7. L. A. Tonkin *et al.*, *EMBO J.* **21**, 6025 (2002).
8. M. Higuchi *et al.*, *Cell* **75**, 1361 (1993).
9. A. Herb, M. Higuchi, R. Sprengel, P. H. Seeburg, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1875 (1996).
10. S. M. Rueter, T. R. Dawson, R. B. Emeson, *Nature* **399**, 75 (1999).
11. R. A. Reenan, C. J. Hanrahan, B. Ganetzky, *Neuron* **25**, 139 (2000).
12. Materials and methods are available as supporting material on *Science* Online
13. P. J. Aruscavage, B. L. Bass, *RNA* **6**, 257 (2000).
14. G. Eaholtz, T. Scheuer, W. A. Catterall, *Neuron* **12** 1041 (1994).
15. Y. Jiang *et al.*, *Nature* **423**, 33 (2003).
16. D. A. Doyle *et al.*, *Science* **280**, 69 (1998).
17. M. Yoshihara, J. T. Littleton, *Neuron* **36**, 897 (2002).
18. A. F. Davis *et al.*, *Neuron* **24**, 363 (1999).
19. E. R. Chapman, R. C. Desai, A. F. Davis, C. K. Tornehl, *J. Biol. Chem.* **273**, 32966 (1998).
20. T. Konno *et al.*, *Proc. R. Soc. London Ser. B* **244**, 69 (1991).
21. K. Imoto *et al.*, *FEBS Lett.* **289**, 193 (1991).
22. M. Zhou, J. H. Morais-Cabral, S. Mann, R. MacKinnon, *Nature* **411**, 657 (2001).
23. A. Kuryatov *et al.*, *J. Neurosci.* **17**, 9035 (1997).
24. B. Maylie, E. Bissonnette, M. Virk, J. P. Adelman, J. G. Maylie, *J. Neurosci.* **22**, 4786 (2002).
25. Y. Jiang, V. Ruta, J. Chen, A. Lee, R. MacKinnon, *Nature* **423**, 42 (2003).
26. G. A. Lopez, Y. N. Jan, L. Y. Jan, *Neuron* **7**, 327 (1991).
27. O. Yifrach, R. MacKinnon, *Cell* **111**, 231 (2002).
28. S. Tamamizu, Y.-H. Lee, B. Hung, M. G. McNamee, J. A. Lasalde-Dominicci, *J. Membr. Biol.* **170**, 157 (1999).
29. We thank M. Lalande, A. Das, B. Graveley, and S. Helfand for critical reading of this manuscript. Supported by NIH R01 GM062291, NSF 0091142, and the Donaghue Medical Research Foundation.

Supporting Online Material
www.sciencemag.org/cgi/content/full/301/5634/832/DC1
Materials and Methods
Figs. S1 to S18
Table S1

13 May 2003; accepted 1 July 2003

Single Molecule Profiling of Alternative Pre-mRNA Splicing

Jun Zhu,[1]* Jay Shendure,[1]* Robi D. Mitra,[2] George M. Church[1]†

Alternative pre-messenger RNA splicing is an important mechanism for generating protein diversity and may explain in part how mammalian complexity arises from a surprisingly small complement of genes. Here, we describe "digital polony exon profiling," a single molecule–based technology for studying complex alternative pre-messenger RNA splicing. This technology allows researchers to monitor the combinatorial diversity of exon inclusion in individual transcripts. A minisequencing strategy provides single nucleotide resolution, and the digital nature of the technology allows quantitation of individual splicing variants. Digital polony exon profiling can be used to investigate the physiological and pathological roles of alternately spliced messenger RNAs, as well as the mechanisms by which these messenger RNAs are produced.

Alternative splicing of eukaryotic pre-mRNAs is a powerful means of regulating gene expression and enhancing proteome diversity (*1*). A recent study estimated that greater than 55% of human genes are alternatively spliced (*2*). Complex cases, such as the neurexin and CD44 loci, have two or more alternative splice sites, and the independence of alternatively spliced cassettes can lead to a multiplicative accumulation of proteins encoded by a single locus (*3*). As a striking example, the *Drosophila* gene encoding the Dscam axon guidance receptor can potentially generate 38,016 different protein isoforms (*4*). Although dozens of cis-regulatory elements and trans-acting factors that affect splicing have been identified (*5*), the molecular mechanisms governing complex spatiotemporal control of alternative pre-mRNA splicing remain poorly defined. Any breakthrough in our understanding will require a reliable technology that allows cost-effective and quantitative monitoring of splicing isoforms.

We describe here an approach to profiling complex alternative splicing via the polymerase colony (polony) technology, which permits the solid-phase parallel amplification of thousands to millions of DNA or RNA molecules, such that each template gives rise to an individual colony of amplification products (*6*, *7*). One strand of each amplicon is covalently linked to an acrylamide matrix and serves as a template for probe hybridization and/or single base extensions (SBEs). Detection of multiple gene-specific sequences in cis is achieved by combinations of spectrally distinct fluorophores and/or repeated cycles of probing. In the case of alternative splicing, combinatorial patterns of exon inclusion or exclusion can be unequivocally determined across multiple polonies in parallel. Furthermore, because each polony arises from a single molecule, the digital nature (*8*) yields a sensitive and accurate means of quantifying individual mRNA isoforms in one or more pools of interest.

As a proof of concept, we considered the human microtubule-associated protein, MAPT, which is located at human chromosome 17q21 and is primarily expressed in brain tissue. Three exons (2, 3, and 10) in MAPT transcripts undergo regulated alternative splicing, and disruption of this process can lead to frontotemporal dementia with parkinsonism (*9*). The inclusion of exon 3 is always dependent on the inclusion of exon 2; thus, six different MAPT transcripts can be generated by alternative splicing (Fig. 1A). Linearity of polony amplification was examined with two in vitro transcribed MAPT variants, and the correlation coefficient between predicted and experimental data is greater than 0.99 (fig. S1). We next amplified polonies from human brain cDNA with the use of an acrydite-anchored MAPT exon 1–specific primer and a free exon 11–specific primer (*10*). After denaturation of the nonanchored strands of the polymerase chain reaction (PCR) amplicons, Cy3-labeled exon 2– and Cy5-labeled exon 3–specific probes were hybridized to the slides, and images were acquired with a standard microarray scanner. Amplicons bearing exon 2 (green) and/or exon 3 (red) yielded discrete polonies. A merged image shows colocalization of exon 2 and exon 3 in some cases (yellow), but no polony bears exon 3 by itself (Fig. 1B). After denaturing away the first set of probes, we applied Cy3-labeled exon 7 and Cy5-labeled exon 10 probes to the same slide. Exon 7 is not alternatively spliced and therefore allowed us to detect all MAPT-transcript polonies, including the "zero" alternative isoform. The exon 10 probe, in contrast, yielded a pattern distinct from exon 2 and exon 3 inclusion patterns. Overall, we identified all six possible alternative splicing variants without ambiguity (Fig. 1B). A confidence interval for the relative abundance of each individual isoform was determined by simply counting the number of polonies bearing a given combination of exons and applying standard statistics (Fig. 1C).

Consistent with previous estimations (*11*), the exon 3–containing variants (+2+3 and +2+3+10) were the least abundant species, comprising less than 10% of total MAPT mRNAs. The remaining isoforms were divided almost equally between isoforms lacking exon 2 ("zero" and +10) and isoforms containing exon 2 (+2 and +2+10). However, we detected a

[1]Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [2]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: church@arep.med.harvard.edu

slightly greater level of inclusion of one or more alternative exons relative to the "zero" isoform (Fig. 1C) than that of a previous study (*11*). Although variation between independent samples cannot be ruled out, the discrepancy is potentially explained by preferential amplification of short and abundant splicing isoforms in semiquantitative reverse transcription PCR (RT-PCR) studies (*12*).

Combining SBEs on polonies (*7*, *13*) with the exon profiling developed here allowed us to correlate single nucleotide changes (SNCs) with splicing defects. A number of SNCs, including SNCs at intronic, synonymous, and nonsynonymous positions, can lead to splicing alterations that underlie human diseases (*14*). We studied SNC-induced splicing alterations in the two nearly identical human survival motor neuron (SMN) genes, which map to chromosome 5q13. Homozygous absence of the telomeric copy, SMN1, correlates with the development of spinal muscular atrophy. SMN2, on the other hand, appears to modify disease severity in a dose-dependent manner (*15*). SMN1 and SMN2 differ by at least five nucleotides, none of which affects protein coding. However, a C→T transition at position +6 of exon 7 (C6T) affects splicing, resulting in the inefficient inclusion of exon 7 in SMN2 mRNAs (*15*) (Fig. 2A). To capture this phenomenon, we amplified SMN1 and SMN2 polonies from cDNA prepared from an Ataxia-Telangiectasia cell line (Fig. 2B). The sequencing primers were designed to hybridize immediately 5′ to the C6T and G254A positions, such that incorporation of a specific fluorescent nucleotide upon SBEs queries bases that are polymorphic between the paralogous copies of SMN (*10*). The C6T polymorphism served to identify the subset of polonies in which exon 7 had been skipped, whereas the G254A polymorphism served as a reference to locate the full set of SMN1- or SMN2-derived polonies. As expected, SMN1 polonies (green) were colocalized in both scans with no exceptions, indicating that exon 7 was never skipped, whereas ~31% of the SMN2 polonies (red) were not present in the C6T image as compared to G254A scan, indicating exon 7 skipping (Fig. 2, B and C). Though demonstrated here in an extreme example, this approach can conceivably be adapted to study more subtle SNC-induced variation in alternative splicing, as well as posttranscriptional modifications such as RNA editing (*16*).

Many loci that undergo complex, combinatorial splicing encode cell-surface receptors, whose diversity could have pronounced consequences for cellular behavior and developmental specification. The CD44 gene, which maps to human chromosome 11p13, is one of the best-studied examples, encoding cell-surface glycoproteins involved in cell-cell contacts, cell-matrix interactions, and signal transduction (*17*). The extracellular proximal domain of CD44, involved in ligand binding, is encoded by 10 alternative spliced exons, v1 through v10, arranged in a central tandem array (Fig. 3A). Unrestricted combinatorial splicing of these exons could produce as many as 1024 isoforms (2^{10}). Specific CD44 isoforms have been evaluated as diagnostic and prognostic markers for cancer, but the results have been variable (*18*).

We speculated that the semiquantitative nature of RT-PCR and antibody-based assays, along with an inability to study exon combinatorics, might contribute to the inconsistencies in the literature. We therefore applied digital polony exon profiling to monitor transformation-induced CD44 splicing changes in a murine mammary epithelial cell line, Eph4. The transformed Eph4 cell line (Eph4βDD), which stably expresses constitutively activated MEK-1 (mitogen-activated protein kinase kinase 1), displays invasive metastatic growth

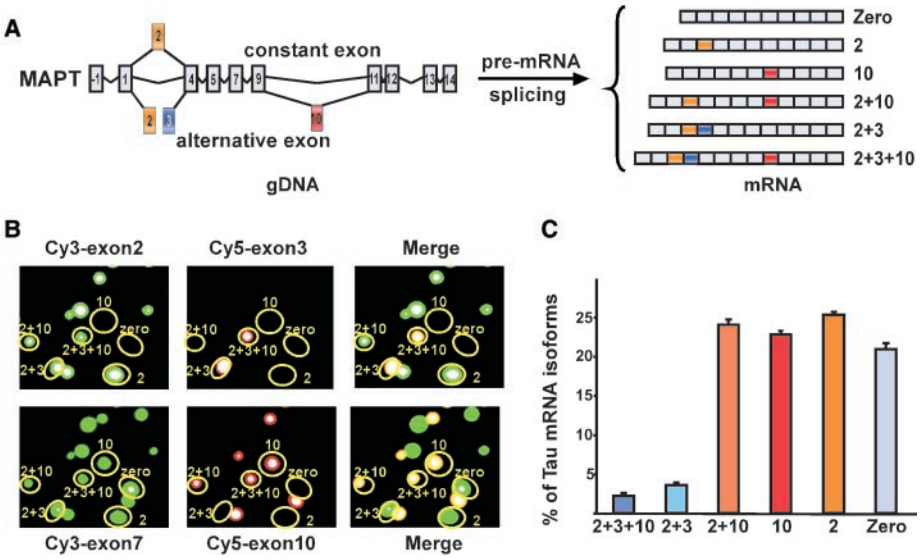

Fig. 1. Polony exon profiling of the MAPT locus. (**A**) Scheme of brain-specific MAPT alternative splicing. Constant exons are shown in light blue, whereas alternatively spliced exon 2, 3, and 10 are shown in orange, blue, and red, respectively. All six possible alternative splicing isoforms are indicated. (**B**) Individual and merged images were obtained by hybridizing exon 2–, 3–, 7–, or 10–specific probes to the same gel. Six MAPT isoforms were identified as indicated. The white center shown in some of the polonies is because of signal saturation. (**C**) Quantification of MAPT variants based on triplex experiments (total polony number $n = 723$).

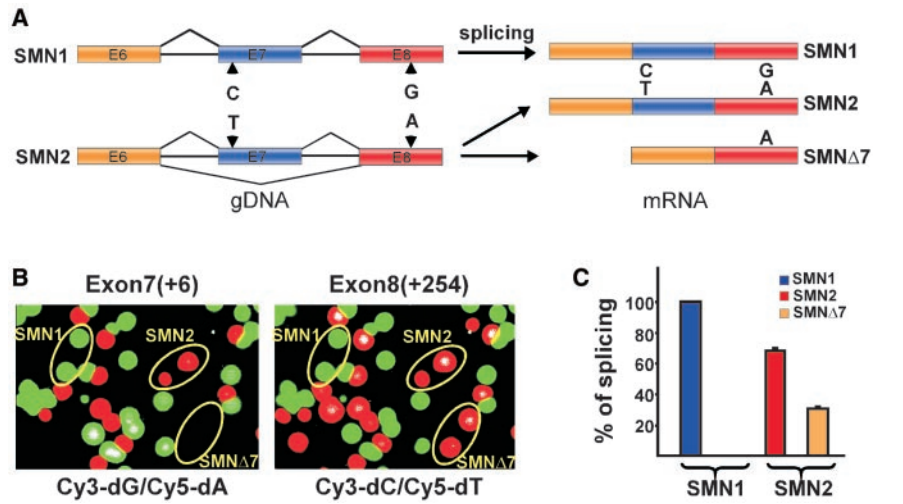

Fig. 2. SNC-related alternative splicing in SMN genes. (**A**) Scheme of SMN1 and SMN2 splicing. Two SNCs, exon 7 (+6, blue) and exon 8 (+254, red), are indicated in the context of SMN genes. (**B**) SBE with exon 7 (+6) and exon 8 (+254) sequencing primers. Examples of three SMN species are indicated in the circles. Fluorescent nucleotides used in each SBE are shown at the bottom of images (Cy3, red; Cy5, green). (**C**) Quantification of SMN1 and SMN2 isoforms based on triplex experiments (total polony number $n = 930$).

properties in vitro and in mice (*19*). Gels containing cDNA derived from Eph4 or Eph4βDD cell lines were amplified in quadruplicate with the use of a constant primer set that targets the full extracellular variable region (Fig. 3A), and the exon composition of polonies was determined by sequential SBEs. Each slide underwent five cycles of primer annealing, SBE, scanning, denaturation and scanning with two exon-specific sequencing primers in each cycle. Image-processing algorithms were used to align images, define polony locations, and identify patterns of exon inclusion or exclusion within individual polonies (Fig. 3, B and C) (*10*).

The multiplexed nature of digital polony exon profiling permits the identification of rare isoforms. Although only ~9000 polonies were queried in this study, the potential exists to detect a single rare isoform in a pool of 10^7 amplified transcripts (*7*). Indeed, we detected 69 distinct mouse CD44 isoforms [Supporting Online Material (SOM) Text], which is more than twice the number previously identified (*20*). Some exon combinations were more common than others, with the 15 most abundant isoforms accounting for ~95% of all CD44 transcripts that contained at least 1 of the 10 extracellular variable exons (table S1). In agreement with previous reports (*21*, *22*), inclusion of variable exons tracks in the 3′ to 5′ direction, such that an isoform containing any given variable exon is highly likely to contain all variable exons 3′ to it (table S1). The primary exceptions involve variable exon 6, which is skipped via joining of v5 and v7, as shown in 4 of the 15 most common isoforms (table S1).

Comparison of normal and transformed cells revealed changes in CD44 splicing profiles. The two most significant changes in transformed cells were a ~1.73-fold down-regulation of the v6-10 isoform ($P < 1 \times 10^{-31}$) and a ~2.88-fold up-regulation of the v4-5/7-10 isoform ($P < 1 \times 10^{-10}$), as the apparent consequences of a decrease in exon 6 inclusion. Supporting this observation, several other v6-skipping isoforms, including some rare combinations, were more abundant in transformed cells (table S1). The increase in the frequency of v5-7 splicing events in Eph4βDD cells was confirmed with the use of RT-PCR (fig. S2). However, not every isoform bearing v6 was down-regulated. For example, v3-10 showed a ~1.3-fold increase, whereas no change was observed for v4-10 (table S1). Thus, the overall down-regulation of v6 is ~1.3-fold ($P < 1 \times 10^{-16}$) (table S2). We also detected statistically significant changes in the overall inclusion frequencies of exons other than v6 (table S2). All changes in overall exon frequencies are relatively modest and would likely be missed by a nondigital assay. Consistent with our results, down-regulation of CD44 v6 expression has been documented upon malignant transformation of a murine epithelial cell (*23*). Our results support CD44 isoforms as potential markers of tumorigenesis and suggest that a digital approach to exon and isoform quantitation may be well suited for monitoring subtle changes in splicing regulation.

Polony technology provides a comprehensive, digital assay for quantifying complex populations of mRNA isoforms and may facilitate our ability to understand the role of alternative splicing in development and disease. This technology is capable of monitoring the cis relationship of multiple alternative spliced exons within the same transcripts, a prerequisite for a complete "parts list" of mammalian biology. Polonies have also been applied to genotyping, long-range haplotyping, and fluorescent in situ sequencing (FISSEQ) (*7*, *13*). Though the current throughput of polony-based exon profiling is limited as compared to microarray-based splicing assays (*24*, *25*), the potential exists to achieve higher levels of multiplexing via quantum-dot labeled probes and/or FISSEQ. We expect that detailed gene expression information, including data on posttranscriptional events such as alternative pre-mRNA splicing, will shed light on the modeling and engineering of functional complexity and plasticity in multicellular organisms.

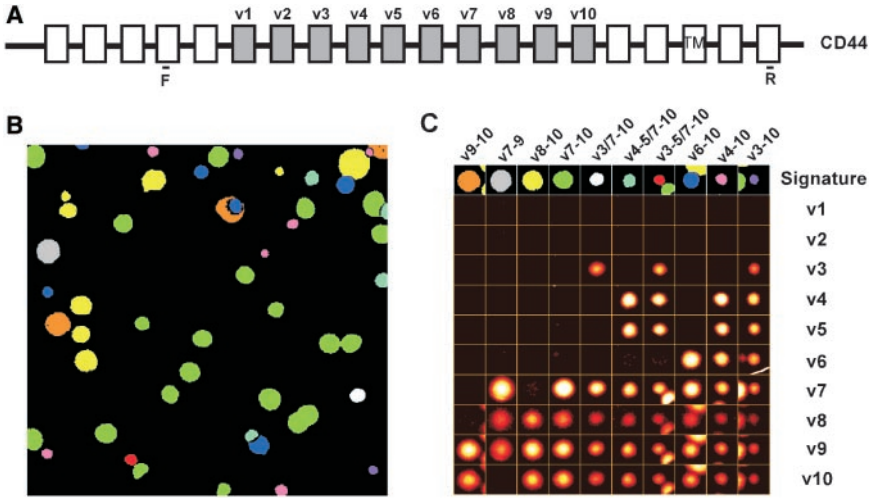

Fig. 3. Alternative splicing profile of mouse CD44 gene. (**A**) Schematic of the murine CD44 locus. Alternatively spliced exons, v1 through v10, are shown in gray. Positions of amplification primers used here (F and R) are as indicated. TM, transmembrane domain. (**B**) Images obtained via probing for each of the 10 individual alternative exons were aligned at single pixel resolution and processed to detect polonies and determine exon composition signatures. Pseudocolors are assigned to unique mRNA isoforms with specific compositions of variable exons. (**C**) Distinct polonies shown in (B) with corresponding regions from each background-subtracted signal image. Four of the profiles each contain a portion of a second, adjacent polony in the image: v9-10 (+ v8-10), v3-5/7-10 (+ v7-10), v6-10 (+ v8-10), and v3-10 (+ v7-10).

References and Notes

1. D. L. Black, *Cell* **103**, 367 (2000).
2. Z. Kan, E. C. Rouchka, W. R. Gish, D. J. States, *Genome Res.* **11**, 889 (2001).
3. M. Missler, T. C. Sudhof, *Trends Genet.* **14**, 20 (1998).
4. D. Schmucker *et al.*, *Cell* **101**, 671 (2000).
5. T. Maniatis, B. Tasic, *Nature* **418**, 236 (2002).
6. R. D. Mitra, G. M. Church, *Nucleic Acids Res.* **27**, e34 (1999).
7. R. D. Mitra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5926 (2003).
8. B. Vogelstein, K. W. Kinzler, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9236 (1999).
9. A. Delacourte, L. Buee, *Curr. Opin. Neurol.* **13**, 371 (2000).
10. Materials and methods are available as supporting material on *Science* Online.
11. N. Sergeant *et al.*, *Hum. Mol. Genet.* **10**, 2143 (2001).
12. C. R. Valentine, *Mutat. Res.* **411**, 87 (1998).
13. R. D. Mitra *et al.*, *Anal. Biochem.*, in press.
14. L. Cartegni, S. L. Chew, A. R. Krainer, *Nature Rev. Genet.* **3**, 285 (2002).
15. C. L. Lorson, E. Hahnen, E. J. Androphy, B. Wirth, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6307 (1999).
16. B. L. Bass, *Annu. Rev. Biochem.* **71**, 817 (2002).
17. J. Lesley, R. Hyman, *Front. Biosci.* **3**, D616 (1998).
18. P. Herrlich *et al.*, *Ann. N. Y. Acad. Sci.* **910**, 106 (discussion 118) (2000).
19. J. Pinkas, P. Leder, *Cancer Res.* **62**, 4781 (2002).
20. J. Bajorath, *Proteins* **39**, 103 (2000).
21. M. V. Bell, A. E. Cowper, M. P. Lefranc, J. I. Bell, G. R. Screaton, *Mol. Cell. Biol.* **18**, 5930 (1998).
22. X. Roca *et al.*, *Am. J. Pathol.* **153**, 183 (1998).
23. T. Soukka *et al.*, *Cancer Res.* **57**, 2281 (1997).
24. J. M. Yeakley *et al.*, *Nature Biotechnol.* **20**, 353 (2002).
25. T. A. Clark, C. W. Sugnet, M. Ares Jr., *Science* **296**, 907 (2002).

26. We thank J. Pinkos and P. Leder for the Eph4 and Eph4βDD cell lines and A. Dudley and S. Munroe for critical comments on the manuscript. Supported by the U.S. Department of Energy and NIH Glue grant 5U54GM62119. The polony technology is covered by U.S. Patent 6,432,360.

Supporting Online Material

www.sciencemag.org/cgi/content/full/301/5634/836/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 and S2

16 April 2003; accepted 20 June 2003