15. Kiss, I. Z., Zhai, Y. & Hudson, J. L. Emerging coherence in a population of chemical oscillators. *Science* **296,** 1676–1678 (2002).
16. Levine, J. D., Funes, P., Dowse, H. B. & Hall, J. C. Resetting the circadian clock by social experience in *Drosophila melanogaster. Science* **298,** 2010–2012 (2002).
17. Winfree, A. T. *The Geometry of Biological Time*, 2nd edn (Springer, New York, 2001).
18. Goldbeter, A. Computational approaches to cellular rhythms. *Nature* **420,** 238–245 (2002).
19. Gonze, D., Halloy, J. & Goldbeter, A. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl Acad. Sci. USA* **99,** 673–678 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank S. Golden for the AMC412 strain and for advice, D. Peoples for advice and technical assistance, B. Houchmandzadeh, J. Paulsson, J. Vilar, C. Weitz and M. Young for discussions, and N. Questembert-Balaban, E. Kussell and M. Vallade for comments on the manuscript. This work was supported partially by Princeton University through the Lewis Thomas Fellowship (I.M.), the National Institutes of Health, the Howard Hughes Medical Institute and the Centre National de Recherche Scientifique through an ATIP and an AC 'Dynamique et réactivité des assemblages biologiques'.

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

Correspondence and requests for materials should be addressed to I.M. (imihalce@spectro.ujf-grenoble.fr).

Evolutionary changes in *cis* and *trans* gene regulation

Patricia J. Wittkopp, Belinda K. Haerum & Andrew G. Clark

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

Differences in gene expression are central to evolution. Such differences can arise from *cis*-regulatory changes that affect transcription initiation, transcription rate and/or transcript stability in an allele-specific manner, or from *trans*-regulatory changes that modify the activity or expression of factors that interact with *cis*-regulatory sequences[1,2]. Both *cis*- and *trans*-regulatory changes contribute to divergent gene expression, but their respective contributions remain largely unknown[3]. Here we examine the distribution of *cis*- and *trans*-regulatory changes underlying expression differences between closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, and show functional *cis*-regulatory differences by comparing the relative abundance of species-specific transcripts in F_1 hybrids[4,5]. Differences in *trans*-regulatory activity were inferred by comparing the ratio of allelic expression in hybrids with the ratio of gene expression between species. Of 29 genes with interspecific expression differences, 28 had differences in *cis*-regulation, and these changes were sufficient to explain expression divergence for about half of the genes. *Trans*-regulatory differences affected 55% (16 of 29) of genes, and were always accompanied by *cis*-regulatory changes. These data indicate that interspecific expression differences are not caused by select *trans*-regulatory changes with widespread effects, but rather by many *cis*-acting changes spread throughout the genome.

D. melanogaster and *D. simulans* diverged about 2.5 Myr ago[6], yet are still able to mate and produce viable (but sterile) offspring. Their compatibility allowed us to identify *cis*- and *trans*-regulatory changes by comparing the regulation of species-specific alleles in a common, hybrid genetic background. Differential expression of two alleles in the same cellular environment indicates functional *cis*-regulatory differences[4,5,7]; thus, asymmetric allelic expression in F_1 hybrids implies *cis*-regulatory divergence. If *trans*-regulation diverges between species, the collection of *trans*-acting factors in hybrids will be different from that in one or both of the parental species. As a result, the relative allelic expression in hybrids will differ from the relative gene expression between species. *Trans*-regulatory divergence was inferred for any gene with significant differences in the ratio of species-specific transcripts between F_1 hybrids and the parental species.

To measure the relative abundances of species-specific transcripts, we analysed RNA and DNA, extracted separately from pools of female flies containing either 14 F_1 hybrids or 7 *D. melanogaster* and 7 *D. simulans* individuals (Fig. 1a). Four complementary DNA samples were generated from each RNA extraction and used to measure the relative abundance of species-specific transcripts. At least two replicate hybrid pools (eight cDNA measurements in total) and four replicate parental pools (16 cDNA measurements in total) were analysed for each gene. The ratio of *D. melanogaster* to *D. simulans* alleles was also measured in genomic DNA from each pool in duplicate and used to correct cDNA ratios for allelic differences in extraction and/or amplification (Supplementary Information). For each gene, the ratio of expression between species (designated Mel/Sim) and the ratio of species-specific transcripts in F_1 hybrids (designated Mel_{F1}/Sim_{F1}) were quantified, normalized and averaged across replicate pools.

Pyrosequencing[8] was used to measure the relative abundance of *D. melanogaster* and *D. simulans* alleles directly in cDNA and DNA samples. A single nucleotide difference that distinguished the species-specific transcripts was identified for each gene, and polymerase chain reaction (PCR) primers were used to amplify a region of DNA containing the divergent site (Fig. 1b). After denaturing the PCR products, an internal primer ('Pyro-reverse' in Fig. 1b) was annealed and extended one base at a time in a Pyrosequencing

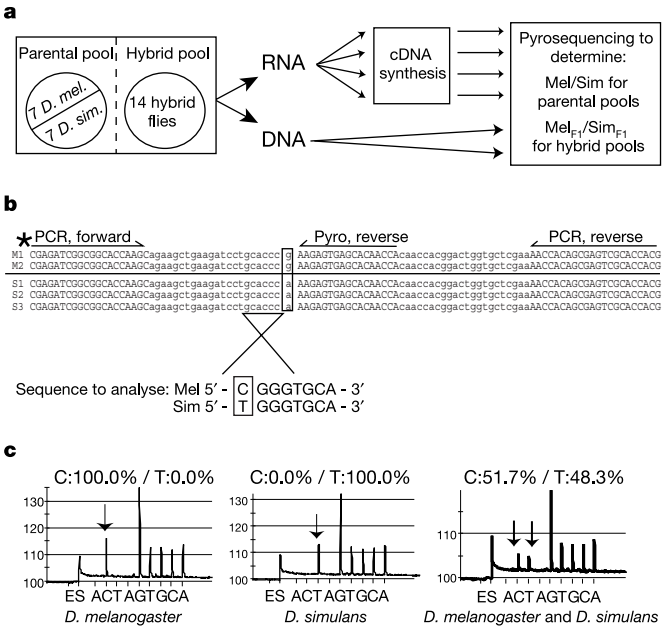

Figure 1 Pyrosequencing[8] measures allelic gene expression. **a**, Overview of sample collection. **b**, Example assay of CG14770. Aligned coding sequences from *D. melanogaster* (M1, M2) and *D. simulans* (S1–S3) identify transcribed fixed differences (box). The location of PCR and Pyrosequencing primers are shown with an asterisk indicating the biotinylated PCR primer. **c**, Pyrosequencing produces a pyrogram with peaks whose height is directly proportional to the quantity of nucleotide added to the extending Pyrosequencing primer. Pyrograms from reactions using genomic DNA from *D. melanogaster* (left), *D. simulans* (middle), and both species (right) are shown. Arrows indicate peaks reflecting nucleotide incorporation at the divergent site. ES indicates the addition of enzymes and substrate, respectively.

reaction, which couples DNA synthesis with a series of enzymatic reactions that produce stoichiometric quantities of light[9]. The amount of light generated by the incorporation of species-specific bases was directly proportional to the relative abundance of the *D. melanogaster* and *D. simulans* alleles in the PCR template (Fig. 1c, Supplementary Fig. 1).

Thirty-four genes previously shown to have a significant expression difference between *D. melanogaster* and *D. simulans* were selected for analysis[10]; of these genes, 29 showed significant expression differences between species in this study, ranging from 1.1-fold to 6.7-fold (Fig. 2a, *t*-test, H_0: Mel/Sim = 1, $P < 0.05$). Because we examined a different developmental stage in different strains of flies from those in the initial study[10], it is not surprising that we observed distinct expression levels. Comparison with an independent microarray study of expression differences between *D. melanogaster* and *D. simulans* adults[11] indicates that this gene set is representative of interspecific expression differences genome-wide (Supplementary Information).

Unequal expression of the *D. melanogaster* and *D. simulans* alleles in a hybrid genetic background indicates differences in *cis*-regulation. Allelic expression of species-specific alleles in F_1 hybrids showed that nearly all (28 of 29) genes with interspecific expression differences had evolved changes in *cis*-regulatory function (Fig. 2b; *t*-test, H_0: $Mel_{F1}/Sim_{F1} = 1$, $P < 0.05$). Differential expression of species-specific alleles ranged from 1.1-fold to 6.7-fold, with the more highly expressed allele derived from *D. melanogaster* half of the time. Two of the five genes with similar expression between species also showed evidence of *cis*-regulatory divergence (Fig. 2a, b, arrows; *t*-test, H_0: $Mel_{F1}/Sim_{F1} = 1$, $P < 0.05$), indicating that compensatory *cis*- and *trans*-regulatory changes have evolved that maintain gene expression. To ensure that differences in allelic expression were caused by *cis*-regulatory changes rather than parent-of-origin effects, we compared the ratio of species-specific transcripts for 24 genes in F_1 hybrids from reciprocal crosses. The direction of cross used to produce the hybrids generally had a negligible effect on the expression levels of the *D. melanogaster* and *D. simulans* alleles (Supplementary Fig. 2).

Cis-regulatory differences seem to be extremely common between orthologous genes in *D. melanogaster* and *D. simulans*. But are changes in *cis*-regulatory function usually the primary cause of interspecific expression differences, or do they contribute relatively little in comparison with genetic changes affecting *trans*-regulation? To address this question, we compared the expression difference between species (Mel/Sim) with the relative abundance of species-specific transcripts in F_1 hybrids (Mel_{F1}/Sim_{F1}; Fig. 3). If *cis*-regulatory divergence completely explains the difference between species, the hybrid and parental expression ratios will be the same ($Mel/Sim = Mel_{F1}/Sim_{F1}$) and points in Fig. 3 will fall on the diagonal. In contrast, if only *trans*-regulatory differences cause the divergent expression, the *D. melanogaster* and *D. simulans* alleles will be equally expressed in F_1 hybrids ($Mel_{F1}/Sim_{F1} = 1$) and

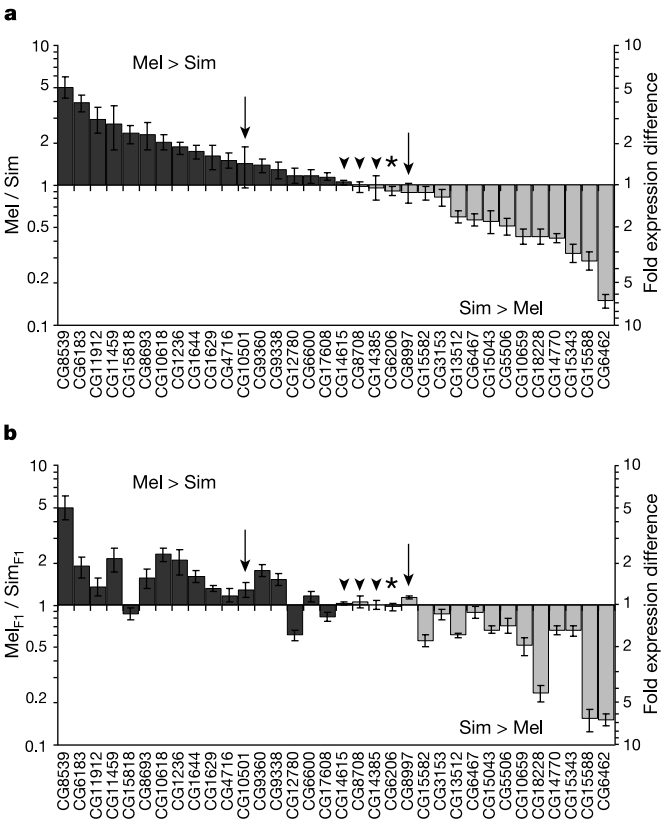

Figure 2 *Cis*-regulatory changes frequently contribute to interspecific expression differences. **a**, Relative gene expression levels between species (Mel/Sim). **b**, Relative allelic expression in F_1 hybrids (Mel_{F1}/Sim_{F1}). The gene order is the same in both panels, with data plotted on a logarithmic scale. The normalized ratio of *D. melanogaster* to *D. simulans* transcripts is shown at the left; the corresponding fold expression changes are shown at the right. Error bars show 95% confidence intervals. Arrows indicate genes with compensatory *cis*- and *trans*-regulatory changes, and arrowheads show genes with conserved *cis*- and *trans*-regulation. Asterisks indicate CG6206, the only gene that showed divergent expression without significant *cis*-regulatory differences.

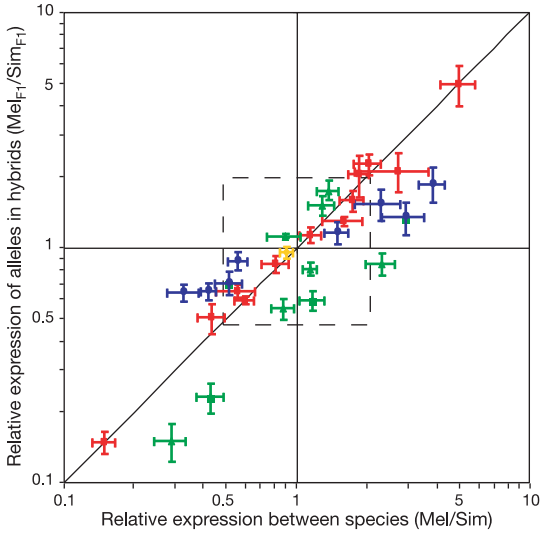

Figure 3 Evolutionary changes in both *cis*- and *trans*-regulation underlie interspecific expression differences. The relative expression between species (Mel/Sim) is plotted against the relative expression of species-specific alleles in hybrids (Mel_{F1}/Sim_{F1}) on a logarithmic scale. The dashed box indicates twofold expression differences. Error bars show 95% confidence intervals. Genes are colour-coded according to our inference of *cis*- and *trans*-regulatory changes: red, *cis*-regulatory changes can completely explain interspecific expression differences; blue, *cis*-regulatory changes account for a fraction of the expression difference, with *trans*-regulatory changes explaining the remaining fraction; green, *cis*- and *trans*-regulatory changes with antagonistic effects have evolved; yellow, CG6206, which displayed no evidence of *cis*- or *trans*-regulatory divergence ($P = 0.09$ and $P = 0.25$, respectively) despite a small significant expression difference between species (Mel/Sim = 0.91, $P = 0.005$). The diagonal line indicates 100% *cis*-regulatory divergence; the bold horizontal line indicates 100% *trans*-regulatory divergence.

points will fall along the horizontal axis. Genes affected by both *cis*- and *trans*-regulatory divergence are indicated in Fig. 3 by points that do not fall on either of these lines.

Cis-regulatory changes could completely explain interspecific expression differences for almost half (12 of 28) of the genes with divergent *cis*-regulatory functions (red in Fig. 3; *t*-test, H_0: Mel/Sim = Mel_{F1}/Sim_{F1}, $P > 0.05$). The remaining 16 genes displayed evidence of both *cis*- and *trans*-regulatory changes ($Mel_{F1}/Sim_{F1} \neq 1$ and $Mel_{F1}/Sim_{F1} \neq$ Mel/Sim). For eight genes, differences in *cis*-regulation were less extreme than the interspecific expression difference (blue in Fig. 3). Genetic changes affecting *trans*-acting factors are responsible for the remaining expression difference between species. The other eight genes showed allelic expression differences in F_1 hybrids that were greater than, or in the opposite direction from, the interspecific differences (green in Fig. 3). Evolutionary changes in *cis*- and *trans*-regulation with opposing effects on gene expression must underlie interspecific differences for these genes. Such changes are consistent with the coevolution of *cis*- and *trans*-regulatory factors. Overall, *cis*-regulatory differences affected 97% of genes with divergent expression between *Drosophila* species, with at least 57% also influenced by *trans*-regulatory changes. The presence of *trans*-regulatory differences was not related to the magnitude of the expression difference between species (Kruskal–Wallis test, $P = 0.9$).

The distribution of *cis*- and *trans*-regulatory changes provides insight into the evolution of regulatory networks. To a first approximation, these networks are composed of two classes of gene[12]: regulatory (for example those encoding transcription factors or signalling molecules) and structural (for example those encoding enzymes or cellular components). Regulatory genes (Fig. 4, filled circles) comprise internal connections in the network, and changes in their expression often have *trans*-acting effects on the expression of many other genes. In contrast, structural genes (Fig. 4, open circles) typically lie at terminal nodes of the network where changes in their expression have more limited effects on expression of other genes. On the basis of the available annotations of molecular function[13,14], our data set contains primarily structural genes (Supplementary Table 1). Analysis of structural genes provides a complete view of the regulatory network, with gene-specific changes at terminal nodes in the network detected as *cis*-regulatory changes and genetic changes located in upstream regulatory genes identifiable as *trans*-regulatory differences.

Two extreme models can be considered to explain the myriad differences in gene expression between species: a few *cis*-regulatory changes affecting regulatory genes could have widespread *trans*-acting effects on the expression of many downstream genes (Fig. 4a), or each structural gene's expression could be altered by its own *cis*-acting genetic changes (Fig. 4b). We found that interspecific expression differences were almost always caused, at least in part, by *cis*-regulatory changes in structural genes, and that differences in *trans*-regulation also affected half of the genes (Fig. 4c). The prevalence of *cis*-regulatory changes suggests that differences in gene expression between *D. melanogaster* and *D. simulans* could have evolved by changing the expression of structural genes one gene at a time. However, regulatory differences observed in extant flies might not have been the original source of expression divergence.

With the availability of DNA microarrays, whole genomes can now be compared for differences in gene expression within and between species. By combining expression analysis with genetic methods, such as quantitative trait locus mapping or the approach presented here, the genetic basis of variable gene expression is beginning to be elucidated[4,5,15–18]. Distinguishing between *cis*- and *trans*-regulatory changes is the first step in this process, and it provides a crucial foothold for identifying the specific nucleotide changes underlying gene expression differences. □

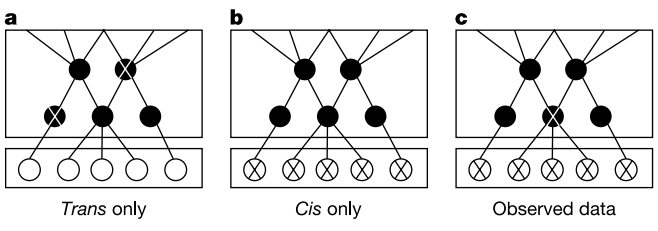

Figure 4 Models of regulatory divergence. A hypothetical regulatory network is shown with filled circles representing regulatory genes, open circles representing structural genes (all with altered expression), and lines representing regulatory interactions. Crosses indicate genes harbouring a regulatory change. **a**, **b**, Two extreme models of regulatory divergence in which either *trans*-regulatory differences (**a**) or *cis*-regulatory differences (**b**) are solely responsible for expression differences. **c**, A distribution of *cis*- and *trans*-regulatory changes consistent with our data.

Methods

Fly strains and crosses

A *D. melanogaster zygotic hybrid rescue* (*zhr*) strain (provided by A. Orr) and the Tsimbazaza *D. simulans* strain (provided by H. Hollocher) were crossed to produce F_1 hybrids. These strains reduce premating isolation between *D. melanogaster* and *D. simulans* and rescue the lethality of hybrid females[19–22]. Female flies from parental lines and F_1 hybrid females from reciprocal interspecific crosses were collected within 4 h of emergence, aged for 24 h, snap-frozen in liquid nitrogen, and stored at −70 °C. The *D. melanogaster* In(1)AB strain and *D. simulans* C167.4 and *male hybrid rescue* (*mhr*) strains (provided by D. Barbash) were also used for DNA sequencing.

Gene selection

Genes were selected on the basis of the microarray data of Rifkin *et al.*[10], which measured changes in gene expression during metamorphosis between *D. melanogaster* and *D. simulans*. Each gene selected showed evidence of lineage-specific selection and had expression in *D. simulans* that was at least twofold different from all four strains of *D. melanogaster* analysed. Because morphological identification of F_1 hybrid females during the larval and pupal stages examined by Rifkin *et al.* is difficult, we analysed adult female flies instead. In addition, because the fly strains used by Rifkin *et al.*[10] do not mate, we used the *zhr* and Tsimbazaza strains.

Extraction of nucleic acids and preparation of cDNA

For each pool of 14 flies, DNA and RNA were extracted separately from a single homogenate by using a modified protocol for the SV Total RNA Isolation System (Promega)[23]. The homogenate was passed over a column that retained DNA; the flow-through (containing RNA and proteins) was then passed over a second column that bound RNA. RNA samples were treated with DNase during extraction and immediately before cDNA synthesis. RNA was reverse transcribed into cDNA with the use of a poly(T) primer and standard protocols. PCR reactions with primers annealing to introns confirmed that cDNA samples were free of genomic DNA. Protocols are available from the authors on request.

Allele quantification with Pyrosequencing

Using primer sequences listed in ref. 10, we amplified and sequenced 200–800 base pairs of transcribed DNA from each gene in two *D. melanogaster* and three *D. simulans* strains. After the identification of species-specific nucleotide differences, PCR primers annealing to conserved sequences were used to amplify regions of sequence including the divergent sites. An internal primer, also matching a conserved sequence, was then annealed. Pyrosequencing reactions, performed in accordance with the manufacturer's instructions (www.pyrosequencing.com), were used to measure the relative abundance of the two alleles in genomic DNA and cDNA samples from both parental and hybrid pools. Primer sequences and PCR annealing temperatures are available from the authors on request. Pyrosequencing software reports a peak height directly proportional to the number of molecules incorporated into the growing DNA chain. Custom Perl scripts were used to calculate the ratio of peak heights associated with incorporation of allele-specific bases (Mel_{F1}/Sim_{F1} or Mel/Sim), which corresponds to the relative abundance of the *D. melanogaster* and *D. simulans* alleles in the starting sample (Supplementary Fig. 1). cDNA ratios were normalized with genomic DNA measurements as described in Supplementary Information. Because both alleles are extracted and measured in a single sample, this method is insensitive to differences in extraction efficiency and eliminates the need for 'control' genes or quantification of total RNA recovery.

Statistical analysis

After normalization, replicate pools were compared for each gene by using analysis of variance. Significant heterogeneity between pools was observed for 31 of 34 parental pools, but for only 9 of 34 hybrid pools (Supplementary Fig. 3). This variation can be caused either by residual genetic variation within inbred strains or, more probably, by environmental variation introduced during fly rearing. However, because no single pool

was a consistent outlier for all genes, data from all pools were included in our analysis. Differences between parental pools increase our confidence intervals for the interspecific expression difference and can reduce our power for detecting *trans*-regulatory differences. To deal with this issue, interspecific expression differences for genes with unusually large variance between the four original pools were measured in an additional three parental pools.

The normalization procedure used to correct cDNA measurements for experimental bias (Supplementary Information) prohibits a standard nested analysis of variance, but a *t*-test provided a simple and robust test of our null hypotheses. Two-tailed *t*-tests were used to identify *cis*-regulatory divergence (H_0: $Mel_{F1}/Sim_{F1} = 1$), interspecific expression differences (H_0: Mel/Sim = 1) and parent-of-origin effects (H_0: $Mel_{F1}/Sim_{F1} = Mel_{F1}/Sim_{F1}$ in reciprocal crosses). To identify *trans*-regulatory divergence, two-sided *t*-tests (with the Cochran correction for unequal variances) and nonparametric Mann–Whitney *U*-tests were used to compare relative expression between hybrid and parental pools. The decision to accept or reject the null hypothesis (H_0: Mel_{F1}/Sim_{F1} = Mel/Sim) was the same for both tests for all except three genes, and *t*-test significance was ultimately used to infer *trans*-regulatory divergence. All statistical analyses were preformed with SAS software v. 8.2 (SAS Institute, Cary, North Carolina) and are shown in Supplementary Table 2.

Received 29 February; accepted 1 June 2004; doi:10.1038/nature02698.

1. Davidson, E. H. *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego, 2001).
2. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Oxford, 2001).
3. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20,** 1377–1419 (2003).
4. Cowles, C. R., Hirschhorn, J. N., Altshuler, D. & Lander, E. S. Detection of regulatory variation in mouse genes. *Nature Genet.* **32,** 432–437 (2002).
5. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297,** 1143 (2002).
6. Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, New York, 1997).
7. Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genomics* **16,** 184–193 (2004).
8. Ahmadian, A. *et al.* Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280,** 103–110 (2000).
9. Neve, B. *et al.* Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *Biotechniques* **32,** 1138–1142 (2002).
10. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33,** 138–144 (2003).
11. Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila transcriptome. Science* **300,** 1742–1745 (2003).
12. Davidson, E. H., McClay, D. R. & Hood, L. Regulatory gene networks and the properties of the developmental process. *Proc. Natl Acad. Sci. USA* **100,** 1475–1480.
13. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25,** 25–29 (2000).
14. The FlyBase Consortium, The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **31,** 172–175 (2003).
15. Yvert, G. *et al. Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35,** 57–64 (2003).
16. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
17. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302 (2003).
18. Montooth, K. L., Marden, J. H. & Clark, A. G. Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila. Genetics* **165,** 623–635 (2003).
19. Davis, A. W. *et al.* Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans. Nature* **380,** 157–159 (1996).
20. Sawamura, K., Yamamoto, M. T. & Watanabe, T. K. Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The *Zygotic hybrid rescue* (*Zhr*) gene of *D. melanogaster. Genetics* **133,** 307–313 (1993).
21. Hollocher, H., Agopian, K., Waterbury, J., O'Neill, R. W. & Davis, A. W. Characterization of defects in adult germline development and oogenesis of sterile and rescued female hybrids in crosses between *Drosophila simulans* and *Drosophila melanogaster. J. Exp. Zool.* **288,** 205–218 (2000).
22. Barbash, D. A., Roote, J. & Ashburner, M. The *Drosophila melanogaster* hybrid male rescue gene causes inviability in male and female species hybrids. *Genetics* **154,** 1747–1771 (2000).
23. Otto, P., Kephart, D., Bitner, R., Huber, S. & Volkerding, K. Separate isolation of genomic DNA and total RNA from single samples using the SV Total RNA Isolation System. *Promega Notes* **69,** 19–24 (1998).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank A. Kumari and S. Madhavarapu for experimental assistance, K. Montooth for statistical advice, and K. Montooth, B. Payseur, A. Fiumera, T. Schlenke and E. Hill for comments on the manuscript. Funding for this project was provided by NIH grants to A.G.C. P.J.W. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation, and B.K.H. was funded by a Howard Hughes Undergraduate Research award.

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

Correspondence and requests for materials should be addressed to P.W. (pw72@cornell.edu).

Evidence for dynamically organized modularity in the yeast protein–protein interaction network

Jing-Dong J. Han[1], Nicolas Bertin[1], Tong Hao[1], Debra S. Goldberg[2], Gabriel F. Berriz[2], Lan V. Zhang[2], Denis Dupuy[1], Albertha J. M. Walhout[1]*, Michael E. Cusick[1], Frederick P. Roth[2] & Marc Vidal[1]

[1]*Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA*
[2]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA*

* Present address: Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

In apparently scale-free protein–protein interaction networks, or 'interactome' networks[1,2], most proteins interact with few partners, whereas a small but significant proportion of proteins, the 'hubs', interact with many partners. Both biological and non-biological scale-free networks are particularly resistant to random node removal but are extremely sensitive to the targeted removal of hubs[1]. A link between the potential scale-free topology of interactome networks and genetic robustness[3,4] seems to exist, because knockouts of yeast genes[5,6] encoding hubs are approximately threefold more likely to confer lethality than those of non-hubs[1]. Here we investigate how hubs might contribute to robustness and other cellular properties for protein–protein interactions dynamically regulated both in time and in space. We uncovered two types of hub: 'party' hubs, which interact with most of their partners simultaneously, and 'date' hubs, which bind their different partners at different times or locations. Both *in silico* studies of network connectivity and genetic interactions described *in vivo* support a model of organized modularity in which date hubs organize the proteome, connecting biological processes—or modules[7]—to each other, whereas party hubs function inside modules.

The biological role of topological hubs, so far considered in static representations of interactome networks without information on the functional states of these networks—that is, dynamic or steady state[8]—might vary depending on the timing and location of the interactions they mediate (Fig. 1a). Because accurate temporal parameters are not yet available for many protein–protein interactions, we estimated temporal characteristics of hubs and their partners by using compilations of yeast messenger RNA expression profiling data[9].

Hubs connected by false-positive interactions[10] would be uncorrelated in mRNA expression with their interaction partners[9,11], and would resemble date hubs. To minimize false positives, we first generated a high-quality yeast interaction data set by intersecting data generated by several different interaction detection methods (see Methods). The resulting 'filtered yeast interactome' (FYI) data set contains 2,493 high-confidence interactions, each observed by at least two different methods (Supplementary Fig. 1). FYI is a high-quality network enriched for genuine positives (Supplementary Information and Supplementary Fig. 2). The FYI network contains 1,379 proteins with an average degree of 3.6 interactions per protein and a large connected component of 778 proteins. Its degree distribution follows the power law that characterizes scale-free networks (Supplementary Fig. 3). FYI hubs were characterized with an expression-profiling compendium of 315 data points for most yeast genes across five different experimental conditions (referred to below as the 'yeast expression compendium'[9]). For each hub we calculated the average of Pearson