REPORTS

successfully accomplishing a variety of pressing tasks. Given similar numbers of workers, environmental conditions, and need, newly founded colonies built comb faster, foraged more, and stored greater amounts of food when their work forces were comprised of many genetically distinct patrilines. Initial differences in labor productivity amplified growth rates over time and led to dramatic fitness gains for genetically diverse colonies (i.e., production of drones, colony growth, and survival). Thus, we expect intense selection favoring polyandry because intracolonial genetic diversity improves the productivity of the work force and increases colony fitness during the risky process of colony founding.

Higher collective productivity of genetically diverse colonies may be rooted in a broader or more sensitive response from worker populations to changing conditions. The probability that a worker will engage in a task has been linked repeatedly to genotype [e.g. (5, 8, 19)]. Consequently, colonies with multiple patrilines would be expected to have worker populations that are able to respond to a broad range of task-specific stimuli and, as a group, should be able to provide appropriate, incremental responses to changes in these stimuli (5). The observation that intracolonial genetic diversity improved productivity in colonies is consistent with predictions made by models of division of labor that rely on genotypic differences in response thresholds among workers (20). Nevertheless, the extent to which genetically uniform colonies lagged behind genetically diverse colonies in the early stages of colony development was surprising, considering that colonies initially lacked comb and food reserves, and presumably, stimuli reflecting these needs could not have been greater. Actual response thresholds of workers are not well documented (20), and it is difficult to know how they are related to the productivity of individuals and the colony as a whole. For example, workers may vary genetically in the rate at which they perform a task once their response threshold is reached or they may not be "good" at tasks for which they have high thresholds (i.e., they lack physiological apparatuses or experience). Alternatively, thresholds may be so high for some tasks that behaviors are effectively missing from a worker's repertoire, thus multiple patrilines would contribute to the diversity of labor in a colony, rather than division of labor among workers.

A key advantage of intracolonial genetic diversity was revealed during infrequent periods when food resources were plentiful (~33 days during our study). Genetically diverse colonies gained weight at rates that far exceeded those of genetically uniform colonies (Fig. 3), whose sluggish foraging rates suggest that intracolonial genetic diversity enhances the discovery and exploitation of food resources by work forces, especially during periods when resources become suddenly and abundantly available. Intracolonial genetic diversity would result in more rapid mobilization of forager work forces if, by broadening the range of response thresholds in colonies, it increased the probability of having sufficient workers functioning as foragers and/or broadened the range of conditions over which foragers inspected, scouted, recruited to or were recruited/reactivated to food resources. Selection for polyandry would be strong if the genetic diversity that it bestows on colonies enhances the sophisticated mechanisms of honey bees for recruiting nest mates to food. Because successful colony founding by honey bees depends so heavily on rallying foragers and the swift accumulation of resources, this could explain, in concert with other benefits unrelated to worker productivity (15, 21), the widespread occurrence of extreme polyandry in all honey bee species.

References and Notes

1. R. H. Crozier, P. Pamilo, *Evolution of Social Insect Colonies* (Oxford Univ. Press, Oxford, 1996).
2. J. Strassmann, *Insectes Soc.* **48**, 1 (2001).
3. D. R. Tarpy, D. I. Nielsen, *Ann. Entomol. Soc. Am.* **95**, 513 (2002).
4. R. H. Crozier, E. J. Fjerdingstad, *Ann. Zool. Fennici* **38**, 267 (2001).
5. G. E. Robinson, R. E. Page Jr., in *Genetics of Social Evolution*, M. D. Breed, R. E. Page Jr., Eds. (Westview Press, Boulder, CO, 1989), pp. 61–80.
6. B. P. Oldroyd, T. E. Rinderer, J. R. Harbo, S. M. Buco, *Ann. Entomol. Soc. Am.* **85**, 335 (1992).
7. S. Fuchs, V. Schade, *Apidologie* **25**, 155 (1994).
8. J. C. Jones, M. R. Myerscough, S. Graham, B. P. Oldroyd, *Science* **305**, 402 (2004).
9. R. E. Page Jr., G. E. Robinson, M. K. Fondrk, M. E. Nasr, *Behav. Ecol. Sociobiol.* **36**, 387 (1995).
10. P. Neumann, R. F. A. Moritz, *Insectes Soc.* **47**, 271 (2000).
11. T. D. Seeley, *Oecologia* **32**, 109 (1978).
12. T. D. Seeley, R. A. Morse, *Insectes Soc.* **25**, 323 (1978).
13. R. D. Fell et al., *J. Apic. Res.* **16**, 170 (1977).
14. Materials, methods, and statistical analyses are available on *Science* Online.
15. T. D. Seeley, D. R. Tarpy, *Proc. R. Soc. Lond. B. Biol. Sci.* **274**, 67 (2007).
16. J. B. Free, P. A. Racey, *Entmol. Exp. Appl.* **11**, 241 (1968).
17. T. D. Seeley, P. K. Visscher, *Ecol. Entomol.* **10**, 81 (1985).
18. P. C. Lee, M. L. Winston, *Ecol. Entomol.* **12**, 187 (1987).
19. G. Arnold, B. Quenet, C. Papin, C. Masson, W. H. Kirchner, *Ethology* **108**, 751 (2002).
20. S. N. Beshers, J. H. Fewell, *Annu. Rev. Entomol.* **46**, 413 (2001).
21. D. R. Tarpy, R. E. Page Jr., *Behav. Ecol. Sociobiol.* **52**, 143 (2002).
22. We thank K. Burke for field assistance, T. and S. Glenn for rearing queens, and P. Barclay, M. Kirkland, and two anonymous reviewers for comments on the manuscript. Funded by a Postdoctoral Fellowship (H.R.M.) from Natural Sciences and Engineering Research Council (Canada) and a grant from the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service (T.D.S.) (no. 2003-35302-13387).

PDZ Domain Binding Selectivity Is Optimized Across the Mouse Proteome

Michael A. Stiffler,[1]* Jiunn R. Chen,[2]* Viara P. Grantcharova,[1]† Ying Lei,[1] Daniel Fuchs,[1] John E. Allen,[1] Lioudmila A. Zaslavskaia,[1]‡ Gavin MacBeath[1]§

PDZ domains have long been thought to cluster into discrete functional classes defined by their peptide-binding preferences. We used protein microarrays and quantitative fluorescence polarization to characterize the binding selectivity of 157 mouse PDZ domains with respect to 217 genome-encoded peptides. We then trained a multidomain selectivity model to predict PDZ domain–peptide interactions across the mouse proteome with an accuracy that exceeds many large-scale, experimental investigations of protein-protein interactions. Contrary to the current paradigm, PDZ domains do not fall into discrete classes; instead, they are evenly distributed throughout selectivity space, which suggests that they have been optimized across the proteome to minimize cross-reactivity. We predict that focusing on families of interaction domains, which facilitates the integration of experimentation and modeling, will play an increasingly important role in future investigations of protein function.

Eukaryotic proteins are modular by nature, comprising both interaction and catalytic domains (1, 2). One of the most frequently encountered interaction domains, the PDZ domain, mediates protein-protein interactions by binding to the C termini of its target proteins (3–6). Previous studies of peptide-binding selectivity have placed PDZ domains into discrete functional categories: Class I domains recognize the consensus sequence Ser/Thr-X-ψ-COOH, where X is any amino acid and ψ is hydrophobic; class II domains prefer ψ-X-ψ-COOH; and class III domains prefer Asp/Glu-X-ψ-COOH (5, 7). More recent information has suggested that these designations are too restrictive and so additional classes have been proposed (8, 9). The idea that domains fall into discrete categories, however, raises questions about functional overlap: Domains within the same class are more likely to cross-react with each other's ligands. To resolve this issue, we characterized and modeled PDZ domain selectivity on a genome-wide scale.

We began by cloning, expressing, and purifying most of the known PDZ domains encoded in

the mouse genome (*10*–*12*) (table S1). Soluble protein of the correct molecular weight was obtained for 157 PDZ domains (fig. S1). Whereas previous efforts to characterize the selectivity of PDZ domains have relied on collections of peptides with randomized sequences (*7*, *9*, *13*, *14*), our goal was to focus on genome-encoded sequences. We therefore synthesized and purified fluorescently labeled peptides derived from the 10 C-terminal residues of mouse proteins. In total, we synthesized 217 such peptides, which we termed our "training set" (table S2) (*15*). Although our training set is not guaranteed to contain ligands for every PDZ domain, it permitted us to obtain a broad view of binding selectivity.

To investigate biophysical interactions between the 157 well-behaved PDZ domains and each of the 217 fluorescent peptides, we devised a strategy that combines the throughput of protein microarrays and the fidelity of fluorescence polarization (FP) with predictive modeling (Fig. 1A). Microarrays of PDZ domains were prepared within individual wells of microtiter plates and probed, in triplicate, with a 1 µM solution of each peptide (Fig. 1B) (*16*). Interactions with a mean fluorescence that was at least three times the median fluorescence on the array were scored as "array positives" (*17*). This process yielded 1301 putative interactions involving 127 PDZ domains. Little can be concluded about the 30 domains for which no array positives were found. For domains that bound at least one peptide, however, the inability to bind other peptides provides important information: These noninteractions were scored as "array negatives."

As with any high-throughput method, there are error rates associated with identifying both positives and negatives. To eliminate array false positives, we retested and quantified every array positive with a solution-phase FP assay (Fig. 1C), which served as our "gold standard" (*17*). By measuring FP at 12 concentrations of PDZ domain, we determined the dissociation constant (K_d) for each of the 1301 array positives (table S3). Interactions that showed saturation binding (*18*) with a $K_d < 100$ µM were considered "positives"; those that did not were considered "negatives." On the basis of these criteria, 85 PDZ domains bound at least one peptide in the training set. Although our K_d cutoff was high, ~90% of the interactions had a $K_d < 50$ µM, and ~60% of interactions had a $K_d < 20$ µM (fig. S2). In addition, FP assays revealed the false-

negative rate of the protein microarray assay to be 6.6% (*19*).

To extract from our data the rules that govern the peptide-binding selectivity of PDZ domains, we built a model that predicts the PDZ domains to which a peptide will bind, given its sequence. Peptide recognition is often modeled with a position-specific scoring matrix (PSSM), $\Theta =$

$\{\theta_{p,q}\}$, where $\theta_{p,q}$ is defined as the probability of observing amino acid q at position p in the subset of peptides that bind to that domain (*20*). This scoring approach is useful for predicting peptides that bind to a single domain, but it is not ideally suited to our purpose for two reasons. First, our peptide sequences are derived from the genome and thus are not random. Second, our goal is to
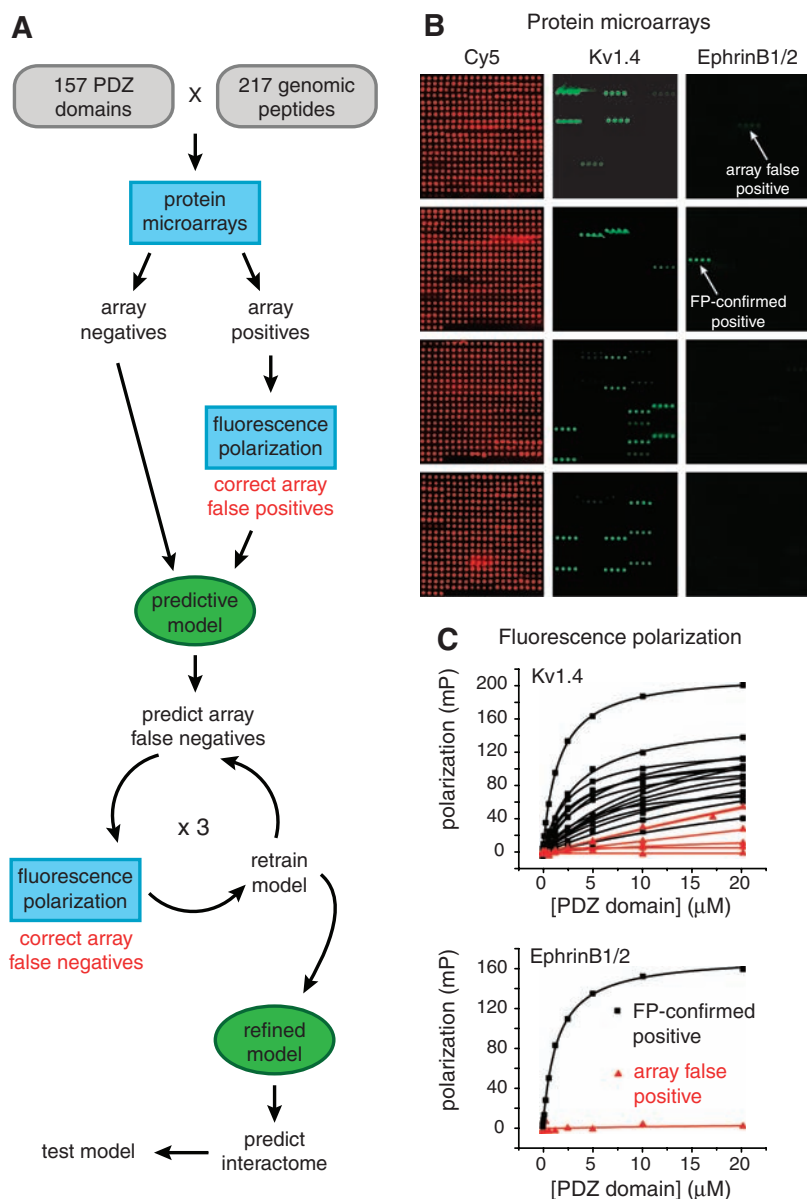

Fig. 1. (**A**) Strategy for constructing a multidomain selectivity model for mouse PDZ domains. Protein microarrays were used to test all possible interactions between 157 mouse PDZ domains and 217 genome-encoded peptides. Array positives were retested and quantified by FP, thereby correcting array false positives. The resulting data were used to train a predictive model of PDZ domain selectivity. The model highlighted putative array false negatives, which were tested by FP, and the corrected data were used to retrain the model. After three cycles of prediction, testing, and retraining, the refined model was used to predict PDZ domain–protein interactions across the mouse proteome. (**B**) Representative images of protein microarrays, probed with fluorescently labeled peptides. PDZ domains were spotted in quadruplicate in individual wells of 96-well microtiter plates. (Four wells were required to accommodate all of the domains.) The red images (Cy5) show the location of the PDZ domain spots. The green images show arrays probed with a promiscuous peptide derived from Kv1.4 (left) and a selective peptide derived from ephrin B1/2 (right). (**C**) FP titration curves obtained for the array positives identified in (**B**).

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA. [2]Department of Molecular and Cellular Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA.

*These authors contributed equally to this work.
†Present address: Merrimack Pharmaceuticals, 1 Kendall Square, Building 700, Cambridge, MA 02139, USA.
‡Present address: Tepnel Lifecodes Corporation, 550 West Avenue, Stamford, CT 06902, USA.
§To whom correspondence should be addressed. E-mail: macbeath@chemistry.harvard.edu

learn how one domain differs from another, (i.e., how selectivity is achieved). This information is not captured in a traditional PSSM because peptide residues that contribute strongly to binding affinity, such as the C-terminal residue, dominate the model, even if they are not important in distinguishing one domain from another.

To construct a single model that includes many PDZ domains, we developed a variation of a PSSM in which a peptide is predicted to bind to PDZ domain i if

$$\varphi_i = \sum_{p,q} A_{p,q}\theta_{i,p,q} > \tau_i \qquad (1)$$

where φ is a binding score, A is an indicator of peptide sequence, ($A_{p,q} = 1$ if the amino acid at position p of the peptide is q and $A_{p,q} = 0$ otherwise), and τ_i is a scoring threshold, specific to each domain. To ensure that our model focuses on PDZ domain selectivity, we constrained $\sum_i \theta_{i,p,q}$ to be 0 for every position p and every amino acid q. Thus, $\theta_{i,p,q}$ is positive if PDZ domain i prefers amino acid q at position p more than the other PDZ domains, negative if it prefers it less, and 0 if it has no bias relative to the other domains. To tailor the threshold appropriately for each domain, we defined τ_i to be the mth percentile of φ_i's for all of the peptides in our training set that bound to PDZ domain i. Empirically, we found that setting $m = 5$ provides a good balance between false-positive predictions and false-negative predictions. Because this model is designed to highlight selectivity across many members of a domain family, we refer to it as a multidomain selectivity model (MDSM).

Our model takes into account the five C-terminal residues of the peptide ligand: positions –4, –3, –2, –1, and 0. Even with 217 data points for each domain, there is insufficient information to train such a high-dimensional model. To avoid overfitting, we implemented a smoothing technique. If two PDZ domains bind a similar subset of peptides, it is reasonable to expect that their $\theta_{p,q}$'s are also similar, unless the data suggest otherwise. Likewise, if two amino acids have similar physicochemical properties, it is reasonable to expect that their $\theta_{i,p}$'s will be similar. Smoothing requires a quantitative measure of pairwise distance. With PDZ domains, distance was defined as the Hamming distance of their binding vectors across the training-set peptides. With amino acids, we relied on previously reported "z scales" to capture their physicochemical properties, where z_1 is considered a descriptor of hydrophilicity, z_2 is a descriptor of molecular weight and surface area, and z_3 is a descriptor of polarity and charge (21). We reduced the equivalent degrees of freedom in our model by smoothing over PDZ domains and over amino acids with a Gaussian kernel during regression (22).

We were able to model 74 of the 85 PDZ domains, which suggests that the majority of

PDZ domains (87%) conform to the assumption that the contribution of each peptide position to selective binding is additive. Having trained the MDSM, we used it to predict false negatives in our microarray data (Fig. 1A). Predicted array false negatives were assayed experimentally by quantitative FP, and the MDSM was retrained using the updated information. This cycle of prediction, experimentation, and retraining was performed three times. In total, we tested 303 predicted array false negatives, of which 133 (44%) were found to be positives, yielding a high-quality, quantitative interaction matrix for mouse PDZ domains (Fig. 2A and table S3). Overall, we found that the average binding affinity of the array false negatives was slightly lower than that of the array true positives. The distributions of binding affinities, however, overlapped considerably (fig. S3).

The refined model performs well on the updated data set, with a true-positive rate of 96% (it correctly identifies 515 of 536 FP-confirmed positives) and a false-positive rate of 15% (it predicts an interaction for 186 of 1229 FP-confirmed negatives) when m is set to 5 (Fig. 2B). The parameters of the MDSM are depicted as a heat map in Fig. 2C and are provided in table S4. As anticipated, position 0 does not contribute strongly to discriminative binding, but the four other positions contribute substantially (Fig. 2C).

To extract biophysical modules out of the resulting interaction network, we designed a modified version of the Markov cluster algorithm (23), tailored to the special situation of a bipartite network (22). The algorithm simulates a random walk on the graph and is based on the observation that random walks tend to be confined within "tight clusters" of nodes. The algorithm identified four tight clusters of PDZ domains and their binding partners (Fig. 2D). For example, the claudins (tight junction proteins) cluster with ZO-1 and ZO-2, whereas the N-methyl-D-aspartic acid (NMDA) receptor subunit isoforms NMDAR2A and NMDAR2B, as well as several voltage-gated potassium channels, cluster with PSD-95, SAP-97, Magi-1, Magi-2, and Magi-3.

Encouraged by the close agreement of our model with the training-set data, we used the MDSM to predict to which proteins in the mouse proteome each of the 74 PDZ domains are able to bind. In total, we surveyed 31,302 peptide sequences corresponding to the C termini of all translated open reading frames (24). We have previously shown that our domain-based in vitro strategy faithfully captures ~85% of the previously reported interactions involving PDZ domains (17). We therefore provide these predictions (18,149 PDZ domain–peptide interactions) as supplemental information (table S5) to help guide future biological investigations (25). We note, however, that not all interactions that are observed in vitro necessarily occur in vivo.

To further assess the accuracy of our model, we selected a "test set" of 48 proteins from the mouse proteome that were predicted to be highly

connected to PDZ domains (table S6). We synthesized fluorescently labeled peptides corresponding to their C termini and assayed them for binding to the 74 PDZ domains in our MDSM with the use of a single-point FP assay (26). These peptides were not included in the training set and so offer a stringent test of our model. In total, 493 new interactions and 3059 noninteractions were identified. Our model predicted 48% (237) of the new interactions and 88% (2680) of the noninteractions when m was set to 5 (Fig. 2E), with a true-positive/false-positive (TP/FP) ratio of 0.63 (237/379). The TP/FP ratio of our model predictions exceeds by a factor of more than 20 the TP/FP ratio of a Bayesian model that integrates information from two large-scale yeast two-hybrid experiments and two large-scale in vivo pull-down experiments in *Saccharomyces cerevisiae*, while maintaining the same true-positive rate (27). We attribute the accuracy of our MDSM to its focus on a related family of domains, rather than on a broad collection of proteins with disparate properties. This argues strongly for a systematic but segmented effort to uncover protein-protein interactions by focusing on families of interaction modules.

We also observed a positive correlation between the model output (φ_i) and binding affinity (fig. S5). We found that smoothing over both PDZ domains and amino acids substantially contributes to the accuracy of the model, boosting the TP/FP ratio by 44% over the model constructed without smoothing, while maintaining the true-positive rate essentially the same (Fig. 2E). Most of the effect was derived from smoothing over PDZ domains, but smoothing over amino acids was also beneficial. To exclude the possibility that the model performance was due to chance correlation, we performed a Y-randomization test (28) in which the interaction data were shuffled. The resulting receiver operating characteristic (ROC) curve was indistinguishable from the no-discrimination line (fig. S6), indicating the effectiveness of our training and test sets.

Having established that the model accurately captures information about the binding selectivity of PDZ domains, we asked which physicochemical properties each domain uses at each position to define its selectivity. For example, if we look at the amino acid preferences of Dlgh3 (1/1) at position –4, we find that the 20 θ's are positively correlated with z_1 but are not correlated with z_2 or z_3 (Fig. 3A). In contrast, z_2, but not z_1 or z_3, correlates with discriminative binding at position –4 for Magi-1 (4/6) (Fig. 3B), whereas z_3, but not z_1 or z_2, correlates with discriminative binding at position –4 for MUPP1 (10/13) (Fig. 3C). These three examples are extremes; in general, PDZ domains rely on all three z scales for discriminative binding. To capture this information for all PDZ domains at all positions, we constructed a correlation matrix between the model parameters and the first three z scales of amino acids (Fig. 3D). Because the contribution to discriminative binding at position 0 is weak, we omitted this

position from our analysis to avoid biasing our results with artificially amplified effects.

To understand the organization of peptide-binding selectivity on a global level, we deconvoluted the correlation matrix through singular-value decomposition and found that the distribution of PDZ domain binding preferences can be largely explained by three principal axes. The space defined by these axes can be thought of as "PDZ domain selectivity space." Each of the first two axes explains ~30% of the variance in the correlation matrix, whereas the third axis explains ~14% (Fig. 3E). The first axis (Fig. 3F) can distinguish canonical class I PDZ domains, which are preferred by peptides with a small, hydrophilic residue at position –2, from canonical class II domains, which are preferred by peptides with a large, hydrophobic residue at posi-
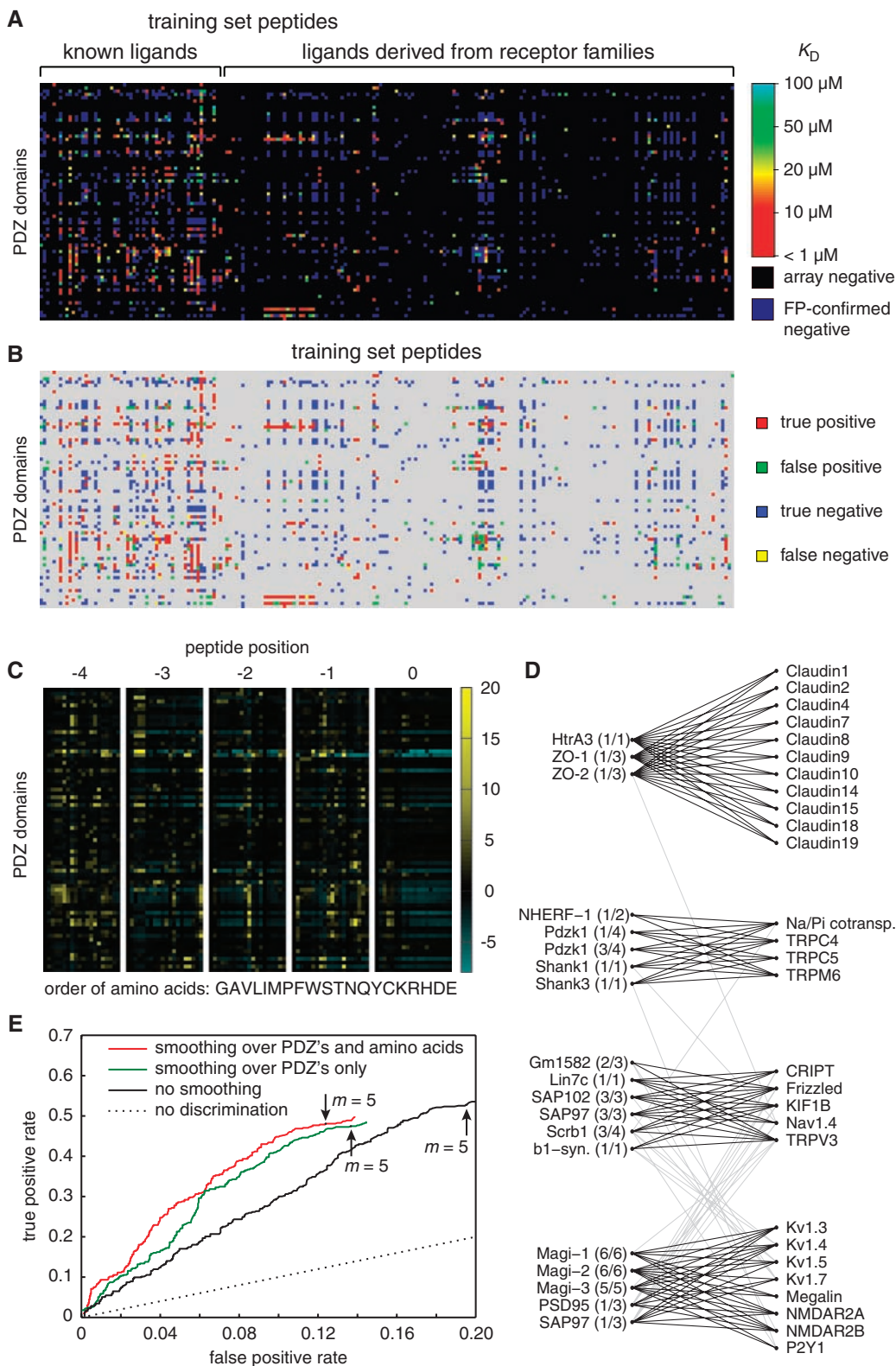

Fig. 2. (A) Graphical view of the training-set data. K_d's of FP-confirmed positives are represented by colors, ranging from high affinity (red) to low affinity (light blue). Array negatives are shown in black, and FP-confirmed negatives are shown in dark blue. Numerical values are provided in table S3. (B) Performance of the MDSM on the training set, with m set to 5. True positives are shown in red, false positives in green, true negatives in blue, and false negatives in yellow. (C) Graphical representation of the MDSM parameters, $\theta_{i,p,q}$. Positive contributions to discriminative binding are graded from black to yellow, and negative contributions are graded from black to light blue. Numerical values are provided in table S4. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (D) Tight clusters embedded in the bipartite interaction network between the 74 PDZ domains and the 217 training-set peptides. (E) ROC curves for three versions of the MDSM, obtained with the test set of 48 peptides. The best performance was obtained after smoothing over both PDZ domains and amino acids. The performance of each version of the MDSM with m set to 5 is indicated with an arrow.

tion –2. Thus, the class I domains PSD-95 (1/3) and Shank3 (1/1) lie at the negative end of the first principal axis (Fig. 3G), whereas the class II domains PDZ-RGS3 (1/1) and Grip1 (6/7) lie at the positive end. Erbin (1/1), which has been shown to bind both class I and class II peptides (28–31), lies between the two extremes. The second and third principal axes (Fig. 3F) add further resolution. In particular, the third axis distinguishes class III domains, such as neuronal nitric oxide synthase (nNOS) (1/1) (preferred by peptides with a negatively charged residue at position –2), from the other PDZ domains. The closer a PDZ domain lies to the positive end of

the third principal axis, the more it falls into the class III designation.

There are, however, two important differences between the standard view of PDZ domain selectivity and the view that emerges from our broad investigation. First, positions –4, –3, –2, and –1 all contribute substantially to the definition of our three principal axes (Fig. 3F). This implies that selectivity is derived from interactions throughout the binding pocket, whereas peptide library screens have shown that affinity is derived largely from the recognition of amino acids at positions –2 and 0 (7). Second, and more importantly, PDZ domains do not fall into

discrete classes but instead lie on a continuum. Indeed, the canonical classes lie only in select portions of this continuum (i.e., at the extremes of the first and third principal axes). Moreover, the PDZ domains represented in our model are evenly distributed throughout selectivity space (Fig. 3G). Zarrinpar *et al.* previously showed that the 23 Src homology 3 domains in *S. cerevisiae* are optimized to avoid cross-reactivity with the mitogen-activated protein kinase signaling protein Pbs2 (32). Here, we find on a much broader scale that a similar principle is in effect among mouse PDZ domains and their ligands. Although the selectivity of protein-protein interactions
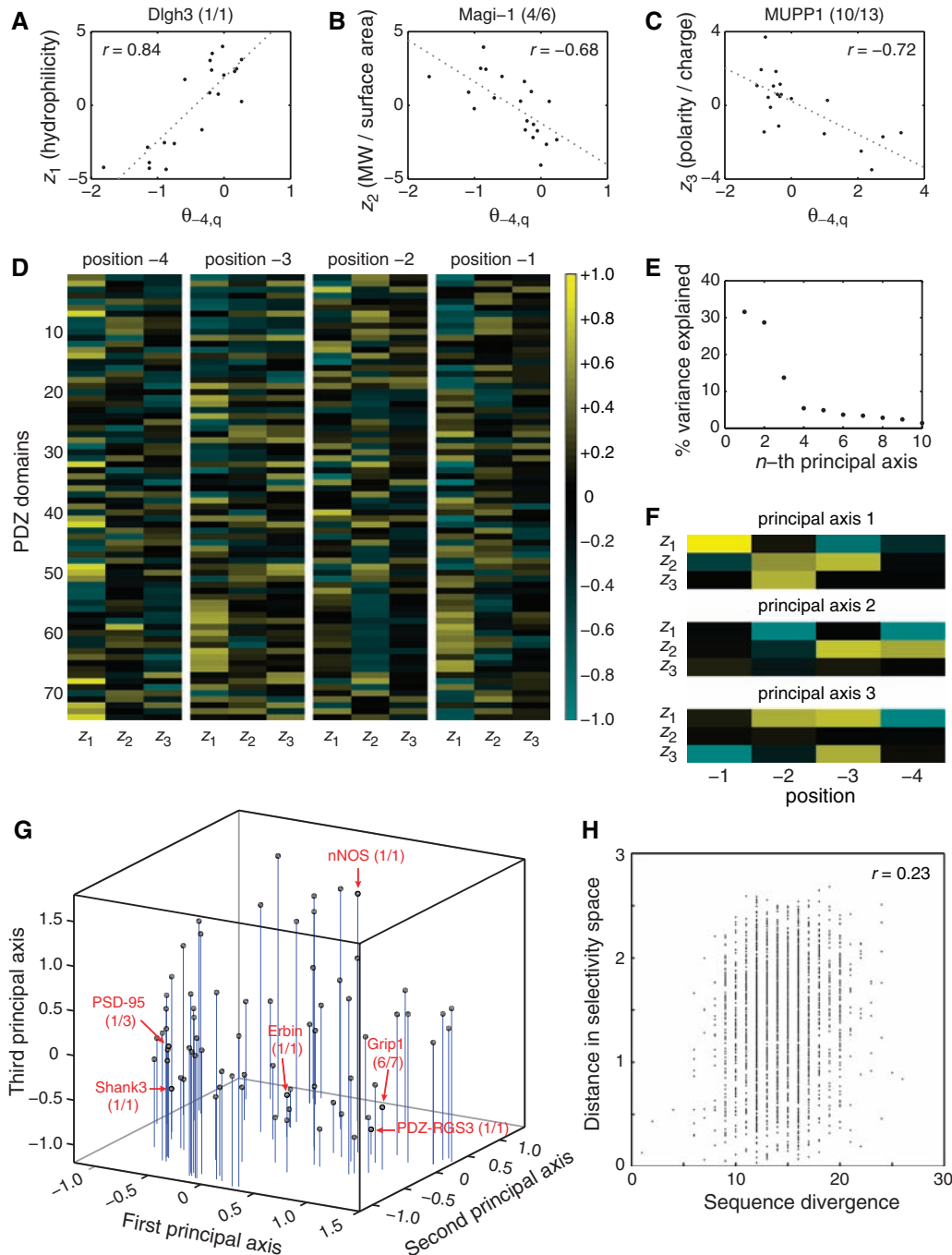

Fig. 3. (**A** to **C**) Correlations between z scales and model parameters at position –4 for three PDZ domains. (**A**) z_1 positively correlates with $\theta_{-4,q}$ for Dlgh3 (1/1). (**B**) z_2 negatively correlates with $\theta_{-4,q}$ for Magi-1 (4/6). (**C**) z_3 negatively correlates with $\theta_{-4,q}$ for MUPP1 (10/13). (**D**) Correlation matrix between the model parameters for all 74 PDZ domains at positions –4, –3, –2, and –1 and the first three z scales of the amino acids. (**E**) Percentage of variance in the correlation matrix that is explained by the 12 principal axes identified through singular-value decomposition. (**F**) Graphical representation of the first three principal axes, used to define PDZ domain selectivity space. (**G**) Distribution of the 74 PDZ domains in selectivity space. Selected PDZ domains are shown, representing class I domains [PSD-95 (1/3) and Shank3 (1/1)], class II domains [Grip1 (6/7) and PDZ-RGS3 (1/1)], and class III domains [nNOS (1/1)]. Erbin (1/1), which has been described as a dual-specificity domain, lies between the class I and class II domains. (**H**) Correlation between pairwise sequence divergence of PDZ domains and their pairwise distances in selectivity space. Sequence divergence was obtained from pairwise alignments performed with Vector NTI version 8 (InforMax, Invitrogen Life Science Software, Frederick, Maryland), using the blosum62mt2 matrix. Pairwise distances in selectivity space are Euclidean distances obtained from the three-dimensional plot in (**G**).

could, in a multicellular organism, be controlled at the level of gene coexpression and protein colocalization, our results indicate that the intrinsic selectivity of PDZ domains is tuned across the mouse proteome to minimize cross-reactivity.

Finally, we observed only a weak correlation (correlation coefficient $r = 0.23$) between the pairwise sequence divergence of PDZ domains and their distances in selectivity space (Fig. 3H). Similarity at the overall sequence level is thus a poor predictor of PDZ domain function. This low correlation suggests that most of the sequence variation among PDZ domains is neutral with respect to peptide-binding selectivity and that only a subset of residues—presumably in the binding pocket of the PDZ domain—is responsible for the distribution of PDZ domains in selectivity space.

References and Notes

1. T. Pawson, *Nature* **373**, 573 (1995).
2. T. Pawson, P. Nash, *Science* **300**, 445 (2003).
3. E. Kim, M. Niethammer, A. Rothschild, Y. N. Jan, M. Sheng, *Nature* **378**, 85 (1995).
4. H. C. Kornau, L. T. Schenker, M. B. Kennedy, P. H. Seeburg, *Science* **269**, 1737 (1995).
5. C. Nourry, S. G. N. Grant, J.-P. Borg, *Sci. STKE* **2003**, re7 (2003).
6. B. Z. Harris, F. W. Lau, N. Fujii, R. K. Guy, W. A. Lim, *Biochemistry* **42**, 2797 (2003).
7. Z. Songyang et al., *Science* **275**, 73 (1997).
8. I. Bezprozvanny, A. Maximov, *FEBS Lett.* **509**, 457 (2001).
9. E. Song et al., *Mol. Cell. Proteomics* **5**, 1368 (2006).
10. I. Letunic et al., *Nucleic Acids Res.* **34**, D257 (2006).
11. J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857 (1998).
12. The "genomic" mode of the Simple Modular Architecture Research Tool (SMART) database (*10, 11*) currently lists 240 PDZ domains identified from the mouse genome sequence. We obtained sequence-verified clones for 203 of them. In addition, we cloned 18 PDZ domains that are listed only in the "normal" mode of the SMART database.
13. G. Fuh et al., *J. Biol. Chem.* **275**, 21486 (2000).
14. Y. Zhang et al., *J. Biol. Chem.* **281**, 22299 (2006).
15. We derived 57 of the training-set peptides from proteins that had previously been shown to interact with PDZ domains. To allow for the possibility of discovering sequences that fall outside the established view of peptide-binding selectivity, we derived the other peptides from different members of 13 families of membrane proteins, regardless of whether their C termini feature canonical PDZ domain binding motifs (table S2).
16. R. B. Jones, A. Gordus, J. A. Krall, G. MacBeath, *Nature* **439**, 168 (2006).
17. M. A. Stiffler, V. P. Grantcharova, M. Sevecka, G. MacBeath, *J. Am. Chem. Soc.* **128**, 5913 (2006).
18. Data for each PDZ-peptide combination [FP, recorded as millipolarization (mP) units] were fit to an equation that describes saturation binding, as previously noted (*17*). Interactions were scored as "positive" if all three of the following criteria were met: (i) The data fit well to the equation ($r^2 > 0.95$); (ii) the difference between FP at 20 µM PDZ domain and FP at 0 µM PDZ domain was >15 mP units; and (iii) the K_d was <100 µM.
19. To estimate the false-negative rate of our microarray assay, we randomly selected 32 PDZ domains and 32 peptides. We then screened all 1024 possible interactions with the use of a single-point FP assay and determined the K_d for all positive interactions. A comparison of the resulting interaction matrix with the microarray data showed a false-negative rate of 6.6%.
20. J. C. Obenauer, L. C. Cantley, M. B. Yaffe, *Nucleic Acids Res.* **31**, 3635 (2003).
21. M. Sandberg, L. Eriksson, J. Jonsson, M. Sjostrom, S. Wold, *J. Med. Chem.* **41**, 2481 (1998).
22. Materials and methods are available as supporting material on *Science* Online.
23. S. van Dongen, thesis, University of Utrecht, Netherlands (2000).
24. Full-length sequences of 31,302 unique mouse proteins (including splicing variants) were downloaded with BioMart from data set NCBIM36 (*Mus musculus* genes) of Ensembl 44. The C-terminal sequence of each entry was extracted using a Python script.
25. Interactions in table S5 are model predictions with *m* set to 20. On the basis of the results of our model validation efforts, we estimate these predictions to have a true-positive rate of 35%, a false-positive rate of 7%, and a TP/FP ratio of 0.83.
26. This single-point assay measures the difference between FP at 20 nM peptide, 20 µM PDZ domain and FP at 20 nM peptide, 0 µM PDZ domain. An analysis of 1710 FP titration curves shows that applying a threshold of 40 mP units to this single-point assay correctly identifies 91% of the positives and 96% of the negatives (fig. S3). Thus, instead of performing an additional 1170 titration curves, we used this single-point assay with a threshold of 40 mP units to evaluate interactions between the 48 test peptides and the 74 PDZ domains in the MDSM.
27. R. Jansen et al., *Science* **302**, 449 (2003).
28. A. Tropsha, *Annu. Rep. Comput. Chem.* **2**, 113 (2006).
29. G. Birrane, J. Chung, J. A. Ladias, *J. Biol. Chem.* **278**, 1399 (2003).
30. F. Jaulin-Bastard et al., *J. Biol. Chem.* **277**, 2869 (2002).
31. F. Jaulin-Bastard et al., *J. Biol. Chem.* **276**, 15256 (2001).
32. A. Zarrinpar, S. H. Park, W. A. Lim, *Nature* **426**, 676 (2003).
33. We thank A. Tropsha for valuable suggestions and the Faculty of Arts and Sciences Center for Systems Biology for support with instrumentation and automation. This work was supported by awards from the Smith Family Foundation, the Arnold and Mabel Beckman Foundation, and the W. M. Keck Foundation and by a grant from the NIH (1 RO1 GM072872-01). M.A.S. was supported in part by the NIH Molecular, Cellular, and Chemical Biology Training Grant (5 T32 GM07598-25), and J.R.C. was the recipient of a Corning CoStar fellowship.

Brain IRS2 Signaling Coordinates Life Span and Nutrient Homeostasis

Akiko Taguchi, Lynn M. Wartschow, Morris F. White*

Reduced insulin-like signaling extends the life span of *Caenorhabditis elegans* and *Drosophila*. Here, we show that, in mice, less insulin receptor substrate–2 (Irs2) signaling throughout the body or just in the brain extended life span up to 18%. At 22 months of age, brain-specific *Irs2* knockout mice were overweight, hyperinsulinemic, and glucose intolerant; however, compared with control mice, they were more active and displayed greater glucose oxidation, and during meals they displayed stable superoxide dismutase–2 concentrations in the hypothalamus. Thus, less Irs2 signaling in aging brains can promote healthy metabolism, attenuate meal-induced oxidative stress, and extend the life span of overweight and insulin-resistant mice.

Reaching old age in good health is not just good luck but the result of a favorable balance between hundreds of disease-causing and longevity-promoting genes; regardless, some common mechanisms that influence life span have emerged (*1*). First, calorie restriction reliably increases animal longevity, and second, reduced insulin-like signaling extends life span in *Caenorhabditis elegans* and *Drosophila melanogaster* (*2, 3*). Calorie restriction and reduced insulin-like signaling might be linked because fasting reduces the intensity and duration of insulin secretion required for glucose homeostasis, and reduced insulin-like signaling promotes the expression of antioxidant enzymes that are associated with longevity (*3–5*). Adapting these principles to humans is challenging because calorie restriction is difficult and because reduced insulin-like signaling can be associated with small stature, metabolic disease, and diabetes.

Insulin and insulin-like growth factor-1 (IGF1) bind to receptors on the surface of all cells that phosphorylate tyrosyl residues on the insulin receptor substrates (IRSs)—chico in *Drosophila* and Irs1, -2, -3, and -4 in mammals. This signaling cascade activates the phosphoinositide-3-kinase (Pik3C) and the thymoma viral proto-oncogene Akt, which regulates many cellular processes, including the inactivation of forkhead box O1 (FoxO1) transcription factor (*6*). Reduced *chico* expression decreases brain and body growth while increasing life span up to 50%, which is related to the increased activity of dFOXO in *Drosophila* (*7, 8*). In mice, the deletion of Irs1 reduces body growth and causes hyperinsulinemia, whereas the deletion of Irs2 (*Irs2*$^{-/-}$ mice) reduces brain growth and causes

Howard Hughes Medical Institute, Division of Endocrinology, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: morris.white@childrens.harvard.edu

ERRATUM

Post date 19 October 2007

Reports: "PDZ domain binding selectivity is optimized across the mouse proteome" by M. A. Stiffler *et al.* (20 July 2007, pp. 364369). The position numbers appeared in the wrong order in Fig. 3F. The corrected panel is shown here.
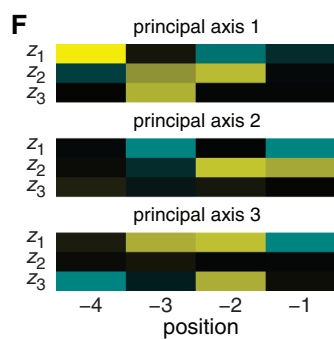

Fig. 3F

LETTERS

edited by Jennifer Sills

Of Aging Mice and Men

LIU *ET AL.* (REPORT, "AUGMENTED WNT SIGNALING in a mammalian model of accelerated aging," 10 August, p. 803) have elegantly shown how alterations in Wnt signals contribute to the suffering of klotho-deficient mice, but not every sick little rodent is a suitable model for human aging. The pathological features and short life span of klotho mutant mice have been shown to reflect hypervitaminosis D, secondary to ablated responses to Fgf-23 (*1*–*3*). The same syndrome appears in Fgf-23 mutants and can be cured by deleting the 1-α-hydroxylase gene that increases the activity of the vitamin. In both mutants, the features represented as evidence of "premature aging" can be eliminated simply by putting the mice on a diet low in vitamin D. Perhaps vitamin D deprivation will turn out to be the long-sought cure for aging, but in the meantime, it would be wise to view with some skepticism the claims that klotho and similar developmental mishaps provide convenient shortcuts for learning about mechanisms of "real" aging.   **RICHARD MILLER**

Geriatrics Center and Department of Pathology, University of Michigan, Ann Arbor, MI 48109–2200, USA.

References
1. M. S. Razzaque, B. Lanske, *Trends Mol. Med.* **12**, 298 (2006).
2. H. Tsujikawa, Y. Kurotaki, T. Fujimori, K. Fukuda, Y. Nabeshima, *Mol. Endocrinol.* **17**, 2393 (2003).
3. B. Lanske, M. S. Razzaque, *Ageing Res. Rev.* **6**, 73 (2007).

Response

THERE ARE MANY AREAS IN AGING RESEARCH IN which there is some disagreement. One question in dispute is the degree to which observations in simple organisms, such as postmitotic worms, can inform our understanding of mammalian aging. Similarly, reasonable people disagree on the role, if any, of cellular senescence in organismal aging. We appreciate that there is also considerable disagreement regarding how much mammalian models of accelerated aging can teach us about the normal aging process.

Our study centered on a set of observations suggesting that the Wnt family of proteins could bind to klotho, a protein whose absence has been linked to an accelerated aging phenotype in mice. Genetic evidence suggests that alleles of klotho are also associated with variation in human longevity (*1*). Nonetheless, we agree with Miller that considerable care must be taken when using the existing accelerated aging models as an indication of the normal aging process. Our opinion is that studying models of rapid aging will be useful in teasing out the underlying mechanisms of how we age, although we understand that Miller does not share that opinion. Hopefully, we will all live long enough to find out who is right.

HONGJUN LIU AND TOREN FINKEL

Cardiology Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA.

Reference
1. D. E. Arking *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 856 (2002).

Replicating Genome-Wide Association Studies

GENOME-WIDE ASSOCIATION STUDIES PROMISE to significantly expand our knowledge of host control of deadly pathogens. Lurking in the background of these studies, however, is a serious methodologic issue. Individuals who participate in the cohorts used in genome-wide association studies are often ethnically and racially different from their fellow citizens who do not participate in these studies (*1*); more important, they are markedly different from the populations of developing countries with the highest burdens of infectious diseases.

The Report "A whole-genome association study of major determinants for host control of HIV-1" (J. Fellay *et al.*, 17 August, p. 944) demonstrates how much can be learned from the study of a highly motivated, largely European cohort. Unfortunately, rather than suggesting that readers strive to replicate the study findings in different populations, J. Fellay *et al.* instead proceed directly to discussion of "directions for therapeutic intervention" and "urgency in carrying out similar studies for other infectious diseases."

In rushing these issues, the authors overlook several important points. Similar studies conducted in different geographic regions may fail to find the same associations, and may even find different associations. The highly polymorphic nature of human MHC, different pathogen strains, or gene-environment interactions could all result in variability of associations across populations in different regions. In some cases, such as the CCR5-Δ32 mutant allele, genetic associations specific to geographic region may indeed aid drug or vaccine target discovery (*2*). However, a premature focus of financial and intellectual resources on a few specific alleles may throw out the baby for the bathwater.

MARK H. KUNIHOLM

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA.
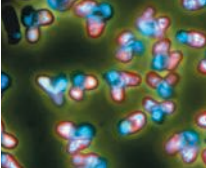
References
1. A. L. Gifford *et al.*, *N. Engl. J. Med.* **346**, 1373 (2002).
2. J. A. Este, A. Telenti, *Lancet* **370**, 81 (2007).

Response

WE RECENTLY REPORTED THAT THREE POLY-morphisms significantly influence host response to HIV-1. Two of these polymorphisms associate with viral load during the asymptomatic set point period, and the third associates with a measure of HIV-1 disease progression.

Kuniholm raises the question of whether the associations could be "replicated" in other geographic regions and suggests that it would have been preferable to evaluate this rather than moving to practical applications of the findings.

Our original study included samples from multiple European populations, both north and south; the effects observed cannot be viewed as the result of a specific cohort or geographic region within Europe. In the original study, we also replicated all three discoveries in a fully independent set of samples.

Kuniholm is correct that genetic effects are sometimes observed in some population groups and not others. The current consensus view is that when a polymorphism is present in different geographic regions, it tends to have a similar effect, but causal variants do vary in frequency among different groups (*1*, *2*). Indeed, one of our associations is known to be rare or absent in some geographic regions. Absence of a relevant genetic variant in a particular population does not in itself limit the applicability of new knowledge: The example of the CCR5-Δ32 variant illustrates this point by demonstrating that a medication of universal use can indeed be developed on the basis of genetic information from one human population. The question of the geographic distribution of causal polymorphisms is an important one, but it is separate from the question of whether the polymorphisms have important clinical effects in the groups under study. Indeed, we are currently expanding our study to include multiple cohorts from the United States and from Africa.

DAVID B. GOLDSTEIN

Center for Population Genomics and Pharmacogenetics, Duke Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710, USA.

References
1. J. P. Ioannidis, E. E. Ntzani, T. A. Trikalinos, *Nat. Genet*. **36**, 1312 (2004).
2. D. B. Goldstein, J. N. Hirschhorn, *Nat. Genet*. **36**, 1243 (2004).

A Measure of Respect for Translational Research

IT WAS A PLEASANT SURPRISE TO SEE A SPECIAL feature ("Careers in translational research," 17 August, p. 966) in *Science* focusing on translational research and its opportunities, risks, and challenges. In their respective articles, S. Carpenter (p. 966) and K. Garber (p. 968) highlight the concerns that translational researchers have about not being able to satisfy traditional measures of scientific success, including number of publications and impact factors. This apprehension is well founded. More worrisome is the scenario in which the onus is placed on the members of the translational community to prove their worth. I think it is too much. Measures of basic research productivity are well established, but the same is not true for translational research. I agree with Wu's advice (p. 967) to "go with what you passionately care about, because it's a long row, no matter how you hoe it," but I also sympathize with June's lament (p. 969): "I have seen several instances since I've been at [Penn] where promising translational researchers had to go back and just do basic research in order to assure their promotion."

It is time to think seriously about how to develop criteria for quantitatively evaluating

translational work. "Bench-to-bedside" and "lab-to-clinic" research will otherwise suffer from a perennial problem of lack of recognition. Considering the risk of failure in translational research, we need to be open-minded and adopt measures that focus not only on success but on honest effort. Achievements such as partnering, patents, clinical trials, and drug screening should be considered on par with publications for assessment and promotion. Successes in translational efforts should be provided with "impact factors" commensurate with the volume of work, time taken, or importance in terms of clinical or pharmacological utility.

ABHAY SHARMA

Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India.

CORRECTIONS AND CLARIFICATIONS

News Focus: "Accidents spur a closer look at risks at biodefense labs" by J. Kaiser (28 September, p. 1852). The highest biocontainment level is "biosafety level 4," not "biosecurity level 4," as stated in the article.

News of the Week: "Lapses in biosafety spark concern" by J. Couzin (14 September, p. 1487). A report by the Centers for Disease Control and Prevention (CDC) incorrectly noted that its last inspection of Texas A&M University's biosafety program prior to July had been in February 2007. CDC has since noted that this was a typo. The inspection took place in February 2006.

Special Issue on Attosecond Spectroscopy: Reviews: "The future of attosecond spectroscopy" by P. H. Bucksbaum (10 August, p. 766). In the second line of the legend to Fig. 1, the phrase "two cojoined coins" should read "two cojoined cones."

Reports: "PDZ domain binding selectivity is optimized across the mouse proteome" by M. A. Stiffler et al. (20 July, p. 364). The position numbers appeared in the wrong order in Fig. 3F. The corrected panel is shown here.

Reports: "Genome plasticity a key factor in the success of polyploid wheat under domestication" by J. Dubcovsky and J. Dvorak (29 June, p. 1862). In the final reference, National Research Institute should have been National Research Initiative.

Reports: "Thrice out of Africa: Ancient and recent expansions of the honey bee, *Apis mellifera*" by C. W. Whitfield et al. (27 October 2006, p. 642). Several critical references were left out of the final manuscript. In discussing the fact that "ample evidence shows that both European and African alleles occur in Africanized populations" (p. 644), we should have referenced two studies that first demonstrated introgression between invading Africanized and resident European honey bees in Texas: M. Pinto et al., *Evolution* **58**, 1047 (2004) and M. Pinto et al., *Genetics* **170**, 1653 (2005). In addition, these studies showed no differences between mitotypes of Africanized and European bees in the later years of Africanization. References to this conclusion should have been cited at the end of the first paragraph on p. 645. We greatly regret that these references were omitted, and for this we extend our apologies to Pinto et al. The North American portion of this effort was built upon the Pinto et al. work. It was only because we could make use of many of the same bees used in the Pinto et al. study that we were able to corroborate the results of Pinto et al. and then expand on them, showing that the lack of correlation between mtDNA and nuclear DNA involved markers distributed throughout the nuclear genome, and examining in more detail the relationships between M-, C-, O-, and A-derived genomes.

F — principal axis 1, principal axis 2, principal axis 3; z_1, z_2, z_3; position −4, −3, −2, −1
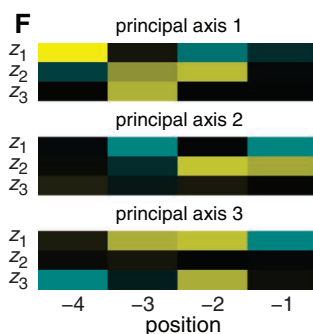
TECHNICAL COMMENT ABSTRACTS

Comment on "Human Neuroblasts Migrate to the Olfactory Bulb via a Lateral Ventricular Extension"

Nader Sanai, Mitchel S. Berger, Jose Manuel Garcia-Verdugo, Arturo Alvarez-Buylla

Curtis et al. (Research Articles, 2 March 2007, p.1243) claimed discovery of a human neuronal migratory stream to the olfactory bulb along a putative lateral ventricular extension. However, high levels of proliferation reported with proliferating cell nuclear antigen were not confirmed using different markers, neuronal chain migration was not demonstrated, and no serial reconstruction shows a true ventricular extension.

Full text at www.sciencemag.org/cgi/content/full/318/5849/393b

Response to Comment on "Human Neuroblasts Migrate to the Olfactory Bulb via a Lateral Ventricular Extension"

Maurice A. Curtis, Monica Kam, Ulf Nannmark, Richard L. M. Faull, Peter S. Eriksson

In contrast to a previous study of Sanai et al., our study had the advantage of using serial sagittal sections of the human basal forebrain, combined with 5-bromo-2'-deoxyuridine labeling, rigorous magnetic resonance imaging, and polymerase chain reaction analysis. We believe these methods convincingly demonstrate the presence of a rostral migratory stream in the human brain that resembles that in other mammals.

Full text at www.sciencemag.org/cgi/content/full/318/5849/393c