclei increases from one to two.
582 Fluorescence intensity is adjusted for uneven illumination according to a series of empty
583 fields imaged using the same settings. Background and autofluorescence is determined from
584 wild-type cells cultured within the same field-of-view, with the mean level of fluorescence
585 within these cells subtracted from measurements. Fluorescence is normalised by cell or
586 nuclear area by dividing total fluorescence of all pixels within their respective boundary by
587 the area of that boundary. Scripts are available upon request.

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

594 Image analysis was semi-automated and performed using custom interactive MATLAB
595 scripts available upon request. Tracks corresponding to individual cell cycles are recorded

596 by user clicking. Only cell cycles for which cells remained in in focus for at least the 2 hours
597 preceding mitosis were collected. Local background subtraction was automatically applied to
598 fluorescence images. Cell segmentation was performed automatically based on thresholding
599 of the brighfield images. Nuclear segmentation was performed automatically based on
600 thresholding of the uch2-mCherry fluorescence images.

601 **Cell size measurements**

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

610 $$\text{volume} = 4/3 * \pi * (\text{area}/\pi)^{3/2}$$

611

612 **Mathematical modelling**

613 We use agent-based simulations of stochastic gene expression coupled to cell size in
614 growing and dividing cells (**Figure 2A**) [24]. We assume cells grow exponentially with a
615 constant growth rate from birth to division that is sampled from a truncated Gaussian
616 distribution with mean m_g and standard deviation σ_g. A cell that is born with birth length L_i
617 grows until it reaches the division length L_f. We use a phenomenological model of cell size
618 control that relates the final size to initial size through a noisy linear map, which captures
619 experimentally observed variability and correlations in cell size [77–79].

620 $$L_f = aL_i + b + \eta_b \qquad (1)$$

621 where a and b are size control parameters ($a = 0$ denotes a sizer mechanism and $a = 1$ an
622 adder mechanism) and η_b is a truncated Gaussian with mean zero and standard deviation
623 σ_b. The dividing cell of length L_f produces two daughter cells with birth sizes L'_i and $L_f - L'_i$,
624 where $L'_i = L_f \eta_d$ and η_d is another truncated Gaussian with mean 0.5 and standard
625 deviation σ_d. The biomolecules such as mRNA molecules (except for DNA) are binomially
626 partitioned in the daughter cells with a probability proportional to the daughter cell size
627 L'_i. As shown in **Figure 2A**, we simulate a fixed number of cells, so upon cell division one of

628 the existing cells (including one of the newly born daughter cells) is chosen randomly and
629 taken out of the simulation. This ensures we are simulating a constant number of cells in
630 time and can produce snap-shot data with the correct cell age and size distribution as
631 observed in the experimental data. This has been used in the simulations used in the ABC
632 inference (**Figure 2**). For the modelling results shown in **Figure 3 and 5**, where the results
633 are conditioned on cell size, we have used a simpler scheme [24], where upon cell division
634 only one of the daughter cells is followed modelling a single lineage (similar to the data
635 generated in a mother machine) [80]. The simulations are performed using a simple
636 algorithm that uses discretised time steps to simulate exactly the Gillespie method [81] with
637 time-dependent parameters [82]. The simulation code for inference is written in the Julia
638 programming language and the rest of simulations are performed in R. The codes are
639 available upon request.

640 Our main gene expression model is the so-called telegraph model or the two-state model
641 [83] where the gene can be in an 'off' or an 'on' state and transcription can only occur when
642 the gene is on (**Figure 2A**). If the gene is always on then this model reduces to a simple
643 non-bursty birth-death process with parameters transcription rate v and mRNA decay rate d.
644 Here the mRNA counts per cell have a Poisson distribution (in the absence of cell cycle
645 effects). In the limit where the duration of the promotor on-state is shorter than the mRNA life
646 time we have the bursty limit characterised by 3 parameters of average burst frequency k_{on},
647 average burst size v/k_{off} and mRNA decay rate d. In this model, birth events are
648 simulated as geometrically distributed increases in mRNA numbers [84]. For **Figure 2B**
649 model selection, we used four variants of this model including the Poisson and bursty limits
650 with or without transcriptional dependence to cell size. Transcriptional scaling is modelled as
651 linear dependence of transcription rate or burst size to cell size ($v(L) = v_0 L$, where, v_0 is a
652 constant and L,is the cell length). The models for the cell cycle regulated genes, assume that
653 there is a point in the cell cycle, where gene expression increases from a basal level to an
654 active level. ABC model selection in **Figure 2E** is performed on the data from the
655 transcription shut-off experiments of the 3 strains of *wt*, *wee1-50*, and *cdc25-22.* The 3
656 models included are all the Poisson limit with different scaling assumptions for transcription
657 and decay rates. Model one assumes transcriptional scaling with a single constant decay
658 rate across the 3 strains, the second model assumes also transcriptional scaling but with 3
659 different constant decay rates within each strain. And the third model assumes constant
660 transcription and decay rate that is proportional to inverse cell size across the 3 strains. The
661 priors used in the model selection, are wide over a physiological range. The model selection
662 results were not too senstivie 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) [6,39–41]. Here, the promoter is assumed to be always active, i.e. we are modelling a non-bursty gene. The gene is modelled 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 modelled as hopping of RNAPII forward with rate β if the next site is available. Termination is modelled 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, elongation or termination rate to cell size. 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 time scale τ_I, elongation time scale through the whole gene τ_E, and termination time scale 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 time scale 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 [cite]. The rates are inversely proportional to the time-scales as $\alpha = 1/\tau_I$, $\beta = L_G/\tau_E$ and $\gamma = 1/\tau_T$. 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. 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 modelling 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:

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

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}

697 and correlation coefficient of ρ_d. We choose $\rho_d = 0.5$ based on analysis of our time-lapse
698 imaging data (**Figure 4**) and the rest of the parameters of our dual noisy linear map model of
699 cell and nuclear control were fitted on the static pom1 mutant size data using the ABC
700 inference.

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

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

715 In this hybrid deterministic and stochastic model, RNAPII dynamics are modelled
716 deterministically by a series of ODEs inside growing and dividing cells, while transcription of
717 mRNA is modelled stochastically. Upon cell division, we assume mRNA are partitioned
718 binomially according to the size of the daughter cells. The free cytosolic and nuclear RNAPII
719 are portioned binomially according to the size of the daughter cells and their nucleus. The
720 scaling in this model comes about from sequestration of RNAPII on the DNA. The model is
721 very robust to different model parameters and assumptions as long as the level of free
722 cytosolic and nuclear RNAPII is much smaller than the DNA bound RNAPII. The qualitative
723 model results for Rpb1 shown in **Figure 5E**, are obtained by using the parameters of the
724 dual noisy linear map model discussed above for the pom1 mutant, tuning RNAPII
725 parameters to obtain about 10% free RNAPII and linear scaling of DNA bound RNAPII, as
726 well as setting the transcription rate to match expression levels of Rpb1. The model without
727 any further tuning recovers deviation from concentration homeostasis observed at small cell
728 sizes, which is observed for the different genes in the WT and Pom1 mutant strains.

729 **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 90s (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 favourable 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 firstly 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 secondly 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.

768 **FIGURE LEGENDS**

769

770 **Figure 1: Gene expression scaling is a single cell attribute of constitutive and**
771 **inducible gene expression. A.** Representative images of an smFISH experiment. The
772 *rpb1*, *rpb2*, and *rpb3* mRNA are labelled in *wee1-50*, *cdc25-22*, and *wt* cells. Overlay of the
773 three channels and DAPI staining of DNA are shown in the last column. The white scale bar
774 represents 5µm. **B.** Log_2 average copies/cell of 7 mRNAs in wee1-50 (green) and cdc25-22
775 (red) cells plotted against log_2 average copies/cell of the same mRNAs in *wt* cells. mRNA
776 common names are shown on the figure. Plain line shows equality and dotted lines values 2
777 fold up or down. **C.** *rpb1* mRNA copies/cell plotted against cell length for *wee1-50* (green), *wt*
778 (blue), *cdc25-22* (red) and *wt* diploid cells (orange). **D.** *ace2* mRNA copies/cell plotted
779 against cell length for *wee1-50* (green), *wt* (blue), *cdc25-22* (red) and wt diploid cells
780 (orange). **E.** *urg1* mRNA copies/cell plotted against cell length for *wee1-50* (green), *wt* (blue),
781 *cdc25-22* (red) cells. Cells were analysed 3h after addition of uracil to the medium. **F.** same
782 as C. with all the cells in grey. The solid line shows median counts in running windows
783 sampled from 100 bootstrap samples of the experimental data and the shaded area
784 represents 95% confidence intervals. Only cells above the dotted lines were considered as
785 expressing the mRNA and were used for the running window analysis. **G.** same as D. with
786 all the cells in grey. The solid line shows median counts in running windows sampled from
787 100 bootstrap samples of the experimental data and the shaded area represents 95%
788 confidence intervals. Only cells above the dotted lines were considered as expressing the
789 mRNA and were used for the running window analysis. **H.** same as E. with all the cells in
790 grey. The solid line shows median counts in running windows sampled from 100 bootstrap
791 samples of the experimental data and the shaded area represents 95% confidence intervals.
792 Only cells above the dotted lines were considered as expressing the mRNA and were used
793 for the running window analysis.

794

795 **Figure 2: Coordination of transcription rates to cell size and not burst frequency nor**
796 **mRNA decay rates mediates scaling. A.** Cartoon of the modelling strategies used in this
797 section (**methods**). Domino-like shapes represent cells/agents with mRNA pictured as
798 yellow dots. Modelling of the cell cycle uses a noisy linear map framework and population
799 size is kept constant by removing a random cell for every division event. The left part of the
800 figures represents the two states model of gene expression. The transcription rate v and
801 degradation rate d can be coupled to cell length (L). **B.** Bar chart showing the sum of particle
802 weights for four models of transcription scaling after model selection analysis. Results for

genes with three types of regulation are shown. Constitutive: constitutively expressed genes, Periodic: cell cycle regulated genes, Regulated: genes regulated by external stimuli. For regulated genes, times after addition of 2,2-dipyridyl (sib1) or uracil (urg1) are shown. The four types of model used in this analysis are marked on the right. Poisson: non-bursty transcription, Bursty: bursty transcription, Constant: constant transcription rates during the cell cycle, Scaling: transcription rates scaling with cell length (**methods**). **C.** Simulation of mRNA numbers during the cell-cycle using a model where transcription is scaling under a poisson regime. Coefficients of variation (CV) are plotted as a function of mean simulated mRNA numbers for each particle of each mRNA. Yellow circles denote the median of simulated data for all particles of a given mRNA. Red circles denote median for all experimental data used for parameter inference. **D.** Expression correlation between pairs of mRNA in single cells. smFISH experimental measurements are marked by orange dots with the 95% confidence interval shown with orange bars. 95% confidence intervals for correlations obtained from model simulations as in C are shown with blue bars. Red bars show the 95% confidence interval for simulations including 20% additional extrinsic noise (**methods**). **E.** Bar chart showing the sum of particle weights for three models of mRNA degradation after model selection analysis. Decay rate scaling: transcription constant and degradation rates scale with cell length, Constant single: *wt*, *wee1-50*, and *cdc25-22* cells have an identical constant degradation rate during the cell cycle with transcriptional scaling. Constant multiple: *wt*, *wee1-50*, and *cdc25-22* cells have distinct constant degradation rates during the cell cycle with transcriptional scaling.

Figure 3: Coordination of RNAPII initiation rates with cell size as the main mechanism of scaling. A. *rpb1* normalised nascent sites intensities plotted against their respective cell length for *wee1-50* (green), *wt* (blue), *cdc25-22* (red). **B.** same as A. with all the cells in grey. 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 (methods). Cells with no nascent sites are excluded. Pearson correlation coefficient of nascent site intensity and length is shown. **C.** Boxplots of normalised nascent intensities for *rpb1* and the cell cycle regulated genes *ace2*, *mid2* and *fkh2*. Intensities are compared between haploid (green), diploid (yellow, orange), and *rpb1* heterozygote deletion cells (purple). For diploid cells the yellow boxplot shows intensities of single nascent sites, and orange boxplots show the total nascent intensities per cell. **D.** Cartoon of the TASEP modelling strategy of RNAPII transcription. A gene is represented by a lattice of length L_G. RNAPII molecules enter the lattice at a rate α (Initiation), hop through the lattice at a rate β (Elongation) and exit the lattice at a rate γ (Termination, **methods**). **E.** Transcription linear scaling captured by the TASEP model where rates α (Initiation), β (Elongation) or γ

839 (Termination) are linked to cell length. Simulations where run with each model with 500
840 random sets of TASEP parameters α, β and γ over times scales of 0.001 to 0.1 hours
841 (**methods**). Deviation from linearity is shown for parameter samples in each of the 3 model
842 variants. For each parameter set 1500 cells are simulated and two linear regressions are
843 done on the mRNA numbers vs cell length data for the smaller and larger half of simulated
844 cells. The deviation from linearity is estimated as the difference between the linear
845 regression coefficients of small vs large cells. Values close to zero indicate a linear scaling
846 and larger values indicate saturation of mRNA numbers at large cells. The dashed line
847 shows the error in estimating the slope of the linear fit. Note that the initiation model shows
848 the most robust linearity. **F.** Fraction of cells with a nascent site ("Transcribing fraction")
849 plotted as a function of mean cell length for the *rpb1* mRNA in a series of strains of different
850 average length. Strains names are labelled next to the data points. Red line shows linear
851 regression and R^2 is shown in the top left corner. **G.** Chromatin immunoprecipitation analysis
852 of RNAPII occupancy in wee1-50 (green), wt (blue), and cdc25-22 (red) cells. RNAP ChIP-
853 seq data for antibodies against Serine 5 CTD phosphorylation (left, middle) and total
854 Rbp1(right) are shown. RNAP occupancy values were normalised using DEseq2 and scaled
855 using occupancy at histone genes (**methods**). **H.** RNAPII occupancy 5'/3' ratio in wee1-50
856 (green), wt (blue), and cdc25-22 (red) cells. RNAP ChIP-seq data for antibodies against
857 Serine 5 CTD phosphorylation (left, middle) and total Rbp1(right) are shown. RNAP
858 occupancy values were normalised to input samples and each gene was divided into 100
859 bins using the deeptools package. Ratios between mean occupancy in bins 1-50 compared
860 to bins 51-100 are shown (**methods**). Note that ratios are not changing with size and are
861 consistently lower than 1.

862 **Figure 4: Nuclear RNA Polymerase II concentration increases with cell size. A.**
863 Confocal images of Rpb1, Rpb9, and Hta2 proteins tagged with GFP in *wt* fission yeast cells.
864 DAPI straining for DNA and overlay of both channels are shown. The white scale bar
865 represents 5μm. **B.** Widefield fluorescence data from live-cell imaging RNAPII subunit Rpb9.
866 Normalised intensity/cell as a proxy for total cellular protein concentration is plotted along
867 time relative to the mitotic phase of the cell cycle. **C.** Widefield fluorescence data from live-
868 cell imaging RNAPII subunit Rpb9. Total pixel intensity in the area of strong fluorescence
869 signal as a proxy for chromatin bound amounts is plotted along time relative to the mitotic
870 phase of the cell cycle. **D.** As in B. but for the histone protein Hta2. **E.** As in C. but for the
871 histone protein Hta2.

872 **Figure 5: The nucleus is the scaling unit. A.** Representative smFISH images for the rpb1
873 and rpb2 mRNA in cdc11-119 grown at non-permissive temperature. The white scale bar
874 represents 5μm. **B.** rpb1 normalised nascent sites intensities per nucleus (blue) or per cell

875 (orange) plotted as a function of each cell respective size. Red lines are regression lines
876 from linear models. The white scale bar represents 5μm. **C**. Representative smFISH images
877 for the *rpb1*, *rpb2*, and *shd1* mRNA in *wt* and *pom1* deletion mutants carrying the Cut11-
878 GFP nuclear marker (left column). Overlay is shown on the right. The white scale bar
879 represents 5μm. **D.** mRNA concentration (mRNA/cell length) for the *rpb1*, *rpb2*, and *shd1*
880 mRNA in *wt* (black) and *pom1* deletion mutants (red). The solid line shows median counts in
881 running windows sampled from a count distribution identical to the experimental data.
882 Shaded area represents 95% confidence intervals. **E.** *rpb1* concentrations measured by
883 smFISH are plotted as a function of cell length. Running average for the experimental data
884 (grey) or derived from model predictions assuming transcription rates scaling with cell length
885 (magenta), nuclear area (green), or predicted from a mechanistic model of RNAPII prediction
886 are shown (light blue). **F.** Cartoon representation of a mechanistic model of RNAPII
887 transcription (see Methods for details and parameters).

888 **Figure S1 related to Figure 1. A.** Cell length distribution for the 111 datasets described in
889 this study. Colours represent different genotypes. Datasets are described in **Table S2**.
890 Coloured lines show the interquartile range for all the *wee1-50* (green), *wt* (blue), *cdc25-22*
891 (red), diploid and heterozygote deletions (orange) and *pom1*Δ (light blue) cells. **B.** Top:
892 Pooled count distribution of 7 constitutively expressed genes in *wee1-50* (green), *wt* (blue),
893 and *cdc15-22* (red) cells. Bottom: Median counts of 7 constitutively expressed genes as a
894 function of cell length in running windows sampled from a count distribution identical to the
895 experimental data. Shaded area represents 95% confidence intervals (**methods**). Data from
896 *wee1-50*, *wt*, *cdc15-22* cells are pooled and gene names are indicated next to their
897 respective data line. **C.** Deviation from linearity is plotted as a function of linearity P-values
898 (see Figure 3E, **methods**). Low P-values mean significant deviation from linearity. Colours
899 are as in panel A. P-values equal 0.05 and 0.001 are shown with dashed lines and
900 annotated. **D.** *rpb1* mRNA copies/cell plotted against cell length for diploid (orange), *cdc25-*
901 *22* (red) and *rpb1* heterozygote diploid (purple). The solid line shows median counts in
902 running windows sampled from a count distribution identical to the experimental data.
903 Shaded area represents 95% confidence intervals (**methods**). **E.** *myo1* mRNA copies/cell
904 plotted against cell length for diploid (orange), *cdc25-22* (red) and *myo1* heterozygote diploid
905 (purple). The solid line shows median counts in running windows sampled from a count
906 distribution identical to the experimental data. Shaded area represents 95% confidence
907 intervals (**methods**). **F.** *pan1* mRNA copies/cell plotted against cell length for diploid
908 (orange), *cdc25-22* (red) and *pan1* heterozygote diploid (purple). The solid line shows
909 median counts in running windows sampled from a count distribution identical to the
910 experimental data. Shaded area represents 95% confidence intervals (**methods**). **G.**

911 Representative images of an smFISH experiment for cell cycle regulated genes. The *mid2*,
912 *fkh2*, and *ace2* mRNA are labelled in *wee1-50*, *cdc25-22* and *wt* cells. Overlay of the three
913 channels and DAPI staining of DNA is shown on the last column. The white scale bar
914 represents 5µm. **H.** Representative images of an smFISH experiment for a gene regulated
915 by external conditions (*urg1*). The *urg1*, *ace2*, and *rpb1* mRNA are labelled in *wt* cells
916 before, 3 hours and 12 hours after addition of uracil to the culture. Overlay of the three
917 channels and DAPI staining of DNA is shown on the last column. The white scale bar
918 represents 5µm. **I.** *mid2* mRNA copies/cell plotted against cell length for *wee1-50* (green), *wt*
919 (blue), cdc25-22 (red). The solid line shows median counts in running windows sampled
920 from a count distribution identical to the experimental data. Shaded area represents 95%
921 confidence intervals (**methods**). Only cells above the dotted lines were considered as
922 expressing the mRNA and were used for the running window analysis. **J.** *fkh2* mRNA
923 copies/cell plotted against cell length for *wee1-50* (green), *wt* (blue), cdc25-22 (red). The
924 solid line shows median counts in running windows sampled from a count distribution
925 identical to the experimental data. Shaded area represents 95% confidence intervals
926 (**methods**). Only cells above the dotted lines were considered as expressing the mRNA and
927 were used for the running window analysis. **K.** *sib1* mRNA copies/cell plotted against cell
928 length for wt cells before (light blue), 20min after (blue) or 45min (pink) after addition of 2,2-
929 dipyridyl (DIP). The solid line shows median counts in running windows sampled from a
930 counts distribution identical to the experimental data. Shaded area represents 95%
931 confidence intervals (**methods**).

932 **Figure S2 related to Figure 2. A.** Representative images of smFISH experiments for the
933 *rpb1* mRNA in cells before (t = 0) and after a 45min (t = 45) treatment with 15 µg/ml of
934 Thiolutin. Note the number of mRNA dots decreases with time and that the transcription foci
935 disappear after treatment. The white scale bar represents 5µm. **B. Left:** mRNA half-lives
936 were measured after Thiolutin treatment in bins of increasing cell length for the *rpb1* (2
937 repeats) and *rpb2* mRNA in *wee1-50*, *wt*, and *cdc25-22* cells (left). **Right:** mRNA half-lives
938 were measured after nmt41 promoter switch-off in bins of increasing cell length for the *rpb1*,
939 *rpb2* and *rtp1* mRNA in *wt* cells.

940 **Figure S3 related to figure 3. A.** Representative images of smFISH experiments for the
941 *rpb1* mRNA in haploid and diploid cells. Nascent transcription sites are indicated with white
942 arrows. The white scale bar represents 5µm. **B.** Fraction of cells with 1, 2, or 3 nascent sites
943 (as in A) in haploid and diploid cells. Note the diploid cells have mostly 2 sites validating our
944 approach specificity and sensitivity to detect nascent transcription sites. **C.** Boxplot of
945 nascent sites intensities for the *rpb1* gene in *wee1-50* (green), *wt* (blue) and *cdc25-22* (red)
946 (related to **Figure 3A**). **D.** Pearson correlation between mRNA numbers and cell length for

947 the simulation data from **Figure 3E**. **E.** Histogram of ratio of initiation time scale ($1/\alpha$) and
948 the sum of elongation time scale and termination time scale ($\frac{L_G}{\beta}+\frac{1}{\gamma}$) for parameter sets
949 resulting in linear scaling and parameter sets resulting in saturating transcription scaling
950 indicating that initiation rate should be limiting for linear transcriptional scaling. **F.** Elongation
951 time of a 2kb gene obtained from the simulations from **Figure 3E** for the models with
952 Initiation rates (α, green) and elongation rates (β, red) coupled with size. Only simulation that
953 achieved linearity are shown. Elongation time from published experimental measurements of
954 elongation rates (1-2kb/min) are shown. Note that the elongation scaling simulations require
955 elongation rates much slower than experimentally measured in order to achieve linearity. **G.**
956 ABC model selection results between the initiation scaling and elongation scaling TASEP
957 models for 3 genes across several strains. **H.** Distribution of the correlations of frequencies
958 of actively transcribing cells with cell length for different parameter sets from the **Figure 3E**
959 simulations. Note the strong positive correlations as observed in experimental data (**Figure**
960 **3F**). **I.** Boxplot of correlations between normalised intensities of the rpb1 mRNA and cell
961 length for all cell genotypes in sliding windows or containing different number of cells. Each
962 box represents a window of a given width. Sliding windows are shifted with increment of 1
963 cell along all cells ordered by length. Note that correlations are mainly positive. **J.** Average
964 gene analysis for the RNAPII ChIP-seq data from Figure 3H. Data were normalised to input
965 samples and split in 200 bins (50: upstream, 100: gene, 50: downstream) using the
966 deeptools package.

967 **Figure S4 related to figure 4. A.** Representative image of a single cell followed for a full
968 cell-cycle using ibidi slides. Brightfield (top), the nuclear marker Uch2 (middle) and the
969 RNAPII subunit Rpb1 bottom. **B.** Rpb1 normalised intensity/cell as a proxy for total cellular
970 protein concentration is plotted along time relative to the mitotic phase of the cell cycle. **C.**
971 Rpb9 total pixel intensity in the area of strong fluorescence signal (green mask in A) as a
972 proxy for chromatin bound amounts is plotted along time relative to the mitotic phase of the
973 cell cycle.

974 **Figure S5 related to figure 5. A.** Ratios of nascent site intensities of bi-nucleated cdc11-
975 119 cells plotted as a function of the ratio of local cytoplasmic volumes (**methods**). The
976 black line is the regression line from a linear model**. B.** Correlation between cell length and
977 nuclear area in pom1Δ (red) and wt (black) cells. Pearson corretions are shown. **C.** As in
978 **Figure 5D** for mRNA numbers. **D.** As in **Figure 5D** for mRNA numbers/nuclear volume
979 plotted as a function of nuclear volume. **E.** Simulation data from the model described in
980 **Figure 5F**. DNA bound RNAPII is plotted as a function of cell length. Magenta lines show a
981 running average. **F.** As in **Figure 5E** for mRNA numbers divided by nuclear area.

982

983 **REFERENCES**

984

985

986 1. Marguerat, S., and Bähler, J. (2012). Coordinating genome expression with cell size.
987 Trends Genet. *28*, 560–5.

988 2. Schmoller, K.M., and Skotheim, J.M. (2015). The Biosynthetic Basis of Cell Size
989 Control. Trends Cell Biol.

990 3. Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J., and Nurse, P. (2010).
991 A coordinated global control over cellular transcription. Curr. Biol. CB *20*, 2010–2015.

992 4. Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W.,
993 Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single Mammalian Cells
994 Compensate for Differences in Cellular Volume and DNA Copy Number through
995 Independent Global Transcriptional Mechanisms. Mol. Cell *58*, 339–52.

996 5. Kempe, H., Schwabe, A., Crémazy, F., Verschure, P.J., and Bruggeman, F.J. (2015).
997 The volumes and transcript counts of single cells reveal concentration homeostasis
998 and capture biological noise. Mol. Biol. Cell *26*, 797–804.

999 6. Ietswaart, R., Rosa, S., Wu, Z., Dean, C., Howard, M., Ietswaart, R., Rosa, S., Wu, Z.,
1000 Dean, C., and Howard, M. (2017). Cell-Size-Dependent Transcription of FLC and Its
1001 Antisense Long Non-coding RNA COOLAIR Explain Article Cell-Size-Dependent
1002 Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-
1003 to-Cell Expression Variation. Cell Syst. *4*, 622-635.e9.

1004 7. Mena, A., Medina, D.A., García-Martínez, J., Begley, V., Singh, A., Chávez, S.,
1005 Muñoz-Centeno, M.C., and Pérez-Ortín, J.E. (2017). Asymmetric cell division requires
1006 specific mechanisms for adjusting global transcription. Nucleic Acids Res. *45*, 12401–
1007 12412.

1008 8. García-Martínez, J., Delgado-Ramos, L., Ayala, G., Pelechano, V., Medina, D.A.,
1009 Carrasco, F., González, R., Andrés-León, E., Steinmetz, L., Warringer, J., *et al.*
1010 (2016). The cellular growth rate controls overall mRNA turnover, and modulates either
1011 transcription or degradation rates of particular gene regulons. Nucleic Acids Res. *44*,
1012 3643–3658.

1013 9. García-Martínez, J., Troulé, K., Chávez, S., and Pérez-Ortín, J.E. (2016). Growth rate
1014 controls mRNA turnover in steady and non-steady states. RNA Biol. *13*, 1175–1181.

1015 10. Schmoller, K.M., Turner, J.J., Kõivomägi, M., and Skotheim, J.M. (2015). Dilution of
1016 the cell cycle inhibitor Whi5 controls budding-yeast cell size. Nature *526*, 268–72.

1017 11. Keifenheim, D., Sun, X.-M., D'Souza, E., Ohira, M.J., Magner, M., Mayhew, M.B.,
1018 Marguerat, S., and Rhind, N. (2017). Size-Dependent Expression of the Mitotic
1019 Activator Cdc25 Suggests a Mechanism of Size Control in Fission Yeast. Curr. Biol.
1020 *27*, 1491-1497.e4.

1021 12. Saint, M., Bertaux, F., Tang, W., Sun, X.-M., Game, L., Köferle, A., Bähler, J.,
1022 Shahrezaei, V., and Marguerat, S. (2019). Single-cell imaging and RNA sequencing
1023 reveal patterns of gene expression heterogeneity during fission yeast growth and
1024 adaptation. Nat. Microbiol. *4*, 480–491.

1025 13. Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic

1026 contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. U. S. A. *99*,
1027 12795–800.

1028 14. Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in
1029 noisy gene expression. Science *342*, 1188–93.

1030 15. Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic
1031 transcriptional dynamics: From single molecules to cell populations. Nat. Rev. Genet.
1032 *14*, 572–584.

1033 16. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II:
1034 emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

1035 17. Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: Control of
1036 transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478.

1037 18. Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA
1038 polymerase II juggernaut. Science (80-.). *352*.

1039 19. Heldt, F.S., Lunstone, R., Tyson, J.J., and Novák, B. (2018). Dilution and titration of
1040 cell-cycle regulators may control cell size in budding yeast. PLoS Comput. Biol. *14*, 1–
1041 26.

1042 20. Lin, J., and Amir, A. (2018). Homeostasis of protein and mRNA concentrations in
1043 growing cells. Nat. Commun. *9*, 4496.

1044 21. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S.
1045 (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat.
1046 Methods *5*, 877–879.

1047 22. Ciechonska, M., Sturrock, M., Grob, A., Larrouy-Maumus, G., Shahrezaei, V., and
1048 Isalan, M. (2019). Ohm's Law for emergent gene expression under fitness pressure.
1049 bioRxiv, 693234.

1050 23. Thomas, P. (2019). Intrinsic and extrinsic noise of gene expression in lineage trees.
1051 Sci. Rep. *9*, 474.

1052 24. Bertaux, F., Marguerat, S., and Shahrezaei, V. (2018). Division rate, cell size and
1053 proteome allocation: impact on gene expression noise and implications for the
1054 dynamics of genetic circuits. R. Soc. Open Sci. *5*, 172234.

1055 25. McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression.
1056 Proc. Natl. Acad. Sci. U. S. A. *94*, 814–9.

1057 26. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory
1058 networks. Proc. Natl. Acad. Sci. U. S. A. *98*, 8614–9.

1059 27. Munsky, B., Trinh, B., and Khammash, M. (2009). Listening to the noise: random
1060 fluctuations reveal gene network parameters. Mol. Syst. Biol. *5*, 318.

1061 28. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., and Koeppl, H.
1062 (2012). Moment-based inference predicts bimodality in transient gene expression.
1063 Proc. Natl. Acad. Sci. U. S. A. *109*, 8340–5.

1064 29. Zechner, C., Unger, M., Pelet, S., Peter, M., and Koeppl, H. (2014). Scalable
1065 inference of heterogeneous reaction kinetics from pooled single-cell recordings. Nat.
1066 Methods *11*, 197–202.

1067 30. Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G., and Legewie, S.
1068 (2018). Estrogen-dependent control and cell-to-cell variability of transcriptional
1069 bursting. Mol. Syst. Biol. *14*, e7678.

1070 31. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011).
1071 Mammalian genes are transcribed with widely different bursting kinetics. Science *332*,
1072 472–4.

1073 32. Komorowski, M., Costa, M.J., Rand, D.A., and Stumpf, M.P.H. (2011). Sensitivity,
1074 robustness, and identifiability in stochastic chemical kinetics models. Proc. Natl. Acad.
1075 Sci. U. S. A. *108*, 8645–50.

1076 33. Tiberi, S., Walsh, M., Cavallaro, M., Hebenstreit, D., and Finkenstädt, B. (2018).
1077 Bayesian inference on stochastic gene transcription from flow cytometry data.
1078 Bioinformatics *34*, i647–i655.

1079 34. Watt, S., Mata, J., López-Maury, L., Marguerat, S., Burns, G., and Bähler, J. (2008).
1080 urg1: a uracil-regulatable promoter system for fission yeast with short induction and
1081 repression times. PLoS One *3*, e1428.

1082 35. Mercier, A., and Labbé, S. (2010). Iron-Dependent remodeling of fungal metabolic
1083 pathways associated with ferrichrome biosynthesis. Appl. Environ. Microbiol. *76*,
1084 3806–3817.

1085 36. Wilusz, C.J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of
1086 gene expression into focus. Trends Genet. *20*, 491–497.

1087 37. Amorim, M.J., Cotobal, C., Duncan, C., and Mata, J. (2010). Global coordination of
1088 transcriptional control and mRNA decay during cellular differentiation. Mol. Syst. Biol.
1089 *6*, 380.

1090 38. Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals
1091 alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

1092 39. Zhao, Y.-B., and Krishnan, J. (2014). mRNA translation and protein synthesis: an
1093 analysis of different modelling methodologies and a new PBN based approach. BMC
1094 Syst. Biol. *8*, 25.

1095 40. Sahoo, M., and Klumpp, S. (2013). Backtracking dynamics of RNA polymerase:
1096 pausing and error correction. J. Phys. Condens. Matter *25*, 374104.

1097 41. Voliotis, M., Cohen, N., Molina-París, C., and Liverpool, T.B. (2008). Fluctuations,
1098 pauses, and backtracking in DNA transcription. Biophys. J. *94*, 334–48.

1099 42. Dobrzynski, M., and Bruggeman, F.J. (2009). Elongation dynamics shape bursty
1100 transcription and translation. Proc. Natl. Acad. Sci. U. S. A. *106*, 2583–8.

1101 43. Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and
1102 Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat. Struct.
1103 Mol. Biol. *14*, 796–806.

1104 44. Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable
1105 subunits of yeast RNA polymerase II stimulate the initiation of transcription at a
1106 promoter in vitro. J. Biol. Chem. *266*, 71–5.

1107 45. Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation
1108 rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831–40.

1109 46. Pelechano, V., Chávez, S., and Pérez-Ortín, J.E. (2010). A Complete Set of Nascent
1110 Transcription Rates for Yeast Genes. PLoS One *5*, e15442.

1111 47. Swinburne, I.A., and Silver, P.A. (2008). Intron Delays and Transcriptional Timing
1112 during Development. Dev. Cell *14*, 324–330.

1113 48. Neumann, F.R., and Nurse, P. (2007). Nuclear size control in fission yeast. J. Cell

1114 Biol. *179*, 593–600.

1115 49. Verde, F., Mata, J., and Nurse, P. (1995). Fission yeast cell morphogenesis:
1116 identification of new genes and analysis of their role during the cell cycle. J. Cell Biol.
1117 *131*, 1529–38.

1118 50. Bahler, J., and Pringle, J.R. (1998). Pom1p, a fission yeast protein kinase that
1119 provides positional information for both polarized growth and cytokinesis. Genes Dev.
1120 *12*, 1356–1370.

1121 51. Hayles, J., Wood, V., Jeffery, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Salas-Pino, S.,
1122 Heichinger, C., and Nurse, P. (2013). A genome-wide resource of cell cycle and cell
1123 shape genes of fission yeast. Open Biol. *3*, 130053.

1124 52. Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T.,
1125 Nam, M., Palmer, G., *et al.* (2010). Analysis of a genome-wide set of gene deletions in
1126 the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. *28*, 617–623.

1127 53. Weston, L., Greenwood, J., and Nurse, P. (2017). Genome-wide screen for cell
1128 growth regulators in fission yeast. J. Cell Sci. *130*, 2049–2055.

1129 54. Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R.,
1130 Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019).
1131 Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.

1132 55. Bartman, C.R., Hamagami, N., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A.,
1133 and Raj, A. (2018). Transcriptional Burst Initiation and Polymerase Pause Release
1134 Are Key Control Points of Transcriptional Regulation. Mol. Cell.

1135 56. Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V.,
1136 Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., *et al.* (2013). Promoters maintain
1137 their relative activity levels under different growth conditions. Mol. Syst. Biol. *9*, 701.

1138 57. Windner, S.E., Manhart, A., Brown, A., Mogilner, A., and Baylies, M.K. (2019).
1139 Nuclear Scaling Is Coordinated among Individual Nuclei in Multinucleated Muscle
1140 Fibers. Dev. Cell *49*, 48-62.e3.

1141 58. Billman, M.R., Rueda, D., and Bangham, C.R.M. (2017). Single-cell heterogeneity and
1142 cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome
1143 open Res. *2*, 87.

1144 59. Youn, M., Takada, I., Imai, Y., Yasuda, H., and Kato, S. (2010). Transcriptionally
1145 active nuclei are selective in mature multinucleated osteoclasts. 1025–1035.

1146 60. Roberts, S.E., and Gladfelter, A.S. (2015). Nuclear autonomy in multinucleate fungi.
1147 Curr. Opin. Microbiol. *28*, 60–65.

1148 61. Dundon, S.E.R., Chang, S.-S., Kumar, A., Occhipinti, P., Shroff, H., Roper, M., and
1149 Gladfelter, A.S. (2016). Clustered nuclei maintain autonomy and nucleocytoplasmic
1150 ratio control in a syncytium. Mol. Biol. Cell *27*, 2000–2007.

1151 62. Wu, Z., Ietswaart, R., Liu, F., Yang, H., Howard, M., and Dean, C. (2016). Quantitative
1152 regulation of FLC via coordinated transcriptional initiation and elongation. Proc. Natl.
1153 Acad. Sci. U. S. A. *113*, 218–23.

1154 63. Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in
1155 Single Mammalian Cells. Cell *163*, 1596–1610.

1156 64. Soltani, M., Vargas-Garcia, C.A., Antunes, D., and Singh, A. (2016). Intercellular
1157 Variability in Protein Levels from Stochastic Expression and Noisy Cell Cycle
1158 Processes. PLoS Comput. Biol. *12*, e1004972.

1159 65. Takahashi, C.N., Miller, A.W., Ekness, F., Dunham, M.J., and Klavins, E. (2015). A
1160 Low Cost, Customizable Turbidostat for Use in Synthetic Circuit Characterization.
1161 ACS Synth. Biol. *4*, 32–38.

1162 66. Heinrich, S., Geissen, E.-M., Kamenz, J., Trautmann, S., Widmer, C., Drewe, P.,
1163 Knop, M., Radde, N., Hasenauer, J., and Hauf, S. (2013). Determinants of robustness
1164 in spindle assembly checkpoint signalling. Nat. Cell Biol. *15*, 1328–39.

1165 67. Trcek, T., Chao, J.A., Larson, D.R., Park, H.Y., Zenklusen, D., Shenoy, S.M., and
1166 Singer, R.H. (2012). Single-mRNA counting using fluorescent in situ hybridization in
1167 budding yeast. Nat. Protoc. *7*, 408–19.

1168 68. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E.,
1169 Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of
1170 transcripts in 3D FISH images. Nat. Methods *10*, 277–278.

1171 69. Rodríguez-López, M., Gonzalez, S., Hillson, O., Tunnacliffe, E., Codlin, S., Tallada,
1172 V.A., Bähler, J., and Rallis, C. (2019). The GATA transcription factor Gaf1 represses
1173 tRNA genes, inhibits growth, and extends chronological lifespan downstream of
1174 fission yeast TORC1. bioRxiv, 700286.

1175 70. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
1176 Wheeler transform. Bioinformatics *25*, 1754–1760.

1177 71. Lock, A., Rutherford, K., Harris, M.A., Hayles, J., Oliver, S.G., Bähler, J., and Wood,
1178 V. (2019). PomBase 2018: user-driven reimplementation of the fission yeast database
1179 provides rapid and intuitive access to diverse, interconnected information. Nucleic
1180 Acids Res. *47*, D821–D827.

1181 72. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with
1182 high-throughput sequencing data. Bioinformatics *31*, 166–169.

1183 73. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change
1184 and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

1185 74. Wiśniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A “proteomic ruler” for
1186 protein copy number and concentration estimation without spike-in standards. Mol.
1187 Cell. Proteomics *13*, 3497–506.

1188 75. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a
1189 flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–
1190 W191.

1191 76. Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman,
1192 N. (2014). Advanced methods of microscope control using μManager software. J.
1193 Biol. Methods *1*, 10.

1194 77. Amir, A. (2014). Cell size regulation in bacteria. Phys. Rev. Lett. *112*.

1195 78. Tanouchi, Y., Pai, A., Park, H., Huang, S., Stamatov, R., Buchler, N.E., and You, L.
1196 (2015). A noisy linear map underlies oscillations in cell size and gene expression in
1197 bacteria. Nature *523*, 357–360.

1198 79. Jun, S., and Taheri-Araghi, S. (2014). Cell-size maintenance: universal strategy
1199 revealed. Trends Microbiol. *23*, 4–6.

1200 80. Wang, P., Robert, L., Pelletier, J., Dang, W.L., Taddei, F., Wright, A., and Jun, S.
1201 (2010). Robust growth of Escherichia coli. Curr. Biol. *20*, 1099–103.

1202 81. Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J.
1203 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. U. S. A. *105*, 17256–61.

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. (2015). Adaptive approximate bayesian computation for complex models. Comput. Stat. *2013*, 2777–2796.

Figure 1 Sun, Bowman et al

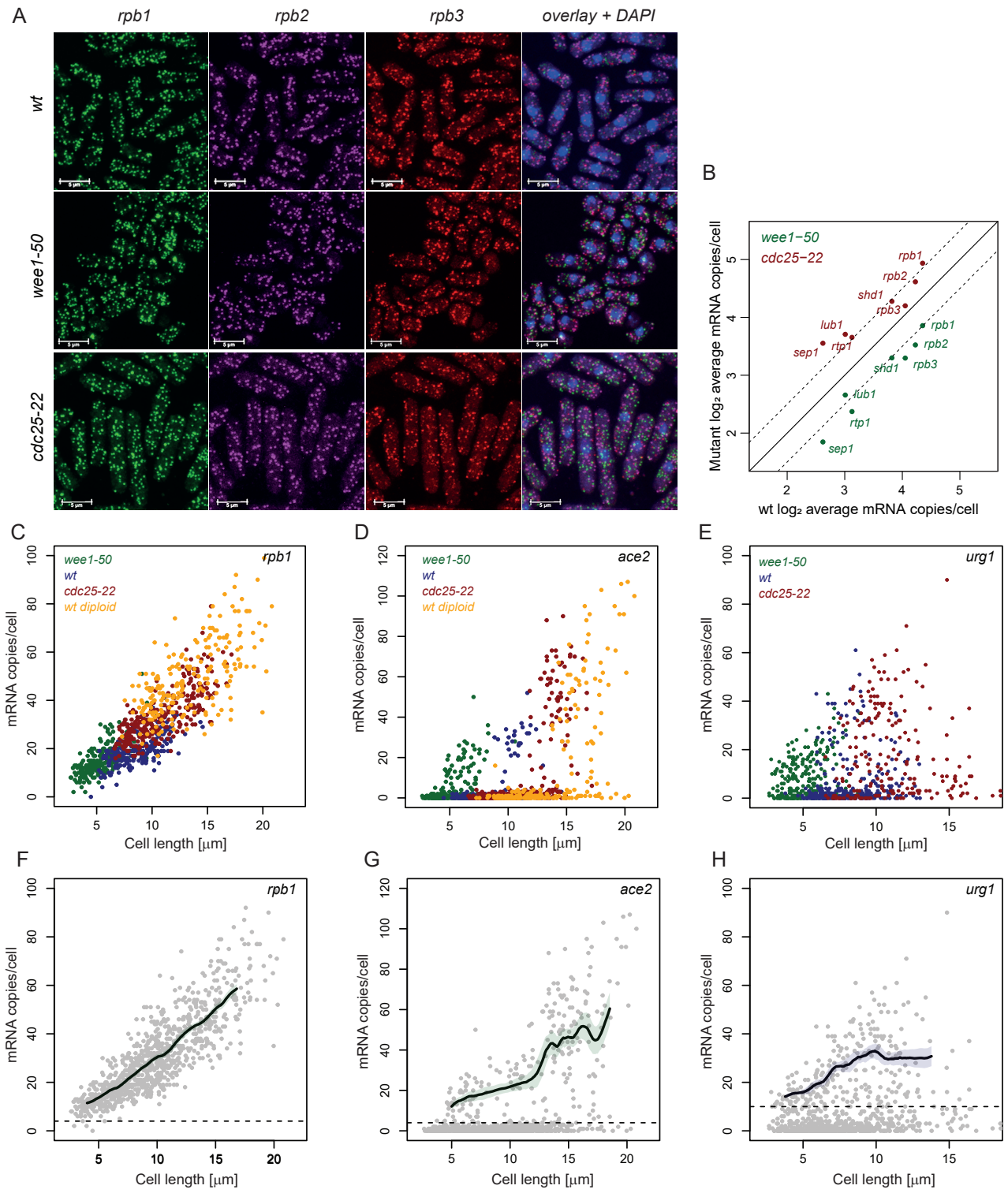


Figure 2 Sun, Bowman et al

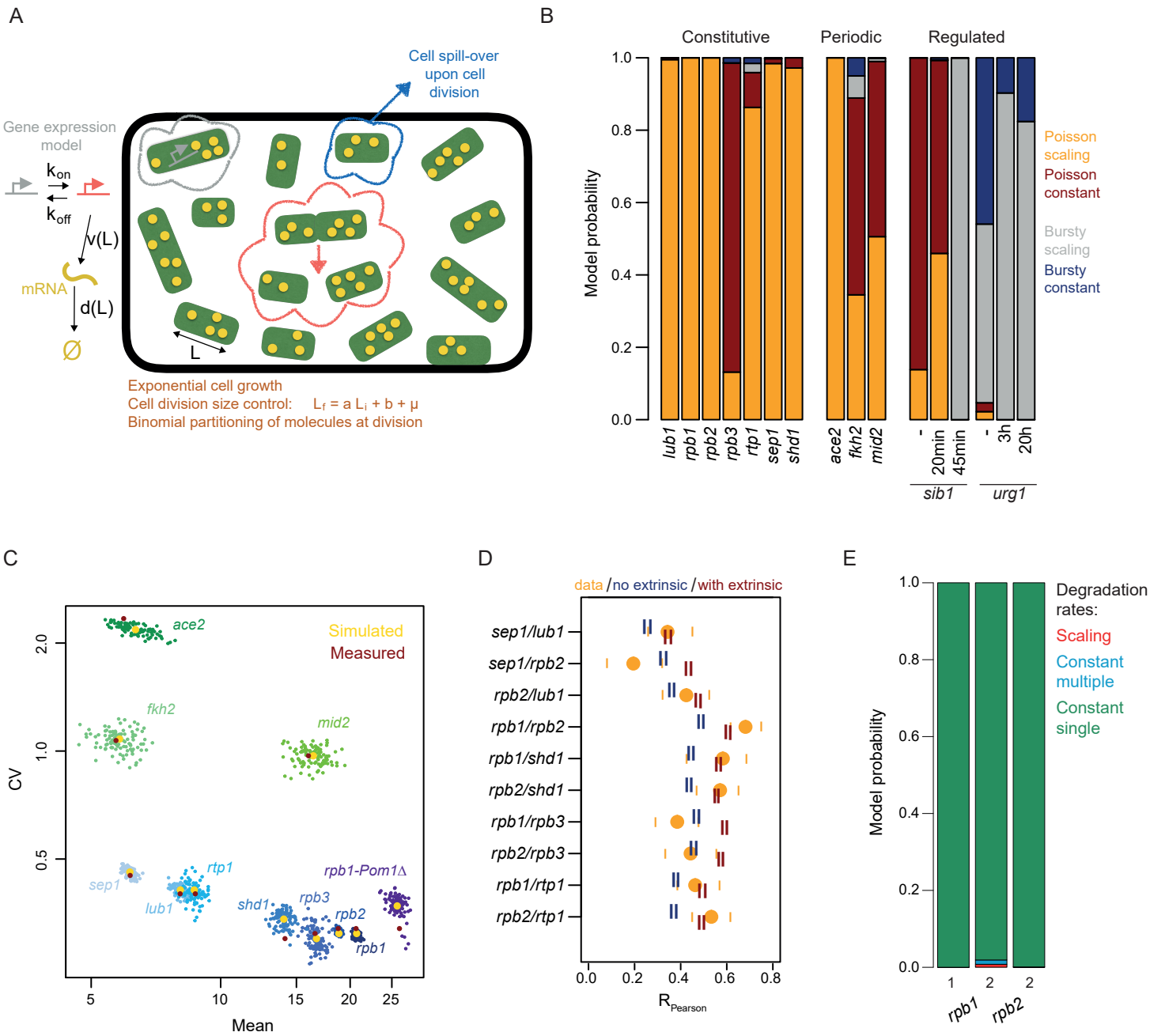


Figure 3 Sun, Bowman et al

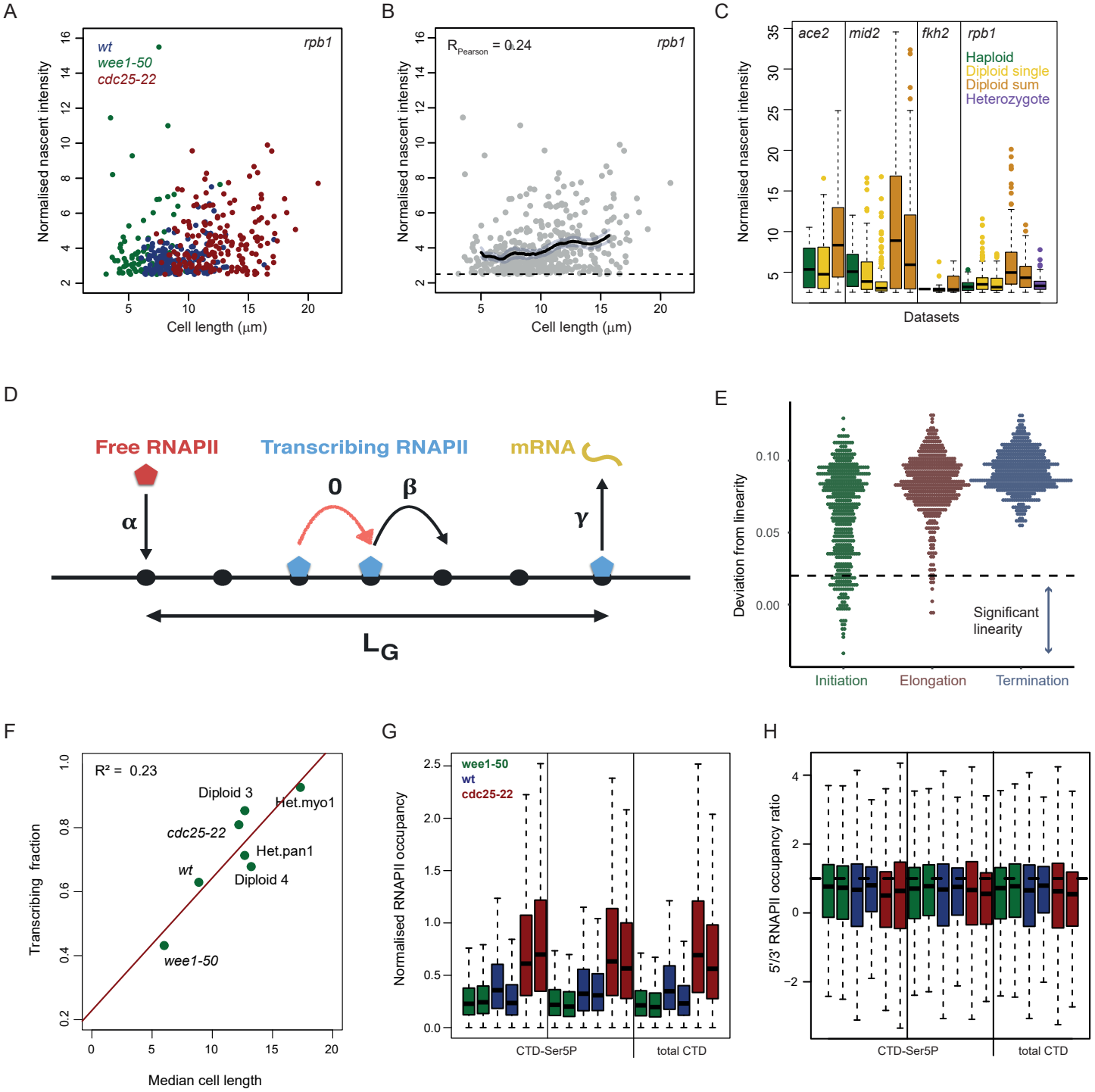


Figure 4 Sun, Bowman et al

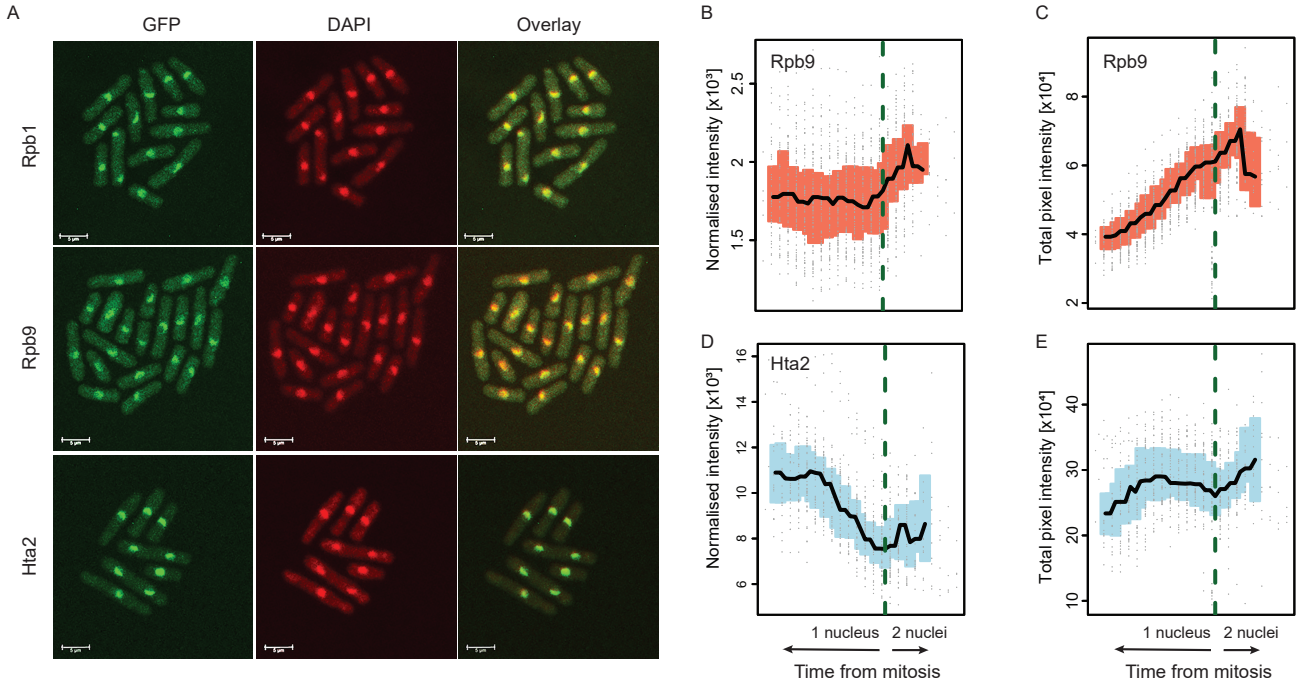


Figure 5 Sun, Bowman et al

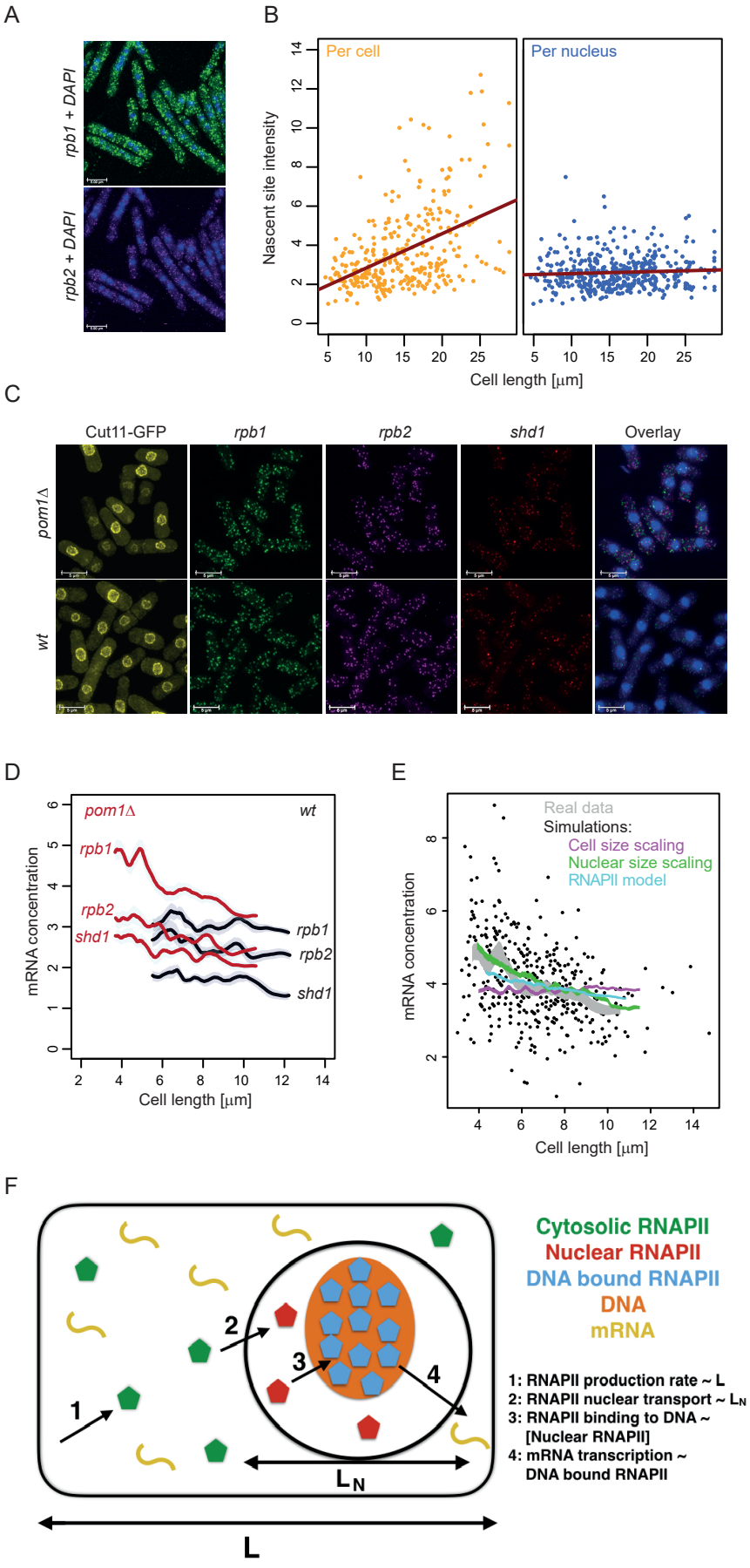


Figure S1 Sun, Bowman et al

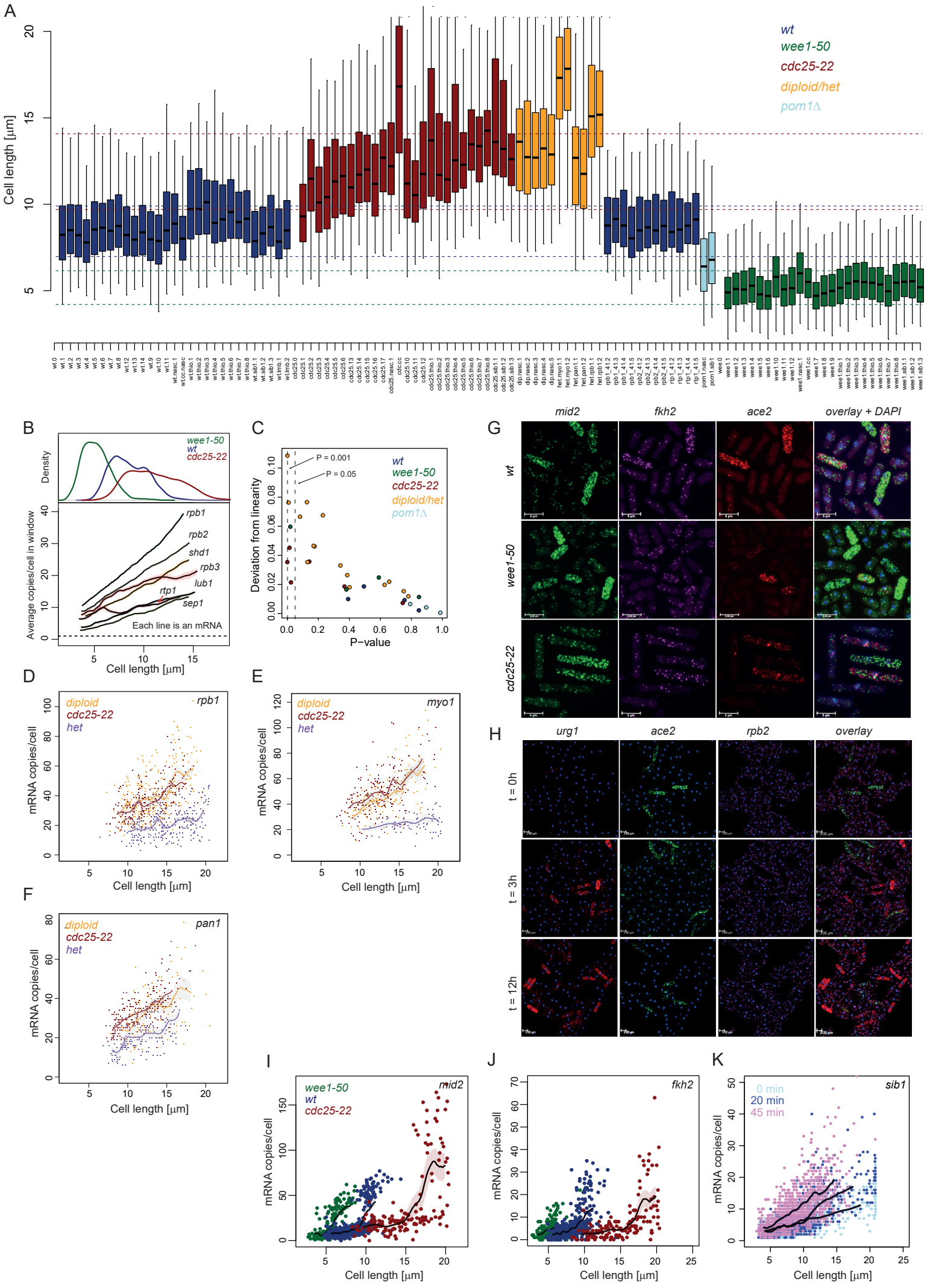
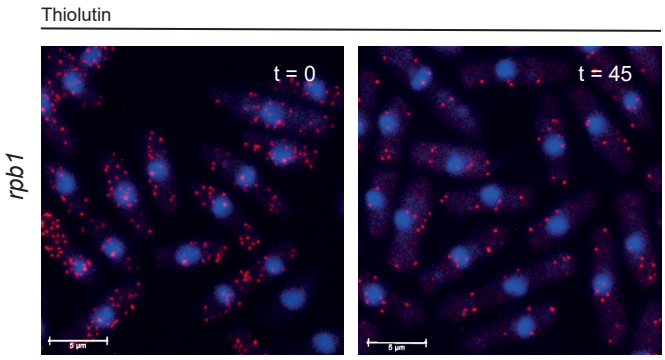


Figure S2 Sun, Bowman et al

A



B

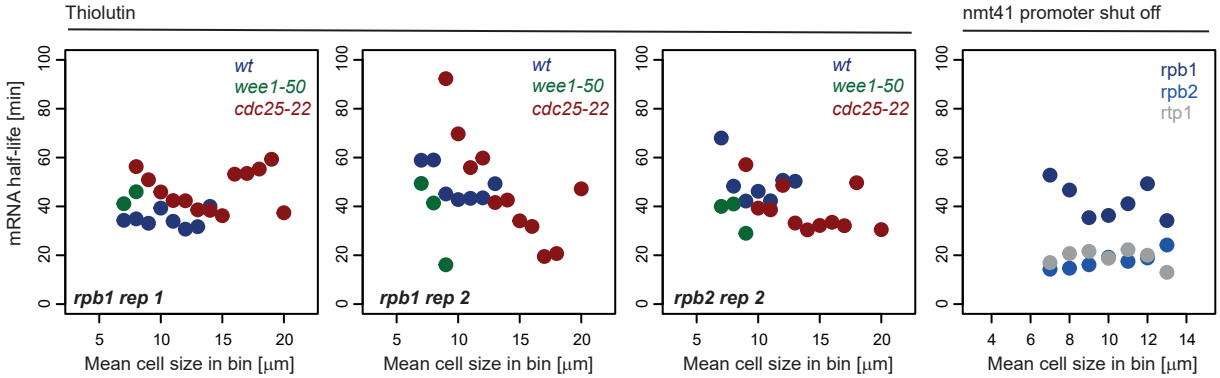


Figure S3 Sun, Bowman et al

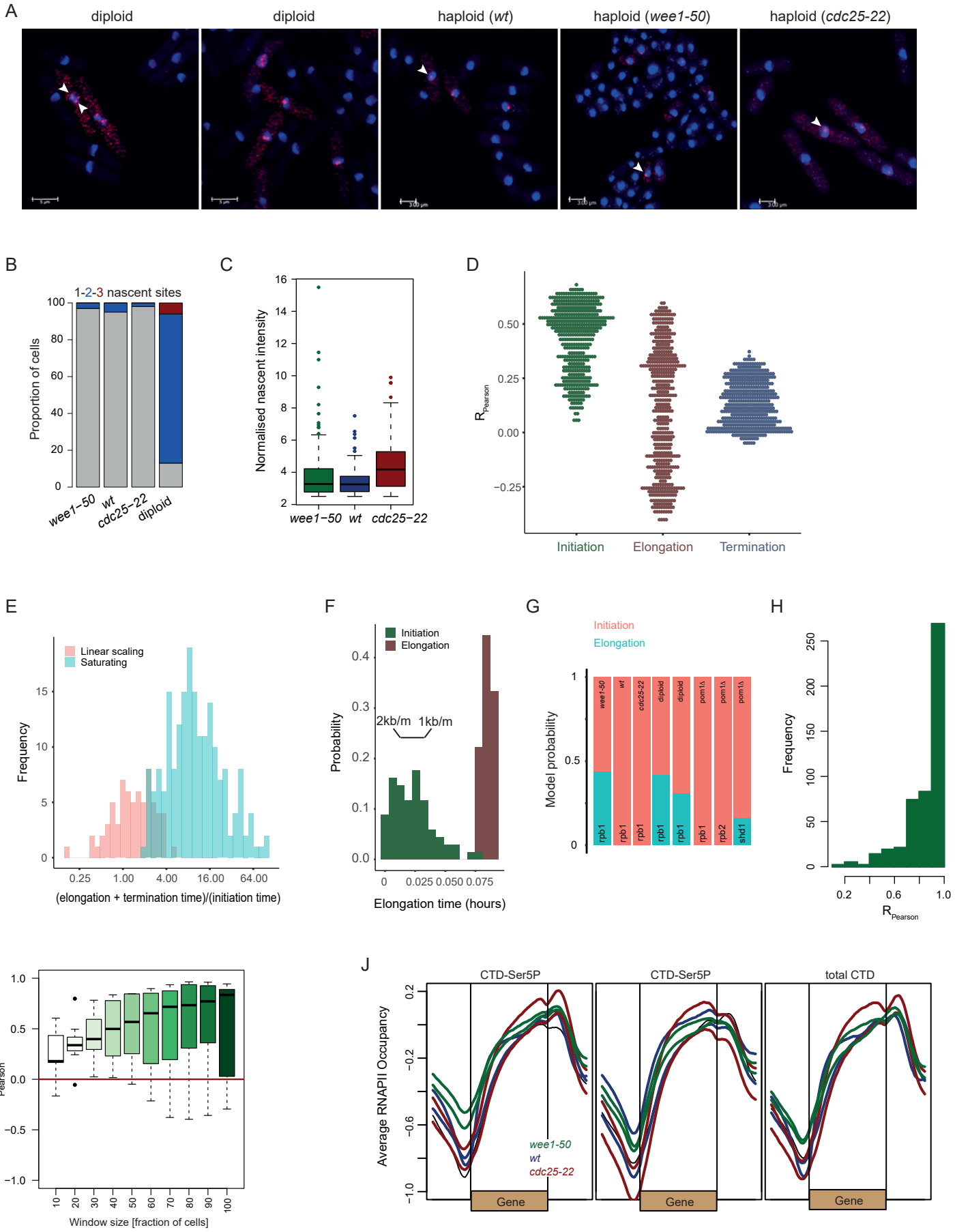


Figure S4 Sun, Bowman et al

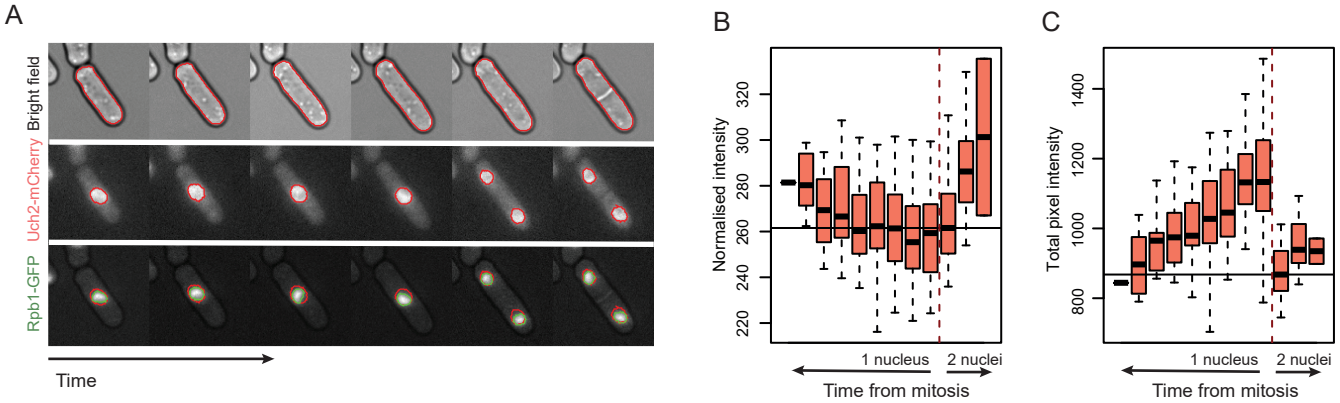
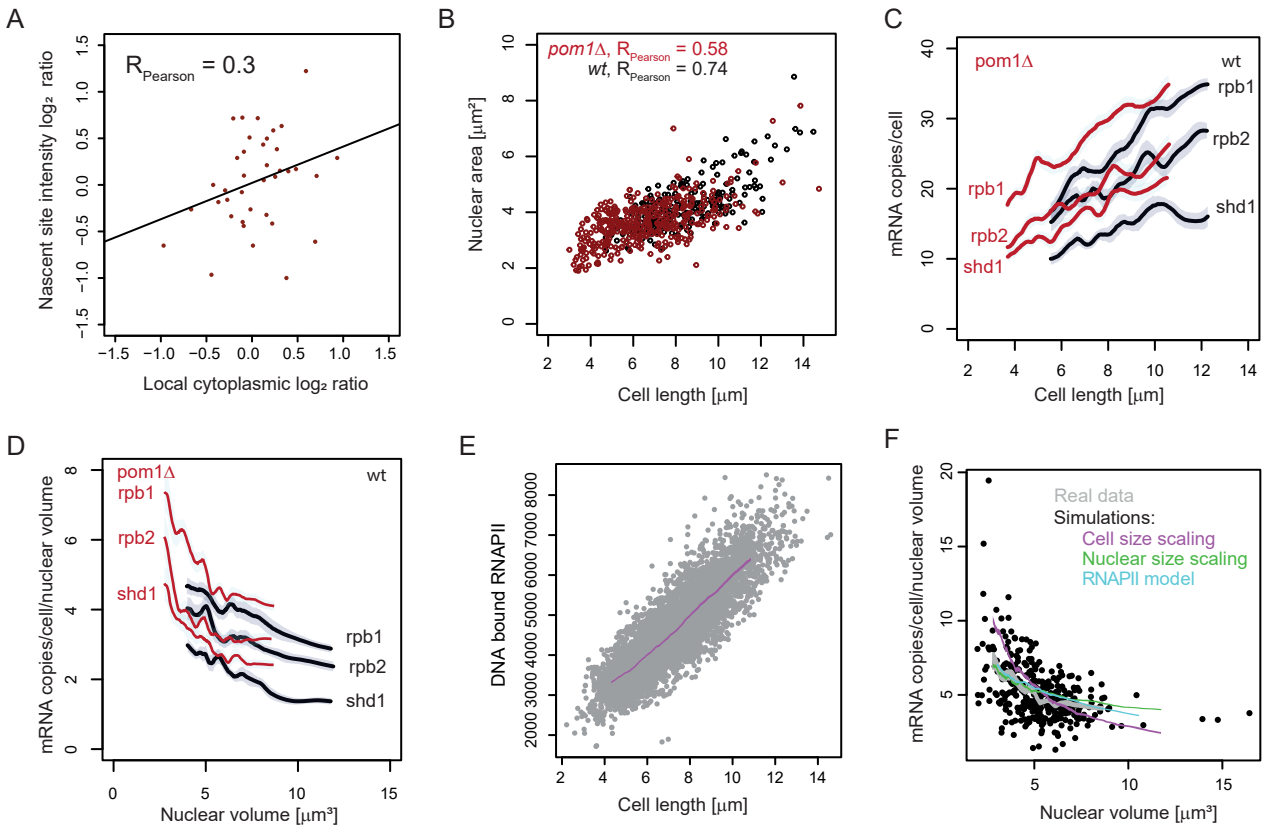



Figure S5 Sun, Bowman et al