Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation

Tetsuya Muramoto a,1, Danielle Cannon a,2, Marek Gierliński b, c, Adam Corrigan a,2, Geoffrey J. Barton b, c, and Jonathan R. Chubb a,2,3

*Division of Cell and Developmental Biology, †Division of Biological Chemistry and Drug Discovery, and ‡Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

Edited by Sanjay Tyagi, University of Medicine and Dentistry of New Jersey, Newark, NJ, and accepted by the Editorial Board March 21, 2012 (received for review October 25, 2011)

Transcription of genes can be discontinuous, occurring in pulses or bursts. It is not clear how properties of transcriptional pulses vary between different genes. We compared the pulsing of five housekeeping and five developmentally induced genes by direct imaging of single gene transcriptional events in individual living Dictyostelium cells. Each gene displayed its own transcriptional signature, differing in probability of firing and pulse duration, frequency, and intensity. In contrast to the prevailing view from both prokaryotes and eukaryotes that transcription displays binary behavior, strongly expressed housekeeping genes altered the magnitude of their transcriptional pulses during development. These nonbinary “tunable” responses may be better suited than stochastic switch behavior for housekeeping functions. Analysis of RNA synthesis kinetics using fluorescence recovery after photobleaching implied modulation of housekeeping-gene pulse strength occurs at the level of transcription initiation rather than elongation. In addition, disparities between single cell and population measures of transcript production suggested differences in RNA stability between gene classes. Analysis of stability using RNaseq revealed no major global differences in stability between developmental and housekeeping transcripts, although strongly induced RNAs showed unusually rapid decay, indicating tight regulation of expression.

transcriptional bursting | RNA turnover | stochastic gene expression

Transcription is not adequately described by the smooth, seamless process implied by standard measures of RNA level. Within individual cells, transcription occurs as a series of irregular pulses, interspersed by long, irregular periods of inactivity. Pulsing (or bursting) is a fundamental feature of transcription, conserved from prokaryotes to mammalian cells (1–5). These phenomena have strong implications for our understanding of transcriptional mechanism and may provide a major source of stochasticity in gene expression (6), a driver of cell diversity in differentiation and disease (7, 8). However, it is unclear how pulsing behaves for different genes with different functional properties.

Pulsing dynamics, and therefore deeper understanding of underlying transcriptional mechanics and regulation of different genes, are masked when averaged over millions of dead cells, as occurs with standard bulk RNA measurement techniques, from Northern blotting to RNA sequencing (RNaseq). The readout from these methods also has a variable contribution from RNA stability. Although strong inferences can be made from heterogeneities in transcript number using hybridization against RNA in single cells (RNAFISH) (9–11), an erroneous inference of strong transcription from both bulk and fixed-cell RNA techniques emerges if RNA is stable. To appreciate how transcription is regulated and how the process differs between different genes, it is crucial to look at the process itself in living cells at the single-gene level. Live-cell methods using fluorescent proteins or luciferase have been very useful in giving us a sense of the instability of transcriptional states, with expression heterogeneity between cells and slow fluctuations (hours to days) between different states (12–14). These methods also reveal that different genes have different fluctuation kinetics. However, available data on transcription bursts and pulses imply active transcriptional states last more in the range of minutes, even for strongly transcribed genes (1). To describe the dynamics of transcription pulses it is therefore necessary to directly observe pulses of RNA production, which requires using high-affinity RNA–protein interactions to deliver fluorescent signals to nascent RNA (15, 16). The resultant accumulation of fluorescence at the site of transcription is viewed under a microscope as a fluorescent spot, which appears and disappears (pulses) at irregular intervals when imaged in living cells.

How is pulsing different for genes with different functions and expression requirements? How is pulsing reactive to cell context for different genes? To address these questions, we compared pulsing of a set of housekeeping and developmentally induced genes in Dictyostelium cells. Dictyostelium are social amoebae, existing as single feeding cells that, upon starvation, initiate a developmental program resulting in chemotactic cell aggregation, followed by differentiation and remodeling of the aggregate into a spore-containing mass suspended above the substrate by a stalk. As with many differentiation steps in disease and development, from prokaryotes to stem cells, initial differentiation is scattered or stochastic, not determined by cell position (7, 17).

We measured pulsing of five developmental and five housekeeping genes at different stages during preaggregative development. Each gene showed its own pulsing properties, measured directly using a variety of parameters. Housekeeping genes strongly modulate pulsing strength during development, allowing greater tunability of transcription at the single-cell level. In contrast, most developmental genes showed binary behavior. Differences in pulsing at different developmental time points are controlled at the level of transcription initiation. Finally, we demonstrate strongly induced developmental transcripts tend to show higher turnover than other transcripts, indicative of tight control of expression.

Results and Discussion

To visualize transcriptional dynamics of single housekeeping and developmentally induced genes, we used MS2 stem loops for nascent RNA detection (15). A 1.3-kb array of 24 MS2 loops was integrated into 5 coding sequences of genes in Dictyostelium cells (Fig. L4). Selection of recombinants used blasticidin resistance

Author contributions: T.M. and J.R.C. designed research; T.M. and D.C. performed research; T.M., M.G., G.J.B., and J.R.C. contributed new reagents/analytic tools; T.M., D.C., M.G., A.C., and J.R.C. analyzed data; and T.M., D.C., M.G., and J.R.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.T. is a guest editor invited by the Editorial Board.

Present address: Quantitative Biology Center, RIKEN, Osaka 565-0874, Japan.

Present address: Department of Cell and Developmental Biology and Medical Research Council Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, United Kingdom.

To whom correspondence should be addressed. E-mail: j.chubb@ucl.ac.uk.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.1117603109/DS1Supplemental.

7350–7355 | PNAS | May 8, 2012 | vol. 109 | no. 19

www.pnas.org/cgi/doi/10.1073/pnas.1117603109
cassettes (bsr) downstream of MS2 repeats. To ensure similar transcript length between genes, termination of transcription after MS2 repeats used the bsr terminator, which directly follows the repeats. The RNA contained only the 1.3-kb MS2 array, plus standard short 3′ and short variable 5′ extensions. We studied transcription of five housekeeping genes [act5 (β-actin), cinD (putative transcription factor), rpl15 (ribosome protein), and scd (fatty acid desaturase)] and five development genes [carA (chemoattractant receptor), csA4 (cell adhesion), cofC (actin-severing), hspF (heat-shock), and zfaA (early development marker)]. Developmental genes show strong increases in mRNA abundance during the first 5 h of development. Housekeeping genes were selected for strong expression in undifferentiated (vegetative) cells, although transcripts can be detected to varying degrees during development. *Dictyostelium* is haploid, so to minimize effects of targeting we selected, where possible, genes from a duplicated region of chromosome 2 in the parental strain. Act5, zfaA, and csA4 genes are not duplicated, but act5 is one of 17 actin genes encoding identical proteins, csA4 mutants have a minimal developmental phenotype (18), and zfaA-MS2 cells show no obvious defects. Expression of MS2 RNA was similar to endogenous RNA in wild-type cells during early development (Fig. S1). Small differences observed likely relate to differences in turnover between MS2 and parental RNAs (see below). The MS2-GFP expression vector was transformed into recombinants to obtain relatively uniform MS2-GFP expressing cell lines, with comparable intensities for each gene (Fig. S2A).

Nascent transcripts were detected in living cells for all 10 genes, and spots appeared and disappeared in pulses (Fig. 1B). It was immediately clear that pulsing behavior, from spot intensity to the proportion of cells with pulses, was highly variable dependent on gene and development stage. Blind-selected fields were captured as 3D stacks every 2.5 min, for 30 min for undifferentiated (0 h), starvation (3 h), and preaggregation (5 h) cells, with around 1,000 cells studied per time point per gene for all 10 selected genes over 3–4 experimental days (Fig. S2B).

**Pulsing Dynamics of Developmental and Housekeeping Genes.** During development, each gene showed characteristic pulsing behavior. In line with bulk expression data (19), housekeeping genes showed a higher proportion of cells with RNA spots at 0 h (Fig. 2A), from 90% of cells for act5 and scd, to 15% for abpE and rpl15. The reverse was true for development genes, with 3% of cells with carA spots at 0 h increasing to 80% at 5 h of development (Fig. 2A). Weaker expressed development genes, cofC and zfaA, barely reached 1% of cells showing spots, although to preserve viability we did not image with single-molecule sensitivity, so would not detect weak transcription. Under no conditions did any gene show spots in 100% of cells. Four of five development genes showed spots in a small percentage (<5%) of undifferentiated cells. Do these rare events imply all cells expressed, masked by use of a limited capture period? Or do they reflect distinct subpopulations of cells?

Both possibilities were identified. MS2 repeats were replaced in targeting vectors with a GFP gene, which allowed GFP expression...
from the same native promoters (Fig. 2B). GFP half-life is several hours in Dictyostelium (20), allowing a readout of accumulated transcription. GFP expressed from housekeeping promoter, act5, showed strong, relatively even expression, implying over timescales of several hours that all cells transcribe act5. The same finding was true for carA. Compared with autofluorescence levels of parental cells, carA-GFP cells display even GFP, with fewer than 1/1000 cells expressing at levels two- to threefold greater than the mean. This finding implies the 3% of 0 h cells showing carA-MS2 spots are not partially differentiated “founders,” but that all cells have low-level transcription. CarA encodes the aggregation stage cAMP receptor. It might be advantageous for undifferentiated cells to perceive cAMP, a differentiation signal and chemoattractant, to ensure against changing nutritional conditions. In contrast, GFP expressed from the hspF promoter showed variegated expression, with most cells showing only autofluorescence, and around 1% of cells with strong GFP expression (Fig. 2B). Heterogeneity for hspF may relate to it being a stress-response gene. These genes display high expression noise in yeast, perhaps a hedging strategy against environmental fluctuations (21, 22).

Dynamics of nascent RNA production were assessed within individual cells displaying spots. We describe dynamics in terms of pulse duration, frequency, and intensity (Fig. 3A). Pulse duration was highly heterogeneous between pulses, cells, and genes. Shorter pulses (<10 min) were most common for all genes, but some pulses began before and ended after capture (>30 min). Pulse-duration distributions approximate to exponential (Fig. S3A). Deviation from exponential appeared greater for strong pulsers (act5, scd, carA, and csaA). To measure durations we included incomplete pulses, to avoid the capture-window skewing data for strong transcribers. With pulse durations from complete pulsations, distributions for these genes tended more to exponential (Fig. S3B) and goodness-of-fit tests find exponential behavior rejected at the 1% level in only one case (csdA 5 h). Behavior of csaA might be explained by responsiveness of the gene to extracellular cAMP, which oscillates during development. Periodic stimulation might override the more irregular transitions of bursting. Intervals between pulses were also distributed exponentially for development and housekeeping genes (Fig. S4A).

The effect of development on pulse duration showed distinct behaviors for developmental and housekeeping genes (Fig. 3B and C). For development genes, pulse duration did not alter greatly during development. Durations were similar at 5 h, when the genes were expressed most strongly, and 0 h, when few cells displayed spots. This finding implies the transcriptional response for developmental genes is binary or all-or-nothing, with parameters of pulses remaining relatively constant but the number of pulsing cells changing. This finding has been observed at the protein level in viral (23), prokaryotic (24), and synthetic (25, 26) gene-induction contexts, and is apparent for pulsing of two additional developmental genes in Dictyostelium (1, 27). In contrast, pulse durations for two strongly expressed housekeeping genes, act5 and scd, diminished strongly during development, with pulses two- to threefold shorter for these genes at 5 h than at 0 h. This finding indicates an alternative, nonbinary response, where individual cells tune the level of transcript produced per transcriptional event during differentiation.

Measurements of other pulse parameters showed similar binary and nonbinary behaviors for development and housekeeping genes. Pulse frequency and intensity showed little or no change for most developmental genes during differentiation (Fig. 3). Pulse frequency for carA increased (Fig. 3E), evidence of nonbinary behavior, offset by decreasing pulse intensity between 3 and 5 h (Fig. 3G). In contrast, all five housekeeping genes showed decreasing pulse intensity (Fig. 3F) and three housekeeping genes showed clear decreases in pulse frequency during development (Fig. 3D), most apparent for the two strongest expressers, act5 and scd. Summing intensity values for each time point within a pulse gave integrated pulse strengths (Fig. 3 H and I). For act5, this parameter decreased 8.1-fold during development. Decreases
in pulse strength were also marked for scd and cinD, although occurred for all housekeeping genes. For developmental genes (Fig. 3f), carA showed an increase in integrated intensity of 2.5-fold, dominated by a small outlier group of very strong long pulses (Fig S5A, 4 of 100). GFP expression in carA-GFP knock-in cells at 5 h of development was, as at 0 h, relatively homogeneous, implying the outliers are not a distinct population of carA expression. An alternative explanation is the outliers reflect poor clearance of transcript, or a “traffic jam” behind stalled polymerase (28).

**Polymase Kinetics at Transcription Sites.** How does the changing spot behavior during development relate to polymerase dynamics at transcription sites? To address this question, we performed fluorescence recovery after photobleaching (FRAP) on spots in undifferentiated (0 h) and 4-h developed act5-MS2 cells. Transcription sites were bleached with localized laser pulses, and 3D images captured every 5 s after bleaching (Fig. 4A). Turnover of MS2 protein on stem loops is negligible (29), so spot recovery reflects new transcription. Postbleach nuclear background intensities for both 0- and 4-h cells were 85% of original intensity (Fig. S5B). The small fraction of bleached MS2-GFP will affect spot intensity, but bleaching is the same for different time points, so we can compare recovery rates and relate these to estimates of transcription rates.

Of 46 spots in 0-h cells, 87% of spots recovered within 120 s postbleach. High proportions of recovery imply spots are highly dynamic, not dominated by long-term stalled polymerase (28). For 85 4-h cells, 52% of spots recovered. These recovery data are in line with nonlinear pulse behavior for housekeeping genes. Pulses are shorter and less intense during differentiation, so less likely to show detectable fresh transcription postbleach.

Normalized spot recovery rates were similar in undifferentiated (0 h) and differentiated cells (4 h) (Fig. 4B). Because both 0- and 4-h spots take the same time to repopulate with nascent RNA, this indicates the time for a polymerase to traverse the gene (the elongation rate) is similar. Raw spot intensities were greater at 0 h (Fig. 4C) and initial and final spot intensities were correlated (Fig. S5C). This finding led to raw final intensities and recovery rates being greater at 0 h (Fig. S5D). However, with recovery data normalized by prebleach intensity (29) (Fig. 4B), trajectories in 0- and 4-h cells were very similar. These data imply the differences in raw intensities between developmental time points (nonbinary behavior) are simply because of differences in the loading of polymerase on the gene (the initiation rate).

With an awareness that trains of polymerases only move as fast as the slowest, we approximated a lower-bound transcription rate for a native single locus of 1.3 kb·min⁻¹. The plateau for 0-h cells is less distinct, but similar trajectories imply similar rates. This transcription rate is within the 1- to 6-kb range of estimates made using bulk approaches or multicopy array FRAP (30), two- to fourfold higher than estimates from single alleles at heterologous loci in human HEK-293 cells (31), and similar to estimates from fluorescence fluctuation measurements at a single yeast gene (16). Transcription from HIV promoters can be 50–100 kb·min⁻¹ (29).

The binary tendency for developmental genes might result from polymerase preloading, so even in undifferentiated cells, strong transcription occurs because polymerase is primed. To test this possibility, we performed ChIP using antibodies against RNA Polymerase II (Pol II). Pol II initiation is marked by phosphorylation of serine 5 (S5) on its carboxy terminal domain. S5 also marks poised or primed polymerase. Pol II then enters progressive transcript elongation characterized by serine 2 phosphorylation (S2) (32). ChIP using modification-specific antibodies revealed S5 and S2 were high in undifferentiated cells for housekeeping genes, declining during development (Fig. S5E). In contrast, development genes showed increases in S2 and S5 during differentiation. Increases were weak for hspF, perhaps evidence of a preloaded state in undifferentiated cells. This finding was not apparent for carA and csdA, implying no population-wide Pol II preloading.

**Comparing Single-Cell Dynamics and Steady-State RNA Level.** How do single-cell nascent RNA dynamics relate to RNA levels assessed by traditional measures? To address this question, we compared pulse properties to RNA levels quantified from Northern blots of developmental time courses. Data comparing spot intensity and pulse duration are plotted against relative RNA level in Fig. 5. RNA levels showed correlations with both spot intensity and pulse duration (Fig. 5B). Fig. 5A and B show housekeeping RNAs were more abundant than developmental RNAs for a given level of nascent RNA. This may be, in part, because housekeeping RNAs are already present at the beginning of development. There will be a delay before developmental RNAs accumulate. These data may also be explained by housekeeping RNAs being more stable than developmentally induced RNAs.

We tested this possibility by treating 5-h developed MS2 (Fig. S6A) and parental (Fig. S6B) cell lines with 125 μg·mL⁻¹ actinomycin D (actD), to block new transcription (33). After treatment, spots disappeared for act5 after 10 min, and scd, carA, and csdA after 20 min. After actD, MS2 RNAs from housekeeping genes act5, cinD, and scd were degraded gradually, with mean lifetimes of 57, 46, and 36 min, respectively (Fig. S6C). MS2 RNAs from developmental genes, carA, csdA, and hspF were degraded more quickly (lifetimes 18, 25, and 25 min). Similar trends were seen for untagged genes in 5-h developed parental cells (Fig. S6B), where overall, housekeeping mRNAs were more stable than developmentally induced mRNAs, apart from scd (housekeeping) and dscA (development). Comparing data on

---

**Fig. 4.** Dynamics of RNA turnover at a single native gene. (A) Recovery of fluorescence at act5 transcription site after photobleaching. Undifferentiated cell, 2D maximum projection of 3D stack. The spot became detectable 30 s postbleach. Box in first image shows bleached region. Timing (minutes:second). (B) Trajectories of FRAP curves from normalized data from undifferentiated (0 h, n = 20) and 4-h developed (n = 14) act5 cells. Data were normalized to prebleach spot intensities. (C) Raw FRAP trajectories for act5 transcription spots in 0- and 4-h cells. Bars show SEM. Field of view for each image is a 17.5 × 17.5 μm.
tagged and untagged alleles indicated MS2-tagged RNAs were less stable than parental RNAs (lifetimes two- to threefold lower). However, with the exception of scd, relative stability of transcripts was retained after tagging (Fig. S6 A–C). As transcripts from tagged genes only differ by their 5′ ends (terminators are identical), these data imply 5′ sequences carry strong determinants of RNA stability. This finding may parallel recent yeast studies revealing regulation of RNA stability at the promoter (34, 35). Overall, this small-scale analysis provides evidence that RNAs from developmentally induced genes tend to be less stable than housekeeping RNAs.

To address whether this tendency is general, we quantified RNA decay using RNAseq, to measure RNA stability at the genome level. Five-hour–developed cells were treated with actD. RNA was collected from the starting population (0 h) the 5-h population and 30 and 60 min after actD treatment. RNAseq was carried out on three biological replicates for each time point. At least two-million reads per time point were obtained (Dataset S1). Similarity between replicates was high (Pearson’s correlation coefficients $r > 0.94$). Eleven RNAs revealed by RNAseq to be stable or unstable were tested by blotting of different extracts (Fig. S6D). Degradation measured by Northern and RNAseq showed a clear correlation ($R^2 = 0.75$) (Fig. S6E).

We defined induction, $R_g$, of a gene as the ratio of intensity (RNA abundance) at 5 h to that at 0 h. We defined degradation, $D_{ref}$, of a gene as the ratio of intensity at 5 h and 5 h + 60 min. The relationship between induction and degradation is shown in Fig. 5C. There is a weak ($r = 0.13$), but significant ($P = 10^{-10}$) correlation between $R_g$ and $D_{ref}$. The basis of this trend appears to be high levels of degradation for very highly induced genes ($R_g > 10$). To assess the significance of this trend, we divided data into bins of 30 datapoints along the horizontal axis. Individually, only the last bin ($R_g > 50$) showed significant increase above $D_{ref}$, the mean degradation of genes with low induction ($R_g < 10$). However, the rising trend starts at lower $R_g$ and the combined significance of the last four bins ($R_g > 9$) is $P < 10^{-10}$. In summary, these data show a subset of highly induced genes are, on average, less stable than genes with low levels of developmental induction.

The dispersion in degradation is large and there are points with high induction and low degradation. In addition, the effect is somewhat dependent on selection of “good” datapoints, where all replicates show a level of expression significantly different from zero ($α = 0.05$). Choosing $α = 0.01$ for accepting points leaves only 463 “good” genes and too few high-induction data-points to assess the increase in degradation. On the other hand, with $α = 0.1$, “good” data are noisier, although the trend combined for $R_g > 10$ is still significant ($P < 10^{-4}$). Among these RNA stability determinants supposed to be at 5′ ends of genes for several loci described here, we were unable to find specific motifs related to general mRNA stability or instability from whole-genome data. There was a clear correlation between transcript length and degradation (Fig. S7A) but not between GC content and degradation (Fig. S7B). We extracted stability data for 28 “core” cAMP pathway genes and plotted this against induction and expression level. Degradation was not correlated to fold-induction (Fig. S7C) or expression level (Fig. S7D) for these genes. The act8 family is a group of 17 genes (including act5) encoding identical β-actin proteins (36). Why do cells need so many genes encoding identical proteins? Perhaps 5′ and 3′ UTRs allow specialization, for example by regulating RNA stability. Fig. S7E shows a plot of induction vs. degradation for the 17 mRNAs. None show extremes of degradation or induction, although two outliers are apparent--act1, the most stable and highly induced and act7, the least stable and most weakly induced. 3′ UTRs of act mRNAs show distinct motifs. Most of the family have a motif with a core GATGAAAG 25 nt downstream of the TAA. Both act1 and act7 lack this motif. A second motif, GTTGGTGCAC, is conserved 140 nt downstream of TAA, although not in act7. Both act1 and act7 are the least-expressed members of the family (Fig. S7F), so these may be genes no longer under selection, or used at a specific developmental state.

Concluding Remarks. Live-cell analysis of nascent RNA dynamics has revealed differences in pulsing behavior between different genes. Each gene has characteristic pulsing kinetics, but distinctions between housekeeping and developmentally induced genes were identified. In keeping with the single-cell literature for prokaryote, eukaryote, viral, and synthetic systems, developmental gene pulsing was observed to be predominantly binary. Individual cells have a varying probability of firing, but properties of pulses within transcribing cells largely fall within a standard range, despite external conditions. In contrast, housekeeping genes showed a strong capacity for intracellular modulation of transcription, with shorter, weaker pulses occurring as cells enter differentiation. Binary responses might be explained by a high cooperativity and positive feedback in upstream signaling, or an obstacle creating an “off” state, such as a recalcitrant nucleosome. Noisy binary behaviors necessitate noise-reduction strategies to allow precision in cell response, or alternative programs to compensate for heterogeneity, which may be a driver for the evolution of so-called “redundancy” if ordered cell behavior is
a requirement. That individual cells can tune the expression level of some genes more precisely implies a modicum of control in an otherwise noisy process. Stochastic switch-like behavior is well suited to cell choices in development or stress responses (37); it may be less suitable for housekeeping functions, where measured responses, rather than probabilistic on-or-off decisions, seem more appropriate. If the absence or presence of a conserved riboswitch were effectively a coin-toss decision, this would not be in the interests of the cell. It will be interesting to address how modulation of housekeeping transcription relates to the derived nonbinary behavior in synthetic systems (38, 39).

For a subset of induced genes, tight control of RNA level was also apparent in our RNA stability analysis. Whereas the bulk of RNAs are largely unstable or stable irrespective of their degree of induction during development, a subset of strongly induced developmental RNAs showed unusually high turnover. Instability allows flexibility, permitting more rapid transition to the next developmental stage, or bet-hedging for rapid access back to vegetative growth.

Materials and Methods

Cell Lines and Imaging. Detailed information can be found in SI Materials and Methods. MS2 repeats were targeted into S. coding regions of 10 Dictyostelium genes. Targeting vectors were transformed into D. AX3 cells and recombiantants selected with 10 μg mL⁻¹ blasticidin S. Correct knock-in clones were identified by Southern blotting then transformed with MS2-GFP expression vector (1) and selected for stable clones with relatively uniform fluorescence. To generate GFP knock-in cell lines, we used in-frame GFP sequences inserted into the same targeting fragments.

For imaging, we used agar overlay (1) and an imaging station optimized for rapid acquisition from photosensitive samples (40). Three-dimensional stacks (41 slices; 250-nm z-step) were captured at multiple positions every 2.5 min with 30-ms exposures. Cells were imaged for 30 min without prior fluorescence exposure. For FRAP, spots were bleached using a 488-nm laser at 40% power (50 iterations over 1.8 μm²) on a Zeiss LSM 710 confocal with a 63×/1.4 NA objective. Stacks (13 slices; step 0.78 μm) were acquired immediately before bleaching and every 5 s for 120 s during recovery. Images are displayed as 2D maximum-intensity projections of 3D stacks.

Data Analysis for RNaseq. Short-read data were obtained for four conditions (0, 5, 5 h, 5 h + 30 min, and 5 h + 60 min actD), in three replicates. Reads were single-end and 50-bp long. We quality-clipped reads, mapped them to Dictyostelium discoideum genome release 9 and corresponding gene models from Ensembl, and obtained normalized gene-expression levels (intensities) quantified as number of fragments (reads) per kilobase of exon per million fragments mapped (FPKM). Intensity errors were substantial for many genes; therefore, we removed noise by rejecting intensities statistically different from zero. The majority of rejected “noisy” data had low expression levels (Fig. S7G). We normalized each gene profile to a calibration reference and selected ribosomal protein genes, which have roughly constant, high expression levels. We calculated mean profiles of all calibration genes (Fig. 57H), then divided all gene profiles by this profile normalized to its own mean. Examples of normalized expression profiles are shown in Fig. S7I.

We defined induction, Rg, of a gene g as the ratio of intensities at 5 h to that at 0-h development. We define degradation, Dg, of a gene g as the ratio of intensities at 5 h and 5 h + 60 min actD. These two quantities describe the base in expression levels during differentiation and amount of mRNA degradation, respectively. Errors were propagated from intensities to find SEs on Rg and Dg. To assess the significance of the trend between Rg and Dg, we divided data into bins of 30 datapoints along the horizontal axis. For each bin, i, we calculated the mean degradation, Di; its SE, ei; and the 95% confidence interval of the mean. We defined a reference level, the mean degradation level at low-induction, Dref, as the mean of all points with Rg < 10.

ACKNOWLEDGMENTS. This work was supported by a Japan Society for the Promotion of Science Postdoctoral Fellowship for Research Abroad (to T.M.) and a Wellcome Trust Senior Research Fellowship and Medical Research Council project grant (to J.R.C.).
Supporting Information

Muramoto et al. 10.1073/pnas.1117603109

SI Materials and Methods

Generation of Cell Lines. MS2 repeats were targeted into S′ coding regions of 10 Dicyostelium genes. To construct MS2 repeat knock-in vectors, we amplified promoter sequences for abpE (actin-binding; −750 to +22), act5 (β-actin; −680 to +21), carA (chemoattractant receptor; −852 to +18), cinD (putative transcription factor; −615 to +18), casA (cell adhesion; −562 to +68), cofC (actin-severing; −816 to +25), hspF (heat-shock; −820 to +27), prl15 (ribosome protein, DDB_G0272893; −945 to +3), scd (fatty acid desaturase, DDB_G0273105; −657 to +25), zfaA (early development marker; −753 to +67) by PCR, and after cloning, checked promoter sequences with dictyBase. Gene regions were amplified and cloned downstream of promoter sequences: abpE (+52 to +1380), act5 (+108 to +1313), carA (+18 to +1321), cinD (+158 to +1313), casA (+80 to +1451), cofC (+56 to +1032), hspF (+18 to +1118), prl15 (+25 to +1071), scd (+212 to +1318), zfaA (+25 to +1071). The act5 and scd targeting vectors were previously described. A BamHI fragment containing bsr downstream of in-frame target sequences for the different times during differentiation. Immediately postbleach, background nuclear intensities for both 0- and 4-h cells were still 85% of original intensity (Fig. S5B). The small fraction of bleached MS2-GFP will affect spot intensity, but bleaching is the same for the different time points, so we can still compare recovery rates and relate these to estimates of elongation rates.

ChIP for RNA pol II. For ChIP, 108 cells were treated with 1.2% (wt/vol) formaldehyde in KK2 for 15 min at room temperature and the reaction was stopped with addition of glycine to a final concentration of 360 mM. Cells were washed twice in PBS, and nuclei isolation was performed using 10 mL of RLB (0.32 M sucrose, 10 mM Tris-Cl pH 7.5, 0.7% SDS) and centrifugation at 1,000 g for 5 min. The isolated nuclei were washed once in RLB. After resuspension in 500 µL of TBS, an equal volume of Sonication buffer [50 mM Tris-Cl pH 8.0, 150 mM NaCl, 2% (vol/vol) Triton X-100, 4 mM EDTA, and 0.1% SDS] was added, then nuclei were sonicated to produce DNA fragments using a Branson sonicator (for 10 s, six times with 6 W). The resulting samples were centrifuged twice at 10,000 × g for 10 min. Chromatin was precleared with protein-G-Sepharose 4 Fast Flow beads (Amersham Biosciences) in the presence of protease inhibitor for 1 h at 4 °C and 37.5 µL of aliquot was retained as an input. Next, 700 µL of ChIP buffer (50 mM Tris-Cl pH 8.0, 150 mM NaCl, 1% Triton X-100, 2 mM EDTA, and 0.05% SDS) was added to 300 µL of precleared chromatin. Immunoprecipitation was carried out using 4 µg of antibodies against H5 (phospho-S2 RNA polymerase II; Covance) and CTD4H8 (phospho-S5 RNA polymerase II; Upstate) overnight at 4 °C. For ChIP using rabbit IgM antibody, beads were incubated for 3 h at 4 °C; washed three times with ChIP buffer, then three times with wash buffer (10 mM Tris-Cl pH 8.0, 500 mM LiCl, 1% deoxycholate, 1% Nonidet P-40, 1 mM EDTA). For ChIP using mouse IgM antibody, the protein-G-Sepharose was incubated with rabbit anti-mouse IgM antibody (Jackson ImmunoResearch) for 1 h at 4 °C and washed with ChIP buffer. The coupling beads were incubated for 3 h at 4 °C with immunoprecipitation samples and washed three times with ChIP buffer then once with IgM-Wash.
buffer (10 mM Tris-Cl pH 8.0, 50 mM LiCl, 0.1% deoxycholate, 0.1% Nonidet P-40, 0.02 mM EDTA). DNA was recovered by incubating with E buffer (50 mM Tris-Cl pH 7.5, 1% SDS, 1 mM EDTA) at 37 °C for 15 min. After reversing the aldehyde cross-links by heating at 65 °C overnight, RNase A was added to final concentrations of 20 µg·mL⁻¹ and incubated at 37 °C for 30 min. The resulting samples were incubated with 200 µg·mL⁻¹ of proteinase K at 56 °C for 1 h. DNA was purified with PCR purification columns (Qiagen), which was also performed for the Input sample. Quantification of DNA sequence associated enriched by antibodies was carried out by real-time PCR with a SYBR Green reagent mix on a realplex2 Thermal Cycler (Eppendorf). Results were normalized to Input. Primers for ChIP from the 5' of the coding region.

**Primers for ChIP (ATG + 1001).**

act5 0837: GGGTCACCCACAAATCTCTTAT
act5 1024: TCCTTGAATCTTCACGGTC
scd 0997: TCTCTTTTGATGATGGTTG
scd 1105: TCTCTTTGTGAATGTTTGGT

**RNAseq.** Total RNA was isolated by RNeasy Mini Kit (Qiagen). Genomic DNA was digested by RNase-Free DNase Set (Qiagen) during RNA purification. RNA was subjected to several QC steps, including integrity (Bioanalyzer.Agilent RNA-nano reagents) and fluorometric determination of concentration and DNA contamination, carried out on the Qubit (Invitrogen QuantRNA); 260:280 was assessed on the Nanodrop1000 (Thermo Scientific). To prepare Illumina RNaseq libraries, 10 µg total RNA was poly-A selected once on oligo(dT) beads (Serabeads). Resultant mRNA was fragmented to an average size of 100 bp using divalent cations at 95 °C for 5 min (Illumina mRNASeq Kit). First-strand cDNA synthesis was carried out using SuperScript III RT (Invitrogen) with the modification of 3 µg random primer 9 (New England Biolabs; S1254S) for optimal cDNA synthesis from the AT-rich transcriptome. Second-strand cDNA and RNAseq libraries were prepared according to the manufacturer’s protocol (Illumina). Briefly, cDNA fragments were blunted and tailed, followed by ligation of Illumina paired-end adaptors. Adapter-ligated fragments were size selected on 2% agarose gels, then subjected to 18 cycles of PCR. At this stage, modified indexed Illumina primers were used to introduce unique barcodes to each sample. The fragment size and concentration of resultant libraries were assessed with Qubit and Bioanalyzer High Sensitivity Chips. Samples were diluted to 10 nmol in EB/0.1% Tween, then pooled and sequenced at a density of six per lane to yield at least two-million reads per sample.

**Data Analysis for RNAseq.** Short-read data were obtained for four conditions (time points 0 h, 5 h, +30 min, and +60 min), in three replicates (R1, R2, and R3). Reads were single-end and 50-bp long. We removed adapter sequences and quality-clipped reads to remove erroneous base calls. Reads were clipped when mean quality score dropped below 20 in a moving window of 3 nucleotides. Clipped reads shorter than 16 bp were rejected. For analysis we used *Dictyostelium discoideum* genome release 9 and corresponding gene model annotation data from Ensembl. We mapped reads to this genome and found splice junctions using TopHat version 1.3.0 (8). We assembled alignments and found transcript abundances using cufflinks version 1.0.3 (9). Random hexamer priming during cDNA generation can cause sequence-specific bias in reads (10). We corrected for this using the build-in option –b in cufflinks. We also used the option –u to estimate the distribution of multiple hits (11). As a result, we obtained normalized gene-expression levels (intensities) quantified as the number of fragments (reads) per kilobase of exon per million fragments mapped (FPKM).

The intensities *I*~g~ were found for each gene, *g*, replicate, *r*, at time point, *t* = 0 h, 5 h, +30 min, +60 min. The intensity profile for each gene was found as the mean of all replicates, *I*~g~(*t*) = $\frac{1}{3}$ $\sum$ *I*~*g*,*r*~(*t*), where *n* is the number of replicates. We also estimated intensity errors, *E*~g~(*t*). For three replicates, the commonly used SD estimator

$$S_g(t) = \sqrt{\frac{1}{n-1} \sum_{r=1}^{n} (I_{g,r}(t) - \bar{I}_g(t))^2},$$

is biased. For *n* = 3, *S*~g~(*t*) is underestimated on average by about 11%. We corrected this bias using a simple approximation proposed by ref. 12:

$$\hat{S}_g(t) = \left[1 + \frac{1}{4(n-1)}\right] S_g(t).$$

We define intensity errors as unbiased SEs:

$$E_g(t) = \frac{\hat{S}_g(t)}{\sqrt{n}}$$

The errors were substantial for many genes, therefore we chose to remove noise. We defined an intensity to be statistically significant when it is statistically different from zero, *I*~g~(*t*) - *t*~*E*~g~(*t*), where *t*~*E*~ is the critical value from Student’s *t*-distribution for a given significance level, *α*, and *n* − 1 degrees of freedom. We assumed *α* = 0.05, for which *t*~2.92~ = 2.92. We accepted a gene’s intensity profile if intensities at all four time points are significant. This conservative approach selects 3,669 adequate profiles from the total number of 10,232. The majority of rejected “noisy” data had low expression levels (Fig. S7G).

We normalized each gene profile to a calibration reference. We selected ribosomal protein genes because they have roughly constant and high expression levels and calculated the mean profile of all these calibration genes (Fig. S7H). Then, we divided all gene profiles by this profile normalized to its own mean. Examples of normalized expression profiles shown in Fig. S7I.

We defined induction, *R*~g~, of a gene *g* as the ratio of intensities at the time of actD treatment to that in the beginning of the experiment:

$$R_g = \frac{I_g(5h)}{I_g(0h)}.$$  

We define degradation, *D*~g~, of a gene *g* as the ratio of intensities at 5 h and 5 h +60 min:

$$D_g = \frac{I_g(5h)}{I_g(60p)}.$$  

Errors were propagated from intensities to find SEs on *R*~g~ and *D*~g~. These two quantities describe increase in expression level during cell development, and the rate of mRNA degradation, respectively. To assess the significance of the trend between *R*~g~ and *D*~g~, we divided data into bins of 30 datapoints along the horizontal axis. For each bin, *i*, we have calculated the mean degradation, *D*~i~ its SE, *ε*~i~, and the 95% confidence interval of the mean. We have also defined a reference level: the mean degradation level at low-induction, *D*~ref~, as the mean of all points with *R*~g~ < 10.


**Fig. S1.** Comparing gene expression in MS2 and parental cell lines. Endogenous gene and MS2 RNA expression were compared in wild-type and the MS2 cell lines, respectively. Expression was assessed by Northern blotting of extracts taken at the indicated developmental times. Equal amount of total RNA was loaded per lane. *act5* was analyzed by RT-PCR because of lack of a specific hybridization probe (there are 17 near-identical act genes).
Fig. S2. Basic imaging statistics. (A) Comparison of background MS2-GFP fluorescence intensity between cell lines. Mean and SD are shown. (B) Number of datapoints analyzed shown for each gene at each stage of development. Within parentheses are numbers of biological replicates on different days. Intensity data were sampled blind from the first imaging fields on a capture day. For weaker expressing genes/conditions, this was not possible, so we used all available data.
Fig. S3. Comparing pulse duration for different genes. (A) Pulse-duration frequency distributions for each gene at 0, 3, and 5 h of development. All transcriptional pulses including those beginning or finishing at the end of imaging period were analyzed to avoid window size skewing data for strong transcribers. (B) Pulse-duration distributions for complete pulses only. With pulse durations from complete pulses, distributions for these genes tended more to exponential. The $\chi^2$ goodness-of-fit tests using the null hypothesis that pulse durations are given by a modified exponential distribution, corrected for the finite imaging window, find the null hypothesis rejected at the 1% level in only one case (csaA 5 h).

Fig. S4. (A) Pulse interval distributions. (B) Distributions of pulse duration and interval from complete cell-cycle data (1) for act5.

Fig. S5. Transcription site intensity data and FRAP. (A) Integrated intensities of individual pulses for act5 and carA at 0 and 5 h. (B) Data in Fig. 4B showing normalized recoveries of spot fluorescence at 0 and 4 h. In addition, the plot contains FRAP recoveries of nuclear fluorescence. Immediately after bleaching, nuclear fluorescence is on average only 15% below starting levels. Nuclear recoveries were normalized to account for photobleaching; intensities were scaled to the normalized cell background (as for the transcription spots), then normalized with respect to the initial nuclear intensity. (C) Scatter plot of initial and final spot intensities for FRAP data from act5 spots at 0 and 4 h. (D) Plot of initial spot intensity and rate of recovery (over 120-s recovery phase) for FRAP data from act5 spots at 0 and 4 h. (E) RNA polymerase modifications at housekeeping and development genes in parental cells. Amount of DNA immunoprecipitated by antibodies against ser2-phosphorylated (S2) and ser5-phosphorylated (S5) Pol II were determined by real-time PCR with primers amplifying the promoter and 5′ regions. Data representative of two biological replicates.
Fig. S6. RNA stability data. (A) Decay of MS2 RNAs for different genes after treatment, commencing at 5 h of development, with 125 μg mL⁻¹ actD. Northern blots probed with MS2 sequence. (B) Northern blots showing RNA decay of five housekeeping (Left) and five developmental (Right) genes after actD treatment, also at 5-h development. Data from parental AX3 cells, using probes against endogenous loci. (C) Intensity information extracted from the blots in A and B, together with mean lifetimes. (D) Testing RNAseq data with Northern blotting of extracts from parental cells. (E) Correlation between Northern and RNAseq data from different extracts. Bars show SE from the three replicates of the RNAseq data.
Fig. S7. Gene functions and mRNA turnover. (A) RNA degradation as a function of coding sequence length. (B) Degradation as a function of GC content. (C–F) Degradation as a function of induction and intensity (expression count) for functional groups of genes (see SI Materials and Methods for gene IDs). Gray points represent all genes with nonzero intensity. Black points bars show 5-h data for 28 genes involved in core cAMP signaling (C and D) and the actB gene family (E and F). Error bars are SEs of the three replicates. Errors in induction and degradation were propagated from intensity (RNA quantity). (G) Selection of low-noise data. The graph shows total fractional variability, $\sqrt{\sum s_i^2(t)/\sum I_g(t)}$ as a function of total intensity, $\sum I_g(t)$, for each gene, g (gray points). Black points show the selection of data where each time point passes the threshold for reproducibility. (H) Calibration data used to normalize gene profiles. Gray lines show individual profiles from ribosomal protein genes. The black line represents their mean. (I) Examples of gene-expression profiles for rpl11 (V18), abpE, csaA, and caA. Three colors represent three replicates. The broken line joins replicate means, $I_g(t)$, and error bars are SEs, $\pm E_g(t)$.

Dataset S1. Read data for RNAseq analysis of RNA decay

Dataset S1 (XLSX)