**Pathway databases and tools for their exploitation: benefits, current limitations and challenges**

Anna Bauer-Mehren, Laura I Furlong* and Ferran Sanz

Research Unit on Biomedical Informatics (GRIB), IMIM-Hospital del Mar, Universitat Pompeu Fabra, Barcelona Biomedical Research Park, Dr Aiguader 88, Barcelona, Spain

* Corresponding author. Research Unit on Biomedical Informatics, Universitat Pompeu Fabra, IMIM-Hospital del Mar, PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain. Tel.: + 34 9331 60521; Fax: + 34 9331 60550; E-mail: lfurlong@imim.es

In past years, comprehensive representations of cell signalling pathways have been developed by manual curation from literature, which requires huge effort and would benefit from information stored in databases and from automatic retrieval and integration methods. Once a reconstruction of the network of interactions is achieved, analysis of its structural features and its dynamic behaviour can take place. Mathematical modelling techniques are used to simulate the complex behaviour of cell signalling networks, which ultimately sheds light on the mechanisms leading to complex diseases or helps in the identification of drug targets. A variety of databases containing information on cell signalling pathways have been developed in conjunction with methodologies to access and analyse the data. In principle, the scenario is prepared to make the most of this information for the analysis of the dynamics of signalling pathways. However, are the knowledge repositories of signalling pathways ready to realize the systems biology promise? In this article we aim to initiate this discussion and to provide some insights on this issue.

Molecular Systems Biology 5; Article number 290; doi:10.1038/msb.2009.47

**Subject Categories:** bioinformatics; signal transduction

**Keywords:** biological pathways; cell signalling; network models; pathway databases; systems biology

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. Creation of derivative works is permitted but the resulting work may be distributed only under the same or similar licence to this one. This licence does not permit commercial exploitation without specific permission.

**Introduction**

The past decades of research have led to a better understanding of the processes involved in cell signalling. Cell signalling refers to the biochemical processes using which cells respond to cues in their internal or external environment (Alberts et al, 2007). With the advent of high throughput experimentation, the identification and characterization of the molecular components involved in cell signalling became possible in a systematic way. In addition, the discovery of the connections between each of these components promoted the reconstruction of the chain of reactions, which subsequently gives rise to a signalling pathway. Ultimately, our ability to interpret the function and regulation of cell signalling pathways is crucial for understanding the ways in which cells respond to external cues and how they communicate with each other.

In this regard, the systematic collection of pathway information in the form of pathway databases and the application of mathematical analysis for pathway modelling are crucial. Several databases containing information on cell signalling pathways have been developed in conjunction with methodologies to access and analyse the data (Suderman and Hallett, 2007). Furthermore, mathematical modelling emerged as a solution to study the complex behaviour of networks (Alves et al, 2006; Fisher and Henzinger, 2007; Karlebach and Shamir, 2008). The models, so far obtained, allow formulating hypothesis that can be tested in the laboratory. Iterative cycles of prediction and experimental verification have resulted in the refinement of our knowledge of cell signalling, and have shed light on different aspects of cell signalling at a systems level (regulatory aspects, such as feedback control circuits or architectural features, such as modularity).

Furthermore, signalling cascades are not isolated units within the cell, but form part of a mesh of interconnected networks through which the signal elicited by an environmental cue can traverse (Yaffe, 2008). Ultimately, each cell is exposed to a variety of signalling cues, and the specificity of the response will be determined by the signalling mechanisms that are activated by the cue (Alberts et al, 2007). Recent research highlights the importance of the, so called, crosstalks between pathways, such as the recently published connections between signalling through the purinergic receptors and the Ca^{2+} sensing (Chaumont et al, 2008); the link between extracellular glycosalx structure and nitric oxide signalling pathway (Tarbell and Ebong, 2008); the interactions between insulin and epidermal growth factor signalling (Borisov et al, 2009) and the crosstalk between phosphoinositide 3 kinase and Ras/extracellular signal-regulated kinase signalling pathways (Wang et al, 2009).

An important goal of this research is to achieve a reconstruction of the network of interactions that gives rise to a signalling pathway in a biologically consistent and meaningful manner that in turn allows the mathematical analysis of the emerging properties of the network. In this regard, comprehensive maps of signalling pathways have been developed by manual curation from literature (Oda et al, 2005; Oda and Kitano, 2006; Calzone et al, 2008). Building such
reference maps requires huge effort and would benefit from information stored in databases and from automatic retrieval and integration methods. Once a reconstruction of the network of interactions is achieved, analysis of the structural features of the network and its dynamic behaviour can take place. A commonly seen architecture of signalling pathways is called ‘bow-tie’, in which many input and output signals are handled by a common layer constituted by a small number of conserved components. This network architecture provides robustness and flexibility to a variety of external cues due to the redundancy of reactions that are part of the input and output layers (Kitano, 2007a). Robustness refers to the ability of an organism to compensate the effects of perturbations to maintain the organism’s functions (Kitano, 2007b). Such perturbations can be changes in the availability of nutrients as well as the presence of mutagens or toxins. Moreover, systems can be subjected to functional disruptions when facing perturbations for which they are not optimized, thus showing points of fragility of the biological system (Kitano, 2007b). For instance, an undesirable effect of a drug can be caused by the unwanted interaction of the drug with molecules that represent points of fragility of the physiological system (Kitano, 2007a). In contrast, drugs can be completely ineffective when the robustness of the system compensates their action. It has been suggested that crosstalks between signalling pathways contribute to the robustness of cells against perturbations (Kitano, 2007a). In addition, the points of fragility of the system are sometimes exploited by pathogens causing diseases, or represent processes that are usually found to malfunction in particular diseases, such as cancer. Diseases that arise from dysfunction in cell signalling are usually not attributed to a single gene but to the failure of emerging control mechanisms in the network. It has been reported that the loss of negative feedback loops characterizes solid tumours (Amit et al, 2007). These diseases are difficult to diagnose and treat unless accurate understanding of the underlying principles regulating the system is in place. Thus, the interpretation of the global properties of signalling pathways has important implications for the elucidation of the mechanisms that lead to complex diseases, and also for the identification of drug targets.

At present, there are several repositories of information on cell signalling pathways that cover a wide range of signal transduction mechanisms and include high quality data in terms of annotation and cross references to biological databases. In principle, the scenario is prepared to make use of the information for the analysis of the behaviour of the signalling pathways. Thus, are the knowledge repositories on signalling pathways ready to realize the systems biology promise? In this article, we aim to initiate this discussion and to provide some insights on this issue.

First, we present an analytical overview of current pathway databases (see Pathway databases). In section ‘Case study: EGFR signalling’, we present the results of an evaluation exercise conducted to determine the accuracy and completeness of current pathway databases in front of an expert-curated pathway used as ‘gold standard’. Moreover, we propose a strategy for the use of pathway data from public databases for network modelling (Box 1; Table I). Finally, in the section ‘Conclusions and perspectives’ we discuss the strengths and limitations of the current pathway databases and their usefulness in practical biological problems and applications.

Pathway databases

Pathway databases serve as repositories of current knowledge on cell signalling. They present pathways in a graphical format comparable to the representation in text books, as well as in standard formats allowing exchange between different software platforms and further processing by network analysis, visualization and modelling tools. At present, there exist a vast variety of databases containing biochemical reactions, such as signalling pathways or protein–protein interactions. The Pathguide resource serves as a good overview of current pathway databases (Bader et al, 2006). More than 200 pathway repositories are listed, from which over 60 are specialized on reactions in human. However, only half of them provide pathways and reactions in computer-readable formats needed for automatic retrieval and processing, and even less support standard formats, such as Biological Pathway Exchange (BioPAX) (http://www.biopax.org) and Systems Biology Markup Language (SBML) (Hucka et al, 2003).

To obtain a complete view of the biological process of interest, combination of information from diverse reactions and pathways is often needed. A recent publication (Adriaens et al, 2008), describes a workflow developed for gathering and curating all information on a pathway to obtain a broad and correct representation. However, the described process heavily relies on manual intervention. Consequently, there is a need for the automation of both the pathway retrieval process and the integration of different data sources. This section is devoted to the description of main pathway databases: Reactome, Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways, Nature Pathway Interaction Database (PID) and Pathway Commons. Table II lists all pathway databases and protein–protein interaction resources that are mentioned in this section.

Reactome

Reactome is currently one of the most complete and best-curated pathway databases. It covers reactions for any type of biological process and organizes them in a hierarchical manner. In this hierarchy, the lower level corresponds to single reactions, whereas the upper level represents the pathway as a whole.

Reactome was first developed as an open source database for pathways and interactions in human. Equivalent reactions for other species are inferred from the human data (Vastrik et al, 2007), providing coverage to 22 non-human species, including mouse, rat, chicken, puffer fish, worm, fly, yeast, and Escherichia coli. Furthermore, other Reactome projects exist focusing on single species, such as the Arabidopsis Reactome (http://www.arabidopsisreactome.org).

All pathway and reaction data in Reactome are extracted from biomedical experiments and literature. For this purpose, PhD-level biologists are invited to work together with the Reactome curators and editors on the curation of data on selected biological processes. Once the first outline of the
Pathway databases and the systems biology promise
A Bauer-Mehren et al

Box 1  Use of data from public pathway databases for modelling purposes

Most public, available pathway databases offer their data in BioPAX format, which was developed for detailed pathway representation and as data exchange format. For storing and sharing of computational models of biological networks, SBML has emerged as standard and is supported by most modelling software. BioPAX and SBML, the two main standards for the representation of biological networks, have been discussed in detail by others (Stromback and Lambrin, 2005; Stromback et al, 2006). In Table I, we briefly list the most important features of the SBML and BioPAX standards. A scenario in which pathway data were directly used for network modelling is proposed here. One or more pathways represented in BioPAX format are automatically retrieved from different databases and imported into a pathway visualization and analysis tool. Then, integration of the different pathways can take place to obtain a comprehensive and biologically meaningful representation of the network. In addition, annotations can be added if required or structural analysis of the network can be carried out. The resulting network, which integrates the original pathways retrieved from the databases, is exported to SBML format and subjected to modelling. If a quantitative approach is chosen, additional information, such as rate constants are required to start the modelling process. In this process, conversion between the two formats is required to achieve inter-operability between pathway and model representations. Some solutions are already available. The BioModels (http://www.ebi.ac.uk/biocompmodels-main/) database, which contains a variety of curated models in SBML format, offers conversion to BioPAX format. The opposite conversion, from BioPAX to SBML, would open the possibility of modelling the pathways stored in public databases. However, the inter-conversion between BioPAX and SBML is not trivial as both formats were developed for different purposes. BioPAX, for instance, does not offer the possibility to store quantitative information needed for kinetic modelling, whereas SBML does not represent relationships between nodes that are not needed for modelling and that are present in BioPAX. Examples of approaches for the conversion from BioPAX to SBML are Bio2SBML (Zinovyev et al, 2008), which is available as Cytoscape plugin, and SyBil, which is part of the model environment for quantitative modelling VCell (Evelo, 2009). Although compatibility of different pathway and network model exchange formats is still not completely achieved, the efforts made towards this goal represent significant contributions to pathway retrieval, integration and subsequent modelling.

Biological process is created and annotated, it is inspected by peer reviewers and potential inconsistencies and errors are fixed. Every two years the data are reviewed to keep it updated (Joshi-Tope et al, 2005; Matthews et al, 2009). Moreover, cross references to different databases, such as UniProt (The UniProt Consortium, 2008), Ensembl (http://www.ensembl.org/index.html), NCBI (http://www.ncbi.nlm.nih.gov), Gene Ontology (GO) (Ashburner et al, 2000), Entrez Gene (Maglott et al, 2007), UCSC Genome Browser (http://genome.ucsc.edu), HapMap (http://www.hapmap.org), PubMed, as well as to other pathway databases, such as KEGG (Kanehisa and Goto, 2000) are provided.

Pathways are presented as chains of chemical reactions and the same data model is used to describe reactions for any biological process, such as transcription, catalysis or binding (Matthews et al, 2007). Altogether, this represents a coherent view of pathway knowledge. The data model is based on classes, such as physical entity or event. Physical entities comprise proteins, DNA, RNA, small molecules but also complexes of single entities. Proteins, RNA and DNA, for which the sequence is known, are linked to the appropriate databases. Chemical entities such as small molecules are linked to ChEBI (http://www.ebi.ac.uk/chebi/init.do). An event can be either a ReactionLikeEvent, which represents reactions that convert an input into an output, or a PathwayLikeEvent, grouping together several related events. Each class possesses properties, such as information on the type of interaction (e.g. inhibition or activation). Reactome explicitly considers the different states an entity can show in a reaction. The phosphorylated and the unphosphorylated version of a protein are, for example, represented as separate entities. In addition, generalization is allowed. This means that if two different entities have exactly the same function in a reaction, such as isoenzymes, the reaction is only described once and the functional equivalent entities belong to the same defined set. Another interesting element of the Reactome data model is the use of candidate sets, which act as placeholders for all possible entities in a reaction, in case the exact entity involved in the reaction is not yet known.

Reactome can either be directly browsed or queried by text search using, for instance, UniProt accession numbers. In addition, some tools for advanced queries are provided. The PathFinder tool allows connecting an input to an output molecule or event by constructing the shortest path between both. The SkyPointer tool can be used to identify events or pathways that are statistically over-represented for a list of genes or proteins. Moreover, Reactome data can be combined
Pathway databases and the systems biology promise
A Bauer-Mehren et al

with other databases such as UniProt, by using the Reactome BioMart (http://www.biomart.org) tool.

In addition to browsing pathways through the Reactome web interface, it is possible to download the data for local visualization and analysis using other tools. Different formats are provided for pathway download, including SBML Level 2, and BioPAX Level 2 and Level 3 (for some reactions only), as well as graphical formats. Pathway files, for instance, in BioPAX format can be directly opened in Cytoscape (Shannon et al., 2003), a software for the visualization and analysis of networks. Moreover, data can be programmatically accessed through a SOAP web service.

KEGG

KEGG is not only a database for pathways but consists of 19 highly interconnected databases, containing genomic, chemical and phenotypic information (Kanehisa and Goto, 2000; Kanehisa et al., 2008). Here we concentrate on the database storing biological pathways. KEGG categorizes its pathways into metabolic processes, genetic information processing, environmental information processing, including signalling pathways, cellular processes, information on human diseases and drug development. However, the best-organized and most complete information can be found for metabolic pathways.

KEGG is not organism specific but covers a wide range of organisms, including human. The pathways are manually curated by experts using literature. In addition to the interconnection of all databases underlying KEGG, links to external databases, such as NCBI Entrez Gene, OMIM, UniProt and GO are provided. Pathways can either be browsed or queried by free text search. The user can search for gene names, chemical compounds or whole pathways. A tutorial on how to browse pathways in KEGG and an overview of the multiple representation formats is available (Aoki-Kinoshita & Minoru Kanehisa, 2007).

Each pathway stored in KEGG can be downloaded in its own XML format named KGML, which is supported by VisANT, a software tool for pathway visualization (Hu et al., 2008b) and indirectly by Cytoscape using scripting plugins. In addition, metabolic pathways are available in BioPAX Level 1, which was especially designed for metabolic reactions, as well as in SBML. For converting KEGG metabolic pathways to SBML, a tool called KEGG2SBML (http://sbml.org/Software/KEGG2SBML) was developed.

KEGG data can also be accessed using the KEGG API or KEGG FTP. Moreover, for making use of the KEGG resources, several applications exist. KegArray, for example, allows the analysis of microarray data in the context of KEGG pathways.

Table 1 Comparison between SBML and BioPAX

| SBML                                                                 | BioPAX                                                                 |
|----------------------------------------------------------------------|------------------------------------------------------------------------|
| **Representation format**                                           | OWL (Web Ontology Language). XML Pathway description with all details on reactions, components, information on cellular location etc. |
| **Main purpose**                                                    | Basic ontology based on three classes                                   |
| **Entities and reactions**                                          | Pathway (set of interactions)                                          |
| Based on species and reactions (Hucka et al., 2003):                | Physical entity with subclasses, such as RNA, DNA, protein, complex and small molecules |
| Species (proteins, small molecules etc.)                            | Interaction with subclasses, such as conversion having biochemicalReaction as subclass, etc. |
| Reactions (how species interact)                                    | Several pathways per BioPAX file possible (each object has its own RDF id and is hence uniquely identifiable) |
| Compartment (in which interactions take place)                      |                                                                         |
| **Number of pathways represented**                                  |                                                                         |
| One model per SBML file                                             |                                                                         |
| **Reaction kinetics**                                               | No kinetics as BioPAX is not meant for modelling but pathway representation |
| Levels                                                              | BioPAX Level 1: representation of chemical reactions involved in metabolism |
| Built in levels with different versions. Each level adds new features, such as the incorporation of controlled vocabularies. At the time of writing, the most stable version is SBML Level 2 | BioPAX Level 2: adds molecular interactions and protein post-translational modifications |
|                                                                         | BioPAX Level 3: any kind of biological reaction, including regulation of gene expression (BioPAX L3 is at the time of writing still in release process) |
| **Pathway database support**                                        | The BioPAX project roadmap envisages two additional levels capturing interactions at the cellular level. (http://www.biopax.org/Docs/BioPAX_Roadmap.html) |
| Reactome                                                            |                                                                         |
| KEGG                                                                |                                                                         |
| **Model database support**                                          |                                                                         |
| BioModels                                                           |                                                                         |
| **Library for reading/writing**                                     |                                                                         |
| libSBML (Bornstein et al., 2008)                                    | PathwayCommons                                                         |
| Standard modelling software, such as CellDesigner or Copasi (Hoops et al., 2006) | Paxtools (http://www.biopax.org/paxtools/) |
| Network visualization software, such as Cytoscape                   |                                                                         |

Network visualization software, such as Cytoscape or VisANT
WikiPathways

A recently developed resource for pathway information that strongly differs from other pathway repositories is WikiPathways. WikiPathways is an open source project based, like Wikipedia, on the MediaWiki software (Pico et al., 2008). It serves as an open and collaborative platform for creation, edition and curation of biological pathways in different species.

WikiPathways aims to achieve a public commitment to pathway storage and curation by keeping pathway creation and curation processes simple. Although the curation process of the previously described databases is subjected to experts, any user with an account on WikiPathways can create new pathways, and edit already existing ones.

The pathway entities are linked to reference databases, based on the criteria provided by the editor. Hence, the identifiers depend on the chosen reference database and can therefore differ between pathways and even within a single pathway.

Pathways in WikiPathways can be browsed by species and categories, for example, Metabolic Process. They can also be searched using gene, protein or pathway name or any free text query. In addition, pathways can be programatically accessed through a web service (http://www.wikipathways.org/index.php/Help:WikiPathways_Webservice).

For pathway data exchange, WikiPathways does not use standard formats like BioPAX or SBML, but offers a much simpler representation called GenMAPP Pathway Markup Language (GPML) that is compatible with visualization and analysis tools, such as Cytoscape, GenMAPP (Salomonis et al., 2007) and PathVisio (van Iersel et al., 2008). The use of GPML is in agreement with the community annotation nature of the project, as it offers a simple pathway representation and several functionalities for building network diagrams. However, inter-operability with other pathway databases is impeded, and substantial efforts towards combining WikiPathways with the other pathway repositories will be required. In this regard, some approaches with the objective of conversion between GPML and standard pathway exchange formats, such as SBML and BioPAX, are under development (Evelo, 2009). In addition, KEGG pathways in KGML format are also available in GPML format ready for download (http://www.pathvisio.org/Download#Step_3) or can be converted into GPML (http://www.bigcat.unimaas.nl/tracprojects/pathvisio/wiki/KeggConverter).

The exponential growth of biological data poses a challenge to the high-quality annotation and curation of databases. In this scenario, the use of wikis for community curation of biological data have emerged in the past years with the goal of increasing quality of data annotation by combining knowledge from multiple experts (Giles, 2007; Waldrop, 2008; Hu et al., 2008a). However, their success will strongly depend on the commitment of the community and WikiPathways authors claim that the initiative represents an experiment, in which the ‘community curation’ approach is being tested (Pico et al., 2008). Thus, WikiPathways can be seen as a complementary tool for the high-quality annotation and curation of databases.

Table II  Online pathway and protein–protein interaction (PPI) databases

| Pathway/PPI database                  | Web link                                      | Standard exchange formats for download                  | Web service API                                      |
|--------------------------------------|-----------------------------------------------|--------------------------------------------------------|-----------------------------------------------------|
| Reactome                             | http://www.reactome.org                       | BioPAX Level 2                                         | SOAP web service API                                  |
|                                      |                                               | BioPAX Level 3 (only some reactions)                    | Detailed user manual available, example client in Java |
|                                      |                                               | SBML Level 2                                           |                                                     |
|                                      |                                               | KGML (default format)                                  |                                                     |
| KEGG                                 | http://www.genome.jp/kegg/pathway.html        | BioPAX Level 1 (only metabolic reactions)               | Example client in Java, Ruby, Perl                    |
|                                      |                                               | SBML (using converter)                                 |                                                     |
|                                      |                                               | GPML (default format)                                  |                                                     |
| WikiPathways                         | http://www.wikipathways.org                   | GPML (default format)                                  | SOAP web service API                                  |
|                                      |                                               | Converters to standards, such as SBML and BioPAX are   |                                                     |
|                                      |                                               | in progress                                            |                                                     |
| NCI/Nature Pathway Interaction Database (PID) | http://pid.nci.nih.gov                      | PID XML (default format)                               | Access through Pathway Commons                        |
| BioCarta                              | http://www.biocarta.com                      | BioPAX Level 2                                         |                                                     |
| Pathway commons                      | http://www.pathwaycommons.org                | BioPAX Level 2 through NCI/ Nature Pathway Interaction Database (PID) | HTTP URL-based XML web service through cPath         |
|                                      |                                               | BioPAX Level 2 (default format for pathways)           | Direct import into Cytoscape                          |
|                                      |                                               | PSI-MI (default format for protein–protein interactions) |                                                     |
| Cancer cell map                      | http://cancer.cellmap.org                     | BioPAX Level 2                                         | HTTP URL-based XML web service via cPath              |
| HumanCyc                             | http://humanycyc.org                         | BioPAX Level 2                                         | Access through Pathway Commons and Pathway Tools (Karp et al., 2002) |
| IntAct                               | www.ebi.ac.uk/intact/                        | BioPAX Level 3                                         | Access through Pathway Commons                        |
| HPRD                                 | http://www.hprd.org                          | PSI-MI                                                 | Access through Pathway Commons                        |
| MINT                                 | http://mint.bio.uniroma2.it/mint/             | PSI-MI                                                 | Access through Pathway Commons                        |
and enhancing source of information for the major pathway
databases, like Reactome or KEGG.

In contrast to the aforementioned databases, the systems
described below combine diverse pathway repositories, and can be
seen as first attempts towards the integration of
pathway information from various sources.

**Nature pathway interaction database**

PID contains data on cell signalling in humans (Schaefer et al.,
2009). PID combines three different sources: the NCI-curated
pathways that are obtained from peer reviewed literature, as well as pathways imported from Reactome and BioCarta.
Similar to Reactome, PID structures pathways hierarchically
into pathways and their sub-pathways that are called sub-
networks in PID.

The PID data model is based on molecular interactions in
which input biomolecules are transformed into output
biomolecules. Each process can be promoted or inhibited by
regulators. Biomolecules are proteins, RNA, complexes or
small molecules. DNA is not a part of the PID data model and
only output RNA and regulator are represented in transcriptional
processes. Each protein is cross-referenced to UniProt,
RNA to Entrez Gene, small molecules to the Chemical
Abstracts Service (CAS) registry number and complexes are
annotated using GO terms. Different states of biomolecules,
such as ‘active/inactive’ or ‘phosphorylated’ are part of the
annotations of the biomolecule. Cellular location, biological
processes and molecular function of the entities are cross-
linked to GO. Moreover, interactions are annotated with the
supporting literature or other evidence, such as *inferred from
array experiment* (Schaefer et al., 2009).

Pathways can be browsed and queried using gene or protein
identifiers, such as Entrez Gene identifier, UniProt accession
numbers or HUGO gene symbols, as well as biological process
terms from GO, among others. The system returns available
results from each of the three sources. Moreover, PID offers
advanced queries. The *connected molecules search* option
allows finding a possible path between two or more molecules.
In the *batch query*, the user can upload lists of gene or protein
identifiers and obtain a list of pathways ranked by the
probability of including the entities of the query list. Using
this application, pathways over-represented in a set of genes,
for example, derived from microarray expression experiments,
can be obtained.

PID provides different pathway representation formats,
including BioPAX Level 2 and a PID proprietary XML format.
Data from Reactome are directly imported using the BioPAX
Level 2 format and is regularly updated. As not all entities or
events stored in Reactome can be presented in BioPAX Level 2,
some information is lost during the import. However, this
might be avoided once BioPAX Level 3 is released. The
BioCarta data are manually assigned to the PID data model, as
BioCarta only offers a graphical cartoon view of the pathways
and does not provide computer readable download format.

**Pathway commons**

Pathway Commons is a compilation of the public pathway
databases Reactome, PID and Cancer Cell Map as well as
protein–protein interaction databases, such as HPRD (Mishra
et al., 2006), HumanCyc, IntAct (Kerrien et al., 2007a) and
MINT (Zanoni et al., 2002). Herein, the pathway hierarchies of
Reactome and PID are conserved.

Pathway Commons serves as an access point for a collection
of public databases and provides technology for integrating
pathway information. Pathway creation, extension and cura-
tion remain the duty of the source pathway databases. As a
consequence, entries in Pathway Commons are cross-linked to
their source database, and links to external databases rely on
the source database.

A regular search is provided and a filter can be set for
restricting the results to source and organism. Furthermore,
Pathway Commons provides a web service API for an
automatic access of the data. In addition, cPath, a Java open-
source software for aggregating, storing and querying pathway
data, is offered. One of its key features is the identifier mapping
system. It handles mapping tables of equivalent entities, such as
the UniProt and the RefSeq accession number of proteins.
These tables can in principle be used to integrate data from
diverse sources that use different identifiers. Moreover, the
system can straightforwardly be extended including self-
created mapping tables. PSI-MI (Kerrien et al., 2007b) and
currently BioPAX Level 2 exchange format are supported.
Furthermore, the complete Pathway Commons database can
be automatically accessed using the Pathway Commons plugin
in Cytoscape.

The systems presented above allow the access to a wide
range of data on biological pathways. However, there is
overlap in the information offered by different databases. In
contrast, for specific pathways some databases offer more
accurate and complete information than others. Hence, the
user might have difficulties in choosing the right database and
in dealing with redundancies and inconsistencies among the
pathways. The integration initiatives exemplified by PID and
Pathway Commons are attempts to solve these problems.
However, the intended integration is not trivial as the data are
fragmented and stored in databases that may differ in the
representation of the biochemical reactions, as well as in the
coverage and accuracy of annotations. In addition, often data
are not provided in interchangeable formats hampering the
automatic integration.

**Case study: EGFR signalling**

The epidermal growth factor receptor (EGFR) signalling
cascade is one of the best-studied and most important
signalling pathways in mammals. It regulates cell growth,
survival, proliferation and differentiation. Recently, a detailed
and comprehensive map of the EGFR signalling pathway has
been reported (Oda et al., 2005). As the map was built manually
by experts using the literature, it can be seen as a reference
representation of the pathway. Other reference maps of
important signalling pathways have been reported previously
(Oda and Kitano, 2006; Calzone et al., 2008; Herrgard et al.,
2008), providing the scientific community with comprehen-
sive maps that can be used for modelling, which in turn will
shed light on important aspects of cell signalling. However,
these initiatives constitute huge efforts and, as judged by the
limited number of already available maps, there is a lag between the amount of data available in public databases and the availability of such references map. Hence, we argue that public pathway databases could be used to build such reference maps of signalling pathways. Most pathway databases are also developed by experts in the field and constitute repositories of high-quality data, with the additional advantage of being already represented in machine readable formats that could, in principle, be easily and automatically retrieved, analysed and fed into modelling software tools.

We selected the EGFR pathway (Oda et al., 2005), hereafter referred to as EGFR map, as a ‘gold standard’ to evaluate the completeness and accuracy of public pathway databases in the representation of the reactions that are part of the EGFR signalling (Figure 1). We based our selection on the following reasons: (i) signalling through EGFR has been studied for more than 40 years and a lot of information about the reactions is already available (Citri and Yarden, 2006); (ii) it has been carefully curated by experts; (iii) it constitutes an excellent example of crosstalk between different signalling events, thus allowing evaluation of the coverage of crosstalks in the public databases and the ability of network analysis tools to retrieve and combine networks in a meaningful manner; (iv) the study of signalling through EGFR has important implications for understanding several cancer types and the development of new therapeutic strategies. Several computational models have been reported on different aspects of EGFR signalling (Khodolenko et al., 1999; Schoeberl et al., 2002; Hornberg et al., 2005; Birtwistle et al., 2007; Borisov et al., 2009; Li et al., 2009). However, it is worth mentioning that, to the best of our knowledge, no model for the whole EGFR map has been reported till now.

In the following paragraphs and in Figures 1–3, we use the same notation of entities as in the SBML file of the EGFR map (Oda et al., 2005). The EGFR map is based on more than 240 publications and contains several crosstalks between the EGFR downstream signalling and other signalling pathways. In the EGFR map, depicted in Figure 1, entities are clustered according to their cellular location and function. The functional units comprise receptor endocytosis, recycling and degradation, small GTPase signalling, MAPK cascade, PIP signalling, cell cycle, Ca\(^{2+}\) signalling and GPCR-mediated transactivation. Seven phenotypic outcomes of EGFR signalling are depicted: ErbB endocytosis, ErbB degradation, apoptosis, actin reorganization, cell cycle, gene transcription and mitogenesis/tumorigenesis.

To address the completeness and accuracy of pathway information available in public databases and its automatic retrieval, we tried to recover the complete EGFR map. For this purpose we queried Reactome version 26 with the term ‘EGFR’ and its UniProt identifier ‘P00533’ and downloaded and visualized the retrieved pathways. Reactome was chosen as it is currently the most detailed pathway repository, and utilizes a data model that accommodates different types of biochemical reactions. For visualization, we chose Cytoscape because of its user-friendly visualization capabilities and its network analysis tools. To map the entities found in Reactome to those in the EGFR map, a mapping through standard identifiers was carried out. We compared the original EGFR map with the EGFR pathway recovered from Reactome (in BioPAX format) and coloured entities and reactions according to their representation in both resources (see Figure 1). Red entities are found to be identical in the EGFR pathway in Reactome and the EGFR map. Purple connotes entities that could be recovered from Reactome but that are differently represented (for instance, if a single protein instead of a complex is described to take part in a reaction).

Only a small proportion of the original EGFR reactions could be recovered from Reactome, and most of them are directly related to signals coming from downstream EGFR signalling. Most of the reactions related to other signalling cascades are not connected with EGFR signalling in Reactome and could therefore not be recovered. Regarding the associated phenotypes, only two were found in Reactome: EGFR endocytosis and EGFR degradation. However, both mechanisms are described in a slightly different manner, in some cases even with more details than in the original EGFR map. In a second step, we tried to extend the EGFR map by querying Reactome with key entities found in the EGFR map to complete signalling cascades, such as the GPCR signalling or the MAPK cascade that are missing in the EGFR pathway in Reactome. All pathways that were added are listed in Table III. We used additional colours to depict the entities that were recovered in this extension process. Green was used for entities found in Reactome and the EGFR map, and turquoise was used for entities that differed in their representation in both sources (the coloured EGFR map is available in XML and pdf formats as Supplementary information). By this extension, we were able to recover four of the five missing phenotypes: actin reorganization, apoptosis, cell cycle and the transcription of target genes. However, some reactions were still missing in the information recovered from Reactome and in some cases gaps or contradictions appear impeding an automatic integration (Figure 2). In this example, reactions in which ERK1 and ERK2 participate are first separately described and later the representation switches towards a combined ERK1/2 entity.

Regarding the reactions that give rise to regulatory loops in the EGFR map, only some of them could be recovered from Reactome. For instance, although the reaction that involves cleavage of pro-HB-EGF by ADAMs is described, its regulation by Pyk2 and c-Src is not included and therefore this positive feedback loop is not coloured in the EGFR map. In total, three of the six negative feedback loops were detected: inhibition of EGFR by SHP1, downregulation of EGFR and phosphorylation of SOS1 by ERK1, which leads to SOS1 inhibition.

Although most of the crosstalks between signalling cascades in the EGFR signalling could be established by the extension process, a significant number were not found because the entities that link the different cascades are missing in Reactome. For example, the important crosstalk of the Ca\(^{2+}\) and the EGFR signalling by the effect of Ca\(^{2+}\) on Pyk2 activity could not be recovered, as Pyk2 is not present in Reactome. Moreover, it is worth mentioning that details about some of the reactions differ between the EGFR map and the data found in Reactome. In part, this can be explained by the fact the former is based on literature curated in 2005, and version 26 of Reactome was released in October, 2008.

The extension process was achieved by searching the database with entities representing the main signalling cascades that are known to be connected with the EGFR
signalling, followed by manual identification of the reactions that connect the pathways. In principle, the process of finding the connections or crosstalks between pathways could be automated using tools available in Cytoscape or Pathway Commons (cPath). The Cytoscape merging function was evaluated for this purpose. This function compares the

---

**Figure 1** **EGFR map.** EGFR map created using CellDesigner ver.2.0 (Oda et al., 2005). This map has been coloured to show the entities and reactions that are in common between the **EGFR map** and information found in Reactome. Red colour denotes that the entities and reactions are equivalent, purple connotes that they are similar but differ in some description details, and white is used for entities and reactions that could not be directly found in Reactome. In a second step, the map was extended by querying Reactome with key entities appearing in the **EGFR map**, which were missing in the representation of the EGFR pathway in Reactome. After this extension process, we coloured new equivalent entities in green and new similar entities (the ones that are differently described in both resources) in turquoise. For comparing the **EGFR map** with the pathways downloaded from Reactome, the Reactome pathways have been imported into Cytoscape and the entities and reactions have been manually compared using the node and edge search functions of Cytoscape. As the SBML version of the **EGFR map** does not contain unique identifiers for the nodes (species), all names have been first matched to Entrez Gene identifiers, which have then been used for comparing entities between the **EGFR map** and Reactome. In the extension process canonical names as well as Entrez Gene and UniProt identifiers have been used to retrieve all the information available in Reactome. In some cases, when no results were obtained in this way, the search was additionally expanded. For example, to find the EGFR crosstalk with the GPCR signalling, we additionally searched for ‘G protein’. GPCR signalling pathways activated by S1P1, S1P2/3, LPA1, LPA2, EP3 and EP2/4 were not found in Reactome. Instead, GPCR signalling through thrombin and glucagon receptors that are related to EGFR signalling (Prenzel et al., 1999; Buteau et al., 2003) are present in Reactome and were incorporated into the **EGFR map** to complete the missing crosstalk.

---

**Figure 2** Comparison of ERK signalling as found in the **EGFR map** and in Reactome. In Reactome, ERK1 or ERK2 are phosphorylated by MKK1 or MKK2, respectively. The same reaction is found in the **EGFR map**, but here ERK1 and ERK2 are represented as a single entity, namely ERK1/2, which combines both proteins and can be phosphorylated by MKK1 and MKK2. In Reactome, dimerization and translocation to the nucleus are still separately described for each entity. Then, the representation switches from separate entities to one combined entity. In addition, the dissociation of MKK1 or MKK2, which is needed before the dimerization of ERK can take place, is not described in Reactome. Manual intervention is needed to correctly map both representations.
attributes of the nodes to automatically connect reactions from different pathways. However, when tested on the reactions in the EGFR map, several problems arose. Most of them appeared as a result of annotation issues. For instance, Figure 3 shows two reactions in which the two IP3 entities are differently annotated. The first IP3 entity is located in the ‘cytosol’, whereas this cellular location annotation is missing for the second IP3, precluding the expected merging of the two reactions. Hence, an automatic integration is impeded and manual intervention is needed. Another factor that hampers finding connections between reactions or pathways is the use of combined entities. For instance, the already mentioned ERK1 and ERK2 proteins first represented as separate entities are later described as a combined ERK1/ERK2 entity (Figure 2). This problem could be solved by considering all the annotations of the nodes while deciding whether two entities are equivalent or not. This would allow comparing nodes that represent states of entities, for instance, post-translational modified proteins or proteins annotated using cellular locations. In summary, reconstruction of crosstalks between signalling pathways is difficult by means of the automatic tools currently available. Manual intervention is required to recover all reactions involved in the pathways and their crosstalks.

The case study presented here shows that a process combining automatic retrieval and manual intervention can be used to reconstruct the EGFR map in its main features. This shows that current pathway databases contain a lot of detailed information though in some cases this information is still incomplete and manual intervention is needed to obtain a complete and correct network representation containing different signalling pathways. This is especially critical for reactions that are part of regulatory feedback loops, as these determine the dynamic behaviour of the signalling pathway. Nevertheless, information obtained for individual reactions and even for some pathways is quite complete and would be accurate enough as a starting point for model building. A proposal for a strategy for the use of pathway data from public databases for network modelling is presented in Box 1.
Conclusions and perspectives

In this paper, we have reviewed the main knowledge resources of human pathways, and we have evaluated the feasibility of using this information for the reconstruction of signalling pathways in a biologically meaningful manner. Moreover, we have presented a scenario in which data from public pathway databases are directly used for modelling (see Box 1). In this regard, we have briefly discussed the main standards for representation of biological networks, BioPAX and SBML. Furthermore, we have discussed the advantages and drawbacks of current methods for pathway retrieval and integration, using the EGFR signalling as an illustrative example.

We encourage the combination of data from different pathway databases, as they are often complementary and, in this way, a better coverage of all the reactions involved in a given pathway will be achieved. However, the integration of pathways from different databases poses a challenge, as different standard formats are used and data models vary.

Even if we choose a single database as a source of pathway information, the retrieval of a signalling pathway in conjunction with its crosstalks to other signalling cascades is a difficult task. Although in this case the data model and representation formats are not an issue, there are still annotation problems that remain to be solved to allow an effective integration of reactions and pathways. In this regard, the communication of these problems to database curators by the users will be of great help to improve the completeness and quality of annotations.

There is a strong need of tools for the automatic integration of different pathways in a biological meaningful way. The analysis presented here stresses that this is not a trivial task. As several annotation problems and inconsistencies exist, manual intervention is needed to achieve the integration. Moreover, other factors have to be considered to decide whether two pathways can be merged: are the pathways found in the same cell type? Or, are they found at the same developmental stage of the cell? Accurate annotation of the reactions taking place in each signalling pathway will be required to appropriately solve these questions.

The case study on the EGFR signalling has highlighted very important issues for the practical use of pathway databases. The information obtained for individual reactions and even for particular pathways is quite complete in most of the cases and would be accurate enough as a starting point for modelling. Although we did not carry out a systematic evaluation of all the reactions found in Reactome, on the basis of the results of this case study we can conclude that public databases contain accurate and quite complete information about the main processes involved in cell signalling pathways. However, for processes for which no such level of detail on the reactions is available, the representation recovered from public databases will be less complete. For example, comparison of manually created Rb/E2F pathway with data from Reactome indicated that the latter does not cover all the reactions (Calzone et al., 2008). We foresee that in the following years, the coverage of the databases will grow as well as the quality of the annotations, which will benefit the scientific community in providing a source of representations of pathways for modelling purposes.

We would like to finish by stressing the importance of the annotation and data representation issues for an effective integration of data from public pathway databases. Researchers involved in pathway annotation and in pathway modelling should engage in collaborative projects to take advantage of the data already available in public databases and work together on representations that fit the needs of both communities.

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements

This work was generated in the framework of the @neurIST and the EU-ADR projects co-financed by the European Commission through the contracts nos. IST-027703 and ICT-215847, respectively. The Research Unit on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB) (www.inab.org). It is also member of the COMBIOMED network. We thank the Departament d’Innovacio, Universitat i Empresa (Generalitat de Catalunya) for a grant to ABM.

Conflict of interest

The authors declare that they have no conflict of interest.
References

Adriaens ME, Jaillard M, Waagmeester A, Coort SLM, Pico AR, Evelo CTA (2008) The public road to high-quality curated biological pathways. Drug Discov Today 13: 856–862

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2007) Molecular Biology of the Cell. New York, USA: Garland Science

Alves R, Antunes F, Salvador A (2006) Tools for kinetic modeling of biochemical networks. Nat Biotechnol 24: 667–672

Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Taricic G, Siwik D, Lahad J, Jacob-Hirsch J, Amatiagi N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y (2007) A module of negative feedback regulators defines growth factor signaling. Nat Genet 39: 503–512

Aoki-Kinoshita K, Kanehisa M (2007) KEGG Primer: An introduction to pathway analysis using KEGG. doi:10.1038/pid.2007.2. NCI-Nature Pathway Interaction Database

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25: 25–29

Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. Nucleic Acids Res 34: D904–D906

Birwistle MR, Hatakeyama M, Yumoto N, Ogumaike BA, Hoek JB, Khodolenko BN (2007) Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. Mol Syst Biol 3: 144

Borisyov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhourt J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, Khodolenko BN (2009) Systems-level interactions between insulin-EGRF networks enable mitogenetic signaling. Mol Syst Biol 5: 256

Bornstein BJ, Keating SM, Jouhrak A, Huchka M (2008) LibSBML: An API library for SBML. Bioinformatics 24: 880–881

Buteau J, Foisy S, Joly E, Prentki M (2003) Glucagon-like peptide 1 induces pancreatic beta-cell proliferation via transactivation of the epidermal growth factor receptor. Diabetes 52: 124–132

Calzone L, Gelay Al, Zinovyev A, Radvanyi F, Barillot E (2008) Molecular Biology of the Cell. New York, USA: Garland Science

Altman S, Antunes V, Toulme E, Richler E, Housley GD, Rassendren H, Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Fisher J, Henzinger TA (2007) Executable cell biology. Bioinformatics 19: 524–531

Joshi-Tope G, Gillespie M, Vastrik I, D’Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L (2005) Reactome: a knowledgebase of biological pathways. Nucleic Acids Res 33: D248–D243

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tomimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36: D480–D484

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27–30

Kanehisa M, Goto S (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25: 25–29

Karp PD, Paley S, Romero P (2002) The Pathway Tools software. Bioinformatics 18: S225–S232

Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn A, Vingan S, Schild H,订y E, Alagona P, Moreau Y, Saillant G, Jensen LJ, Bork P, Deragon M, Veltman E, Helmes C, Hermjakob H, Apweiler R (2007) The IntAct molecular interaction database. Nucleic Acids Res 35: D561–D565

Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn A, Vingan S, Bader G, Xenarios I, Wojcik J, Sherman D, Tyers M, Salama J, Moore S, Ceol A, Chatr-aryamontri A, Oesterheld M, Stumpflen V, Salwinski L, Neroth J, Cerami E et al (2007b) Broadening the horizon—level 2.5 of the HuPO-PSI format for molecular interactions. BMC Biol 5: 44

Kholodenko BN, Demin OV, Moehren G, Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem 274: 30169–30181

Kitano H (2007a) A robustness-based approach to systems-oriented drug design. Nat Rev Drug Discov 6: 202–210

Kitano H (2007b) Towards a theory of biological robustness. Mol Syst Biol 3: 137

Li H, Ung CY, Ma XH, Li BW, Low BC, Chen YZ (2009) VisANT: an integrative framework for networks in systems biology. Mol Syst Biol 5: D411–D414

Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, Hanumanthu G, Gupta M, Upendran S, Gupta S, Mahesh M, Jacob B, Mathew P, Chatterjee P, Arun KS et al (2006) Human protein reference database—2006 update. Nucleic Acids Res 34: D411–D414

Oda K, Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. Mol Syst Biol 2: 2006.0015

Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive map of the toll-like receptor signaling network. Mol Syst Biol 1: 2005.0010

Hu Z, Snitkin ES, DeLisi C (2008b) VisANT: an integrative framework for networks in systems biology. Brief Bioinform 9: 317–325

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, the rest of the SBML Forum, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19: 524–531
Pico AR, Kelder T, Iersel MPv, Hanspers K, Conklin BR, Evelo C (2008) WikiPathways: pathway editing for the people. PLoS Biol 6: e184
Prenzel N, Zwick E, Daub H, Leserer M, Abraham R, Wallasch C, Ullrich A (1999) EGF receptor transactivation by G-protein-coupled receptors requires metalloproteinase cleavage of proHB-EGF. Nature 402: 884–888
Salomonis N, Hanspers K, Zambon A, Vranizan K, Lawlor S, Dahlquist K, Doniger S, Stuart J, Conklin B, Pico A (2007) GenMAPP 2: new features and resources for pathway analysis. BMC Bioinformatics 8: 217
Schaefer CF, Anthony K, Krupa S, Buchhoff J, Day M, Hannay T, Buetow KH (2009) PID: the pathway interaction database. Nucleic Acids Res 37: D674–D679
Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol 20: 370–375
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
Stromback L, Jakoniene V, Tan H, Lambrix P (2006) Representing, storing and accessing molecular interaction data: a review of models and tools. Brief Bioinform 7: 331–338
Stromback L, Lambrix P (2005) Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. Bioinformatics 21: 4401–4407
Suderman M, Hallett M (2007) Tools for visually exploring biological networks. Bioinformatics 23: 2651–2659
Tarbell JM, Ebong EE (2008) The endothelial glyocalyx: a mechanosensor and -transducer. Sci Signal 1: pt8
The UniProt Consortium (2008) The universal protein resource (UniProt). Nucleic Acids Res 36: D190–D195
van Iersel M, Kelder T, Pico A, Hanspers K, Coort S, Conklin B, Evelo C (2008) Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics 9: 399
Vastrik I, D’Eustachio P, Schmidt E, Joshi-Tope G, Gopinath G, Croft D, de Bono B, Gillespie M, Fassal R, Lewis S, Matthews L, Wu G, Birney E, Stein L (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8: R39
Waldrop M (2008) Wikiomics. Nature 455: 22, 25
Wang C-C, Cirit M, Haugh JM (2009) PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. Mol Syst Biol 5: 246
Yaffe MB (2008) Signaling networks and mathematics. Sci Signal 1: eg7
Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G (2002) MINT: a molecular INTeraction database. FEBS Lett 513: 135–140
Zinovyev A, Viaa E, Calzone L, Barillot E (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. Bioinformatics 24: 876–877

Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group. This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Licence.