The minimal intrinsic stochasticity of constitutively expressed eukaryotic genes is sub-Poissonian

Douglas E. Weidemann[1,2], Abhyudai Singh[3,4], Ramon Grima[5] and Silke Hauf[1,2]

[1]*Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA*

[2]*Fralin Life Sciences Institute, Virginia Tech, Blacksburg, VA 24061, USA*

[3]*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA*

[4]*Department of Biomedical Engineering, University of Delaware, Newark, DE 19716, USA*

[5]*School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JR, Scotland, UK*

*Correspondence: silke@vt.edu

Running head: Sub-Poissonian noise

Abstract

Stochastic variation in gene products ("noise") is an inescapable by-product of gene expression and can have profound consequences on cellular physiology. The least variation is seen for constitutively expressed genes, whose mRNA numbers in the population are commonly assumed to be Poisson-distributed, i.e. showing a variance equal to the mean. Here, we show that specific constitutively expressed cell division genes in *S. pombe* show considerably lower variance in mRNA numbers, classifying them as sub-Poissonian. Importantly, the sub-Poissonian nature is more pronounced in the cytoplasm than the nucleus. Modeling indicates that the dynamics of cytoplasmic mRNA degradation and nuclear export are key in reducing the variance-to-mean ratio in the cytoplasm. Our findings suggest that constitutively expressed genes are not a homogenous group with respect to expression noise. Instead, they can differ in gene expression dynamics to create a range of mRNA distributions, whose lower limit is sub-Poissonian.

Introduction

All gene expression is “noisy”. Gene products fluctuate stochastically, which is an inescapable consequence of the nature of gene expression[1–4]. Gene expression, like other intracellular processes, relies on molecule-molecule (in the case of gene expression often protein-DNA) interactions. When both molecules in such a reaction are abundant, their interaction is frequent and not limiting for the reaction itself. However, for several stages of gene expression, one binding partner may be present in very low numbers. For example, to initiate transcription, proteins must bind a gene promoter, which typically only has one to four copies per cell. As a result, stochasticity becomes apparent and the intracellular number of mRNA molecules that has been transcribed from a specific gene varies significantly over time and between cells[5–8].

Gene expression noise differs substantially between genes. For genes with high variation, some cells in the population may contain dozens or hundreds of mRNA molecules expressed from that gene, and others none[7,9–13]. In a few cases, this large variation has been shown to be advantageous, since it allows “bet-hedging”, i.e. some cells in the population are well prepared for a change in environmental conditions, whereas others don’t produce gene products that are not currently needed[14–17]. Highly variable expression of a single gene is also exploited in development, where it may specify cell fate and generate a mix of cells with different phenotypes[18–20]. At the other end of the spectrum, “housekeeping” genes, i.e. genes required for core cellular processes and often required for survival, are expressed stably with low noise[21–23]. Cell cycle genes have also been found to show low intrinsic noise[24].

High-noise regimes have been extensively studied. One contributing factor to noisy expression is toggling between an inactive and active transcription state, so that mRNAs are produced in “bursts”, leading to large variation in mRNA abundance[25]. The large noise created by bursty expression can be modulated downstream of mRNA synthesis[26,27]. For example, slow mRNA export from the nucleus may buffer bursty mRNA synthesis and thereby lower noise in the cytoplasm relative to noise in the nucleus[11,28–30]. By contrast, molecular processes downstream of mRNA synthesis, such as multi-step mRNA decay, may also enhance noise created during transcription[31].

Less is known about the low-noise regime. Low noise genes are thought to be expressed from constitutive promoters that stay 'on', and do not toggle into an 'off' state[32,33]. Widely used models for this constitutive expression assume that mRNA is produced with a single rate-limiting step from an active promoter and that the wait time between synthesis events follows an exponential distribution[33–36]. This has been experimentally confirmed for two constitutively expressed genes in budding yeast[37]. Under these assumptions, the steady-state mRNA numbers in the cell population will follow a Poisson distribution, independent of the precise nature of the mRNA degradation process[38]. A Poisson distribution is characterized by the variance being equal to the mean, or the Fano factor (the variance divided by the mean) being equal to 1. The Fano factor describes the "noise strength"[1], or the variance relative to the variance of a Poisson distribution with the same mean. Consistent with this proposed expression model, several constitutive genes in yeast show distributions of mRNA numbers that closely follow a Poisson distribution[6,12,39,40]. This distribution is typically considered the "noise floor", i.e. the lower limit of intrinsic stochasticity in gene expression[11,40–43].

To shed more light on low-noise gene expression, we have investigated a group of constitutively expressed *S. pombe* genes, that are important for cell division and show low mRNA and protein noise[44,45]. The protein products of these genes contribute to the spindle assembly checkpoint (SAC, also known as mitotic checkpoint). This signalling pathway operates during cell division to detect chromosomes that are not correctly attached to the mitotic spindle and halts the execution of anaphase as a response[46–49]. SAC signalling involves several protein-protein interactions, and the relative ratio of SAC proteins is important for SAC function[44,50,51]. Hence, low expression noise of these genes supports SAC function, making it important to understand how low noise is achieved.

Strikingly, rather than Poissonian mRNA distributions, we found sub-Poissonian mRNA distributions for these genes, i.e. Fano factors less than 1. In the cytoplasm, the Fano factor could be as low as 0.5. We also examined other, non-SAC, low noise genes, as well as re-examined published results, and found a spectrum of sub-Poissonian to Poissonian mRNA distributions for constitutively expressed genes. The cytoplasmic mRNA distributions of these genes can have Fano factors that are lower, the same, or higher than those of the nuclear mRNA distributions. Altogether,

this suggests that constitutively expressed genes are not a homogenous group, but instead can differ substantially in mRNA noise. Importantly, we conclusively establish that the lower limit of intrinsic stochasticity for constitutively expressed genes is sub-Poissonian, not Poissonian.

Results

Gene expression is regulated independently for each *S. pombe* SAC gene

SAC genes need to cooperate in one signalling pathway, and their relative stoichiometries are important for function[44,50,51]. Stoichiometries could be more easily maintained if expression of these genes was coupled. i.e. if higher or lower expression of one coincided with higher or lower expression of another. Yet, isolated observations suggest that their expression is independent[44,52]. We systematically tested for interdependence at both the mRNA and protein levels (Fig. 1). Except for the encoded protein itself, the levels of the SAC proteins Mad1, Mad2, Mad3, and Bub1 were largely unaffected by single deletion of *mad1*, *mad2*, *mad3*, or *bub1* (Fig. 1A), and so were the mRNA levels measured by quantitative PCR (Fig. 1B). Mad1 and Mad2 form a tight 2:2 complex that is central for SAC function[53,54]. Yet, *mad1* and *mad2* mRNA concentrations did not correlate in single cells (Fig. 1C); nor did those of *mad1* and *mad3* (Fig. 1D). This agrees with findings in budding yeast that the concentrations of constitutively expressed mRNAs coding for subunits of stable protein complexes do not necessarily correlate[39]. We did observe a slight increase in mRNA concentration of *mad1, mad2*, and *mad3* by quantitative PCR when expression of *bub1* was increased by 2- to 3-fold, and a similar trend for *mad3* in a *mad1* overexpression (Fig. 1B). However, the changes were so subtle that we conclude that SAC genes are expressed mostly or wholly independently.

SAC genes show sub-Poissonian distributions of mRNA numbers, more narrow than those of other low noise genes

Without cross-talk between SAC expression levels, it is important that expression of these genes is stable, so that levels do not deviate beyond the range that allows for SAC function. We therefore assessed variability in SAC gene mRNA numbers at the single cell level by single-molecule mRNA

fluorescence in-situ hybridization, smFISH[55]. The FISH spot intensities in the cytoplasm were homogenously distributed in a narrow range, consistent with representing single mRNA molecules (Fig. S1). Spots with higher intensities were almost exclusively observed for highly expressed genes and only in the nucleus, presumably representing the transcription site (Fig. S1, S9). We counted spots as a single mRNA if they had an intensity below the 95th percentile of spot intensities in the cytoplasm (Fig. S1C). This avoids that technical noise from uneven labelling is interpreted as biological noise. For brighter spots, the number of mRNA contained in each spot was estimated as value of its intensity divided by the median intensity of a cytoplasmic spot. This ensures that several mRNAs at the same position are accurately counted as such.

The cell-to-cell variation in mRNA numbers is a result of intrinsic and extrinsic noise sources[1,56,57]. A considerable extrinsic influence is cell size. Mean mRNA numbers increase with cell size (see Fig. 2B, S2 for examples), so that mRNA concentrations (number divided by cell volume) remain approximately constant as cells grow[40,43,58–60]. To exclude the influence of size, we either considered only cells within a given size range (Fig. 2B, S2A) or mathematically corrected the variance for the size effect[58] (Fig. 2C, S2B; see Methods). When excluding the influence of cell size, all four SAC genes (*mad1 mad2*, *mad3*, and *bub1*) showed Fano factors of the mRNA distributions that were clearly below 1 and for *mad1* reaching as low as 0.5 (Fig. 2B,C; S2A,B; Supplemental Information, A.2). This was independent of whether we used cell length or cell volume as the size parameter (Fig. S2C), and for *mad1* was the case regardless of whether a GFP tag was present or not (Fig. 2B,C; S2A,B; Supplemental Information, A.2). Further statistical analysis showed that the Fano factor of the SAC genes being below 1 could not be explained by finite sample size effects (Supplemental Information, B.4). A Poisson distribution (Fano factor = 1, Fig. 2A) would have been expected if SAC genes were expressed with a single rate-limiting step, as is typically assumed for constitutively expressed genes. Hence, this observation suggests that other, potentially more complex mechanisms account for the ultra-low Fano factor of the SAC gene mRNA distributions.

The distribution of mRNA numbers in *S. pombe* for several other genes had previously been examined, and some genes showed mRNA distributions close to a Poisson distribution [12,40]. When

we corrected these published data for cell size in the same way we had done for the SAC genes, five genes (*sep1*, *lub1*, *rpb2*, *shd1*, *rpb1*) showed single experiments with evidence of sub-Poissonian mRNA distributions, but none of them as consistently across multiple experimental replicates as we observed for the SAC genes (Fig. 2D). After averaging Fano factors from all experimental replicates, none of these genes had a Fano factor significantly below 1 (Supplemental Information Fig. A-2). We re-tested two of these genes (*sep1* and *rpb1*), as well as two higher noise genes (*SPAC2H10.01*, *SPAC27D7.09c*), with our workflow and found the mRNA numbers and variance to be highly consistent with those obtained previously, cross-validating the data (Fig. 2E, S2D). We did not observe a correlation between mean mRNA numbers and cell size-corrected Fano factors across these genes (Fig. 2F). SAC genes show low mean mRNA numbers (means of 2–5). Other genes with similarly low mean mRNA number can show considerably larger Fano factors. In summary, the SAC genes show the lowest Fano factor in mRNA number distributions among all *S. pombe* genes for which cell size and mRNA numbers have been reported.

Characterization of SAC gene structure shows short promoters, variable 5' UTR lengths, and typical 3' UTR lengths

The sub-Poissonian mRNA number distribution of SAC genes raised the question whether these genes share any peculiar characteristics, possibly in the promoter or the mRNA. We first mapped the 5' and 3'UTRs, because they may influence mRNA dynamics, and because the 5'UTR length provides insight into the position of the promoter (Fig. 3). The 5'UTRs of *mad1* and *mad2* are extremely short (~10 nucleotides), below the 6th percentile of all *S. pombe* genes[61]. The 5'UTR of *mad3* (~270 nucleotides) was longer than the interquartile range observed in *S. pombe*, whereas that of *bub1* (~60 nucleotides) was within the interquartile range. Our results were highly consistent with a recent genome-wide study identifying transcription start sites in *S. pombe* (Fig. S3A)[61]. The very short 5'UTRs were surprising, since short 5'UTRs have been shown to impair translation from the first start codon in the mRNA[62]. The only gene ontologies that were enriched for genes with such short 5'UTRs were those of ribosomal proteins (Fig. S3B). Although the short 5'UTRs of *mad1* and mad2 are unusual, the strong differences in 5'UTR length between the SAC genes imply that

there is no direct correlation with the sub-Poissonian mRNA distributions.

Promoters have a large influence on gene expression noise[15,42,63–68]. The location, size, and characteristics of most promoters in *S. pombe* remain poorly defined[61,69]. To map the promoter regions of the SAC genes, we transferred these genes with a variable length of sequence upstream of the transcription start site (3 to 770 nucleotides) to two different intergenic loci: (i) the 3.8 kb region next to *wis1*, one of the largest intergenic regions in *S. pombe*, and (ii) the 0.4 kb region next to *leu1*, which is a frequently used integration site in *S. pombe*[70] (Fig. 4A). We used two loci to be able to control for effects from the surrounding region. We found that expression only started to break down when less than 50 to 100 base pairs of upstream sequence were kept (Fig. 4B,C; S4A, S5). Monitoring mRNA levels by qPCR and protein levels by either immunoblotting or imaging gave similar results (Fig. 4B,C). This suggests that the expression breakdown indeed reflects the inability of the coding sequence to be transcribed. The results from the two intergenic loci were overall consistent, with some quantitative differences observed for *mad2* (Fig. 4B,C; S4A).

Based on the short length of the candidate promoter regions, expression is likely driven by a core promoter only. Consistently, *mad1* and *bub1* had TATA box-like sequences in the ~40 base pair upstream sequence that, when removed, drastically lowered expression (Fig. S5). This was surprising, though, because TATA boxes are typically found in promoters causing large noise[15,63,64,71]. Mutating the candidate TATA box in *mad1* clearly lowered expression (Fig. 4D,E), suggesting it functions as a genuine TATA box. This was not the case for *bub1* (Fig. 4D,E), but *bub1* contains another TATA box-like sequence upstream (Fig. S5), which may have compensated. We conclude that the expression of SAC genes is likely driven by core promoter sequences only.

Regulatory sequences of *mad2* and *mad3* are sufficient for narrow mRNA distributions in the Poisson-range, but not for sub-Poissonian mRNA distributions

To test whether the identified promoters are sufficient for expression with low Fano factors, we placed the coding sequences of two high-noise genes under the *mad2* and *mad3* promoters at the exogenous locus (Fig. 5A). We used *rad21*, which is a cell cycle-regulated gene whose Fano factor can increase to around 12 during expression late in the cell cycle (Fig. 5B; S6A,B)[72–74], and *nmt1*,

which is a thiamine-responsive, highly variable gene[12,45]. When expressing *rad21* and *nmt1* from the *mad2* and *mad3* promoter, the cell cycle-regulation of *rad21* disappeared (Fig. 5C; S6), and the Fano factor of the mRNA distributions of *rad21* and *nmt1* was lowered to a value slightly above 1 (Fig. 5D). This was close to, but not quite as low, as the *mad2* and *mad3* coding sequences expressed from their own promoter at the exogenous locus (Fig. 5D). Overall, this confirms that promoter sequences have a large influence[67], but suggests that a combination of gene elements is required for the very narrow mRNA distributions observed for SAC genes.

To explore the subtle differences in noise between endogenous and exogenous loci, and between the endogenous and exogenous coding sequences, we determined the size-corrected Fano factor separately for the nucleus, where mRNAs are produced, and for the cytoplasm, where mRNAs are ultimately translated. We assigned mRNAs to either the nucleus or the cytoplasm after segmenting nuclei in 3D based on DNA staining and correcting for chromatic aberration (Fig. S7A). This revealed interesting differences. For *mad3* at its endogenous locus, the Fano factor of the mRNA distribution was lower in the cytoplasm than in the nucleus, whereas it was similar for *mad2* (Fig. 5E). When *mad3* was expressed from the exogenous locus, though, its Fano factor now also became similar between nucleus and cytoplasm. Furthermore, *rad21* and *nmt1* expressed from the same regulatory sequences tended to show a higher Fano factor in the cytoplasm than in the nucleus (Fig. 5E). Together, this suggested that an interplay between promoter, coding sequence, and surrounding sequences may be necessary for sub-Poissonian mRNA distributions, and that the sub-Poissonian mRNA expression of SAC genes is maintained or further suppressed in the cytoplasm.

Most SAC genes show lower mRNA Fano factors in the cytoplasm than in the nucleus

To test for the generality of this observation, we determined size-corrected nuclear and cytoplasmic Fano factors for all SAC genes and the other genes examined previously, all expressed from their endogenous locus (Fig. 6A; Supplemental Information A.3). Strikingly, all SAC genes, with the exception of *mad2*, showed lower Fano factors in the cytoplasm than in the nucleus (Fig. 6A, Supplemental Information Table A-I). Other low noise genes (*sep1* and *rpb1*) showed a similar or

slightly higher Fano factor in the cytoplasm than in the nucleus, and high noise genes (*SPAC2H10.01* and *SPAC27D7.09c*) showed a considerably higher Fano factor in the cytoplasm than in the nucleus. These results suggest that for most SAC genes the Fano factor of their mRNA distribution is further lowered post-transcriptionally.

These data with segmented nuclei also allowed us to distinguish mononucleated and binucleated cells (Fig. 6B-G; Supplemental Information A.4). In *S. pombe*, the nuclear envelope does not break down during cell division, and chromosome segregation is followed by splitting of the nucleus into two, and ultimately cell division[75]. G1 phase in *S. pombe* is short and DNA replication occurs mostly prior to cell division during the binucleated stage[76]. Thus, binucleated cells are in late mitosis, G1, or S phase, whereas almost all mononucleated cells are in G2 (Fig. 6B). We observed significant Fano factor differences between mononucleated and binucleated cells for almost all genes, with stronger differences in the cytoplasm than the nucleus for the low noise genes (Fig. 6C-G, S8, Supplemental Information, Table A-IV). In mononucleated cells, the *rpb1* gene shows a pattern very similar to the SAC genes with a sub-Poissonian distribution of mRNA numbers in the cytoplasm and a lower Fano factor in the cytoplasm than the nucleus (Fig. 6F). The Fano factor suppression between nucleus and cytoplasm for SAC genes went away during the binucleated stage, and *rpb1* and *cdc13* showed drastic increases in cytoplasmic (but not nuclear) Fano factor in binucleated cells (Fig. 6F,G). Overall, this most likely reflects dramatic changes in cellular physiology during cell division and DNA replication, which could impact mRNA synthesis, nuclear export, mRNA degradation, or possibly all of those[77–83]. It also illustrates that it is important to minimize cell cycle effects when analyzing gene expression noise and its mechanistic basis[43,84–86].

The Fano factor of the *mad1* mRNA distribution is lower in the cytoplasm than at the transcription site or in the nucleoplasm

Our observations suggest that the Fano factor of SAC gene mRNA distributions is further lowered post-transcriptionally. To strengthen this conclusion, we additionally examined mRNA directly at the transcription start site (Fig. 7A). Using a similar strategy to one previously employed in *E. coli*[87], we labelled the transcription start site by integrating lac operator repeats into an intergenic region

3.3 kb and 6.4 kb away from *mad1* and *rpb1*, respectively, and expressing a Lac inhibitor-NLS-GFP fusion in these cells (Fig. 7A). This resulted in typically one GFP spot in the nucleus (Fig. 7B). Both mean mRNA numbers and cell size-corrected Fano factors remained highly similar to cells without the transcription site labelled (Fig. S9A), suggesting that the integration did not alter expression from the endogenous locus.

We classified mRNAs in the direct vicinity of the GFP spot as nascent (Fig. 7A; S7B). Occasionally, a transcription site was just outside of the region that we had classified as nuclear based on DNA labelling and segmentation. In this case, we isotropically expanded the nucleus to include the transcription site (Fig. S7B) and adjusted our classification of nuclear and cytoplasmic mRNA accordingly. Excluding these cells from the analysis as an alternative method yielded similar results (compare Fig. 7E and Fig. S9B). For the highly expressed *rpb1* gene, this procedure captured all high-intensity mRNA spots as nascent (Fig. 7B,C; S9C), confirming the validity of the approach. This procedure then allowed us to distinguish nascent, mature nuclear, and cytoplasmic mRNAs (Fig. 7A). For both *mad1* and *rpb1*, the number of mature nuclear mRNAs was lower than the number at the transcription start site, indicating fast nuclear mRNA export (Fig. 7D), as has been previously found for budding yeast[88]. Consequently, the number of all nuclear mRNAs is not a good proxy for the number of mature nuclear RNAs (Fig. 7D; Supplemental Information Table B-II).

The analysis of the cell size-corrected Fano factors in the different compartments showed that both *mad1* and *rpb1* are transcribed in a way that yields a sub-Poissonian distribution of mRNA numbers at the transcription site, even stronger for *rpb1* than for *mad1* (Fig. 7E; S9D; Supplemental Information A.2). The Fano factor of mature nuclear mRNA was close to 1. In agreement with our previous observations (Fig. 6F), the Fano factor of cytoplasmic *mad1* mRNA in mononucleated cells was clearly lower than either in the nucleoplasm or at the transcription site (Fig. 7E; Supplemental Information A.3). For *rpb1*, the Fano factor of cytoplasmic mRNA was lower than that of mature nuclear mRNA, but slightly larger than that at the transcription site. As we had seen before, suppression of the Fano factor in the cytoplasm for *mad1* disappeared in binucleated cells (Fig. 7E). The results were independent of whether we used cell length or cell volume as size parameter (Fig. S9E). Furthermore, the same trends were observed when binning cells by their

size and analyzing the Fano factor of the mRNA distribution in cells of similar size (Fig. S9F). Overall, these results show that both these constitutively expressed genes show characteristics of transcription that lead to a sub-Poissonian distribution of mRNA numbers at the transcription start site. In addition, the results confirm that the Fano factor of the cytoplasmic mRNA distribution can be significantly lower than that in the nucleoplasm, or even at the transcription site.

Modeling suggests that multiple rate-limiting steps in transcription and degradation, combined with fast nuclear export, can explain the sub-Poissonian mRNA distributions

How can sub-Poissonian mRNA distributions be explained mechanistically? A simple, widely used model for constitutive gene expression assumes single rate-limiting steps for mRNA synthesis and degradation with exponential wait times between single synthesis or degradation events[33–36]. Under these assumptions, the Fano factor of mature mRNA in both the nucleus and cytoplasm will be 1[38]. This may be consistent with data for some constitutive genes, but clearly not all.

To obtain a distribution of mRNA numbers that is narrower than the Poisson distribution, the wait times between single mRNA synthesis events need to be more homogeneous than in an exponential distribution[89,90]. Assuming multiple steps with similar timescales in mRNA synthesis, rather than just one rate-limiting step, will lead to such narrower than exponential wait times[91–94]. Biologically, this assumption is well justified because RNA polymerase II (Pol II) binding to the promoter is known to be followed by several additional steps, such as promoter opening, promoter escape, promoter-proximal pausing and pause release[95,96]. We constructed a model of this type that also includes mRNA export from the nucleus to the cytoplasm and degradation of the mRNA in the cytoplasm (Fig. 7F). Degradation of mRNA in the cytoplasm is also known to involve multiple steps[97,98], and it has been argued that assuming several rate-limiting steps is necessary to fit transcriptome-wide mRNA decay data of several budding yeast genes[99]. Hence, the model includes the possibility of multiple rate-limiting steps in mRNA degradation.

We derived mathematical expressions for the Fano factor in steady-state conditions to determine the minimal Fano factors possible for this model (Fig. 7G and Supplemental Information). As expected, increasing the number of transcription initiation steps decreases the possible Fano factor

and allows for Fano factors below 1 (Fig. 7G). Interestingly, increasing the number of rate-limiting steps in mRNA degradation allows for lower Fano factors in the cytoplasm than in the nucleus (Fig. 7G), fitting what we observed experimentally (Fig. 6,7). In this model, whether the cytoplasmic Fano factor is lower or higher than the nuclear Fano factor depends on the nuclear export rate (Fig. 7H). At slow nuclear export rates, the Fano factor in the cytoplasm is larger than that in the nucleus. Only at fast nuclear export rates will the cytoplasmic Fano factor become smaller than that in the nucleus (Fig. 7H). The threshold value at which the switch occurs becomes lower with an increasing number of rate-limiting steps in transcription and mRNA degradation (Fig. 7H; Supplemental Information Table B-XIII). Hence, our mathematical model shows that multiple rate-limiting steps in transcription and mRNA degradation, along with efficient nuclear export of mRNA, can explain both the sub-Poissonian mRNA distributions as well as the lower Fano factor in the cytoplasm than in the nucleus. Furthermore, nuclear export can be an efficient lever to tune cytoplasmic mRNA variability in this sub-Poisson regime.

At least three rate-limiting steps in synthesis and mRNA degradation are required to explain all SAC gene data

To determine how many rate-limiting steps in transcription and mRNA degradation are required to explain the experimental data, we used Bayesian model selection. We found strong evidence that 3 or 4 rate-limiting steps in both transcription and mRNA degradation are required to explain the data for *mad1*, *mad1*-GFP and *bub1*-GFP (Supplemental Information Table B-XI). The expression of other genes (*mad2*-GFP, *mad3*-GFP, *rpb1*, *sep1*-GFP) could be equally well explained by other combinations of the number of rate-limiting steps, with the provision that there are at least 2 steps for transcription. For each gene, we then estimated the rate constants for promoter remodeling (k_A), promoter freeing (k_B), and nuclear export (k_C) relative to the effective rate of cytoplasmic degradation, using a model with three rate-limiting steps in both transcription and mRNA degradation (Supplemental Information Table B-XII, Fig. B-1). This indicated differences between genes. The non-SAC genes, *rpb1* and *sep1*, were best fit with higher rates of promoter remodeling (k_A); the two genes not showing a reduction of the Fano factor in the cytoplasm, *mad2* and *sep1*,

were best fit with lower rates of nuclear export (k_C) (Supplemental Information, Fig. B-2). Consistent with a lower nuclear export rate, *mad2* and *sep1* have a higher fraction of their mRNA in the nucleus (Fig. S8G), information that was not used in the fitting (Supplemental Information B.3). Altogether, our results imply that the sub-Poissonian mRNA distributions observed for some constitutively expressed genes can be biologically explained by multiple rate-limiting steps in transcription and mRNA degradation. Furthermore, our results indicate that constitutively expressed genes are not a homogenous group, but instead differ measurably in their expression characteristics with consequences on mRNA noise.

Discussion

The expression characteristics of low noise genes have received much less attention than those of high noise genes. Here, we showed that the mRNA distribution of constitutively expressed, low noise genes can be narrower than a Poisson distribution, which establishes that the intrinsic "noise floor" of constitutively expressed genes is sub-Poissonian, not Poissonian. This reveals that low noise gene expression is more varied than previously thought and that constitutively expressed genes are not a homogenous group. Our findings imply that the widely used model for constitutive expression with single rate-limiting steps in transcription and mRNA degradation is an oversimplification. This knowledge is relevant because researchers have at times eliminated data from their analysis when the mRNA distributions appeared sub-Poissonian[100]. Based on our results, this may have excluded valid and interesting data. We suggest that—in their most basic, unregulated form—the multi-step nature of transcription and mRNA degradation leads to sub-Poissonian noise. However, certain promoter architectures or added-on regulation, e.g. by a transcription factor or by an RNA-binding protein, can introduce additional intrinsic noise and thereby move the gene closer to a Poisson-type expression.

Our findings were surprising because prior reports of sub-Poissonian mRNA distributions are extremely sparse. In bacteria, one group has reported that expression from the *tetA* promoter results in a sub-Poissonian mRNA distribution[101], and that the wait times between synthesis events from this and two other bacterial promoters are more narrowly distributed than exponential,

consistent with two or three rate-limiting steps in transcription[101–103]. Others concluded that gene expression noise in bacteria is generally Poissonian or super-Poissonian[5,41,104]. In eukaryotes, mRNA distributions of constitutively expressed genes have generally been reported to be Poissonian[6,39,40]. However, some of these values were obtained without correcting for cell size, which may have made a sub-Poissonian distribution appear Poissonian (see Fig. S2A for an example). In addition, unrecognized cell cycle effects may have muddied the data. For example, the cytoplasmic mRNA distribution of *rpb1* (coding for the largest subunit of Pol II) has a Fano factor clearly below 1 in mononucleated *S. pombe* cells, but clearly above in binucleated (Fig. 7D), which averages to around 1 across the population (Fig. 6A)—and hence its expression was assumed Poissonian[40]. Interestingly, one report provides evidence for a sub-Poissonian mRNA distribution of the orthologous *S. cerevisiae RPB1* gene[105]. However, finite sample effects would need to be excluded for a formal conclusion. Furthermore, nascent mRNA distributions at the transcription site of budding yeast genes have been reported to be sub-Poissonian[39,106]. However, the transcription site is special in that it contains both full-length and shorter mRNAs. As a result, even if the time between synthesis events is exponentially distributed, the distribution of nascent mRNAs at the transcription site may appear sub-Poissonian[91].

Overall, although prior data from eukaryotes does not specifically support sub-Poissonian distributions, it does not exclude them either. Since we used one of the gold standards to assess mRNA numbers[107], which yielded solid evidence of sub-Poissonian distributions (Figs. 2,6,7; Supplemental Information), and since unrecognized biological or technical variability would only inflate the measured variability, we find it hard to escape the conclusion that the expression of constitutive genes can be sub-Poissonian.

Whether mRNA distributions are sub-Poissonian or Poissonian has at least two biological implications. For one, it makes a difference in noise strength. Whether this is functionally relevant, i.e. whether increasing sub-Poissonian to Poissonian expression lowers fitness, remains to be determined. Generally, noise characteristics of genes can be under natural selection[22,108–111]. It will be interesting to determine which classes of genes show sub-Poissonian expression and if and how this relates to their function. Secondly, the mRNA distributions provide a window into the gene

expression process. A promoter transcribing with a single rate-limiting step must be fundamentally different from one exhibiting multiple rate-limiting steps. To what extent and how constitutive promoters differ from each other is still surprisingly little understood[112]. Distinguishing sub-Poissonian and Poissonian low noise genes may help to classify promoters and identify their functionally important elements. It is worth noting, though, that the elements defining sub-Poissonian expression may not be neatly confined to the promoter (Fig. 5). This highlights the important role that chromatin context[68,113] and non-coding features of the coding sequence[114] play in gene expression. Further, it remains a possibility that sub-Poissonian distributions are brought about by negative feedback, as has been examined repeatedly in models[115–120]. For the genes examined here, we are not aware of any known process establishing negative feedback, and we therefore favor multiple steps in transcription and mRNA degradation as the origin of their sub-Poissonian mRNA distributions.

Another relevant observation is that most SAC genes showing sub-Poissonian mRNA distributions also show a reduction of the Fano factor between nucleus and cytoplasm. Such a reduction of variance has been seen for bursty genes, where it was attributed to slow nuclear export[11,28,29]. In contrast, our model predicts that the nuclear export rate needs to be sufficiently fast for the Fano factor in the cytoplasm to drop below the one in the nucleus (Fig. 7H). Our analysis also shows that the cytoplasmic Fano factor can become lower with an increasing number of rate-limiting steps in mRNA degradation (Fig. 7G,H), which seems at odds with published results that multiple rate-limiting steps in mRNA degradation can increase the Fano factor in the cytoplasm[31]. The difference between these published results and ours is the underlying type of expression. Lowering of the cytoplasmic Fano factor by slow nuclear export or increase by multiple rate-limiting steps in mRNA degradation holds true for super-Poissonian expression (nuclear Fano factor larger than 1), but the relationships change in sub-Poissonian expression (nuclear Fano factor smaller than 1). Hence, the sub-Poissonian regime is not only quantitatively, but also qualitatively distinct, with its own rules.

Lastly, among the SAC genes we have studied, *mad2* is an interesting exception. It does show a sub-Poissonian mRNA distribution, but it does not show the Fano factor reduction between nucleus and cytoplasm seen for the other SAC genes (Fig. 6, S8). This suggests that its nuclear export or

mRNA degradation characteristics differ qualitatively from that of the other SAC genes. Indeed, Bayesian inference suggested that the nuclear export rate of *mad2* is slower than that of other SAC genes (Supplemental Information Table B-XII). We speculate that this could be a consequence of co-translational assembly of the Mad1/Mad2 complex. The Mad1 and Mad2 proteins form an extremely tight 2:2 complex[53,54]. The Mad1 dimer at the core of this complex assembles co-translationally[45,121]. We consider it possible that the binding of Mad2 to the Mad1 dimer also needs to be co-translational. In this case, *mad2* mRNA may associate with Mad1 protein, which in turn may affect its export characteristics or the number of rate-limiting steps in its degradation.

In summary, we have established that the low noise regime of eukaryotic gene expression reaches lower than previously thought. This opens avenues to understand the underlying molecular differences and identify gene elements that minimize noise.

Materials and Methods

Tables

S. pombe strains are shown in Table S1, sgRNA sequences in Table S2, FISH probes in Table S3, and qPCR primers in Table S4.

Strain construction

SAC genes *mad1*, *mad2*, *mad3*, and *bub1* were tagged with ymEGFP at the endogenous locus without inserting any other exogenous sequences[45]. This was accomplished either by replacement of the counter-selectable rpl42-hphNT1 cassette in an rpl42::cyh^{R}(sP56Q) strain[121] or by using CRISPR-mediated targeting[122]. The *mad2*-ymEGFP strain contains a single, silent (AGG to AGA) PAM site mutation at amino acid position 173 of Mad2. The *mad3*-ymEGFP strain contains a single, silent (TTG to TTA) PAM site mutation at amino acid position 199 of Mad3. Yeast codon-optimized, monomeric enhanced GFP (ymEGFP) was derived from yEGFP[123,124] by mutation of Alanine 206 to Arginine (A206R), which is expected to reduce dimerization[125]. Tagging of *sep1*, *SPAC2H10.01*, *SPAC27D7.09c*, and *rad21* with ymEGFP was performed by conventional PCR-based gene targeting[126]. Integration at the exogenous *wis1* and *leu1* loci used CRISPR-mediated targeting (CRISPR); sgRNA sequences are listed in Table S2.

For transcription site labeling, lacO repeats were integrated as described by Rohner and colleagues[127]. Insertion close to *mad1* is between the adjacent *trz2* and *but1* genes (3.3 kb away from the *mad1* stop codon), insertion close to *rpb1* is between the adjacent *toa1* and *hrd3* genes (6.4 kb away from the *rpb1* start codon). We tried to be minimally disruptive to gene expression by choosing relatively large intergenic regions, not directly adjacent but sufficiently close to the genes of interest. A resistance cassette (hphMX4 from pAG32) was inserted at the respective locus, and then swapped for the lacO repeats and *LEU2* as resistance marker (from pSR13). The swap was facilitated by CRISPR/Cas9-targeting of the resistance cassette.

Cell growth

YEA (yeast extract with adenine, 30 g/L glucose, 5 g/L yeast extract, 0.15 g/L adenine hemisulfate dihydrate) was used as rich medium; EMM (Edinburgh minimal medium, MP Biomedicals, 4110032) was used as minimal medium. Strains were thawed on YEA plates, and then grown for approximately 24 hr in liquid EMM, all at 30°C. When cultures were diluted to low densities, 50 % pre-conditioned medium (made by filtering EMM cultures) was added to EMM. L-Leucine (0.2 g/L) was added to EMM medium as needed for auxotrophic strains.

Single-molecule mRNA fluorescence in-situ hybridization (smFISH)

Cultures were grown to ~0.7-1.5 x 10^7 cells/ml and 2 x 10^8 cells were fixed with either 2 % or 4 % formaldehyde. After 30 min, formaldehyde was washed out with three washes of ice-cold Buffer B (1.2 M sorbitol, 100 mM potassium phosphate buffer pH 7.5). When not immediately continuing with the next step, cells were stored in 1 mL of Buffer B at 4°C. Fixed cells were resuspended in 1 mL spheroplast buffer (1.2 M sorbitol, 0.1 M potassium phosphate, 20 mM vanadyl ribonuclease complex [NEB S1402S], 20 µM beta-mercaptoethanol), and 1.2–5 µL 100T zymolyase (US Biological Z1005; 10 mg/mL) were added to digest the cell wall. Cells were kept at 30°C until the cell walls were sufficiently digested (determined by counting the fraction of cells that lysed when placed in deionized water). Typically, 30-50% of cells lysed was taken as evidence for sufficient digestion. Digestion was stopped by washing the cells three times with 1 mL of Buffer B. Next, cells were incubated for 20 min in 1 mL of 0.01 % Triton X-100 in 1x phosphate-buffered saline (PBS), followed by three more washes of Buffer B and one wash with 10 % formamide in 2x saline-sodium citrate (SSC) buffer (ThermoFisher AM9770). Cells were resuspended in the formamide/SSC solution and split evenly into two replicate samples for hybridization. For each sample, 3.75 pmol of Stellaris™ RNA FISH probes (CAL Fluor red 610 probes targeting ymEGFP or cdc13, or Quasar 570 probes targeting mad1 or rpb1; Biosearch Technologies, LGC) were combined with 2 µL each salmon-sperm DNA (Life Technologies, 15632-011) and yeast tRNA (Thermo Fisher, AM7119) and diluted in Buffer F (20 % formamide, 10 mM sodium phosphate buffer pH 7.2) to a final volume of 50 µL. This mixture was heated to 95°C for 3 min, allowed to cool to room temperate, and added

to an additional 50 µL of Buffer H (4x SSC buffer, 4 mg/mL acetylated BSA [Sigma, B8894], 20 mM vanadyl ribonuclease complex). Each sample was resuspended in 100 µL of this hybridization solution. When possible, as a positive control, one of the replicates of each sample was hybridized with probes for a higher abundance mRNA which would be expected to show FISH spots in all cells if the experiment was successful. After overnight incubation in the dark at 37°C, the hybridization solution was removed and cells were incubated for 6 min in 10 % formamide/2x SSC heated to 37°C, 6 min in 0.1 % Triton X-100/2x SSC, and finally 10 min in 1 µg/ml DAPI in 1x PBS. After DAPI staining, the cells were washed once with 1x PBS before final resuspension in 1x PBS. Cells were stored in the dark at 4°C until imaging.

For imaging, cells were mounted in SlowFade Diamond Antifade Mountant (Thermo Scientific, S36972) with #1.5 glass coverslips and RNase-free slides. Slides were imaged with a Zeiss AxioImager M1 equipped with Xcite Fire LED illumination (Excelitas), a Zeiss Plan FLUAR 100x/1.45 oil objective, and an ORCA-Flash4.0LT sCMOS camera (Hamamatsu). FISH optimized red or gold filters (Chroma 49306 and 49304, respectively) were used to image the FISH probes, while standard GFP, CFP, and DAPI filters were used to capture images of GFP-labeled transcription sites, cell boundaries, and nuclei, respectively. The images for each channel consisted of a 6 µm z-stack containing 31 images at 0.2 µm intervals.

Image processing of smFISH

Images were dark noise-subtracted and flatfield-corrected. Cells were segmented in two dimensions using a custom FIJI macro based on using trainable WEKA segmentation[128] while nuclei were segmented in 3D dimensions from DAPI images using a custom FIJI macro adapted from https://github.com/haesleinhuepf/cca_benchmarking (Robert Haase, MPI-CBG, Dresden). Cell segmentation was manually corrected and cells missing nuclear segmentation, or whose nuclei were not fully contained within the image stack, were removed. Cell length was measured from the segmented cell outlines using the bounding box method. RNA FISH spots and GFP transcription site spots were detected with FISHquant v3a[129]. FISH or GFP images were filtered with the default 3D_LoG filter (size = 5, sigma = 1). Spots were initially found using local maximum detection and

an automatically determined minimum intensity threshold; the quality score filtering option was not used. The default setting was used for minimum distance between spots (160 nm). After point spread function (PSF) fitting, final thresholds were set manually for the PSF sigma xy, sigma z, and z position to exclude non-RNA or non-transcription site spots such as hot pixels. Spot detection accuracy was checked manually for a subset of cells from each image. Additional R scripts were used to classify each spot as nuclear or cytoplasmic based on the 3D nuclear DAPI segmentation.

In transcription-site labeling experiments, cells with a GFP spot in the cytoplasm, distant from the nucleus, were excluded from the analysis (only 3 cells across both strains and experiments). In addition, cells without a GFP spot or with more than two GFP spots per nucleus were excluded from the analysis. This removed 3 – 13 % of the mononucleated cells and 15 – 37 % of the binucleated cells. In the binucleated cells, this was typically due to the absence of a detectable GFP spot in one or both of the two nuclei. To identify nascent mRNA FISH spots, the 3D distance of each FISH spot to its nearest transcription site (TS) spot was determined. The distribution of distances was bimodal (Fig. S7B), and a distance cutoff was used to separate nascent from mature mRNA. The distance cutoff was calculated separately for each image generated by a clustering method that locates the minimum between the two distance distributions.

In cells where a transcription site was located just outside the region initially segmented as nuclear based on DNA staining, the segmentation of the nucleus was isotropically expanded to include the transcription site (NucSegmentMethod = 3DNucExpand). This seemed justified, since nuclei in cells where this was the case were on average slightly smaller than those in cells where this was not the case, suggesting that the initial nuclear segmentation had not captured the entire nucleus. Alternatively, these cells were excluded from the analysis (NucSegmentMethod = CF, cells filtered), which did not change the overall results or conclusions.

The number of mRNA molecules per cell were calculated with three different approaches. “Spot counts” are the number of FISH spots detected, not taking into account their intensity. “Intensity counts” were obtained by dividing the amplitude of each spot’s PSF by the median amplitude of the cytoplasmic spots in each image. Since for the genes in this study the vast majority of spots in the

cytoplasm can be assumed to consist of single mRNA molecules, the result is a normalized intensity measure where a typical single-RNA FISH spot will have a value of 1 (Fig. S1C). These normalized intensities were summed to produce an intensity-based estimate of the number of mRNA molecules per cell. The third method ("hybrid counts") combines features of both these methods: the number of mRNAs was calculated by partially converting the normalized intensity estimates into integers (Fig. S1). If a spot had a normalized intensity below the 95th percentile of the cytoplasmic spots in that sample, it was assumed to contain one mRNA molecule. For brighter spots, the number of mRNA molecules per spot was set equal to the normalized intensity (Fig. S1C). The numbers for all spots were summed to obtain the number of mRNA molecules per cell. We consider "hybrid count" to be the most reliable method, since it ignores random intensity noise in "single mRNA" spots from differences in labeling efficiency, but counts more than one mRNA molecule when the intensity of a spot is clearly higher than the typical "single mRNA" spot (as is the case at *rpb1* transcription sites, for example). We have used the "spot count" method for comparison to the Marguerat group data, since we did not have intensity values available for their data.

Size-corrected Fano factor

Cell-size corrected Fano factors were calculated as by Padovan-Merhar and colleagues[58], using linear regression to assess the influence of cell size. We either used cell length as a proxy for cell size, or we calculated volume from cell length and width, making the assumption that *S. pombe* cells are shaped like a cylinder with hemispheres at the two ends.

To determine the Fano factor in a sliding window across cell sizes, cells within a window of 1 μm cell length were combined, and mRNA variance, mean, and Fano factor were calculated. The window was moved by 0.1 μm across all cell sizes. For mono- and binucleated cells combined, data were retained when there were at least 50 cells in the window for pooled data, or 40 cells in the window for single replicates, for at least 20 consecutive windows. For mononucleated cells, these threshold numbers were: 40, 30, 10; for binucleated cells, they were 30, 20, 6.

Analysis of published smFISH data

We re-analyzed smFISH data of *S. pombe* by the Marguerat group[12,40], who had also determined both mRNA numbers and cell size. Genes that appeared cell cycle-regulated were excluded (*ace2, fkh2, mid2, SPAPB17E12.14c, SPAPB1E7.04c*). One dataset (wt.1) for *rpb1* from Sun *et al.* was excluded because 5 % of the cells lacked any mRNA, whereas the highest percentage of cells without *rpb1* mRNA in any other dataset was 0.6 %. A few cells across all datasets had highly unusual cell widths (smaller than 1.5 µm or larger than 4 µm) and were excluded. Cells with a length above 16 µm were excluded as well. The size-corrected Fano factor was calculated in the same way as for our data.

Immunoblotting

Between 6 x 10^7 and 1 x 10^8 cells were washed with water, pelleted, resuspended in 500 µL 0.22 M NaOH, 0.12 M beta-mercaptoethanol, and incubated on ice for 15 min. Trichloroacetic acid was added to a final concentration of 6 % followed by another 10 min incubation on ice. The samples were pelleted, optionally washed with acetone, and either resuspended in 100 µl of 1x HU buffer pH 6.8 (4 M urea, 2.5 % SDS, 100 mM Tris-HCl, 0.5 mM EDTA, 0.005 % bromophenol blue, 10 % glycerol, 0.1 M DTT) or in 1.5x NuPage LDS sample buffer (ThermoFisher NP0007) with 5 % beta-mercaptoethanol. Samples were heated to 70ºC for 7 min, and further lysed with glass beads (Sigma G8772) for 1–2 min at 30 s^{-1} in a Retsch MM 400 ball mill. Extracts were collected by centrifugation and incubated at 70ºC for another 3 min. Samples were run on NuPAGE 4-12 % Bis-Tris polyacrylamide gels (Invitrogen) with MOPS running buffer and NuPAGE antioxidant. Proteins were transferred to Immobilon-P PVDF membrane using an Amersham Semi-Dry Transfer Unit and transfer buffer (39 mM glycine, 48 mM Tris-base, 5–10 % methanol) containing 0.01 % SDS and 1:1,000 NuPAGE antioxidant. The membrane was blocked with 4 % nonfat milk in TBS-T (150 mM NaCl, 20 mM TRIS, 0.05 % Tween-20) for 15–30 min at room temperature. Antibodies were also in 4 % nonfat milk in TBS-T, and membranes were washed with TBS-T between antibody incubation steps. Primary antibodies were: mouse anti-GFP (Roche, 11814460001, 0.4 µg/mL), rabbit anti-Cdc2 (Santa Cruz, SC-53, 0.2 µg/mL), rabbit anti-Mad1[44] (2 µg/mL), rabbit anti-Mad2[130],

rabbit anti-Mad3[131] (serum at 1:500), rabbit anti-Bub1 (peptide: CKPKAGGPGRRRSSN, serum at 1:2,000); secondary antibodies were: goat anti-mouse HRPO or goat anti-rabbit HRPO (Dianova, 115-035-003 and 111-035-003, at 1:10,000). For imaging, membranes were incubated SuperSignal West Dura Extended Duration Substrate or SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoFisher) and chemiluminescence was recorded with a ChemiDoc XRS+ (Bio-Rad). Protein expression was quantified relative to the control sample using the control sample dilution series and Cdc2 as a loading control.

Quantitative PCR

RNA extraction, cDNA library preparation, and quantitative PCR (qPCR) were performed as previously described[45]. Briefly, cells were grown in either EMM (TATA-box mutant experiments), EMM with 0.2 g/L leucine (one *mad1* deletion strain), or YEA (all other) to concentrations of 0.7-1.5 x 10^7 cells/ml and then flash frozen in liquid nitrogen. RNA extraction was followed by DNase-treatment and SuperScript IV reverse transcription with oligo $d(T)_{20}$ primers. *act1* and *cdc2* were used as reference genes in all qPCRs.

RACE-PCR

Cells were grown in YEA and snap frozen in liquid nitrogen (~1 x 10^8 cells). RNA was extracted using acidic phenol chloroform and treated with DNase (Roche DNase I) to remove any remaining DNA contamination. For 3' UTR sequencing, DNase-treated RNA was reverse transcribed with the FirstChoice RLM-RACE kit (Invitrogen AM1700). Either the FirstChoice RLM-RACE kit with SuperScript IV (Invitrogen) or the GeneRacer kit (Invitrogen L150201) with SuperScript III (Invitrogen) was used to prepare the cDNA for 5' UTR sequencing. cDNA was amplified with two rounds of nested PCR targeting the gene of interest. The PCR product was inserted into a pBlueScript vector using Gibson assembly (NEB, E2611) and transformed into NEB 5-alpha competent *E. coli* (NEB, C2992). Vectors were sequenced by Sanger sequencing. Sequences were discarded if the UTR appeared to begin (5' UTR) or end (3' UTR) in the coding sequence of the gene. When possible, to limit the analysis to mature mRNA, sequences were checked for the presence of introns and discarded if introns were found.

SAC protein-GFP imaging in live cells

Cells were grown in EMM or EMM with 0.2 g/L leucine to between 4 x 10^6 and 1.5 x 10^7 cells/ml and mounted in Ibidi chambers (8-well, 80827) coated with lectin (Sigma, L1395, at 50 µg/mL). Cells were imaged on a DeltaVision Elite microscope equipped with a PCO edge sCMOS camera. Imaging was performed with an Olympus 60x/1.42 Plan APO oil objective and GFP filters. Images were taken as 7.2 µm z-stacks with 0.1 µm step size. To bleach autofluorescence, an initial GFP image stack was acquired with 0.1 sec exposures and discarded, before the main GFP image stack was acquired using 0.25-0.40 sec exposures (depending on the protein being imaged). GFP images were deconvolved using SoftWoRx software with two cycles of the ratio method (conservative) and a camera intensity offset of 50. Representative regions of cells were selected and single image slices from the z-stack were extracted at approximately the midplane of the cells. Image intensity was further scaled in ImageJ and Photoshop, being consistent for each GFP-tagged protein.

GO enrichment analysis

GO-slim enrichment analysis was performed with the PANTHER overrepresentation test (http://www.pantherdb.org[132]) (released 2022-02-02), using the GO Ontology database (DOI:10.5281/zenodo.6399963; released 2022-03-22). From the data reported by Thodberg and colleagues[61] only transcripts of non-mitochondrial protein-coding genes were kept. For genes with multiple TSSs, only the TSS with the highest pooled TPM value was retained. Start codon positions were downloaded from Pombase[133] on 2022-06-22. All genes for which the TSS was located downstream of the currently annotated start codon were excluded. This resulted in a final list of 4,648 genes, which was used as reference data for the GO enrichment analysis. Of those genes, 296 had a 5' UTR of 15 nucleotides or shorter, and 451 had a 5' UTR of 20 nucleotides or shorter. Only the most specific subclass from the GO-slim analysis is shown. Fisher's exact test with FDR correction was used to identify statistically significant results.

Acknowledgments

We thank Andrea Ciliberto and Peter Swain for comments on the manuscript, Eric Esposito, Anupreet Kour, Anna Makarov, and Jessie Rogers for contributing experimental data, James Holehouse and Svitlana Braichenko for data processing and useful discussions, Tatiana Boluarte for constructing yeast strains, Hunter Haynie, and Haoyun Yang for preliminary experiments, Erod Keaton Baybay for help in establishing the image analysis pipeline, Yimeng Liu for implementing nuclear thresholding, Ben Unruh and Shihoko Kojima for support in running qPCRs, Samir Chethan for help with image quantification, and Daniel Zenklusen for sharing their yeast FISH protocol. Research reported in this publication was supported by NIGMS (NIH) under award number R35GM119723 (DEW, SH) and R35GM148351 (AS), by NSF grant ECCS-1711548 (AS), by the Leverhulme Trust RPG-2020-327 (RG), and by a Dean's Discovery Fund of Virginia Tech's College of Science (DEW, AS, SH).

Author contributions

Conceptualization, DEW and SH; Software, DEW, RG and AS; Formal analysis, DEW and RG; Investigation, DEW; Writing–Original Draft, SH; Writing–Review & Editing, DEW, RG and AS; Visualization, DEW, RG and SH; Supervision, SH; Funding acquisition: SH, RG and AS.

Competing Interest

None declared.

References

1. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6, 451–464 (2005).

2. Raj, A. & Oudenaarden, A. van. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135, 216–26 (2008).

3. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231–237 (2002).

4. Maheshri, N. & O'Shea, E. K. Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression. *Annu Rev Bioph Biom* 36, 413–434 (2007).

5. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* 123, 1025–1036 (2005).

6. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15, 1263–71 (2008).

7. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology* 4, e309 (2006).

8. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional Pulsing of a Developmental Gene. *Curr Biol* 16, 1018–1025 (2006).

9. So, L. *et al.* General properties of transcriptional time series in Escherichia coli. *Nat Genet* 43, 554–560 (2011).

10. Halpern, K. B. *et al.* Bursty Gene Expression in the Intact Mammalian Liver. *Mol Cell* 58, 147–156 (2015).

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

12. Saint, M. *et al.* Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nature microbiology* 4, 480–491 (2019).

13. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435 (2017).

14. Kussell, E. & Leibler, S. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science* 309, 2075–2078 (2005).

15. Blake, W. J. *et al.* Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Molecular Cell* 24, 853–865 (2006).

16. Carey, J. N. *et al.* Regulated Stochasticity in a Bacterial Signaling Network Permits Tolerance to a Rapid Environmental Change. *Cell* 173, 196-207.e14 (2018).

17. Fraser, D. & Kærn, M. A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol Microbiol* 71, 1333–1340 (2009).

18. Wernet, M. F. *et al.* Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* 440, 174–180 (2006).

19. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544–547 (2008).

20. Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular Decision Making and Biological Noise: From Microbes to Mammals. *Cell* 144, 910–925 (2011).

21. Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biology* 2, e137 (2004).

22. Batada, N. N. & Hurst, L. D. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature Genetics* 39, ng2071 (2007).

23. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular Systems Biology* 4, 170 (2008).

24. Sun, M. & Zhang, J. Allele-specific single-cell RNA sequencing reveals different architectures of intrinsic and extrinsic gene expression noises. *Nucleic Acids Res* 48, 533–547 (2019).

25. Rodriguez, J. & Larson, D. R. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu Rev Biochem* 89, 1–24 (2020).

26. Filatova, T., Popović, N. & Grima, R. Modulation of nuclear and cytoplasmic mRNA fluctuations by time-dependent stimuli: Analytical distributions. *Math Biosci* 347, 108828 (2022).

27. Szavits-Nossan, J. & Grima, R. Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle. *Phys Rev E* 105, 014410 (2022).

28. Bahar Halpern, K. *et al.* Nuclear Retention of mRNA in Mammalian Tissues. *Cell Reports* 13, 2653–2662 (2015).

29. Stoeger, T., Battich, N. & Pelkmans, L. Passive Noise Filtering by Cellular Compartmentalization. *Cell* 164, 1151–61 (2016).

30. Singh, A. & Bokes, P. Consequences of mRNA Transport on Stochastic Variability in Protein Levels. *Biophys J* 103, 1087–1096 (2012).

31. Hansen, M., Desai, R. V., Simpson, M. L. & Weinberger, L. S. Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Systems* (2018) doi:10.1016/j.cels.2018.08.002.

32. Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression. *Trends Cell Biol* 19, 630–637 (2009).

33. Sanchez, A. & Golding, I. Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science* 342, 1188–1193 (2013).

34. Peccoud, J. & Ycart, B. Markovian Modeling of Gene-Product Synthesis. *Theor Popul Biol* 48, 222–234 (1995).

35. Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 105, 17256–61 (2008).

36. Thattai, M. & Oudenaarden, A. van. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* 98, 8614–8619 (2001).

37. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science* 332, 475–478 (2011).

38. Thattai, M. Universal Poisson Statistics of mRNAs with Complex Decay Pathways. *Biophys J* 110, 301–305 (2016).

39. Gandhi, S. J., Zenklusen, D., Lionnet, T. & Singer, R. H. Transcription of functionally related constitutive genes is not coordinated. *Nat Struct Mol Biol* 18, 27–34 (2010).

40. Sun, X.-M. *et al.* Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size. *Curr Biol* (2020) doi:10.1016/j.cub.2020.01.053.

41. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* 329, 533–538 (2010).

42. Jones, D. L., Brewster, R. C. & Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* 346, 1533–1536 (2014).

43. Foreman, R. & Wollman, R. Mammalian gene expression variability is explained by underlying cell state. *Mol Syst Biol* 16, e9146 (2020).

44. Heinrich, S. *et al.* Determinants of robustness in spindle assembly checkpoint signalling. *Nature Cell Biology* 15, ncb2864 (2013).

45. Esposito, E. *et al.* Mitotic checkpoint gene expression is tuned by codon usage bias. *Embo J* 41, e107896 (2022).

46. Musacchio, A. The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr Biol* 25, R1002–R1018 (2015).

47. Lara-Gonzalez, P., Pines, J. & Desai, A. Spindle assembly checkpoint activation and silencing at kinetochores. *Semin Cell Dev Biol* (2021) doi:10.1016/j.semcdb.2021.06.009.

48. Kops, G. J. P. L., Snel, B. & Tromer, E. C. Evolutionary Dynamics of the Spindle Assembly Checkpoint in Eukaryotes. *Curr Biol* 30, R589–R602 (2020).

49. Fischer, E. S. Kinetochore-catalyzed MCC formation: A structural perspective. *Iubmb Life* (2022) doi:10.1002/iub.2697.

50. Chung, E. & Chen, R.-H. Spindle Checkpoint Requires Mad1-bound and Mad1-free Mad2. *Mol Biol Cell* 13, 1501–1511 (2002).

51. Ryan, S. D. *et al.* Up-regulation of the mitotic checkpoint component Mad1 causes chromosomal instability and resistance to microtubule poisons. *Proc National Acad Sci* 109, E2205–E2214 (2012).

52. Öztürk, M. *et al.* Proteome effects of genome-wide single gene perturbations. *Nat Commun* 13, 6153 (2022).

53. Chen, R.-H., Brady, D. M., Smith, D., Murray, A. W. & Hardwick, K. G. The Spindle Checkpoint of Budding Yeast Depends on a Tight Complex between the Mad1 and Mad2 Proteins. *Mol Biol Cell* 10, 2607–2618 (1999).

54. Sironi, L. *et al.* Crystal structure of the tetrameric Mad1–Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. *Embo J* 21, 2496–2506 (2002).

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

56. Raser, J. M. & O'Shea, E. K. Noise in Gene Expression: Origins, Consequences, and Control. *Science* 309, 2010–2013 (2005).

57. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* 297, 1183–1186 (2002).

58. Padovan-Merhar, O. *et al.* Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Molecular Cell* 58, 339–352 (2015).

59. Zhurinsky, J. *et al.* A Coordinated Global Control over Cellular Transcription. *Current Biology* 20, 2010–2015 (2010).

60. Kempe, H., Schwabe, A., Crémazy, F., Verschure, P. J. & Bruggeman, F. J. The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Molecular Biology of the Cell* 26, 797–804 (2015).

61. Thodberg, M. *et al.* Comprehensive profiling of the fission yeast transcription start site activity during stress and media response. *Nucleic acids research* (2018) doi:10.1093/nar/gky1227.

62. Kozak, M. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expression* 1, 111–5 (1991).

63. Faure, A. J., Schmiedel, J. M. & Lehner, B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst* 5, 471-484.e4 (2017).

64. Hornung, G. *et al.* Noise–mean relationship in mutated promoters. *Genome Research* 22, 2409–2417 (2012).

65. Dadiani, M. *et al.* Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. *Genome Research* 23, 966–976 (2013).

66. Sharon, E. *et al.* Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Research* 24, 1698–1706 (2014).

67. Hocine, S., Vera, M., Zenklusen, D. & Singer, R. H. Promoter-Autonomous Functioning in a Controlled Environment using Single Molecule FISH. *Sci Rep-uk* 5, 9934 (2015).

68. Chen, X. & Zhang, J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell Systems* 2, 347–354 (2016).

69. Li, H. *et al.* Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *Rna Biol* 12, 525–37 (2015).

70. Matsuyama, A. *et al.* pDUAL, a multipurpose, multicopy vector capable of chromosomal integration in fission yeast. *Yeast* 21, 1289–1305 (2004).

71. Raser, J. M. & O'Shea, E. K. Control of Stochasticity in Eukaryotic Gene Expression. *Science* 304, 1811–1814 (2004).

72. Birkenbihl, R. P. & Subramani, S. The rad21 Gene Product of Schizosaccharomyces pombe Is a Nuclear, Cell Cycle-regulated Phosphoprotein (∗). *J Biol Chem* 270, 7703–7711 (1995).

73. Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nat Genet* 36, 809–817 (2004).

74. Peng, X. *et al.* Identification of Cell Cycle-regulated Genes in Fission Yeast. *Mol Biol Cell* 16, 1026–1042 (2005).

75. Hayles, J. & Nurse, P. Introduction to Fission Yeast as a Model System. *Cold Spring Harb Protoc* 2018, pdb.top079749 (2017).

76. Carlson, C. R., Grallert, B., Stokke, T. & Boye, E. Regulation of the start of DNA replication in Schizosaccharomyces pombe. *J Cell Sci* 112, 939–946 (1999).

77. Krenning, L., Sonneveld, S. & Tanenbaum, M. Time-resolved single-cell sequencing identifies multiple waves of mRNA decay during the mitosis-to-G1 phase transition. *Elife* 11, e71356 (2022).

78. Peterson, J. R., Cole, J. A., Fei, J., Ha, T. & Luthey-Schulten, Z. A. Effects of DNA replication on mRNA noise. *Proceedings of the National Academy of Sciences* 112, 15886–15891 (2015).

79. Voichek, Y., Bar-Ziv, R. & Barkai, N. Expression homeostasis during DNA replication. *Science* 351, 1087–1090 (2016).

80. Müller, C. A. & Nieduszynski, C. A. DNA replication timing influences gene expression level. *J Cell Biol* 216, 1907–1914 (2017).

81. Trcek, T., Larson, D. R., Moldón, A., Query, C. C. & Singer, R. H. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* 147, 1484–97 (2011).

82. Hausnerová, V. V. & Lanctôt, C. Transcriptional Output Transiently Spikes Upon Mitotic Exit. *Sci Rep-uk* 7, 12607 (2017).

83. Tsirkas, I. *et al.* Transcription-replication coordination revealed in single live cells. *Nucleic Acids Res* 50, 2143–2156 (2022).

84. Fu, X. *et al.* Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions. *Elife* 11, (2022).

85. Jia, C. & Grima, R. Coupling gene expression dynamics to cell size dynamics and cell cycle events: exact and approximate solutions of the extended telegraph model. *Iscience* 105746 (2022) doi:10.1016/j.isci.2022.105746.

86. Skinner, S. O. *et al.* Single-cell analysis of transcription kinetics across the cell cycle. *Elife* 5, e12175 (2016).

87. Wang, M., Zhang, J., Xu, H. & Golding, I. Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality. *Nature microbiology* (2019) doi:10.1038/s41564-019-0553-z.

88. Saroufim, M.-A. *et al.* The nuclear basket mediates perinuclear mRNA scanning in budding yeast. *J Cell Biol* 211, 1131–1140 (2015).

89. Pedraza, J. M. & Paulsson, J. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* 319, 339–343 (2008).

90. Braichenko, S., Holehouse, J. & Grima, R. Distinguishing between models of mammalian gene expression: telegraph-like models versus mechanistic models. *J Roy Soc Interface* 18, 20210510 (2021).

91. Cao, Z., Filatova, T., Oyarzún, D. A. & Grima, R. A stochastic model of gene expression with polymerase recruitment and pause release. *Biophys J* (2020) doi:10.1016/j.bpj.2020.07.020.

92. Karmakar, R. & Das, A. K. Effect of transcription reinitiation in stochastic gene expression. *J Statistical Mech Theory Exp* 2021, 033502 (2021).

93. Lloyd-Price, J. *et al.* Dissecting the stochastic transcription initiation process in live Escherichia coli. *Dna Res Int J Rapid Publ Reports Genes Genomes* 23, 203–214 (2016).

94. Cramer, P. Organization and regulation of gene transcription. *Nature* (2019) doi:10.1038/s41586-019-1517-4.

95. Booth, G. T., Wang, I. X., Cheung, V. G. & Lis, J. T. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Research* 26, 799–811 (2016).

96. Cao, D. & Parker, R. Computational modeling of eukaryotic mRNA turnover. *Rna* 7, 1192–1212 (2001).

97. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* 136, 763–776 (2009).

98. Deneke, C., Lipowsky, R. & Valleriani, A. Complex Degradation Processes Lead to Non-Exponential Decay Patterns and Age-Dependent Decay Rates of Messenger RNA. *Plos One* 8, e55442 (2013).

99. Ochiai, H. *et al.* Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Sci Adv* 6, eaaz6699 (2020).

100. Muthukrishnan, A.-B. *et al.* Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Research* 40, 8472–8483 (2012).

101. Kandhavelu, M. *et al.* In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps. *BMC Systems Biology* 5, 149 (2011).

102. Mäkelä, J. *et al.* In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucleic Acids Res* 41, 6544–6552 (2013).

103. Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of Transcriptional Bursting in Bacteria. *Cell* 158, 314–326 (2014).

104. Lionnet, T., Wu, B., Grünwald, D., Singer, R. H. & Larson, D. R. Nuclear Physics Quantitative Single-Cell Approaches to Nuclear Organization and Gene Expression. *Cold Spring Harb Sym* 75, 113–126 (2010).

105. Choubey, S., Kondev, J. & Sanchez, A. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. *Plos Comput Biol* 11, e1004345 (2015).

106. Torre, E. *et al.* Rare Cell Detection by Single-Cell RNA Sequencing as Guided by Single-Molecule RNA FISH. *Cell Syst* 6, 171-179.e5 (2018).

107. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proceedings of the National Academy of Sciences* 108, E67–E76 (2011).

108. Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F. & Wittkopp, P. J. Selection on noise constrains variation in a eukaryotic promoter. *Nature* 521, 344–347 (2015).

109. Duveau, F. *et al.* Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. *eLife* 7, e37272 (2018).

110. Zhang, Z., Qian, W. & Zhang, J. Positive selection for elevated gene expression noise in yeast. *Molecular Systems Biology* 5, 299 (2009).

111. Mittal, C., Lang, O., Lai, W. K. M. & Pugh, B. F. An integrated SAGA and TFIID PIC assembly pathway selective for poised and induced promoters. *Gene Dev* 36, 985–1001 (2022).

112. Hong, C. K., Ramu, A., Zhao, S. & Cohen, B. A. Effect of genomic and cellular environments on gene expression noise. *Biorxiv* 2022.08.31.506082 (2022) doi:10.1101/2022.08.31.506082.

113. Nieuwkoop, T., Finger-Bou, M., Oost, J. van der & Claassens, N. J. The Ongoing Quest to Crack the Genetic Code for Protein Production. *Mol Cell* 80, (2020).

114. Singh, A. Negative Feedback Through mRNA Provides the Best Control of Gene-Expression Noise. *IEEE Transactions on NanoBioscience* 10, 194–200 (2011).

115. Iyer-Biswas, S. & Jayaprakash, C. Mixed Poisson distributions in exact solutions of stochastic autoregulation models. *Phys Rev E* 90, 052712 (2014).

116. Sturrock, M., Li, S. & Shahrezaei, V. The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression. *J Theor Biol* 424, 55–72 (2017).

117. Bokes, P., Hojcka, M. & Singh, A. MicroRNA Based Feedforward Control of Intrinsic Gene Expression Noise. *Ieee Acm Transactions Comput Biology Bioinform* 18, 272–282 (2019).

118. Czuppon, P. & Pfaffelhuber, P. Limits of noise for autoregulated gene expression. *Journal of mathematical biology* 77, 1153–1191 (2018).

119. Swain, P. S. Efficient Attenuation of Stochasticity in Gene Expression Through Post-transcriptional Control. *J Mol Biol* 344, 965–976 (2004).

120. Bertolini, M. *et al.* Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science* 371, 57–64 (2021).

121. Roguev, A., Wiren, M., Weissman, J. S. & Krogan, N. J. High-throughput genetic interaction mapping in the fission yeast Schizosaccharomyces pombe. *Nat Methods* 4, 861–866 (2007).

122. Jacobs, J. Z., Ciccaglione, K. M., Tournier, V. & Zaratiegui, M. Implementation of the CRISPR-Cas9 system in fission yeast. *Nat Commun* 5, 5344–5344 (2014).

123. Watson, A. T., Garcia, V., Bone, N., Carr, A. M. & Armstrong, J. Gene tagging and gene replacement using recombinase-mediated cassette exchange in Schizosaccharomyces pombe. *Gene* 407, 63–74 (2008).

124. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast* 21, 661–670 (2004).

125. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science* 296, 913–916 (2002).

126. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast Chichester Engl* 14, 943–51 (1998).

127. Rohner, S., Gasser, S. M. & Meister, P. Modules for cloning-free chromatin tagging in Saccharomyces cerevisae. *Yeast* 25, 235–239 (2008).

128. Arganda-Carreras, I. *et al.* Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* 33, 2424–2426 (2017).

129. Mueller, F. *et al.* FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat Methods* 10, 277–278 (2013).

130. Yamada, H. Y., Matsumoto, S. & Matsumoto, T. High dosage expression of a zinc finger protein, Grt1, suppresses a mutant of fission yeast slp1(+), a homolog of CDC20/p55CDC/Fizzy. *J Cell Sci* 113, 3989–3999 (2000).

131. Sewart, K. & Hauf, S. Different Functionality of Cdc20 Binding Sites within the Mitotic Checkpoint Complex. *Current biology : CB* 27, 1213–1220 (2017).

132. Mi, H. *et al.* Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc* 14, 703–721 (2019).

133. Harris, M. A. *et al.* Fission stories: using PomBase to understand Schizosaccharomyces pombe biology. *Genetics* 220, iyab222 (2021).

Figure 1
Weidemann et al.

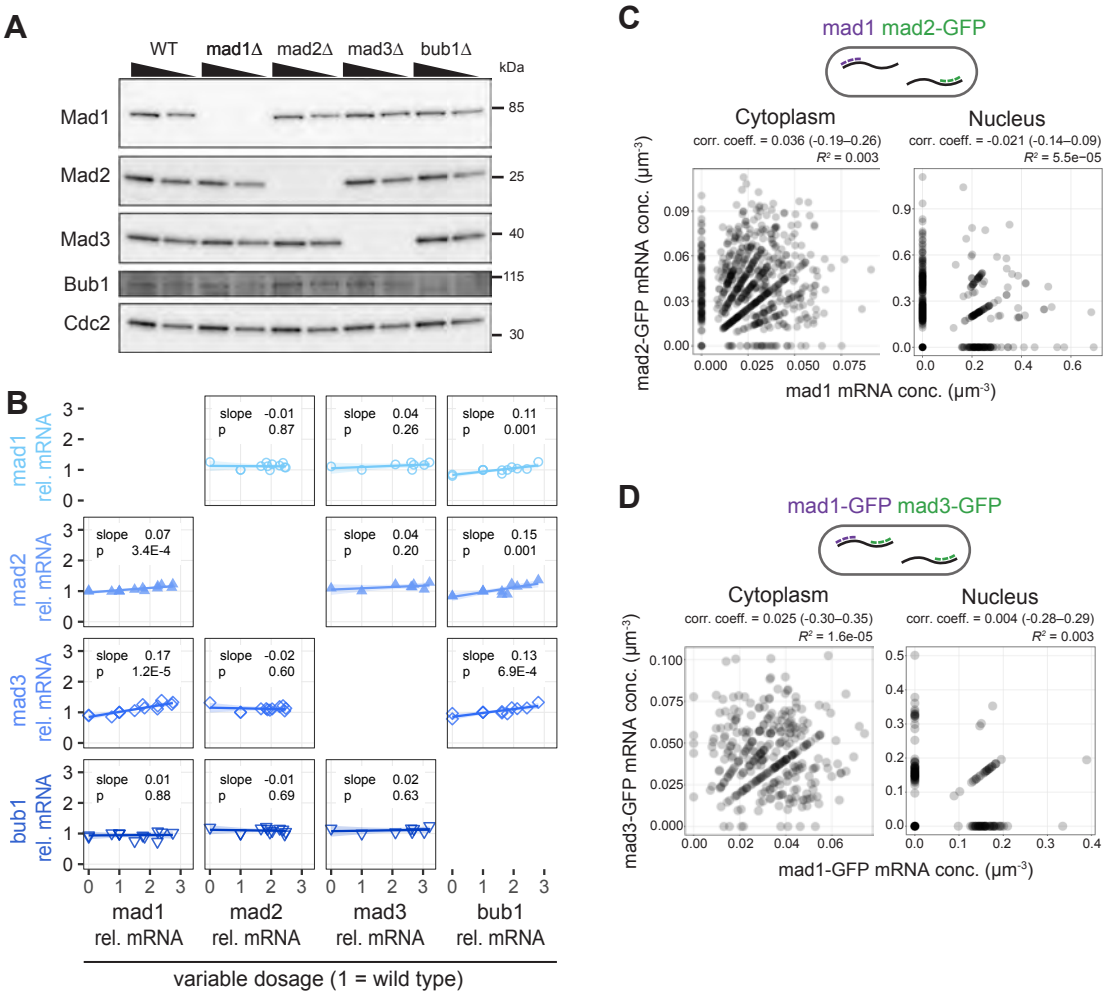

Figure 1 - Expression of individual SAC genes is not correlated with other SAC genes
(A) Immunoblot of cell extracts from the indicated strains; 70 % of the extract is loaded in each second lane. Antibodies against the endogenous proteins. Cdc2 serves as loading control.
(B) Strains in which SAC genes were either deleted (dosage = 0) or slightly overexpressed (dosage > 1) by expression from a second genetic locus were analyzed for the concentration of SAC gene mRNAs by quantitative PCR; the mRNA concentration is given relative to that in a wild-type strain. Parameters of linear regression are shown.
(C) Both mad1 and mad2-GFP mRNA concentrations were determined in the same cells by single-molecule mRNA FISH (n = 734 cells). Cells expressed untagged mad1 and mad2-GFP from their respective endogenous locus; mad1 mRNA was targeted with mad1-specific probes, mad2 mRNA with probes against GFP. Nuclei were stained by DAPI. The correlation coefficient was calculated as Kendall's tau-b; the 95 % confidence interval is given in brackets.
(D) Both mad1-GFP and mad3-GFP mRNA concentrations were determined in the same cells by single-molecule mRNA FISH (n = 329 cells). Cells expressed mad1-GFP and mad3-GFP from their respective endogenous locus; mad1-specific probes and probes against GFP were used. GFP dots co-localizing with mad1 were considered mad1-GFP, all others were considered mad3-GFP. Nuclei were stained by DAPI. The correlation coefficient was calculated as Kendall rank correlation coefficient, tau-b; the 95 % confidence interval is given in brackets.

Figure 2
Weidemann et al.

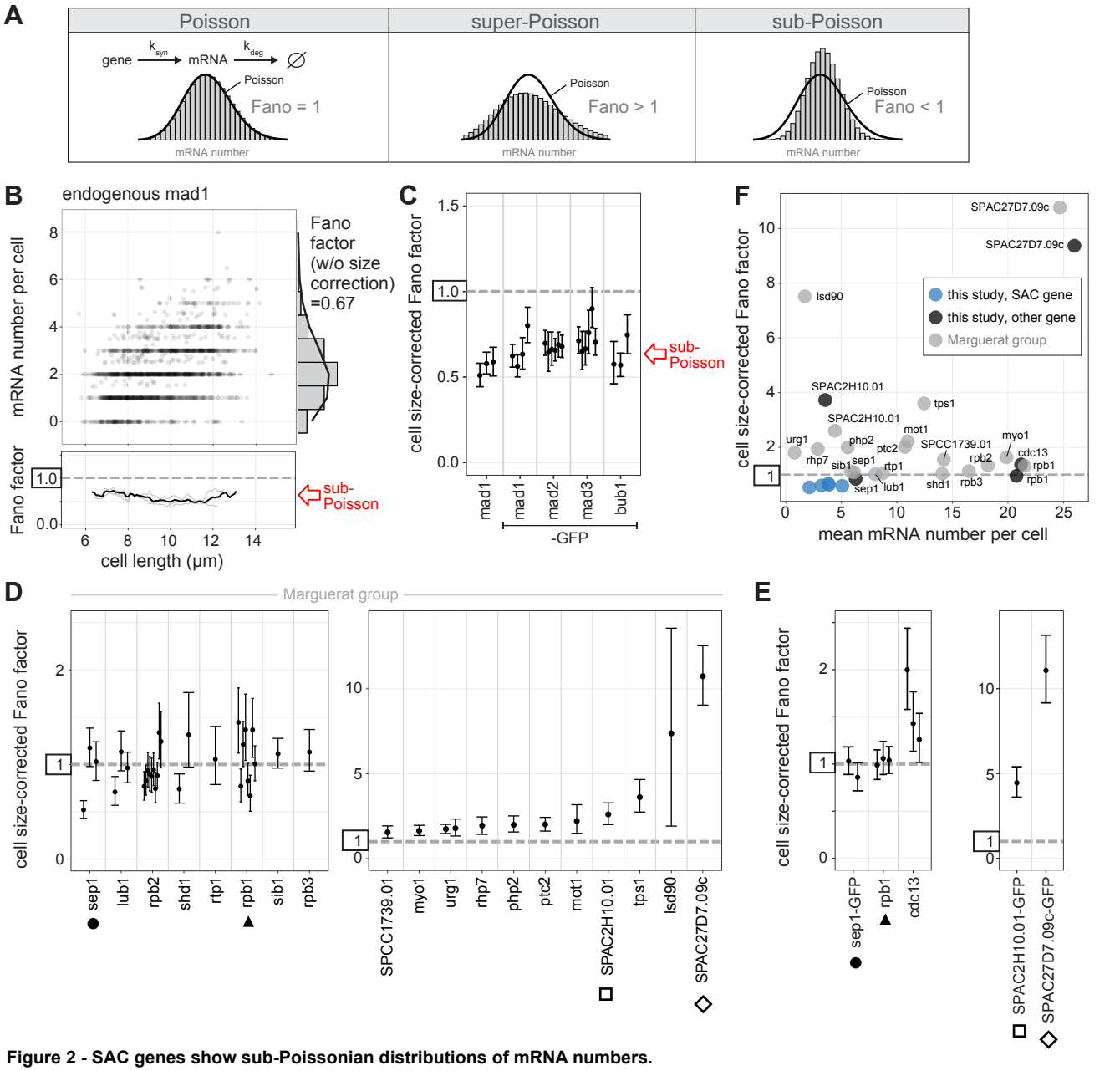

Figure 2 - SAC genes show sub-Poissonian distributions of mRNA numbers.
(A) Constitutive gene expression with a single rate-limiting step in transcription and mRNA degradation yields a Poisson distribution of mRNA numbers per cell (variance = mean, Fano factor = 1). Super-Poissonian distributions have a larger variance than the mean (Fano factor > 1), sub-Poissonian distributions have a lower variance than the mean (Fano factor < 1). Here, a negative binomial and a binomial distribution are used as illustration. Black lines are Poisson distributions with the same mean as the data in the histogram.
(B) Top: Scatter plot of cell length versus mRNA number per cell for untagged mad1 (n = 1,619 cells; data from 3 replicates combined). Right: Histogram of mRNA number across all cells with fit to Poisson distribution (black line). Bottom: The Fano factor was determined in a sliding window spanning 1 µm of cell length. The Fano factors for single replicates are shown in light gray; the Fano factor for the pooled data in black. See Fig. S2 for other SAC genes.
(C) Cell size-corrected Fano factors and their 95 % confidence interval were calculated from smFISH data as in Padovan-Merhar et al., Mol Cell 2015. At least three replicate experiments for each gene; the number of cells in single replicates ranged from 209 to 778.
(D) Cell size-corrected Fano factors and their 95 % confidence interval were calculated as in (C) using published smFISH data by the Marguerat group (Saint et al., Nature Microbiology 2019; Sun et al., Curr Biol 2020). Genes that were also analyzed in this study (see E) are marked. Between 1 and 10 replicate experiments per gene; the number of cells in single replicates ranged from 103 to 441.
(E) Cell size-corrected Fano factors and their 95 % confidence interval for non-SAC genes were calculated as in (C). Genes also analyzed by the Marguerat group (see D) are marked. Between 1 and 3 replicate experiments per gene; the number of cells in single replicates ranged from 236 to 676.
(F) Cell size-corrected Fano factors were plotted against the mean mRNA number per cell for individual genes. Data from replicates for each gene were pooled to determine the cell size-corrected Fano factor. Genes analyzed in this study are shown in black; SAC genes from this study in blue; genes analyzed by the Marguerat group are shown in gray. Between 236 and 3,501 cells per gene for our data, and between 103 and 2,564 cells per gene for the Marguerat group's data. For data from this study the "spot count" for mRNA number was used for better comparability to the Marguerat data; panels (B), (C), and (E) use the "hybrid count" for mRNA number (see Fig. S1).

Figure 3
Weidemann et al.

Figure 3 - 5' UTR lengths of SAC genes range from untypically short to untypically long
(A) Lengths of 5' and 3' untranslated regions (UTRs) determined by RACE-PCR; black, endogenous gene; green, ymEGFP-tagged gene. For the 5' UTRs, the spread in lengths and the most frequent length is given at the bottom.
(B) Lengths of SAC gene 5' UTRs in comparison to genome-wide data from Thodberg et al., 2019, displayed as empirical cumulative distribution (4,710 transcripts).

Figure 4
Weidemann et al.

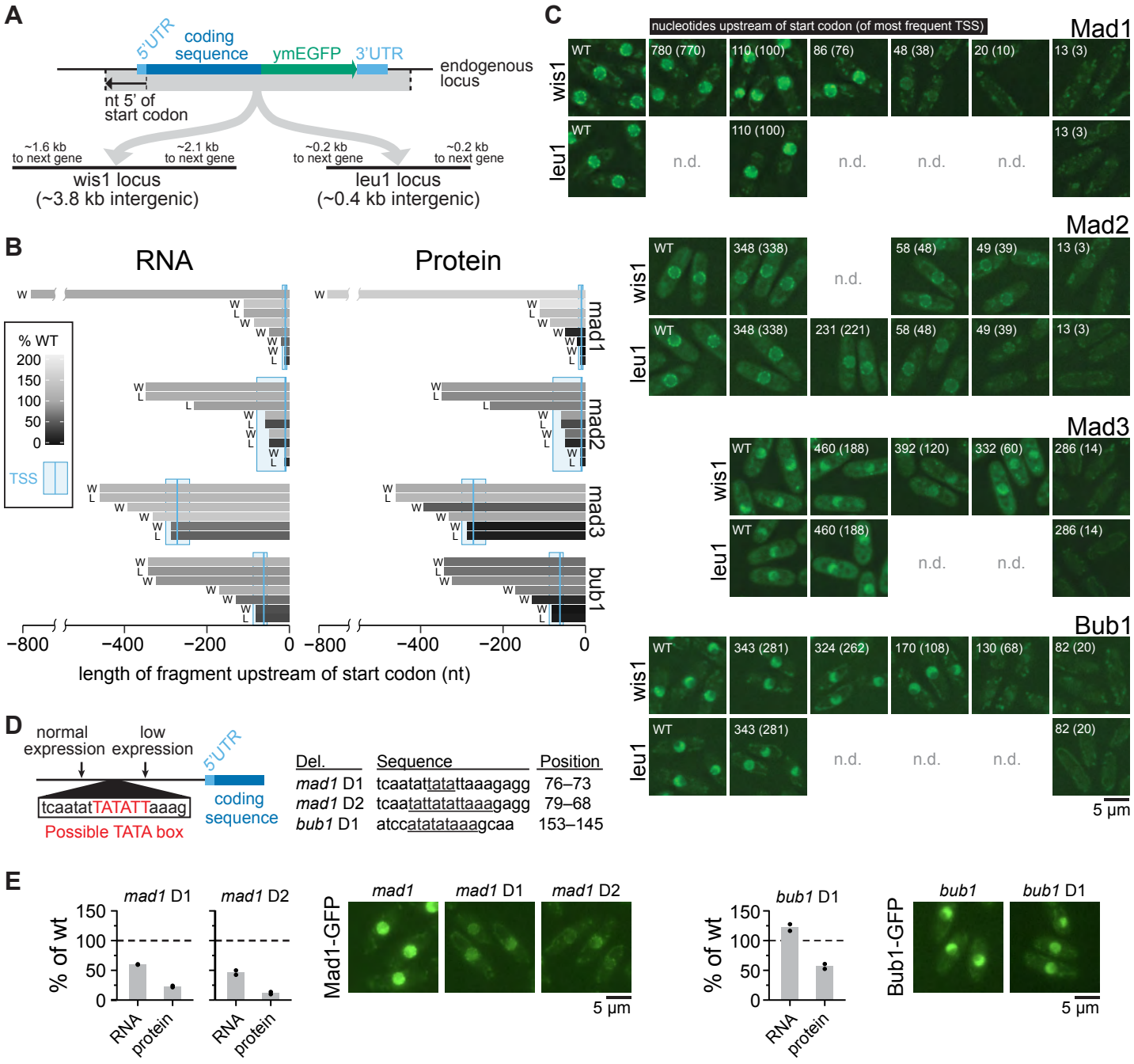

Figure 4 - SAC genes have short promoter regions

(A) Diagram showing design of integrating SAC genes at two exogenous, intergenic loci.

(B) Amount of RNA (by qPCR) and protein (by immunoblot quantification, see Fig. S4) from exogenous locus expression (W = wis1 locus, L = leu1 locus) relative to expression from the endogenous locus. The length of bars shows the length of the upstream region (promoter and 5'UTR), shades of gray show the mRNA or protein amount. qPCR primers and immunoblot antibody targeted the GFP tag.

(C) Representative live-cell images showing GFP fluorescence from GFP-tagged genes at the endogenous locus (WT) or gene fragments inserted at an exogenous locus (leu1 or wis1). Numbers indicate distance in nucleotides that the exogenous locus gene fragment extends upstream of the start codon, or upstream of the most prominent TSS (in brackets). Gamma correction for images to make both high and low expression cells visible: Mad1 = 2.0, Mad2 = 1.8, Mad3 = 1.0, and Bub1 = 1.5.

(D) Schematic (left) and table (right) of deletions to test the functionality of possible TATA box sequences. Deleted bases are underlined in the table and their positions are given in bp upstream of the start codon.

(E) Relative expression of RNA (by qPCR) and protein (by immunoblot quantification, see Fig. S4) from deletion mutants in (D) normalized to strains with wild-type promoter sequences (two biological replicates each). One-sample t-tests: mad1 D1 RNA p = 0.007, mad1 D1 protein p = 0.01, mad1 D2 RNA p = 0.04, mad1 D2 protein p = 0.01, bub1 D1 RNA p = 0.16, bub1 D1 protein p = 0.06. Representative live-cell images showing GFP fluorescence of the wild-type gene or the TATA box mutants, all expressed from the exogenous leu1 locus (gamma correction: 1.5).

Figure 5
Weidemann et al.

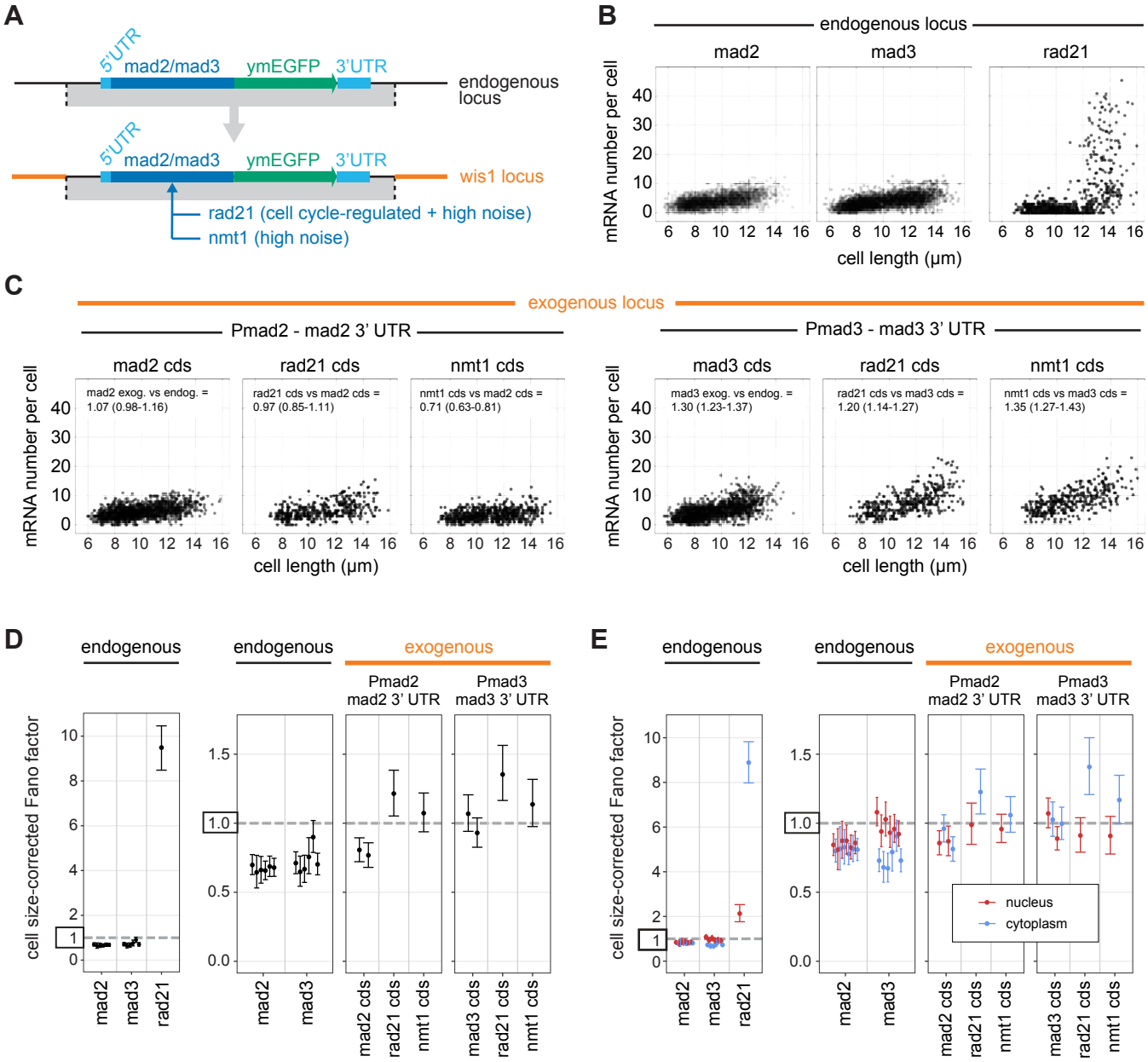

Figure 5 - Small surrounding regions are sufficient for low-noise expression, but not for sub-Poissonian mRNA distributions.
(A) Design of the experiment: *mad2*-GFP and the region surrounding it or *mad3*-GFP and the region surrounding it were integrated at the large intergenic locus adjacent to the wis1 gene. At this location, the coding-sequences of *mad2* or *mad3* were then replaced with those of *rad21*, a gene showing cell cycle-regulated and high-noise expression, or *nmt1*, a high-noise gene. The region around *mad2*-GFP ranged from 48 nucleotides 5' of the most prominent TSS to 521 nucleotides 3' of the stop codon; the region around *mad3*-GFP ranged from 270 nucleotides 5' of the most prominent TSS to 279 nucleotides 3' of the stop codon.
(B) mRNA number relative to cell length at the endogenous locus with the endogenous promoter (*mad2*: 3,501 cells from 6 independent experiments; *mad3*: 2,993 cells from 6 independent experiments; *rad21*: 678 cells from 1 experiment; data for *mad2* and *mad3* are the same as in Fig. 2).
(C) mRNA number relative to cell length at the exogenous locus for expression from the *mad2* promoter (*Pmad2*) and *mad2* downstream region or *mad3* promoter (*Pmad3*) and *mad3* downstream region (*Pmad2-mad2* cds: 1,308 cells from 2 independent experiments; *Pmad2-rad21* cds: 375 cells from 1 experiment; *Pmad2-nmt1* cds: 557 cells from 1 experiment; *Pmad3-mad3* cds: 1,586 cells from 2 independent experiments; *Pmad3-rad21* cds: 368 cells from 1 experiment; *Pmad3-nmt1* cds: 342 cells from 1 experiment). Generalized linear mixed model estimates for the change in mRNA number per cell between conditions with 95 % confidence interval in brackets are given on top; also see Fig. S6C.
(D) Cell size-corrected Fano factors and their 95 % confidence interval were calculated from smFISH data as in Fig. 2 using the data shown in (B) and (C). Between 1and 6 replicate experiments for each genotype; the number of cells in single replicates ranged from 209 to 794.
(E) As in (D), but nuclei were segmented using DAPI staining, and the cell size-corrected Fano factor was calculated separately for mRNA localizing to the nucleus and localizing to the cytoplasm.

Figure 6
Weidemann et al.

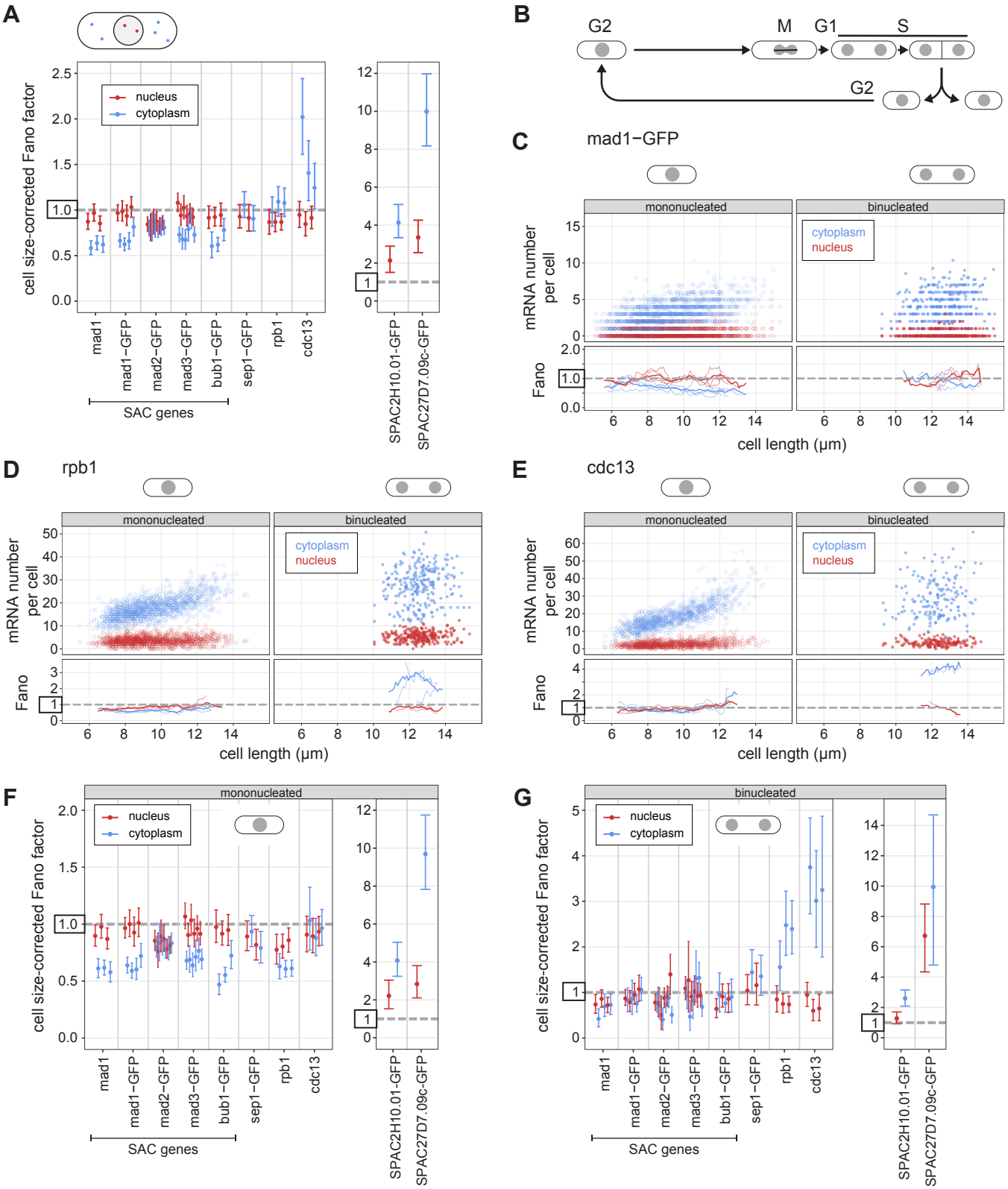

Figure 6 - The Fano factor of the mRNA distribution can be lower in the cytoplasm than in the nucleus.
(A) Cell size-corrected Fano factors and their 95 % confidence interval were calculated from smFISH data. Same experiments as in Fig. 2, but nuclei were segmented using DAPI staining, and the cell size-corrected Fano factor was calculated separately for mRNA localizing to the nucleus and localizing to the cytoplasm. Between 1and 6 replicate experiments for each genotype; the number of cells in single replicates ranged from 209 to 778. **(B)** Schematic of the *S. pombe* cell cycle. **(C-E)** Top: Scatter plot of cell length versus mRNA number in cytoplasm (blue) or nucleus (red). Mono- and binucleated cells are shown separately. Bottom: The Fano factor was determined in a sliding window spanning 1 µm of cell length. The Fano factors for single replicates are shown as thin lines; the Fano factor for the pooled data as thick line. See Fig. S8 for other SAC genes. **(F,G)** As in (A), but analyzed separately for mononucleated (F) and binucleated (G) cells.

Figure 7
Weidemann et al.

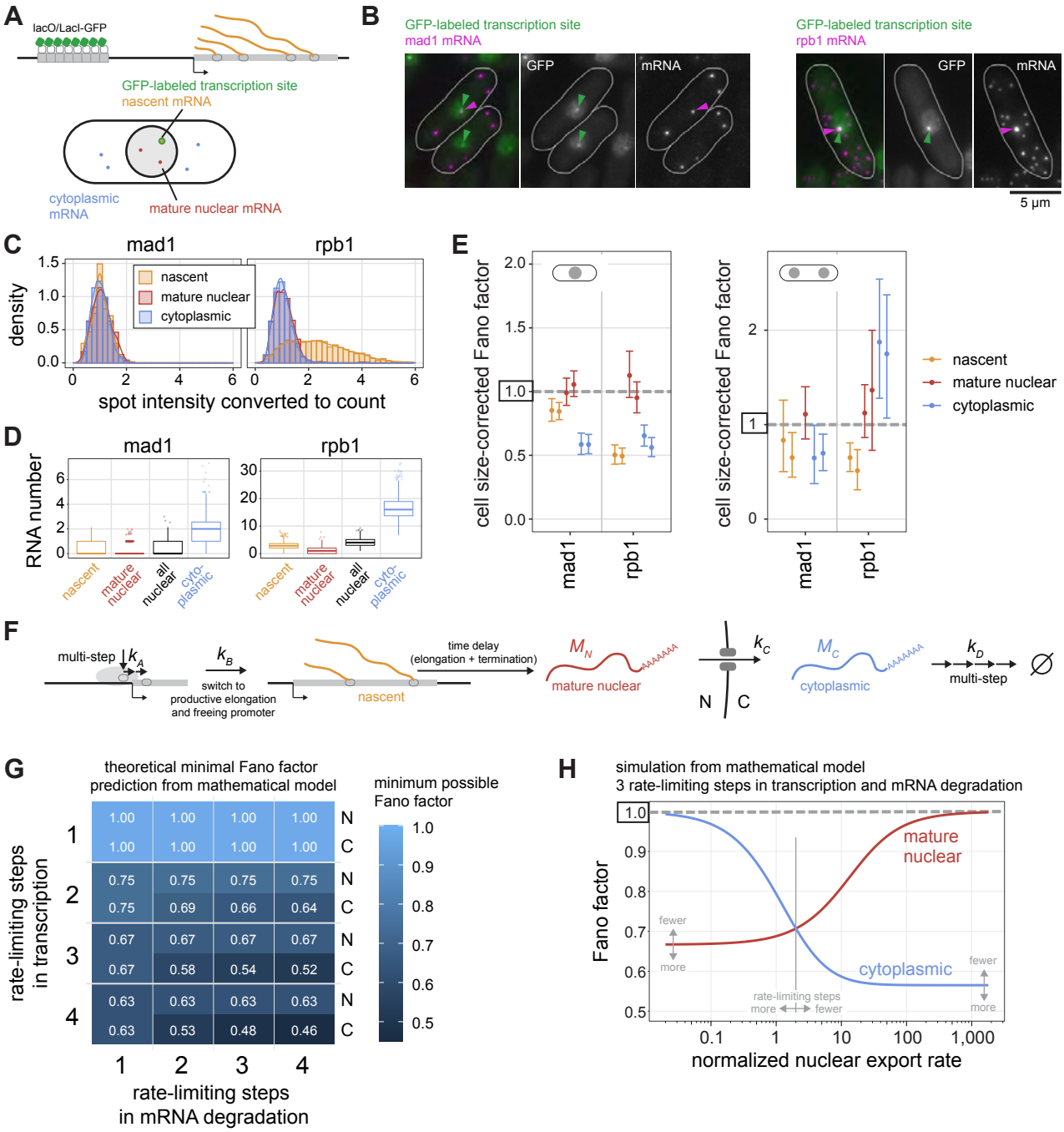

Figure 7 - A sub-Poissonian mRNA distribution in the cytoplasm requires multiple rate-limiting steps in mRNA degradation and is modulated by nuclear export kinetics. (A) Diagram of transcription site labeling, and classifying mRNA into nascent (associated with the GFP-labeled gene), mature nuclear, or cytoplasmic. **(B)** Representative images of mad1 or rpb1 mRNA FISH in cells with the transcription site labeled by GFP. **(C)** Histogram and density distribution of FISH spot intensity at the transcription site (nascent), at other positions in the nucleus (mature nuclear), or in the cytoplasm. Data were normalized to the median of the spot intensity in the cytoplasm for each image. Pooled data from two independent experiments; n = 2,217 spots for mad1; n = 20,740 spots for rpb1. See Fig. S9 for the independent experiments. **(D)** Numbers of mRNA in the different compartments in mononucleated cells; n = 884 cells for mad1, n = 921 cells for rpb1. Boxplots show median and interquartile range; whiskers extend to values no further than 1.5 times the interquartile range from the first and third quartile, respectively. **(E)** Cell size-corrected Fano factors and their 95 % confidence interval were calculated from mRNA FISH data for nascent, mature nuclear, or cytoplasmic mRNA. Mono- and binucleated cells were analyzed separately. Two replicate experiments for each genotype; one of the experiments did not have enough binucleated cells with mature nuclear mad1 mRNA for a reliable estimate of the cell size-corrected Fano factor. The number of cells in single replicates ranged from 343 to 541 for mononucleated and 32 to 98 for binucleated cells. **(F)** Outline of the parameters in the mathematical model, including multiple rate-limiting steps in mRNA synthesis and degradation. **(G)** Analytical results from the mathematical model: minimum possible Fano factors in nucleus (N) and cytoplasm (C) depending on the number of rate-limiting steps in transcription and mRNA degradation. **(H)** Expectation for the nuclear and cytoplasmic Fano factors for a model with three rate-limiting steps in both transcription and mRNA degradation; $k_A = k_B = 10$; $k_D = 3$. The x-axis shows the nuclear export rate k_C, which is defined relative to the effective mRNA degradation rate. The nuclear export rate at which the Fano factors are equal scales inversely with the number of rate-limiting steps in transcription and mRNA degradation (Supplemental Information, Table B-XIII). Gray arrows illustrate in which direction the curves are expected to change for fewer or more rate-limiting steps in transcription and mRNA degradation.