Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology

Chris Bakal, *et al.*

Science **316**, 1753 (2007);
DOI: 10.1126/science.1140324

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology

Chris Bakal,[1,2,3]*† John Aach,[1]* George Church,[1] Norbert Perrimon[1,2]

Although classical genetic and biochemical approaches have identified hundreds of proteins that function in the dynamic remodeling of cell shape in response to upstream signals, there is currently little systems-level understanding of the organization and composition of signaling networks that regulate cell morphology. We have developed quantitative morphological profiling methods to systematically investigate the role of individual genes in the regulation of cell morphology in a fast, robust, and cost-efficient manner. We analyzed a compendium of quantitative morphological signatures and described the existence of local signaling networks that act to regulate cell protrusion, adhesion, and tension.

Morphogenesis commonly relies on the spatial and temporal regulation of distinct groups of genes acting in local signaling networks. The morphology of a single cell also results from the spatiotemporally regulated activity of signaling proteins. For example, Rac-type guanosine triphosphatases (GTPases) promote the formation of protrusive lamellipodia at the leading edge of motile cells, whereas Rho-type GTPases promote cortical tension and cell retraction at the rear of the cell through the activation of the actomyosin machinery (*1*). Both protrusive activity and cell body retraction are tightly coupled to the assembly and disassembly of adhesive structures through Rho signaling (*2*). Many signaling proteins must act both upstream and downstream of specific Rho GTPases in spatially distinct subcellular local networks to translate extracellular signals to changes in GTPase activation and ultimately in cellular morphology. However, the components of these networks and the precise role they play in regulating cell shape remain largely unclear.

We performed a genetic screen of 249 gene-overexpression or double-stranded RNA (dsRNA) treatment conditions (TCs) using the *Drosophila* BG-2 cell line to determine the roles of genes acting in local networks to control distinct aspects of cell morphology. BG-2 cells are highly motile and exhibit many of the traits observed in mammalian fibroblasts and epithelial cells, including the formation of integrin-based adhesions, polarized lamellipodia, and coordinated retraction of the cell body (*3*, *4*); but unlike many mammalian cell types, the growth of BG-2 cells is not inhibited by contact with other cells. To analyze the morphologies of cells in each TC, we used protocol and image-processing techniques designed to detect clear and complete boundaries of individual cells and to quantitatively analyze the shapes of these boundaries along with the intensities and textures of their interiors. First, we stochastically labeled samples with green fluorescent protein (GFP) [supporting online material (SOM)] to enable individual cells to stand out in the crowded and overgrown samples, acquired images of these cells using conventional fluorescence microscopy, and used software to identify the boundaries of individual cells (Fig. 1A and SOM). Although these techniques make lower numbers of cells per sample available to subsequent analysis as compared to other published methods (*5*, *6*), we preferred them because they yielded detailed, high-quality empirical boundaries instead of algorithmically determined approximations. For each individual cell, we computed 145 different quantitative features that reflected basic aspects of cell geometry, detailed aspects of cellular protrusions, or the distribution of GFP intensity within the cellular boundaries (Fig. 1B and SOM). Altogether we analyzed 12,601 individual cells from our 249 TCs.

To transform our 145 features into biologically meaningful morphological indicators, we trained a set of neural networks (NNs) to use informative subsets of the features to discriminate cells from particular reference TCs from sets of other reference TCs (Fig. 1D and SOM). We targeted seven TCs for NN training because they produced phenotypes that were qualitatively distinctive and discernable from control cells (SOM). For example, overexpression of an N-terminally truncated form of the Rho guanine nucleotide exchange factor (RhoGEF) SIF (ΔN-SIF), the *Drosophila* ortholog of mammalian Tiam-1, stimulates extensive lamellipodia formation, cell spreading, and a general loss of tension, demonstrated by the flat and thin appearance of the ΔN-SIF cells (Fig. 2) (*7*). After training a ΔN-SIF NN to distinguish ΔN-SIF cells from the cells of our six other target TCs, we applied this NN to all 12,601 cells in our data set to score each of them for this distinctive morphology (Fig. 1D). In addition to ΔN-SIF, we trained NNs for TCs treated with overexpression constructs for RacV12, RacF28L, RhoV14, RhoF30L, CG3799 full-length, and ΔN-RhoGEF3. The seven TCs selected for NN training included several for which the underlying mechanisms responsible for their phenotypes are not understood.

Finally, for each NN and TC, we calculated a NN *Z* score (NNZ), which is the variance-adjusted difference between the mean NN score of all cells in the TC and the mean NN score of all cells in our data set. Each NNZ is thus an index of the morphology of an entire TC (SOM) (Fig. 1E). For example, for cells in which CG10188 (a RhoGEF) was targeted by dsRNA, the NNZ for the ΔN-SIF NN was 0.76, which indicates that CG10188 dsRNA induces a morphology that is slightly more "ΔN-SIF–like" than an equal number of randomly chosen cells (which would have a NNZ of 0). In contrast, the NNZ for ΔN-SIF cells using the ΔN-SIF classifier is 26.77. Together, the seven NNZs computed for each TC constituted a quantitative morphological signature (QMS) of the TC (Fig. 1E). A QMS is thus a high-order representation of the morphology of cells in a TC as a vector of seven specific quantitative similarities and dissimilarities with seven panels of reference cells with distinctive phenotypes.

Two-dimensional hierarchical clustering (SOM) of 249 QMSs revealed that TCs, and in particular RNA interference (RNAi) against individual genes, fell into several distinct clusters. QMSs with similar qualitative phenotypes clustered tightly together (Fig. 2). We define "phenoclusters" as genes grouped at the highest node in the clustering for which the cluster distance metric (an average of uncentered Pearson correlation coefficients) was greater than 0.80 and term this a "cluster distance cutoff" (CDC) (Fig. 2) (*8*). A value of 0.80 was chosen as the CDC because smaller cutoffs resulted in groupings of visually diverse morphologies, whereas higher thresholds resulted in the segregation of visually similar morphologies into distinct clusters.

A large phenocluster was composed of TCs that clustered because their QMSs have high ΔN-RhoGEF3 NNZs (cluster 6). All cells in this cluster were extremely round and had very few or no protrusions of any type (Fig. 2 and fig. S18). QMSs for *p190RhoGAP*, *SCAR*, *slingshot*, *armadillo*, *ankyrin*, *Sop2*, and *RhoGEF3* RNAi were clustered together. Moreover, we observed an enrichment in this cluster for RNAi against genes involved in Rap signaling (three of six genes in the data set). The finding that depletion of either SCAR, the cofilin phosphatase Slingshot, or Sop2 resulted in defects in protrusion is consistent with the known function of these three proteins (*4*, *9*–*14*). Gef26 and its downstream target Rap1 function in the formation of adherence junctions in *Drosophila* (*15*–*17*) and have recently been observed to be required for cell spreading and migration of *Drosophila* macrophages (*18*). In mammalian systems, p190RhoGAP acts downstream of integrins and adhesions to promote cell

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Howard Hughes Medical Institute, Boston, MA 02215, USA. [3]MIT Computer Science and Technology Laboratory, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.

spreading (*19*). Thus, our method successfully identified two distinct but coupled signaling pathways that regulate the formation of protrusions. We anticipate that *Drosophila* RhoGEF3 plays a critical role in the regulation of adhesion, because overexpression of a N-terminally deleted form or inhibition by dsRNA have similar QMSs. ΔN-RhoGEF3 is not likely to be constitutively activated (*20*) but may promote cell rounding by acting in a dominant-negative manner toward endogenous RhoGEF3 signaling.

A second phenocluster contained a group of TCs that co-clustered with RhoF30L and had high RhoF30L NNZs (cluster 8). These cells differed qualitatively from cells in the ΔN-RhoGEF3 phenocluster by virtue of the fact that cells in this cluster did have some visible, but poorly formed, lamellipodial protrusions (Fig. 2 and fig. S18). dsRNAs in this cluster may target genes that specifically promote the formation of lamellipodia, and RhoF30L (an activated form of Rho that cycles between GTP- and GDP-bound states) may inhibit this process. In support of this notion, dsRNAs targeting *twinstar*, *capt*, and *ARC-p20* were part of this cluster, representing an enrichment in genes that have been identified in previous screens for genes required for lamellipodia organization (three of seven genes) (*10*). Given the fact that lamellipodia formation occurs after the formation of adhesion, protrusion, and actin-filament nucleation, this suggests that phenotypic profiling can not only simultaneously monitor the activity of coupled signaling pathways (within phenoclusters) but can also monitor the temporal hierarchical relationships that exist among local signaling networks.

The largest phenocluster was a group of TCs that shared high ΔN-SIF, RacV12, and RacF28L NNZs (cluster 18) and that corresponded to large flat cells, typically with extensive lamellipodia (Fig. 2 and fig. S18). This phenotype is consistent with repeated observations of cells overexpressing activated Rac mutant proteins or activated RacGEFs such as SIF or Tiam-1 (*7*, *21*). QMSs for *cenG1A*, *cenB1A*, *CG16728*, and *CG13692* RNAi were members of this phenocluster, representing an enrichment in ArfGAPs (four of six genes). Mammalian ArfGAPs such as GIT1 and

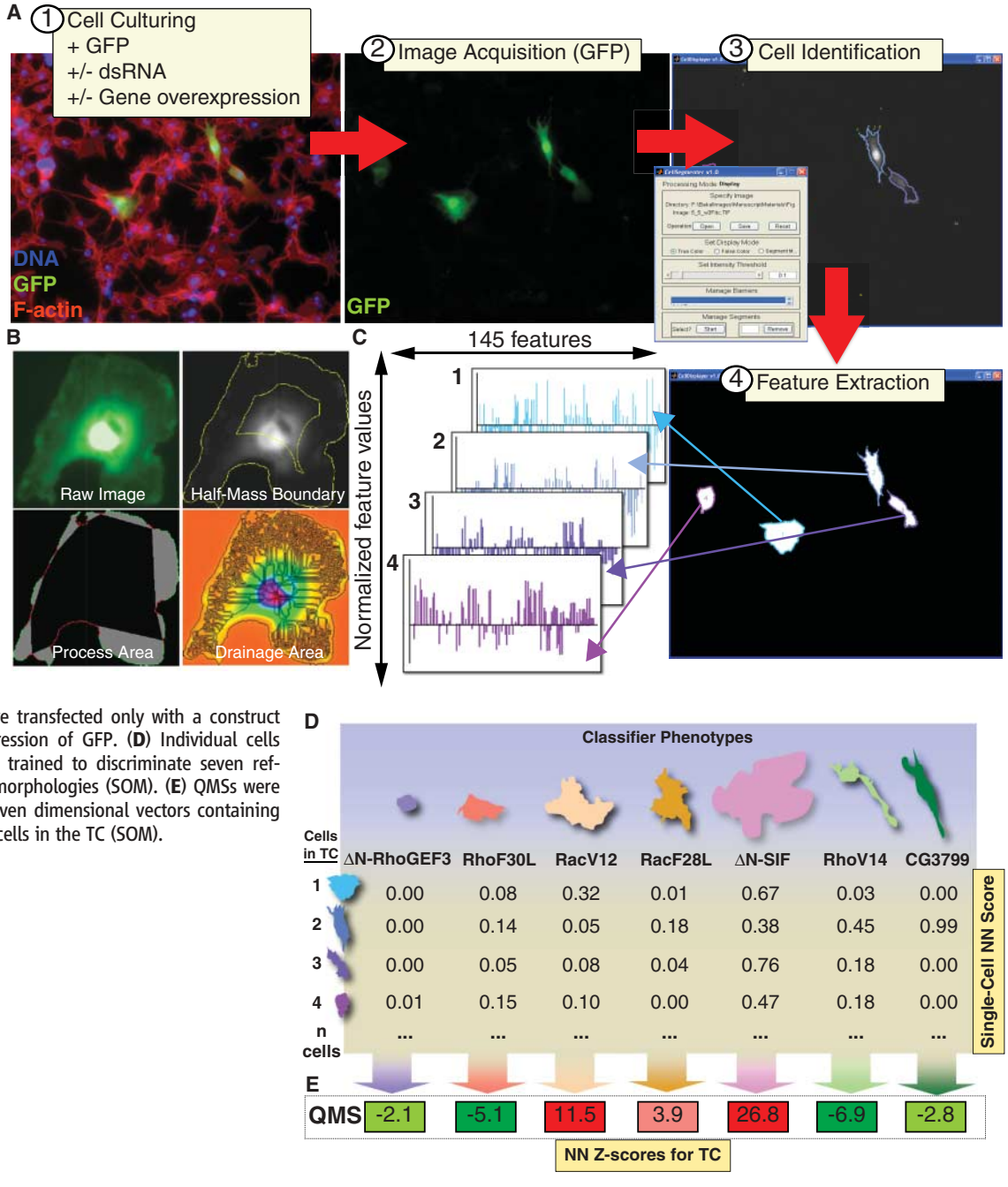

Fig. 1. Phenotypic profiling workflow. (**A**) Cultured *Drosophila* BG-2 cells were transfected with plasmids encoding GFP and either cotransfected with plasmids encoding red fluorescent protein–tagged proteins or incubated in the presence of dsRNA for 4 days. Images of GFP-labeled cells were acquired by standard fluorescence microscopy, and individual cell images with clear and complete boundaries were selected with custom-developed software (SOM). (**B**) Graphical representations of some of the features computed from each individual cell image (SOM). (**C**) 145 different features relevant to cell morphology and GFP signal intensity were derived from individual cells and expressed as *Z* scores relative to their values over a subset of GFP control cells which were transfected only with a construct coding for constitutive expression of GFP. (**D**) Individual cells were then scored with NNs trained to discriminate seven reference TCs with distinctive morphologies (SOM). (**E**) QMSs were computed for each TC as seven dimensional vectors containing the NNZs of the individual cells in the TC (SOM).

GIT2 promote the disassembly of integrin-based focal adhesions by binding the tyrosine-phosphorylated form of Paxillin, which in turn results in ArfGTPase activation and a concomitant down-regulation of Rac activity and adhesion turnover (*22*–*24*). In addition to ArfGAPs, this phenocluster also contained QMSs for dsRNAs targeting *paxillin* (fig. S18) and α-actinin, and was defined by high levels of Rac activity (Fig. 2). Based on this and sequence analysis, we suggest that *CG16728* is the *Drosophila* ortholog of mammalian GIT ArfGAPs. Proteomic analysis has recently revealed that GIT1 is part of a supramolecular complex that, in addition to Paxillin, includes β_2 centaurin and gelsolin and directly binds moesin (*25*). *Drosophila cenB1A* (the *Drosophila* ortholog of β_2 *centaurin*), *Gelsolin*, and *Moesin* dsRNA were also members of same phenocluster as *CG16728* and *paxillin*, further demonstrating that our methodology is capable of identifying both functionally and physically coupled signaling components. Taken together, this suggests that the large, flat, and spread morphology of cells in this phenocluster was partially due to inhibiting the disassembly of adhesion and that QMSs can be used to functionally annotate genes.

We have demonstrated that quantitative morphological profiling of single cells combined with RNAi-based genetic screening technology results in the identification of local signaling networks with spatially, temporally, and functionally defined characteristics that act in a hierarchical manner to regulate cell shape and migration. These methods can be used not only in the context of genetic screens but also in large-scale screens of small-molecule libraries or screens involving the overexpression of cDNAs. Because our approach is a fast and cost-effective way to query the activity of multiple signaling proteins and pathways, quantitative morphological profiling may also be useful as a diagnostic tool in the analysis of clinical samples. Furthermore, akin to gene-expression data, we can now use morphological phenotypic data for computa-

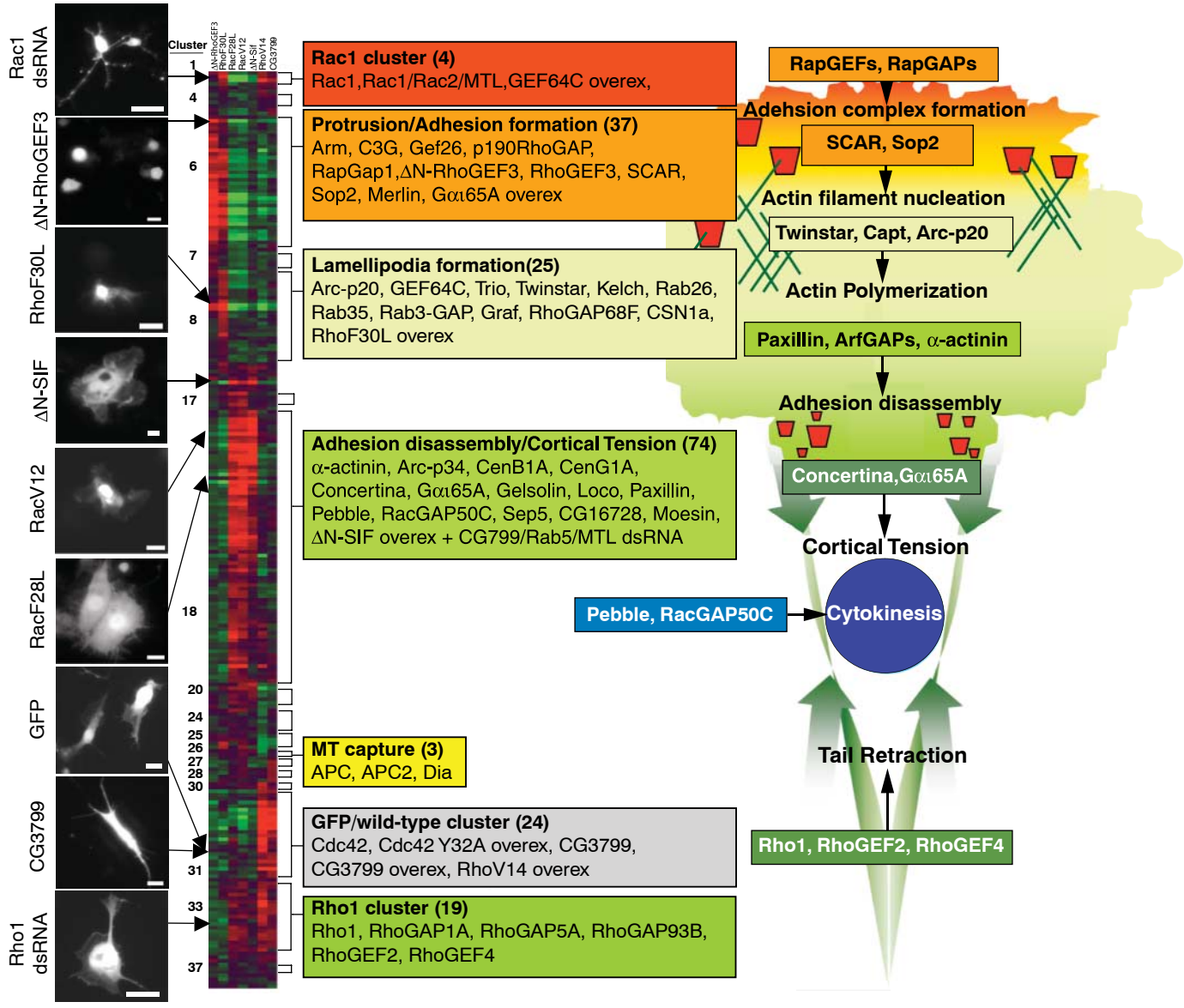

Fig. 2. Identification of local networks that regulate distinct aspects of morphology. Hierarchical clustering of the genes in the data set (the *y* axis) by how cells scored on the ΔN-SIF, ΔN-RhoGEF3, CG3799, RacF28L, RacV12, RhoF30L, and RhoV14 NNs (the *x* axis) is shown. We define phenoclusters of genes as clusters with a CDC >0.80, which results in 41 total clusters comprising 17 multigene clusters and 24 singletons. All multigene clusters are identified in brackets on the right-hand side of the clustergram. For some clusters, we describe prominent TCs, and the number of TCs within these clusters is indicated in parentheses. Examples of individual cells and their positions in the clustergram are shown on the left-hand side of the clustergram. Based on their gene membership, a number of clusters were determined to have specialized roles in cell morphology. A complete listing of clusters is provided in table S8. Scale bars, 10 μm.

tional approaches that aim to model the dynamic nature of signaling networks, while the RNAi component pushes us closer to causal mechanistic linkages.

References and Notes

1. A. Hall, *Biochem. Soc. Trans.* **33**, 891 (2005).
2. K. A. DeMali, K. Wennerberg, K. Burridge, *Curr. Opin. Cell Biol.* **15**, 572 (2003).
3. Y. Takagi, K. Ui-Tei, T. Miyake, S. Hirohashi, *Neurosci. Lett.* **244**, 149 (1998).
4. A. Biyasheva, T. Svitkina, P. Kunda, B. Baum, G. Borisy, *J. Cell Sci.* **117**, 837 (2004).
5. Z. E. Perlman *et al.*, *Science* **306**, 1194 (2004).
6. A. E. Carpenter *et al.*, *Genome Biol.* **7**, R100 (2006).
7. M. Sone *et al.*, *Science* **275**, 543 (1997).
8. K. C. Gunsalus *et al.*, *Nature* **436**, 861 (2005).
9. J. A. Zallen *et al.*, *J. Cell Biol.* **156**, 689 (2002).
10. S. L. Rogers, U. Wiedemann, N. Stuurman, R. D. Vale, *J. Cell Biol.* **162**, 1079 (2003).
11. P. Kunda, G. Craig, V. Dominguez, B. Baum, *Curr. Biol.* **13**, 1867 (2003).
12. N. Zebda *et al.*, *J. Cell Biol.* **151**, 1119 (2000).
13. M. Nishita *et al.*, *J. Biol. Chem.* **279**, 7193 (2004).
14. M. D. Welch, A. H. DePace, S. Verma, A. Iwamatsu, T. J. Mitchison, *J. Cell Biol.* **138**, 375 (1997).
15. A. L. Knox, N. H. Brown, *Science* **295**, 1285 (2002).
16. H. Wang *et al.*, *Dev. Cell* **10**, 117 (2006).
17. S. R. Singh *et al.*, *Dev. Growth Differ.* **48**, 169 (2006).
18. S. Huelsmann, C. Hepper, D. Marchese, C. Knoll, R. Reuter, *Development* **133**, 2915 (2006).
19. W. T. Arthur, K. Burridge, *Mol. Biol. Cell* **12**, 2711 (2001).
20. K. Murayama *et al.*, *J. Biol. Chem.* (26 December 2006).
21. E. E. Sander, J. P. ten Klooster, S. van Delft, R. A. van der Kammen, J. G. Collard, *J. Cell Biol.* **147**, 1009 (1999).
22. D. J. Webb *et al.*, *Nat. Cell Biol.* **6**, 154 (2004).
23. M. C. Brown, L. A. Cary, J. S. Jamieson, J. A. Cooper, C. E. Turner, *Mol. Biol. Cell* **16**, 4316 (2005).
24. N. Nishiya, W. B. Kiosses, J. Han, M. H. Ginsberg, *Nat. Cell Biol.* **7**, 343 (2005).
25. M. W. Mayhew *et al.*, *J. Proteome Res.* **5**, 2417 (2006).

26. We thank P. Bradley and A. Kiger for contributions to the stochastic labeling method, as well as B. Mathey-Prevot, A. Friedman, R. Griffin, B. Berger, S. Altschuler, the Drosophila RNAi Screening Center, and the Harvard Research Information Technology Group. This work is supported in part by grants from the National Human Genome Research Institute Centers of Excellence in Genomic Science. N.P. is an investigator of the Howard Hughes Medical Institute. C.B. is a fellow of the Leukemia and Lymphoma Society.

Supporting Online Material

www.sciencemag.org/cgi/content/full/316/5832/1753/DC1
Materials and Methods
SOM Text
Figs. S1 to S19
Tables S1 to S9
References

24 January 2007; accepted 15 May 2007
10.1126/science.1140324

Restriction of an Extinct Retrovirus by the Human TRIM5α Antiviral Protein

Shari M. Kaiser,[1,2] Harmit S. Malik,[3] Michael Emerman[2,3]*

Primate genomes contain a large number of endogenous retroviruses and encode evolutionarily dynamic proteins that provide intrinsic immunity to retroviral infections. We report here the resurrection of the core protein of a 4-million-year-old endogenous virus from the chimpanzee genome and show that the human variant of the intrinsic immune protein TRIM5α can actively prevent infection by this virus. However, we suggest that selective changes that have occurred in the human lineage during the acquisition of resistance to this virus, and perhaps similar viruses, may have left our species more susceptible to infection by human immunodeficiency virus type 1 (HIV-1).

A large portion of primate genomes is composed of endogenous retroviruses that can be thought of as an archaeological record of past infections. Both chimpanzee and gorilla genomes harbor more than 100 copies of *Pan troglodytes* endogenous retrovirus (PtERV1), whereas it is absent from the human genome (*1*). Comparison of individual PtERV1 proviruses in gorilla and chimpanzee genomes suggest that this virus was active 3 to 4 million years ago, after the separation of chimpanzee and human lineages (*1*). This raises an evolutionary conundrum as to why sister species, but not humans, acquired germline copies of this retrovirus even though all three species cohabited when PtERV1 was an active exogenous virus (*1*). One mechanism of active restriction from retroviral infections is conferred by the TRIM5α protein, which binds directly to the incoming retroviral capsid (CA) core and targets its premature disassembly or destruction (*2*, *3*). Each primate species encodes a TRIM5α with a different antiviral specificity (*4*). For example, TRIM5α encoded by rhesus macaques renders them resistant to infection by HIV-1, but human TRIM5α affords no such protection (*5*). Indeed, the antiviral specificity of TRIM5α has rapidly evolved by dramatic episodes of positive selection during the past 30 million years of primate evolution (*6*). The branch leading to the human lineage shows one of the strongest signatures of positive selection (*6*), which suggests that at least one major pathogenic retroviral assailant has challenged the human lineage in the past 4 to 5 million years. Taken together, these findings suggest that TRIM5α evolution was shaped by a species-specific history of ancestral retroviral challenges. Although human TRIM5α has relatively poor activity against retroviruses compared with the gene from other primates, it potently blocks a γ-retrovirus N-MLV, which is related to PtERV1 (*7*, *8*). We therefore tested the hypothesis that TRIM5α may have protected early humans from invasion by PtERV1.

All copies of PtERV1 in the chimpanzee genome have been inactivated by accumulated detrimental mutations (*1*). However, the numerous proviral copies of PtERV1 present in the chimpanzee genome allow us to reconstruct the ancestral sequence of the *gag* gene of this ancient, extinct retrovirus in silico (see supporting online material). Analysis of the reconstructed PtERV1 ancestral sequence reveals about 50% identity with murine leukemia virus (MLV), and several characteristic conserved elements are intact (Fig. 1A). Phylogenetic analysis of chimpanzee and gorilla PtERV sequences shows that a single-source virus likely infected both chimpanzees and gorillas because viral sequences from both species form a monophyletic group (Fig. 1B).

We next used site-directed mutagenesis to reconstruct the ancestral p12 and CA coding regions (ignoring synonymous changes) starting from one chimpanzee PtERV1 provirus cloned from the genome. We focused on CA because it is the functional target of TRIM5α and included p12 because of functional interactions that exist between p12 and CA in other γ-retroviruses (*9*). Because TRIM5α interacts with the retroviral CA only in the multimeric structure characteristic of mature retroviral particles (*10*), we generated the PtERV1 capsid core in the context of an infectious virus capable of only a single round of infection. This was achieved by constructing a chimeric virus between PtERV1 and MLV that encodes a *gag/pol* gene expressing the reconstructed p12 and CA proteins of PtERV1 with the remainder of the viral structural proteins and enzymes of MLV (Fig. 1C). Our MLV/PtERV1 chimeric virus was indeed infectious (Fig. 1D), which demonstrates that regions of a 3- to 4-million-year-old primate endogenous retrovirus can be successfully resurrected.

We tested human TRIM5α restriction of PtERV1 by infecting cells that express an exogenous copy of human TRIM5α. A much younger human endogenous retrovirus, HERV-K, was also recently resurrected (*11*, *12*) but was not restricted by human TRIM5α (*12*). In contrast, expression of human TRIM5α in a heterologous cell type resulted in a dramatic reduction of infectivity of the MLV/PtERV1 chimera by a factor of more than 100 compared with cells that do not express TRIM5α (Fig. 2A). These data indicate that humans possess an intrinsic immunity gene capable of effectively neutralizing an extinct retrovirus that never successfully fixed into the human genome.

Specificity of TRIM5α for a particular retroviral capsid is largely determined by amino acids within the C-terminal B30.2 domain. Within this

[1]Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA. [2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [3]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

*To whom correspondence should be addressed. E-mail: memerman@fhcrc.org