beater's Possum using Population Viability Analysis (Centre for Resource and Environmental Studies, Australian National University, Canberra, 1995).
21. M. D. Fox, D. Adamson, in A Natural Legacy, H. F. Recher, D. Lunney, I. Dunn, Eds. (Pergamon Press, Sydney, Australia, 1986), pp. 235–255.
22. J. A. Friend, Proc. Ecol. Soc. Aust. 16, 369 (1990).
23. C. Bailey, Landscope 11, 41 (1996).
24. J. E. Kinnear, M. L. Onus, N. R. Sumner, Wildl. Res. 25, 81 (1998).
25. G. B. A. Da Fonseca, Biol. Conserv. 34, 17 (1985).
26. V. G. Roig, in Latin American Mammalogy: History, Biodiversity, and Conservation, M. A. Mares, D. J. Schmidly, Eds. (Univ. Oklahoma Press, Norman, OK, 1991), pp. 239–279.
27. G. B. Corbet, H. N. Southern, Eds., The Handbook of British Mammals (Blackwell Scientific, London, 1977).
28. A. J. Mitchell-Jones et al., The Atlas of European Mammals (Academic Press, London, 1999).
29. J. A. Scott, The Butterflies of North America: A Natural History and Field Guide (Stanford University Press, Stanford, CA, 1986).
30. D. D. Murphy, S. B. Weiss, Biol. Conserv. 46, 183 (1988).
31. R. A. Mittermeier, N. Myers, P. R. Gil, C. G. Mittermeier, Hot Spots: Earth's Biologically Richest and Most Endangered Terrestrial Ecosystems (CEMEX and Conservation International, Monterrey, Mexico, 1999).
32. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, Nature 403, 853 (2000).
33. G. Ceballos, J. H. Brown, Conserv. Biol. 9, 559 (1995).
34. C. Santiapillai, K. MacKinnon, in Rhinoceros Biology and Conservation, O. A. Ryder, Ed. (San Diego Zoological Society, San Diego, CA, 1993), pp. 257–264.
35. We thank I. Salazar, G. Oliva, and J. Pacheco for helping gathering the data and with the spatial analyses. We are indebted to J. Brown and G. Daily for extensive discussions of the ideas presented here and for reviewing the manuscript. J. Diamond, G. Luck, M. Mayfield, H. Mooney, S. Pimm, J. Ranganathan, and C. Sekercioglu also kindly criticized drafts of this paper. The comments by M. Lomolino and an anonymous reviewer improved the manuscript. Funded by the Universidad Nacional Autónoma de México, the Koret Foundation, the Joyce Mertz-Gilmore Foundation, and the Winslow Foundation.

Genomewide Analysis of mRNA Processing in Yeast Using Splicing-Specific Microarrays

Tyson A. Clark,[1–3] Charles W. Sugnet,[2–4] Manuel Ares, Jr.[1–3]*

Introns interrupt almost every eukaryotic protein-coding gene, yet how the splicing apparatus interprets the genome during messenger RNA (mRNA) synthesis is poorly understood. We designed microarrays to distinguish spliced from unspliced RNA for each intron-containing yeast gene and measured genomewide effects on splicing caused by loss of 18 different mRNA processing factors. After accommodating changes in transcription and decay by using gene-specific indexes, functional relationships between mRNA processing factors can be identified through their common effects on spliced and unspliced RNA. Groups of genes with different dependencies on mRNA processing factors are also apparent. Quantitative polymerase chain reactions confirm the array-based finding that Prp17p and Prp18p are not dispensable for removal of introns with short branchpoint-to-3' splice site distances.

Protein-coding information in eukaryotic genomes is fragmented into exons, which must be recognized and joined by the process of RNA splicing. Splicing takes place in the nucleus within a dynamic ribonucleoprotein complex called the spliceosome (1). The spliceosome transforms information within transcripts of the eukaryotic genome to create sequences not found in DNA. By its nature and position in the gene expression pathway, splicing expands the possible interpretations of genomic information and does so under developmental and environmental influence (2). Our understanding of the process of splicing is derived from studies on relatively few introns. As eukaryotic genomes are sequenced, it has become necessary to ask how the process of splicing is integrated into ge-

nome function and evolution. Compared with higher eukaryotes, yeast contains relatively few spliceosomal introns, and most have been correctly annotated (3, 4). Hence, we chose to perform genomewide study of splicing in the yeast Saccharomyces cerevisiae.

To discriminate between spliced and unspliced RNAs for each intron-containing yeast gene, we used DNA microarrays (5, 6). Oligonucleotides were designed to detect the splice junction (specific to spliced RNA and not found in the genome), the intron (present in unspliced RNA), and the second exon (common to spliced and unspliced RNA) for each intron-containing gene as shown in Figure 1A. The oligonucleotides were printed on glass slides to create splicing-sensitive microarrays for yeast (7).

To determine whether oligonucleotide arrays can function as genomewide sensors of splicing, we compared RNA of cells carrying the temperature-sensitive splicing mutation prp4-1 with RNA of wild type during a shift from 26°C to 37°C (7). Prp4p is an integral component of the spliceosome (8, 9). Plots of fluorescence (10) for each oligonucleotide for the wild-type (Cy3) versus the prp4-1 mutant

(Cy5) with time are shown in Fig. 1B. Even at the permissive temperature of 26°C, many intron probes (red spots) display Cy5/Cy3 ratios >1, indicating accumulation of intron-containing RNA in the mutant strain. After the shift to the restrictive temperature, the Cy5/Cy3 ratio increases for most intron probes. In contrast, the ratio decreases for many splice junction probes (green spots), a sign that spliced RNAs become depleted in the mutant. The Cy5/Cy3 ratios for about a thousand intronless genes remain largely unaffected (yellow spots). This indicates that the array reports catastrophic splicing defects and can measure the kinetics of splicing inhibition genomewide.

Despite their conservation, numerous mRNA processing factors are not essential in yeast. To analyze more subtle changes in splicing, we studied 18 mutant strains lacking nonessential genes implicated in mRNA processing (Table 1). Plots of mutant versus wild-type fluorescence intensities for prp18Δ, cus2Δ, and dbr1Δ are shown in Fig. 1C. The effect of each deletion on spliced and unspliced RNA is different. Most severe is prp18Δ, which causes widespread intron accumulation and loss of splice junction sequences relative to wild type (Fig. 1C, left). The cus2Δ mutation enhances defects in U2 small nuclear RNA (snRNA) or Prp5p (11, 12) but causes little intron accumulation (Fig. 1C, center). Although not required for splicing, Dbr1p debranches the lariat, and its loss results in the dramatic accumulation of intron lariats (13). In the dbr1Δ strain, most introns accumulate, and there is little effect on spliced mRNAs (Fig. 1C, right). This demonstrates that qualitative differences in splicing phenotype can be distinguished by using splicing sensitive microarrays.

Changes in spliced and unspliced RNA levels due to loss of an mRNA processing factor may arise directly from splicing inhibition or may be due to secondary events that alter transcription or RNA decay. For example, signal from a splice junction probe may increase for a gene whose transcription is induced, even though splicing is inhibited. To

[1]Department of Molecular, Cell, and Developmental Biology, [2]Center for Molecular Biology of RNA, Sinsheimer Laboratories, University of California, Santa Cruz, CA 95064, USA. [3]Center for Biomolecular Sciences and Engineering, [4]Department of Computer Sciences, Baskin School of Engineering, University of California, Santa Cruz, CA 95064, USA.

*To whom correspondence should be addressed. E-mail: ares@biology.ucsc.edu

Fig. 1. Genomewide analysis of *S. cerevisiae* splicing. **(A)** Array design. Arrays contain three oligonucleotide probes for each intron-containing gene, as well as probes for control intronless genes. Intron probes (red) detect unspliced RNA and lariats. Splice junction probes (green) detect spliced mRNA. Exon probes (blue) detect both spliced and unspliced RNAs. Data are normalized to intronless genes (yellow). **(B)** Scatter plots of probe intensities during heat shift of *prp4-1*. Raw intensity ($\log_{10}$ scale) of each spot without background subtraction or normalization is shown for heat shifted wild-type (wt) (Cy3, *x* axis) and mutant cells (Cy5, *y* axis), color-coded for probe type as in Fig. 1A. **(C)** Scatter plots of probe intensities for deletion mutants. Data are plotted as in (B).
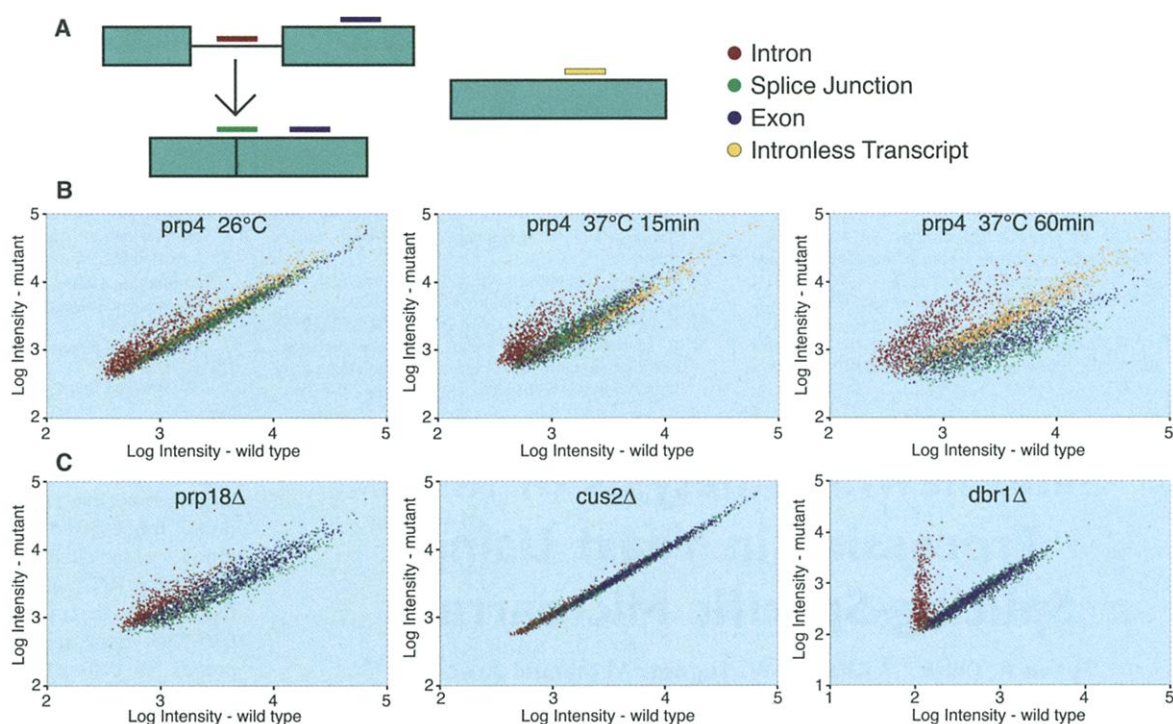
account for such effects, we devised two gene-specific indexes that relate changes in spliced and unspliced RNA to changes in total transcript level. The splice junction index (SJ) relates gain (or loss) of splice junction probe signal to gain (or loss) of total gene-derived signal as measured by the corresponding exon 2 probe. Similarly, the intron accumulation (IA) index relates changes in signal from the intron probe to its corresponding exon 2 probe (7, 14). We calculated both indexes for each intron-containing gene, clustered the indexes, and compared the relationships of the mutant strains revealed by their genomewide splicing phenotypes (Fig. 2A).

A striking conclusion from this comparison is that different mutations have distinct effects on spliced (SJ index cluster) and unspliced (IA index cluster) RNA. This means that the SJ index detects a different set of consequences of mRNA processing factor loss than the IA index. Furthermore, there appears to be no general formula to describe the relationship between the loss of spliced RNA and the accumulation of unspliced RNA. Early studies assumed a simple relationship between these processes (15) and have used the change in ratio of unspliced to spliced RNA or the increase in unspliced RNA to the total as a measure of splicing inhibition. This finding also indicates that information may be gleaned by considering the indexes separately (Fig. 2A).
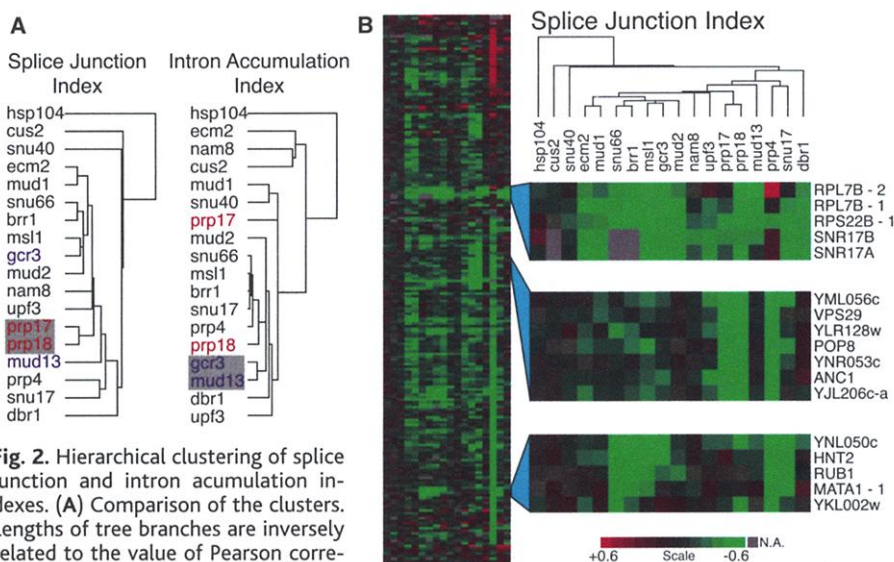
To test this, we examined the clusters in light of known functional relationships between mRNA processing factors. The IA indexes derived from loss of the two subunits of the nuclear cap binding complex Mud13p and Gcr3p (16, 17) cluster together (r = 0.88), whereas their SJ indexes do not. This indicates that the genomewide effect of their loss on intron accumulation is much more similar than their effect on splice junctions

and also is distinct from the effects of other mutations on intron accumulation (Fig. 2A). This could be due to a function of the complete nuclear cap–binding complex specific to intron-containing RNA. The failure of *mud13Δ* and *gcr3Δ* SJ indexes to cluster may be explained if one subunit has a partial function specific to spliced RNA that does not require the other subunit (18). Also nota-

Fig. 2. Hierarchical clustering of splice junction and intron accumulation indexes. **(A)** Comparison of the clusters. Lengths of tree branches are inversely related to the value of Pearson correlation coefficients of joined nodes. Shaded boxes highlight genes that are known to function together (see text). **(B)** The SJ index cluster. The 18 mutants are clustered on the horizontal axis with intron-containing genes on the vertical axis. Green squares represent a decrease in SJ index. Index values range from −3.2 to +4.3 ($\log_2$).
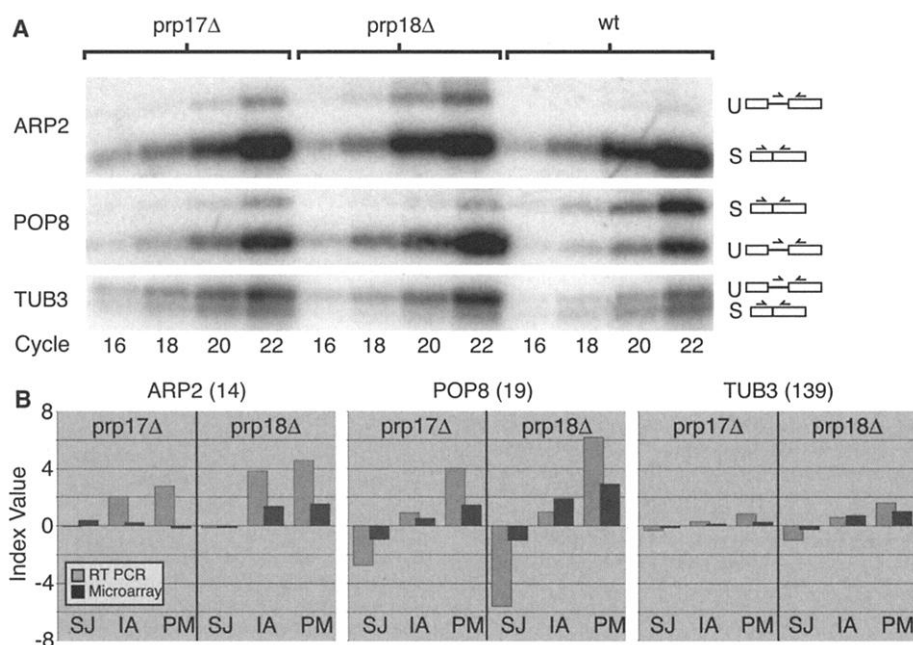
Fig. 3. RT-PCR validation of microarray data. (**A**) RT-PCR of *ARP2*, *POP8*, and *TUB3* transcripts in *prp17Δ*, *prp18Δ*, and wild-type yeast. Separate primers for spliced and unspliced RNA are used with a common downstream primer in excess. PCR products were quantified (7) by using ImageQuant software (Molecular Dynamics). (**B**) Comparison of RT-PCR and microarray data. All values are log$_2$. PhosphorImager counts for each PCR product were normalized to the average of the two intronless genes to adjust for differences in mRNA levels of the different samples. The normalized values from PCR were treated as intensity measures for intron or splice junction array probes. The ratios for total gene-derived (exon 2-containing) RNA were obtained from the ratios of the sums of the normalized spliced and unspliced counts for each gene. The PM index derived from the PCR data represents counts in unspliced RNA divided by counts in spliced RNA in the same lane (7). Numbers next to gene names indicate the distance from brp-to–3′ ss in nucleotides.

Table 1. mRNA processing genes used in this study. All strains used except *prp4-1* and its wild-type reference were derived from BY4741 (7). All genes are nonessential except *PRP4*. ORF, open reading frame; bold indicates names of genes used in this study. Additional information concerning these genes is available at the Stanford Genome Database (32).

Gene	ORF	Product
GCR3, *STO1*, *CBC1*, *CBC80*	YMR125w	Nuclear cap–binding complex subunit
MUD13, *CBC2*, *CBC20*	YPL178w	Nuclear cap–binding complex subunit
NAM8, *MRE2*, *MUD15*	YHR086w	U1 snRNP protein
MUD1	YBR119w	U1 snRNP A protein
MUD2	YKL074c	Commitment complex protein
MSL1, *YIB1*	YIR009w	U2 snRNP B″ protein
CUS2	YNL286w	U2 snRNP protein
SNU17, *IST3*	YIR005w	U2 snRNP protein
PRP4	YPR178w	U4/U6 snRNP protein
SNU40	YHR156c	U5 associated
SNU66	YOR308c	U4/U6.U5 tri-snRNP
ECM2, *SLT11*	YBR065c	U2/U6 associated, second step
PRP18	YGR006w	U5 snRNP protein, second step
PRP17	YDR364c	Second step
BRR1	YPR057w	snRNP biogenesis/recycling
UPF3	YGR072w	Nonsense-mediated decay
DBR1, *PRP26*	YKL149c	Debranching enzyme
HSP104	YLL026w	Splicing and heat shock

ble is the dissimilarity in the intron accumulation patterns of mutants lacking Prp17p and Prp18p, in contrast to their much more similar effects on splice junction levels (r = 0.82). This implies that the fate of incompletely spliced transcripts is different in these mutants, despite the expectation (supported by the SJ index) that they work together at or near the same step in splicing (*19*).

We next asked whether intron-containing genes depend on mRNA processing factors to different extents. The genomewide response to loss of individual factors is complex, suggesting a variety of dependencies (Fig. 2B,

left). The top panel (Fig. 2B, right) shows a group of genes that appear to be affected by the loss of most nonessential factors. The middle panel shows a small cluster of genes that are primarily affected by the loss of Prp17p and Prp18p, but not greatly affected by the loss of other factors. The bottom panel shows a group whose splicing is weakly affected by loss of Prp17p and Prp18p, but more severely decreased in strains lacking Snu66p, Brr1p, and Msl1p. Each intron-containing gene shares a distinct set of factor dependencies for RNA splicing with a relatively small number of other genes. These dependencies also do not align in snRNP-specific fashions, because patterns produced by loss of Mud1p and Nam8p, both U1 snRNP proteins, are distinct from each other, as are those of the U2 snRNP proteins Ecm2p, Cus2p, and Msl1p. In contrast, Mud1p and Ecm2p produce similar patterns (r = 0.83), suggesting a cooperative function between a U1 and a U2 snRNP protein.

To test the robustness of an array-based observation, we validated a small fraction of the array data relevant to a prevailing hypothesis for Prp18p function using reverse-transcriptase polymerase chain reaction (RT-PCR) (Fig. 3). Based on splicing of mutant *ACT1* reporter substrates in vitro, Prp18p is hypothesized to be dispensable for splicing when the branchpoint (brp)-to–3′ splice site (ss) distance is ≤17 nucleotides (nt) and is increasingly required in vitro as this distance increases (*20*, *21*). A comparison of brp-to–3′ ss distances with either SJ or IA index values from *prp18Δ* experiments for natural introns shows no correlation [(7) Suppl. figs. 1, 2]. Because *prp18Δ* clusters with *prp17Δ*, we included both for validation (Fig. 3B). Some genes with short brp-to–3′ ss distances are relatively unaffected by loss of Prp17p and Prp18p [e.g., *RUB1*, 12 nt, Fig. 2B, bottom right panel, PCR (*22*)]. However, two introns with short distances are detectably affected (Fig. 3B). *POP8*, with a brp-to–3′ ss distance of only 19 nt, was the intron most affected by loss of Prp18p (Fig. 3B). Conversely, several introns with long brp-to–3′ ss distances were not drastically affected. *TUB3*, containing the intron with the largest distance (139 nt), was only weakly affected (Fig. 3B). With respect to the genes we tested, RT-PCR has greater sensitivity and dynamic range than the array; however, the two kinds of data provide the same trends (Fig. 3B). This confirms changes in splicing detected by the array and suggests that hypotheses concerning mRNA processing factor function can be refined by using this approach.

To test this, we evaluated additional hypotheses concerning mRNA processing factor function in light of the array data (*7*). We find that the expectation that nonsense-mediated decay is generally important for reducing the levels of

unspliced RNA in the cytoplasm (23, 24) is not supported by the observation that the majority of these do not accumulate significantly in a upf3Δ strain [suppl. fig. 3 (7)]. The expectation based on intronic small nucleolar RNA–processing phenotypes that accumulation of introns in the dbr1Δ mutant should be inversely related to intron size (25) seems not to hold either, most likely because of Dbr1p-independent mechanisms of intron turnover (suppl. fig. 4). We do not observe correlation between a nonconsensus 5' splice site or a U-rich region near the 5' splice site and strong dependence on Nam8p (26) for splicing in vivo (Suppl. figs. 5 and 6). We also see no correlation between the presence of a U residue upstream of the branch-point sequence (27) or the presence of a poly-pyrimidine tract before or after the branchpoint and strong dependence on Mud2p (suppl. figs. 7 and 8). These data indicate that using any one intron as a reporter may cause the importance of a factor to be overemphasized or missed. Genomewide analysis allows perturbations of splicing to be evaluated on every intron at once, in effect using the entire genome as a reporter.

These studies present the first genome-wide view of splicing for any organism. The ability to distinguish differently spliced forms of RNA by using oligonucleotide microarrays opens the way for expression profiling that accounts for alternative splicing and splicing regulation in higher cells. Estimates suggest that 40 to 60% of human genes produce alternatively spliced transcripts (28, 29). In a growing number of key cases, alternatively spliced mRNAs produce proteins of distinct or even antagonistic function [e.g. (30)]. Improved expression profiling technologies must resolve changes in alternative splicing not simply by estimating exon representation [e.g. (31)], but by providing direct evidence for exon joining. The results we describe here demonstrate that oligonucleotide arrays designed to detect specific splicing products will be key to accurate parallel analysis of alternative splicing in higher organisms.

References and Notes

1. J. P. Staley, C. Guthrie, Cell 92, 315 (1998).
2. D. L. Black, Cell 103, 367 (2000).
3. C. A. Davis, L. Grate, M. Spingola, M. Ares Jr., Nucleic Acids Res. 28, 1700 (2000).
4. M. Spingola, L. Grate, D. Haussler, M. Ares Jr., RNA 5, 221 (1999).
5. J. L. DeRisi, V. R. Iyer, P. O. Brown, Science 278, 680 (1997).
6. D. J. Lockhart, E. A. Winzeler, Nature 405, 827 (2000).
7. Supplementary on Science Online at www.sciencemag.org/cgi/content/full/296/5569/907/DC1. Microarray data are available at the Gene Expression Omnibus (GEO) at www.ncbi.nlm.nih.gov/geo/, under accession numbers GSE34 and GSE35.
8. J. Banroques, J. N. Abelson, Mol. Cell. Biol. 9, 3710 (1989).
9. S. P. Bjorn, A. Soltyk, J. D. Beggs, J. D. Friesen, Mol. Cell. Biol. 9, 3698 (1989).
10. Fluorescently labeled target sequence sample preparation and hybridization were performed as described (5) by using 20 μg of total RNA primed with a mixture of oligo(dT) and random hexamers (7). Ar-

rays were scanned and analyzed by using a GenePix 4000A scanner and GenePix Pro 3.0 software from Axon Instruments (Union City, CA).
11. D. Yan et al., Mol. Cell. Biol. 18, 5000 (1998).
12. R. Perriman, M. Ares Jr., Genes Dev. 14, 97 (2000).
13. K. B. Chapman, J. D. Boeke, Cell 65, 483 (1991).
14. The splice junction (SJ) index is the ratio of the mutant/wild-type ratios derived from the normalized signals from the splice junction probe to the exon 2 probe: SJ index = SJ$_{mut}$/SJ$_{wt}$ divided by E2$_{mut}$/E2$_{wt}$, obtained by subtracting the log$_2$ ratio of the exon 2 probe from the log$_2$ ratio of the splice junction probe. The intron accumulation (IA) index is obtained by subtracting the log$_2$ ratio of the exon 2 probe from the log$_2$ ratio of the intron probe. Because probe performance may not be directly related to absolute transcript amount, these indexes depend idiosyncratically on the sequences of the probes. We also calculated the precursor/mature (PM) index, which is obtained by subtracting the log$_2$ ratio of the splice junction probe from the log$_2$ ratio of the intron probe. This index mimics the unspliced/spliced ratio used in classical splicing studies (15).
15. C. W. Pikielny, M. Rosbash, Cell 41, 119 (1985).
16. H. V. Colot, F. Stutz, M. Rosbash, Genes Dev. 10, 1699 (1996).
17. P. Fortes et al., Mol. Cell. Biol. 19, 6543 (1999).
18. E. C. Shen, T. Stage-Zimmermann, P. Chui, P. A. Silver, J. Biol. Chem. 275, 23718 (2000).
19. M. H. Jones, D. N. Frank, C. Guthrie, Proc. Natl. Acad. Sci. U.S.A. 92, 9687 (1995).
20. X. Zhang, B. Schwer, Nucleic Acids Res. 25, 2146 (1997).
21. Zhang and Schwer define the brp-to–3' ss distance starting from 2 bases downstream of the brp adenosine to the Y of the 3' ss YAG sequence. Therefore, their distance of 12 nt corresponds to 17 nt from the actual branched A to the 3' splice site G.
22. T. A. Clark, C. W. Sugnet, M. Ares Jr., unpublished observations.
23. F. He, S. W. Peltz, J. L. Donahue, M. Rosbash, A. Jacobson, Proc. Natl. Acad. Sci. U.S.A. 90, 7034 (1993).
24. M. J. Lelivelt, M. R. Culbertson, Mol. Cell. Biol. 19, 6710 (1999).
25. S. L. Ooi, D. A. Samarsky, M. J. Fournier, J. D. Boeke, RNA 4, 1096 (1998).
26. O. Puig, A. Gottschalk, P. Fabrizio, B. Seraphin, Genes Dev. 13, 569 (1999).
27. J. C. Rain, P. Legrain, EMBO J. 16, 1759 (1997).
28. E. S. Lander et al., Nature 409, 860 (2001).
29. B. Modrek, C. Lee, Nature Genet. 30, 13 (2002).
30. L. H. Boise et al., Cell 74, 597 (1993).
31. D. D. Shoemaker et al., Nature 409, 922 (2001).
32. http://genome-www.stanford.edu/Saccharomyces/
33. Thanks go to H. Igel for technical assistance; and to J. DeRisi, T. Ferea, T. Furey, G. Hartzog, D. Haussler, H. Heynecker, R. Perriman, T. Powers, L. Shiue, and R. Samaha for advice. Funding was provided by a W. M. Keck Foundation Grant to the Center for Molecular Biology of RNA, by National Cancer Institute grant CA77813 to M.A. and D. Haussler, and by NIH grant GM40478 to M.A. T.C. was supported by a University of California Biotechnology training grant, and an NIH Training Grant. C.S. is a Predoctoral Fellow of the Howard Hughes Medical Institute.

Specificity and Stability in Topology of Protein Networks

Sergei Maslov[1] and Kim Sneppen[2]

Molecular networks guide the biochemistry of a living cell on multiple levels: Its metabolic and signaling pathways are shaped by the network of interacting proteins, whose production, in turn, is controlled by the genetic regulatory network. To address topological properties of these two networks, we quantified correlations between connectivities of interacting nodes and compared them to a null model of a network, in which all links were randomly rewired. We found that for both interaction and regulatory networks, links between highly connected proteins are systematically suppressed, whereas those between a highly connected and low-connected pairs of proteins are favored. This effect decreases the likelihood of cross talk between different functional modules of the cell and increases the overall robustness of a network by localizing effects of deleterious perturbations.

With the growth of experimental information about basic biochemical mechanisms of life, molecular networks operating in living cells are becoming better defined. Direct physical interactions between pairs of proteins form one such network, which serves as a backbone for functional and structural relationships among its nodes and defines pathways for the propagation of various signals such as

phosphorylation and allosteric regulation of proteins. The information about specific binding of proteins to each other has recently grown by an unprecedented amount as a result of high-throughput two-hybrid experiments (1, 2). The production and degradation of proteins participating in the interaction network is controlled by the genetic regulatory network of the cell, formed by all pairs of proteins in which the first protein directly regulates the abundance of the second. The majority of known cases of such regulation happens at the level of transcription, in which a transcription factor positively or negatively regulates the RNA transcription of the controlled protein. The large-scale structure of

[1]Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA. [2]Department of Physics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

*To whom correspondence should be addressed. E-mail: maslov@bnl.gov