Nucleic Acids Research Advance Access published November 10, 2010

Nucleic Acids Research, 2010, ***1–6***
doi:10.1093/nar/gkq1039

Pathway Commons, a web resource for biological pathway data

Ethan G. Cerami[1,2], Benjamin E. Gross[1], Emek Demir[1], Igor Rodchenkov[3], Özgün Babur[1], Nadia Anwar[1], Nikolaus Schultz[1], Gary D. Bader[3] and Chris Sander[1]

[1]Computational Biology Center, Memorial Sloan-Kettering Cancer Center 1275 York Avenue, Box 460, New York, NY 10065, [2]Tri-Institutional Training Program in Computational Biology and Medicine, New York, USA and [3]Banting and Best Department of Medical Research, The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada

Received August 24, 2010; Revised October 7, 2010; Accepted October 11, 2010

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010

ABSTRACT

Pathway Commons (http://www.pathwaycommons.org) is a collection of publicly available pathway data from multiple organisms. Pathway Commons provides a web-based interface that enables biologists to browse and search a comprehensive collection of pathways from multiple sources represented in a common language, a download site that provides integrated bulk sets of pathway information in standard or convenient formats and a web service that software developers can use to conveniently query and access all data. Database providers can share their pathway data via a common repository. Pathways include biochemical reactions, complex assembly, transport and catalysis events and physical interactions involving proteins, DNA, RNA, small molecules and complexes. Pathway Commons aims to collect and integrate all public pathway data available in standard formats. Pathway Commons currently contains data from nine databases with over 1400 pathways and 687 000 interactions and will be continually expanded and updated.

INTRODUCTION

Pathway information captures knowledge of biological processes at the molecular level and can be an important tool for interpreting the growing amount of biological data from genomic studies. For example, molecular profiling data, molecular interaction screens, genome re-sequencing projects and genome wide association studies are frequently interpreted using prior information about biological processes with pathway enrichment analysis (1). Pathway and network information can also be usefully combined with high-throughput genomic data and clinical phenotype data to investigate the network properties of specific disease types (2) and to build classifiers for disease subtypes (3). Ideally one would use all available pathway information for these analyses. Unfortunately, this information is currently highly fragmented among many databases, making it difficult for individual researchers to access a complete pathway data set (4). Pathway Commons (http://www.pathwaycommons.org) is a freely available database that collects, normalizes and integrates publicly available biological pathway and molecular interaction data about cellular processes and provides it to researchers via a convenient point of access. Pathway Commons currently integrates data from nine public databases and contains over 1400 pathways and 687 000 interactions (Table 1), and aims to include all public pathway data available in standard formats. Pathway Commons provides a web-based interface that enables biologists to browse and search, a download site that provides integrated bulk sets of pathway information in standard or convenient formats and a web service that software developers can use to conveniently query and access all data.

DATA

Pathway Commons collects data in the standard BioPAX format for biological pathway exchange (5). Depending on the source database, data may include proteins, small molecules, DNA, RNA, complexes and their cellular locations, different types of physical interactions, such as molecular interaction, biochemical reaction, catalysis, complex assembly and transport, post-translational protein modifications, original citations, experimental evidence and links to other database information, such as protein sequence annotation. Future versions of Pathway Commons will expand this to include genetic interactions and gene regulatory networks. Pathway Commons redistributes data from primary databases,

Please address correspondence to: pc-info@pathwaycommons.org

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Pathway Commons currently includes pathway and interaction information from nine sources

Data Source	Format	Size	Updated	Focus (species)	Reference or URL
BioGRID	PSI–MI 2.5	347 508 Interactions	August 2010 (3.0.67)	Model organisms	(20)
Cancer Cell Map	BioPAX L2	10 Pathways 2104 Interactions	May 2006	Human	http://cancer.cellmap.org
HPRD	PSI–MI 2.5	40 618 Interactions	13 April 2010 Version 9	Human	(21)
HumanCyc	BioPAX L2	266 Pathways 4879 Interactions	16 June 2010 Version 14.1	Human	(22)
IMID	BioPAX L2	1729 Interactions	March, 2009	Human	http://www.sbcny.org/
IntAct	PSI–MI 2.5	154 567 Interactions	8 August 2010 Version 3.1, r14760	All	(23)
MINT	PSI–MI 2.5	117 202 Interactions	28 July 2010	All	(24)
NCI/Nature PID	BioPAX L2	186 Pathways 13 879 Interactions	10 August 2010	Human	(25)
Reactome	BioPAX L2	1015 Pathways 5397 Interactions	18 June 2010 Version 33	Human	(5)
All Integrated	BioPAX L2	1477 Pathways 687 883 Interactions		Multiple	http:///www.pathwaycommons.org

New sources are periodically added and listed on the Pathway Commons website. Note that pathway and interaction statistics represent non-unique counts from source databases, as these records are not currently merged from multiple sources (only molecules are currently merged).

thus users must cite all primary sources to support the curation teams that share pathway data. All data is made available under the original license terms of the primary databases, which is specified in BioPAX files and with each record available from the website.

USING PATHWAY COMMONS

Browse and search using the website

The Pathway Commons web interface is designed primarily for scientists who want to browse or search pathways and interactions across multiple pathway databases. Users can answer questions such as ‘What proteins interact with my favorite protein?’, ‘What pathways involve my favorite protein?’, ‘Is my favorite protein involved in transport events or biochemical reactions?’, or ‘What enzymes use my favorite metabolite as a substrate?’. The system is designed to make it easy for users to find a pathway or molecule of interest, and then, for pathways, drill down to view embedded components, such as biochemical reactions, complexes and proteins, or for molecules, work their way up to view interactions or pathways that the molecule participates in. A simple search form on the home page enables a user to search for pathways or molecules. If the user is interested in browsing pathways, they enter a keyword, such as the gene symbol ‘BRCA1’, and retrieve the list of pathways containing the keyword ‘BRCA1’, and the list of pathways that contain the *BRCA1* gene. If the user is interested in molecules, they enter a keyword, and retrieve the list of proteins, genes, or small molecules that contain it. Keywords and frequently used molecule identifiers are recognized [i.e. gene symbols, UniProt, Entrez Gene and RefSeq identifiers (IDs) for genes and proteins and BioCyc and KEGG Ligand IDs for small molecules]. Search results are ranked by relevance and summarized on a page that allows users to narrow their query by data source, for example, selecting only pathways from the Reactome database (5). By default, all Pathway Commons queries run on all included molecular networks allowing users to search multiple data sources. However, users can also restrict searches to specific data sources or organisms by setting filters (available from a link under the main search box).

Clicking on a search result brings the user to a Pathway or Molecule record page. Every record page allows users to easily navigate to related pathway components through hyperlinks. Links to the source pathway database are provided for every record. Additional information about the record can be found by following hyperlinks on each web page to external sites, such as PubMed, Entrez, UniProt and Gene Ontology (6–8). Each molecule page provides a ‘Neighborhood Map’ that graphically displays the interactions and complexes that the molecule participates in using a reduced simple interaction format (SIF) (described below). Users can click the thumbnail neighborhood map for an enlarged network view, or visualize the complete BioPAX network within Cytoscape, an open source network visualization and analysis software platform (9). Users can also download the record in different formats for import into software, such as Cytoscape (download in BioPAX format), a spreadsheet (Gene Sets) or Gene Set Enrichment Analysis (GSEA—Gene Set: Gene Symbols) (10). Finally, stable links are provided to molecule pages, where stable molecule identifiers are available. Stable links to pathway pages are not currently available, as there is no community standard for globally identifying pathway records, and many source databases do not maintain their own stable identifiers for pathway records. However, where source databases do maintain stable pathway records IDs, such as in the case of Reactome and HumanCyc, Pathway Commons plans to make stable pathway links available for these data sources in the future.

Data quality of Pathway Commons is dependent on the quality of the pathways from source databases. Pathway Commons allows users to filter data by various criteria, including data source, which allows viewing a restricted subset of high quality data. In the future, Pathway Commons will implement published algorithms (11,12)

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010

to automatically assess data quality and enable this as an additional filter.

Query relevance score and syntax

Pathway Commons can be queried with any keyword or recognized molecule identifier, and will return a ranked list of entities, interactions and pathways containing the keyword. Results are ranked using scores derived from the Lucene text indexing software used to implement the search (http://lucene.apache.org/). Specific fields are also indexed or boosted to improve search results (including data source, gene names, organism, record type and database IDs). If a user enters a gene symbol, exact matches will be highlighted at the top of the results page. The full Lucene query syntax is supported, which enables quoted string search, Boolean query logic and field specific searches, although we anticipate that most users will query the system using simple keyword searches.

Download large data sets

All of the information in Pathway Commons, including individual BioPAX exports from data providers and the integrated network of all Pathway Commons information, per organism or per data source, can be downloaded in different formats (described below) for computational analysis (linked from the Download tab on the home page). Detailed information about each file format is available on the download site and summarized below.

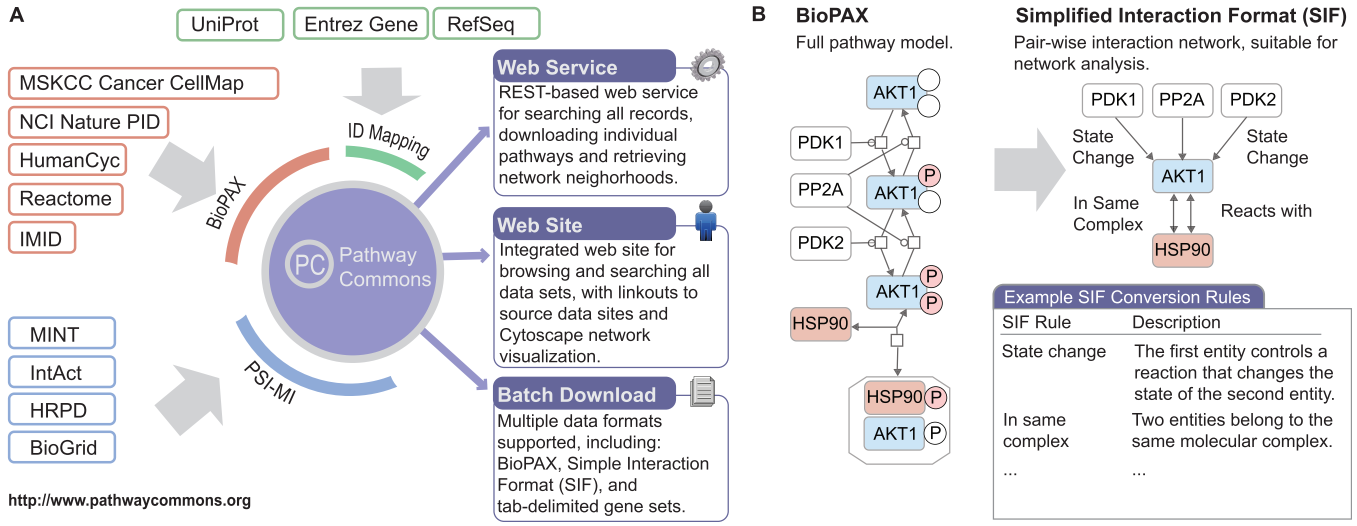
Simple Interaction Format

Many network analysis algorithms require pairwise interaction networks as input. A BioPAX network often contains more complex relationships with multiple participants, such as biochemical reactions. To make it easier to use all of the pathway information in Pathway Commons with typical network analysis tools, we developed a set of rules to reduce BioPAX interactions to pairwise relationships (Figure 1). Since SIF interactions are always binary it is not possible to fully represent all of BioPAX, thus this translation is lossy in general. Nonetheless, the SIF network is useful for those applications that require pairwise interaction input (2). SIF format can be easily imported into popular network analysis tools, such as Cytoscape (9) (Figure 2).

Gene set enrichment formats

Over-representation analysis (ORA) is frequently used to assess the statistical enrichment of known gene sets (e.g. pathways) in a discrete or ranked list of genes (1). This type of analysis is useful for summarizing large gene lists and is commonly applied to genomics data sets. A popular software system for analyzing ranked gene lists is Gene Set Enrichment Analysis (GSEA) (13). The 'Gene sets' used by GSEA are stored for convenience in the Molecular Signature Database (MSigDB) (10) in the Gene Matrix Transposed file format (*.gmt). All pathways in Pathway Commons are mapped to this format and pathway participants are specified as official gene symbols (if an official gene symbol is not available, the participant is not exported). All participants for a pathway must come from the same species as the pathway; therefore, some participants from cross-species pathways are removed. Thus, data from Pathway Commons can be used directly with GSEA software. Pathway Commons also makes available a second type of gene set format that contains more information about genes. This resembles the MSigDB format described above, except that all participants are micro-encoded with multiple identifiers.

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010



Web Service: REST-based web service for searching all records, downloading individual pathways and retrieving network neighborhoods.

Web Site: Integrated web site for browsing and searching all data sets, with linkouts to source data sites and Cytoscape network visualization.

Batch Download: Multiple data formats supported, including: BioPAX, Simple Interaction Format (SIF), and tab-delimited gene sets.

SIF Rule	Description
State change	The first entity controls a reaction that changes the state of the second entity.
In same complex	Two entities belong to the same molecular complex.
...	...

Figure 1. (**A**) Overview of Pathway Commons. Pathway Commons (http://www.pathwaycommons.org) provides a central, convenient point of access to multiple publicly available pathway and interaction databases. It does so by integrating data sources formatted in the BioPAX or PSI–MI standards, and making all data sets available via a unified web site, a single web service interface and a batch download site. (**B**) Pathway Commons supports multiple output formats, including BioPAX and SIF. To support SIF, we have developed a set of inference rules for converting the full BioPAX model to a simplified pairwise interaction network. Left: The full BioPAX model of the AKT pathway, shown as an SBGN process diagram (19). Upper right: the same pathway, converted to a SIF representation. Lower right: name and description of SIF rules used in the conversion. The full set of SIF rules is available on the Pathway Commons website (http://www.pathwaycommons.org/pc/sif_interaction_rules.do).

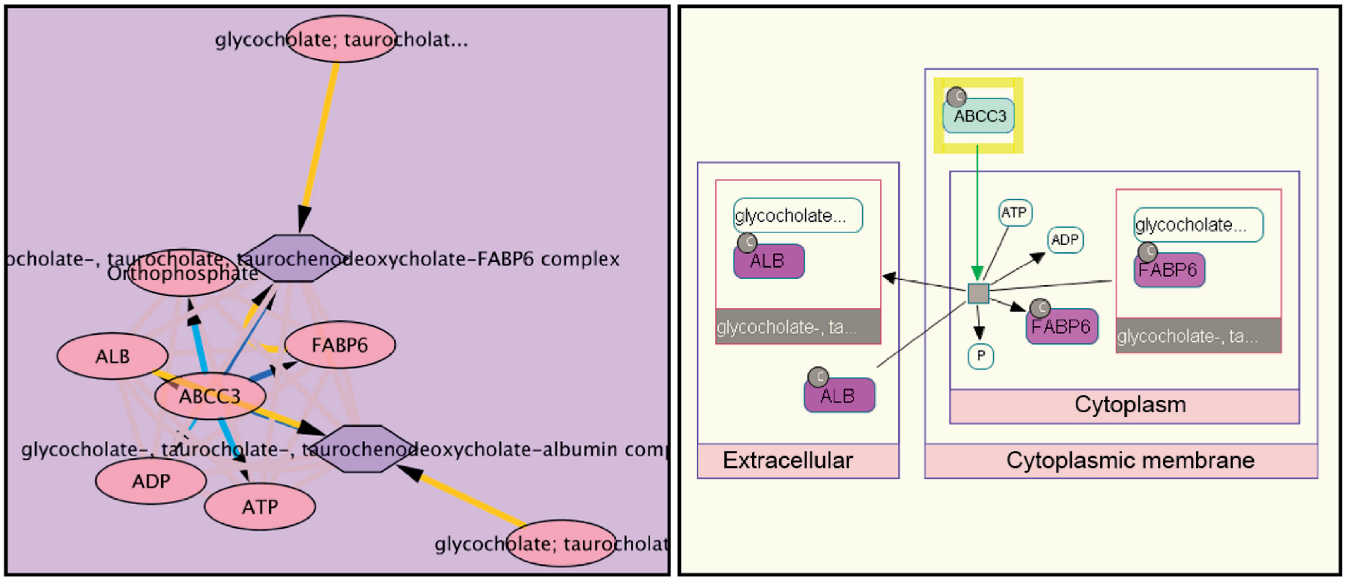

glycocholate; taurocholat...
ocholate–, taurocholate, taurochenodeoxycholate–FABP6 complex
Orthophosphate
ALB
FABP6
ABCC3
glycocholate–, taurocholate–, taurochenodeoxycholate–albumin comp
ADP
ATP
glycocholate; taurocholat
ABCC3
ATP
ADP
glycocholate...
ALB
glycocholate-, ta...
glycocholate...
FABP6
glycocholate-, ta...
FABP6
P
ALB
Cytoplasm
Extracellular
Cytoplasmic membrane

Figure 2. Neighborhood of ABCC3—a member of the superfamily of ATP-binding cassette (ABC) transporters—in Pathway Commons. Left: SIF view of the neighborhood displayed in the molecule page of Pathway Commons. Right: Process diagram of the same neighborhood drawn by ChiBE (14), which can query PC using the web service interface. It contains a single event mediated by ABCC3—transportation of glycocholate from cytoplasm to extracellular region.

This format is useful for exporting pathway information to any ORA tool.

BioPAX OWL (RDF/XML)

BioPAX is the native format of Pathway Commons and offers complete access to all the details that can be stored in the system. This format is ideal for users wishing to import all pathway data for an organism into a local database, or to access specific data not available in other formats. Since BioPAX is defined using the standard OWL XML language, this export can be used with RDF/OWL tools such as reasoners or triplestores. All pathways and interactions within Pathway Commons are available in BioPAX Level 2 (5). Due to the richness of representation in BioPAX, reading and using such a large BioPAX document requires knowledge of the format and software development tools available for processing it, such as Paxtools, a Java library for working with BioPAX (http://www.biopax.org/paxtools.php).

Automate querying using the web service

The Pathway Commons web service allows software developers to access pathway data programmatically. Web services include a full-text search (supporting the same Lucene query syntax as the website, described above), searches to retrieve records using a Pathway Commons identifier, searches to retrieve pathways or neighbors of a molecule and searches that traverse the BioPAX data structure (e.g. get pathways that contain a molecule). Results are returned in the same formats available for bulk download (described above). The web service is currently implemented using a RESTful architecture, which means it is accessed by specifying a URL. Most web service commands were designed to be atomic rather than to address a complex research question in one iteration. Developers combine these calls to build up more complex queries, as exemplified in multiple projects including, the Pathway Commons Cytoscape plugin, ChiBE (14) and Paxtools (Figure 2). Complete web service documentation is available at Pathway Commons.

A community of pathway data providers

Pathway Commons does not compete with or duplicate efforts of pathway databases or software tool providers. Pathway Commons adds value to these existing efforts by providing a shared resource for publishing, distributing and querying pathway information. Existing database groups provide pathway curation and Pathway Commons provides a mechanism and the technology for sharing. An important aspect of Pathway Commons is clear author attribution. Curation teams at existing databases must be supported as much as possible by researchers to ensure they can keep performing their valuable work.

A major benefit of working together as a community to collect and integrate data, is that duplication of effort is reduced and best practices are more readily shared. Automatic data integration will never reach the quality level of rigorous data curation because expert decisions are needed to resolve conflicts. Thus, improved data integration must be a community effort to develop and implement best practices, such as consistent use of protein identifiers and pathway representation conventions. Pathway Commons is providing tools to help move towards this goal, including BioPAX, the BioPAX validator, the Paxtools Java library and the pathway data integration and cPath database software underlying the Pathway Commons website.

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010

IMPLEMENTATION

Pathway data integration

All data in Pathway Commons are provided by source databases, which curate or collect it. Pathway Commons regularly collects data from all source databases and integrates it. Pathway Commons relies on the use of standard formats, such as BioPAX (5) and PSI–MI (15), by databases for pathway data integration. Once a source database makes their data available in a standard format, it can be collected by Pathway Commons. PSI–MI data are first converted to the standard BioPAX format, which is the native language of Pathway Commons. The integration pipeline has three steps: aggregation and validation of data from the data sources, identifier normalization and merging. Aggregation is automated using scripts to download data from source databases. A validation process checks that the syntax of all input files is correct and that recognized identifiers are used, but in general assumes that data is semantically correct and has already been semantically validated using available PSI–MI or BioPAX validation web services by the source database. IDs are normalized by mapping those used by source databases to ones used by Pathway Commons. For instance, protein IDs are mapped to UniProt IDs and a full description of each protein is fetched from UniProt. Merging occurs at the level of physical entities, such as proteins and small molecules. This is accomplished by linking interaction and pathway records together if they use the same physical entities (recognized by shared use of normalized physical entity database identifiers, such as from UniProt for proteins). Pathway Commons does not currently merge records at the interaction or pathway level—for example, if two data sources describe the mTOR pathway, Pathway Commons does not create a single unified mTOR pathway, but rather makes available both original pathway versions.

Software

The Pathway Commons website is implemented using the freely available, open source (LGPL) cPath database software (16). Thus, users can install Pathway Commons locally. cPath is implemented using Java SE, MySQL, Apache Lucene and other popular open source software libraries. Current cPath development can be followed at http://code.google.com/p/pathway-commons/.

FUTURE DIRECTIONS

Recently, a new version of BioPAX (Level 3) was released, which adds support for gene regulation, genetic interactions and protein degradation and significantly improves support for signaling pathways. Data providers are now making data available in this format and Pathway Commons is being upgraded to support it. New web services will also be available, including full BioPAX property traversing, graph queries (17) and result auto-completion. These services will be more semantic web friendly, by using standard URIs wherever possible. We are also working to improve pathway integration. Pathway Commons currently merges small molecule and protein references from different pathway sources based on standard identifiers. This coarse grained integration can be improved by matching identical physical entities at the level of molecular state, considering features such as subcellular location and post translational modifications. Similarly, interactions and reactions can be matched based on their participants. Matching across data sources is still difficult, however, due to differences in level of detail used for representing physical entities and curation errors. As a first step, we are developing a fuzzy graph alignment tool that can match similar pathways. This will help discover inconsistencies within and conflicts between data sources, and report these to database curation teams at the source, catalyzing increasingly better pathway alignment and ultimately, development of a high quality map of all biological processes. For pathway and network visualization, we are also planning to replace or augment our static network neighborhood maps with an interactive network visualization using the Cytoscape Web software (http://cytoscapeweb.cytoscape.org) that we have developed (18).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the dedicated effort of curation teams that are creating and sharing biological pathway information.

FUNDING

National Institutes of Health, National Human Genome Research Institute (grant number P41HG004118); National Institute of General Medical Sciences (grant number 2R01GM070743-06); National Institute of General Medical Sciences Training Grant (grant number 1T32 GM083937); National Cancer Institute caBIG^{TM} program. Funding for open access charge: Institutional funding from Memorial Sloan-Kettering Cancer Center.

Conflict of interest statement. None declared.

REFERENCES

1. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
2. Cerami,E., Demir,E., Schultz,N., Taylor,B.S. and Sander,C. (2010) Automated network analysis identifies core pathways in glioblastoma. *PLoS One*, **5**, e8918.
3. Chuang,H.Y., Lee,E., Liu,Y.T., Lee,D. and Ideker,T. (2007) Network-based classification of breast cancer metastasis. *Mol. Sys. Biol.*, **3**, 140.
4. Bader,G.D., Cary,M.P. and Sander,C. (2006) Pathguide: a pathway resource list. *Nucleic Acids Res.*, **34**, D504–D506.
5. 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.
6. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010

7. Wu,C.H., Apweiler,R., Bairoch,A., Natale,D.A., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.
8. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.*, **25**, 25–29.
9. 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.
10. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
11. Goldberg,D.S. and Roth,F.P. (2003) Assessing experimentally derived interactions in a small world. *Proc. Natl Acad. Sci. USA*, **100**, 4372–4376.
12. Bader,J.S., Chaudhuri,A., Rothberg,J.M. and Chant,J. (2004) Gaining confidence in high-throughput protein interaction networks. *Nat. Biotechnol.*, **22**, 78–85.
13. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
14. Babur,O., Dogrusoz,U., Demir,E. and Sander,C. (2010) ChiBE: interactive visualization and manipulation of BioPAX pathway models. *Bioinformatics*, **26**, 429–431.
15. 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.
16. Cerami,E.G., Bader,G.D., Gross,B.E. and Sander,C. (2006) cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics*, **7**, 497.
17. Dogrusoz,U., Cetintas,A., Demir,E. and Babur,O. (2009) Algorithms for effective querying of compound graph-based pathway databases. *BMC Bioinformatics*, **10**, 376.
18. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
19. Le Novere,N., Hucka,M., Mi,H., Moodie,S., Schreiber,F., Sorokin,A., Demir,E., Wegner,K., Aladjem,M.I., Wimalaratne,S.M. *et al.* (2009) The systems biology graphical notation. *Nat. Biotechnol.*, **27**, 735–741.
20. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
21. 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.
22. Romero,P., Wagg,J., Green,M.L., Kaiser,D., Krummenacker,M. and Karp,P.D. (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.
23. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
24. Ceol,A., Chatr Aryamontri,A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.
25. Schaefer,C.F., Anthony,K., Krupa,S., Buchoff,J., Day,M., Hannay,T. and Buetow,K.H. (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.*, **37**, D674–D679.

Downloaded from nar.oxfordjournals.org at University of Toronto Library on November 19, 2010