D502–D507 *Nucleic Acids Research, 2010, Vol. 38, Database issue* *Published online 30 October 2009*
doi:10.1093/nar/gkp820

DRYGIN: a database of quantitative genetic interaction networks in yeast

Judice L. Y. Koh[1,*], Huiming Ding[1], Michael Costanzo[1], Anastasia Baryshnikova[1], Kiana Toufighi[1], Gary D. Bader[1], Chad L. Myers[2], Brenda J. Andrews[1] and Charles Boone[1,*]

[1]Banting and Best Department of Medical Research and Department of Molecular Genetics, The Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, ON, Canada M5S 3E1 and [2]Department of Computer Science and Engineering, University of Minnesota, 4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN 55455, USA

Received August 19, 2009; Accepted September 16, 2009

ABSTRACT

Genetic interactions are highly informative for deciphering the underlying functional principles that govern how genes control cell processes. Recent developments in Synthetic Genetic Array (SGA) analysis enable the mapping of quantitative genetic interactions on a genome-wide scale. To facilitate access to this resource, which will ultimately represent a complete genetic interaction network for a eukaryotic cell, we developed DRYGIN (Data Repository of Yeast Genetic Interactions)—a web database system that aims at providing a central platform for yeast genetic network analysis and visualization. In addition to providing an interface for searching the SGA genetic interactions, DRYGIN also integrates other data sources, in order to associate the genetic interactions with pathway information, protein complexes, other binary genetic and physical interactions, and Gene Ontology functional annotation. DRYGIN version 1.0 currently holds more than 5.4 million measurements of genetic interacting pairs involving ~4500 genes, and is available at http://drygin.ccbr.utoronto.ca

INTRODUCTION

Genetic interactions tend to occur between genes that operate in functionally related pathways, and therefore are highly informative for understanding how cellular processes are carried out (1). Recent developments in large-scale Synthetic Genetic Array (SGA) genetic analysis enable a quantitative, genome-scale mapping of the genetic interaction network of budding yeast *Saccharomyces cerevisiae* (M. Costanzo *et al.*, in press). SGA automates yeast genetic analysis, facilitating the isolation of double mutants, which enables the assessment of their fitness (2,3). Image processing methodology, combined with mathematical modeling and correction of experimental variances, allow systematic measurements of the growth phenotype, and enable the identification of double mutants that display negative genetic interactions (synthetic lethality or synthetic sick interactions) where the double-mutant fitness is lower than expected for the combined effect of the single mutant growth phenotypes, or positive genetic interactions (within-pathway or suppression interactions) where the double-mutant fitness is greater than expected (A. Baryshnikova *et al.*, manuscript in preparation). To date, the approach has harvested more than 5.4 million gene pair measurements, with ongoing efforts targeting to complete genetic interaction mapping of the entire yeast genome comprising more than 6000 genes.

Decades of studies have contributed to unraveling the genetics of *S. cerevisiae*, and it remains the best-understood eukaryote. We found significant conservation of up to 29% in both literature-curated and experimentally derived negative genetic interactions between *S. cerevisiae* and its distantly related fission yeast *Schizosaccharomyces pombe* (4), indicating that a substantial portion of the genetic interactions are conserved for million of years, and justifying that a genome-wide map of the genetic interaction network in *S. cerevisiae* will be useful for extending our understanding of the genetic wiring in more complex human system.

A number of databases, such as BIOGRID (5), SGD (6) and MIPS (7), collate *S. cerevisiae* genetic interactions from a spectrum of high-throughput as well as focused interaction studies, and provide access to the binary genetic interactions in unified, consistent formats. While these resources are valuable assets for the yeast research

*To whom correspondence should be addressed. Tel: +1 416 978 3514; Fax: +1 416 978 8287; Email: judice.koh@utoronto.ca
Correspondence may also be addressed to Charles Boone. Email: charlie.boone@utoronto.ca

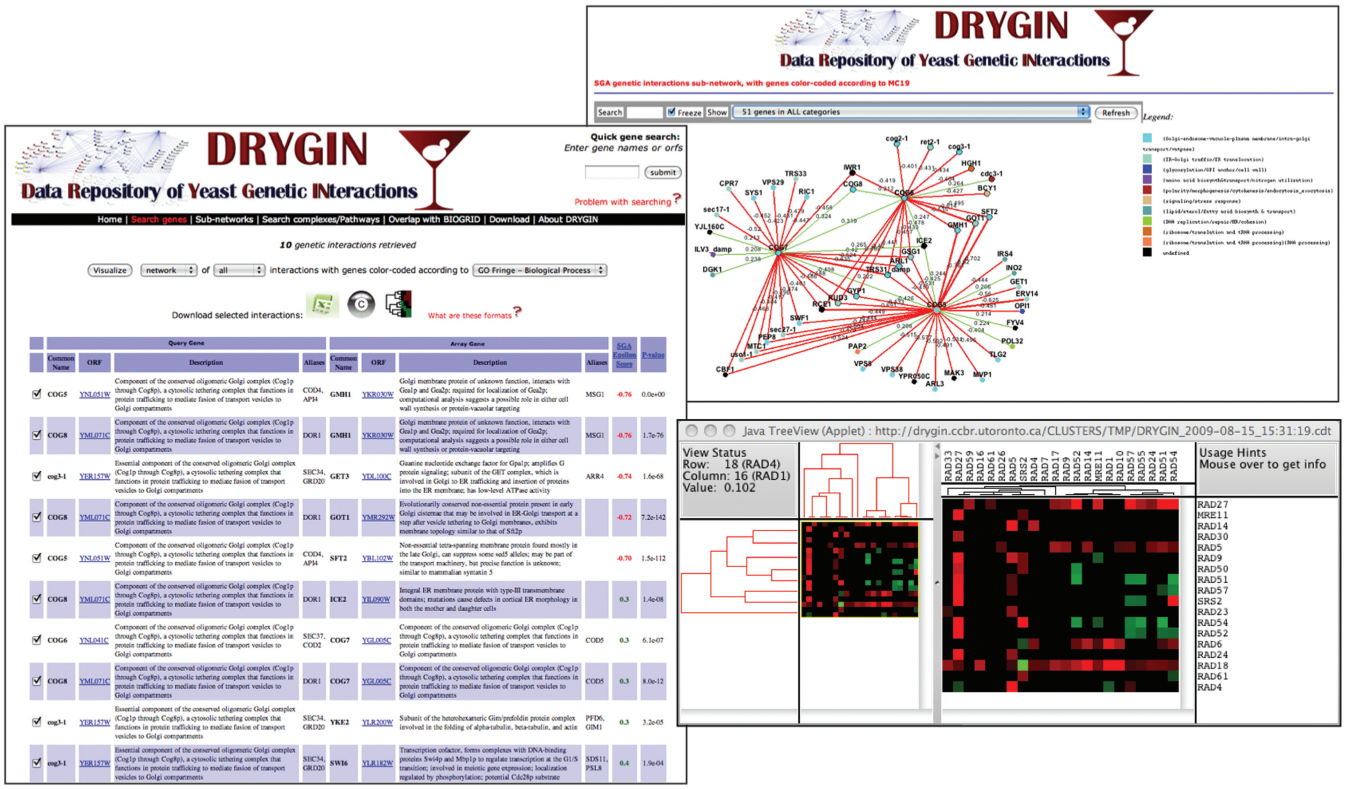

Figure 1. Sample search and result pages generated by DRYGIN. Querying 'COG' in 'Search gene', with the wildcard option, generates the tabular result and the graphical network display. The hierarchical cluster is generated from querying 'RAD' in 'Sub-network' with wildcard option, with the positive genetic interactions colored in green, and the negative interactions in red.

community, they were mainly designed as repositories for binary interactions (where the interactions are either present or absent). To facilitate access to quantitative measurements of genetic interaction and enable customized analyses leveraging on the power of quantitative genetic interactions and its association with other types of molecular interactions, a central platform is preferable to enable querying, ranking and visualization of the quantitative genetic interaction network.

Here, we describe DRYGIN (Data Repository of Yeast Genetic Interactions)—a web-based query system for the first comprehensive database of quantitative genetic interactions in *S. cerevisiae*. Currently, DRYGIN contains genetic interaction measurements of more than 5.4 million gene pairs, involving 1712 query genes, corresponding to 1673 open-reading frames (ORFs) and 3885 array genes, corresponding to 3885 ORFs. DRYGIN imports general and functional information pertaining to genes from a variety of external data sources to map the genetic interactions onto physical interactions, protein complexes, biochemical and metabolic pathways, and to benchmark quantitative genetic interactions against binary genetic interactions.

At the backend of DRYGIN is a PostgresSQL relational database system, that enables efficient querying of large volume of interaction data. The front-end of DRYGIN is an Apache 2.0 web server hosting web interface developed using a combination of HTML, CGI Perl, Perl DataBase Interface (DBI), Cascading Style Sheets (CSS) and Javascript for easy navigation, and incorporating Java applets for graphical visualization of genetic networks.

Various search options in DRYGIN allow users to query the database with different objective modes, based on identifiers such as gene name, ORF, complex name, pathway descriptors, and free text. User-defined cutoffs are applied on the genetic interaction score (ε) and/or *P*-value depicting the score confidence, allowing interactions to be retrieved and sorted in order of the interaction score, *P*-value or gene name. Selected or complete query results can be visualized and downloaded in tabular, network, or two-dimensional hierarchical cluster heat map views (Figure 1). In the network view, nodes (genes) are color-coded according to biological process annotations from Gene Ontology (GO) or a customized set of terms that annotate 4373 genes into one or more of 17 generalized biological processes (M. Costanzo *et al.*, in press).

MATERIALS AND METHODS

Quantitative genetic interactions

High-throughput SGA analysis, coupled with a new scoring method for normalizing experimental variations in genetic interaction arrays, have enabled us to extract quantitative measurements of genetic interactions. The

genetic interactions are modeled after the Fisher definition of epistasis as quantitative deviations from the expected multiplicative combination of independently functioning genes (8), and which define genetic interaction strength as follows:

$$\varepsilon = f_{ab} - f_a f_b$$

f_a and f_b denote the single mutant fitness of gene a and gene b respectively, and f_{ab} is the fitness of the double knockout mutant of gene a and b. The distribution of ε values follows a distribution such that interactions with higher absolute magnitude and therefore greater interaction strength are rare (Figure 2). Cutoffs for the genetic interactions can be imposed on the ε score or/and on the P-value depicting score confidence. Details of the scoring method are given in (A. Baryshnikova *et al.*, manuscript in preparation) , but briefly, replicate double mutant colonies are used to assess confidence in the interaction measurements. Statistical analysis have shown that both reproducible and functionally informative interactions are determined at *P*-value <0.05 and $|\varepsilon| > 0.08$, and therefore these are used as the default cutoffs in DRYGIN queries.

Among the query genes screened are 214 temperature-sensitive (TS) query strains, corresponding to 183 ORFs. Also, 120 hypomorphic alleles are constructed by applying the DAmP (decreased abundance by mRNA perturbation) technique to replace the 3′ UTR and lower transcription of the targeted gene (9). The TS and DAmP query strains enriched the collection with essential ORFs, complementing the value of the genome-wide genetic interaction map with functional information from the essential gene set.

Genetic, physical interactions and pathways

An effective methodology for extracting significant sub-networks of biological interest is to assemble a variety of molecular interaction networks into contiguous pathway models. Such integrative network analysis approach aids in the identification of potential false positives and false negatives, through combining overlapping networks derived from different studies and experimental techniques while also increasing the coverage of the genome by combining orthogonal interactions from complementary networks (10).

Even though the exact modes of mechanism are not completely understood, previous studies have shown that positive genetic interactions often occur between genes that function in the same pathway, when a single-mutation is adequate to deactivate the pathway, and the resulting double-mutant phenotype resembles that of a single-mutant. In addition, positive interactions also occur between genes in different complexes and pathways, which presumably indicates some kind of functional relationship between them. On the other hand, negative genetic interactions have been thought to involve non-essential genes in compensatory pathways that impinge on a shared essential function, and are orthogonal to physical interactions (3). Recent studies, however, reveal negative genetic interactions sometimes overlap with physical interactions, especially when they occur within protein complexes that contain essential components (11).

To enable users to correlate the genetic interactions involving genes of their interests to physical interactions and pathways, DRYGIN integrates 75 609 physical interactions from BIOGRID (version 2.0.53), and

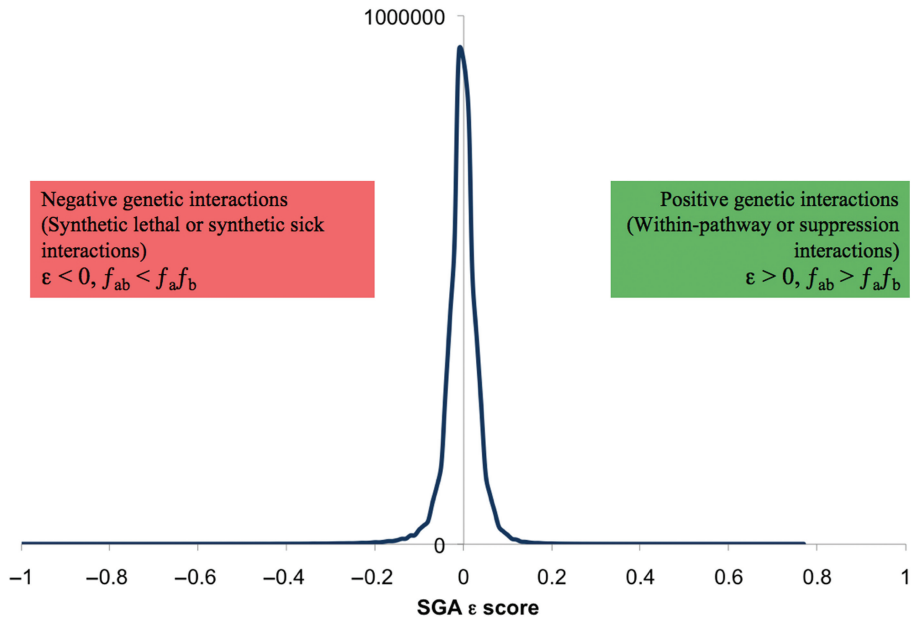

Figure 2. Distribution of genetic interaction scores in DRYGIN. f_a and f_b denote the single mutant fitness of gene a and b, respectively, and f_{ab} is the fitness of the double knockout mutant of gene a and b.

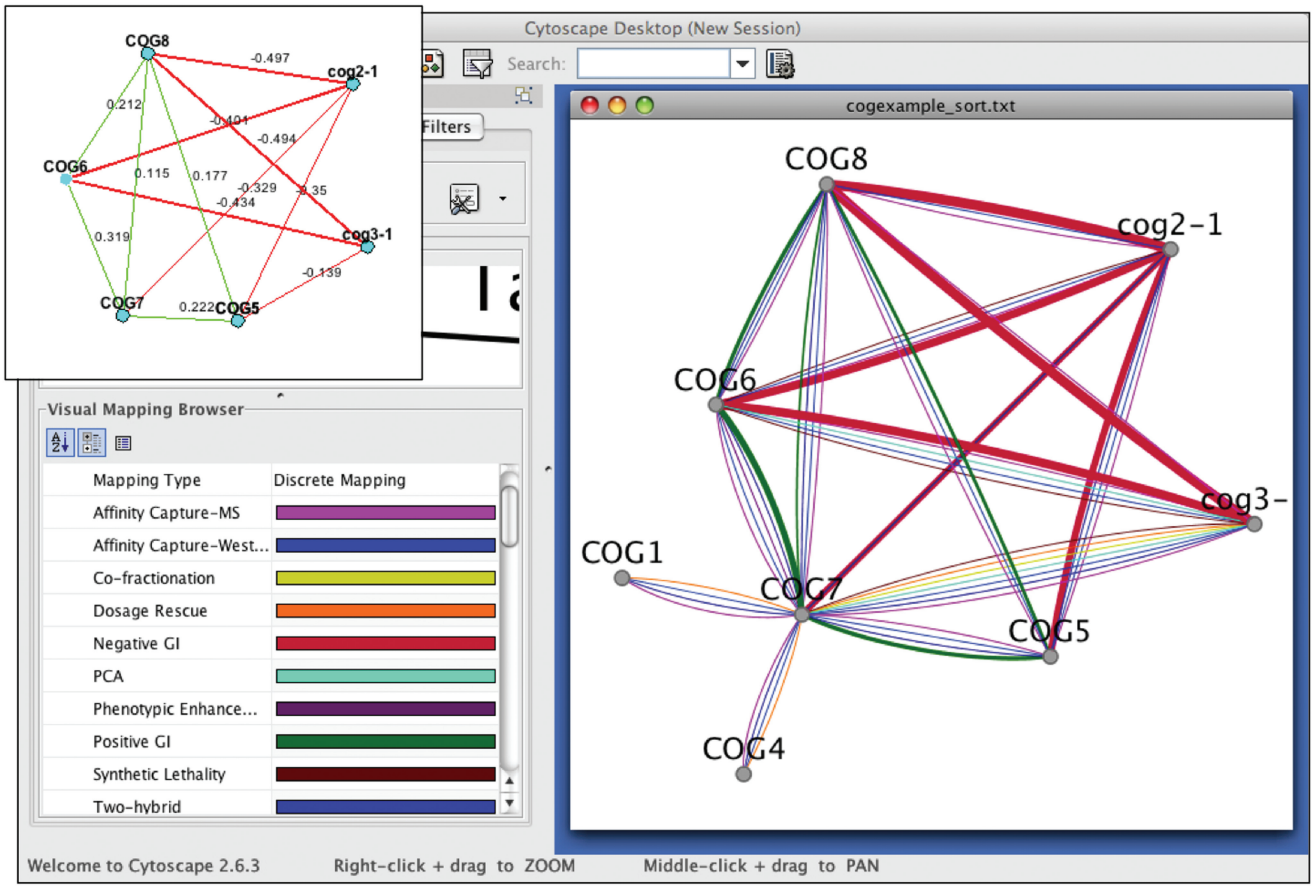

Figure 3. Querying 'Golgi' in 'Search complexes/Pathways' retrieves 14 genetic interactions between 6 COG genes (COG1 and COG4 were not screened in SGA) of the Golgi transport complex, and display these interactions in a Java graph applet (top left box). To search for other types of physical or genetic interactions within the Golgi transport complex, query the components in 'Overlap with BIOGRID'. A network view of the merged network of multiple types of interactions is generated as a Cytoscape session file and deployed through Java Web Start.

provides a visualization of the merged genetic/physical network through Java web start to Cytoscape (12). The physical interactions imported were compiled from various evidence sources, including large-scale two-hybrid (Y2H), tandem affinity purification followed by mass spectrometry (TAP-MS), and protein fragment complementation assay (PCA). In addition to the physical interactions, 50 611 binary genetic interactions are also imported from BIOGRID for benchmarking the quantitative genetic interactions. Figure 3 shows a Cytoscape session generated from DRYGIN. The genetic interactions involving 8 COG genes are overlaid with 49 BIOGRID interactions, corresponding to physical and genetic interactions from 10 different evidence sources.

Studies have shown that protein complexes are significantly enriched in pure positive or negative genetic interactions (11,13), meaning positive and negative genetic interactions tend to form coherent groups of positive or negative interactions, linking genes within a particular pathway or between pathways or protein complexes. This phenomenon coined 'monochromatic' can be used to draw insights into functional modularity in the yeast cell (13). To facilitate such studies of functional modularity, definitions of protein complexes from SGD (6) and the yeast protein complex catalogue (14) are assembled into DRYGIN, constituting 1758 genes and 412 complexes. It has also been shown that some 'dichromatic' complexes are explained by the presence of sub-structure (A.Baryshnikova *et al.*, manuscript in preparation), in which case the integration of protein complex definitions in DRYGIN also aid in the extracting views of functional subunits in complexes. One example is the Golgi transport complex exhibits two subunits that can be separated by positive and negative interactions (Figure 3).

In addition, DRYGIN also contain pathway definitions are extracted from the KEGG (15), and Reactome (16), constituting 329 metabolic and biochemical pathways, and 1461 genes. All genes/proteins assigned to the same pathway are considered functionally related, and not differentiated based on the presence or absence of physical interactions.

The integrative approach of overlaying genetic interactions with other types of biological interactions is intended to help decipher the underlying dynamics of how genes interact with one another genetically as well as other

interaction modes to deliver the yeast cell functions. External data sources as well as tools have been integrated into DRYGIN to support these analyses, with the goal of equipping its users with the ability to interpret and draw biological insights from the integrative sub-network models. All external data sources integrated to DRYGIN are updated periodically to maintain the currency of the information.

RESULTS

System interfaces and visualization

The primary accessibility of DRYGIN is via the Web-interface, with the objective of providing easy and efficient accessibility to a genome-wide database of quantitative genetic interactions to users for interpretation, and to assist biologists in experimental planning. A number of query and visualization tools are incorporated into DRYGIN version 1.0.

Search genes. Search options for genes include by gene name, common name, aliases, or description. Enabling wild-card query retrieves any match containing the specified keyword/s. Gene searches are checked against recent release of gene descriptors in SGD. Users may choose to retrieve all interactions involving one or more genes (Search gene) or only interactions within a group of genes (Sub-networks).

Network visualization. Retrieved interactions can be visualized in graphical networks using a network visualization tool that is enhanced from the Java graph viewer (17). We introduced various feature improvements to the graph viewer to enable more meaningful representation of the quantitative genetic interactions (Figure 1). The genes/proteins are represented as nodes, and colored according to the biological process assignments of the genes. The genetic/physical interactions are shown as edges, with the positive genetic interactions colored in green and negative interactions in red. The width of edges is proportional to the absolute magnitude of the genetic interaction score (ε). For more complicated network views, DRYGIN provides an interface to the Cytoscape network visualization and analysis software using the Java Web Start.

Search complexes and pathways. Search options for complexes and pathways include by name, gene or description. List of complexes or pathways matching the query, and the genetic interactions involving the proteins (genes) in the complex are retrieved and can be visualized in graphical network format.

Download. DRYGIN supports downloading of user-specified subsets or the full set of interactions in tab-delimited files as well as in various established molecular interactions format including Cytoscape SIF (12), and 'Clustered Data File' (CDT) format for visualizing the hierarchical clustering matrix and dendrograms in Treeview or Java Treeview (18).

DISCUSSION

Future developments

DRYGIN is intended to be an active resource for quantitative genetic interactions in *S. cerevisiae*. Mechanisms are in place to support future regular updates of internal as well as external data sources, and maintenance. Future additions to the system aim at further automating the updating processes to enable more frequent updates of new genetic interactions and other data sources.

At this stage, the semantics of how genetic interactions overlay with other types of molecular interactions is unknown. Therefore, in DRYGIN version 1.0, we focus on the integration of physical interactions and pathways. Future developments consider the integration of gene expression profiles, as well as other molecular networks.

We also consider expanding the export formats of DRYGIN to conform to other established data formats of molecular interactions in the biological network community. One example is the PSI-MI XML format that is compatible to the International Molecular Exchange Consortium (IMEx) (19).

Conclusion

Biotechnological developments have enabled the quantification of genetic interactions in yeast, so that genetic interactions are no longer assessed simply by presence or absence, but by the measurable strengths, which can be ranked and prioritized for experimental testing. The development of DRYGIN aimed at providing a common platform for accessing and visualizing quantitative genetic interactions. With the integration of external data providing descriptive and exploratory views of the yeast interactome, we envision DRYGIN database will be central *S. cerevisiae* resource assisting experimentalists in the generation of testable hypotheses.

ACKNOWLEDGEMENTS

The authors thank Renee L. Brost, Bilal Sheikh and other anonymous users of DRYGIN beta version for testing and suggestions, and Chris S.H. Tan for naming the database.

FUNDING

Genome Canada through the Ontario Genomics Institute [2004-OGI-3-01 to J.L.Y.K., H.D., M.C., A.B., K.T., G.D.B., B.J.A., C.B.]. Funding for open access charge: Genome Canada through the Ontario Genomics Institute as per research agreement 2004-OGI-3-01.

Conflict of interest statement. None declared.

REFERENCES

1. Boone,C., Bussey,H. and Andrews,B.J. (2007) Exploring genetic interactions and networks with yeast. *Nat. Rev.*, **8**, 1641–1649.
2. Tong,A.H.Y., Evangelista,M., Parson,A.B., Xu,H., Bader,G.D., Page,N., Robinson,M., Raghibizadeh,S., Hogue,C.W.V., Bussey,H.

et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2365–2368.
3. Tong,A.H., Lesage,G., Bader,G.D., Ding,H., Xu,H., Xin,X., Young,J., Berriz,G.F., Brost,R.L., Chang,M. *et al*. (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**, 808–813.
4. Dixon,S.J., Fedyshyn,Y., Koh,J.L.Y., Prasad,T.S., Chahwan,C., Chua,G., Toufighi,K., Baryshnikova,A., Hayles,J., Hoe,K. *et al*. (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly-related eukaryotes. *Proc. Natl Acad. Sci. USA*, **105**, 16653–16658.
5. Stark,C., Breitkreutz,B.J., Reguly,T., Boucher,L., Breitkreutz,A. and Tyers,M. (2006) BIOGRID: a general repository for interaction datasets. *Nucleic Acids Res*., **34**, D535–D539.
6. Hirschman,J.E., Balakrishnan,R., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R., Fisk,D.G., Hong,E.L., Livstone,M.S., Nash,R. *et al*. (2006) Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. *Nucleic Acids Res*., **34**, D442–D445.
7. Mewes,H.W., Dietmann,S., Frishman,D., Gregory,R., Mannhaupt,G., Mayer,K., Muensterkötter,M., Ruepp,A., Spannagl,M., Stuempflen,V. *et al*. (2008) MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res*., **36**, D196–D201.
8. Fisher,R.A. (1918) The correlations between relatives on the supposition of Mendelian inheritence. *Trans. R. Soc. Edinb*., **52**, 399–433.
9. Schuldiner,M., Collins,S., Thompson,N., Denic,V., Bhamidipati,A., Punna,T., Ihmels,J., Andrews,B., Boone,C., Greenblatt,J. *et al*. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, **123**, 507–519.
10. Beyer,A., Bandyopadhyay,S. and Ideker,T. (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet*., **8**, 699–710.
11. Bandyopadhyay,S., Kelley,R., Krogan,N.J. and Ideker,T. (2008) Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput. Biol*., **4**, e1000065.
12. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*., **13**, 2498–2504.
13. Segrè,D., DeLuna,A., Church,G.M. and Kishony,R. (2005) Modular epistasis in yeast metabolism. *Nat. Genet*., **3**, 77–83.
14. Pu,S., Wong,J., Turner,B., Cho,E. and Wodak,S.J. (2009) Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res*., **37**, 825–831.
15. Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. *et al*. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res*., **36**, D480–D484.
16. Matthews,L., Gopinath,G., Gillespie,M., Caudy,M., Croft,D., de Bono,B., Garapati,P., Hemish,J., Hermjakob,H., Jassal,B. *et al*. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*., **37**, D619–D622.
17. Chang,A.N., MacDermott,J. and Samudrala,R. (2005) An enhanced Java graph applet interface for visualizing interactomes. *Bioinformatics*, **21**, 1741–1742.
18. Saldanha,A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.
19. Hermjakob,H., Montecchi-Palazzi,L., Bader,G., Wojcik,J., Salwinski,L., Ceol,A., Moore,S., Orchard,S., Sarkans,U., von Mering,C. *et al*. (2004) The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. *Nat. Biotechnol*, **22**, 177–183.