

The BioGRID interaction database: 2013 update

Andrew Chatr-aryamontri¹, Bobby-Joe Breitkreutz², Sven Heinicke³, Lorrie Boucher², Andrew Winter⁴, Chris Stark², Julie Nixon⁴, Lindsay Ramage⁴, Nadine Kolas², Lara O'Donnell², Teresa Reguly², Ashton Breitkreutz², Adnane Sellam¹, Daici Chen¹, Christie Chang³, Jennifer Rust³, Michael Livstone³, Rose Oughtred³, Kara Dolinski³ and Mike Tyers^{1,2,4,*}

¹Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Quebec, Canada H3C 3J7, ²Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5, ³Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA and ⁴School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK

Received October 2, 2012; Accepted October 25, 2012

ABSTRACT

The Biological General Repository for Interaction Datasets (BioGRID: <http://thebiogrid.org>) is an open access archive of genetic and protein interactions that are curated from the primary biomedical literature for all major model organism species. As of September 2012, BioGRID houses more than 500 000 manually annotated interactions from more than 30 model organisms. BioGRID maintains complete curation coverage of the literature for the budding yeast *Saccharomyces cerevisiae*, the fission yeast *Schizosaccharomyces pombe* and the model plant *Arabidopsis thaliana*. A number of themed curation projects in areas of biomedical importance are also supported. BioGRID has established collaborations and/or shares data records for the annotation of interactions and phenotypes with most major model organism databases, including Saccharomyces Genome Database, PomBase, WormBase, FlyBase and The Arabidopsis Information Resource. BioGRID also actively engages with the text-mining community to benchmark and deploy automated tools to expedite curation workflows. BioGRID data are freely accessible through both a user-defined interactive interface and in batch downloads in a wide variety of formats, including PSI-MI2.5 and tab-delimited files. BioGRID records can also be interrogated and analyzed with a series of new bioinformatics tools, which include a post-translational modification viewer, a graphical viewer, a REST service and a Cytoscape plugin.

INTRODUCTION

The architecture and function of cellular interaction networks underpin the complex behavior of living systems. The network responses to both internal cues and exogenous stimuli and how environmental and/or genetic perturbations affect these responses are critical for understanding the molecular basis of human disease (1–3). Significant efforts have been made to chart the interaction networks of model organisms (4–7), based on advances in experimental techniques that allow the systematic exploration of biological interactions, both *in vivo* and *in vitro* (8,9). The integration of these various experimental datasets has begun to enable computational models of cellular interaction networks and the prediction of individual gene function in the regulation of cellular physiology.

The systematic curation of biological data, including protein and genetic interactions, is essential for computational biology and for the interpretation of genetic variation and disease associations revealed by genome-sequencing efforts (10,11). Biological interaction databases allow curated experimental datasets that would otherwise be dispersed in the biomedical literature to be accessed and exploited. These databases thus act as central repositories that provide a wealth of interaction data in a unified and common format, and thereby facilitate the exploration, visualization and integrative analysis of biological interaction networks. The Biological General Repository for Interaction Datasets (BioGRID: <http://thebiogrid.org>) is an open access database committed to the annotation of genetic and physical interactions between genes or gene products across all major model organism species. BioGRID is now a widely used resource that provides interaction datasets directly to the biological and computational communities, as well as to

*To whom correspondence should be addressed. Tel: +1 514 343 6668; Fax: +1 514 343 5839; Email: md.tyers@umontreal.ca; tyers@lunenfeld.ca; m.tyers@ed.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

several model organism database (MOD) partners. BioGRID data records can be used by the biomedical research community to generate and explore specific hypotheses about gene and network function, and as a benchmark for newly generated experimental high-throughput datasets.

DATA CONTENT AND ACCESS

Since our 2011 NAR Database report (12), the number of interactions curated and amassed in BioGRID has increased by >30%. As of September 2012 (version 3.1.92), BioGRID contains 527 569 protein and genetic interactions, of which 360 375 are non-redundant interactions. These interactions correspond to 309 819 (209 354 non-redundant) protein interactions and 217 750 (157 849 non-redundant) genetic interactions (Table 1). The data were directly extracted from 33 858 manually annotated peer-reviewed publications, which were identified from the corpus of biomedical literature by keyword searches, text-mining approaches and manual inspection of candidate abstracts. All BioGRID interaction records are directly mapped to experimental evidence in the supporting publication, as classified by a structured set of evidence codes (12).

BioGRID curation is focused on the parallel approaches of model organism-oriented curation and themed curation in human biology and disease. In addition to housing curated interaction data for more than 30 organisms, BioGRID has achieved exhaustive annotation of the literature for the budding yeast *Saccharomyces cerevisiae* (304 198 interactions), the fission yeast *Schizosaccharomyces pombe* (20 034 interactions) and the model plant *Arabidopsis thaliana* (16 664 interactions) (Table 1). These datasets are updated monthly and are directly linked from the respective MODs, *Saccharomyces* Genome Database (SGD) (13), PomBase (14) and The Arabidopsis Information Resource (TAIR) (15).

The complete manual annotation of all human interaction data documented in the biomedical literature remains a daunting task due to the sheer number of potentially relevant publications, now well in excess of 12 million papers in PubMed. To enable meaningful insights into human interaction networks, we have undertaken comprehensive curation of interactions in particular areas of biomedical interest. Current focused projects include central signaling conduits implicated in development and disease, such as the target of rapamycin (TOR), Wnt and TGF- β networks, disease-centric networks in breast cancer and HIV, and vital global processes such as the chromatin modification (CM) (16) and ubiquitin-proteasome systems (UPS). For example, the complex network of chromatin modifications that controls gene expression is dictated by at least 470 human genes annotated by the Gene Ontology (GO) process term 'chromatin remodelling' (16). Based on searches and text mining with this gene set, we recently curated more than 15 000 prioritized publications to yield 57 141 protein interactions from 7561 papers. In another example of a global cellular function, conjugation of the small conserved protein ubiquitin to myriad substrates controls the stability, activity and localization of most of the proteome (17). We manually annotated a set of 1140 genes that mediate the core functions of the UPS, including E1, E2 and E3 enzymes, deubiquitinating enzymes, ubiquitin-binding domain proteins, and proteasome core and auxiliary subunits. We have currently curated more than 5800 publications that bear evidence for 48 679 interactions (24 400 non-redundant interactions) in the UPS. These and other anticipated themed datasets will facilitate the prediction of individual gene function and network behavior within the major cellular regulatory systems.

DATA CURATION

Curation for BioGRID is performed by a dedicated team of PhD-level curators. A web-based interaction

Table 1. Increase in BioGRID data content since 2011 NAR Database Update

Organism	Type	August 2010 (3.0.67)			September 2012 (3.1.92)		
		Nodes	Edges	Publications	Nodes	Edges	Publications
<i>Arabidopsis thaliana</i>	PI	1735	4719	747	5915	16 476	1118
	GI	88	174	55	107	188	62
<i>Caenorhabditis elegans</i>	PI	2813	4663	12	2927	5010	93
	GI	1030	2112	5	1109	2326	22
<i>Drosophila melanogaster</i>	PI	7396	24 480	167	7998	35 843	314
	GI	982	9994	1466	1023	9934	1468
<i>Homo sapiens</i>	PI	9467	48 368	10 203	14 896	123 436	17 134
	GI	479	463	178	1291	1609	237
<i>Saccharomyces cerevisiae</i>	PI	5783	90 769	5444	6003	114 506	6601
	GI	5357	146 081	5606	5561	189 692	6686
<i>Schizosaccharomyces pombe</i>	PI	1441	4019	769	1773	6019	968
	GI	1340	11 527	953	1907	14 015	1158
Other organisms	ALL	2288	2985	830	8435	15 978	2724
Total	ALL	30 665	347 966	23 451	44 515	527 569	33 858

Data drawn from monthly release 3.0.67 and 3.1.92 of BioGRID. Nodes refer to gene or proteins, edges refer to interactions. PI, protein interaction; GI, genetic interaction.

management system (IMS) is used to build prioritized publication queues for different projects and facilitate the curation process through structured pull-down menus. The history of all curated data is tracked to each individual curator. Curators also help guide direct deposition by authors, which is particularly useful for pre-publication annotation of large-scale datasets and allows immediate public release of the data upon publication.

Within the past 2 years, BioGRID curators have begun to use text-mining tools to prioritize the relevant literature for each curation project (18). In turn, BioGRID supports the text-mining community by providing a gold-standard collection of manually curated interactions for the BioCreative challenge (19–22), a community-wide effort for evaluating text mining and information extraction systems applied to the biological domain. We have also established collaborations with WormBase (23) and the development team for the Textpresso text-mining tool (24). For example, the curation queue for the Wnt-signaling network is prioritized based on text-mining

results by Textpresso support vector machine (SVM) analyses, and ‘Textpresso for Wnt’ has also been set up as a text-mining interface to facilitate our curation. The overall curation pipeline of BioGRID is illustrated in Figure 1.

BioGRID actively collaborates with the extensive MOD community on different aspects of curation. For example, in collaboration with SGD, BioGRID curators have used the Yeast Phenotype Ontology (YPO) developed at SGD to assign structured phenotypes to over 200 000 budding and fission yeast genetic interactions. Collaborations are also underway with WormBase (23), ZFIN (25), FlyBase (26), MGI (27) and CGD (28) to coordinate interaction curation, and thereby leverage expertise and in-house MOD data that are relevant to biological interactions. For example, GO evidence codes generated by the MODs are often derived from publications that are likely to contain interaction data. Collaborations with the MODs have also led to an improved curation approach for higher organisms by implementation of

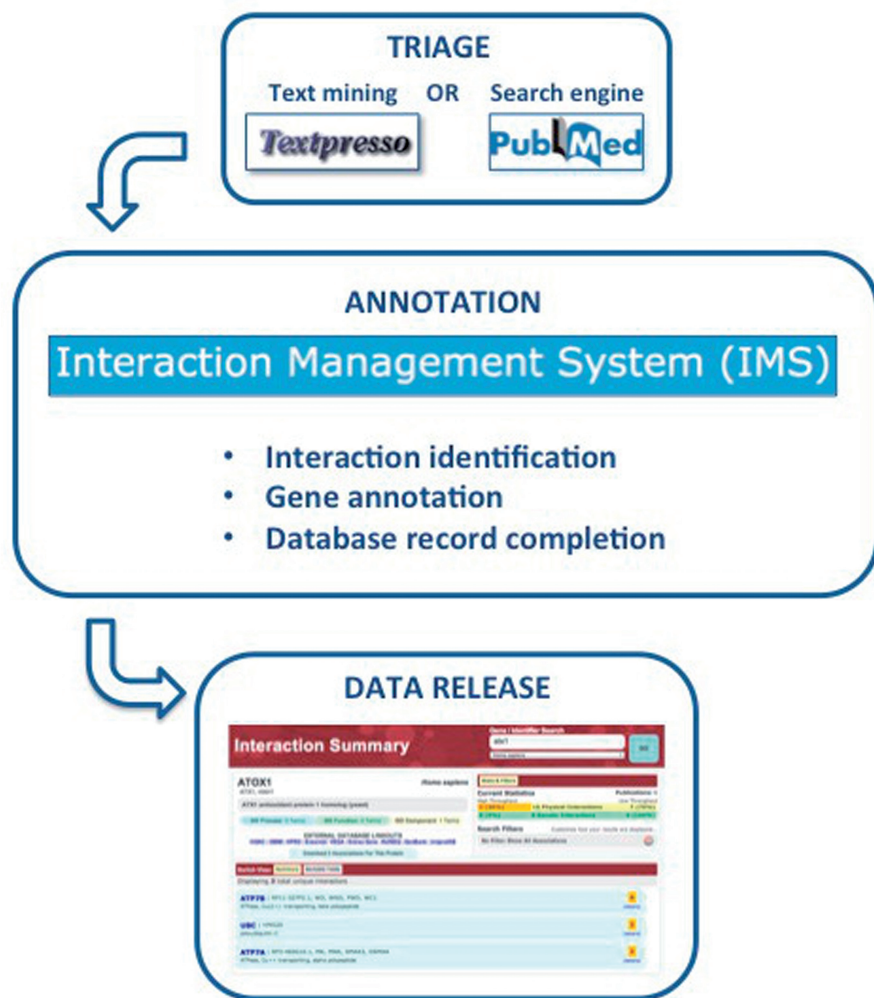


Figure 1. BioGRID curation pipeline. The curation workflow consists of three major steps: (i) triage of the literature of interest by text-mining tools and/or interaction-directed PubMed queries; (ii) curation, annotation and tracking of interaction data through the web-based IMS and (iii) monthly public release of interaction data records.

species-specific phenotype ontologies and the broadening of interaction terms to capture more complex genetic interaction data. The different biology of various organisms used in biomedical research presents a formidable challenge in the annotation and interpretation of genetic interactions, and in the reconciliation of structured phenotypes across all species. In order to meet this challenge, in conjunction with WormBase (23), and supported by other MODs such as SGD (13), CGD (28), PomBase (14), FlyBase (26), TAIR (15) and ZFIN (25), we have developed a universal genetic interaction (GI) ontology that enables the annotation of more complex phenotypic outcomes associated with genetic interactions from higher organisms. The genetic interaction ontology has been submitted to the PSI-MI editorial committee (29) and will be made publicly available with the next official PSI-MI ontology release.

DATABASE ARCHITECTURE AND USER INTERFACE

In order to ensure consistent reliability and accessibility of the BioGRID web interface, we have migrated the BioGRID to a cloud-based server system with a third party provider that provides up-to-date hardware, facile operating system upgrades and improved fault tolerance. BioGRID 3.2 supports ~28 million systematic names, aliases, official symbols and external identifiers from Ensembl (30), UniprotKB (31), NCBI Entrez-Gene (32), Genbank (33), SGD (13), WormBase (23), FlyBase (26), HGNC (34), MGD (27), TAIR (15), VectorBase (35), BeeBase (36), ZFIN (25) and HPRD (37), among other sources. BioGRID currently also supports annotation for more than 85 organisms and contains interaction data for more than 30 different species. The BioGRID web service (webservice.thebiogrid.org) has been completely redesigned to run off the new decentralized database architecture for better access and maintenance by developers. The new web service will facilitate the incorporation of BioGRID data in other databases and applications. Additional new documentation in the BioGRID wiki provides comprehensive instructions for this resource.

The BioGRID 3.2 web interface has been upgraded to include an integrated post-translational modification (PTM) viewer. This viewer highlights PTM sites on protein sequences and incorporates much of the functionality available in PhosphoGRID (Figure 2). PTM sites are colored within protein sequences according to the modification type, with clickable functions that display details such as publications, evidence codes and enzymatic relationships. The BioGRID currently supports both phosphorylation and ubiquitination sites and will expand to cover other major PTMs across all supported species.

To facilitate exploration of the biological datasets in BioGRID, we have developed a new gene tag feature for specific annotation, including membership in network-specific cohorts, gene functions or detailed attributes such as PTM site information. These gene tags can be used to build customized datasets for downloads and to define criteria for building and maintaining

project-specific datasets, as for example defined by themed curation projects. These datasets may be maintained in concert with monthly BioGRID updates and are subject to strict version controls that allow reference to specific builds for data analysis. Project-specific datasets—such as for the CM and UPS datasets—will be accessed through custom gateways within the BioGRID that encompass genes, interactions, publications and biological context for the project.

Graphical network representation provides an intuitive summary of an interaction query dataset and, when appropriately configured with a dynamic interface, can be used to inspect and further query a network of interest. However, a drawback of current network visualization software is that the graphical output becomes cluttered as network complexity increases. To address this issue, we have developed a new dynamic BioGRID interaction viewer that is based on a simple visual layout and which has user-friendly filters (Figure 3). In the BioGRID viewer, all interaction nodes are distributed in a circular layout with the query gene in the center. The properties of individual interactions are visualized by moving the tooltip over the interactor of interest to highlight gene information, including species type, gene acronym, gene identifier and number of interactions in BioGRID. To facilitate retrieval of data types of interest, the viewer provides the user with a check-box filter to reduce the complexity of the graph. The user can thus choose to view only those interactions supported by particular types of experimental evidence, low- versus high-throughput data or species-specific data. The network can also be extended to include all known interactions between interactors for the query. Results of the filtered and/or extended query are downloadable in tab2 format with a single click. The BioGRID viewer is based on an open source widget, downloadable from GitHub (<https://github.com>), and is embeddable in any web page. Network images in the viewer are rendered from interactions retrieved from the BioGRID REST service. The viewer is implemented using the d3js (<http://d3js.org/>) library and requires a browser that supports JavaScript and SVG, which includes modern browsers such as Chrome, Firefox and Safari.

DATA ACCESS AND DISTRIBUTION

BioGRID datasets are updated and archived every month and can be freely accessed through widely used community resources over the internet and a number of dedicated bioinformatic tools. Records are now available interactively through the BioGRID web search page for download in a variety of XML (PSI-MI 1.0, PSI-MI 2.5) and tabular (tab, tab2 and mitab) formats and are also available through NCBI Entrez-Gene (32), DroID (38) and GermOnline (39), through several major MODs such as FlyBase (26), TAIR (15), SGD (13) and PomBase (14), and through meta-databases such as STRING (40), iRefIndex (41) and Pathway Commons (42). BioGRID datasets can also be directly interrogated through network visualization and analysis suites, including the original Osprey viewer (43), Cytoscape (44)

CDC28 *Saccharomyces cerevisiae*
 CDK1, SRM5, HSL5, YBR160W

State & Filters
Current Statistics PTM Publications: 19
 Total Proteins with PTMs: 1
 Total PTM Sites: 4
PTM Site Distribution
 Phosphorylation - 4 Sites from 19 Publications

EXTERNAL DATABASE LINKOUTS
 SGD | Entrez Gene | RefSeq | GeneDB | GeneDB | PhosphoSitePlus | UniProt | TrEMBL | PDB | Ensembl | KEGG | NCBI | EMBL | GenBank | BioGRID | ChEMBL | Pubmed | UniProt | TrEMBL | PDB | Ensembl | KEGG | NCBI | EMBL | GenBank | BioGRID | ChEMBL | Pubmed

POST TRANSLATIONAL MODIFICATION - SUMMARY
 NP_009718 | Catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK); associates with G1 cyclins (CLNs) and G2/M cyclins (CLBs) which direct the CDK to specific substrates

1 MSGELANYKR LEKVGEGT**Y**G VVYKALDLRP GQGQRVVALK KIRLSESEDEG VPSTAIRESIS LLKELKDDNI
 71 VRLYDIVHSD ANKLYLVFEP LDLDLKRYME GIFPKDQLGA DIVKFKFMMQL CKGIAYCHSH RILHRDLKPKQ
 141 NLLINKDGNL KLGDFGLARA FGVPLRAY**T**H EIVTLWYRAP EVLLGGKQYS TGVDTWSIGC IFAEMCWRKP
 211 IFSGDSEIDQ IPKIFRVLGT PNEAIWFDIV YLPDFKPSFP QWRKRDLSQV VPSLDPRGID LLDKLLAYDF
 281 INRISARRAA IHPYFQES

Location	Residue	PTM	Source(s)	Relationship(s)	Notes
19	Y	Phosphorylation	Albuquerque CP (2008) Bocher RN (1993) Gned F (2009) Gruhler A (2005) Helbig AO (2010) Holt LJ (2009) Huber A (2009) Keaton MA (2007) Sorgor PK (1992) Soufi S (2009) Soulard A (2010)	SWE1 (Kinase)	
46	S	Phosphorylation	Russo GL (2000) Russo GL (2001)	CKA1 (Kinase) CKA2 (Kinase) CKA3 (Kinase) CKA4 (Kinase)	
168	Y	Phosphorylation	Li X (2007)	-	
169	T	Phosphorylation	Cross FR (1998) Gruhler A (2005) Helbig AO (2010) Holt LJ (2009) Li X (2007) Lim HT (1996) Ross KE (2000) Schaber H (2002) Thuret JY (1996)	CKA1 (Kinase)	

Protein	Relationship	Location	Residue	PTM	Relationship Source(s)	Identity
SWE1 (Kinase)	Phosphorylates	19	Y	Phosphorylation	Bocher RN (1993) Keaton MA (2007)	catalytic
CKA1 (Kinase)	Phosphorylates	46	S	Phosphorylation	Russo GL (2000) Russo GL (2001)	catalytic
CKA2 (Kinase)	Phosphorylates	46	S	Phosphorylation	Russo GL (2000)	catalytic

Figure 2. PTM display features. (A) Button to reveal PTM sites. (B) Statistics for different types of PTMs. (C) Pop-up with links to publications that document PTM evidence and relationships. (D) Tabular view of PTM site locations and links to publications. (E) Tabular view of PTM relationships and links to publications. (F) Custom gene tags.

and GeneMANIA (45). Notably, BioGRID data have recently been dynamically integrated into the ProHits LIMS system (46) in order to allow real-time comparison of experimental mass spectrometry data to published data housed in BioGRID.

In 2012, Google Analytics reported that the BioGRID received on average 69 237 page views and 10 110 unique visitors per month, versus 64 298 page views and 9928

unique visitors per month in 2011. BioGRID data files were downloaded on average 6900 times per month, compared with 6400 downloads per month in 2011. These statistics do not include the widespread dissemination by the various partner websites listed above that host BioGRID interaction data. The BioGRID user base is located primarily in the USA (37%), followed by UK (8%), Germany (7%), Canada (6%), Japan (6%),

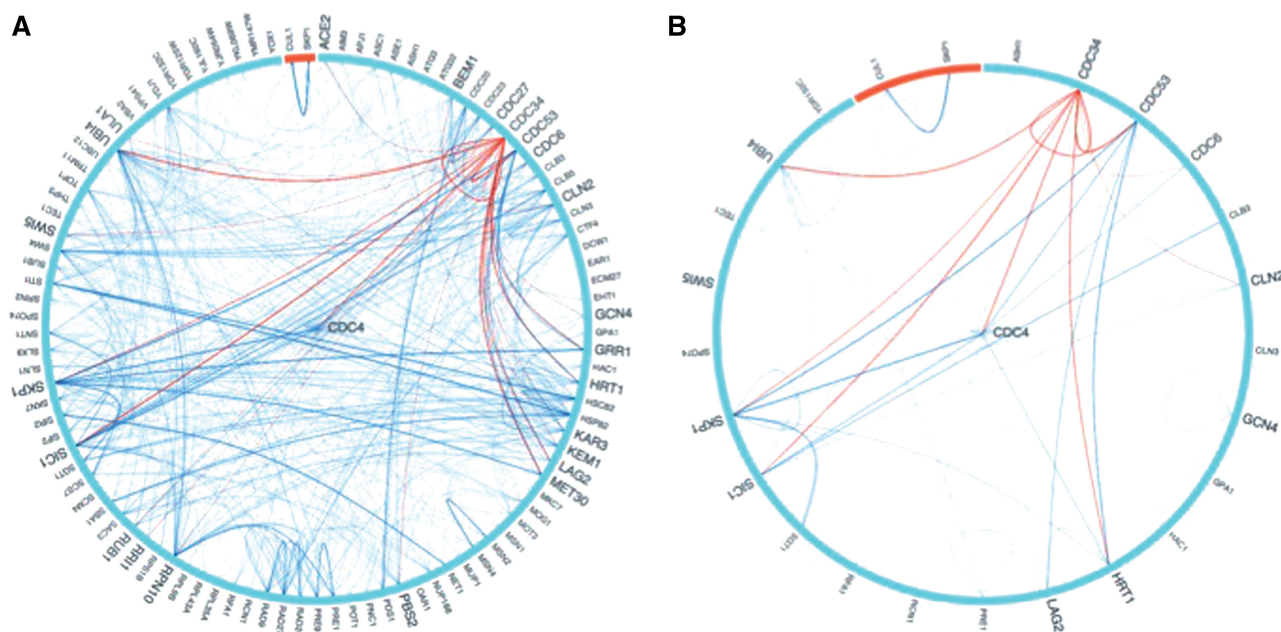


Figure 3. The BioGRID viewer. (A) In the example shown the viewer returns all the interactions of the query gene/protein and interactions between the hit genes/protein. Properties of individual interactions are revealed by hovering over the interaction node of interest. (B) Network view can be simplified by several available filters. In this example, all genetic interactions have been filtered out. See text for further details.

China (5%), France (4%), India (4%), Spain (2%) and all other countries (25%).

In order to facilitate the access and interoperability of BioGRID data with multiple platforms, we recently developed a BioGRID representational state transfer (REST) service and a BioGRID plug-in for the widely used Cytoscape visualization system (47). The BioGRID REST service grants full URL-based access to the BioGRID data and also provides the user with specific parameters to filter the data by various attributes. For example, the REST service drives a related tool called BioGRID Webgraph that generates network representations from user-provided gene lists. The dedicated Cytoscape plug-in acts as a web service client that provides facile import and filtering of the full BioGRID dataset for visualization and analysis in Cytoscape (44).

FUTURE DEVELOPMENTS

The BioGRID will continue to provide the biomedical and biological research communities with up-to-date, high-quality and extensively annotated protein and genetic interaction data, along with the requisite software tools to search, visualize and analyze interaction datasets. BioGRID will also continue to participate in the IMEx consortium of interaction databases (48). In addition to ongoing curation of interactions for the major model organism species, we will expand species coverage in order to facilitate interolog analyses, in particular to enable comparison of interaction networks across model organism species and humans. We have recently initiated the systematic annotation of protein and genetic interactions for *Candida albicans*, which is an important emerging model system and a prevalent

human pathogen. We have also initiated the annotation of the human HIV1 interactome in the context of the Linking Animal Models to Human Disease Initiative (see <http://www.lamhdi.org>). These and other nascent projects will be facilitated by the development of more efficient text-mining tools through collaborations with Textpresso and others. This cross-species and themed approach to curation will enable new insights into human biology and disease by integration of interaction data from multiple model organism systems.

ACKNOWLEDGEMENTS

The authors thank Mike Cherry, Paul Sternberg, Bill Gelbart, David Botstein, Henning Hermjakob, Shoshana Wodak, Anne-Claude Gingras, Gary Bader, Chris Sander, Val Wood, Gavin Sherlock, Ivan Sadowski, Lincoln Stein, Judy Blake, Monty Westerfield, Maryann Martone, Mark Ellisman and Olga Troyanskaya for many helpful discussions. Particularly thanks are due to Chris Grove and other colleagues at WormBase for ongoing collaborative development of the genetic interaction ontology.

FUNDING

The National Institutes of Health National Center for Research Resources [R01RR024031 and R24RR032659 to M.T. and K.D.]; the Biotechnology and Biological Sciences Research Council [BB/F010486/1 to M.T.]; the Canadian Institutes of Health Research [FRN 82940 to M.T.]; the European Commission FP7 Program [2007-223411 to M.T.] and a Genome Québec International Recruitment Award and a Canada Research Chair in Systems and Synthetic Biology [to

M.T.J. Funding for open access charge: The National Institute of Health.

Conflict of interest statement. None declared.

REFERENCES

- Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, **487**, 491–495.
- Ryan, O., Shapiro, R.S., Kurat, C.F., Mayhew, D., Baryshnikova, A., Chin, B., Lin, Z.Y., Cox, M.J., Vizeacoumar, F., Cheung, D. *et al.* (2012) Global gene deletion analysis exploring yeast filamentous growth. *Science*, **337**, 1353–1356.
- Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K. *et al.* (2012) Global landscape of HIV-human protein complexes. *Nature*, **481**, 365–370.
- Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G. *et al.* (2010) A global protein kinase and phosphatase interaction network in yeast. *Science*, **328**, 1043–1046.
- Babu, M., Vlasblom, J., Pu, S., Guo, X., Graham, C., Bean, B.D., Burston, H.E., Vizeacoumar, F.J., Snider, J., Phanse, S. *et al.* (2012) Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature*, **489**, 585–589.
- Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S. *et al.* (2012) A census of human soluble protein complexes. *Cell*, **150**, 1068–1081.
- LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C. *et al.* (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*, **438**, 103–107.
- Bensimon, A., Heck, A.J. and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.*, **81**, 379–405.
- Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X. *et al.* (2011) Next-generation sequencing to generate interactome datasets. *Nat. Methods*, **8**, 478–480.
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S. *et al.* (2008) Big data: the future of biocuration. *Nature*, **455**, 47–50.
- Ideker, T., Dutkowsky, J. and Hood, L. (2011) Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell*, **144**, 860–863.
- Stark, C., Breitkreutz, B.J., Chatr-aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X. *et al.* (2011) The BioGRID interaction database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.
- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R. *et al.* (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.
- Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bahler, J., Kersey, P.J. *et al.* (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**, D695–D699.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
- Turinsky, A.L., Turner, B., Borja, R.C., Gleeson, J.A., Heath, M., Pu, S., Switzer, T., Dong, D., Gong, Y., On, T. *et al.* (2011) DAN-CER: disease-annotated chromatin epigenetics resource. *Nucleic Acids Res.*, **39**, D889–D894.
- Varshavsky, A. (2012) The ubiquitin system, an immense realm. *Annu. Rev. Biochem.*, **81**, 167–176.
- Hirschman, L., Burns, G.A., Krallinger, M., Arighi, C., Cohen, K.B., Valencia, A., Wu, C.H., Chatr-Aryamontri, A., Dowell, K.G., Huala, E. *et al.* (2012) Text mining for the biocuration workflow. *Database*, **2012**, April 18 (doi:10.1093/database/bas020; epub ahead of print).
- Krallinger, M., Leitner, F., Vazquez, M., Salgado, D., Marcelle, C., Tyers, M., Valencia, A. and Chatr-aryamontri, A. (2012) How to link ontologies and protein-protein interactions to literature: text-mining approaches and the BioCreative experience. *Database*, **2012**, March 21 (doi:10.1093/database/bas017; epub ahead of print).
- Arighi, C.N., Roberts, P.M., Agarwal, S., Bhattacharya, S., Cesareni, G., Chatr-Aryamontri, A., Clematide, S., Gaudet, P., Giglio, M.G., Harrow, I. *et al.* (2011) BioCreative III interactive task: an overview. *BMC Bioinformatics*, **12**(Suppl. 8), S4.
- Krallinger, M., Vazquez, M., Leitner, F., Salgado, D., Chatr-Aryamontri, A., Winter, A., Peretto, L., Briganti, L., Licata, L., Iannuccelli, M. *et al.* (2011) The protein-protein interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. *BMC Bioinformatics*, **12**(Suppl. 8), S3.
- Chatr-Aryamontri, A., Winter, A., Peretto, L., Briganti, L., Licata, L., Iannuccelli, M., Castagnoli, L., Cesareni, G. and Tyers, M. (2011) Benchmarking of the 2010 BioCreative Challenge III text-mining competition by the BioGRID and MINT interaction databases. *BMC Bioinformatics*, **12**(Suppl. 8), S8.
- Yook, K., Harris, T.W., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., 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.
- Muller, H.M., Kenny, E.E. and Sternberg, P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.
- Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D.G., Knight, J., Mani, P., Martin, R., Moxon, S.A. *et al.* (2011) ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res.*, **39**, D822–D829.
- McQuilton, P., St Pierre, S.E. and Thurmond, J. (2012) FlyBase 10: the basics of navigating FlyBase. *Nucleic Acids Res.*, **40**, D706–D714.
- Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A. and Richardson, J.E. (2012) The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.*, **40**, D881–D886.
- Inglis, D.O., Arnaud, M.B., Binkley, J., Shah, P., Skrzypek, M.S., Wymore, F., Binkley, G., Miyasato, S.R., Simison, M. and 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, A.F., Vinod, N., Bader, G.D., 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.
- Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.
- UniProt Consortium. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **40**, D71–D75.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetverin, V., Church, D.M., Dicuccio, M., Federhen, S. *et al.* (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **40**, D13–D25.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res.*, **39**, D32–D37.
- Seal, R.L., Gordon, S.M., Lush, M.J., Wright, M.W. and Bruford, E.A. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
- Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., Campbell, K.S., Christophides, G.K., Christley, S., Dyalynas, E. *et al.* (2009) VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.*, **37**, D583–D587.

36. Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., Childs, K.L., Anzola, J.M., Milshina, N. and Elsik, C.G. (2011) Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. *Nucleic Acids Res.*, **39**, D658–D662.
37. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
38. Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G. 3rd and Finley, R.L. Jr (2011) DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.*, **39**, D736–D743.
39. Lardenois, A., Gattiker, A., Collin, O., Chalmel, F. and Primig, M. (2010) GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database*, **2010**, Dec 10 (doi:10.1093/database/baq030; epub ahead of print).
40. Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
41. Razick, S., Magklaras, G. and Donaldson, I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, **9**, 405.
42. Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D. and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
43. Breitkreutz, B.J., Stark, C. and Tyers, M. (2003) Osprey: a network visualization system. *Genome Biol.*, **4**, R22.
44. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
45. Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.
46. Liu, G., Zhang, J., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z.Y., Breitkreutz, B.J., Ding, Y., Colwill, K., Pasculescu, A. *et al.* (2010) ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nat Biotechnol*, **28**, 1015–1017.
47. Winter, A.G., Wildenhain, J. and Tyers, M. (2011) BioGRID REST Service, BiogridPlugin2 and BioGRID WebGraph: new tools for access to interaction data at BioGRID. *Bioinformatics*, **27**, 1043–1044.
48. Orchard, T., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkman, F.S.L., Cesareni, G. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.