BioGRID: A Resource for Studying Biological Interactions in Yeast: Table 1.

Article · January 2016

DOI: 10.1101/pdb.top080754

Topic Introduction

BioGRID: A Resource for Studying Biological Interactions in Yeast

Rose Oughtred,[1] Andrew Chatr-aryamontri,[2] Bobby-Joe Breitkreutz,[3] Christie S. Chang,[1] Jennifer M. Rust,[1] Chandra L. Theesfeld,[1] Sven Heinicke,[1] Ashton Breitkreutz,[3] Daici Chen,[2] Jodi Hirschman,[1] Nadine Kolas,[3] Michael S. Livstone,[1] Julie Nixon,[4] Lara O'Donnell,[3] Lindsay Ramage,[4] Andrew Winter,[4] Teresa Reguly,[3] Adnane Sellam,[2] Chris Stark,[3] Lorrie Boucher,[3] Kara Dolinski,[1,5] and Mike Tyers[2,3,4,5]

[1]*Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544;* [2]*Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Québec H3C 3J7, Canada;* [3]*Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada;* [4]*School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom*

The Biological General Repository for Interaction Datasets (BioGRID) is a freely available public database that provides the biological and biomedical research communities with curated protein and genetic interaction data. Structured experimental evidence codes, an intuitive search interface, and visualization tools enable the discovery of individual gene, protein, or biological network function. BioGRID houses interaction data for the major model organism species—including yeast, nematode, fly, zebrafish, mouse, and human—with particular emphasis on the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* as pioneer eukaryotic models for network biology. BioGRID has achieved comprehensive curation coverage of the entire literature for these two major yeast models, which is actively maintained through monthly curation updates. As of September 2015, BioGRID houses approximately 335,400 biological interactions for budding yeast and approximately 67,800 interactions for fission yeast. BioGRID also supports an integrated post-translational modification (PTM) viewer that incorporates more than 20,100 yeast phosphorylation sites curated through its sister database, the PhosphoGRID.

BACKGROUND

The Biological General Repository for Interaction Datasets (BioGRID; http://www.thebiogrid.org) is an open source database that curates and disseminates collections of protein and genetic interactions from major model organism species from yeast to human (Stark et al. 2006; Chatr-Aryamontri et al. 2013). The BioGRID was originally developed as a budding yeast–specific database to house and visualize protein interaction data from high-throughput proteomic studies (Ho et al. 2002; Breitkreutz et al. 2003a; Stark et al. 2006). Subsequently, comprehensive curation of protein and genetic interactions from the entire budding yeast literature was undertaken to compare emerging high-throughput interaction data to the extensive body of interaction data reported in thousands of focused studies (Reguly et al. 2006). Importantly, the evidence for each interaction in BioGRID is recorded as a structured evidence code derived from the primary experimental data. These evidence codes are concordant and interoperable with high-level stratification of the detailed Proteomics Standards Initiative-Molecular Interaction (PSI-MI) ontology (Hermjakob et al. 2004a; Kerrien et al. 2007). All curated data within BioGRID is fully archived as monthly releases and all records are date-stamped

[5]Correspondence: kara@genomics.princeton.edu; md.tyers@umontreal.ca

Cite this introduction as *Cold Spring Harb Protoc*; doi:10.1101/pdb.top080754

and mapped to individual curators to ensure data integrity. Curation efforts at BioGRID have since been expanded to capture biological interaction data from each of the major model organism species. These data sets serve as a readily accessible resource for interrogation of biological interactions, discovery of gene function, and computational analysis of interaction networks (Dolinski et al. 2013).

BioGRID CURATION STATISTICS

The September 2015 release of BioGRID (version 3.4.128) contains more than 812,000 interactions curated from both high-throughput data sets and low-throughput focused studies found in the literature. These interactions have been distilled from more than 45,000 publications covering 57 different organisms, including the budding yeast *Saccharomyces cerevisiae*, the fission yeast *Schizosaccharomyces pombe*, the yeast *Candida albicans SC5314*, the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, the mouse *Mus musculus*, the plant *Arabidopsis thaliana*, and *Homo sapiens* (Stark et al. 2011; Chatr-Aryamontri et al. 2013). BioGRID interaction data sets are shared with the respective model organism databases (Cherry et al. 2012; Inglis et al. 2012; Lamesch et al. 2012; Wood et al. 2012; Yook et al. 2012; Marygold et al. 2013), with other interaction databases (Luc and Tempst 2004; Razick et al. 2008; Chautard et al. 2009; Matthews et al. 2009; Cerami et al. 2011; Franceschini et al. 2013), and with metadatabases (Benson et al. 2004; Matthews et al. 2009). Complete coverage of the entire literature for *S. cerevisiae* and *S. pombe*, as well as for the model plant *A. thaliana*, is maintained through continuous monthly updates. As of the latest BioGRID release, approximately 335,400 (225,700 unique) interactions have been curated for *S. cerevisiae* genes/proteins from more than 13,000 publications, and approximately 67,800 (55,400 unique) interactions have been curated for *S. pombe* genes from nearly 2100 publications (Table 1). Of these interactions, 63% of budding yeast and 85% of fission yeast interactions derive from genetic experiments, and for both organisms, some 80% of interactions are derived from high-throughput data sets. Recently, more than 400 interactions have also been curated from nearly 40 papers for the pathogenic yeast model species, *C. albicans*. All yeast genetic interactions include associated phenotypes curated using the structured Ascomycete Phenotype Ontology (APO) developed by the *Saccharomyces* Genome Database (SGD; Engel et al. 2010). In addition, more than 20,100 phosphorylation sites mapped onto nearly 3200 budding yeast proteins are documented in a sister database called PhosphoGRID (http://www.phosphogrid.org; Sadowski et al. 2013) and are available through a new posttranslational modification (PTM) viewer integrated within BioGRID.

USING THE BioGRID DATABASE

The research community can access these extensive interaction data sets using the BioGRID web interface (http://www.thebiogrid.org), which provides users with a tabular interaction summary for each query gene or protein, as well as a link to the abstract for each curated publication and associated PubMed identifier. Details including interaction type, evidence code, and data source are provided in a condensed format on each summary page. Interaction data may also be viewed using an interactive network visualization tool embedded within BioGRID, downloaded in bulk for local analysis, or captured through stand-alone visualization applications, such as Osprey and Cytoscape (Breitkreutz et al. 2003b; Shannon et al. 2003; Cline et al. 2007).

For detailed instructions on how to use the BioGRID website to query genetic or protein interactions for any gene of interest, how to visualize the associated interactions using an embedded interactive network viewer, and how to download data files for either selected interactions or the entire BioGRID interaction data set, see Protocol: **Use of the BioGRID Database for Analysis of Yeast Protein and Genetic Interactions** (Oughtred et al. 2015).

 Cite this introduction as *Cold Spring Harb Protoc;* doi:10.1101/pdb.top080754

TABLE 1. Summary of current yeast interactions curated in BioGRID

	Total interactions	Curated publications	Protein interactions	Genetic interactions	Unique phenotypes
Saccharomyces cerevisiae	335,427 (225,753)	13,060	125,534 (81,115)	209,893 (151,379)	602
HTP	258,619 (196,923)	345	80,657 (66,045)	177,962 (133,341)	57
LTP	81,739 (44,028)	12,951	45,011 (21,300)	36,728 (26,165)	600
Schizosaccharomyces pombe	67,795 (55,428)	2084	10,062 (7,261)	57,733 (48,922)	318
HTP	55,686 (48,510)	49	3947 (3,759)	51,739 (44,782)	13
LTP	12,144 (7,552)	2074	6147 (3,768)	5997 (4399)	317
Candida albicans	414 (375)	43	146 (113)	268 (263)	11
HTP	254 (254)	2	0 (0)	254 (254)	1
LTP	160 (121)	41	146 (113)	14 (9)	10

BioGRID yeast curation statistics as of September 2015 (BioGRID version 3.4.128). To date, more than 403,600 total interactions have been curated from more than 15,100 publications. These cover 6504 *S. cerevisiae* proteins, 3964 *S. pombe* proteins, and 372 *C. albicans SC5314* proteins. The number of unique interactions is given in parentheses, and the number of interactions derived from high-throughput or low-throughput studies is given for each category. The number of unique phenotypes refers to the number of nonredundant phenotypes curated for genetic interactions using the Ascomycete Phenotype Ontology (APO).

HTP, high-throughput; LTP, low-throughput.

SCOPE OF THE BioGRID

The BioGRID will continue to expand the curation of protein and genetic interactions from the biomedical literature, as well as associated attributes such as PTMs, protein variants, phenotypes, and chemical or drug interactions. In addition to BioGRID, the following members of the International Molecular Exchange (IMEx) consortium (http://www.imexconsortium.org/) (Orchard et al. 2012) also actively curate and freely disseminate yeast protein interaction data:

- DIP (Database of Interacting Proteins: http://dip.doe-mbi.ucla.edu/) (Xenarios et al. 2002; Salwinski et al. 2004) and
- IntAct (IntAct Molecular Interaction Database: http://www.ebi.ac.uk/intact/) (Hermjakob et al. 2004b; Kerrien et al. 2012).

The MPact yeast protein interaction database also supports the PSI-MI standard (Guldener et al. 2006) (http://mips.helmholtz-muenchen.de/genre/proj/mpact) and provides protein interaction data contained in the MIPS Comprehensive Yeast Genome Database (CYGD) (Guldener et al. 2005).

BioGRID is currently unique among interaction databases in that it is the only open access resource that curates both genetic and protein interactions for different yeast species from published high- and low-throughput studies. In particular, BioGRID provides comprehensive interaction curation coverage for the budding and fission yeast literature, which is updated via continuous monthly increments that are fully archived (Reguly et al. 2006). BioGRID is also the only current resource that fully captures phenotypes using the APO for all curated yeast genetic interactions.

The ongoing development of the PhosphoGRID database allows full integration of documented yeast phosphorylation sites into the BioGRID PTM viewer, and thereby provides an integrated display of all curated phosphosites and their corresponding interactions in a single resource. In conjunction with WormBase, BioGRID has also developed a new Genetic Interactions (GI) ontology that will allow consistent curation of genetic interactions across various model organisms and interaction databases, enabling better comparisons of phenotypic information across different species (CA Grove, R Oughtred, A Winter, et al., unpubl.). The GI ontology has recently been incorporated into the PSI-MI ontology, with the intention that the GI ontology will be adopted as a community standard.

The BioGRID will continue to facilitate the use of yeast model systems for understanding the role of biological interaction networks in human biology and disease. Protein and genetic interactions, and entire interaction networks, are often physically and functionally conserved (Bandyopadhyay et al. 2006), such that the detailed interrogation of interactions in genetically tractable systems can prove extremely informative in biomedical contexts. To this end, efforts are now underway at BioGRID for parallel curation of yeast and human interactions that are implicated in human disease. These focused curation drives may be either biological process-centric, such as curation of the ubiquitin–proteasome

system (UPS) that controls the stability, localization, and activity of most of the proteome, or disease-centric, such as curation of interaction networks implicated in HIV or other infectious diseases, neurobiological disorders, and metabolic enzymes, all of which are currently in progress. New features planned for BioGRID include the incorporation of ubiquitination sites, which are comprehensively curated as part of the UPS project, into the integrated PTM display. Many expansive interaction networks implicated in prevalent human diseases will be the focus of future curation drives across multiple model organism species, from yeast to human. The integration of these network data sets with other data types, including expression data, quantitative phenotype data, and high-resolution sequence data, should help to enable predictive medicine and future drug discovery efforts.

ACKNOWLEDGMENTS

The authors thank Chris Grove and Paul Sternberg at WormBase for ongoing collaborative development of the Genetic Interaction Ontology. This work was supported by National Institutes of Health (NIH) grants R01OD010929 and R24OD011194 to M.T. and K.D., the Biotechnology and Biological Sciences Research Council (grant number BB/F010486/1 to M.T.), the Canadian Institutes of Health Research (grant number FRN 82940 to M.T.), and a Genome Québec International Recruitment Award and a Canada Research Chair in Systems and Synthetic Biology to M.T.

REFERENCES

Bandyopadhyay S, Sharan R, Ideker T. 2006. Systematic identification of functional orthologs based on protein network comparison. *Genome Res* **16:** 428–435.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2004. GenBank: Update. *Nucleic Acids Res* **32:** D23–D26.

Breitkreutz BJ, Stark C, Tyers M. 2003a. The GRID: The General Repository for Interaction Datasets. *Genome Biol* **4:** R23.

Breitkreutz BJ, Stark C, Tyers M. 2003b. Osprey: A network visualization system. *Genome Biol* **4:** R22.

Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. 2011. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* **39:** D685–D690.

Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, et al. 2013. The BioGRID interaction database: 2013 update. *Nucleic Acids Res* **41:** D816–D823.

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S. 2009. MatrixDB, a database focused on extracellular protein–protein and protein–carbohydrate interactions. *Bioinformatics* **25:** 690–691.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. 2012. *Saccharomyces* Genome Database: The genomics resource of budding yeast. *Nucleic Acids Res* **40:** D700–D705.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2:** 2366–2382.

Dolinski K, Chatr-Aryamontri A, Tyers M. 2013. Systematic curation of protein and genetic interaction data for computable biology. *BMC Biol* **11:** 43.

Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, Dwight SS, Fisk DG, Hirschman JE, Hitz BC, Hong EL, et al. 2010. *Saccharomyces* Genome Database provides mutant phenotype data. *Nucleic Acids Res* **38:** D433–D436.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, et al. 2013. STRING v9.1: Protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41:** D808–D815.

Guldener U, Munsterkotter M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J, Perez-Ortin JE, et al. 2005. CYGD: The Comprehensive Yeast Genome Database. *Nucleic Acids Res* **33:** D364–D368.

Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, Mewes HW, Stumpflen V. 2006. MPact: The MIPS protein interaction resource on yeast. *Nucleic Acids Res* **34:** D436–D441.

Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, von Mering C, et al. 2004a. The HUPO PSI's molecular interaction format—A community standard for the representation of protein interaction data. *Nat Biotechnol* **22:** 177–183.

Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, et al. 2004b. IntAct: An open source molecular interaction database. *Nucleic Acids Res* **32:** D452–D455.

Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415:** 180–183.

Inglis DO, Arnaud MB, Binkley J, Shah P, Skrzypek MS, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G. 2012. The *Candida* genome database incorporates multiple *Candida* species: Multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. *Nucleic Acids Res* **40:** D667– D674.

Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, Bader GD, Xenarios I, Wojcik J, Sherman D, et al. 2007. Broadening the horizon—Level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol* **5:** 44.

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al. 2012. The IntAct molecular interaction database in 2012. *Nucleic Acids Res* **40:** D841–D846.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. 2012. The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res* **40:** D1202–D1210.

Luc PV, Tempst P. 2004. PINdb: A database of nuclear protein complexes from human and yeast. *Bioinformatics* **20:** 1413–1415.

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ, FlyBase c. 2013. FlyBase: Improvements to the bibliography. *Nucleic Acids Res* **41:** D751–D757.

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.

Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FS, Cesareni G, et al. 2012. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat Methods* **9:** 345–350.

Oughtred R, Chatr-aryamontri A, Breitkreutz B-J, Chang CS, Rust JM, Theesfeld CL, Heinicke S, Breitkreutz A, Chen D, Hirschman J, et al. 2015. Use of the BioGRID database for analysis of yeast protein and genetic interactions. *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot088880.

Razick S, Magklaras G, Donaldson IM. 2008. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* **9:** 405.

Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hon GC, Myers CL, Parsons A, Friesen H, Oughtred R, Tong A, et al. 2006. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J Biol* **5:** 11.

Sadowski I, Breitkreutz BJ, Stark C, Su TC, Dahabieh M, Raithatha S, Bernhard W, Oughtred R, Dolinski K, Barreto K, et al. 2013. The PhosphoGRID *Saccharomyces cerevisiae* protein phosphorylation site database: Version 2.0 update. *Database (Oxford)* doi: 10.1093/database/bat026.

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. 2004. The database of interacting proteins: 2004 update. *Nucleic Acids Res* **32:** D449–D451.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res* **34:** D535–D539.

Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, et al. 2011. The BioGRID interaction database: 2011 update. *Nucleic Acids Res* **39:** D698–D704.

Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, Aslett M, Lock A, Bahler J, Kersey PJ, et al. 2012. PomBase: A comprehensive online resource for fission yeast. *Nucleic Acids Res* **40:** D695–D699.

Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. 2002. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* **30:** 303–305.

Yook K, Harris TW, Bieri T, Cabunoc A, Chan J, Chen WJ, Davis P, de la Cruz N, Duong A, Fang R, et al. 2012. WormBase 2012: More genomes, more data, new website. *Nucleic Acids Res* **40:** D735–D741.

BioGRID: A Resource for Studying Biological Interactions in Yeast

Rose Oughtred, Andrew Chatr-aryamontri, Bobby-Joe Breitkreutz, Christie S. Chang, Jennifer M. Rust, Chandra L. Theesfeld, Sven Heinicke, Ashton Breitkreutz, Daici Chen, Jodi Hirschman, Nadine Kolas, Michael S. Livstone, Julie Nixon, Lara O'Donnell, Lindsay Ramage, Andrew Winter, Teresa Reguly, Adnane Sellam, Chris Stark, Lorrie Boucher, Kara Dolinski and Mike Tyers

Cold Spring Harb Protoc; doi: 10.1101/pdb.top080754