

Termination mechanisms. On the basis of our data, we envision that rising insolation triggers the initial disintegration of a massive, isotropically compensated ice sheet, which in turn triggers a slowing of MOC and hence a lowering of surface-ocean heat flux to the North Atlantic. Along with sea-ice formation, this collapse generates a cold anomaly in the North Atlantic, which weakens the AM through atmospheric teleconnections (26, 27) and also moves the Intertropical Convergence Zone (ITCZ) to the south (44–46). Antarctic temperature increase could result from CO₂ rise, from the bipolar seesaw mechanism (20, 47–51), and from southward shifts in atmospheric circulation patterns (52).

A number of mechanistic ties between this set of events and CO₂ rise seem plausible. First, simple southward movement of climatic zones [observed for ITCZ (45) and southern Brazil (52)] could include a southward shift in the westerlies (53), resulting in enhanced wind-driven upwelling in the ocean around Antarctica, promoting ventilation of respired CO₂, atmospheric CO₂ rise, and observed productivity peaks (54). Second, warming from the bipolar seesaw mechanism could melt sea ice in the Southern Ocean, also promoting CO₂ ventilation (55). Third, warming associated with southerly shifts in climate zones could reduce Patagonian glaciation, lowering the flux of dust and iron from Patagonia to the Southern Ocean, reducing the efficiency of the biological pump (56). There are limits (imposed by bounds on the glacial-interglacial change in the carbonate compensation depth) on the extent to which alkalinity-based mechanisms can contribute to the CO₂ rise (57, 58). However, within these limits, it is plausible that alkalinity-based mechanisms may contribute. Given the broad synchrony between terminations and CO₂ rise, alkalinity-based feedbacks between sea level and atmospheric CO₂, such as the coral reef hypothesis (59), may well contribute once the sea level begins to rise. Archer *et al.* (57) argued that no single mechanism could explain the full glacial-interglacial range in CO₂. Here, we present a scenario in which CO₂ rise could be caused by a set of mechanisms all ultimately linked to the rise in boreal summer insolation. Both rising insolation and rising CO₂ (60–62), generated with multiple positive feedbacks, drove the termination.

References and Notes

1. C. Emiliani, *J. Geol.* **63**, 538 (1955).
2. W. S. Broecker, J. van Donk, *Rev. Geophys. Space Phys.* **8**, 169 (1970).
3. J. D. Hays, J. Imbrie, N. J. Shackleton, *Science* **194**, 1121 (1976).
4. Materials and methods are available as supporting material on Science Online.
5. H. Heinrich, *Quat. Res.* **29**, 142 (1988).
6. H. Cheng *et al.*, *Geology* **34**, 217 (2006).
7. Y. J. Wang *et al.*, *Science* **294**, 2345 (2001).
8. J. F. McManus, D. W. Oppo, J. L. Cullen, *Science* **283**, 971 (1999).
9. M. J. Kelly *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **236**, 20 (2006).
10. C. A. Dykoski *et al.*, *Earth Planet. Sci. Lett.* **233**, 71 (2005).
11. Y. J. Wang *et al.*, *Nature* **451**, 1090 (2008).

12. C. H. Hendy, *Geochim. Cosmochim. Acta* **35**, 801 (1971).
13. W. Dansgaard, *Tellus* **16**, 436 (1964).
14. K. R. Johnson, B. L. Ingram, *Earth Planet. Sci. Lett.* **220**, 365 (2004).
15. D. Fleitmann *et al.*, *Quat. Sci. Rev.* **26**, 170 (2007).
16. A. Sinha *et al.*, *Geology* **33**, 813 (2005).
17. C. Hu *et al.*, *Earth Planet. Sci. Lett.* **266**, 221 (2008).
18. R. B. Alley *et al.*, *Nature* **362**, 527 (1993).
19. G. Bond *et al.*, *Nature* **365**, 143 (1993).
20. W. S. Broecker, D. M. Peteet, D. Rind, *Nature* **315**, 21 (1985).
21. J. F. McManus, R. Francois, J. M. Gherardi, L. D. Keigwin, S. Brown-Leger, *Nature* **428**, 834 (2004).
22. J. C. H. Chiang, M. Biasutti, D. S. Battisti, *Paleoceanography* **18**, 1094 10.1029/2003PA000916 (2003).
23. G. H. Denton, R. B. Alley, G. C. Comer, W. S. Broecker, *Quat. Sci. Rev.* **24**, 1159 (2005).
24. W. S. Broecker, *Global Planet. Change* **54**, 211 (2006).
25. H. F. Blanford, *Proc. R. Soc. London* **37**, 3 (1884).
26. T. P. Barnett, L. Dümenil, U. Schlese, E. Roeckner, *Science* **239**, 504 (1988).
27. R. Zhang, T. L. Delworth, *J. Clim.* **18**, 1853 (2005).
28. G. H. Denton, W. S. Broecker, R. B. Alley, *Pages News* **14**, 14 (2006).
29. A. L. Berger, *Quat. Res.* **9**, 139 (1978).
30. J. Imbrie *et al.*, in *Milankovitch and Climate, Part I*, A. Berger, J. Imbrie, J. Hays, G. Kukla, B. Saltzman, Eds. (Reidel, Norwell, MA, 1984), pp. 269–305.
31. L. G. Thompson *et al.*, *Science* **298**, 589 (2002).
32. R. B. Alley, P. U. Clark, P. Huybrechts, I. Joughin, *Science* **310**, 456 (2005).
33. G. E. Birchfield, W. S. Broecker, *Paleoceanography* **5**, 835 (1990).
34. W. R. Peltier, *Science* **265**, 195 (1994).
35. M. E. Raymo, *Paleoceanography* **12**, 577 (1997).
36. J. R. Petit *et al.*, *Nature* **399**, 429 (1999).
37. K. Kawamura *et al.*, *Nature* **448**, 912 (2007).
38. L. Loulergue *et al.*, *Nature* **453**, 383 (2008).
39. J. Jouzel *et al.*, *Science* **317**, 793 (2007).
40. M. Suwa, M. Bender, *Quat. Sci. Rev.* **27**, 1093 (2008).
41. N. J. Shackleton, *Science* **289**, 1897 (2000).
42. J. P. Severinghaus *et al.*, *Science* **324**, 1431 (2009).
43. M. Bender, T. Sowers, L. Labeyrie, *Global Biogeochem. Cycles* **8**, 363 (1994).
44. L. C. Peterson, G. H. Haug, K. A. Hughen, U. Röhl, *Science* **290**, 1947 (2000).
45. X. F. Wang *et al.*, *Nature* **432**, 740 (2004).
46. J. C. H. Chiang, C. M. Bitz, *Clim. Dyn.* **25**, 477 (2005).
47. T. J. Crowley, *Paleoceanography* **7**, 489 (1992).
48. T. F. Stocker, D. G. Wright, L. A. Mysak, *J. Clim.* **5**, 773 (1992).
49. W. S. Broecker, *Paleoceanography* **13**, 119 (1998).
50. T. Blunier, E. J. Brook, *Science* **291**, 109 (2001).
51. T. F. Stocker, S. J. Johnsen, *Paleoceanography* **18**, 1087 (2003).
52. X. F. Wang *et al.*, *Geophys. Res. Lett.* **34**, L23701 10.1029/2007GL031149 (2007).
53. J. R. Toggweiler, J. L. Russell, S. R. Carson, *Paleoceanography* **21**, 2005 (2006).
54. R. F. Anderson *et al.*, *Science* **323**, 1443 (2009).
55. R. F. Keeling, B. B. Stephens, *Paleoceanography* **16**, 112 (2001).
56. J. H. Martin, S. E. Fitzwater, *Nature* **331**, 341 (1988).
57. D. Archer, A. Winguth, D. Lea, N. Mahowald, *Rev. Geophys.* **38**, 159 (2000).
58. D. M. Sigman, E. A. Boyle, *Nature* **407**, 859 (2000).
59. B. N. Opdyke, J. C. G. Walker, *Geology* **20**, 733 (1992).
60. D. Paillard, *Rev. Geophys.* **39**, 325 (2001).
61. J. P. Severinghaus, *Nature* **457**, 1093 (2009).
62. P. U. Clark, A. M. McCabe, A. C. Mix, A. J. Weaver, *Science* **304**, 1141 (2004).
63. We thank the late Gary Comer for his generous support and J. Severinghaus, V. Masson-Delmotte, and P. Clark for valuable suggestions. This work was supported by U.S. NSF grant 0502535, Gary Comer Science and Education Foundation grant CC8, NOAA grants to G.H.D., and National Natural Science Foundation of China grants 40631003 and 40771009.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5950/248/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2
References

17 June 2009; accepted 14 September 2009
10.1126/science.1177840

Reactome Array: Forging a Link Between Metabolome and Genome

Ana Beloqui,^{1*} María-Eugenia Guazzaroni,^{1*} Florencio Pazos,² José M. Vieites,¹ Marta Godoy,² Olga V. Golyshina,³ Tatyana N. Chernikova,³ Agnes Waliczek,³ Rafael Silva-Rocha,² Yamal Al-ramahi,¹ Violetta La Cono,⁴ Carmen Mendez,⁵ José A. Salas,⁵ Roberto Solano,² Michail M. Yakimov,⁴ Kenneth N. Timmis,^{3,6} Peter N. Golyshin,^{3,7,8,††} Manuel Ferrer^{1††}

We describe a sensitive metabolite array for genome sequence-independent functional analysis of metabolic phenotypes and networks, the reactomes, of cell populations and communities. The array includes 1676 dye-linked substrate compounds collectively representing central metabolic pathways of all forms of life. Application of cell extracts to the array leads to specific binding of enzymes to cognate substrates, transformation to products, and concomitant activation of the dye signals. Proof of principle was shown by reconstruction of the metabolic maps of model bacteria. Utility of the array for unsequenced organisms was demonstrated by reconstruction of the global metabolisms of three microbial communities derived from acidic volcanic pool, deep-sea brine lake, and hydrocarbon-polluted seawater. Enzymes of interest are captured on nanoparticles coated with cognate metabolites, sequenced, and their functions unequivocally established.

Functional genomics has greatly accelerated research on the genomic basis of life processes in health and disease and provided a quantum advance in our understanding of such processes, their regulation, and underlying

mechanisms (1). Functional assignments and metabolic network reconstructions have generally depended on both the genome sequence of the organism(s) in question and bioinformatic analyses based on homology to known genes and proteins

(2–6). However, many genes in databases have questionable annotations or are not annotated at all (7–9), which hinders effective exploitation of the rapidly growing volume of genome sequence data. Metabolomics provides new insights into the metabolic state of a cell under a given set of environmental parameters, or in response to a parameter change, independently of a genome sequence, although problems of metabolite identification and quantification exist (10–12). Functionally associating the metabolic profile obtained with the enzymes and pathways responsible still depends heavily on sequenced-based metabolic reconstructions. There is thus a need for a new method to causally link metabolites with cognate enzymes, which, in addition to delivering global descriptions of metabolic responses to given environmental conditions, simultaneously provides annotation of the enzymes featured. The “reactome array” we describe here forges this link between genome and metabolome and provides a global metabolic phenotype of a cell extract derived from a clonal population of cells or a mixture of cell types, as is found in clone libraries, tissues, or multicellular organisms. The array constitutes a generic tool for metabolic phenotyping of cells and annotation of proteins and has applications in diverse aspects of biology and medicine.

The reactome array. We synthesized 1676 metabolites—known from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (13) to collectively represent central metabolic pathways in all organisms—and 807 other substrate molecules (14). We coupled these to the dye Cy3 and to a linker molecule used to immobilize the compounds on glass slides to produce the array (fig. S1). The compounds printed on the array are characterized by two key features (Fig. 1). One is that the Cy3 dye component of the compound as synthesized is inactive, as a result of intramolecular resonance quenching. Productive catalytic action of an enzyme on the substrate results in the release of the reaction product and the dye, thereby relieving the quenching such that the dye becomes activated and gives a fluorescent signal (movie S1). The arrayed metabolites (~0.7 fmol/spot; $P < 0.011$, $N = 9$) provided detections down to absolute concentrations of 5.7 ng protein/ml ($P < 0.043$, $N = 12$) (fig.

S2). Even at the lowest enzyme concentrations, no signal interference by proteins or other substances in cell lysates down to a target:total protein ratio of 1:10⁶ (total concentration, 100 μg/ml) was detected (fig. S2). Such sensitivity is sufficient for activity determination of enzymes in many types of sample, including clinical samples. Nonproductive interactions of proteins with the metabolites—binding without chemical change—do not release either the substrate or the dye and thus do not relieve dye quenching. The reactome array contained a total number of 2483 metabolites (tables S1 and S2) that collectively serve as substrates in all metabolic pathways figuring in the KEGG and PubMed databases and the University of Minnesota Biocatalysis/Biodegradation Database.

The nature of the linker molecule, a Co(II) complex containing a poly(A) tail, is such that productive reaction of an enzyme with the substrate leads to release of both histidine “tags” anchored to the Co(II), thereby exposing an active cobalt cation which ligates and immobilizes the enzyme on the array spot (Fig. 1 and fig. S1). This feature is exploited to capture for characterization enzymes determined from the array as being of interest. In this case, individual compounds are attached via their linkers to gold nanoparticles (Fig. 2A), which serve as high-capacity capture probes for cognate enzymes (a concentration of 9×10^{10} particles/ml of diameter $\sim 2.9 \pm 0.8$ nm, corresponds to a surface area of $\sim 141 \pm 3$ cm²/ml that displays 62.5 pmol of substrate with a binding capacity of 3 to 18 pmol protein per ml). Nanoparticles are mixed with the cell lysate to allow enzyme reaction with the substrate; this in

turn results in its selective capture by the nanoparticle, which is readily recovered. The captured protein is then released in essentially pure form for characterization by imidazole treatment of the nanoparticle (14).

The reactomes of *Pseudomonas putida* and *Streptomyces coelicolor*. To validate the reactome array, we used it to determine the metabolic profiles of cultures of the bacteria *P. putida* strain KT2440 (Pp) and *S. coelicolor* strain M145 (Sc). The genomes of both organisms have been sequenced and, according to current annotations, of the 1676 KEGG metabolites printed in the array, 595 should be metabolized by Pp and 598 by Sc, whereas 1081 and 1033, respectively, should not (13). The reactome profiles are displayed in fig. S3. The fluorescence signal (table S3; $P < 0.0255$, $N = 9$) was used as the score for the receiver operating characteristic (ROC) analysis (15), based on the assumption of a positive relationship between spot intensity in the array and metabolism of the compound by Pp or Sc extracts (figs. S4 and S5). Over an optimal value of 1.0, the score obtained of 0.74 ($P = 1.73e^{-71}$) and 0.85 ($P = 5.77e^{-172}$), for Pp and Sc, respectively, indicates that the array readily discriminates compounds metabolized by extracts of both organisms from those that are not. These values are reasonably high, given that the composition of the array is based on the assumption that the reactions carried out by Pp and Sc are those annotated in KEGG, and that this is obviously not true, because the reactome array has revealed the existence of reactions not annotated in KEGG.

Reactome-based (re)annotations. At present, the rate of generation of genomic sequence data

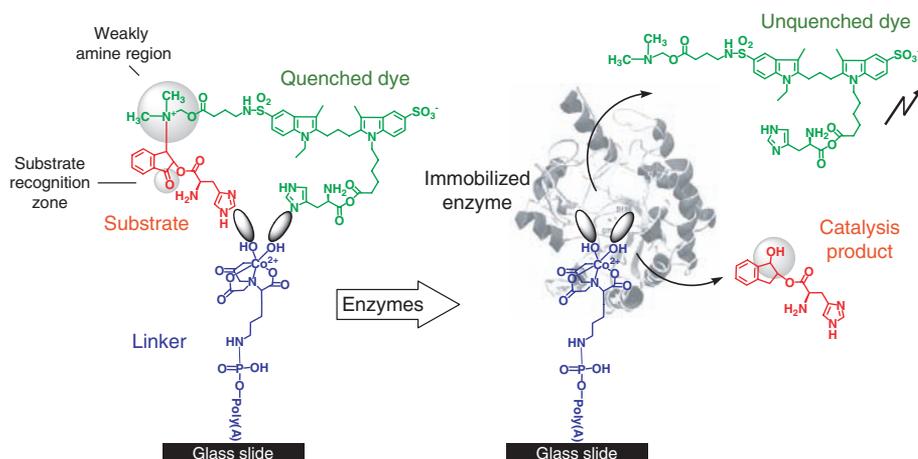


Fig. 1. The reactome strategy. The generic structure of reactome metabolites involves three linked components: the enzyme substrate-metabolite, the quenched dye, and the linker used to immobilize the complex on the array or on nanoparticles. The substrate-metabolite is linked to the quenched dye through a labile nitrogen bond, and both the dye and the substrate are anchored to the Co(II)-containing poly(A) linker by histidine “tags.” Details of the synthetic strategy are provided in fig. S1. An enzyme-catalyzed chemical change in the substrate at a position adjacent to the weakly amine region causes rupture of the labile nitrogen:metabolite bond and release of the quenched Cy3 dye. This in turn provokes release of the reaction product and the histidine “tags” anchored to the Co(II), thereby exposing an active cobalt cation that ligates and immobilizes the enzyme on the array spot. The released dye is no longer quenched and gives a fluorescent signal. The nature of the reaction and the catalysis product is defined by the position to which the quenched dye and the substrate are linked (table S2).

¹CSIC, Institute of Catalysis, 28049 Madrid, Spain. ²CSIC, Centro Nacional de Biotecnología, 28049 Madrid, Spain. ³HZI-Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany. ⁴Institute for Coastal Marine Environment, CNR, Messina 98122, Italy. ⁵Universidad de Oviedo, 33006 Oviedo, Spain. ⁶Institute of Microbiology, Carolo-Wilhelmina Technical University of Braunschweig, 38106 Braunschweig, Germany. ⁷School of Biological Sciences, Bangor University, Gwynedd LL572 UW, UK. ⁸Centre for Integrated Research in the Rural Environment, Aberystwyth University-Bangor University Partnership (CIRRE).

*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: mferrer@icp.csic.es (M.F.); p.golysin@bangor.ac.uk (P.N.G.)

outstrips our ability to analyze and usefully exploit this data (16). There is a lack of functional information about many open reading frames (ORFs) and an incorrect annotation of some others (7). To establish sequence: function relationships for proteins reacting with the array, we created a suite of 528 batches of gold nanoparticles, each batch covered with an individual metabolite found to be metabolized by the *P. putida* extract and collectively representing the entire Pp reactome. Each type of nanoparticle was separately reacted with the Pp cell lysate and recovered by centrifugation, and then the unbound enzyme was removed by washing with buffer. Captured enzyme(s) was released by imidazole and identified by trypsin digestion and mass spectrometry peptide sequencing (Fig. 2A). 549 proteins were captured by the nanoparticles, of which 191 enzymes acting on 158 of the 525 metabolites were unambiguously identified (table S4) and their functions assigned.

As shown in the left panel of Fig. 2B, overall the captured enzymes exhibited a noticeable bias toward functions relating to xenobiotic (33%) and amino acid metabolism (31%). New functional assignments and annotations were defined for 31 enzymes (16%) that had been previously predicted from genome sequence analysis but not characterized as proteins (Fig. 2B, middle panel). These include oxidoreductases, synthases, ligases, kinases, deacetylases, transferases, decarboxylases, nucleosidases, and DNA-metabolizing proteins and were assigned to pathways for xenobiotic (55%), amino acid (19%), carbohydrate (11%), and secondary metabolite (7%) metabolism. Biochemical characterization of 23 out of 31 hypothetical proteins identified by array mapping and further captured in essentially pure form confirmed their activities (Fig. 2C and fig. S6A). This analysis revealed enzymatic potential for *p*-cymene and 4-chlorobiphenyl metabolisms in *P. putida* that have not thus far been predicted by genome analysis of Pp (fig. S6B). The analysis of a further 25% resulted in improved annotations, and the activities of the remaining 59% of enzymes were consistent with current annotations in the databases (Fig. 2B, right panel). All together, these findings expand the extensive potential catabolic landscape of this bacterium (17) and will facilitate future genome annotations, because the majority of these genes are highly conserved in other organisms.

The reactomes of microbial communities of diverse habitats. The results obtained with model organisms for which genome sequences are available confirmed that the reactome array constitutes a means of obtaining a genome-wide overview of the metabolic status of a population of cells. Because all metabolic pathways documented in the main databases are represented by metabolites on the reactome array, and because a subset of these—the core metabolites—are characteristic of all life forms, it should enable such overviews to be obtained of the cells of essentially any organism, and even of communities of multiple species. In the latter case, the quality of the metabolic

overview obtained will depend upon the complexity of the natural or experimental community, the sensitivity of the reactome array, and the pos-

sible inclusion on the array of additional metabolites specific to the habitat and metabolism of the habitat community members.

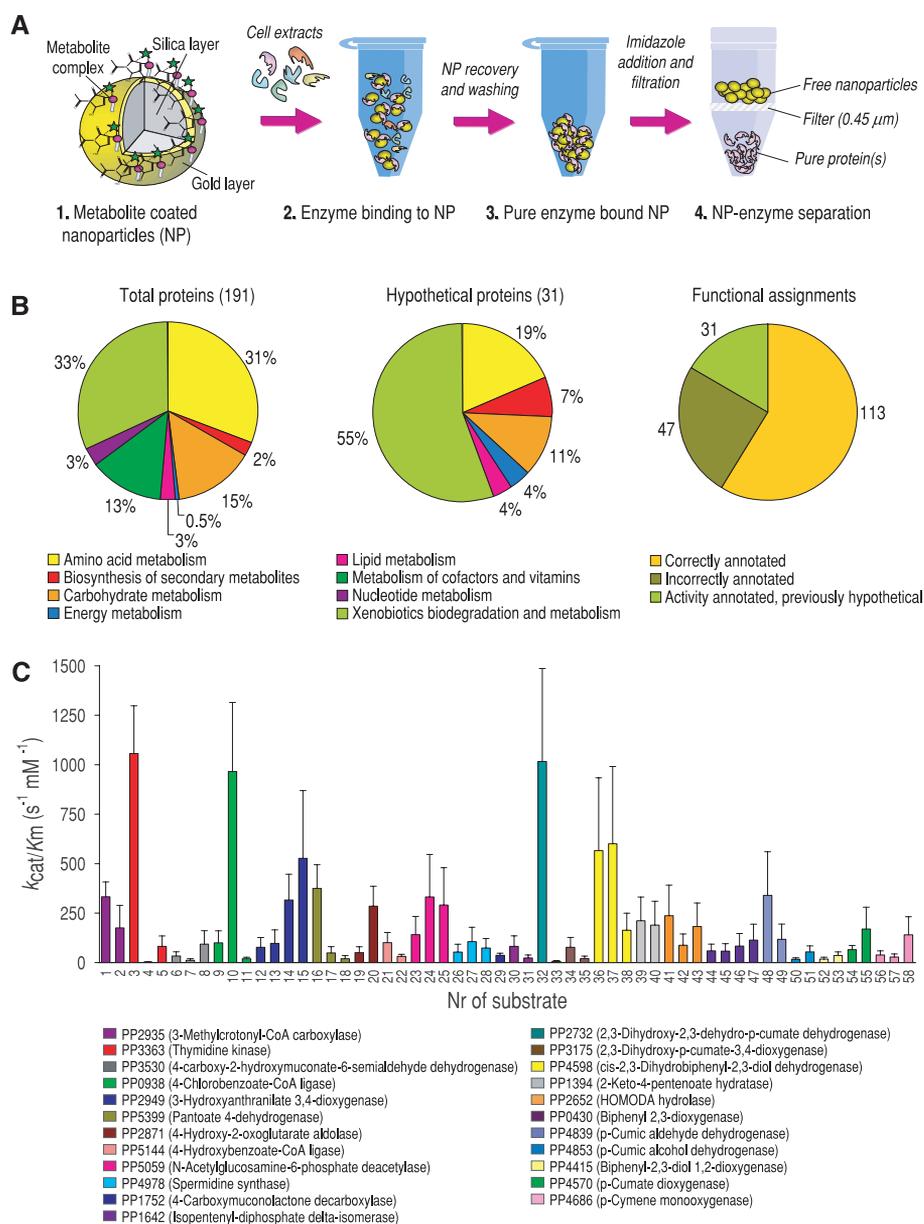


Fig. 2. Nanoparticle capture and functional characterization of hypothetical proteins of *P. putida*. **(A)** The nanoparticle (NP) strategy for protein capture and analysis is summarized in the scheme. NP coated with a single metabolite complex is allowed to react with the cell lysate. An enzyme able to transform the substrate becomes captured by the Co(II) cation of the linker and immobilized on the NP. After recovery of the NP by centrifugation, and washing to remove unbound enzymes, the specifically captured enzyme(s) is released by imidazole treatment of the NP, separated in pure form from the NP by filtration, sequenced, and functionally characterized. In the case where more than one enzyme binds to the substrate on the NP, a further separation will be necessary. **(B)** Functional assignments of *P. putida* proteins detected by the array. On the left is shown the distribution of functional classes of the 191 enzymes captured on NPs and further unambiguously identified and characterized. This provided a direct activity annotation and thereby allowed us to determine which previous genomic annotations were correct. The middle element shows the distribution of the functional classes of the 31 (previously) hypothetical proteins. On the right is shown the distribution of correctly and incorrectly genome-annotated enzymes and of new direct activity annotated proteins. **(C)** Substrate spectra and steady-state k_{cat}/K_m kinetic parameters of 23 previously hypothetical proteins that were NP-capture purified and that are color coded as indicated in the lower part of the figure. Substrate names (14) and K_m and k_{cat} values are shown in fig. S6. Other substrates tested but for which no activity was detected are shown in table S4C. Data represent the mean \pm SEM of four independent measurements.

In some instances, environmental samples will not contain enough biomass for the array analysis, either because organism concentrations are very low or because only small samples can be acquired. In this case, the genetic information present in the samples can be harvested and archived in a laboratory host microbe, like *Escherichia coli* K-12, to create genomic libraries of multiple species, so-called “metagenomic libraries.” We investigated whether useful information on the metabolic profiles of microbial communities can be obtained from such libraries, despite the differences in physiology and expression apparatus of the donor organisms and those of *E. coli*, different

habitat conditions, and so on. If so, this would allow functional exploration of communities characteristic of extreme and difficult-to-sample environments, such as hydrothermal vents, and of slow-growing communities in nutrient-poor habitats.

We obtained samples from three very distinct habitats: a low-nutrient, heavy metal-rich, acidic geothermal pool on Vulcano Island in the Tyrrhenian Sea (VUL); a nutrient-rich, organic pollutant-contaminated surface seawater near Kolguev Island in the Barents Sea (KOL); and the seawater-brine interface of the L’Atalante deep-sea hypersaline, anoxic lake (L’A) in the Eastern Mediterranean Sea. To obtain sufficient genomic DNA for li-

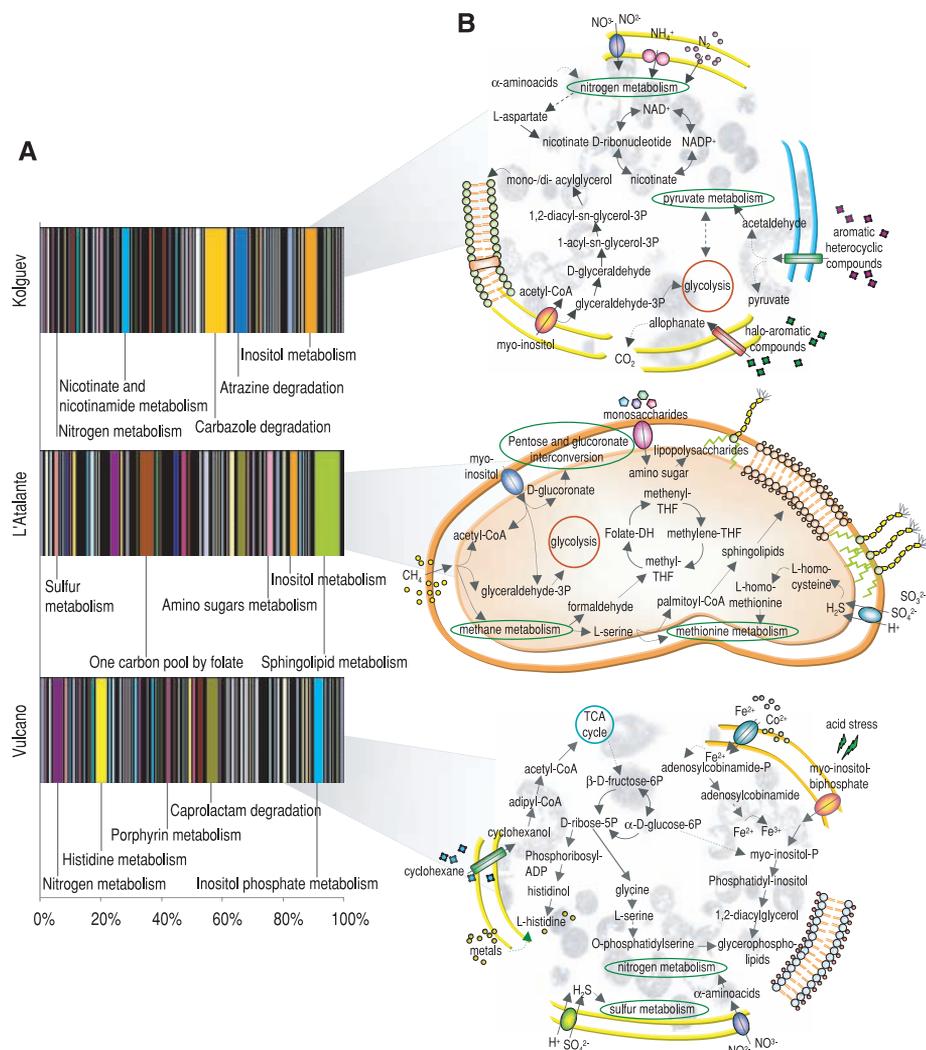


Fig. 3. Functional profiling and metabolic reconstruction of the metagenome library representations of the L’A, KOL, and VUL communities. **(A)** We used the relative proportion of z scores (log scale) provided in fig. S10 to rank KEGG functions in the three metagenome libraries. The detailed metabolic profiles (reactome “barcodes”) are clearly distinct; the most distinctive metabolic characteristics are labeled. **(B)** Depiction of the distinguishing metabolic features of the communities in the context of their pan-metabolic networks. The L’A community consists essentially of two related species, so we have represented its metabolic network in the context of a single cell, whereas the other communities are diverse, with their pan-metabolic networks represented as a “meta cell,” with fragmented membranes indicating facile exchange of metabolites between different community members of the “multicellular community organism.” The metabolic networks depicted of the L’A, KOL, and VUL communities were constructed from the reactome array data from the 1676 metabolites that can be automatically found in KEGG. Tight coupling between major metabolic pathways is indicated.

brary construction, we stimulated multiplication of the cells in the samples by adding growth substrates: Fe(II) to 100 mM plus yeast extract to 0.02% (VUL); crude oil to 0.2% (KOL); and D-glucose to 2 mM (L’A), and subsequently harvested the resulting microbial communities, extracted their total DNA, and established metagenome libraries in *E. coli*. To determine which species of microorganisms were present in each community, we used the DNA obtained from each sample as substrate for amplification of 16S rRNA genes by polymerase chain reaction (PCR). The amplicons thereby obtained were cloned to produce 16S rRNA gene libraries, and representative numbers of the clones were sequenced and taxonomically assigned through use of standard GenBank database homology search tools and phylogeny analysis packages (14). The compositions of the three communities are depicted in fig. S7. The VUL community was not very diverse and was dominated by acidophilic iron-oxidizing eubacterial species of *Acidithiobacillus* and acidophilic species of heterotrophic *Bacillus* and *Alicyclobacillus*. Species of the extremely acidophilic archaeon *Ferroplasma* were present in small amounts. In contrast, the KOL community was diverse and included various petroleum-degrading organisms belonging to the gamma- and epsilon-proteobacteria, such as species of *Pseudomonas*, the hydrocarbonoclastic bacterium *Thalassolituus*, and the denitrifying bacterium *Arco bacter*. The least diverse community was that of L’A, which consisted essentially of two strains distantly related to *Halanaerobium kushneri*.

The pan-reactomes of the three microbial communities were determined by application of cell lysates of pooled clones to the arrays (table S3; $P < 0.0381$, $N = 9$). Cell lysates of the parental, nonrecombinant *E. coli* strain used for library construction, and grown and treated under the same conditions, were used to produce a baseline pattern that was subtracted from those obtained with the library lysates to give the pan-reactome patterns (fig. S8). The VUL reactome consisted of enzymatic reactions involving 807 substrates, the KOL reactome involved 1493 compounds, and the L’A reactome included 2386 compounds (fig. S9). A total of 484 substrates were metabolized by extracts of all three communities, but the detailed metabolic profiles were clearly distinct (Fig. 3 and figs. S10 to S12), presumably reflecting niche-specific metabolic profiles expected of communities occupying distinct habitats.

The L’Atalante hypersaline lake has a high density, and its surface acts as a trap for organic detritus (“marine snow”) and larger bioorganic material sedimenting through the water column (18). The major metabolic activities within the anoxic brine are sulfate reduction and methanogenesis (16). Metabolism in the brine-seawater interface is therefore expected to reflect (i) the slow input of diverse organic materials from above; (ii) the input of methane and sulfide from below, and other electron acceptors and donors and various metals, whose concentrations vary

steeply across the interface and which are cycled through various redox states and balances by the stratified microbial communities; and (iii) the steep gradients of salinity and redox potential.

Some of the features that distinguish the L'A reactome from those of VUL and KOL are high activity levels of sulfur, one-carbon pool-folate, amino-sugar and sphingolipid metabolisms, and low-nitrogen, porphyrin, nicotinate-nicotinamide and riboflavin metabolisms, pentose phosphate and glycolysis-gluconeogenesis pathway activity, and glycerol(phospho)lipid metabolism (Fig. 3 and fig. S12). The predominance of sulfur metabolism reflects the role of sulfide, produced by sulfate reduction in the underlying brine, as the primary source of energy in this system. The predominance of folate reactions (second highest Z_i value) (fig. S10) probably reflects a high proportion of one-carbon biochemistry (19). The predominance of inositol and amino-sugar pathways may reflect high-level production of osmoprotectants, as cellular adaptation to the prevailing high salt conditions and associated osmotic stress. Enzymes of sphingolipid metabolism represented almost 9% of the total reactome and 66% of all lipid biosynthetic pathways present on the array. Sphingolipids are thought to protect the cell surface against harmful environmental factors and can function as important signaling molecules involved in responses to a variety of stresses (20). Although the sample cultivations are expected to have resulted in population-level reductions of some microbial species present in the original source environment, sequence analysis of the 16S rRNA gene library created from the L'A sample (fig. S7) revealed that all of 96 clones analyzed belonged to a single group of halophilic anaerobes, the *Halanaerobium*-like organisms, that includes both fermenting and sulfur- and thiosulphate reducers (*H. congolensis*) (21). A pyrosequencing analysis of the L'A metagenome also indicated the presence of a methanogenic euryarchaeon related to *Methanosarcina* (14), which is consistent with methanogenic activity in the habitat. Thus, despite the biases inherent in both enrichment cultivations and metagenome libraries, there is reasonable consistency among habitat characteristics, the microbial community reactome, and the phylogeny and presumed physiology of the microbes present in the L'A sample.

KOL is a petroleum oil-enriched seawater sample from near the Port of Kolguev Island in the Barents Sea. Water in this area has a temperature of about 1°C in summer, when the sample was taken, and is chronically polluted from on-shore oil mining operations and shipping traffic. The main feature of the KOL metagenome reactome was a rich profile of hydrocarbon and organic utilization pathway activities, including those for the (halo)aromatics carbazole and atrazine (Fig. 3 and fig. S12). This is consistent both with the conditions prevailing in the source habitat and the sample enrichment culture derived from it and with the composition of the enrichment (determined by sequence analysis of a 16S rRNA gene

library prepared from it), in which members of the genus *Thalassolituus*, hydrocarbonoclastic bacteria specialized in the degradation of aliphatic and aromatic fractions of crude oil (22), represent some 33% of the entire community (fig. S7). Other distinguishing features of the KOL reactome were a prevalence of carbon fixation and denitrification and nitrogen metabolism activities as energy-source mechanisms and weak signals for sulfur and methane metabolism. These are consistent with microbial photosynthetic, nitrogen fixation, and denitrification activities typically carried out in coastal surface waters and with the presence of *Roseobacter* sp. (4%), and *Arcobacter* (18%), *Pseudomonas* (40%) spp., and *Paracoccus denitrificans*, known specialists in phototrophic carbon fixation and nitrogen metabolism, respectively (table S5). Nicotinate and nicotinamide metabolism was also evident, which is consistent with a metabolism in which NAD-dependent enzymes, like dehydrogenases (23), play important roles.

The VUL sample represents an enrichment for ferrous iron- and sulfur-oxidizing microbes derived from the acidic sulfur- and iron-rich sediments of a hydrothermal (25 to 75°C) pool. The reactome of the metagenome library of VUL confirmed that nitrogen and sulfur pathways are major activities in this environment, consistent with two species, *Acidithiobacillus* and *Alicyclobacillus*, bacteria with iron- and sulfur-oxidizing activities and the ability to fix nitrogen (24), making up some 74% of the microbial community (fig. S7). One major feature of the reactome is the predominance of histidine pathway enzymes, which represent 24% of the entire amino acid metabolism activities and 4.3% of all metabolic activities measured by the array: This represents 4 to 5 times as many as that found with the other two communities (Fig. 3 and fig. S12). The hydrothermal pool contains iron-rich minerals, like jarosite $[KFe(III)_3(OH)_6(SO_4)_2]$ and goethite $[FeO(OH)\cdot nH_2O]$, at concentrations ranging from 3 to >500 mg per liter, and other heavy metals, which have maximum solubilities at low pH. Because intracellular histidine diminishes the pH-dependent toxicity of heavy metals, by sequestering them in the cytoplasm, the elevated histidine activity of the VUL sample might reflect adaptive responses to a selective pressure for histidine-mediated detoxification of heavy metals (25). The VUL community showed a high level of porphyrin metabolism, which represented about 20% of the entire cofactor and vitamin metabolism (up to 8 times as high as that in the other communities), and included cobalt-salvaging coenzyme B₁₂ (cobalamin) precursor metabolism (fig. S13). It is possible that cobalt, an essential metal cofactor, is acquired and used more efficiently by the acidic community than the others (26, 27). A feature of the VUL reactome not expected for the acidic hydrothermal pool was the high level of enzymatic activity with organic compounds like caprolactam, which suggests the presence of these or structurally related compounds in the source sediment, originating either from natural vent activity or anthropogenic pollution.

Applications of the reactome array. The reactome array reported here represents a tool to obtain a detailed and quantitative profile of the metabolic activity of a cell population or tissue, a cell consortium or organ, and an organism or consortium of organisms. It provides a "metabolic barcode" that can be compared with those of other samples. Because many of the metabolites on the array are connected in known pathway sequences, it is in principle possible to reconstruct much of the metabolic network operating in a cell or organism without any prior genomic information. Where genomic information is available, the array forges a link between metabolome and genome (fig. S3). The most detailed and cohesive metabolic network construction will be with a clonal population of cells, and the least cohesive with a diverse consortium of cell types, because of greater metabolic range and diversity, and cellular discontinuities in the global metabolism. Nevertheless, it is even possible to obtain useful information on the metabolic activity of microbial communities from environmental samples (fig. S14). This can lead to the discovery of unknown metabolic activities in organisms and communities (table S6), identification of metabolic components of niche specificity, and exposure of predominant microbial pathways shaping the characteristics of individual habitats.

The reactome array, and customized versions of it, may find applications in the large-scale metabolic characterization of cell lines, organisms, mutants, transgenes, and libraries thereof, or of mutant enzyme libraries, or in diagnostic tests. In evaluation of environmental pollution, the array would provide a measure of pollutant-metabolizing enzymes and thus a measure of the concentration of bioavailable pollutant. The rapidity of the reactome array and its format, which enables automation and high throughput, should facilitate its application in diagnostic procedures. The array may also contribute to enzyme discovery, because proteins can be identified and isolated directly from environmental samples or from source organisms, thereby circumventing expression cloning biases.

References and Notes

1. T. G. Whitham *et al.*, *Science* **320**, 492 (2008).
2. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
3. M. T. P. Gilbert *et al.*, *Science* **317**, 1927 (2007).
4. J. F. Biddle *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10583 (2008).
5. E. A. Dinsdale *et al.*, *Nature* **452**, 629 (2008).
6. M. G. Kalyuzhnaya *et al.*, *Nat. Biotechnol.* **26**, 1029 (2008).
7. T. Lombardot *et al.*, *Nucleic Acids Res.* **34**, D390 (2006).
8. J. Raes, P. Bork, *Nat. Rev. Microbiol.* **6**, 693 (2008).
9. T. A. Gianoulis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1374 (2009).
10. P. J. Turnbaugh, J. I. Gordon, *Cell* **134**, 708 (2008).
11. T. R. Northen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3678 (2008).
12. J. M. Vieites *et al.*, *FEMS Microbiol. Rev.* **33**, 236 (2009).
13. M. Kanehisa *et al.*, *Nucleic Acids Res.* **32**, D277 (2004).
14. Materials and methods are available as supporting material on Science online.
15. T. Fawcett, *Pattern Recognit. Lett.* **27**, 861 (2006).
16. R. F. Service, *Science* **311**, 1544 (2006).
17. J. I. Jimenez *et al.*, *Environ. Microbiol.* **4**, 824 (2002).
18. D. Daffonchio *et al.*, *Nature* **440**, 203 (2006).

19. E. Y. Huang *et al.*, *J. Bacteriol.* **179**, 5648 (1997).
 20. Y. A. Hannun, C. Luberto, *Trends Cell Biol.* **10**, 73 (2000).
 21. P. W. van der Wielen *et al.*, *Science* **307**, 121 (2005).
 22. M. M. Yakimov *et al.*, *Curr. Opin. Biotechnol.* **18**, 257 (2007).
 23. A. F. Pronk *et al.*, *J. Bacteriol.* **177**, 75 (1995).
 24. O. V. Golyshina, K. N. Timmis, *Environ. Microbiol.* **7**, 1277 (2005).
 25. D. A. Pearce, F. Sherman, *J. Bacteriol.* **181**, 4774 (1999).
 26. J. D. Woodson *et al.*, *J. Bacteriol.* **185**, 7193 (2003).
 27. Y. Jiao *et al.*, *Appl. Environ. Microbiol.* **71**, 4487 (2005).
 28. This research was supported by the BIO2006-11738, CSD2007-00005, GEN2006-27750-C-4-E, BFU2008-

04398-E/BMC, and KBBE-226977 projects. A.B and Y.A-R thank the Spanish MEC for the FPU and FPI fellowships. F.P. thanks the Spanish MEC for the BIO2006-15318 project. K.N.T, O.V.G., and P.N.G acknowledge the Federal Ministry for Science and Education (BMBF) for a grant in the framework of the BiotechGenoMik program, and K.N.T. thanks the Fonds der Chemischen Industrie for generous support. Authors are deeply indebted to A. Yanenko for sampling Kolguev Island coastal water and to the captain and crew of Research Vessel Urania for their assistance in deep-sea sampling in the Mediterranean Sea and to J. Manuel Franco for statistical analyses.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5950/252/DC1
 Materials and Methods
 SOM Text
 Figs. S1 to S14
 Tables S1 to S6
 References
 Movie S1

26 March 2009; accepted 19 August 2009
 10.1126/science.1174094

Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses

Ido Amit,^{1,2,3,4} Manuel Garber,^{1,*} Nicolas Chevrier,^{2,3*} Ana Paula Leite,^{1,5*} Yoni Donner,^{1,*} Thomas Eisenhaure,^{2,3} Mitchell Guttman,^{1,4} Jennifer K. Grenier,¹ Weibo Li,^{2,3} Or Zuk,¹ Lisa A. Schubert,⁶ Brian Birditt,⁶ Tal Shay,¹ Alon Goren,^{1,7} Xiaolan Zhang,¹ Zachary Smith,¹ Raquel Deering,^{2,3} Rebecca C. McDonald,^{2,3} Moran Cabili,¹ Bradley E. Bernstein,^{1,3,7} John L. Rinn,¹ Alex Meissner,¹ David E. Root,¹ Nir Hacohen,^{1,2,3,†} Aviv Regev^{1,4,8,‡}

Models of mammalian regulatory networks controlling gene expression have been inferred from genomic data but have largely not been validated. We present an unbiased strategy to systematically perturb candidate regulators and monitor cellular transcriptional responses. We applied this approach to derive regulatory networks that control the transcriptional response of mouse primary dendritic cells to pathogens. Our approach revealed the regulatory functions of 125 transcription factors, chromatin modifiers, and RNA binding proteins, which enabled the construction of a network model consisting of 24 core regulators and 76 fine-tuners that help to explain how pathogen-sensing pathways achieve specificity. This study establishes a broadly applicable, comprehensive, and unbiased approach to reveal the wiring and functions of a regulatory network controlling a major transcriptional response in primary mammalian cells.

Regulatory networks controlling gene expression serve as decision-making circuits within cells. For example, when immune dendritic cells (DCs) are exposed to viruses, bacteria, or fungi, they respond with transcriptional programs that are specific to each pathogen (1) and are essential for establishing appropriate immunological outcomes (2). These responses are initiated through specific receptors, such as Toll-like receptors (TLRs), that distinguish broad pathogen classes and are propagated through well-characterized signaling cascades (2). However, little is known about how the transcriptional network is wired to produce specific outputs.

Two major observational strategies have associated regulators with their putative targets on a genome scale (3): Cis-regulatory models rely on the presence of predicted transcription factor

binding sites in the promoters of target genes (3–5), whereas trans-regulatory models are based on correlations between regulator and target expression (3–6). Because promoter binding sites and correlated expression are weak predictors of functional regulator-target linkages, such approaches are limited in their ability to produce reliable models of transcriptional networks (3). A complementary strategy is to systematically perturb every regulatory input and measure its effect on the expression of gene targets. This strategy has been successfully used in yeast (7–9) and sea urchin (10), but not in mammals.

A perturbation-based strategy for network reconstruction. We developed a perturbation strategy for reconstructing transcriptional networks in mammalian cells and used it to determine a network controlling the responses of DCs

to pathogens (Fig. 1). First, we profiled gene expression at nine time points after stimulation with five pathogen-derived components and identified specific and shared genes that respond to each stimulus (fig. S1A). We used these profiles to identify 144 candidate regulators whose expression changed in response to at least one stimulus (11) (fig. S1B, top). We also identified a signature of 118 marker genes (fig. S1B, bottom) that captures the complexity of the response. We generated a validated lentiviral short hairpin RNA (shRNA) library for 125 of the 144 candidate regulators (fig. S1C, top), used it to systematically perturb each of the regulators in DCs, stimulated the cells with a pathogen component, and profiled the expression of the 118-gene signature (12) (fig. S1C, bottom). Finally, we used the measurements from the perturbed cells to derive a validated model of the regulatory network (fig. S1D).

Gene expression programs in response to TLR agonists. We measured genome-wide expression profiles in DCs exposed to PAM3CSK4 (PAM), a synthetic mimic of bacterial lipopeptides; polyinosine-polycytidylic acid [poly(I:C)], a viral-like double-stranded RNA; lipopolysaccharide (LPS), a purified component from Gram-negative *Escherichia coli*; gardiquimod, a small-molecule

¹Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. ²Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA. ³Harvard Medical School, Boston, MA 02115, USA. ⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. ⁵Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ⁶NanoString Technologies, 530 Fairview Avenue N., Suite 2000, Seattle, WA 98109, USA. ⁷Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129, USA. ⁸Howard Hughes Medical Institute.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: nhacohen@partners.org

‡These authors contributed equally to this work.

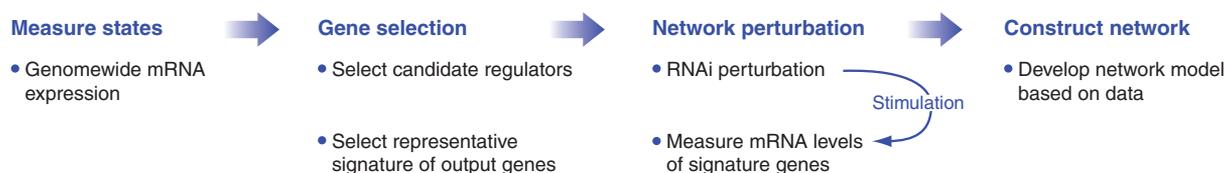


Fig. 1. A systematic strategy for network reconstruction. The strategy consists of four steps (left to right): state measurement using arrays; selection of regulators and response signatures; network perturbation with shRNAs against each regulator, followed by measurement of signature genes; and network reconstruction from the perturbational data.

RETRACTION

Post date 12 November 2010

The Research Article “Reactome array: Forging a link between metabolome and genome” (1) described the synthesis of some 2000 quenched fluorescent dye-metabolite compounds, and their use to create an array to obtain a global overview of the metabolic network operating in a population of cells at the time of sampling. Upon productive interaction with an enzyme in a sample, an array compound releases the dye, which fluoresces and its signal is captured. To our profound regret, peer inspection of the paper after publication revealed errors and omissions in the information provided on the chemistry underlying array compound synthesis, and the processing of array data obtained. After an investigation, the Ethics Committee of the CSIC in Madrid has recommended the withdrawal of the paper. Given the errors in the paper, and the skepticism about the array that they have generated, we retract the paper. We apologize to *Science*, our institutions, and the scientific community for any inconvenience caused by our paper and its retraction.

Ana Beloqui,¹ María Eugenia Guazzaroni,¹ Florencio Pazos,² José M. Vieites,¹
 Marta Godoy,² Olga V. Golyshina,³ Tatyana N. Chernikova,³ Agnes Waliczek,³
 Rafael Silva-Rocha,² Yamal Al-ramahi,¹ Violetta La Cono,⁴ Carmen Mendez,⁵
 José A. Salas,⁵ Roberto Solano,² Michail M. Yakimov,⁴ Kenneth N. Timmis,^{3,6}
 Peter N. Golyshin,^{3,7,8} Manuel Ferrer^{1*}

¹CSIC, Institute of Catalysis, 28049 Madrid, Spain. ²CSIC, Centro Nacional de Biotecnología, 28049 Madrid, Spain. ³HZIHelmholtz Centre for Infection Research, 38124 Braunschweig, Germany. ⁴Institute for Coastal Marine Environment, CNR, Messina 98122, Italy. ⁵Universidad de Oviedo, 33006 Oviedo, Spain. ⁶Institute of Microbiology, Carolo-Wilhelmina Technical University of Braunschweig, 38106 Braunschweig, Germany. ⁷School of Biological Sciences, Bangor University, Gwynedd LL572 UW, UK. ⁸Centre for Integrated Research in the Rural Environment, Aberystwyth University–Bangor University Partnership (CIRRE).

*To whom correspondence should be addressed. E-mail: mferrer@icp.csic.es

Reference

1. A. Beloqui *et al.*, *Science* **326**, 252 (2009).



Reactome Array: Forging a Link Between Metabolome and Genome

Ana Beloqui *et al.*
Science **326**, 252 (2009);
DOI: 10.1126/science.1174094

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of February 24, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

</content/326/5950/252.full.html>

Supporting Online Material can be found at:

</content/suppl/2009/10/08/326.5950.252.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

</content/326/5950/252.full.html#related>

This article **cites 26 articles**, 15 of which can be accessed free:

</content/326/5950/252.full.html#ref-list-1>

This article has been **cited by** 16 article(s) on the ISI Web of Science

This article has been **cited by** 3 articles hosted by HighWire Press; see:

</content/326/5950/252.full.html#related-urls>

This article appears in the following **subject collections**:

Biochemistry

</cgi/collection/biochem>