vestigate longer-term trends in $[CO_3^{2-}]$ and PCO_2. The observations made here also provide potential to refine the estimates of deep-ocean $[CO_3^{2-}]$ based on the assumption that foraminiferal shell weight is an index of dissolution (*12*). Such estimates clearly require prior knowledge of initial weight, which might be attained from a nearby shallow-core calibration or possibly by using an atmospheric PCO_2 record. It is interesting to note that the shell weights of undissolved foraminiferal calcite appear to trace surface-ocean $[CO_3^{2-}]$, whereas those of dissolved foraminiferal calcite appear to trace $[CO_3^{2-}]$ of the deep ocean.

References and Notes

1. W. S. Broecker, T. Takahashi, H. J. Simpson, T. H. Peng, *Science* **206**, 409 (1979).
2. J. A. Kleypas *et al.*, *Science* **284**, 118 (1999).
3. D. A. Wolf-Gladrow, U. Riebesell, S. Burkhardt, J. Bijma, *Tellus* **51B**, 461 (1999).
4. C. Langdon *et al.*, *Global Biogeochem. Cycles* **14**, 639 (2000).
5. U. Riebesell *et al.*, *Nature* **407**, 364 (2000).
6. F. Marubini, H. Barnett, C. Langdon, M. J. Atkinson, *Mar. Ecol. Prog. Ser.* **220**, 153 (2001).
7. I. Zondervan, R. E. Zeebe, B. Rost, U. Riebesell, *Global Biogeochem. Cycles* **15**, 507 (2001).
8. S. J. Zhong, A. Mucci, *Chem. Geol.* **78**, 283 (1989).
9. P. Zuddas, A. Mucci, *Geochim. Cosmochim. Acta* **62**, 757 (1998).
10. H. Spero, J. Bijma, D. W. Lea, B. E. Bemis, *Nature*, **390**, 497 (1997).
11. J. Bijma, H. Spero, D. W. Lea, in *Uses of Proxies in Paleoceanography: Examples from the South Atlantic,* G. Fischer, G. Wefer, Eds. (Springer-Verlag, Berlin, 1999), pp. 489–512.
12. W. Broecker, E. Clark, *Paleoceanography* **16**, 531 (2001).
13. H. Elderfield, G. Ganssen, *Nature* **405**, 442 (2000).
14. W. S. Broecker, personal communication.
15. For weight analysis, 20 to 24 individual shells were picked, with care given to ensuring the highest degree of morphological consistency. The shells were weighed using a seven-place microbalance (precision = 0.1 μg), the mean weight being taken to represent that population. Size analysis of foraminiferal tests was performed on digital photomicrographs using OPTIMAS image analysis software. Two-dimensional analysis of individual sample populations allows measurement of mean major axis length (*L*) and cross-sectional area (*A*). These values may be used to normalize the mean weight of the sample to the average size for that sieve-size fraction. In this case, normalized $L = 436$ μm and $A = 0.113$ mm^2.
16. W. Broecker, E. Clark, *Geochem. Geophys. Geosyst.* **2**, 2001GC000185 (2001).
17. _____, *Geochem. Geophys. Geosyst.* **3**, 2001GC000231 (2002).
18. D. W. Oppo, S. J. Lehman, *Science* **259**, 1148 (1993).
19. C. J. Bertram, H. Elderfield, N. J. Shackleton, J. A. Macdonald, *Paleoceanography* **10**, 563 (1995).
20. $[CO_3^{2-}]$ values were calculated from TA, TCO_2, T, P, S, $[PO_4]$, and $[SiO_4]$ taken from the North Atlantic WOCE database (*26*) using the CO2SYS program (version 01.05) (*27*). Adjustment to preindustrial values was made by recalculation via CO2SYS using preindustrial PCO_2 (= 0.8 × modern PCO_2) values. Values of $[CO_3^{2-}]$ used in the calibration were calculated for a particular sample via the temperature (*28*) for June at 30-m water depth (assumed calcification season and depth of *G. bulloides*) for the sample site.
21. The CO_2 record used here is plotted on the time scale calculated by tuning Vostok $\delta^{18}O$ to the June insolation curve at 65°N (*29*). Using this time scale allows for better correlation between ice-cores and marine records.
22. Empirical values of $K_{0,1,2}$ in terms of T and S were taken from Millero (*30*). Carbonate alkalinity, CA, ($[HCO_3^-]+2[CO_3^{2-}]$) is considered equal to 96% of TA, allowing for the contributions from borate and water (plus minor others) alkalinity. The value of 96% is taken from Zeebe and Wolf-Gladrow (*31*) as a typical seawater example (TA ≈ PA = 2300 μmol kg^{-1}, DIC = 2000 μmol kg^{-1}, T = 25°C, S = 35).
23. Information on sampling methods, Mg/Ca analysis, temperature, and salinity calculations and $PCO_{2(aq)}$ uncertainty estimations are available at *Science* Online.
24. W. S. Broecker, T.-H. Peng, *Tracers in the Sea* (Lamont-Doherty Geolocical Observatory, Columbia University, Palisades, NY, 1982).
25. For a discussion of the response time of PCO_2 and ΣCO_2 in surface waters, see p. 153 of (*24*).
26. WOCE (World Ocean Circulation Experiment) data are available through the Carbon Dioxide Information Analysis Centre (CDIAC) Web site, at http://cdiac.esd.ornl.gov/.
27. E. Lewis, D. W. R. Wallace, Oak Ridge National Laboratory, ORNL/CDIAC-105 (1998).
28. S. Levitus, T. Boyer, *World Ocean Atlas 1994 Volume 4: Temperature*, NOAA Atlas NESDIS 4 (U.S. Department of Commerce, Washington, DC, 1994).
29. N. J. Shackleton, *Science* **289**, 1897 (2000).
30. F. J. Millero, *Geochim. Cosmochim. Acta.* **59**, 661 (1995).
31. R. E. Zeebe, D. Wolf-Gladrow, *CO_2 in Seawater: Equilibrium, Kinetics, Isotopes*. D. Halpern, Ed., Elsevier Oceanography Series (Elsevier, Amsterdam, 2001), vol. 65.
32. NEAPACC, North East Atlantic Palaeoceanography and Climate Change data (CD-ROM) (Natural Environment Research Council, Swindon, UK).
33. G. G. Bianchi, thesis, University of Cambridge (1999).
34. P. P. E. Weaver *et al.*, *Paleoceanography* **14**, 336 (1999).

35. We thank J. Bijma and Y. Rosenthal for encouraging us to consider the carbonate ion effect, W. Broecker and R. Fairbanks for very helpful discussions supported by the ARCHES program, and F. Millero and J. Bijma for comments on the manuscript. L. Booth and M. Greaves provided invaluable laboratory support. Financial support was provided by the Natural Environment Research Council and the Cambridge-MIT Institute.

Supporting Online Material
www.sciencemag.org/cgi/content/full/297/5582/833/DC1
Methods
Fig. S1
References

12 April 2002; accepted 1 July 2002

Single-Cell Gene Expression Profiling

Jeffrey M. Levsky, Shailesh M. Shenoy, Rossanna C. Pezo, Robert H. Singer*

A key goal of biology is to relate the expression of specific genes to a particular cellular phenotype. However, current assays for gene expression destroy the structural context. By combining advances in computational fluorescence microscopy with multiplex probe design, we devised technology in which the expression of many genes can be visualized simultaneously inside single cells with high spatial and temporal resolution. Analysis of 11 genes in serum-stimulated cultured cells revealed unique patterns of gene expression within individual cells. Using the nucleus as the substrate for parallel gene analysis, we provide a platform for the fusion of genomics and cell biology: "cellular genomics."

The first step in the translation of genomic sequence into physiology or pathophysiology is transcription. Transcriptional regulation has been studied almost exclusively on nucleic acids extracted from cultured cells or tissues by Northern blot (*1*), differential display (*2*), serial analysis of gene expression (SAGE) (*3*), or forms of microarray (*4, 5*). Here, we describe a complementary approach that monitors mRNA synthesis by visualizing specific sites of transcription (*6, 7*). The use of transcription sites for expression profiling allows analysis of coordinated transcription events and organization of gene expression. To achieve sensitive and specific detection of RNAs by fluorescence in situ hybridization (FISH), we used oligomer DNA probes that were each tagged with a single fluorophore at multiple sites (*8*). To detect many mRNAs simultaneously, we used combinations of these probes labeled with spectrally distinct colors (Fig. 1A) (*9*). A combinatorial approach to labeling probes (Fig. 1B) provided a large number of virtual "colors" for distinguishing many transcripts (supporting online text and table S1). Spectral "barcodes" with a minimum of two distinct fluorophores were used to increase specificity. Schemes based on color combinations have been applied to detection of entire chromosomes for cytogenetics (*10, 11*) and for analysis of subchromosomal regions of DNA (*12*). However, the targets in these genomic assays are orders of magnitude larger than transcripts and do not contain functional information.

Cell culture and preparation, fluorescence microscopy, and image acquisition were performed and images were subjected to computational analysis to detect sites of transcription (*9*). Transcription sites have a

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA.

*To whom correspondence should be addressed. E-mail: rhsinger@aecom.yu.edu

Downloaded from www.sciencemag.org on July 25, 2010

10x Genomics EX1015

discernible three-dimensional volume and shape, as well as the highest mRNA signal due to the presence of multiple nascent transcripts at the location of the transcription unit in chromatin (Fig. 1B). The color combination, or spectral barcode, defined transcript identity (Fig. 1C). Colocalization of multiple colors in the image represented independent hybridization events that signified the presence of a transcription site. Signals were analyzed for individual cells on the basis of contiguity of the 4′,6′-diamidino-2-phenylindole (DAPI) counterstain (*9*). We identified 10 sites of transcription simultaneously in starved and serum-stimulated human colon adenocarcinoma (DLD-1) cells (*13*) (Fig. 2, A to C). Even accessibility to probe was assured by presence of transcription sites in more than 97% of cells (176/181). Sites of the most prevalent gene, γ-actin, occurred in 80% of nuclei (144/181). Before the mix of probes for 10 genes was made, each gene was first hybridized individually to ensure that measured expression was independent of multiplexing (*14*).

Assaying expression by transcription site visualization preserves population heterogeneity. For the DLD-1 cell line, the number of transcription sites per cell was not uniform throughout the population. There was a strong correlation with nuclear cross-sectional area and total DNA signal, measured as fluorescence intensity of DAPI counterstain (Fig. 3A) (r^2 value, 0.56). One additional transcription site was observed for each 11.9 μm^2 of nuclear area [95% confidence interval (CI), 10.5 to 13.7 μm^2]. In contrast, starved and stimulated normal fibroblasts did not show a correlation of sites and size. Analysis of 162 fibroblast nuclei fit a linear correlative model quite poorly (r^2 value, 0.0025). Starvation of the cancer cells did not completely abolish increased numbers of transcription sites associated with increased DNA content, as it did in normal fibroblasts. Although levels of transcription were far reduced in starved DLD-1 and cell cycle state was grossly synchronized, differences among cells were detectable in situ. The ability to assay this heterogeneity is a clear advantage in the evaluation of mixed tissue samples, such as is common in clinical pathology.

Population profiles were determined for cells that were previously subjected to microarray and other molecular analyses (*15, 16*). This provided a direct comparison of data based on extracted RNA with single-cell expression levels, a process we termed "FISH & Chips." Thirty min after serum stimulation, the average DLD-1 nucleus had detectable expression of five individual genes and roughly six to seven transcription sites, including both alleles of some genes (Fig. 3B). Genes with frequently occurring sites often had two active alleles, with an approximate 1:1 ratio of two-allele to one-allele nuclei. Genes with lower levels usually had only a single allele expressed (Fig. 3C). This showed that variable expression levels can be visualized in situ and scored for intensity, number of alleles, or dosage. Possible applications include monitoring allele silencing and determining the ploidy of heterogeneous samples.

The ability to collect a binary "snapshot" of genes that are "on" or "off" at a single moment in time for many nuclei offered new insights into transcript regulation. Of the 10 mRNAs analyzed, several pairs of genes were coexpressed with higher probability than would be predicted by random association (table S2). Relative to β-actin–negative cells, nuclei positive for β-actin expression were 5.2 times as likely to express γ-actin (99% CI, 1.6 to 16.7) and 3.2 times as likely to express EGR-1 (99% CI, 1.3 to 7.7). Remarkably, all three of these assayed genes contain one or more sequences to which serum response factor binds (*17–20*). This implies that genes with similar promoter elements exhibit correlated activation at the single-cell level. Such activity could result from nonhomogeneous distribution of a transcription factor in certain cells within the population or increased accessibility of similarly regulated

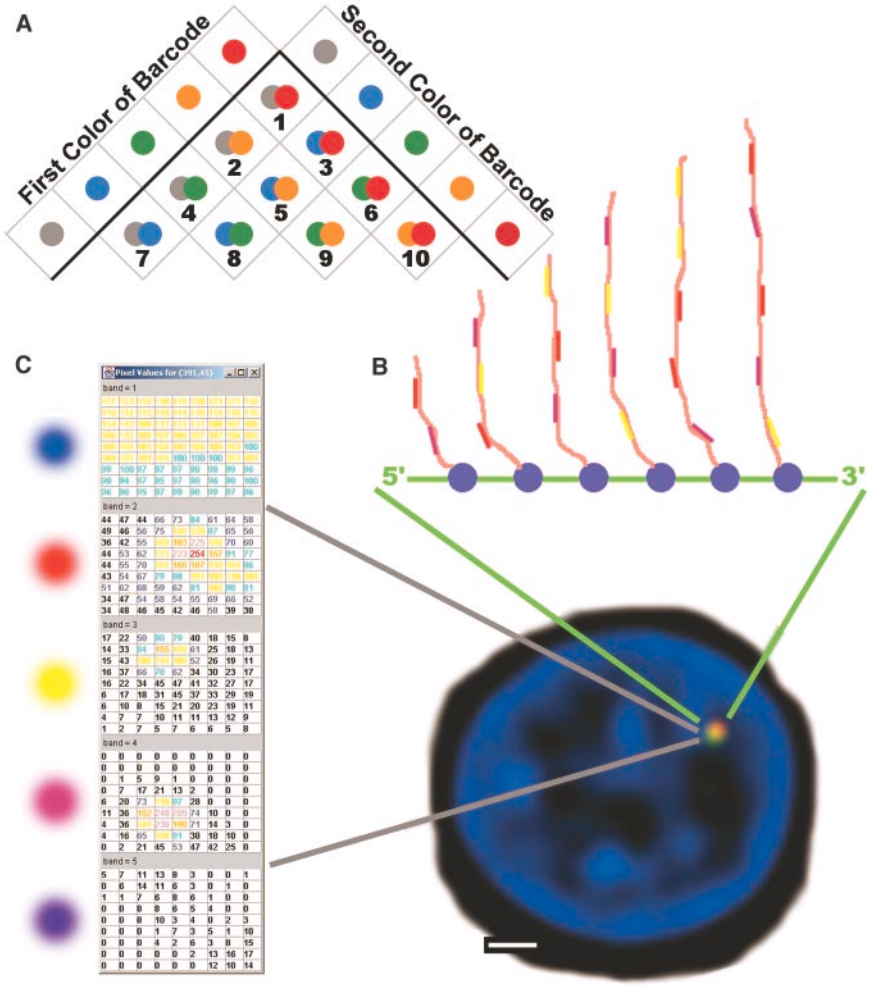


Fig. 1. Conceptual diagram of barcoding. (**A**) Schematic representation of the use of color groups to encode unique gene identities. Combinations shown are only two colors chosen from a total of four used throughout our experiments. A minimum of two colors is used to reduce the chance of single-color false positive signals. Each detected site is therefore representative of two or more independent hybridization events that are spatially colocalized, thus increasing fidelity. (**B**) Schema of the actual probe hybridizations at the transcription site leading to detection by barcode. The gene locus is represented (green) with several polymerases (blue) transcribing nascent messages (pink). A shotgun approach is used such that each specific probe may be of any of the colors in the barcode (red, yellow, magenta). The transcription site is shown below with hybridization data against the nuclear background (blue). Scale bar, 3 μm. (**C**) An example of the signal readout interpreted by the transcription detection algorithm, showing the pixel intensities for the area of the transcription sites in all five color bandwidths. In this example, there is signal for each of the probe components used (red, yellow, and magenta) and only background levels for the color not in the barcode (purple).

Downloaded from www.sciencemag.org on July 25, 2010

genes to their activating factors. Except for the genes just described, cell expression profiles were variable, without a consistent pattern of inclusion and exclusion. If early serum-activated transcription resulted from a distinctly ordered pathway, subsets of cells would have been in different stages of the response. This was not the case. Pairwise analyses of transcription sites showed no distinct mutual exclusions or anticorrelations (the 99% CI for the odds ratio of each of the 45 measured gene pairs exceeded 1). Hierarchical log-linear models showed no higher order associations; cells were positive for different sets of transcripts. At the single-cell level, physiological and random variation, possibly based on the level or activity of a transcription factor, may cause the serum-induced genes to be expressed at different times.

To assess timed expression activation,

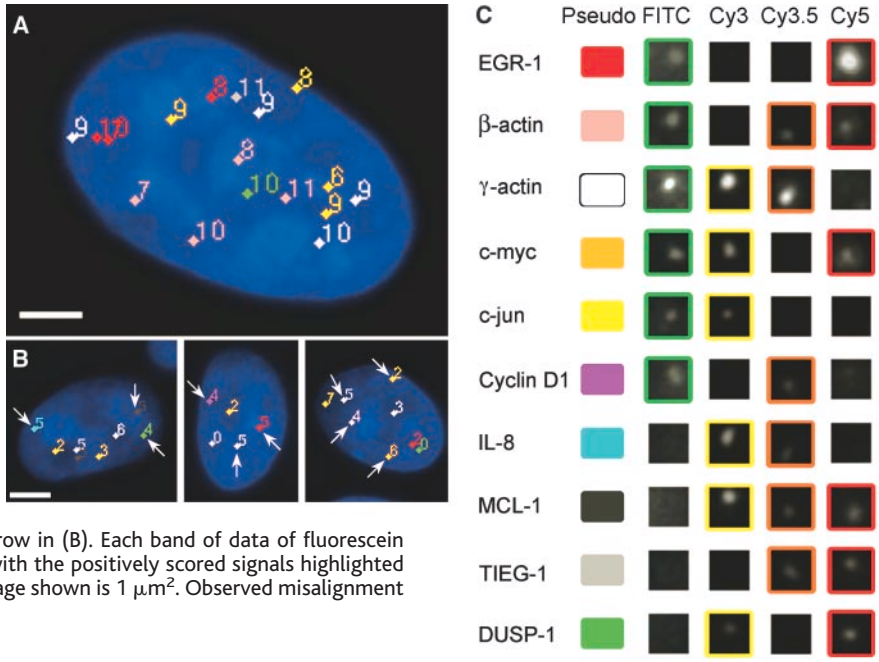


Fig. 2. Simultaneous detection of many genes. (**A**) A single human colon adenocarcinoma cell (DLD-1) G_2 nucleus (by DAPI signal) with a pseudo-colored representation of 17 transcription sites detected in situ. The image is "flattened" such that all 12 0.5-μm Z-sections are displayed on the background, which is the DAPI counterstain from the middle image of the stack. Gene identity is denoted by color and the Z-location is recorded by the adjoining number. Lower numbers represent closer proximity to the cover slip. Scale bar, 3 μm. (**B**) Three G_1 DLD-1 nuclei from the same field, which together express all 10 genes assayed. Arrows indicate sites that are shown below magnified from the original data. From left to right, the 10 marked transcription sites are IL-8, MCL-1, DUSP-1, cyclin D1, γ-actin, EGR-1, TIEG-1, β-actin, c-myc, and c-jun. Scale bar, 3 μm. (**C**) Chart of the 10 genes detected in (B). The "Pseudo" column shows the arbitrary pseudo-color used to denote the gene identity of each transcription site in the renderings above. Columns at right show the actual signal recorded at the appropriate Z-section for the transcription sites shown with an arrow in (B). Each band of data of fluorescein isothiocyanate (FITC), Cy3, Cy3.5, and Cy5 is shown, with the positively scored signals highlighted by surrounding boxes. Each area of the unprocessed image shown is 1 μm^2. Observed misalignment is due to chromatic shift between filter sets.

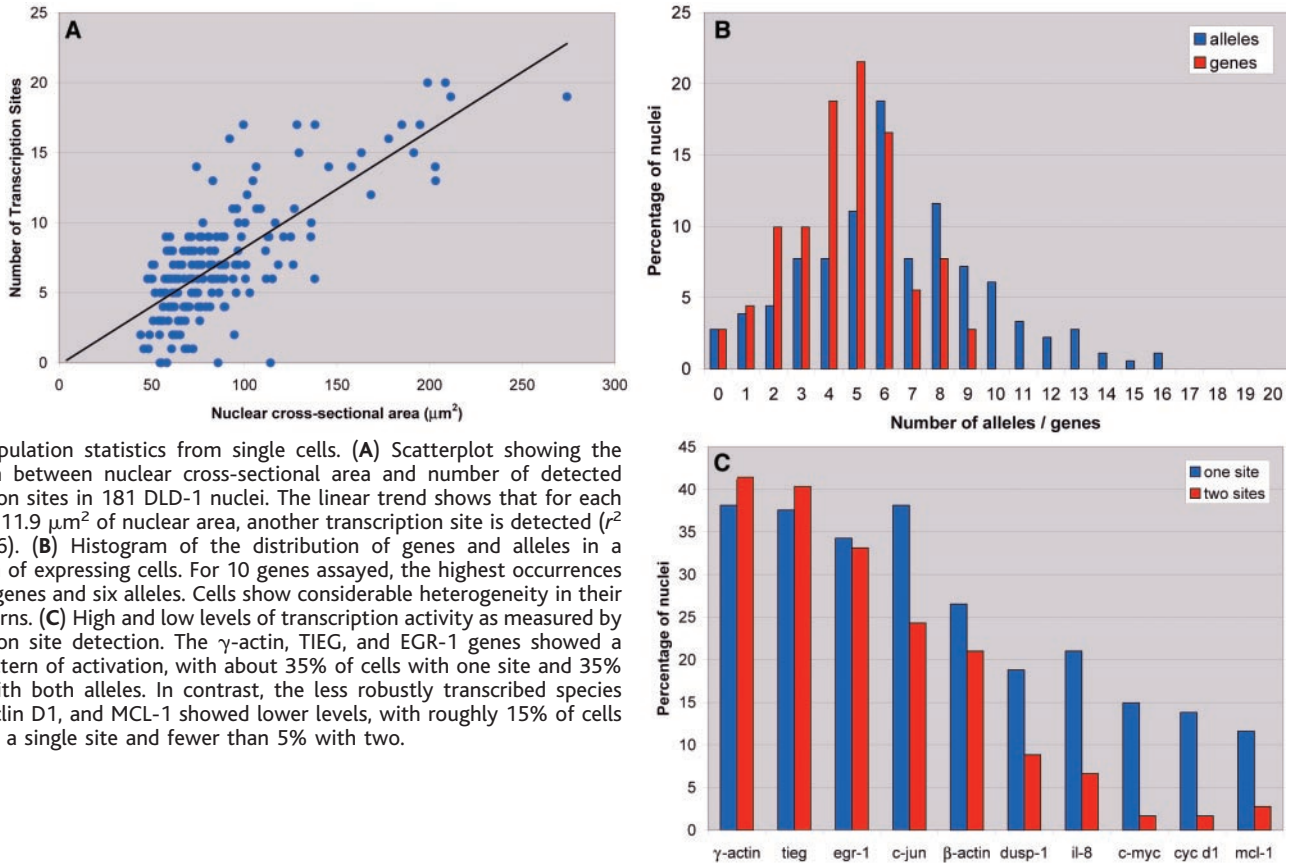


Fig. 3. Population statistics from single cells. (**A**) Scatterplot showing the correlation between nuclear cross-sectional area and number of detected transcription sites in 181 DLD-1 nuclei. The linear trend shows that for each additional 11.9 μm^2 of nuclear area, another transcription site is detected (r^2 value, 0.56). (**B**) Histogram of the distribution of genes and alleles in a population of expressing cells. For 10 genes assayed, the highest occurrences were five genes and six alleles. Cells show considerable heterogeneity in their gene patterns. (**C**) High and low levels of transcription activity as measured by transcription site detection. The γ-actin, TIEG, and EGR-1 genes showed a similar pattern of activation, with about 35% of cells with one site and 35% of cells with both alleles. In contrast, the less robustly transcribed species c-myc, cyclin D1, and MCL-1 showed lower levels, with roughly 15% of cells containing a single site and fewer than 5% with two.

Downloaded from www.sciencemag.org on July 25, 2010

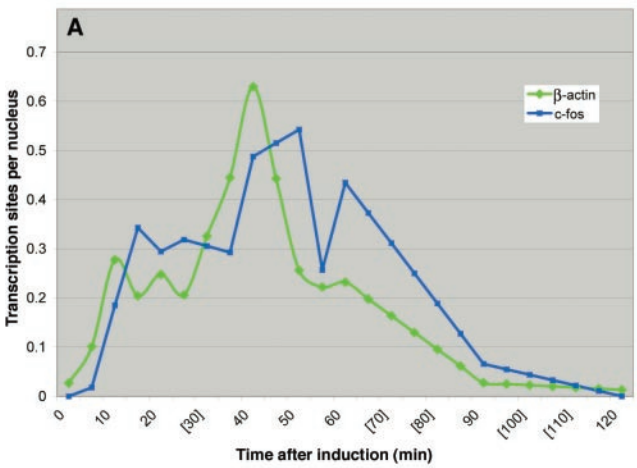
we analyzed the transcription of 11 serum-responsive genes over 14 time points in normal human fibroblasts (*21*). We profiled 2199 nuclei that contained 10,134 transcription sites. For each time point, 55 gene-gene comparisons were made to explore the possibility of ordering of transcription timing (a total of 770 pairwise comparisons). As in the DLD-1 experiments, there were no anticorrelated or mutually excluding gene pairs, indicating that a linear pathway of gene induction is highly unlikely among the early response genes.

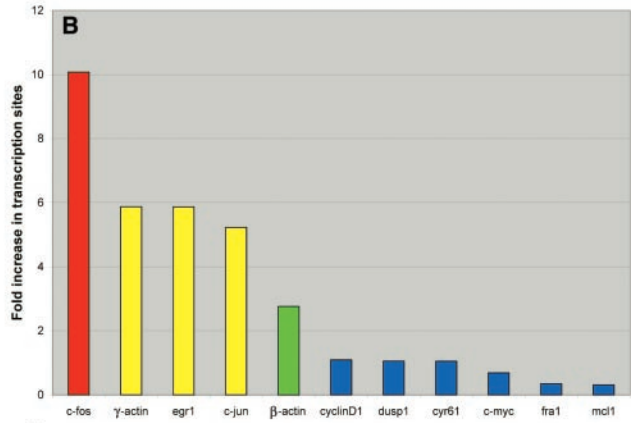
All the induced genes showed similar transcription site activation and deactivation kinetics over a 90-min period (Fig. 4). Several transcripts we visualized easily showed no changes by microarray [less than a factor of 2.2, the detection limit of the technology (*16*)]. These include β-actin (Fig. 4A), γ-actin, and c-jun—all shown to be serum-responsive genes (*22*). Incremental changes may not be detectable using total mRNA. However, with transcription site analysis, small changes in expression could be observed, irrespective of the total abundance of the RNA. Therefore, microarray and in situ assays yield complementary information. A metaphor for the difference is that transcription site profiling measures the transcriptional "thermostat" and microarrays view the "ambient temperature" of whole-cell RNA levels.

All genes showed measurable activation in the population of cells at 5 or 10 min after induction, relative to starved, unstimulated cells (Fig. 4B). Over the first 10 min of induction, one allele was activated. The average cell assayed had 3.2 different genes activated with a total of only 4.0 transcription sites; single allele expression outnumbered the expression of both alleles by a ratio of 3 to 1. By 40 min of induction, the distribution was even: Cells averaged 3.4 genes with a single allele and 2.8 with both alleles active, a total of 9.0 sites per nucleus. From 60 min to 90 min, the number of sites dropped to 2.9. By 120 min, nearly all sites had returned to baseline prevalence (Fig. 4C). This time course fit well with the previous results from a single time point in DLD-1 cells. At 30 min of stimulation, DLD-1 cells showed an average of 6.5 genes and fibroblasts showed 6.1 genes. This implies that serum-responsive transcription follows similar kinetics in different cell types. Differences were also noted, as DLD-1 cells expressed considerably more β-actin (1.2 sites per cell versus 0.33) and somewhat more γ-actin (0.69 sites per cell versus 0.45) than fibroblasts.

Lags were apparent between the kinetics of transcription site activations and their effect on cellular levels of transcripts as assessed by microarray (*14*). This is a function of transcription rate, message stability, and abundance level at the start of induction. Previous analyses measured a combination of these factors, whereas FISH detects only new transcription. Levels of cyclin D1 transcripts reportedly increase later than the early immediate response genes, and it was thus classified as a delayed immediate response gene (*23*). In our studies, cyclin D1 was just as early as other genes with maximal activation at 50 min after stimulation, as compared to the 11-gene average of 46 min. C-myc showed fairly low levels of induction on microarray (*16*), although we have observed frequent c-myc sites (*14*). Levels of c-myc are therefore likely to be controlled posttranscriptionally.

These studies demonstrate the power of cellular transcriptional profiling. The instantaneous transcriptional activity of genes in single cells allows observation of causes and effects of expression. Eventually, the physiological state of cells within tissues will become synonymous with a pattern of gene expression. This will provide a quantitative approach to factors influencing gene expression patterns, such as occur in cytopathology, development and cell differentiation, infectious disease, or response to drug treatment. Investigation of functional genomics may now be approached at the cellular level. We expect that the enormous information inherent to the expression of many genes in large cell populations will aid the understanding of relationships among genes in single nuclei and their cooperative and cumulative roles in physiology and disease.





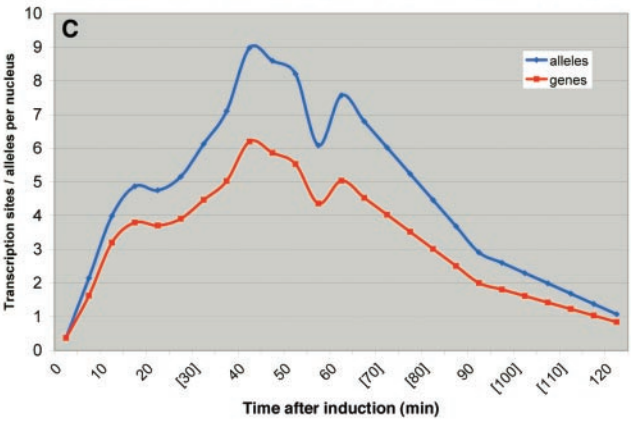


Fig. 4. Transcriptional response of fibroblasts to serum. (**A**) The kinetics of activation of transcription for the 11 genes assayed was relatively similar, in contrast with results obtained through microarray analysis. Here, two serum-responsive genes, β-actin and c-fos, are shown to be similar, yet only c-fos activation was detected on microarray (*16*). (**B**) Increases in visualized transcription sites 10 min after serum induction, relative to starved controls. (**C**) Activation curve for all assayed genes over a time course of serum induction, showing how allele and gene prevalence change as the cascade progresses. In (A) and (C), extrapolated time points are shown in brackets.

Downloaded from www.sciencemag.org on July 25, 2010

References and Notes

1. J. C. Alwine *et al.*, *Methods Enzymol.* **68**, 220 (1979).
2. P. Liang, A. B. Pardee, *Science* **257**, 5072 (1992).
3. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 5235 (1995).
4. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 5235 (1995).
5. D. J. Lockhart *et al.*, *Nature Biotechnol.* **14**, 13 (1996).
6. J. B. Lawrence, R. H. Singer, L. M. Marselle, *Cell* **57**, 3 (1989).
7. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 5363 (1998).
8. E. H. Kislauskis, Z. Li, R. H. Singer, K. L. Taneja, *J. Cell Biol.* **123**, 165 (1993).
9. See supporting data on *Science* Online.
10. M. R. Speicher, S. Gwyn-Ballard, D. C. Ward, *Nature Genet.* **12**, 4 (1996).
11. E. Schrock *et al.*, *Science* **273**, 5274 (1996).
12. J. Brown *et al.*, *Nature Med.* **7**, 4 (2001).
13. The 10 transcripts detected were those of early growth response protein 1 (EGR-1, GenBank accession number NM_001964), β-actin (NM_001101), γ-actin (NM_001614), c-myc (NM_002467), c-jun (NM_002228), cyclin D1 (NM_001758), interleukin-8 (IL-8, NM_000584), myeloid cell leukemia 1 (MCL1, NM_021960), transforming growth factor β immediate early gene (TIEG, NM_005655), and dual specificity kinase 1/MAP kinase phosphatase (DUSP, NM_004417). These genes were first detected singly, using Cy3 and Cy5 probes. Then they were barcoded to distinguishable combinations of fluorophores for multiplexed detection (coding scheme in Fig. 2C). This allows comparison of sensitivity of detection with different color codes to ensure reliability independent of color code. Full sequences of probes are provided in supporting online material.
14. J. M. Levsky, S. M. Shenoy, R. C. Pezo, R. H. Singer, data not shown.
15. G. G. Habets *et al.*, *Methods Enzymol.* **332** (2001).
16. V. R. Iyer *et al.*, *Science* **283**, 5398 (1999).
17. S. Y. Ng, P. Gunning, S. H. Liu, J. Leavitt, L. Kedes, *Nucleic Acids Res.* **17**, 2 (1989).
18. J. L. Schwachtgen, C. J. Campbell, M. Braddock, *DNA Seq.* **10**, 6 (2000).
19. H. P. Erba, R. Eddy, T. Shows, L. Kedes, P. Gunning, *Mol. Cell. Biol.* **8**, 4 (1988).
20. No serum response elements have been reported in the other genes listed in (*13*).
21. As in (*13*), but also including c-fos (GenBank accession number NM_005252), cysteine-rich angiogenic factor (Cyr61, NM_001554) and Fos-related antigen 1 (Fra-1, NM_005438) in the place of IL-8 and TIEG. Full sequences of probes are provided in supporting online material. Experiments were performed for logarithmically growing cells, starved (0 min), and 5, 10, 15, 20, 25, 35, 40, 45, 50, 60, 90, and 120 min after induction. Data were extrapolated for the 30-min time point and points between 60, 90, and 120 min.
22. H. R. Herschman, *Annu. Rev. Biochem.* **60** (1991).
23. A. Lanahan, J. B. Williams, L. K. Sanders, D. Nathans, *Mol. Cell. Biol.* **12**, 9 (1992).

24. We thank C. B. Hall for help with statistical analyses, L. M. Lifschitz for consultation, and S. A. Braut, A. R. Kurland, and J. E. Levsky for technical assistance. Supported by NIH grants T32GM07288 and T32GM07491 (J.M.L.) and grants R33CA83208 and GM54887 (R.H.S.). General information is available at www.SingerLab.org.

Supporting Online Material
www.sciencemag.org/cgi/content/full/297/5582/836/DC1
Materials and Methods
SOM Text
References and Notes
Tables S1 and S2

26 March 2002; accepted 14 June 2002

White Collar–1, a DNA Binding Transcription Factor and a Light Sensor

Qiyang He,[1]* Ping Cheng,[1]* Yuhong Yang,[1] Lixing Wang,[1] Kevin H. Gardner,[2] Yi Liu[1]†

Blue light regulates many physiological processes in fungi, but their photoreceptors are not known. In *Neurospora crassa*, all light responses depend on the Per-Arnt-Sim (PAS) domain–containing transcription factor *white collar–1* (*wc-1*). By removing the WC-1 light, oxygen, or voltage domain, a specialized PAS domain that binds flavin mononucleotide in plant phototropins, we show that light responses are abolished, including light entrainment of the circadian clock. However, the WC-1–mediated dark activation of *frq* remains normal in this mutant, and the circadian clock can be entrained by temperature. Furthermore, we demonstrate that the purified *Neurospora* WC-1–WC-2 protein complex is associated with stoichiometric amounts of the chromophore flavin-adenine dinucleotide. Together, these observations suggest that WC-1 is the blue-light photoreceptor for the circadian clock and other light responses in *Neurospora*.

Blue and near-ultraviolet light are known to regulate many physiological processes in a large number of organisms, including circadian clock functions from fungi to mammals. Currently, phototropins, cryptochromes, and a flavin-adenine dinucleotide (FAD)–containing adenylyl cyclase in *Euglena* are the three known types of eukaryotic blue-light photoreceptors, all of which are flavin-containing proteins (*1–3*). Most fungal photoresponses, ranging from growth responses to phototropism to carotenoid induction to circadian clock entrainment, are mediated by blue light (*4*). For *Neurospora crassa,* previous evidence has suggested that a flavin-containing blue-light photoreceptor is responsible for mediating all known light responses [reviewed in (*4*)].

The *white collar–1* (*wc-1*) and *wc-2* genes are required for all aspects of the known *Neurospora* blue-light responses, including induction of the circadian clock gene *frequency* (*frq*), which has been shown to mediate the light entrainment of the circadian clock (*4–10*). WC-1 and WC-2 are transcriptional activators that can bind to promoter elements upstream of light-inducible genes via their GATA-type zinc-finger DNA binding domains (*5*, *6*, *11*). Both proteins also contain Per-Arnt-Sim (PAS) domains, which are important for the in vivo formation of a nuclear complex between WC-1 and WC-2 (WCC) (*9*, *12*). In addition to their critical roles in light responses, WC-1 and WC-2 are also essential components of the *Neurospora frq-wc*–based circadian feedback loops because they activate the transcription of *frq* in the dark (*7*, *13–16*).

Sequence analysis of the WC-1 protein revealed that one of its three PAS domains (Fig. 1A) belongs to a specialized class of these domains known as a light, oxygen, or voltage (LOV) domain. These LOV domains have been best characterized in plant phototropins, in which each LOV domain binds a flavin mononucleotide (FMN) molecule and is able to undergo fully reversible photocycles (*1*, *17–19*). The crystal structure of a LOV-FMN complex revealed 11 residues in the vicinity of FMN that are highly conserved in LOV domains (*20*), including that of WC-1, suggesting that WC-1 may bind FMN and function as a blue-light photoreceptor.

We made a construct in which the LOV domain was deleted from the wild-type *wc-1* gene (Fig. 1A) and introduced it into a *wc-1* null strain (wc-1.lov) (*21*). The apparent molecular weight of mutant WC-1 was smaller than that of wild-type WC-1, and mutant WC-1 was present in a lower amount than wild-type WC-1 (Fig. 1B). The amount of FRQ in the wc-1.lov mutant was near normal in 22 hours of constant darkness (DD22) but did not increase in cultures grown in constant light (LL), indicating that the light induction of *frq* is lost in the mutant. This result was confirmed with Northern blot analyses of *frq* and *al-3* (a gene required for carotenoid biosynthesis) RNA amounts (Fig. 1C). As for FRQ in the dark, the amounts of *frq* mRNA were comparable in both strains (Fig. 1C, middle panel). These data indicate that the WC-1 LOV domain is essential for light-activated transcription of *frq* but not *frq* transcription activation in the dark.

[1]Department of Physiology, [2]Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: Yi.Liu@UTsouthwestern.edu

Downloaded from www.sciencemag.org on July 25, 2010

View publication stats

Supporting Online Material for Levsky, *et al.*, *Single Cell Gene Expression Profiling*

Materials and Methods

Probe Preparation

Three to five 50-mer DNA probes were synthesized for each target transcript. Sequences were designed using OLIGO 4.0 (Molecular Biology Insights) and verified for specificity using our own JAVA-coded BLAST client that queries the publicly available GenBank library at the NCBI. Amine-modified thymidine bases (five per probe) were used as substrates for amidation using commercially available succinimidyl ester conjugates of fluorescent dyes. In practice, we have used commonly available cyanine dyes as well as fluorescein and rhodamine derivatives. For experiments shown here the fluorophores were Cy3, Cy3.5, Cy5 (Amersham), Oregon Green 488 and Alexa Fluor 488 (Molecular Probes) (Table S1, column 2).

Probe Sequences

Three to five oligomer DNA probes were synthesized for each transcript detected:

Early Growth Response Protein 1 (EGR-1, accession NM_001964):

1 ATCATCTCCTCCAGCTTAGGGTAGTTGTCCATGGTGGGCGAGTGAGGAAA
2 TTCAATTGTCCTGGGAGAAAAGGTTGCTGTCATGTCCGAAAGCCCTGTGG
3 CCTCTCCTATGGCCATCTCCTCCTCCTGTCCTTTAAGTCTCTTGTGTTTC
4 CGATTTTCTAGCATTGAAGGGAGCAAGGGGTCAGGCATATGATGGAGATG

β-actin (NM_001101)

1 TGTCTTCCACAAGCACGTTCTTACCTTAGCCACGAAGTGACCAAGCCACA
2 TTCTTGCCAGCCTCTAGAGAAATCCCAGAACACTCAGCCCTGACACGTTA
3 TTACGGCAGCACTTTTATTTTTCCTTACACAATGACGTGTTGCTGGGGCC

γ-actin (NM_001614)

1 GTGAACTTTGGGGGATGCTCGCTCCAACCGACTGCTGTCACCTTCACCGT
2 TCCTTAGAGAGAAGTGGGGTGGCTTTTAGGATGGCAAGGGACTTCCTGTA
3 CTTTTATTCAACTGGTCTCAAGTCAGTGTACAGGTAAGCCCTGGCTGCCT

c-myc (NM_002467)

1 TCGTAGTCGAGGTCATAGTTCCTGTTGGTGAAGCTAACGTTGAGGGGCAT
2 CCACATACAGTCCTGGATGATGATGTTTTTGATGAAGGTCTCGTCGTCCG
3 TGACCTTTTGCCAGGAGCCTGCCTCTTTTCCACAGAAACAACATCGATTT
4 CTGGTGCATTTTCGGTTGTTGCTGATCTGTCTCAGGACTCTGACACTGTC
5 GGCCTTTTCATTGTTTTCCAACTCCGGGATCTGGTCACGCAGGGCAAAAA

c-jun (NM_002228)

1 GGCGTTGAGGGCATCGTCATAGAAGGTCGTTTCCATCTTTGCAGTCATAG
2 TCGGCCAGGTTCAGGGTCATGCTCTGTTTCAGGATCTTGGGGTTACTGTA
3 GTTTTCACTTTTTCCTCCAGCCGGGCGATTCTCTCCAGCTTCCTTTTTCG
4 CGTGGTTCATGACTTTCTGTTTAAGCTGTGCCACCTGTTCCCTGAGCATG

CyclinD1 (NM_001758)

1 CTACGCGCGGATGGTTTCCACTTCGCAGCACAGGAGCTGGTGTTCCATA
2 TTGACAGGAAGCGGTCCAGGTAGTTCATGGCCAGCGGGAAGACCTCCTCTTCGCACTA
3 CTTCTTAGAGGCCACGAACATGCAAGTGGCCCCCAGCAGCTGCAGGCGGCTCTA
4 ATCTTGAGCTTGTTCACCAGGAGCAGCTCCATTTGCAGCAGCTCCTCGGGCCGTC

Interleukin-8 (IL-8, NM_000584)

1 ACAGTGAGATGGTTCCTTCCGGTGGTTTCTTCCTGGCTCTTGTCCTAGAA
2 CTAAGTTCTTTAGCACTCCTTGGCAAAACTGCACCTTCACACAGAGCTGC
3 TTCTGTGTTGGCGCAGTGTGGTCCACTCTCAATCACTCTCAGTTCTTTGA
4 GAATTCTCAGCCCTCTTCAAAAACTTCTCCACAACCCTCTGCACCCAGTT

Myeloid Cell Leukemia 1 (MCL1, NM_021960)

1 AATCCTGCCCCAGTTTGTTACGCCGTCGCTGAAAACATGGATCATCACTC
2 AACCCATCCCAGCCTCTTTGTTTAACTAGCCAGTCCCGTTTTGTCCTTAC
3 AGCAGCACATTCCTGATGCCACCTTCTAGGTCCTCTACATGGAAGAACTC
4 TCCACCCTACCATCTTCACTAAATCTAAAAGTCCTCCTCCATAGCTTCCC

Transforming Growth Factor β Immediate Early Gene (TIEG, NM_005655)

1 TTGGCAGTATCTGAGAGTGACTTGAAGTGTACAGTAGATGGCGCTGGTGC
2 GGCATCAGCTGTATGACGAATCACACTTGTTGCCTGAGCTTTGGGGAGTT

3 GGCATCAGCTGTATGACGAATCACACTTGTTGCCTGAGCTTTGGGGAGTT
4 TGCGTCCTCGTGTGGGCCTTCAGATGGGAACTTTTAAAGTATGTCTTGCC

Dual Specificity Kinase 1 / MAP Kinase-Phosphatase (DUSP, NM_004417)
1 TGTCCTCCACAGGGATGCTCTTGTACTGGTAGTGACCCTCAAAATGGTTG
2 ATTAGTCCTCATAAGGTAAGCAAGGCAGATGGTGGCTGACCGGGAAATGC
3 GCTGAAGTTGGGAGAGATGATGCTTCGCCTCTGCTTCACAAACTCAAAGG
4 TAAATAAGGACCAGCCCTCTCGAGCCCCTCCCAGAGTTATTGCATTTCTC

c-fos (NM_005252)
1 CCATGCTGGAGAAGGAGTCTGCGGGTGAGTGGTAGTAAGAGAGGCTATCC
2 GGATGAACTCTAGTTTTTCCTTCTCCTTCAGCAGGTTGGCAATCTCGGTC
3 GATGCTGGGAACAGGAAGTCATCAAAGGGCTCGGTCTTCAGCTCCATGCT
4 TAGGTGAAGACGAAGGAAGACGTGTAAGCAGTGCAGCTGGGAGTACAGGT

Cysteine-Rich Angiogenic Factor (Cyr61, NM_001554)
1 CCATTCCAAAAACAGGGATCCGCTTCAGTGAGCTGCCTTTTCCAACTGCA
2 TTTCACAAGGCGGCACTCAGGGTTGTCATTGGTAACTCGTGTGGAGATAC
3 CGCAGTACTTGGGCCGGTATTTCTTCACACTCAAACATCCAGCGTAAGTA
4 CCCCGCCCATTTTCTCCATGATTCTGATTCTGACACTCTTCTCCCTTGTT

Fos-related Antigen 1 (Fra-1, NM_005438)
1 CTCCTCTTCCTCCGGGCTGATCTGTTCACAAGGCCTTCGACGTACCCCTG
2 ATCTCTCGCTGCAGCCCAGATTTCTCATCTTCCAGTTTGTCAGTCTCCGC
3 TGAAGACCAGGCTGGGGGTGAAAGGAGTTAGGGAGGGTGTGGTCATGAGT
4 AGGATGGGTCTCCGCTGCTGCTGCTACTCTTGCGATGAGCTGAGGCACAA

Cell Culture and Hybridization

DLD-1 cells (ATCC CCL-221) and normal human fibroblasts (ATCC CRL-2091) were grown under standard conditions and plated onto cover slips. To augment and synchronize transcription, cells were serum-starved for 24 hours and serum-stimulated (10% FBS) for 30 min in the presence of cycloheximide (10ug/ml). Cytoplasm was extracted with .5% Triton X-100 to increase probe penetration and decrease cytoplasmic fluorescence. The nuclei were fixed in 4% paraformaldehyde for 15 min at room

temperature, washed in PBSM (phosphate buffered saline with 5mM $MgCl_2$), and stored in 70% ethanol at 4C. For hybridization, cells were rehydrated into PBSM and equilibrated in 50% formamide / 2X SSC for 10 min at room temperature. Hybridization was performed for 3 hours in 50% formamide / 2X SSC / 10% BSA at 37C. The concentrations of probes labeled in each color for each gene were determined empirically. Cover slips were then incubated twice in 50% formamide / 2X SSC for 20 min at 37C. Next, the samples were processed through decreasing concentrations of SSC, and finally returned to PBSM for DAPI counterstaining. Slides were mounted in 90% glycerol in PBS with phenylenediamine.

Imaging

Three-dimensional stacks of images were acquired using AX70 and BX51 microscopes (Olympus) with a PZ54 E piezoelectric translator (Physik Instrumente) and CH-350(502) and CoolSNAP-HQ Charge-Coupled Device (CCD) cameras (Roper Scientific), using IPLab software version 3.07 (Scanalytics). We used PlanApo 60x, 1.4 NA objectives (Olympus) and HiQ band pass filters for DAPI (#31000), FITC (#41001), Cy3 (#SP-102v1), Cy3.5 (#SP-103v1) and Cy5 (#41008) (Chroma Technology).

Image Processing

Transcription site detection algorithms and nuclear segmentation were developed in our own software package, using the Java Development Kit 1.3.0 and the Java Advanced Imaging Extension 1.0.2 (Sun Microsystems). CCD-captured image volumes for an arbitrary number of fluorescent colors are read and compiled into a single data image. Then, feature extraction by convolution filtering is performed to highlight areas of punctate signaling using kernels devised and tested empirically. Areas of peak intensity are chosen by processing the entire image in all color bands. These regions are sorted by barcode, or the combination of signals present, based on an adjustable threshold. Finally, the identity of the signal is interpreted based on a color coding scheme. A minimum number of color signals can be imposed in order to better distinguish between true signals from multiple fluorophores and noise from the assay. The data is exported to a spreadsheet program for analysis.

A second step of analysis involves determining the co-localization of the detected signals inside individual nuclei. The nuclear counterstain (DAPI image) is binarized using a selectable threshold and contiguous areas are "flood-filled". Bodies that are not totally surrounded by background are ruled out as they are only partially imaged and cannot be completely analyzed. Then, the set of punctate signals is superimposed on the nuclear regions. Signal outside any body can be ignored to reduce further processing. Signals within bodies are scored for number (or simply for absence and presence) for each detectable species in the color code. Statistics are compiled for the entire field of view as well as for each body in a form exportable for further analysis.

Supporting Text

Number of Barcodes

Simple combinations of colors, or barcodes, are used to encode unique transcript identities. For c total colors, B_c, the number of barcodes, is given by:

$$B_c = \sum_{i=1}^{c} cCi = \sum_{i=1}^{c} \frac{c!}{(c-i)!\cdot i!} = 2^c - 1$$

As the number of fluorophores increases, the number of detectable transcripts increases exponentially. Examples are given below (Table S1, row 1).

To increase the number of detectable transcripts beyond the capability provided by using color alone we used a ratio-labeling scheme. This approach has been demonstrated previously for detection of DNA targets by *in situ* hybridization *(1, 2)*. The high fidelity output of our computerized algorithms allows the precise discrimination of gradations of color intensity values for RNA FISH. Our preliminary studies have shown that computer-analyzed transcription site profiling can allow the discrimination of at least two gradations of color intensity values in the detection of sites. For c total colors, R_c, the number of ratio-labeled barcodes, is defined by the following formula given N_i discernable gradations of intensity, C_i possible sets of ratios for a single color combination, and B_i binary color combinations, for i colors:

$$R_c = \sum_{i=2}^{c} N_i \cdot C_i \cdot B_i = \sum_{i=2}^{c}\left(N_i \cdot \sum_{j=1}^{[i/2]} iCj \cdot cCi \right) = \sum_{i=2}^{c}\left(N_i \cdot \sum_{j=1}^{[i/2]}\left[\frac{i!}{(i-j)!\cdot j!} \right] \cdot \frac{c!}{(c-i)!\cdot i!} \right)$$

Note that a minimum of two colors is always used. Example values are shown below (Table S1, rows 3-6).

Increases in the number of detectable targets can also be accomplished by separating existing fluorochromes that have similar emission spectra using computerized algorithms for spectral decomposition *(3)*, or through the use of additional fluorophores. Using these means and the resolution afforded by digital imaging, a large cohort of genes could be assayed, which will provide a more comprehensive understanding of transcription.

Supporting Tables

Table S1. Numbers of barcodes. Using a given number of colors, the resulting quantity of distinct combinations that can be used for barcodes is determined by the formula given above in supporting online text (row 1). In order to increase fidelity of the assay, barcodes are limited to a minimum of two colors (row 2). In order to increase the number of detectable species, we apply a ratio labeling protocol yielding barcode numbers as above in supporting online text (rows 3-6). These predictions include only separation of two levels of intensity (N_i=2). The numbers are far greater given a higher number of discernable intensity thresholds.

Detection Scheme	3 Fluor.	4 Fluor.	5 Fluor.	6 Fluor.
Qualitative Barcode	7	15	31	63
Qualitative Barcode, min. 2 colors	4	11	26	57
+ 2-Color Ratios, $N_2 = 2$	7	17	36	72

+ 3-Color Ratios, $N_3 = 2$	12	37	86	172
+ 4-Color Ratios, $N_4 = 2$	-	56	181	457
+ 5-Color Ratios, $N_5 = 2$	-	-	210	631

Table S2. Pair wise correlations of genes expressed in stimulated DLD-1 cells. Single cell gene expression correlations are analyzed by scoring the number of nuclei positive for any two given genes, the number positive for one, and the number of double-negatives to determine an odds ratio. These are the only statistically significant, positively-correlated gene pairs for the ten genes and 45 possible pairs, as determined to 99% confidence.

Gene Combination	Odds Ratio.	99% CI
γ-actin & DUSP-1	**5.49**	**(1.09, 27.75)**
γ-actin & β-actin	**5.21**	**(1.62, 16.71)**
β –actin & DUSP-1	**4.17**	**(1.63, 10.62)**
MCL-1 & DUSP-1	**3.86**	**(1.25, 11.91)**
β –actin & MCL-1	**3.57**	**(1.06, 12.02)**
β –actin & EGR-1	**3.22**	**(1.34, 7.74)**
MCL-1 & IL-8	**3.19**	**(1.04, 9.80)**
γ-actin & EGR-1	**2.74**	**(1.04, 7.28)**

Supporting References and Notes

1. P. M. Nederlof, S. van der Flier, J. Vrolijk, H. J. Tanke, A. K. Raap, *Cytometry* 13, 8 (1992).
2. J. G. Dauwerse *et al.*, *Hum. Mol. Genet.* **1**, 8 (1992).
3. E. Schrock *et al.*, *Science* **273**, 5274 (1996).