

Received 22 October 2002; accepted 30 January 2003; doi:10.1038/nature01482.

1. Hauri, E. H. & Hart, S. R. Rhenium abundances and systematics in oceanic basalts. *Chem. Geol.* **139**, 185–205 (1997).
2. Walker, R. J., Prichard, H. M., Ishiwatari, A. & Pimentel, M. The osmium isotopic composition of convection upper mantle deduced from ophiolite chromites. *Geochim. Cosmochim. Acta* **66**, 329–345 (2002).
3. Bennett, V. C., Nutman, A. P. & Esat, T. M. Constraints on mantle evolution from ¹⁸⁷Os/¹⁸⁸Os isotope compositions of Archean ultramafic rocks from southern West Greenland (3.8 Ga) and Western Australia (3.46 Ga). *Geochim. Cosmochim. Acta* **66**, 2615–2630 (2002).
4. Esser, B. K. & Turekian, K. K. The osmium isotopic composition of the continental crust. *Geochim. Cosmochim. Acta* **57**, 3093–3104 (1993).
5. Peucker-Ehrenbrink, B. & Jahn, B. Rhenium-osmium isotope systematics and platinum group element concentrations; loess and the upper continental crust. *Geochem. Geophys. Geosyst.* **2**, 2001GC000172 (2001).
6. Sun, W. D., Bennett, V. C., Eggins, S. M., Arculus, R. J. & Perfit, M. R. Rhenium systematics in submarine MORB and back-arc basin glasses: Laser ablation ICP-MS results. *Chem. Geol.* (in the press).
7. McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
8. Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust; a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
9. Becker, H. Re-Os fractionation in eclogites and blueschists and the implications for recycling of oceanic crust into the mantle. *Earth Planet. Sci. Lett.* **177**, 287–300 (2000).
10. Morford, J. L. & Emerson, S. The geochemistry of redox sensitive trace metals in sediments. *Geochim. Cosmochim. Acta* **63**, 1735–1750 (1999).
11. Jaffe, L. A., Peucker-Ehrenbrink, B. & Petsch, S. T. Mobility of rhenium, platinum group elements and organic carbon during black shale weathering. *Earth Planet. Sci. Lett.* **198**, 339–353 (2002).
12. Korzhinsky, M. A., Tkachenko, S. I., Shmulovich, K. I., Taran, Y. A. & Steinberg, G. S. Discovery of a pure rhenium mineral at Kudriavoy Volcano. *Nature* **369**, 51–52 (1994).
13. Ishihara, S. Rhenium contents of molybdenites in granitoid-series rocks in Japan. *Econ. Geol. Bull. Soc. Econ. Geol.* **83**, 1047–1051 (1988).
14. Alves, S., Schiano, P., Capmas, F. & Allègre, C. J. Osmium isotope binary mixing arrays in arc volcanism. *Earth Planet. Sci. Lett.* **198**, 355–369 (2002).
15. Alves, S., Schiano, P. & Allègre, C. J. Rhenium-osmium isotopic investigation of Java subduction zone lavas. *Earth Planet. Sci. Lett.* **168**, 65–77 (1999).
16. Woodland, S. J., Pearson, D. G. & Thirlwall, M. F. A platinum group element and Re-Os isotope investigation of siderophile element recycling in subduction zones: Comparison of Grenada, Lesser Antilles Arc and the Izu-Bonin Arc. *J. Petrol.* **42**, 171–198 (2002).
17. Lassiter, J. C. & Luhr, J. F. Osmium abundance and isotope variations in mafic Mexican volcanic rocks: Evidence for crustal contamination and constraints on the geochemical behavior of osmium during partial melting and fractional crystallization. *Geochem. Geophys. Geosyst.* **2**, 2000GC000116 (2001).
18. Chesley, J. T., Ruiz, J., Righter, K., Ferrari, L. & Gomez-Tuena, A. Source contamination versus assimilation: An example from the Trans-Mexican Volcanic Arc. *Earth Planet. Sci. Lett.* **195**, 211–221 (2002).
19. Borg, L. E., Brandon, A. D., Clynne, M. A. & Walker, R. J. Re-Os isotopic systematics of primitive lavas from the Lassen region of the Cascade Arc, California. *Earth Planet. Sci. Lett.* **177**, 301–317 (2000).
20. Righter, K., Chesley, J. T. & Ruiz, J. Genesis of primitive, arc-type basalt: Constraints from Re, Os, and Cl on the depth of melting and role of fluids. *Geology* **30**, 619–622 (2002).
21. Bennett, V. C., Norman, M. D. & Garcia, M. O. Rhenium and platinum group element abundances correlated with mantle source components in Hawaiian picrites; sulphides in the plume. *Earth Planet. Sci. Lett.* **183**, 513–526 (2000).
22. Lassiter, J. C. Evidence for Re loss during magma degassing: Implications for Re concentration variations in OIB and MORB. *Eos* **83**(47) (Fall Meet. Suppl.), 1447 (2002).
23. Eggins, S. M. Origin and differentiation of picritic arc magmas, Ambae (Aoba), Vanuatu. *Contrib. Mineral. Petrol.* **114**, 79–100 (1993).
24. Kamenetsky, V. S., Crawford, A. J., Eggins, S. M. & Muehe, R. Phenocryst and melt inclusion chemistry of near-axis seamounts, Valu Fa Ridge, Lau Basin; insight into mantle wedge melting and the addition of subduction components. *Earth Planet. Sci. Lett.* **151**, 205–223 (1997).
25. Kamenetsky, V. S. *et al.* Parental basaltic melts and fluids in eastern Manus backarc basin; implications for hydrothermal mineralisation. *Earth Planet. Sci. Lett.* **184**, 685–702 (2001).
26. Sylvester, P. J. & Eggins, S. M. Analysis of Re, Au, Pd, Pt and Rh in NIST glass certified reference materials and natural basalt glasses by laser ablation ICP-MS. *Geostand. NewsL.* **21**, 215–229 (1997).
27. Barth, M. G., McDonough, W. F. & Rudnick, R. L. Tracking the budget of Nb and Ta in the continental crust. *Chem. Geol.* **165**, 197–213 (2000).
28. Peucker-Ehrenbrink, B. & Hannigan, R. E. Effects of black shale weathering on the mobility of rhenium and platinum group elements. *Geology* **28**, 475–478 (2000).
29. Hofmann, A. W. Mantle geochemistry; the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
30. Schiano, P., Birck, J. L. & Allègre, C. J. Osmium-strontium-neodymium-lead isotopic covariations in mid-ocean ridge basalt glasses and the heterogeneity of the upper mantle. *Earth Planet. Sci. Lett.* **150**, 363–379 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Kamenetsky, C. Allen and J. M. G. Shelley for help with sample preparation and laser ablation ICP-MS analyses. We also thank S.-s. Sun for discussions that initiated this work, and D. G. Pearson for comments and suggestions that improved the manuscript. W.S. was supported by an Australian international graduate student scholarship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.C.B. (e-mail: vicke.bennett@anu.edu.au).

Genetics of gene expression surveyed in maize, mouse and man

Eric E. Schadt*†, Stephanie A. Monks*‡, Thomas A. Drake§, Aldons J. Lusis||, Nam Che||, Veronica Colinayo||, Thomas G. Ruff¶, Stephen B. Milligan*, John R. Lamb*, Guy Cavet*, Peter S. Linsley*, Mao Mao*, Roland B. Stoughton* & Stephen H. Friend*#

* Rosetta Inpharmatics, LLC, 12040 115th Avenue N.E., Kirkland, Washington 98034, USA

‡ Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA

§ Department of Pathology and Laboratory Medicine, and || Department of Microbiology, Molecular Genetics and Immunology, Department of Medicine, and Department of Human Genetics, UCLA, Los Angeles, California 90095, USA

¶ Monsanto Company, St Louis, Missouri 63167, USA

Merck Research Laboratories, Merck & Co., Inc., POB 4, WP14-2500, West Point, Pennsylvania 19486, USA

† These authors contributed equally to this work

Treating messenger RNA transcript abundances as quantitative traits and mapping gene expression quantitative trait loci for these traits has been pursued in gene-specific ways. Transcript abundances often serve as a surrogate for classical quantitative traits in that the levels of expression are significantly correlated with the classical traits across members of a segregating population. The correlation structure between transcript abundances and classical traits has been used to identify susceptibility loci for complex diseases such as diabetes¹ and allergic asthma². One study recently completed the first comprehensive dissection of transcriptional regulation in budding yeast³, giving a detailed glimpse of a genome-wide survey of the genetics of gene expression. Unlike classical quantitative traits, which often represent gross clinical measurements that may be far removed from the biological processes giving rise to them, the genetic linkages associated with transcript abundance affords a closer look at cellular biochemical processes. Here we describe comprehensive genetic screens of mouse, plant and human transcriptomes by considering gene expression values as quantitative traits. We identify a gene expression pattern strongly associated with obesity in a murine cross, and observe two distinct obesity subtypes. Furthermore, we find that these obesity subtypes are under the control of different loci.

Liver tissues from 111 F₂ mice constructed from two standard inbred strains, C57BL/6J and DBA/2J, were profiled using a mouse gene oligonucleotide microarray. The expression values from these experiments were treated as quantitative traits and carried through a linkage analysis using evenly spaced markers across the autosomal chromosomes. Of the 23,574 genes represented on the microarray, 7,861 were detected as significantly differentially expressed (type I error = 0.05) in the parental strains or in at least 10% of the F₂ mice profiled. Using standard interval mapping techniques, quantitative trait loci (QTL) with log of the odds ratio (LOD) scores greater than 4.3 (*P*-value < 0.00005) were identified for 2,123 genes, with a maximum LOD score of 80.0 (*P*-value ≪ 10⁻²⁰). On average, gene expression QTL (eQTL) with LOD scores greater than 4.3 explained 25% of the transcription variation of the corresponding genes observed in this F₂ set, with this percentage increasing to nearly 50% for LOD scores greater than 7.0. Applying a simple Bonferroni correction to a 0.05 genome-wide significance level for each trait, given that 7,861 traits were tested, we would expect only 393 false positives at a LOD score threshold of 4.3.

Without filtering based on significant differential expression over the set of mice profiled, we detected 4,339 eQTL over 3,701 genes with LOD scores greater than 4.3, 11,021 genes with at least one

eQTL with a LOD score greater than 3.0, and 17,415 eQTL with LOD scores over 2.0. The number of eQTL with LOD scores exceeding 7.0 (P -value $\leq 10^{-7}$) increased by 50% when genes that were not detected as significantly differentially regulated, as defined above, were considered. This result indicates that although individual tests of hypotheses on the differential regulation of a single gene may not be significant, viewing the behaviour of that gene by genotype over 111 animals provides significantly more information on the biological activity of that gene, an observation also noted for yeast³. Of the 965 genes with LOD scores greater than 7.0, 157 had a maximum fold change of 3.0, indicating a class of genes whose high LOD scores reflect tight transcriptional control (small variance), not large expression differences.

Figure 1a plots the percentage of eQTL at different LOD score thresholds across 920 evenly spaced bins, each 2 cM wide, covering the mouse genome. The number of eQTL in each bin was divided by

the total number of eQTL and plotted. eQTL hotspots are apparent on chromosomes 2, 6, 7, 9, 10, 16 and 17, where for each of these hotspot locations, greater than 1% of the total number of eQTL identified genome-wide localize to a 4-cM window. With 460 4-cM windows over the 19 autosomal chromosomes, the probability that greater than 1% of the eQTL would localize to one such window is less than 1.0×10^{-16} . Furthermore, gene expression is seen to be a complex trait, given that 40% of genes with at least one eQTL with a LOD score greater than 3.0 had more than one eQTL, and close to 4% of such genes had more than three eQTL. It is the clustering of eQTL to specific loci and the relationship between these genes with respect to expression, rather than single eQTL significance levels, that when taken together lead to highly significant and interesting patterns that can be associated with phenotypes related to common diseases (see below).

Of the 18,460 genes that could be mapped to a chromosome

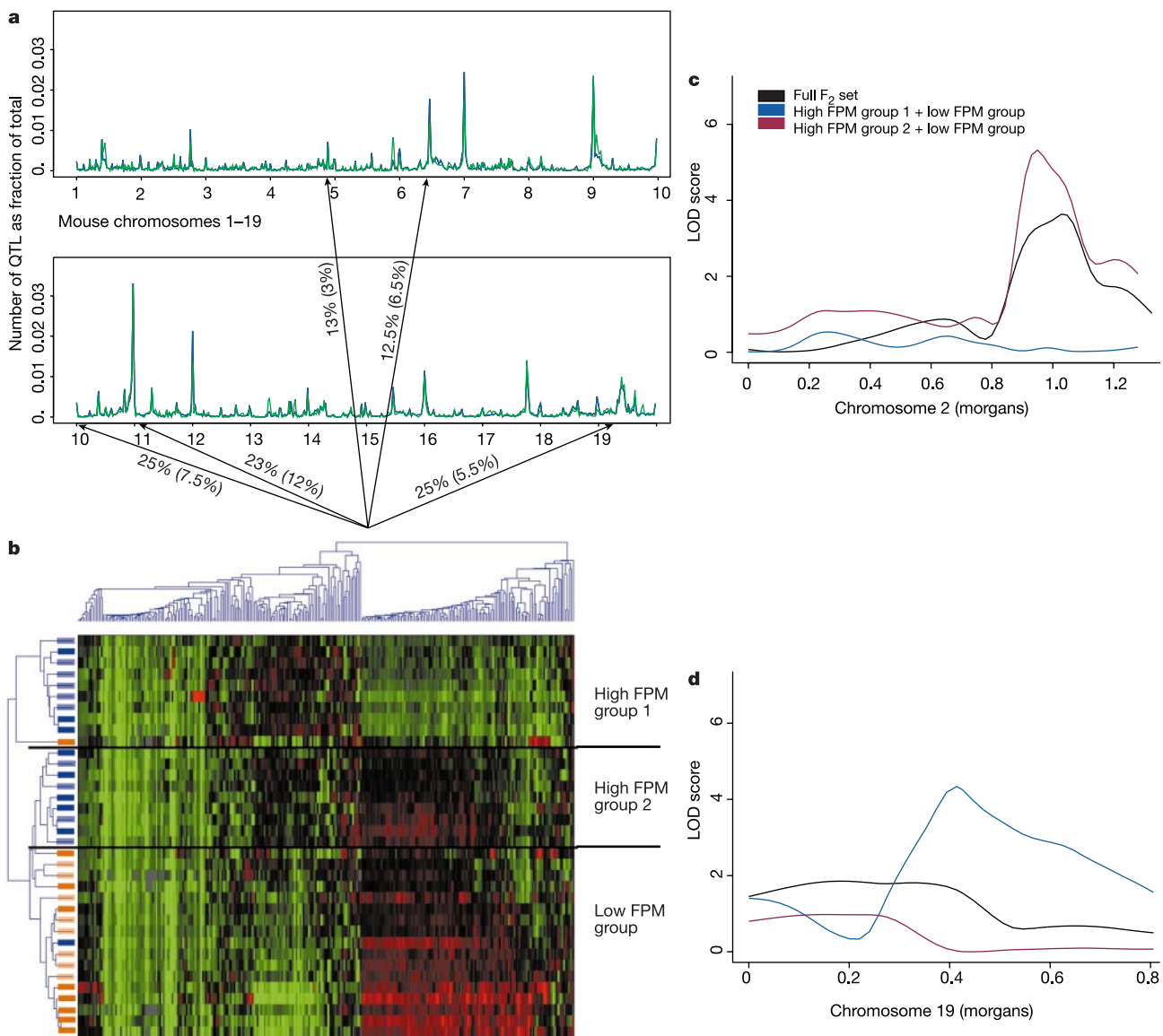


Figure 1 Murine gene expression quantitative trait loci (eQTL) distributions and the molecular basis for fat pad mass (FPM) in a murine F₂ cross. **a**, Percentage of eQTL in 2-cM bins spanning the murine autosomal chromosomes at two LOD score thresholds. **b**, Colour matrix display for hierarchically clustered genes (x axis) and extreme FPM mice (y axis). Dark/light blue bars indicate mice in the upper/lower half of the high FPM group; dark/light orange bars indicate mice in the lower/upper half of the low FPM group.

See text for a definition of the five arrows. **c, d**, Chromosome 2 (**c**) and 19 (**d**) log of the odds ratio (LOD) score curves for the FPM trait. QTL analysis was performed on three different sets: (1) black curves, all F₂ mice; (2) blue curves, F₂ mice classified as high FPM group 1 and low FPM group; (3) red curves, F₂ mice classified as high FPM group 2 and low FPM group.

position using the Celera Mouse Genome Database, 3,007 had eQTL with LOD scores greater than 4.3, and 784 had eQTL with LOD scores greater than 7.0. Approximately 34% of the mapped genes with eQTL exceeding 4.3 had a physical location coincident with the eQTL position, whereas 71% of the mapped genes with eQTL exceeding 7.0 had a physical location coincident with its eQTL position. The trend observed here is that eQTL with high LOD scores are *cis*-acting in most cases, whereas moderately significant QTL are *trans*-acting in most cases. This is consistent with our expectation that first-order effects (DNA variations in a gene that affect transcription of the gene itself) are easier to detect than second-order effects (genes acting on the transcription of other genes).

Figure 2 highlights a range of interesting gene-centred polymorphisms known to exist between the DBA and B6 strains (*cis*-acting transcription regulation). The *C5* gene (NM_010406) has a 2-base-pair (bp) deletion in the coding region in DBA mice that leads to rapid decay of the transcript², compared with B6. A LOD of 27.4 centred over the *C5* gene on chromosome 2 is readily detected (black curve). The *Alad* (AK002300) gene is present in two copies in DBA and one copy in B6 (ref. 4). A LOD of 9.3 centred over the *Alad* gene (red curve) represents the differential dosing that occurs between the two strains. The *St7* (NM_022332) gene is differentially spliced at several locations⁵, and, for a stable splice form at the 3' location of the gene, the probe for this gene fortuitously overlapped the region alternatively spliced out in DBA, but not in B6. The differential splicing event is detected by the major QTL (LOD score of 20.1) for *St7*, which is centred over the *St7* gene (blue curve) on chromosome 6. Finally, the *Nnmt* (NM_010924) gene, important for drug metabolism, has been identified as being polymorphic with respect to transcription between the DBA and B6 strains⁶. This polymorphism is confirmed by a major QTL (LOD score of 15.3) for *Nnmt*, centred over the *Nnmt* gene (green curve) on chromosome 9. We note that single nucleotide polymorphisms (SNPs) covered by 60-residue oligonucleotide probes would not be expected to significantly affect transcript abundance measurements among these samples⁷.

Next we investigated a way of combining gene expression, genetics and clinical trait data to elucidate the complexity of common diseases. Major loci controlling complex phenotypes

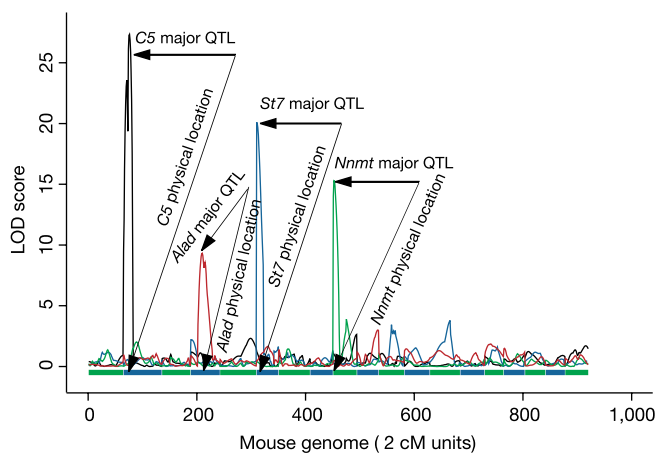


Figure 2 *Cis*-acting eQTL identified for several DNA polymorphisms known to induce transcription polymorphisms. The LOD score curves over the murine autosomal chromosomes are shown for the expression traits of four genes: (1) black curve, *C5* (NM_010406); (2) red curve, *Alad* (AK002300); (3) blue curve, *St7* (NM_022332); and (4) green curve, *Nnmt* (NM_010406). The alternating green and blue strip at the base of the curves represents chromosome boundaries, starting with chromosome 1 and ending with chromosome 19.

such as obesity may potentially affect scores of genes, if not hundreds. We would expect those genes involved in the more downstream aspects of pathways associated with common diseases to have eQTL linked to the major causative loci for those diseases. In addition, there may be heterogeneity among the causative loci for a given disease in a population of interest. When present, this heterogeneity impacts the ability to detect linkages to the causative loci, as the significance of any one locus is diminished when the population is considered as a whole.

Because the mice described above had been on a high-fat, atherogenic diet for 4 months before livers were profiled⁸, they model the spectrum of obesity, diabetes and atherosclerosis in a natural population. The 280 genes depicted in the two-dimensional cluster in Fig. 1b were selected as the most differentially expressed set of genes in mice comprising the upper and lower 25th percentiles of the subcutaneous fat-pad-mass (FPM) trait. When clustering on this set of genes over the high/low FPM mice, the mice cluster almost perfectly into high FPM and low FPM groups (Fig. 1b). Furthermore, there seem to be two distinct expression patterns for mice in the high FPM group, indicating some degree of heterogeneity in the high FPM mice.

Arrows in Fig. 1b highlight five regions in Fig. 1a where more than 50% of the genes in the FPM set genetically link. Taking into account the eQTL distribution for all genes highlighted in Fig. 1a, each of the five hotspot regions defined by the FPM cluster are very significantly enriched for eQTL for the genes in the FPM set. For instance, 25% of the genes in the FPM set have eQTL in a 10-cM region on chromosome 19. This represents an almost fivefold increase over what would be expected by chance, given that roughly 5.5% of genes from the genome-wide gene set have eQTL to this same 10-cM region.

The patterns in Fig. 1b serve to define the obesity trait, FPM, beyond what would be possible without the expression data. There are clearly two distinct patterns associated with high FPM mice. For the FPM trait, a genome-wide scan revealed four classical QTL (clinical QTL or cQTL) with LOD scores greater than 2.0 (ref. 8). To further elucidate this clinical trait, the 111 F₂ animals for which clinical and gene expression data existed were classified into one of the three groups shown in Fig. 1b. Subsequently, separate genetic analyses were performed on two sets of animals: (1) those classified as high FPM group 1 or low FPM; and (2) those classified as high FPM group 2 or low FPM. Figure 1c, d depicts the results of these analyses for two chromosomes. The chromosome 2 FPM QTL was the largest of the four QTL originally identified for this trait, when all animals were considered together. This QTL vanishes when considering the high FPM group 1 with the low FPM group and increases by almost 2 LOD units over the original when considering the high FPM group 2 with the low FPM group. Figure 1d depicts another interesting locus for which the original analysis on the full set of mice yielded no significant QTL for the FPM trait, but where the high FPM group 2 considered with the low FPM group gave rise to a QTL with a significant LOD score, whereas the high FPM group 1 considered with the low FPM group was less significant than that of the full set. These results indicate that the chromosome 2 and 19 QTL each significantly affect only a subset of the F₂ population, a form of heterogeneity that clearly demonstrates the complexity underlying traits such as obesity.

We next focused our efforts on objectively identifying candidate genes for common diseases. An expanded view of the clinical traits and a portion of the gene expression traits linking to the chromosome 2 locus discussed above is given in Fig. 3. Notably, a group of major urinary protein genes represented in the Fig. 1b FPM cluster are linked to the chromosome 2 locus, in addition to seven other loci (all with a LOD score exceeding 2.0), four of which localize together with adiposity or FPM traits. The *Mup1* gene stands out because it was the most highly correlated or had eQTL co-localized with eQTL from many other genes known to be involved in obesity-related

pathways, including retinoid X receptor (Rxr)- γ , Rxr-interacting protein, acyl-coenzyme A oxidase 1, leptin receptor, peroxisome proliferator activated receptor (Ppar)- γ and Lpr6. This demonstrates that the chromosome 2 locus draws together adiposity, FPM, cholesterol and triglyceride levels and is linked to genes with proven roles in obesity and diabetes. Furthermore, the *Mup* genes are members of the lipocalin protein family, and although they are known to have a central role in pheromone-binding processes that affect mouse physiology and behaviour⁹, variations in *Mup* expression have been associated with variations in body weight¹⁰, bone length¹⁰ and levels of very-low-density lipoprotein¹¹.

Most of the genes linked to the chromosome 2 locus do not physically reside on chromosome 2, and so, are at least partially regulated by one or more loci in the chromosome 2 hotspot region. Of the 423 genes linked to the chromosome 2 locus for which we have chromosome mapping information, there are only four eQTL with LOD scores greater than 3.0 that correspond to genes whose physical locations are within 2 cM of the peak. However, only two of the four genes (NM_025575 and NM_15731) had significant genetic interactions with the subcutaneous FPM trait. Gene NM_025575 codes for a dolichyl-diphospho-oligosaccharide-protein glycosyltransferase and gene NM_015731 codes for a cation-transporting ATPase; these genes may be considered as primary causative candidates for the linkage activity at the chromosome 2 locus.

The region supporting the chromosome 2 locus is homologous to human chromosome 20q12-q13.12, a region that has previously been linked to human obesity-related phenotypes^{12,13}. The human orthologues for genes NM_025575 and NM_015731 highlighted in the Fig. 3 reside in the human chromosome 20 region. Although other genes, such as melanocortin 3 receptor (*Mcr3r*), have been suggested as possible candidates for obesity at this locus¹³, our data suggest that the genes NM_025575 and NM_015731 may be responsible for the underlying QTL, as discussed above. We note that expression levels of *Mcr3r* are not linked to the chromosome 2 locus, and there were no SNPs annotated in the exons or introns of this gene between the C57/BL6 and DBA/2J strains in the most recent build of the Celera RefSNP database, further suggesting that *Mcr3r* may not be the chromosome 2 QTL.

To pursue further our survey of the heritability of gene expression, we focused on *Zea mays*, a classical genetic organism that has been studied intensely and where previous reports have

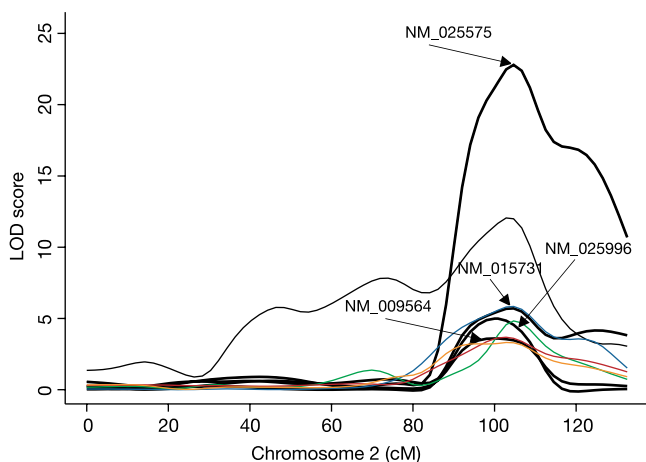


Figure 3 Clinical QTL (cQTL) for obesity-related traits localize together with eQTL. LOD score curves for four obesity-related traits are shown: (1) blue curve, subcutaneous FPM; (2) green curve, perimetrial FPM; (3) red curve, omental FPM; (4) orange curve, adiposity; (5) thin black curve, joint LOD score curve for these four clinical traits considered simultaneously. LOD score curves for four candidate genes that may explain the obesity cQTL are also shown (thick black curves).

demonstrated that protein¹⁴, enzymatic¹⁵ and metabolite¹⁶ levels could be associated with regions of the genome that control complex traits. QTL results from treating gene expression in ear leaf tissue as a quantitative trait in 76 F₂-derived F₃ progeny constructed from two typical inbred lines of maize—a stiff stalk synthetic type and a Lancaster type—largely parallel those in the mouse described above. Of the 18,805 genes detected as being significantly differentially regulated (type I error = 0.05) in at least 10 samples, there were 6,481 genes with at least one eQTL exceeding a LOD score of 3.0 (with a maximum LOD score of 41.3), and a total of 7,322 QTL overall. Most genes had a single eQTL, and just over 80% of the eQTL exceeding a LOD score of 7.0 were localized together with the physical gene giving rise to the eQTL, when the physical location of the gene could be determined.

Figure 4 represents a new form of genetic interaction. Two genes, each with a significant eQTL on different chromosomes, are seen to have uncorrelated expression values over the 76 samples. However, the genes appear to be interacting by genotype, representing an interaction between genes that is similar to epistasis, but more general because it occurs between the eQTL of different genes. Although interactions such as this (statistically significant at a type I error of 0.05) occurred among fewer than 10% of the genes with significant eQTL (LOD scores greater than 3.0), this form of genetic interaction clearly offers a potentially powerful means to implicate genes as being involved in the same or related pathways. Furthermore, this type of interaction would have been missed by all standard clustering methods applied to the expression data alone, but the relationship was readily detected using statistical genetics models designed to identify interactions among QTL.

As a preliminary survey into the genetics of gene expression in humans, 56 individuals from four CEPH (Centre d'Etude du Polymorphisme Humain) reference families¹⁷ were selected for expression profiling of lymphoblastoid cell lines using a standard human gene oligonucleotide microarray¹⁸. The four families, CEPH/Utah pedigrees 1362, 1375, 1377 and 1408, consisted of large 'sibships' along with parents and grandparents. Heritability analysis was performed for gene expression on a subset of 2,726

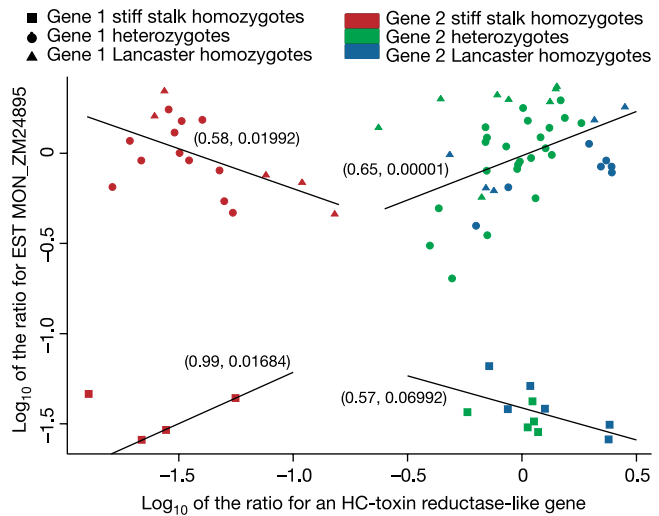


Figure 4 Genes with no overall correlation with respect to expression demonstrate interesting patterns of genetic interaction. The scatter plot shows the mean log₁₀ ratio for two *Zea mays* genes that are uncorrelated overall, each with a significant eQTL (LOD of 24.3 for the gene on the x axis and 24.9 for the gene on the y axis) falling on two separate chromosomes. Patterns are apparent in the plot despite the overall random correlation, as the four groups in each quadrant of the plot are correlated. The least squares regression line is shown for each quadrant, with the correlation coefficient values and corresponding *P*-values given in parentheses. EST, expressed sequence tag. HC, *Helminthosporium carbonum*.

genes that were significantly differentially regulated (type I error = 0.05) within 8 or more of the 16 pedigree founders. We deemed the sample size too small to perform systematic linkage analysis across all genes. We found that for the differentially expressed genes, 29% had a detectable genetic component (type I error < 0.05). This result offers an important glimpse into the genetics of gene expression in humans, with such a large percentage of genes detected with significant heritabilities in such a small sample of 'normal' individuals. We propose that this group of genes may make good therapeutic targets for complex human diseases, given that their degree of genetic control was so readily identifiable in such a small number of families. A complete list of those genes with a significant heritable component and additional information on tests for gender and age effects are provided in the Supplementary Information.

The identification of eQTL for genes expressed in a handful of tissues begins to provide an insight into the genetic networks that underlie the complexity of living systems. The several hundred genes with LOD scores exceeding 20 in the mouse represent a new class of quantitative traits, with linkage significances not commonly seen before in mammalian systems. The causal nature of genetics allows the anchoring of multiple genes under the common control of single or multiple loci, as shown in Fig. 3, thereby providing roots for the graphs that can more completely depict the complicated network of gene interactions at work in complex phenotypes.

The class of genes discussed for Figs 1b and 3 provide objective evidence that many of the genes co-localized to a single QTL hotspot are associated with the obesity-related traits. The patterns of expression serve to refine the obesity phenotype and allow the enrichment of subpopulations that are homogeneous with respect to the underlying causes of obesity in the population. Identifying such subpopulations has significant consequences for drug discovery, as each subpopulation may be more effectively treated by a compound that targets a pathway specifically associated with the disease in that subpopulation.

Combining gene expression, genotype and clinical data in a segregating population has the potential to affect the more significant rate-limiting steps in the drug discovery process: objectively classifying individuals according to disease subtypes and identifying the drivers of the pathways or the causal factors underlying those disease subtypes. In the past, dissecting complex traits using genetics has met with limited success, and until now gene expression has served as an indirect marker for complex traits. We have demonstrated that the combination of gene expression and genetics data has the potential to overcome these barriers. The addition of gene expression data can be used to refine the disease phenotype, directly implicate pathways and genes comprising those pathways associated with the disease phenotype, and identify the key drivers of the pathways underlying the disease phenotype. □

Methods

Preparation of labelled complementary DNA

Lymphoblastoid cell lines from CEPH/Utah pedigree families 1362, 1375, 1377 and 1408 were obtained from Coriell Cell Repositories. Other lymphoblastoid cell lines were established from normal donors by immortalization with Epstein-Barr virus (EBV) as described previously¹⁹. Cells were cultured in RPMI 1640 medium containing 15% fetal bovine serum, and penicillin/streptomycin antibiotics (Invitrogen). Cells were maintained in the log phase of cell growth for at least two days and were collected at densities of 0.4–0.9 × 10⁶ cells ml⁻¹.

An F₂ intercross was constructed from C57BL/6J and DBA/2J strains of mice. All mice were housed under conditions meeting the guidelines of the Association for Accreditation of Laboratory Animal Care. Mice were on a rodent chow diet up to 12 months of age, and then switched to an atherogenic high-fat, high-cholesterol diet for another 4 months. More details on this cross are described in ref. 8. Parental and F₂ mice were killed at 16 months of age. At death, the livers were immediately removed, flash-frozen in liquid nitrogen and stored at -80 °C.

An F₃ cross was constructed from standard inbred lines of *Z. mays*: a stiff stalk strain and Lancaster type²⁰. After the initial cross, the F₁ plants were self-crossed to obtain the F₂ progeny, and then the F₂ were self-crossed to obtain the F₃ progeny. The plants were field grown, and at 5 days after flowering, ear leaf tissues were collected from ten F₃ progeny for

each F₂ line. The ear leaf tissues were frozen on dry ice in the field on collection.

The hybridizations for each species were performed in duplicate with fluor reversal. In all cases, total cellular RNA was then purified using an RNeasy Mini kit according to the manufacturer's instructions (Qiagen). For each species, competitive hybridizations were performed by mixing fluorescently labelled antisense RNA (cRNA; 5 µg) from each sample with the same amount of cRNA from a reference pool comprising equal amounts of cRNA from related samples (for more details, see Supplementary Information).

Probe selection for the gene expression arrays

The human microarray contained 24,479 non-control oligonucleotide probes for human genes as described previously¹⁸. The mouse microarray contained 23,574 non-control oligonucleotide probes for mouse genes and 2,186 control probes. Full-length mouse sequences were extracted from murine Unigene clusters, combined with RefSeq mouse sequences and RIKEN full-length sequences. The maize ear leaf microarray carried 24,473 non-control oligonucleotide probes for genes of interest and 1,287 control probes. More than 50,000 non-control probes were selected from maize and rice. These probes were used on ink-jet microarrays to collect expression data for RNA samples from a range of maize tissues. The top-ranking 24,473 probes were carried forward for use on the ear leaf array.

For all microarrays, to select a probe 60 nucleotides in length for each gene sequence, we used a series of filtering steps, taking into account repeat sequences, binding energies, base composition, distance from the 3' end, sequence complexity and potential cross-hybridization interactions⁷. All microarrays used in this study were custom ink-jet microarrays fabricated by Agilent Technologies.

Analysis of expression data

Array images were scanned using the Agilent Dual Laser Microarray scanner (Agilent Technologies) and processed as previously described²¹ to obtain background noise, single-channel intensity and associated measurement error estimates. The colour displays given in panel b of the Fig. 1b show log₁₀ (expression ratio) as: (1) red when the red channel is upregulated relative to the green channel; (2) green when the red channel is downregulated relative to the green channel; (3) black when the log₁₀ (expression ratio) is close to zero; and (4) grey when data from one or both of the channels for a given probe is unreliable.

Linkage and data analysis

Variance components analysis²² was used to estimate the heritability of gene expression, as measured by the mean log₁₀ expression ratio, for each of the 2,726 mRNAs that were significantly differentially expressed in the founders, and to test whether the heritability was significantly different from zero. Genes were defined as differentially regulated if eight or more founders had a *P*-value for differential expression less than 0.05.

A complete linkage map for all chromosomes except Y in mouse was constructed at an average density of 13 cM using microsatellite markers. Linkage maps were constructed and QTL analysis was performed using MapMaker QTL²³ and QTL Cartographer²⁴. Log of the odds ratio (LOD) scores were calculated at 2-cM intervals throughout the genome for each of the 23,574 genes represented on the mouse microarray.

A complete linkage map for all chromosomes in *Z. mays* was constructed at an average density of 12 cM using microsatellite markers. Linkage maps were constructed based on genotypes from the F₂ progeny using MapMaker QTL. Mapping of QTL for gene expression traits was carried out as described previously²⁵. For the interaction result presented in Fig. 4, individual specific LOD score vectors were calculated for each QTL, and correlations were computed between all vector pairs. See the Supplementary Information for more details on the segregation and linkage analyses performed over all species.

Received 7 October 2002; accepted 10 January 2003; doi:10.1038/nature01434.

1. Eaves, I. A. *et al.* Combining mouse congenic strains and microarray gene expression analyses to study a complex trait: the NOD model of type 1 diabetes. *Genome Res.* **12**, 232–243 (2002).
2. Karp, C. L. *et al.* Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nature Immunol.* **1**, 221–226 (2000).
3. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
4. Claudio, L., Lee, T., Wolff, M. S. & Wetmur, J. G. A murine model of genetic susceptibility to lead bioaccumulation. *Fundam. Appl. Toxicol.* **35**, 84–90 (1997).
5. Huang, Y. H., Chen, Y. T., Lai, J. J., Yang, S. T. & Yang, U. C. PALS db: Putative Alternative Splicing database. *Nucleic Acids Res.* **30**, 186–190 (2002).
6. Yan, L., Otterness, D. M., Kozak, C. A. & Weinshilboum, R. M. Mouse nicotinamide N-methyltransferase gene: molecular cloning, structural characterization, and chromosomal localization. *DNA Cell Biol.* **17**, 659–667 (1998).
7. Hughes, T. R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnol.* **19**, 342–347 (2001).
8. Drake, T. A. *et al.* Genetic loci determining bone density in mice with diet-induced atherosclerosis. *Physiol. Genomics* **5**, 205–215 (2001).
9. Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. & Novotny, M. V. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* **10**, 997–1004 (2001).
10. Metcalf, D. *et al.* Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* **405**, 1069–1073 (2000).
11. Swift, L. L., Valyi-Nagy, K., Rowland, C. & Harris, C. Assembly of very low density lipoproteins in mouse liver: evidence of heterogeneity of particle density in the Golgi apparatus. *J. Lipid Res.* **42**, 218–224 (2001).
12. Borecki, I. G., Rice, T., Perusse, L., Bouchard, C. & Rao, D. C. An exploratory investigation of genetic linkage with body composition and fitness phenotypes: the Quebec Family Study. *Obesity Res.* **2**, 213–219 (1994).

13. Lemberas, A. V. *et al.* Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J. Clin. Invest.* **100**, 1240–1247 (1997).
14. Damerval, C., Maurice, A., Josse, J. M. & de Vienne, D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**, 289–301 (1994).
15. Causse, M. *et al.* Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. *Mol. Breed.* **1**, 259–272 (1995).
16. Byrne, P. F. *et al.* Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl Acad. Sci. USA* **93**, 8820–8825 (1996).
17. Dausset, J. *et al.* Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
18. van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
19. Tosatio, G. Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr. Protoc. Immunol.* **1**, 7.22.1–7.22.3 (1991).
20. Troyer, A. F. Background of US hybrid corn. *Crop Sci.* **39**, 601–626 (1999).
21. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
22. Amos, C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
23. Lincoln, S. E., Daly, M. J. & Lander, E. S. *MAPMAKER/QTL User's Manual* (Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 1993).
24. Basten, C. A., Weir, B. S. & Zeng, Z. B. *QTL Cartographer User's Manual* (Department of Statistics, North Carolina State University, Raleigh, 1999).
25. Fisch, R. D., Ragot, M. & Gay, G. A generalization of the mixture model in the mapping of quantitative trait loci for progeny from a biparental cross of inbred lines. *Genetics* **143**, 571–577 (1996).

Supplementary Information accompanies the paper on Nature's website (♦ <http://www.nature.com/nature>).

Acknowledgements Rosetta Inpharmatics is a wholly owned subsidiary of Merck & Co., Inc.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (♦ <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to E.E.S. (e-mail: eric_schadt@merck.com) or S.H.F. (e-mail: stephen_friend@merck.com).

Glycine binding primes NMDA receptor internalization

Yi Nong*†‡, Yue-Qiao Huang*†‡, William Ju*‡§, Lorraine V. Kalia*||, Gholamreza Ahmadian*§, Yu Tian Wang*§¶ & Michael W. Salter*†||

* Programme in Brain and Behaviour, Hospital for Sick Children, and † Department of Physiology, § Department of Pathobiology and Laboratory Medicine, and || Institute of Medical Science, University of Toronto, Toronto, Ontario M5G 1X8, Canada

¶ Department of Medicine and Brain Research Centre, Vancouver Hospital and Health Sciences Centre, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

‡ These authors contributed equally to this work

NMDA (N-methyl-D-aspartate) receptors (NMDARs) are a principal subtype of excitatory ligand-gated ion channel with prominent roles in physiological and disease processes in the central nervous system¹. Recognition that glycine potentiates NMDAR-mediated currents² as well as being a requisite co-agonist of the NMDAR subtype of 'glutamate' receptor³ profoundly changed our understanding of chemical synaptic communication in the central nervous system. The binding of both glycine and glutamate is necessary to cause opening of the NMDAR conductance pore¹. Although binding of either agonist alone is insufficient to cause current flow through the channel, we report here that stimulation of the glycine site initiates signalling through the NMDAR complex, priming the receptors for clathrin-dependent endocytosis. Glycine binding alone does not cause the receptor to

be endocytosed; this requires both glycine and glutamate site activation of NMDARs. The priming effect of glycine is mimicked by the NMDAR glycine site agonist D-serine, and is blocked by competitive glycine site antagonists. Synaptic as well as extrasynaptic NMDARs are primed for internalization by glycine site stimulation. Our results demonstrate transmembrane signal transduction through activating the glycine site of NMDARs, and elucidate a model for modulating cell–cell communication in the central nervous system.

The turnover of cell surface NMDARs is much slower than that of other ligand-gated ion channels, such as AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)^{4–9} and GABA_A (γ -aminobutyric acid A)¹⁰ receptors, and NMDARs are often considered 'stable' in the plasma membrane. However, recent evidence has raised the possibility that NMDARs may undergo regulated transport to and from the cell surface^{11–13}. Here, we recorded whole-cell currents from neurons acutely isolated from the CA1 region of the hippocampus. We probed NMDAR activity with regularly timed test applications of NMDA (50 μ M) plus glycine (1 μ M), and observed that the NMDAR-mediated currents were stable for recording periods of 45 min or longer (Fig. 1a). In contrast, on administering a conditioning stimulus for NMDARs (1 mM NMDA plus 100 μ M glycine; 2–3 min duration) there was a sharp decline in the NMDAR currents followed by a progressive decrease to a level $57 \pm 3\%$ (mean \pm s.e.m.; $n = 6$ cells) of the control NMDAR current amplitude (Fig. 1b). The depression of NMDAR currents was stable and the currents did not recover even during the longest recordings, up to 1 h. The depression of NMDAR currents was associated with no change in reversal potential, no loss of neuronal integrity and no alteration in single channel conductance or mean open time (see Supplementary Information 1).

It is possible that the decline in NMDAR currents was due to a decrease in the number of cell-surface receptors. Thus, we determined whether the depression of NMDAR currents might be due to clathrin-dependent receptor internalization. For this we used an amphiphysin SH3 domain recombinant protein^{6,14}, which prevents recruitment of dynamin to clathrin-coated pits and thereby blocks endocytosis. We found that when wild-type amphiphysin SH3 domain protein was administered intracellularly during whole-cell recording, the co-agonist conditioning stimulation produced a small, transient decline in NMDAR current amplitude but no sustained decrease in the currents (Fig. 1c). In contrast, when we administered an amphiphysin SH3 domain protein with a double point mutation (G684R, P687L)¹⁵ that renders it incapable of blocking endocytosis, the co-agonist stimulation caused a sustained decline in NMDAR currents (Fig. 1c) that was not different from that in control recordings.

To determine directly whether conditioning of NMDARs leads to loss of the receptors from the cell surface as a consequence of clathrin-dependent receptor endocytosis, we used a cell enzyme-linked immunosorbent assay (ELISA) (Fig. 1d) with hippocampal neurons in primary culture. Using control extracellular bathing solution we found that there was little internalization of NMDARs during the 15-min experimental period (Fig. 1d). On the other hand, exposing the cells to the conditioning stimulus for 5 min caused $27 \pm 3\%$ ($n = 6$) of the pre-labelled NMDARs to be internalized over the subsequent 10-min period (Fig. 1d). To test whether the internalization of NMDARs required clathrin assembly we used extracellular hypertonic sucrose solution, known to block clathrin-mediated endocytosis¹⁶, which was found to prevent the internalization (Fig. 1d, top). NMDAR internalization was produced when the concentration of NMDA in the co-agonist conditioning stimulus was reduced to 50 μ M (Fig. 1d, bottom). In contrast, neither the application of NMDA nor glycine alone caused internalization of NMDARs. Furthermore, applying the competitive NMDAR antagonist AP5 (D-2-amino-5-phosphonovaleric acid)¹⁷—which binds to the glutamate-binding site but does not