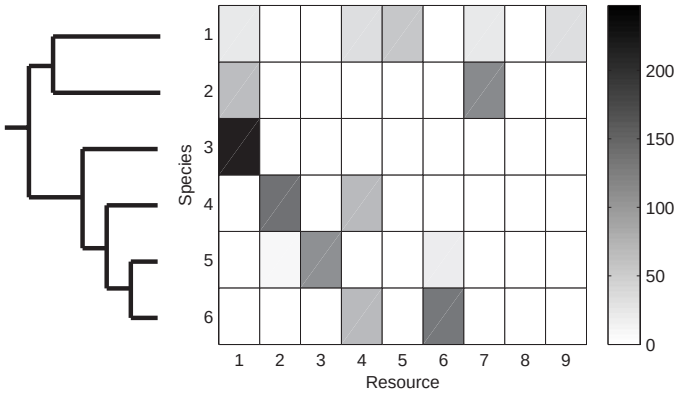

Fig. 4. Matrix of resource usage by the species shown in Fig. 3. Each entry is shaded according to the number of times a particular resource is used during the life cycle of an average member of the species. Although there is some overlap in resource use, each species dominates on at least one resource. The branching pattern shown to the left of the usage matrix indicates the relation between the species based on the phylogeny in Fig. 3 (relative branch lengths not to scale).

simultaneously outcompete two existing species, thereby reducing the total number of species in the community, although species richness often subsequently rebounds. Thus, the species richness in these digital communities reflects a dynamic steady state rather than some optimal end point.

Moving from a phylogenetic perspective to one based on ecological performance, Fig. 4 displays the resource utilization patterns of the same community of species illustrated by the phylogenetic depth profile in Fig. 3. Resource usage levels are averaged over all organisms that belong to a given species based on the clustering algorithm, which uses only phylogenetic data (not ecological phenotypes). The six species are all clearly phenotypically distinct. Moreover, each species consumes a larger share of at least one resource than does any other species. Species 3 is specialized on the use of resource 1, whereas the other species use two or more resources. Species 1 has the most generalized phenotype, using five resources including two that it dominates (resources 5 and 9). The precise partitioning of resources varies from one replicate experiment to another and does not follow any obvious rule. We have also demonstrated that our speciation results are not dependent on the particular ancestral organism we have used and that the pattern of species richness is not substantially altered when species that evolved at other resource levels are introduced into a community [Supporting Online Material (SOM) Text].

In summary, our experiments demonstrate an evolutionary effect of productivity on species richness, with adaptive radiation leading to maximum diversity at intermediate resource inflows in communities with fixed total population size. The decline in species richness at high productivity occurs because selection shifts from favoring exploitation of unused resources to favoring maximum replication rate when resources are superabundant. Importantly, this pattern does not require spatial heterogeneity or predation, although these and other ecological factors may augment or oppose the adaptive radiation.

References and Notes

1. P. J. Morin, *Nature* **406**, 463 (2000).
2. M. L. Rosenzweig, Z. Abramsky, in *Species Diversity in Ecological Communities*, R. E. Ricklefs, D. Schluter, Eds. (Univ. Chicago Press, Chicago, IL, 1993), pp. 52–56.
3. D. Tilman, S. Pacala, in *Species Diversity in Ecological Communities*, R. E. Ricklefs, D. Schluter, Eds. (Univ. Chicago Press, Chicago, IL, 1993), pp. 13–25.
4. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
5. R. B. Waide *et al*., *Annu. Rev. Ecol. Syst.* **30**, 257 (1999).
6. G. G. Mittelbach *et al*., *Ecology* **82**, 2381 (2001).
7. T. Fukami, P. J. Morin, *Nature* **424**, 423 (2003).
8. M. C. Horner-Devine, M. A. Leibold, V. H. Smith, B. J. M. Bohannan, *Ecol. Lett.* **6**, 613 (2003).
9. R. Kassen, A. Buckling, G. Bell, P. B. Rainey, *Nature* **406**, 508 (2000).
10. D. Tilman, *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, NJ, 1982).
11. R. D. Holt, J. Grover, D. Tilman, *Am. Nat.* **144**, 741 (1994).
12. P. A. Abrams, *Ecology* **76**, 2019 (1995).
13. M. A. Leibold, *Am. Nat.* **147**, 784 (1996).
14. R. B. Helling, C. N. Vargas, J. Adams, *Genetics* **116**, 349 (1987).
15. R. F. Rosenzweig, R. R. Sharp, D. S. Treves, J. Adams, *Genetics* **137**, 903 (1994).
16. P. E. Turner, V. Souza, R. E. Lenski, *Ecology* **77**, 2119 (1996).
17. D. S. Treves, S. Manning, J. Adams, *Mol. Biol. Evol.* **15**, 789 (1998).
18. D. E. Rozen, R. E. Lenski, *Am. Nat.* **155**, 24 (2000).
19. P. B. Rainey, M. Travisano, *Nature* **394**, 69 (1998).
20. M. Doebeli, U. Dieckmann, *Am. Nat.* **156**, S77 (2000).
21. M. L. Friesen, G. Saxer, M. Travisano, M. Doebeli, *Evolution* **58**, 245 (2004).
22. F. M. Cohan, *Syst. Biol.* **50**, 513 (2001).
23. C. O. Wilke, C. Adami, *Trends Ecol. Evol.* **17**, 528 (2002).
24. C. Ofria, C. O. Wilke, *Artif. Life* **10**, 191 (2004).
25. R. E. Lenski, C. Ofria, R. T. Pennock, C. Adami, *Nature* **423**, 139 (2003).
26. T. F. Cooper, C. Ofria, in *Artificial Life VIII: Proceedings of the Eighth International Conference on Artificial Life*, International Society for Artificial Life, Sydney, Australia, 9 to 13 December 2002, R. Standish, M. A. Bedau, H. A. Abbass, Eds. (MIT Press, Cambridge, MA, 2003), p. 227.
27. Materials and methods are available as supporting material on *Science* Online.
28. E. Zuckerkandl, L. Pauling, in *Evolving Genes and Proteins*, V. Bryson, H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 97–166.
29. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).

30. Supported by grant DEB-9981397 from NSF.

Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5680/84/DC1
Materials and Methods
SOM Text

2 February 2004; accepted 3 June 2004

Self-Assembling Protein Microarrays

Niroshan Ramachandran,[1] Eugenie Hainsworth,[1,2] Bhupinder Bhullar,[1] Samuel Eisenstein,[1] Benjamin Rosen,[1] Albert Y. Lau,[1] Johannes C. Walter,[3] Joshua LaBaer[1]*

Protein microarrays provide a powerful tool for the study of protein function. However, they are not widely used, in part because of the challenges in producing proteins to spot on the arrays. We generated protein microarrays by printing complementary DNAs onto glass slides and then translating target proteins with mammalian reticulocyte lysate. Epitope tags fused to the proteins allowed them to be immobilized in situ. This obviated the need to purify proteins, avoided protein stability problems during storage, and captured sufficient protein for functional studies. We used the technology to map pairwise interactions among 29 human DNA replication initiation proteins, recapitulate the regulation of Cdt1 binding to select replication proteins, and map its geminin-binding domain.

To exploit the growing number of expression-ready cDNA clone collections, high-throughput (HT) methods to study protein function are needed (*1*–*5*). The development of protein microarrays offers one compelling approach (*6*–*8*). Protein microarrays are currently available in two general formats. Antibody arrays contain an array of antibodies that measure the abundance of specific proteins (or other molecules) in samples (*9*). Our work focuses on target protein arrays, which present arrayed proteins of interest. They can be used to examine target protein interactions with other molecules, such as drugs, antibodies, nucleic acids, lipids, or other proteins. In addition, the arrays can be interrogated to find substrates for enzymes (*10*, *11*).

The current approach to generate target protein microarrays is to produce proteins separately and then spot them on the arrays with the use of a variety of linkage chemistries (*6*–*8*, *12*). Despite these demonstrations of feasibility, target protein microarrays have not been widely adopted. In part, this may be due to the labor and technical issues associated with HT protein production. Challenges remain to find HT expression systems for mammalian proteins with good yield and purity, and under conditions conducive to functional protein studies. Moreover, once isolated, there are concerns regarding protein stability during storage, either before or after spotting on the array.

Building upon the successful use of in vitro translated protein in standard scale applications (*13*–*15*), our approach substitutes the use of purified proteins with the use of cDNAs encoding the target proteins at each feature of the microarray. The proteins are transcribed and translated by a cell-free system and immobilized in situ by means of epitope tags fused to the proteins. This approach eliminates the need to express and purify proteins separately and produces proteins at the time of the assay, abrogating concerns about protein stability during storage. Mammalian proteins can be expressed in a mammalian milieu, providing access to vast collections of cloned cDNAs.

With a nucleic acid programmable protein array (NAPPA), we aimed to exploit this biochemical strategy and enable proteome-scale experiments. This required a high-density format that minimized the use of cell-free extract, and for convenience and accessibility, a readily available matrix (such as standard glass microscope slides) that did not require specially micromachined wells (*10*) and that used existing technology for printing and reading DNA microarrays.

Through testing a variety of cDNA printing schemes, we found that an optimum balance was required between binding DNA efficiently and maintaining a DNA conformation that supported efficient transcription and translation (fig. S1). The most efficient strategy coupled a psoralen-biotin conjugate to the expression plasmid DNA with the use of ultraviolet light, which was then captured on the surface by avidin (Fig. 1).

The addition of a C-terminal glutathione *S*-transferase (GST) tag to each protein enabled its capture to the array through an antibody to GST printed simultaneously with the expression plasmid (fig. S2). Other protein fusion tags and capture molecules can be easily substituted for the GST fusion and antibodies to GST used here (*16*). The resulting array was dried and stored at room temperature.

[1]Harvard Institute of Proteomics, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 320 Charles Street, Cambridge, MA 02141, USA. [2]Technology and Engineering Center, [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: josh@hms.harvard.edu

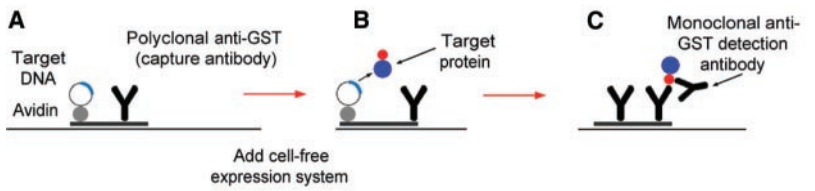

Fig. 1. NAPPA approach. Biotinylation of DNA: Plasmid DNA is cross-linked to a psoralen-biotin conjugate with the use of ultraviolet light (*17*). (**A**) Printing the array. Avidin (1.5 mg/ml, Cortex), polyclonal GST antibody (50 μg/ml, Amersham), and Bis (sulfosuccinimidyl) suberate (2 mM, Pierce) are added to the biotinylated plasmid DNA. Samples are arrayed onto glass slides treated with 2% 3-aminopropyltriethoxysilane (Pierce) and 2 mM dimethyl suberimidate.2HCl (Pierce). (**B**) In situ expression and immobilization. Microarrays were incubated with 100 μl per slide rabbit reticulocyte lysate with T7 polymerase (Promega) at 30°C for 1.5 hours then 15°C for 2 hours in a programmable chilling incubator (Torrey Pines). (**C**) Detection. Target proteins are expressed with a C-terminal GST tag and immobilized by the polyclonal GST antibody. All target proteins are detected using a monoclonal antibody to GST (Cell Signaling Technology) against the C-terminal tag confirming expression of full-length protein.

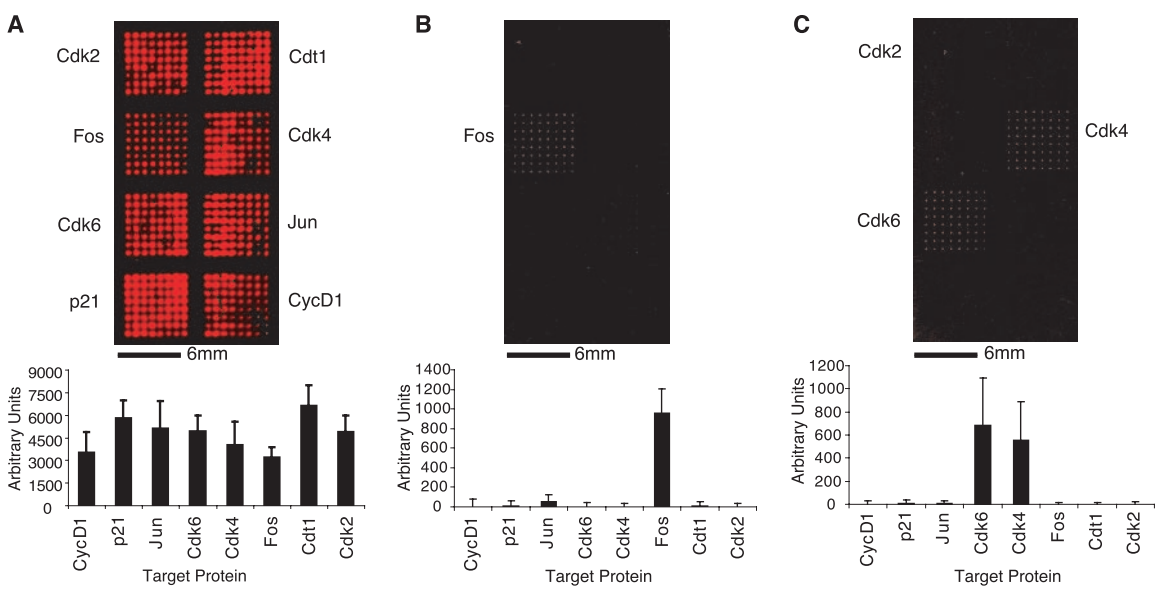

Fig. 2. Expression of target proteins and detection of protein interactions on a NAPPA microarray format. (**A**) Eight target plasmid DNAs encoding C-terminal GST fusion proteins in pANT7_cGST (fig. S2) were immobilized onto the glass slide at a density of 512 spots per slide (900-μm spacing). The target proteins were expressed with 100 μl rabbit reticulocyte lysate supplemented with T7 polymerase. Signals were detected with antibody to GST and tyramide signal amplification (TSA) reagent (PerkinElmer). To verify that the detected proteins were the expected target proteins, and to confirm that there was no cross-talk across the slide, we used target protein–specific antibodies, which detected only their relevant spots (fig. S3). (**B** and **C**) The eight genes were queried for potential interactors with (B) Jun and (C) p16. Query DNA encoding an N-terminal HA tag was added to the reticulocyte lysate before expressing the target proteins (fig. S2). Target and query proteins were coexpressed and the interaction was detected with an antibody to HA (12CA5). The bar graphs show average intensity (+SD) from 64 samples for each interaction. Images were quantified using ScanAlyze software (Michael Eisen, Lawrence Berkeley National Laboratory, CA). The signals were corrected for local background.

To activate and use the array, a cell-free coupled transcription and translation system (such as reticulocyte lysate containing T7 polymerase) was added as a single continuous layer (not discrete spots) covering the arrayed cDNAs on the microscope slide. To test the system, expression plasmids encoding eight genes were immobilized onto an array at a density of 512 spots per slide (900-μm spacing). Expression of target protein was confirmed with an antibody to GST (different from the capture GST antibody), and the signals were measured with a standard glass slide DNA-microarray scanner (Fig. 2A and fig. S3). We observed an easily detectable signal for all proteins [average signal-to-noise ratio (±SD) = 53 ± 14], demonstrating that 100 μl of reticulocyte lysate is sufficient to support protein expression in all 512 spots of the array simultaneously (*17*). Not surprisingly, there was modest variation in protein expression from gene to gene (coefficient of variation = ~24%). We have subsequently found that these differences can often be corrected by adjusting the amount of printed plasmid template. By comparing signal intensities to control spots containing purified GST, we estimate that about 10 fmol (~675 pg) of protein were produced and captured at each spot, which compares favorably to existing methods (*8*).

NAPPA is well suited to the detection of protein-protein interactions because both the target proteins (bound to the array) and the query protein (used to probe the array) can be transcribed and translated in the same extract. As validation, the query protein Jun was tagged with a hemagglutinin (HA) epitope and coexpressed with the target proteins (Fig. 2B). The interaction was visualized with an antibody to HA, which revealed that Jun query protein bound to the Fos target [dissociation constant K_d ~ 50 nM (*12*)]. To determine if binding selectivity is preserved, we tested the Cdk inhibitor p16, which binds selectively to Cdk4 and Cdk6 but not the closely related Cdk2. As shown in Fig. 2C, this specificity was recapitulated with NAPPA.

To apply NAPPA to a biological question, we studied the human DNA replication complex. Experiments in yeast, *Xenopus*, and human cells have led to a detailed model for the initiation of eukaryotic DNA replication. Origins of replication are "licensed" in the G_1 phase of the cell cycle when the origin replication complex (ORC) recruits the initiation factors, Cdt1 and Cdc6, and the minichromosome maintenance complex (MCM2-7) to form the prereplication complex (pre-RC). In S phase, the pre-RC is converted into an active replication fork by the protein kinases Cdc7 and Cdk2, a process that involves origin binding of at least two additional initiation factors, MCM10 and Cdc45, leading to DNA synthesis (*18*).

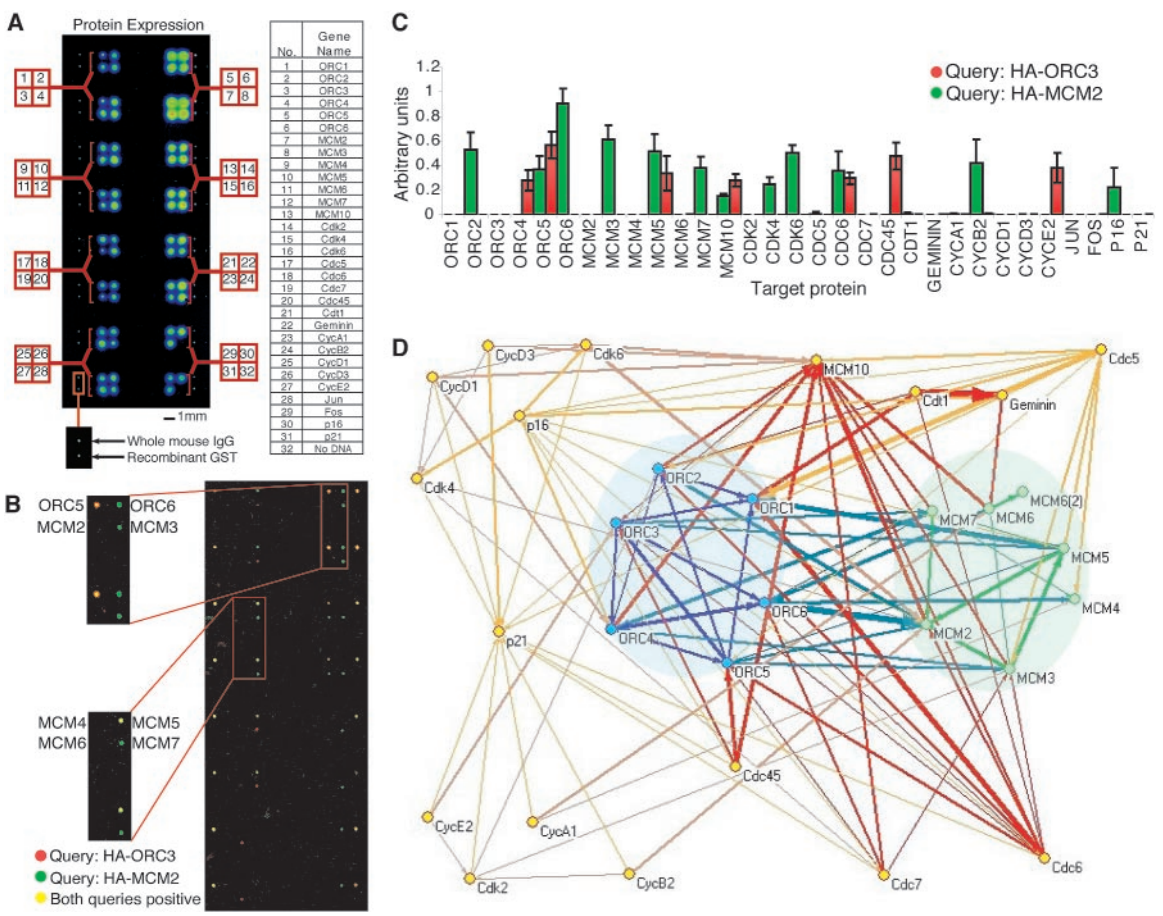

Fig. 3. Expression of human DNA replication proteins and interaction mapping. (**A**) Target DNAs representing 29 human DNA replication proteins and two positive controls were immobilized and expressed on the array in duplicate. The legend (right panel) lists all genes expressed on the array. Expression of all target proteins was confirmed by an antibody to GST (left panel). Two protein registration markers, purified recombinant GST (22 μg/ml, Sigma) and whole-mouse immunoglobulin G (IgG) (550 μg/ml, Pierce), were also printed as registration spots and were able to monitor protein expression and slide variation (inset, bottom). (**B**) Replicate slides from (A) were probed with each member of the DNA replication proteins expressed as HA-tagged query proteins, repeating each query protein on two slides. (B) is a superposition of images of slides probed with HA-ORC3 (red) and HA-MCM2 (green); target spots that interacted with both queries appear as yellow. Interactions were detected with an antibody to HA and quantified with ScanAlyze. (**C**) The signals for all interactions, including (B), were calculated by subtracting local background and then standardized with the intensity of whole-mouse IgG registration marker. Interactions were considered positive when the signal was greater than three times the standard deviation of the background for all instances of the interaction. The bar graph shows signal intensity for interactions with ORC3 (red) and MCM2 (green) shown in (B). (**D**) Interaction map shows interactions among the ORC and MCM complex in blue (lines and shaded oval) and green (lines and shaded oval), respectively. Intercomplex interactions are shown in dark blue. Interactions with proteins involved in the formation of pre-RC and preinitiation complex are shown in red and additional regulatory proteins are shown in brown. All other interactions are shown in orange. The arrows of the connector show the direction (from target to query) of the interaction and the weight given to the connector depicts the strength of the signal.

Sequence-verified human genes for 29 proteins involved in DNA replication initiation (in addition to Fos and Jun as positive controls) were immobilized and expressed on NAPPA (Fig. 3A). Signals were readily detected for all of the target proteins, showed high reproducibility between duplicates, and ranged from 270 pg (4 fmols) to 2600 pg (29 fmols), a sevenfold range that falls well within the range observed in protein-spotting protein microarrays [10 to 950 pg (*8*)]. Each of the 29 DNA replication proteins was used as a query to probe a pair of duplicate arrays to generate a 29 × 29 protein interaction matrix. Examples of the interaction data are shown in Fig. 3, B and C.

We found 110 interactions among the proteins in the replication complex, averaging 7.7 interactions per protein (range of 3 to 16; Fig. 3D and table S1). Detected interactions included 47 previously identified by any method including genetic, two-hybrid, and biochemical interactions (based on our literature survey), and 63 previously undetected interactions. Of the gold-standard interactions that had been demonstrated biochemically with purified proteins, we detected 17 of 20 (85%) (*18*–*20*). We also detected 19 of the 36 reported interactions (53%) on the basis of coimmunoprecipitation (IP) (*18*, *21*–*23*). A difference here is expected because NAPPA only detects binary interactions, whereas IP also reports interactions mediated by bridging proteins. In fact, a NAPPA network in which two proteins shared a common binding partner could be identified for each of the 17 IP interactions not detected by NAPPA. Overlap was lowest (42%) with interactions reported by yeast two-hybrid (*18*, *20*, *22*, *24*).

A variety of biochemical experiments have identified two stable complexes, ORC and MCM2-7, in the pre-RC of many species (*18*). Consistent with this, the microarray experiments detected many interactions (28% of all detected interactions) within and between these two complexes (Fig. 3D) including 10 unique interactions among the six ORC subunits (Fig. 3D, blue) in agreement with the current ORC model (*25*). Similarly, we observed most known interactions within the MCM complex except those involving MCM6, which was among the proteins that showed low expression as both the target and query (Fig. 3D, green). Interactions among Cdc6, Cdt1, and the ORC proteins required for pre-RC formation were not previously understood. Here, we find that Cdc6 interacts directly with all of the ORC proteins except ORC4 and Cdt1 interacts specifically with ORC1 and ORC2 (Fig. 3D, red).

In the S phase, the loading of Cdc45 to the chromatin is postulated to activate the helicase activity of the bound MCM2-7 complex (*26*, *27*). Interestingly, we did not observe any direct interactions between Cdc45 and the MCM2-7 proteins. Cdc45 interacted with MCM10, which in turn interacted with several MCM2-7 proteins (Fig. 3D, red), suggesting that MCM10 could act to recruit Cdc45 to the MCM2-7 complex. This is consistent with recent experiments showing that MCM10 is indeed required for Cdc45 binding to chromatin (*28*).

Cdc6 and Cdt1 are both necessary to recruit the MCM2-7 complex onto chromatin (*18*). We detected many interactions among these proteins but none between Cdt1 and the MCM2-7 proteins, although they coimmunoprecipitate (*29*, *30*). Cdt1 and MCM2 share Cdc6 as a binding partner (fig. S4), suggesting that Cdc6 could bridge Cdt1 to the MCM2-7 complex. The open format of NAPPA supports the expression of proteins in addition to the target and query, allowing the examination of multiprotein complexes and their regulation. By exploiting this feature, we demonstrated that MCM2 bound to Cdt1 only in the presence of coexpressed Cdc6 (Fig. 4A). Thus, it is likely that Cdc6 acts as a bridging protein, although enzymatic or allosteric effects cannot be ruled out, showing that simple regulatory mechanisms can be recapitulated in the protein microarray format.

To further examine Cdt1 protein function, we focused on its interaction with geminin. Previous work has mapped the binding to a relatively large domain of Cdt1 [177 to 380 amino acids (aa) (*20*)]. We used NAPPA to map more precisely the binding domain of geminin on human Cdt1 by generating a series of end deletion fragments of Cdt1, expressing the partial length proteins on the array, and probing the array with HA-geminin as query protein (Fig. 4B). Using this approach, we localized a ~15-aa sequence (198 to 212 aa) that was necessary for binding. We then tested a 78–aa fragment (135 to 212 aa) containing this sequence and demonstrated that it was sufficient for geminin binding, albeit somewhat more weakly.

There still remain technical challenges for NAPPA. First, as in virtually all other HT protein interaction techniques, there is the possibility that bridging proteins or inhibitors (e.g., from the cell-free expression system) may affect some interactions. Second, the use of peptide tags, also common to most interaction methods, may lead to steric effects that block important binding domains, although with NAPPA tags can be configured on either end of the proteins. Third, some posttranslational modifications may be absent in NAPPA. However, the open format of NAPPA allows for the addition of enzymes or extracts if needed. Finally, because NAPPA lacks the spatial and temporal compartmentalization of the cell and because the folding and activity of proteins in vitro may not always reflect protein activity in vivo, it will be important to confirm that previously unidentified interactions make biological sense.

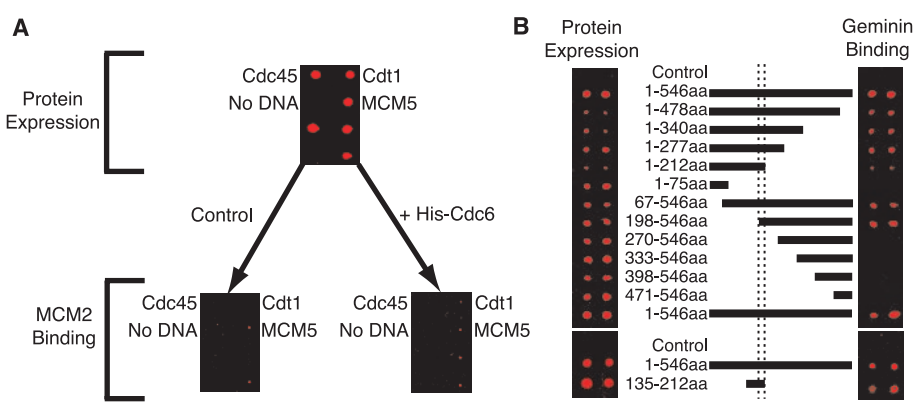

Fig. 4. Characterization of Cdt1. (**A**) Cdt1 regulation. NAPPA was used to test whether Cdc6 could act as a bridging protein between Cdt1 and MCM2. Target proteins Cdt1, Cdc45 (negative control), and MCM5 (positive control) were expressed in duplicate (top panel) and confirmed by an antibody to GST. The target proteins were probed with either HA-MCM2 alone (left panel) or in the presence of coexpressed His-Cdc6 (right panel). The binding of MCM2 was detected with an antibody to HA. (**B**) Cdt1 deletion mapping. NAPPA was used to map the binding domain of geminin on Cdt1. Fragments from various regions of Cdt1 (as indicated) were generated by PCR and cloned into target expression vectors. The partial or full-length polypeptides were expressed and detected on the array with an antibody to GST (left panel). To identify the binding region of geminin, the array was queried with HA-geminin (right panel) and developed with an antibody to HA. Vertical lines (dashed) delimit the small (15-aa) region common to all the fragments that bound geminin. A small 78-aa domain (135 to 212 aa) containing the noted 15 aa (198 to 212 aa) was expressed along with full-length Cdt1 (bottom left), which was again queried with geminin (bottom right).

References and Notes

1. A. Walhout *et al.*, *Methods Enzymol.* **328**, 575 (2000).
2. L. Brizuela, P. Braun, J. LaBaer, *Mol. Biochem. Parasitol.* **118**, 155 (2001).
3. P. Braun *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2654 (2002).
4. J. Reboul *et al.*, *Nature Genet.* **34**, 35 (2003).

5. J. LaBaer, G. Marsischky, in *Proteome Analysis—Interpreting the Genome*, D. W. Speicher, Ed. (Elsevier, Philadelphia, PA, 2004), pp. 287–301.
6. A. Lueking *et al.*, *Anal. Biochem.* **270**, 103 (1999).
7. G. MacBeath, S. Schreiber, *Science* **289**, 1760 (2000).
8. H. Zhu *et al.*, *Science* **293**, 2101 (2001).
9. B. Haab, M. Dunham, P. Brown, *Genome Biol.* **2** (2001).
10. D. J. Cahill, E. Nordhoff, *Adv. Biochem. Eng. Biotechnol.* **83**, 177 (2003).
11. G. Jona, M. Snyder, *Curr. Opin. Mol. Ther.* **5**, 271 (2003).
12. J. R. Newman, A. E. Keating, *Science* **300**, 2097 (2003).
13. K. D. Lustig *et al.*, *Methods Enzymol.* **283**, 83 (1997).
14. V. R. Rao *et al.*, *J. Biol. Chem.* **274**, 37893 (1999).
15. M. He, M. Taussig, *Nucleic Acids Res.* **29**, E73 (2001).
16. N. Ramachandran, data not shown.
17. Materials and methods are available as supporting material on *Science* Online.
18. S. P. Bell, A. Dutta, *Annu. Rev. Biochem.* **71**, 333 (2002).
19. S. K. Dhar, L. Delmolino, A. Dutta, *J. Biol. Chem.* **276**, 29067 (2001).
20. K. Yanagi, T. Mizuno, Z. You, F. Hanaoka, *J. Biol. Chem.* **277**, 40871 (2002).
21. S. Vashee, P. Simancek, M. D. Challberg, T. J. Kelly, *J. Biol. Chem.* **276**, 26666 (2001).
22. M. Kneissl, V. Putter, A. A. Szalay, F. Grummt, *J. Mol. Biol.* **327**, 111 (2003).
23. T. W. Christensen, B. K. Tye, *Mol. Biol. Cell* **14**, 2206 (2003).
24. A. M. Merchant, Y. Kawasaki, Y. Chen, M. Lei, B. K. Tye, *Mol. Cell. Biol.* **17**, 3261 (1997).
25. The Genome Knowledge Base Web site displays curated information regarding DNA replication and other biological pathways and is available at www.genomeknowledge.org and www.genomeknowledge.org/cgi-bin/eventbrowser?DB=gk_current&ID=68616.
26. L. Zou, B. Stillman, *Science* **280**, 593 (1998).
27. T. Masuda, S. Mimura, H. Takisawa, *Genes Cells* **8**, 145 (2003).
28. J. A. Wohlschlegel, S. K. Dhar, T. A. Prokhorova, A. Dutta, J. C. Walter, *Mol. Cell* **9**, 233 (2002).
29. P. J. Gillespie, A. Li, J. J. Blow, *BMC Biochem.* **2**, 15 (2001).
30. J. G. Cook, D. A. Chasse, J. R. Nevins, *J. Biol. Chem.* **279**, 9625 (2003).
31. We thank the following labs for allowing us to use their arrayers and scanners: J. Mekalanos's lab at Harvard Medical School (M. Dziejman, E. Balon, and D. Sturtevant); L. Kunkel's lab at Children's Hospital (J. Haslett, T. Burleson, and M. Han); R. Pratt's lab at Brigham and Women's Hospital (L. Anderson); and The Institute of Chemistry and Cell Biology (R. King and M. Narovlyansky). We thank K. Struhl for providing us with the antibody to HA; D. Moreira, A. Rolfs, P. Braun, and A. Tutter for their support; and E. Harlow for comments on the manuscript. Funded by NIH/National Cancer Institute grant for Functional Proteomics of Breast Cancer (R21 CA99191-01). Molecular interaction data have been deposited in the Biomolecular Interaction Network Database with accession codes 134165 to 134277.

Supporting Online Material
www.sciencemag.org/cgi/content/full/305/5680/86/DC1
Materials and Methods
Figs. S1 to S4
Table S1

8 March 2004; accepted 25 May 2004

Lack of a Fusion Requirement for Development of Bone Marrow–Derived Epithelia

Robert G. Harris,* Erica L. Herzog,* Emanuela M. Bruscia, Joanna E. Grove, John S. Van Arnam, Diane S. Krause†

Analysis of developmental plasticity of bone marrow–derived cells (BMDCs) is complicated by the possibility of cell-cell fusion. Here we demonstrate that epithelial cells can develop from BMDCs without cell-cell fusion. We use the Cre/lox system together with β-galactosidase and enhanced green fluorescent protein expression in transgenic mice to identify epithelial cells in the lung, liver, and skin that develop from BMDCs without cell fusion.

Cells marked as bone marrow–derived (BMDCs) can be found as mature cells of many nonhematopoietic tissues, including lung, liver, kidney, skin, and muscle. The BMDCs are often identified by the presence of the Y chromosome in sex-mismatched bone marrow transplants (male into female) or by detection of gene products such as enhanced green fluorescent protein (EGFP) that are present in the donor but not the recipient. Interpretation of these results has been complicated by observations that in vitro coculture of embryonic stem cells and somatic cells can result in spontaneous cell fusion (*1*, *2*), giving rise to cells of mixed phenotype and genotype. The in vivo appearance of marrow-derived hepatocytes, cardiomyocytes, and Purkinje cells is due, at least in part, to fusion of BMDCs with these cell types (*1*–*7*). However, because some of these cell types are known to form heterokaryons in settings of profound tissue injury (*8*–*10*), the incidence of this process must be examined in nonfusogenic organs under physiologic conditions.

This study was designed to evaluate for fusion events, including those that may have been masked by reductive division. We did this using the Cre/lox recombinase system to examine whether fusion occurs between BMDCs and host cells after bone marrow (BM) transplantation. We used mice of the Z/EG Cre-reporter strain (*11*, *12*) (fig. S1) as marrow donors for transplantation into mice that ubiquitously express Cre. In this model, any cell resulting from fusion of a BMDC with a host cell should express EGFP.

We transplanted lethally irradiated female mice that ubiquitously expressed Cre recombinase with BM from male Z/EG donor mice, β-actin–Cre (Cre) donor mice (*13*), or Z/EG and β-actin–Cre F1 (Z × C F1) donor mice. The transplants from Z × C F1 mice into Cre mice served as positive controls for EGFP expression in donor-derived cells, and the Cre-into-Cre transplants served as negative controls. Tissues from the recipients were analyzed 8 to 12 weeks after transplantation for the presence of BM-derived (Y chromosome–positive) epithelial cells and EGFP expression.

For analysis of the lungs, single-cell suspensions were analyzed by flow cytometry and fluorescence microscopy. For fluorescence-activated cell sorting (FACS), the M1 gate was determined by the fluorescence intensity greater than 99% of the cells in an EGFP-negative population (Fig. 1A, horizontal black line). Given this gate, 66% of all cells from the lungs of Z/EG × Cre F1 mice were EGFP-positive (Fig. 1A, green). The EGFP-negative population mostly comprised blood and endothelial cells (*14*). In the EGFP-positive control transplants (Fig. 1A, blue), 10% of the cells were EGFP-positive. Subsequent immunocytochemistry on FACS-sorted EGFP-positive cells showed that 6% of these were both EGFP- and cytokeratin-positive (Fig. 1, B and C). The high percentage of EGFP-positive, cytokeratin-negative cells was due to contaminating blood cells in the lung digests. From these data, we conclude that at least 0.6% of total lung cells were EGFP-positive epithelial cells. In experimental transplants from Z/EG mice into Cre mice (Fig. 1A, orange and red), ~0.8% of the experimental cells had a fluorescence intensity greater than the EGFP-negative baseline population, which is equal to the percentage in negative-control β-actin–Cre (Fig. 1A, black) animals. Immunofluorescence analysis indicated that these cells were uniformly negative for EGFP (*14*).

We also looked for coexpression of cytokeratin with EGFP by immunofluorescence (Fig. 1, D to G) and for coexpression of cytokeratin with the X and Y chromosomes by fluorescence in situ hybridization (FISH) (Fig. 1H) in cytospins of unsorted cells from the lung digests. EGFP was expressed in the Z/EG × Cre (Fig. 1D) and Z × C F1–into-Cre control mice (Fig. 1G), but not in Z/EG-into-Cre experimental (n = 12 mice) or Cre-into-Cre negative-control recipients (Fig. 1, E and F). In all cases, cytokeratin-positive lung cells contained

Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT 06520, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: diane.krause@yale.edu