plants retain both of the adjacent cpDNA *Nco*I sites that are outside the experimental construct. Most digested DNA samples from the kanamycin-resistant plants show two hybridization fragments that mirror those in the transplastome (Fig. 3a,b in ref. 1). The generally large size of the integrants is confirmed by sequence data from two of the kanamycin-resistant plants (Fig. 3e,f in ref. 1). For example, there are 1,775 bp of vector chloroplast DNA and 5,917 bp of nonvector chloroplast DNA (see Supplementary Information to ref. 1) between the junction with nuclear DNA and the *aad*A gene in kr1. Similarly, there are 1,165 bp of vector chloroplast DNA and 934 bp of nonvector chloroplast DNA adjacent to the junction site downstream of *neo* in the nuclear integrant of kr17. Consequently, there can be no doubt that most of these integrants contain more DNA of chloroplast origin than was present in our experimental cassette (see Fig. 1). Integrants that are shorter than the chloroplast transformation vector may also be present.

Regarding the concern over multiple integrants, we do not yet understand the complexity of the transfer process, but we do know that single Mendelian loci are involved in all but four of the kanamycin-resistant plants from the screen. Multiple integrations do not require multiple transposition events. The lysis of a single plastid would release tens to hundreds of plastid genomes into the cytoplasm, some of which could integrate into a common genomic location. This process could be analogous to the high-copy number of transgenes delivered into the cell via biolistic transformation, so it is not surprising to find a proportion of multiple integrants among the kanamycin-resistant plants.

We did not conclude that chloroplast-specific genes, such as *aadA* in our experiment, will not function when transposed to the nucleus. What we did show, in all cases where we selected for nuclear kanamycin resistance, was that the relocated *neo* gene was accompanied by the adjacent *aadA* gene (and other flanking native chloroplast DNA). We noted that the latter gene was not expressed to confer spectinomycin resistance.

A News & Views commentary[4] that accompanied our article in *Nature* suggested that we undertake a much larger screen to search for spectinomycin resistance to determine whether a chloroplast specific gene rarely could be expressed after integration into an appropriate nuclear environment. This is an evolutionary experiment in the true sense, but the likely scale of such a screen is daunting. However, such an approach may greatly increase our understanding of the evolution of nuclear-encoded plastid genes.

Finally, we stated in our paper that there was likely to be an equilibrium between the ingress of chloroplast DNA sequences and their elimination. The fact that Daniell and Parkinson know of no such mechanism merely demonstrates that there is still much to learn about the processes of genome evolution.

Chun Y. Huang,
Department of Molecular Biosciences,
The University of Adelaide,
South Australia, 5005, Australia,
Michael A. Ayliffe,
CSIRO Plant Industry,
GPO Box 1600,
ACT 2601, Australia,
and Jeremy N. Timmis
The University of Adelaide,
South Australia, 5005, Australia

1. Huang, C.Y., Ayliffe, M.A. & Timmis, J.N. *Nature* **422**, 72–76 (2003).
2. Ayliffe, M.A. & Timmis, J.N. *Theor. Appl. Genet.* **85**, 229–238 (1992).
3. Ayliffe, M.A. & Timmis, J.N. *Mol. Gen. Genet.* **236**, 105–112 (1992).
4. Maliga, P. *Nature* **422**, 31–32 (2003).
5. Martin, W. *Nat. Genet.* **33**, 442 (2003).
6. Lilly, J.W. *et al. Plant Cell* **13**, 245-254 (2001).

Global architecture of genetic interactions on the protein network

To the editor:

Recent and ongoing experiments are uncovering large networks of both protein-protein and genetic interactions in cells[1–3]. Although integration of these networks into a single model has immense promise as a tool for understanding basic cellular mechanisms and disease, it is critical to first understand their high-level correlations and interdependencies. Here, we report that highly connected 'hubs' in the protein-protein network are surprisingly likely to be involved in genetic interactions, particularly with other hubs.

We constructed a physical network for the yeast *Saccharomyces cerevisiae* consisting of 15,114 pairwise interactions among 4,716 proteins as recorded in the DIP[4] database, as of the March 2, 2003 release. Protein-protein interactions in this database represent a pooled collection of several yeast two-hybrid and co-immunoprecipitation experiments. Next, the physical network was combined with a genetic network consisting of 1,312 unique genetic interactions drawn from MIPS[5] (April 3, 2003) and a systematic screen by Tong *et al.*[3] Approximately 70% of genetic interactions were synthetic lethal interactions (mutations in two nonessential genes that are lethal when combined) while the remaining

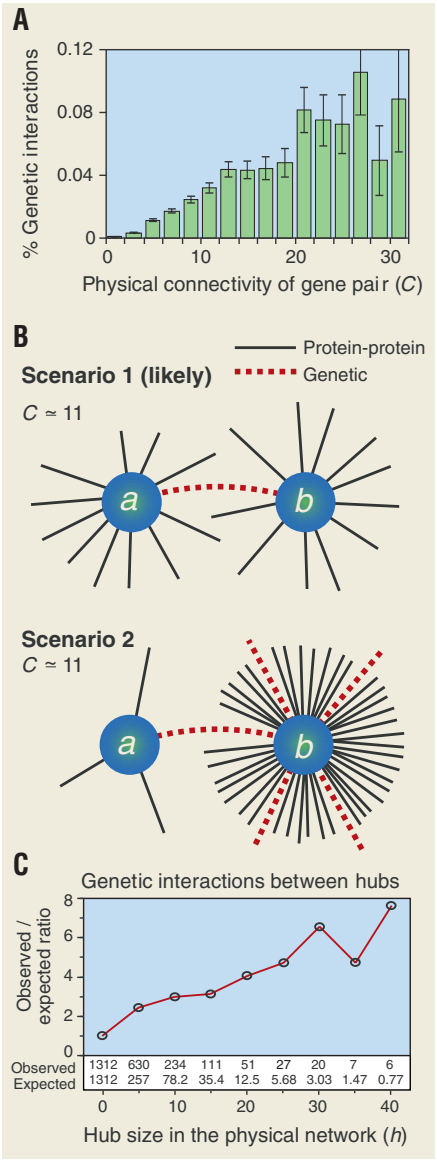

Figure 1. Overlap in connectivity between physical and genetic interaction networks. (A) Frequency of genetic interaction (percent ± standard error) versus degree of physical connectivity *C*, binned over all gene pairs in the network. The overall correlation is preserved when essential genes are excluded from the network; whether suppressor or synthetic lethal interactions are examined separately or pooled; or if *C* is computed using an arithmetic instead of geometric mean (data not shown). (B) Possible interpretations for panel A. In scenario 1, physical interaction hubs (*a* and *b*) are biased to genetically interact with each other; in scenario 2, hubs *b* genetically interact with many partners *a* of both low and high physical connectivity. (C) Ratio of observed/expected numbers of genetic interactions between protein hubs (proteins with ≥ *h* physical interactions) computed for increasing values of *h*. If hubs are nonbiased (scenario 2), then the expected number of genetic interactions between hubs is $g \cdot (n \text{ choose } 2) / ((n \text{ choose } 2) + n \cdot m)$, where *n*, *m*, and *g* are the numbers of hubs, non-hubs and genetic interactions involving a hub, respectively, for a particular hub size *h*. Further information is available at http://web.wi.mit.edu/ideker/pub/nbt/.

30% were classified as suppressors (one mutation is lethal but combination with a second restores viability). Although these data sets are almost certainly incomplete and may also contain a substantial number of false positives[6], we take the currently available physical and genetic interactions as representative of the true underlying networks.

To investigate the overlap between the genetic and physical networks, we recorded the number of connections k that each protein had in the physical network, and we computed a combined physical connectivity $C = \sqrt{k_a k_b}$ for every possible protein pair (a,b). Protein pairs were then sorted according to C and examined for whether they were directly linked by a genetic interaction. As shown in Figure 1A, the frequency of genetic interaction increases markedly with physical connectivity. For example, genetic interactions were roughly 16-fold as likely to occur between proteins with $C = 12$ versus $C = 4$ interactions.

This result has two interpretations: first, highly connected protein 'hubs' in the physical network tend to genetically interact with each other; or second, hubs genetically interact with many partners, regardless of their partners' physical connectivities (Fig. 1B). To investigate this further, we defined proteins having more than h physical interactions as 'hubs' and looked at the frequencies with which hubs genetically interact with other hubs versus non-hubs. Indeed, genetic interactions between hubs occur more frequently than expected, and this bias becomes more pronounced at larger hub sizes h (Fig. 1C). Therefore, not only are single highly connected proteins essential to the cell[7], but so are combinations of such proteins. It is also interesting that this bias toward genetic interactions between hubs contrasts with a previously identified bias against physical interactions between hubs[8], indicating that while there is some correlation between the physical and genetic networks, they do not coincide.

Genetic interactions are often used to screen for genes acting in a common cellular function. However, it seems plausible that the tendency toward genetic interactions between physical network hubs is due not to common functionality, but rather to the additive effects of disrupting central components of two functions that may be very different. Geneticists may therefore wish to exclude these genetic interactions from further analysis. Conversely, genetic interactions among lower-connectivity proteins are less frequent and may indicate proteins that cooperate directly through physical interactions in the network. It is these genetic interactions that, when prioritized for further study, will perhaps provide clearer insight into biological pathways.

Owen Ozier, Nada Amin, and Trey Ideker,
Whitehead Institute for Biomedical Research,
Cambridge, MA 02142
(trey@wi.mit.edu).

1. Legrain, P. & Selig, L. *FEBS Lett* **480**, 32–36 (2000).
2. Hartman, J.L., Garvik, B. & Hartwell, L. *Science* **291**, 1001–1004 (2001).
3. Tong, A.H. *et al. Science* **294**, 2364–2368 (2001).
4. Xenarios, I. *et al. Nucleic Acids Res.* **30**, 303–305 (2002).
5. Mewes, H.W. *et al. Nucleic Acids Res.* **30**, 31–34 (2002).
6. von Mering, C. *et al. Nature* **417**, 399–403 (2002).
7. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. *Nature* **411**, 41–42. (2001).
8. Maslov, S. & Sneppen, K. *Science* **296**, 910–913 (2002).

Developing countries and systems biology

To the editor:

Systems biology is taking root as an integrating tool to comprehend how the genome and proteome respond to environmental changes through the signal-transduction mechanisms of the cell[1]. The development of mathematical and computational tools in such approaches represents a significant opportunity for researchers in developing countries.

The combined laboratories of Gifford and Young have recently shown the feasibility of a computational approach to correctly assign all the regulators of a complex network, such as the cell cycle[2]. These authors were able to devise an algorithm to automatically reconstruct the corresponding transcriptional regulatory architecture without previous knowledge of the molecules involved. This work was possible because the transcriptional mechanism that governs the cell cycle follows the same type of 'network motifs' or patterns of interconnections that pertain to a wide range of complex networks encompassing ecological systems, neuronal synapses, electric circuits, and the Internet, among others[3].

These so called scale-free networks, display a topology conforming to a power-law distribution where very few nodes become hubs dominating the entire network by attaching themselves to others that are already well interconnected[4]. The fact that besides the transcriptional network, the genome[5], proteome[6], metabolome[7], and the physiome[8] all follow a power-law makes it likely that an effort to name all the cellular signal-transduction machinery will also follow suit and be amenable to study with these new mathematical and computational tools.

In this regard, it is noteworthy that the Alliance for Cellular Signalling (AfCS; http://www.cellularsignaling.org/) has been launched to "answer questions about signaling networks…to facilitate quantitative modeling" on two types of cell paradigms: B lymphocytes and cardiac myocytes[9]. I believe that the ability of researchers in developing countries to access this information should be exploited to allow their recruitment in efforts to determine how the entire signal transduction machinery is built and connected in those particular types of important cells.

While scientists and physicians in developing countries face logistical problems in obtaining materials in traditional research, they can take advantage of the emerging plethora of biomedical knowledge available over the Internet. Through a synthesis of genomics, proteomics, signal transduction, network theory, computing science, non-linear dynamics, and clinical practice, such groups could design appropriate algorithms to visualize dynamic maps of the molecular networks that underlie clinical manifestations of disease (in a similar manner to the integration of gene circuits reviewed by Hasty *et al.*[10]). These, in turn, could be used as important aids in early diagnosis, prognosis, and risk management. In this respect, the Lymphoma/Leukemia Molecular Profiling Project has already shown the validity of using a set of genes, or 'gene expression signatures' as molecular predictors of certain types of cancer, such as diffuse large-B-cell lymphoma[11].

With the above developments in mind, and the aid of a national database compiled by my colleague Manuel Bemporad[12], Dr. Bemporad and I have identified the most productive physicians, molecular biologists, biochemists, theoretical physicists, mathematicians, and computer scientists, and asked them to form three types of interconnected networks: *CliniRed*, *BioRed*, and *CompuRed*. These networks have two aims: first, to analyze the wealth of data coming from international research programs, such as those mentioned above, to select those molecular markers that are relevant to clinical practice; and second, to organize a long-term project to develop the tools that will make possible this new type of medicine in the country.

We hope that in this way, we will offer to some of our very best talent an opportunity to do advanced applied research on a very pertinent goal—the fight against the major causes of morbidity and mortality in Venezuela; and all this with an unlimited supply of knowledge and techniques coming from abroad at no cost.