Promoter-mediated transcriptional burst dynamics underlying the expansion of an actin gene family

Edward Tunnacliffe[1], Adam Corrigan[1], Jonathan R Chubb[1]

[1]University College London

During the evolution of gene families, functional diversification of proteins often follows gene duplication. However, many gene families expand while preserving protein sequence. Why do cells maintain multiple copies of the same gene? Here we have addressed this question for an actin family with 17 genes encoding an identical protein. The genes have divergent flanking regions and are scattered throughout the genome. Surprisingly, almost the entire family showed similar developmental expression profiles, with their expression also strongly coupled in single cells. Using live cell imaging, we show that differences in gene expression were apparent over shorter timescales, with family members displaying different transcriptional bursting dynamics. Strong 'bursty' behaviours contrasted steady, more continuous activity, indicating different regulatory inputs to individual actin genes. To determine the sources of these different dynamic behaviours, we reciprocally exchanged the upstream regulatory regions of gene family members. This revealed that dynamic transcriptional behaviour is directly instructed by upstream sequence, rather than features specific to genomic context. A residual minor contribution of genomic context modulates the gene OFF rate. Our data suggest promoter diversification following gene duplication expands the range of stimuli that can regulate the expression of essential genes. These observations contextualize the significance of transcriptional bursting.

transcriptional bursting | stochastic gene expression | single cell transcriptomics | Dictyostelium | gene family

Introduction

Gene duplication is recognised as an important process for generating complexity in evolution (1). Following duplication, gene sequences are present in at least two copies in the genome. Assuming these sequences are identical and subject to the same regulatory constraints, they will perform the same function— they are redundant. Over time, duplicate genes typically diverge in sequence and function, however in some cases strong selection acts to maintain identical amino acid or nucleotide sequences over long periods of evolution. Examples include histones, where humans have 14 genes for histone H4, each encoding the same protein (2). Similarly, ribosomal RNA genes are present in hundreds to thousands of copies in eukaryotes with extremely high sequence conservation between family members (3).

Why does an organism require so many genes encoding an apparently identical end product? One explanation is that a large amount of gene product is required and multiple genes allow more transcription. However, while histone genes can be under coordinate control during the cell cycle (4), they have different promoter elements and show varying contributions to total histone content in normal and cancer cells, suggesting regulatory differences between family members.

To understand how differences in gene regulation have influenced the evolution of multigene families, we investigated the actin gene family of the amoeba *Dictyostelium discoideum*. This organism has more than 30 actin genes, of which 17 (the *act8* group) encode an identical amino acid sequence (5). *Act8* family genes produce more than 95% of total actin (6) and are dispersed throughout the genome (5). This actin family organisation is a broadly applicable evolutionary strategy, shared by species that diverged more than 400 million years ago (Table S1)(7).

Dictyostelium cells are highly motile, so may require lots of actin, perhaps beyond the production capacity of a single gene. However, estimates of their actin content are of the same order as skeletal muscle, which derives its actin from only one gene (8, 9). Divergent flanking sequences (10, 11) and different genomic contexts of the *act8* genes suggest different regulatory dynamics and responses— for example during development. The expansion of the family may also buffer against gene expression noise— it may be undesirable for the expression of an essential protein to be unpredictable, and additional genes may average out noise.

Here, we evaluate the potential for different regulatory dynamics within the gene family. The family shows comparatively similar expression profiles over development, and strong coupling between genes in single cells. However, the genes differ in the dynamics of their transcriptional bursts and show different bursting responses upon induction of development. Switching promoters of actin genes demonstrates that transcriptional dynamics are instructed predominantly by upstream sequence, rather than genomic context.

Results

Developmental dynamics of actin gene expression

Having multiple genes encoding the same protein dispersed throughout the genome may have enabled diversification and refinement of actin expression. Consistent with this view, there is considerable diversity of upstream regulatory sequence between *act8* family members, with TATA and 3' UAS motifs conserved across most of the family, while other motifs such as a G-box are found in only some promoters (Fig. S1A). While some promoters

---

Significance

Gene transcription occurs in discontinuous bursts. Although bursts are conserved in all forms of life, the causes and implications of bursting are not clear. Here we delineate a specific cause of bursts, and contextualize the significance of bursting, using analysis of a gene family encoding 17 identical actin proteins. Although the genes show similar developmental expression, which is coupled in single cells, they show strong differences in bursting dynamics. These distinct bursting patterns indicate that different signals regulate the individual genes, and imply expansion of the gene family allowed diversification of actin gene regulation. By exchanging the promoters of genes, we show that the dominant driver of bursting dynamics is the gene promoter, not the genome context.

Reserved for Publication Footnotes

137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
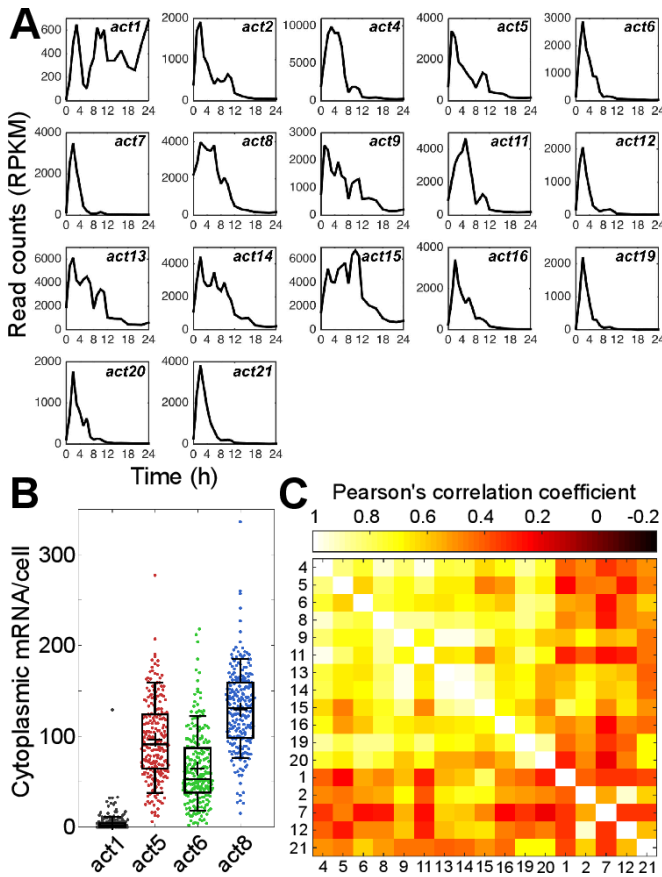195
196
197
198
199
200
201
202
203
204

205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272

Fig. 1. Developmental regulation and single cell coupling of actin gene expression (A) Uniquely mapped RNAseq read counts (reads per kilobase per million mapped reads) for all 17 *act8* genes over development. (B) Transcript counts of *act8* family genes in undifferentiated cells measured by smFISH. Each dot is a cell. Mean shown as cross in box plot. (C) High correlations between *act8* genes in single cells. Panel shows heatmap of all pairwise comparisons of *act8* family expression levels in undifferentiated cells from scRNAseq data.

Fig. 2. Distinct bursting patterns of actin genes. (A) Live imaging of transcription dynamics of three different actin genes. Intensity of transcription spots fluctuates over time to varying extents for each gene (time in min:sec). (B) Transcription spot intensity traces for populations of cells. Each panel represents an individual experiment for each actin gene. Each row represents the intensity trace for a single cell. Black indicates the cell was outside the field of view. Each block in a panel is a field of view. Imaging was carried out over 3 experimental days for each gene, with multiple fields of view per experiment: *act1* (152 cells), *act5* (476), *act6* (384), *act8* (275). (C) Probability density functions of spot intensity distributions for all experiments. Horizontal lines represent quartiles and median.

contain several elements, others such as *act8,* show little complexity, with large runs of A and T. Conserved elements were also found at the 3' end of *act* genes (Fig. S1A), which could enable further regulatory diversification, although earlier studies showed no strong differences in RNA turnover within the family (12).

To test whether actin genes are differentially regulated, we used RNA sequencing data to determine developmental profiles of gene expression (13). Fig. 1A shows the developmental expression patterns of all 17 *act8* genes. All genes are induced upon differentiation onset with a peak between 1-3h. Most genes show decreased expression during mid development (6-10 h) and by 16h, show little expression. Promoter differences may explain the subtle variations in developmental expression— genes with similar promoters, such as *act9*, *act13* and *act14*, show more similar expression during development. An exception to the general pattern is *act1,* which has additional peaks in expression later in development, although the gene shows comparatively few read counts. Despite minor differences, the remaining 16 genes display broadly similar expression, suggesting the generation of different expression profiles during development is unlikely to have been a major influence in the expansion of the family.

Single cell coupling of actin gene expression

To obtain single cell resolution and more accurate quantitation of actin gene expression, we measured the relative abundance of *act8* family transcripts using single molecule RNA FISH (smFISH). To sample genes with contrasting promoter architecture,
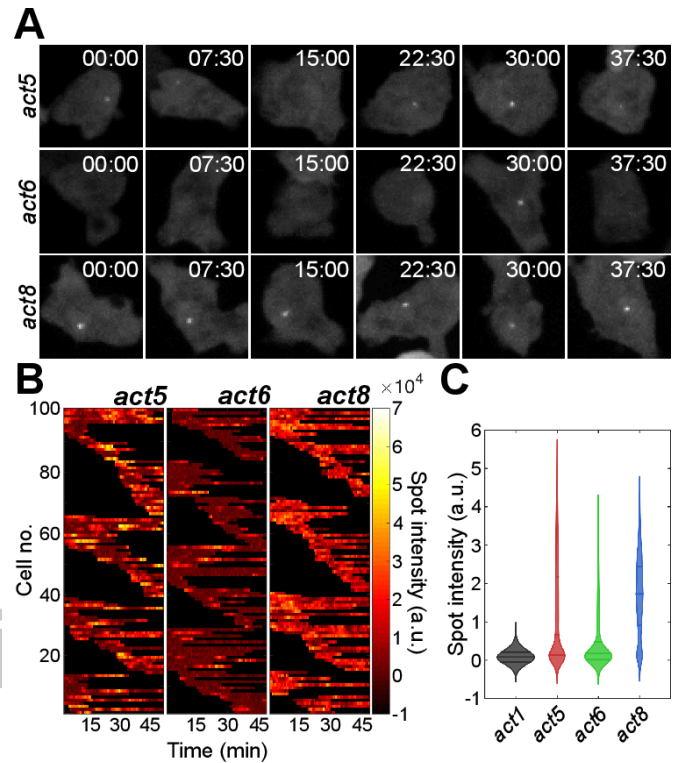
we chose *act1, act5, act6* and *act8*. 24 MS2 stem loops (14) were targeted at the 5' of coding sequences, causing the MS2 loops to be included in the transcribed RNA. Cells with MS2-tagged genes were probed with a fluorescent oligonucleotide complementary to the MS2 array. Cytoplasmic particles corresponding to single RNAs were counted (Fig. 1B). The *act5, act6* and *act8* genes were all strongly expressed, with *act8* the strongest and *act6* the weakest, but each gene showed 10s to 100s of RNAs per cell. In contrast, *act1* expression was close to background (cells without MS2), indicating that in undifferentiated cells, actin expression at full capacity is not required. In line with this, we found that disruption of up to 4 *act8* family genes had no effect on cell doubling times and a 6 gene mutant showed only a weak growth defect (Fig. S1B and C).

Each *act* gene showed considerable expression variability (Fig. 1B). Although variability is useful in cellular decision-making (15), it can also be disruptive, effectively making some cells overexpress the gene, while others are deficient. Noisy gene expression may create variability for one gene, but uncorrelated fluctuations in the 16 other genes could dilute this noise, allowing an optimal actin level for each cell. To test this reasoning, we used single-cell RNA sequencing (scRNAseq) data (16), to compare the expression of *act* genes in individual cells. Single cell read counts for all *act8* genes were clustered as heatmaps showing correlation values between pairs of genes (Fig. 1C). Some genes showed only weak correlations, consistent with the possibility of multiple family members diluting out stochastic variation. How-
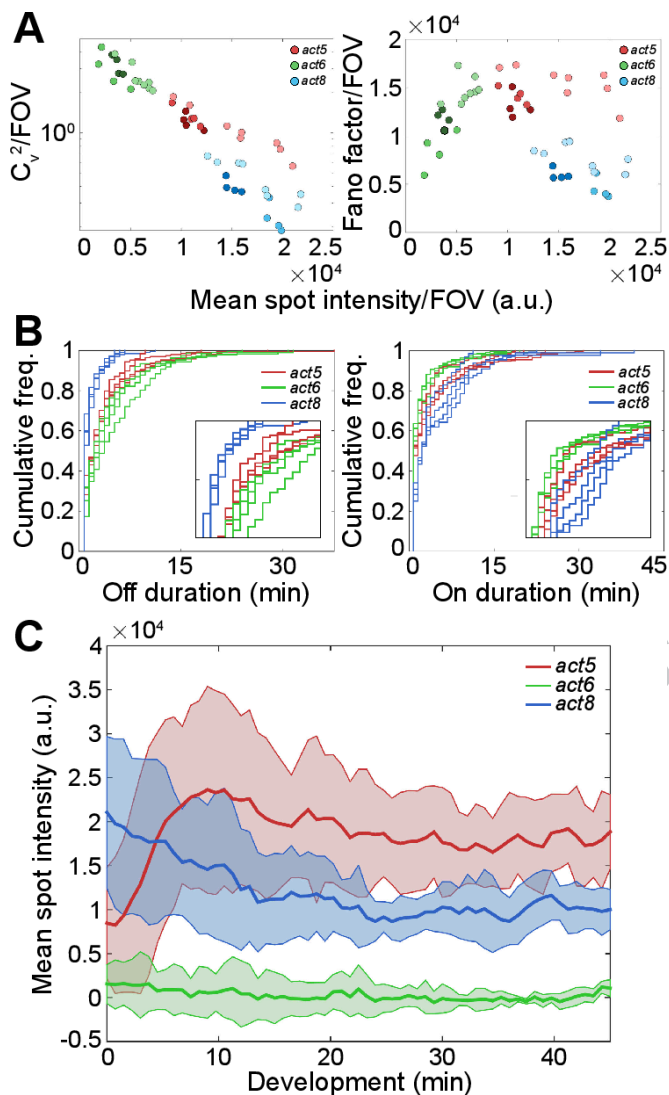
Fig. 3. Differential regulation of actin transcription bursts. (A) Correlations between spot intensity and different measures of variability (noise/ C_v^2 and noise strength/Fano). Each data point represents a field-of-view, with shades of each colour showing data from different experimental days. (B) Burst ON and OFF durations, with the threshold imposed at spot intensity of 5000 AU (See Fig. S2B). Data are shown as cumulative frequency plots. (C) Different responses of *act* genes to induction of development (starvation). Each line is the mean spot intensity per gene averaged across 4 independent datasets: *act5* (415 cells) *act6* (467),and *act8* (375). Shaded areas show SD.

ever, most genes were strongly correlated in their expression in undifferentiated single cells, and across multiple developmental time points (Fig. S1D). These data imply most *act8* family genes are co-ordinately regulated in single cells, with cell-cell variability (high versus low expression) a decision taken over most of the family. This suggests the expansion of the family was not selected primarily for buffering molecular noise in gene expression.

We infer no strong differences in *act* gene post-transcriptional regulation. Protein expression from *act* genes (monitored using mNeonGreen knock-ins) reflected the differences in transcript level seen by smFISH, with no differences observed in the variability of protein expression or localisation to actin structures (Fig. S1E).

Actin genes show different transcriptional bursting patterns

With only small differences in *act* gene developmental expression, despite divergent control sequences, we tested whether differential regulation is apparent over smaller timescales, using

MS2-tagged *act* genes. Upon transcription, nascent MS2-tagged RNA can be detected with a MCP-GFP fusion protein as a fluorescent spot at the transcription site. In live cells, the spot intensity varies over time (Fig. 2A) reflecting the fluctuating transcriptional activity of the gene (12).

Different actin genes exhibit very different transcriptional behaviours in undifferentiated cells (Fig. 2A). No *act1* spots were detected, indicating the gene is inactive or active below the detection threshold (5 RNAs (17)). Both *act5* and *act6* showed bursts of activity followed by periods of inactivity, with spot intensity greater for *act5* than *act6*. In contrast, *act8* activity fluctuations were small compared to the other genes. These example images were representative of hundreds of cells (Fig. 2B). *Act8* was strongly ON in most cells, most of the time, whereas act6 was mostly OFF, with the occasional moderate intensity burst. *Act5* was between these extremes, retaining frequent switching between ON and OFF. These dynamics are summarised in spot intensity distributions (Fig. 2C). The bursting of *act5* and *act6* is apparent in the thin upper tails of the distributions, and a large proportion of cells in the OFF state. In contrast, *act8* was mostly ON, with an approximate normal distribution of spot intensities in the active state. *Act5* and *act6* transcription showed significantly higher variance than *act8* (Fig. S2A).

To assess the type of control mechanisms regulating *act* genes, we determined how different measures of variability change with increasing gene activity. Increasing mean expression while decreasing noise (C_v^2, $\left(\frac{\sigma}{\mu}\right)^2$) indicates burst frequency control of gene expression (18, 19). In contrast, increasing noise *strength* (Fano factor, $\frac{\sigma^2}{\mu}$) with mean expression indicates burst size control. All three genes showed a strong negative correlation between spot intensity and noise (Spearman's rank correlation, *act5*: r = -0.92, p = 0, *act6*: r = -0.79, p = 5 x 10^{-4}, *act8*: r = -0.83, p = 1 x 10^{-4}) (Fig. 3A). In contrast, only *act6* intensity correlated with noise strength (*act5*: r = -0.13, p = 0.62; act6: r = 0.83, p = 1 x 10^{-4}; *act8*: r = -0.36, p = 0.17) suggesting *act6* modulates both burst frequency and size to change gene expression, while *act5* and *act8* modulate frequency alone.

Noise decreases with increasing gene expression (20). The difference in noise between *act5* and *act8* is unlikely to be caused by differences in expression level, as for a given spot intensity, *act5* variability was higher than *act8* (Fig. 3A). The noise-mean relationship of *act5* and *act6* can be explained by the same exponential function, as the data will lie on the same linear regression line (Fig. 3A), indicating the difference in variance between *act5* and *act6* arises because of reduced noise at higher expression.

We directly measured bursting dynamics from live cell data, by imposing a threshold separating ON and OFF phases, using *act1* traces to estimate the measurement noise (Fig. S2B,C). ON and OFF durations are represented as cumulative frequency plots (Fig. 3B), with *act8* showing short OFF and long ON phases, *act6* the opposite, with *act5* intermediate.

Signal regulation of bursting

What generates the different *act* gene bursting patterns? One view is that the likelihood of a burst relates to the concentration or activity of a transcription factor (TF), with the TF responding to signalling. Different bursting patterns would result from different genes responding to different signals. Alternatively, all genes respond to the same signals, with different TF binding properties generating different burst dynamics. The complete lack of homology between the *act8* and *act5* promoters argues against this second view. However, if the first model is valid, there should be stimuli with different effects on different *act* genes.

Cell size and cell speed both correlate with *act5, act6* and *act8* transcription, suggesting larger, faster cells are more likely to express actin (Fig. S3A,B). However, these data do not distinguish
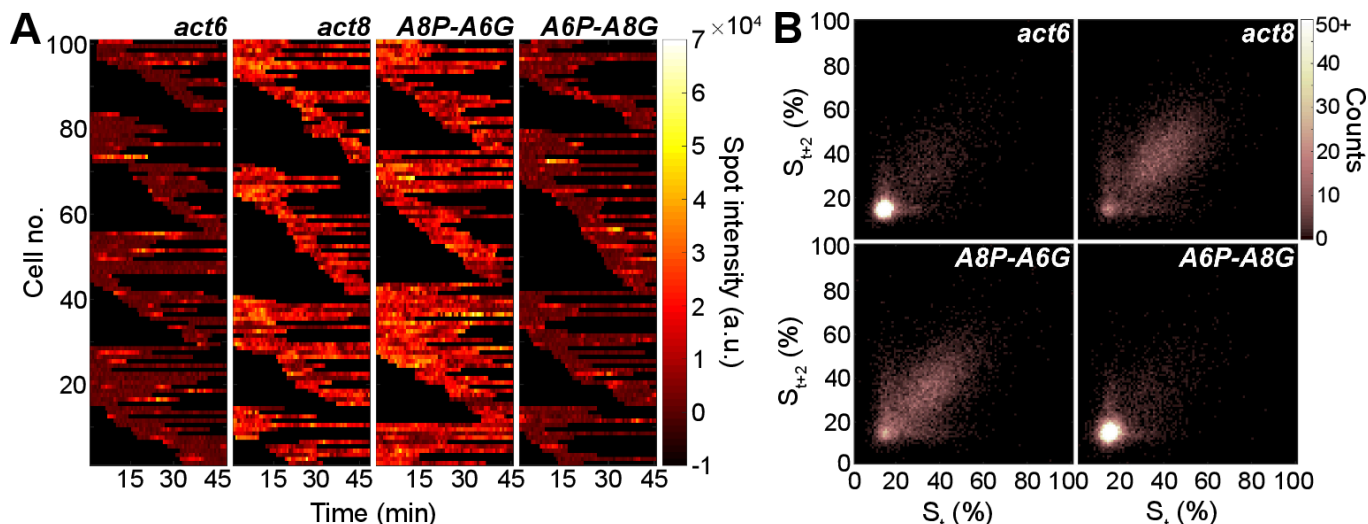
Fig. 4. Promoter regulation of actin transcription bursts. (A) Transcription dynamics of endogenous and promoter-switched genes. Each panel shows one experiment for *act6, act8* and the promoter switched genes *A6P-A8G and A8P-A6G*. Each row represents the spot intensity trace for a single cell. Promoter switched cell lines were imaged over 3 experimental days: *A6P-A8G* (714 cells) *and A8P-A6G* (760). (B) Co-occurrence matrices for endogenous and promoter-switched cell lines representing transitions between imaging frames. The xy coordinates determined by spot intensity at time t (S_t) and at time t+2 (S_{t+2}), expressed as a percentage of the maximum range of spot intensities across all experiments.
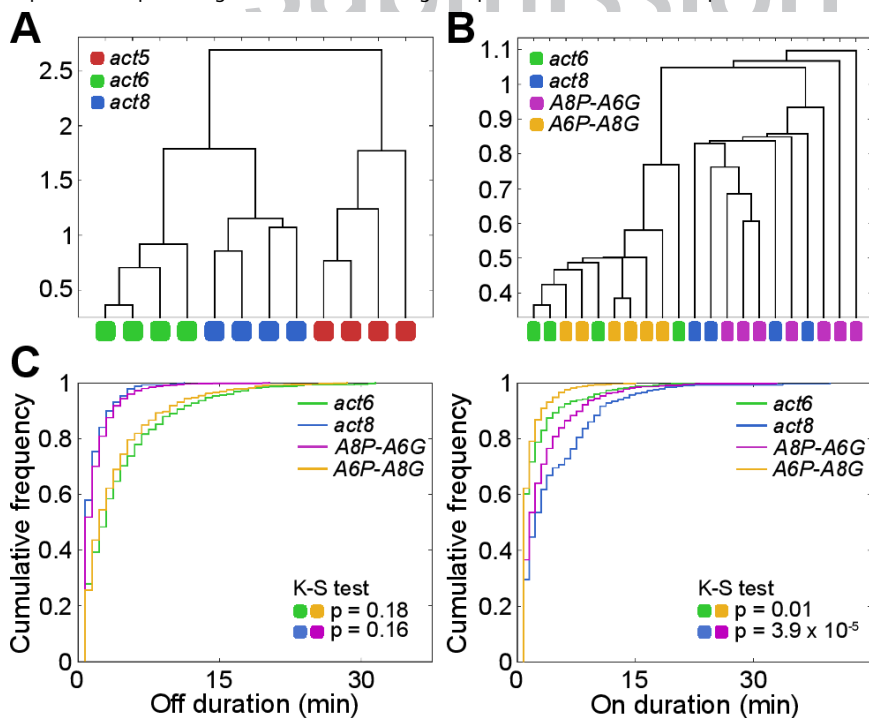

Fig. 5. Actin transcriptional bursting is predominantly promoter-driven (A) Classification of the training set data of *act5, act6* and *act8*. Dendrogram showing clustering by Euclidean distance using "bag-of-patterns". (B) Transcription dynamics of endogenous and promoter-switched genes were clustered using the same parameters as in A. (C) Cumulative distributions of burst ON and OFF durations for endogenous and promoter-switched genes.

between the genes. For undifferentiated cells, the major requirements for actin are in cytokinesis, phagocytosis and macropinocytosis. To identify signals that might differentially regulate actin bursts, we screened a panel of different culture environments to test how these and other cellular processes relate to actin protein expression (Table S2). Although subtle effects were observed, the relative ordering of expression level for *act5, act6* and *act8* did not change, so these different culture cues are unlikely to explain strong differences in bursting.

Differentiation in *Dictyostelium* is induced by starvation, followed by extracellular cAMP and other signals. The increase in *relative* mRNA for all *act8* genes during early differentiation (Fig. 1A) is not suggestive of differences in the regulation of

act transcription by these signals. However, global mRNA synthesis rates during differentiation are only ≈15% of rates for undifferentiated cells (21), so the relative actin increase might reflect continuing *act* transcription while much of the genome falls silent, not a sudden transcription surge. To directly visualise transcription during the actin mRNA spike, we imaged *act*-MS2 cells during early starvation.

Starvation triggered different responses for different *act* genes. At the onset of starvation, relative levels of transcription were similar to those of undifferentiated cells (Fig. 3C). However, *act5* showed a strong increase in the average transcription site intensity (Fig. 3C, S4A,B). In contrast, *act8* showed a decline in output, with *act6* transcription remaining low. The stronger

output of *act5* occurred with more cells showing longer, more frequent bursts (Fig. S4C). These data indicate different *act* genes respond differently to specific stimuli.

Evaluating the contributions of promoter and genomic context to bursting

What are the nuclear determinants of bursting dynamics? Many factors acting over a wide range of length scales can control transcription (22-25). Cis-acting elements, chromatin structure and modification, genomic context and nuclear organisation are all potential inputs. However, the ordering and magnitude of these inputs to transcription are unclear, due to extensive crosstalk between the multiple levels of regulation.

Given the diversity in promoter architecture and genome position across the *act8* family, we specifically tested the role of promoter sequences and genome context in regulating bursting dynamics. We exchanged the promoters of *act6* and *act8*, the genes with the most distinct bursting behaviours, by replacing the promoter of one gene with that of the other at the endogenous locus. The sequences exchanged were the proximal 388bp (*act8*) and 543bp (*act6*) of promoters. Unlike earlier promoter replacement studies (26), we used live cells to directly monitor transcription dynamics. The *act6* and *act8* genes are exactly the same length, are on different chromosomes, have no introns and can be deleted without phenotype (Fig. S1C). This reciprocal switch enabled us to determine the effects of promoter sequence on gene activity, independently of native genomic context.

Analysis of the transcription dynamics of these switched cell lines identified the promoter as the dominant factor in the regulation of transcription. Fig. 4A shows that the '*A8P-A6G*' cell line, with the *act8* promoter upstream of the *act6* gene, had remarkably similar transcription dynamics to endogenous *act8* – steady gene activity with low cell-to-cell variability. Similarly, the '*A6P-A8G*' gene, with the *act6* promoter upstream of *act8*, was mostly OFF with infrequent short bursts of activity, resembling normal *act6*. Spot intensity distributions were similar between endogenous and promoter-switched genes with the same promoter (Fig. S5A). There may be some residual control from the genomic context, apparent when comparing *act8* and *A8P-A6G*, where the spot intensity distribution of *A8P-A6G* is more skewed towards zero than the *act8* distribution and there is a small, but significant, difference in variance for these genes (Fig. S5B). However, at a coarse level, the promoter is more instructive for nascent transcript levels.

To determine the drivers of transcription dynamics, we used co-occurrence matrices to assess changes in gene activity over short time lags, with each point representing the spot intensity at time t and t+l, where l is a lag (2 frames). Matrices for *act6* and *act8* both show thin bands parallel to both axes representative of significant short-term changes in activity (Fig. 4B). Data points closer to the diagonal represent more slowly fluctuating gene activity, and most *act8* data is here. Plots for both genes show a clear separation between data where the gene slowly fluctuates, and switches rapidly. For the promoter-switched genes, the overall shape of the distributions indicate the promoter underpins most of the structure of the data, however, for *A8P-A6G*, the separation between slow and fast switching states is less clear, indicating *A8P-A6G* and *act8* dynamics are not identical.

To further evaluate the contributions of the promoter and genomic locus to bursting, we used an unbiased method of classification based on dynamic features of time series data. The method, "bag-of-patterns" (27), simplifies time-series into collections of discrete symbolic "words" representing the local structure of the data. The "letters" within words represent different bins of signal intensity, and the sequences of letters within words reflect the spot behaviour as the intensity fluctuates between bins. Individual time-series are defined by the relative word usage, tallied as histograms (Fig. S6C). Comparison of time-series involves calculation of Euclidean distances between word histograms, with final representation as a dendrogram. Three parameters, word length (w), number of bins (α) and duration of time-series sub-sequence (n) were optimised. We used *act5*, *act6* and *act8* data as training sets to identify parameter combinations capable of clustering the data according to gene identity (Fig. 5A and S6D).

To test the contributions of promoter and genomic context to bursting, we used the optimised parameters to classify *act6*, *act8*, *A8P-A6G* and *A6P-A8G* according to their transcription dynamics. Genes sharing a promoter rather than a genomic locus display more similar dynamic behaviours. *Act6* and *A6P-A8G* clustered more closely together compared to *act8* and *A8P-A6G* (Fig. 5B and S6E). Two *A8P-A6G* datasets could not be clustered, however all 4 that were clustered fit within the *act8* branch. Overall, the majority of the data clustered according to promoter identity, indicating the promoter is the dominant driver of bursting dynamics.

Comparisons between endogenous and promoter-switched genes showed that OFF periods of transcription were not different between pairs of genes sharing a promoter (KS test: *act6* v *A6P-A8G*, p = 0.18; *act8* v *A8P-A6G*, p = 0.16) (Fig. 5C, S7). However, *A6P-A8G* and *A8P-A6G* both spend significantly less time in the ON state than *act6* and *act8* respectively (KS test: *act6* v *A6P-A8G*, p = 0.01; *act8* v *A8P-A6G*, p = 3.9 x 10^{-5}). This suggests that while promoter sequence controls the majority of the dynamic behaviour, features specific to the genomic locus may influence burst duration.

Discussion

To gain insight into the processes driving expansion of an actin gene family in the absence of protein sequence diversification, we evaluated the contribution of differential gene expression. Our data reveal the expansion of the *Dictyostelium act8* gene family is not likely to have occurred to allow different temporal patterns of actin expression during development. Neither is expansion likely to have had a large contribution from a need to buffer noisy gene expression. Instead, we found the genes differ greatly in the dynamics of their transcriptional bursting. We interpret this effect as meaning that the expansion has occurred, at least in part, to enable actin expression to be regulated by an expanded set of signalling cues. In support of this idea, we show that *act8* family members show different bursting responses to starvation, a trigger for the onset of differentiation. Genes such as *act8* would act as the "workhorses" of the family, delivering much of the cellular actin, with genes such as *act5* showing thermostat-like control, to top up levels as appropriate. Expansion may represent a solution to diversifying regulation of a gene with a high transcript load, in a genome too compact for extensive upstream control by multiple enhancers. We then tested whether bursting is driven by the specific promoter sequence of each gene, or via other features of genomic context, by switching promoters of actin genes on different chromosomes. The dominant contribution to bursting dynamics came from the promoter.

Residual effects on bursting dynamics that might be attributed to genomic context were detected in an increased OFF rate observed when a promoter operates at a non-native site. This could conceivably be due to destabilisation of local chromatin conformation by insertion of a non-local DNA sequence. For example, looping between the 5' and 3' ends of a gene can facilitate transcription re-initiation which may contribute to the repetitive transcription events of a burst (28). Incompatibility between 5' and 3' ends could inhibit looping and increase the OFF rate. This incompatibility scenario assumes specificity provided by the 5' end of the gene- returning the emphasis on control back to the promoter. We therefore propose that the increased OFF rate of the switched genes represents an upper limit on the contribution to bursting dynamics from the genomic context.

Our data do not imply chromatin structure and nuclear organisation are unimportant for bursting. We propose that the effects of chromatin and nuclear structure on a gene are first instructed by DNA sequence. This view is consistent with many studies showing that chromatin modifications and nuclear organisation are imposed by transcription itself (29, 30). We emphasise that these dominant effects of the promoter have been observed for two genes in an apparently simple developmental eukaryote. This view may have to be modified when considering a metazoan genome with long-range transcriptional control. *Dictyostelium* has a small genome size with short intergenic regions— over 60% of the genome encodes protein. Yet the organism has DNA and H3K9 methylation, mediator, a nuclear lamin and late-replicating peripheral heterochromatin— standard features of metazoan chromatin modification, topology and organisation. Metazoan genes for which genomic context is perhaps most strongly implicated are the globin and Hox clusters (25, 31). Here the genomic context is presumably maintained under strong selective pressure to keep the family together for coordinate control. The idea that the *act8* family has undergone dispersal to sample the diversity of different genomic contexts is not strongly supported by our data.

Methods

Molecular biology and cell line generation

For live cell imaging of transcription, we targeted an MS2 cassette to the *act1*, *act6* and *act8* genes respectively. Targeting vectors were designed so that the MS2 sequence was at a similar position in the coding sequence of all genes, 18-24 bp downstream of the ATG. *Act5*-MS2 cells have been described previously (17). To switch promoters, the same promoter fragments used as targeting arms in the *act6* and *act8*-MS2 vectors were cloned next to the MS2 repeats in the targeting vector for the other gene. To ensure targeting to the correct locus, we cloned regions upstream of the promoters as the 5' homology arms of targeting vectors. In MS2 cell lines, the selectable marker was removed by transient CRE expression, to allow MS2 transcription to use endogenous terminators. For visualising nascent RNA, MS2-tagged cells were transfected with an extrachromosomal vector expressing the MCP-GFP fusion protein (17). To monitor protein levels we targeted codon-optimised mNeonGreen (32) to the 3' end of the endogenous genes in AX3 cells, followed by removal of the selectable marker. For transcription imaging, we used *Dictyostelium* Ax3 cells with a stably expressed red fluorescent nuclear marker, H2Bv3-mCherry (17).

Imaging and data analysis

Cells were imaged in a low fluorescent medium in 8-well chambers (Nunc Lab-Tek II) and imaged on an UltraVIEW VoX spinning disc confocal microscope (Perkin Elmer) with an EM-CCD camera (C9100-13, Hamamatsu) using a 60x 1.4 NA objective. For imaging during development, cells were washed free of media and viewed under non-nutrient agar (12). Cell tracking, spot identification and motility/size analysis used custom-built Matlab software (17). For smFISH, we used the protocol of (33). We used a single probe, end-labelled with Quasar 670, which binds the spacer between each MS2 stem-loop (17). Cells were imaged using the UltraVIEW with a 100x objective and 640 nm laser. Cytoplasmic mRNA counts were determined using FISH-quant (34).

To implement bag-of-patterns for our transcription dynamics data we concatenated individual cell tracks to generate time-series of around 1000-5000 frames for each cell line per imaging session. Concatenated tracks were log-transformed and normalised to give mean = 0 and SD = 1 (Fig. S6 A,B), for comparison of time series with different offsets and amplitudes. A MATLAB package was used to derive word histograms (https://cs.gmu.edu/~jessica/sax.htm) before collating the bag-of-patterns for individual time-series (27). Parameter sets were defined empirically using *act5*, *act6* and *act8* data as a training set (Fig. S6C) before applying these to promoter-switched datasets.

1. Zhang J (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18(6):292-298.
2. Marzluff WF, Gongidi P, Woods KR, Jin J, & Maltais LJ (2002) The Human and Mouse Replication-Dependent Histone Genes. *Genomics* 80(5):487-498.
3. Eickbush TH & Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175(2):477-485.
4. Holmes WF, *et al.* (2005) Coordinate control and selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. *The Journal of biological chemistry* 280(45):37400-37407.
5. Joseph JM, *et al.* (2008) The actinome of Dictyostelium discoideum in comparison to actins and actin-related proteins from other organisms. *PloS one* 3(7):e2654.
6. Vandekerckhove J & Weber K (1980) Vegetative Dictyostelium cells containing 17 actin genes express a single major actin. *Nature* 284(5755):475-477.
7. Parikh A, *et al.* (2010) Conserved developmental transcriptomes in evolutionarily divergent species. *Genome biology* 11(3):R35.
8. Yates LD & Greaser ML (1983) Quantitative determination of myosin and actin in rabbit skeletal muscle. *Journal of molecular biology* 168(1):123-141.
9. Uyemura DG, Brown SS, & Spudich JA (1978) Biochemical and structural characterization of actin from Dictyostelium discoideum. *The Journal of biological chemistry* 253(24):9088-9096.
10. Romans P, Firtel RA, & Saxe CL, 3rd (1985) Gene-specific expression of the actin multigene family of Dictyostelium discoideum. *Journal of molecular biology* 186(2):337-355.
11. Hori R & Firtel RA (1994) Identification and characterization of multiple A/T-rich cis-acting elements that control expression from Dictyostelium actin promoters: the Dictyostelium actin upstream activating sequence confers growth phase expression and has enhancer-like properties. *Nucleic acids research* 22(23):5099-5111.
12. Muramoto T, *et al.* (2012) Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proceedings of the National Academy of Sciences of the United States of America* 109(19):7350-7355.
13. Rosengarten RD, *et al.* (2015) Leaps and lulls in the developmental transcriptome of Dictyostelium discoideum. *BMC genomics* 16:294.
14. Bertrand E, *et al.* (1998) Localization of ASH1 mRNA particles in living yeast. *Molecular cell* 2(4):437-445.
15. Symmons O & Raj A (2016) What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Molecular cell* 62(5):788-802.
16. Antolović V, Miermont A, Corrigan AM, & Chubb JR (2017) Generation of Single-Cell Transcript Variability by Repression. *Current biology* 27(12):1811-1817.e1813.
17. Corrigan AM, Tunnacliffe E, Cannon D, & Chubb JR (2016) A continuum model of transcriptional bursting. *eLife* 5.
18. Carey LB, van Dijk D, Sloot PMA, Kaandorp JA, & Segal E (2013) Promoter sequence determines the relationship between expression level and noise. *PLoS biology* 11(4):e1001528.
19. To T-L & Maheshri N (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* 327(5969):1142-1145.
20. Bar-Even A, *et al.* (2006) Noise in protein expression scales with natural protein abundance.

Nat Genet 38(6):636-643.
21. Mangiarotti G, Altruda F, & Lodish HF (1981) Rates of synthesis and degradation of ribosomal ribonucleic acid during differentiation of Dictyostelium discoideum. *Molecular and cellular biology* 1(1):35-42.
22. Lenstra TL, Rodriguez J, Chen H, & Larson DR (2016) Transcription Dynamics in Living Cells. *Annual review of biophysics* 45:25-47.
23. Fukaya T, Lim B, & Levine M (2016) Enhancer Control of Transcriptional Bursting. *Cell* 166(2):358-368.
24. Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91(3):243-248.
25. Bartman CR, Hsu SC, Hsiung CC, Raj A, & Blobel GA (2016) Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular cell* 62(2):237-247.
26. Hocine S, Vera M, Zenklusen D, & Singer RH (2015) Promoter-Autonomous Functioning in a Controlled Environment using Single Molecule FISH. *Scientific reports* 5:9934.
27. Lin J, Khade R, & Li Y (2012) Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems* 39(2):287-315.
28. Hebenstreit D (2013) Are gene loops the cause of transcriptional noise? *Trends in genetics : TIG* 29(6):333-338.
29. Soares LM, *et al.* (2017) Determinants of Histone H3K4 Methylation Patterns. *Molecular cell* 68(4):773-785 e776.
30. Nozawa RS, *et al.* (2017) SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* 169(7):1214-1227 e1218.
31. Rodriguez-Carballo E, *et al.* (2017) The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes & development* 31(22):2264-2281.
32. Shaner NC, *et al.* (2013) A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nature methods* 10(5):407-409.
33. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, & Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* 5(10):877-879.
34. Mueller F, *et al.* (2013) FISH-quant: automatic counting of transcripts in 3D FISH images. *Nature methods* 10(4):277-278.