Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics

David M. Suter,[1]* Nacho Molina,[2]* David Gatfield,[1,3] Kim Schneider,[1] Ueli Schibler,[1]†‡ Felix Naef[2]†‡

In prokaryotes and eukaryotes, most genes appear to be transcribed during short periods called transcriptional bursts, interspersed by silent intervals. We describe how such bursts generate gene-specific temporal patterns of messenger RNA (mRNA) synthesis in mammalian cells. To monitor transcription at high temporal resolution, we established various gene trap cell lines and transgenic cell lines expressing a short-lived luciferase protein from an unstable mRNA, and recorded bioluminescence in real time in single cells. Mathematical modeling identified gene-specific on- and off-switching rates in transcriptional activity and mean numbers of mRNAs produced during the bursts. Transcriptional kinetics were markedly altered by cis-regulatory DNA elements. Our analysis demonstrated that bursting kinetics are highly gene-specific, reflecting refractory periods during which genes stay inactive for a certain time before switching on again.

Polymerase II–mediated transcription of mammalian genes is a complex process consisting of several consecutive steps (*1*). Studies in prokaryotes (*2*–*4*), yeast (*5*–*7*), and higher eukaryotes (*8*–*11*) have suggested that genes are transcribed in a discontinuous fashion, resulting in stochastic production of RNA and protein molecules. Stochastic gene expression has been linked to phenotypic variability, for example, in the resistance to antibiotics in bacterial populations or in the control of developmental transitions in metazoans (*8*, *12*–*14*). Transcriptional bursts can be abstracted in the random telegraph model of gene expression (*2*, *9*, *15*–*17*), whereby transcription switches between "on" and "off" states (Fig. 1A). We monitored transcription kinetics by single-cell time-lapse bioluminescence imaging of mouse fibroblasts expressing a short-lived luciferase reporter gene controlled by endogenous loci, circadian regulatory sequences, or artificial promoters (Fig. 1B). Mathematical modeling allowed us to reconstruct temporal changes in mRNA and protein copy numbers as well as the gene activity state ("on" or "off") at a resolution of 5 min over extended time periods.

Our strategy to monitor transcription is based on the following principle: If both the mRNA transcribed from a gene and its protein product are short-lived, the fluctuations of protein expression should closely reflect the transcriptional bursting kinetics (Fig. 1A). Although this approach cannot capture abortive transcription events, it should reflect the production of mature mRNAs. We engineered a short-lived nuclear luciferase (NLS-luc) encoded by a short-lived mRNA, and fused it to blasticidin deaminase, separated by the 2A peptide of foot-and-mouth disease virus, to generate two polypeptides from one coding sequence (*18*) (Fig. 1B). This cassette was integrated into three types of constructs: (i) a gene trap (GT) lentivector, allowing transcription to be controlled by the endogenous locus at the insertion site (*19*) (the resulting cell lines were dubbed "GT:gene name"); (ii) vectors conferring circadian expression by the *Bmal1* promoter (*20*) (Bmal1/NLS-luc) or the *Dbp* (D site of albumin promoter binding protein) gene (Dbp/NLS-luc); and (iii) vectors carrying artificial promoters consisting of one or two copies of different CCAAT-box variants followed by a TATA box (*21*) (table S1). These constructs were used to generate

[1]Department of Molecular Biology, Sciences III, University of Geneva, and National Centre of Competence in Research Frontiers in Genetics, 30 Quai Ernest Ansermet, 1211 Geneva, Switzerland. [2]Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne and Swiss Institute of Bioinformatics, AAB 021 Station 15, CH-1015 Lausanne, Switzerland. [3]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland.

*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡To whom correspondence should be addressed. E-mail: ueli.schibler@unige.ch (U.S.); felix.naef@epfl.ch (F.N.)

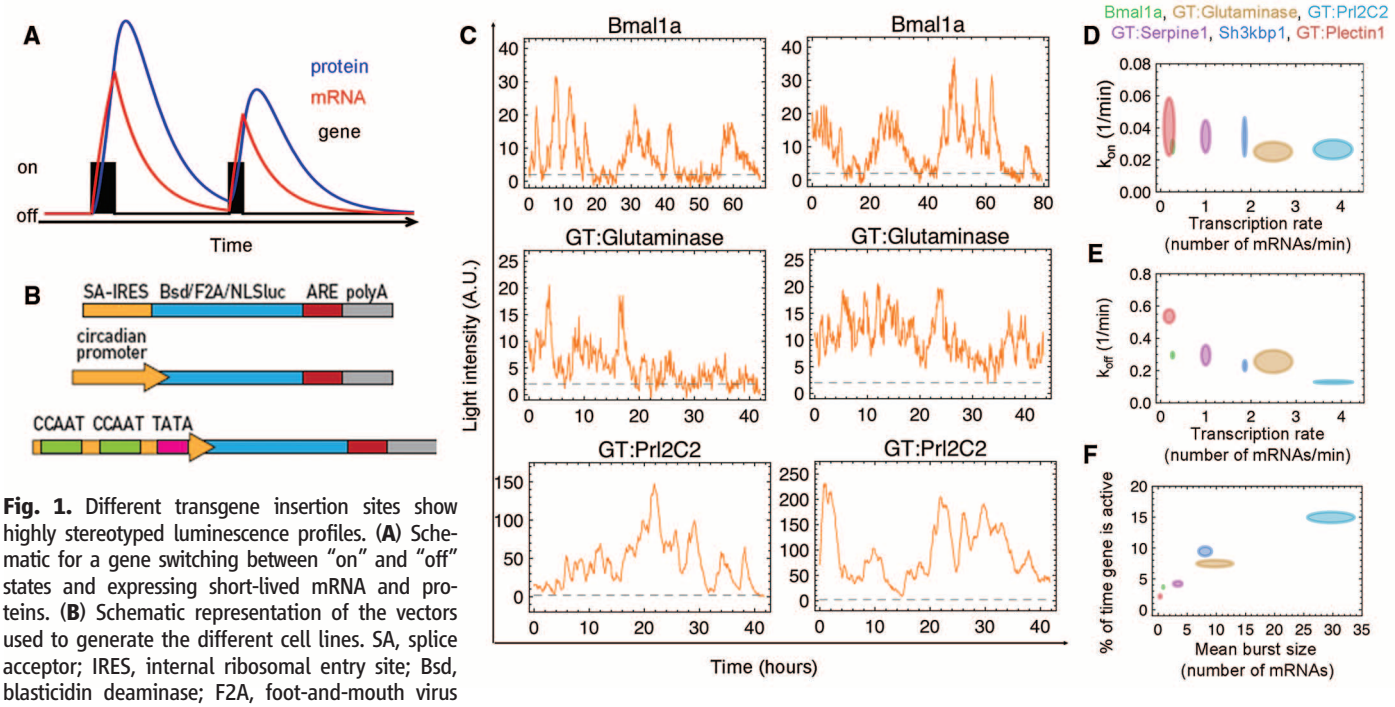

Fig. 1. Different transgene insertion sites show highly stereotyped luminescence profiles. (**A**) Schematic for a gene switching between "on" and "off" states and expressing short-lived mRNA and proteins. (**B**) Schematic representation of the vectors used to generate the different cell lines. SA, splice acceptor; IRES, internal ribosomal entry site; Bsd, blasticidin deaminase; F2A, foot-and-mouth virus peptide 2A; NLS-luc, destabilized nuclear luciferase; polyA, polyadenylation signal; ARE, AU-rich element. (**C**) Examples of single-cell traces. (**D** and **E**) Relationships between transcription rate and k_{on} (D) or k_{off} (E) confirm gene specificity of transcriptional kinetics. (**F**) Mean burst size versus percentage of time during which the gene is active. Ellipses represent means ± 2SD. Color keys are shown on top of (D).

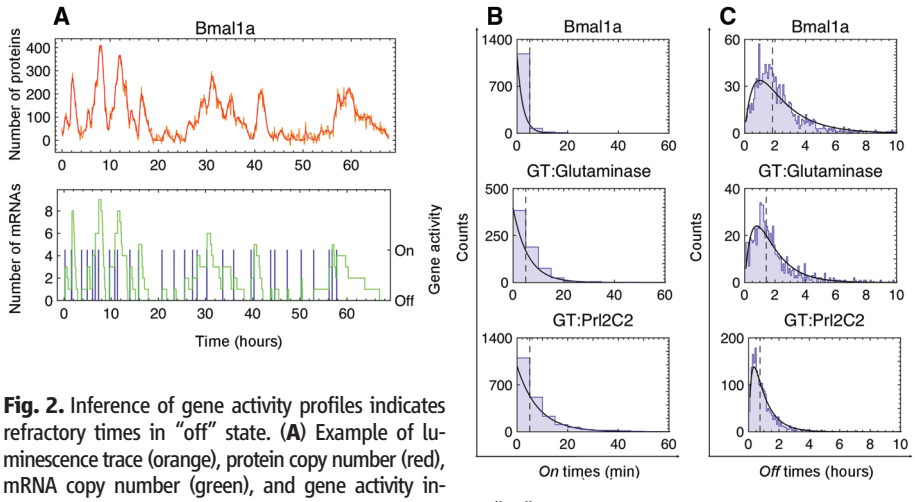

Fig. 2. Inference of gene activity profiles indicates refractory times in "off" state. (**A**) Example of luminescence trace (orange), protein copy number (red), mRNA copy number (green), and gene activity inferences (blue) for the Bmal1a cell line. (**B**) Distribution of "on" intervals; black lines show exponential fits. (**C**) Distribution of "off" intervals; black lines show best fits to "two-step" model (*21*). Black vertical dotted lines show medians of distributions.

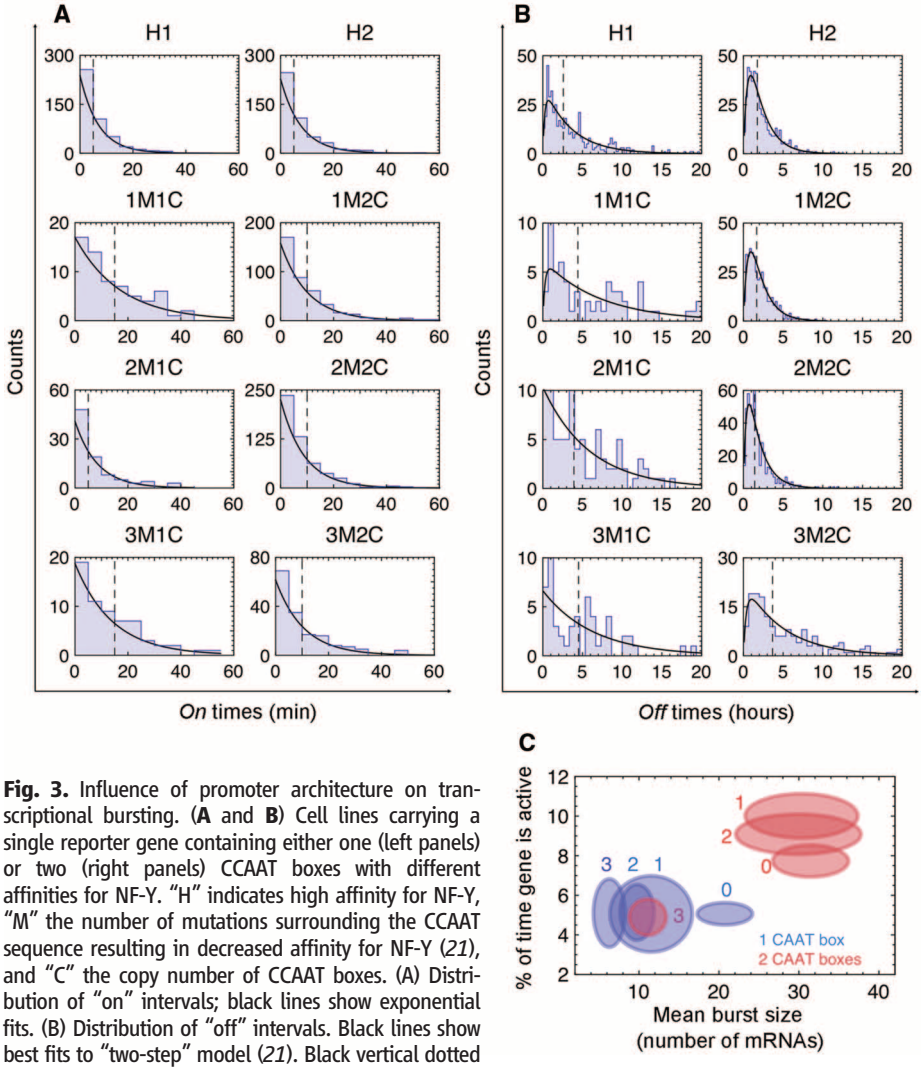

Fig. 3. Influence of promoter architecture on transcriptional bursting. (**A** and **B**) Cell lines carrying a single reporter gene containing either one (left panels) or two (right panels) CCAAT boxes with different affinities for NF-Y. "H" indicates high affinity for NF-Y, "M" the number of mutations surrounding the CCAAT sequence resulting in decreased affinity for NF-Y (*21*), and "C" the copy number of CCAAT boxes. (A) Distribution of "on" intervals; black lines show exponential fits. (B) Distribution of "off" intervals. Black lines show best fits to "two-step" model (*21*). Black vertical dotted lines show medians of distributions. (**C**) Mean burst sizes versus percentage of time during which the gene is active. Ellipses represent contours of mean burst sizes ± 2SD. The numbers 0 to 3 refer to the numbers of mutations adjacent to the CCAAT box sequence.

monoclonal NIH-3T3 cell lines. We also isolated primary mouse fibroblasts from heterozygous and homozygous Per2::luciferase knock-in mice (*22*). Throughout the recordings, we used non-dividing cells to minimize confounding noise sources. All cell lines are listed in table S2.

Figure 1C and fig. S1A show examples of temporal bioluminescence profiles for the cell lines mentioned above. Examination of these data by wavelet analysis (*21*) revealed frequency spectra characteristic for each cell line (fig. S2). As translation is invariably initiated at the internal ribosomal entry site (IRES), discontinuous translation was unlikely to account for the observed differences in temporal bioluminescence profiles. To confirm the transcriptional origin of discontinuous luciferase expression, we blocked transcription in three different cell lines and observed a smooth decay of the luminescence signal in all cells (*21*) (fig. S3). Next, we generated a cell line containing a single transgene that specified a mRNA harboring 24 repeats of MS2-binding sites in its 3′ untranslated region (*21*) whose transcription is driven by two CCAAT boxes. Transcriptional bursting was confirmed by discontinuous appearance of spots in the nucleus upon expression of a yellow fluorescent protein–tagged MS2 (fig. S4).

To further specify the kinetic model, we measured protein and mRNA stabilities by observing luminescence decays in cell populations after translation or transcription inhibition, respectively (*21*) (fig. S5). The photon emissions were calibrated to luciferase protein concentrations by correlating the mean numbers of photons generated by a cell to the mean number of luciferase molecules per cell. This was accomplished for five cell lines spanning a range of expression levels by comparing luminescence emission for a known number of cells with that from a recombinant luciferase having the same specific activity as NLS-luc (*21*) (fig. S6). Knowing the number of molecules per cell, we imaged single cells under conditions used for the time-lapse imaging experiments (fig. S6A) to calibrate photon emission versus protein copy numbers. We also calibrated the number of mRNA molecules using real-time quantitative polymerase chain reaction (fig. S7) (*21*). These calibrations enabled us to develop a stochastic model to quantitatively analyze the data. We computed the probability of the time traces according to the random telegraph model, and computed optimal model parameters using a maximum likelihood approach. We estimated that switching rates between the "on" and "off" gene activity states (k_{on} and k_{off}), the transcription (k_m) and translation (k_p) rates (fig. S8). To visualize the kinetic parameters of transcription for each gene, we plotted k_m against k_{on} and k_{off}. These parameters fell into clear clusters for each cell line, thereby confirming that the kinetic signatures were highly gene-specific (Fig. 1, D and E). Mean burst sizes (i.e., the mean number of transcripts produced per burst) (Fig. 1F and fig. S9B) were in the range of previous quan-

tifications (*5*, *9*, *10*, *23*). To verify that the transcription rate of the tagged mRNA reflected that of the untagged allele, we measured the amount of *Ctgf* mRNA in the GT:ctgf cell line (*21*) and obtained a relative transcription rate of 88 ± 3% as compared to the tagged *Ctgf* allele. Hence, the insertion of a luciferase cassette did not markedly affect the *Ctgf* transcription rate.

Our model could also reconstruct the most probable temporal sequence of protein accumulation, mRNA accumulation, and gene activity states (Fig. 2A and fig. S9A). From this, we computed the distributions of time intervals during which the gene remained "on" and "off," respectively. The "on" intervals followed exponential distributions, suggesting a first-order off-switching of gene transcription (Fig. 2B and fig. S9C). In contrast, the "off" intervals showed a local maximum that was best described by assuming two sequential exponential processes, indicating a refractory period in the "off" state before the gene can be switched on again (Fig. 2C and figs. S9D and S10A). The burst frequency and size of the circadian genes *Bmal1a/b*, *Dbp*, and *Per2::luc* oscillate during a circadian cycle (fig. S11), and we observed a clear phase advance of the former as compared to the latter in the Bmal1a/b and Dbp cell lines (fig. S11). Therefore, burst size and frequency can be uncoupled during up- and down-regulation of a gene.

We next investigated the role of promoter architecture, previously reported to affect on- and off-switching of gene activity in yeast (*24*). To keep the genomic environment invariable, we used NIH-3T3 Flp-In cells to engineer cell lines carrying a single copy of a different artificial construct at the same genomic locus (*21*). To drive NLS-luc expression, the transgenes contain one or two CCAAT boxes (*25*) of different affinities for the ubiquitously expressed NF-Y transcriptional activator upstream of a TATA box, (*21*) (table S1). We generated and analyzed a total of eight different cell lines in which the overall strength of the artificial promoters correlated with both the number of CCAAT boxes and their affinity for NF-Y (fig. S12). Indeed, using two CCAAT boxes instead of one, or choosing CCAAT boxes of higher affinity for NF-Y, increased the mean burst sizes (Fig. 3C, *x* axis). This reflected mainly an increase in the number of transcripts produced during the "on" times, because the duration of "on" times was largely unaffected (Fig. 3A). In addition, the decrease in the duration of "off" intervals (Fig. 3B and fig. S10A) resulted in a higher percentage of activity time in the presence of two CCAAT boxes (fig. S10B).

Finally, we investigated the effect of histone acetylation and chromosomal location on transcription kinetics (Fig. 4), as chromatin environment plays a major role in gene activity (*9*, *12*, *26*). First, we blocked histone deacetylation in four cell lines with trichostatin A (TSA) and recorded single-cell luminescence traces. This treatment abolished the circadian component of the Bmal1a cell line, yet did not drastically affect general aspects of the bursting patterns (fig. S13). In addition, the values for k_{on} and k_{off} were preserved for Bmal1a as well as GT:NcKap1 (fig. S8). In contrast, GT:Prl2c2 and H1 had markedly altered transcription rates (fig. S8) and burst sizes (Fig. 4C) in the presence of TSA, but discontinuous gene activity persisted. Thus, although TSA leads to a spectrum of responses, the qualitative bursting characteristics were not strongly affected. Furthermore, the different chromosomal insertion sites of the transgene in the Bmal1a and Bmal1b cell lines did not markedly influence the temporal activity profiles (Fig. 1, C and F, and figs. S1A and S9B). These results suggest that the chromatin environment plays a secondary role in shaping bursting patterns.

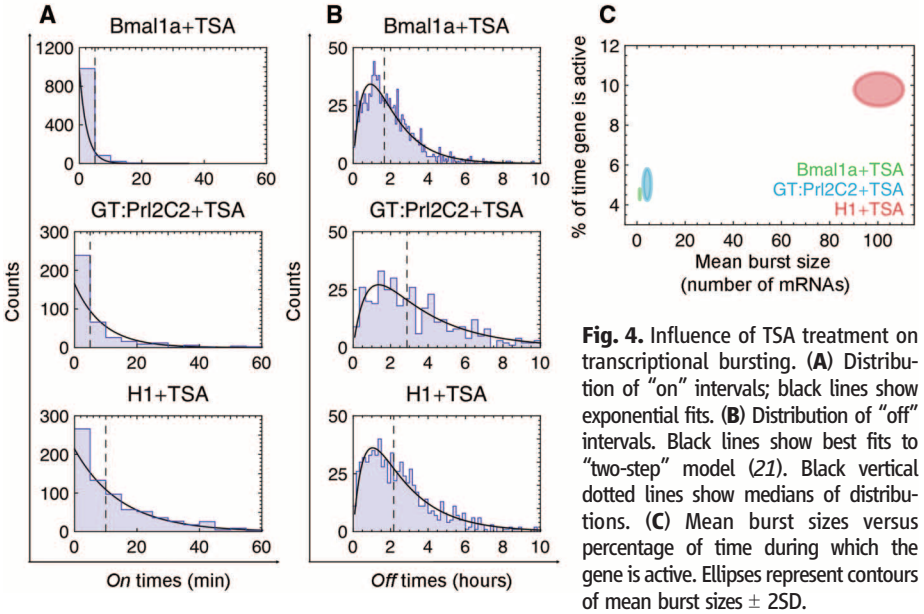

Fig. 4. Influence of TSA treatment on transcriptional bursting. (**A**) Distribution of "on" intervals; black lines show exponential fits. (**B**) Distribution of "off" intervals. Black lines show best fits to "two-step" model (*21*). Black vertical dotted lines show medians of distributions. (**C**) Mean burst sizes versus percentage of time during which the gene is active. Ellipses represent contours of mean burst sizes ± 2SD.

Our findings show that the bursting kinetics of single-allele mammalian genes measured in individual cells at high temporal resolution are highly gene-specific. Cis-acting regulatory elements play a dominant role in shaping transcriptional kinetics.

References and Notes

1. N. J. Fuda, M. B. Ardehali, J. T. Lis, *Nature* **461**, 186 (2009).
2. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
3. L. Cai, N. Friedman, X. S. Xie, *Nature* **440**, 358 (2006).
4. Y. Taniguchi *et al.*, *Science* **329**, 533 (2010).
5. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
6. R. Z. Tan, A. van Oudenaarden, *Mol. Syst. Biol.* **6**, 358 (2010).
7. L. Cai, C. K. Dalal, M. B. Elowitz, *Nature* **455**, 485 (2008).
8. A. Raj, S. A. Rifkin, E. Andersen, A. van Oudenaarden, *Nature* **463**, 913 (2010).
9. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
10. J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, *Curr. Biol.* **16**, 1018 (2006).
11. A. Paré *et al.*, *Curr. Biol.* **19**, 2037 (2009).
12. A. Raj, A. van Oudenaarden, *Annu Rev Biophys* **38**, 255 (2009).
13. A. N. Boettiger, M. Levine, *Science* **325**, 471 (2009).
14. M. F. Wernet *et al.*, *Nature* **440**, 174 (2006).
15. J. M. Pedraza, J. Paulsson, *Science* **319**, 339 (2008).
16. D. R. Larson, R. H. Singer, D. Zenklusen, *Trends Cell Biol.* **19**, 630 (2009).
17. J. Peccoud, B. Ycart, *Theor. Popul. Biol.* **48**, 222 (1995).
18. M. D. Ryan, J. Drew, *EMBO J.* **13**, 928 (1994).
19. W. L. Stanford, J. B. Cohn, S. P. Cordes, *Nat. Rev. Genet.* **2**, 756 (2001).
20. E. Nagoshi *et al.*, *Cell* **119**, 693 (2004).
21. See supporting material on *Science* Online.
22. S. H. Yoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5339 (2004).
23. S. Yunger, L. Rosenfeld, Y. Garini, Y. Shav-Tal, *Nat. Methods* **7**, 631 (2010).
24. L. M. Octavio, K. Gedeon, N. Maheshri, *PLoS Genet.* **5**, e1000673 (2009).
25. R. Mantovani, *Gene* **239**, 15 (1999).
26. X. Darzacq *et al.*, *Annu. Rev. Biophys.* **38**, 173 (2009).

27. We thank A. Liani and Y.-A. Poget for building a bioluminescence reader, C. Bauer and J. Bosset for help with time-lapse imaging, P. Salmon and D. Trono for providing lentiviral constructs, A. Oates for critical reading of the manuscript, and N. Roggli for the artwork. The computations were performed at Vital-IT (www.vital-it.ch). The U.S.'s laboratory was supported by the Swiss National Science Foundation (SNF 31-113565, SNF 31-128656/1, and the NCCR program grant Frontiers in Genetics), the European Research Council (ERC- 2009-AdG 20090506), the State of Geneva, and the Louis Jeantet Foundation of Medicine. F.N.'s laboratory was supported by the Swiss National Science Foundation (SNF grants 3100A0-113617 and 31-130714) and the Ecole Polytechnique Fédérale de Lausanne (EPFL).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1198817/DC1
Materials and Methods
Tables S1 to S3
Figs. S1 to S20
Movies S1 to S3
References

10 December 2010; accepted 4 March 2011
Published online 17 March 2011;
10.1126/science.1198817