D546–D551 *Nucleic Acids Research, 2006, Vol. 34, Database issue*
doi:10.1093/nar/gkj107

TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations

Mathias Krull[1,*], Susanne Pistor[1], Nico Voss[1], Alexander Kel[1], Ingmar Reuter[1], Deborah Kronenberg[2], Holger Michael[2], Knut Schwarzer[2], Anatolij Potapov[1,2], Claudia Choi[1], Olga Kel-Margoulis[1] and Edgar Wingender[1,2]

[1]BIOBASE GmbH, Halchtersche Strasse 33, D-38304 Wolfenbüttel, Germany and [2]Department of Bioinformatics, Medical School, University of Göttingen, Goldschmidtstrasse 1, D-37077 Göttingen, Germany

Received September 15, 2005; Revised and Accepted October 17, 2005

ABSTRACT

TRANSPATH® is a database about signal transduction events. It provides information about signaling molecules, their reactions and the pathways these reactions constitute. The representation of signaling molecules is organized in a number of orthogonal hierarchies reflecting the classification of the molecules, their species-specific or generic features, and their post-translational modifications. Reactions are similarly hierarchically organized in a three-layer architecture, differentiating between reactions that are evidenced by individual publications, generalizations of these reactions to construct species-independent 'reference pathways' and the 'semantic projections' of these pathways. A number of search and browse options allow easy access to the database contents, which can be visualized with the tool PathwayBuilder™. The module PathoSign adds data about pathologically relevant mutations in signaling components, including their genotypes and phenotypes. TRANSPATH® and PathoSign can be used as encyclopaedia, in the educational process, for vizualization and modeling of signal transduction networks and for the analysis of gene expression data. TRANSPATH® Public 6.0 is freely accessible for users from non-profit organizations under http://www.gene-regulation.com/pub/databases.html.

INTRODUCTION

Living cells have to interact intimately with their environment through exchanging a huge variety of signals. In particular the cells of higher eukaryotes have not only to react to extra-organismic stimuli but also to all the signals transmitted from other cells and organs of the same organism in order to coordinate their activities. In many, if not most, cases responding to such a signal implies to adapt the presently active genetic program of the receiver cell. To achieve this, the signal has to be forwarded to the nucleus which occurs via more or less complex signaling pathways. Many of these pathways have the property of a cascade, enabling the input signal to be multiplied on several steps.

Many of these signaling pathways, which may combine to networks and circuits through all kinds of cross-talks, have been unraveled until now and actually have been a matter of intense biochemical and molecular biological research since decades. The need to store this huge body of information in a computer-readable format was recognized early, when CSNDB started its pioneering work in this field (1,2), followed by TRANSPATH® (3), and a couple of further databases and information resources on the same subject such as PATIKA (4), aMAZE (5) or STKE (www.stke.org). During the last years, the TRANSPATH® database was subject to considerable expansion of its contents as well as changes of its data structure (6–8), the latest of them to be reported here. Among the most important changes, the method how variants and homologous molecules are represented, and the way to describe post-translational modifications (PTMs) and their effects will be reported.

DATA STRUCTURE

Basically, all signaling pathways in TRANSPATH® have been modeled as bipartite graphs with molecules and reactions as node classes. Each molecule is linked to the reaction entries in which it participates, and each reaction entry is linked to the entries of molecules involved. Both components, molecules

*To whom correspondence should be addressed. Tel: +49 5331 8584 32; Fax: +49 5331 8584 70; Email: mathias.krull@biobase-international.com

and reactions, are hierarchically organized, molecular components even in multiple hierarchies.

Molecule hierachies

The published data on signaling pathways and the molecular components involved differ largely in their granularity. Occasionally, certain features may have been assigned to whole families only, and this may have been done not because it is known that all members of that family share this property, but rather it is unkown exactly which family member exhibited this behavior in an investigated context. In this sense, all proteins or, more precisely, polypeptides, encoded by one gene form one such group as well, and only in a subset of reported experiments we can assign the reported property to an individual polypeptide. To cover these different granularities, we defined the molecule types 'family', 'isogroup' (summarizing all products of one gene) and 'basic' (representing an individual polypeptide to which we can assign, e.g. a sequence and a molecular weight) (Figure 1). In fact, the type 'family' is subject to a much finer diversification in a kind of ontology described elsewhere (9).

For example, alternative splicing of mRNA encoded by the human gene CDC25B, a cell cycle-relevant phosphatase, results in several different polypeptides. Three of them, Cdc25B1(h), Cdc25B2(h) and Cdc25B3(h) are included in TRANSPATH® as type 'basic' molecule entries with an assigned sequence. Since experimental results often do not differentiate between these isoforms, a molecule entry of type 'isogroup', Cdc25B(h), that represents the properties of all isoforms of this gene, has been created. In some cases, publications assign a certain signaling behavior to Cdc25 without indicating which gene is encoding the detected protein: CDC25A, CDC25B or CDC25C. To be able to capture this information correctly, the type 'family' molecule entry Cdc25(h) has been introduced.

The requirement for another hierarchy comes from the fact that many 'canonical' pathways that are commonly accepted by the research community have been put together using data from different organisms. Frequently, the individual publication does not even state explicitly the species a certain molecule or gene that was used in the reported experiments has been obtained from. Thus, we have to consider the possibility of 'unknown' values for certain attributes such as the biological species. In addition, abstraction onto the level of 'orthologous grouping' helps to gain and keep overview of a certain area. For these reasons, the molecule hierarchy described above is superimposed by a correspondingly structured layer of orthologs, leading to molecule entries of 'orthofamily', 'orthogroup' and 'orthobasic' type (Figure 1).

In addition, a new concept to represent PTMs of the molecules at the different levels has been implemented. The most basic feature of this representation is again that we need to differentiate between very distinct degrees of granularity

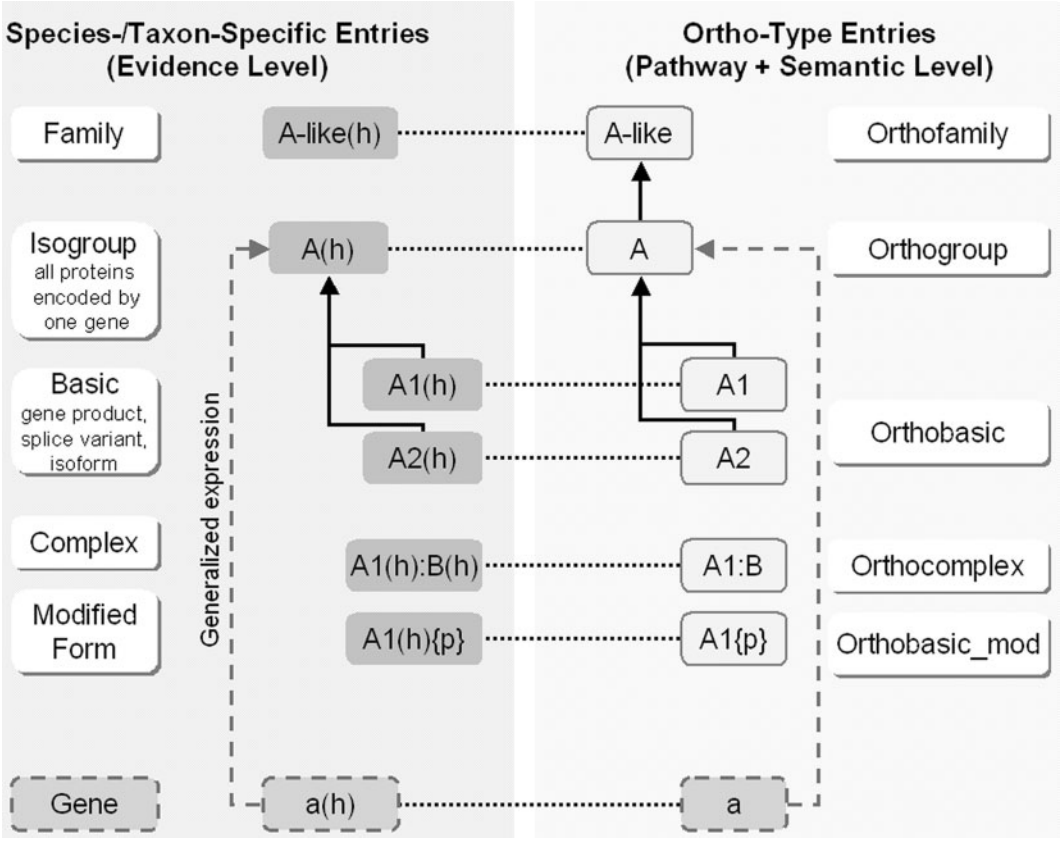

Figure 1. Hierarchical relations between the different molecule types and their assignment to the three-layer reaction hierarchy. Lines indicate links between items. Modified forms are represented with the ending _mod, e.g. basic_mod, orthocomplex_mod. This can be applied to any type. Relations between the gene and molecule table are shown with dashed lines.

of the available information: the property 'modification status unkown' has to be differentiated from 'modification status (knowingly) irrelevant' or 'unmodified', going deeper through observations like 'any phosphorylation' required for a certain purpose down to the detailed description of 'phosphorylation at Ser-n' in a certain context. Further details of representing PTM in TRANSPATH® have been described earlier (7). TRANSPATH® provides information about the combinatorial formation of multiprotein complexes. The same protein (polypeptide) may be a part of several different complexes and thus participate in different physiological functions within the cell.

Having particular protein variants, PTMs and multiprotein complexes as the separate entries in the Molecule table, we use them as educts or products for reactions. This data model allows to correctly reflect specific functions of particular protein molecules.

Reaction hierarchy

We have now consistently modeled a three-layer reaction hierarchy throughout the database. The most basic level stores individual signaling reactions with all mechanistic details available in the corresponding publication, such as the phosphorylation reaction

$$\mathrm{IRAK(v.s.)} + 2\mathrm{ATP} - \mathrm{IRAK\text{-}4(h)} \rightarrow \mathrm{IRAK(v.s.)\{pS376\}\{pT387\}} + 2\mathrm{ADP}$$

(accession number XN000024244; note that writing IRAK-4(h) as interrupt of the reaction arrow characterizes it as a catalyst), or the binding reactions

$$\mathrm{Tollip(m)} + \mathrm{IRAK(h)} \Leftrightarrow \mathrm{Tollip(m):IRAK(h)}\ (\mathrm{XN000003468})$$

and

$$\mathrm{IRAK\text{-}4(h)} + \mathrm{IRAK(h)} \Leftrightarrow \mathrm{IRAK\text{-}4(h):IRAK(h)}\ (\mathrm{XN000024237}).$$

As can be seen, information about the species/taxon origin of each molecule is included [(v.s.) for vertebrate species], revealing that many of these experiments have been conducted in a heterologous set-up. Examples are the transfection of cells with constructs of mixed species origin for co-immunoprecipitation studies (as for XN000003468) or the use of bacterially expressed proteins in an *in vitro* kinase assay with a rat enzyme and a human substrate. This reaction level is called 'evidence' level.

Since several publications may have dealt with the same reaction in a different experimental set-up, the evidence level inevitably contains some redundancy. Moreover, if one publication may have shown the PTM of molecule A(species1) by B(species2), another report may have demonstrated that A(species3) physically interacts with B(species4) as it is shown in the examples above. Since the interaction is a prerequisite for the modification, these pieces of information are summarized into one reaction step on the next level. It is called 'pathway' level since the corresponding reactions entries have been optimized for constructing whole pathways out of them. In this step, abstraction is also done for the biological species, so that reactions on the pathway level usually involve 'ortho-' entries. However, they still model each reaction in a mechanistically 'complete' manner. Thus, the reaction given above would now read

$$\mathrm{Tollip:IRAK} + \mathrm{IRAK\text{-}4} + 2\mathrm{ATP} \rightarrow \mathrm{Tollip:IRAK\{pS376\}\{pT387\}:IRAK\text{-}4} + 2\mathrm{ADP}\ (\mathrm{XN000024253})$$

As can be seen, IRAK is a part of a complex with Tollip and is actually phosphorylated within this complex. The kinase IRAK-4 becomes also a part of the multiprotein complex. This level is optimally used for constructing species-independent 'Reference Pathways'.

The third level is the 'semantic' projection of the contents on the pathway level. In this kind of representation, only those molecules are displayed which process and forward the signal. Small ubiquitous molecules such as ATP or H_2O are omitted from the reaction equation, as is information about complex and PTM status of the molecules. In turn, the roles of these 'semantic' reactions have to be annotated explicitly with terms like 'binding' and 'phosphorylation'. The example given above would now read:

$$\mathrm{IRAK\text{-}4} \rightarrow \mathrm{IRAK}\ (\mathrm{XN000024254}).$$

In addition to these three reaction types, we have defined 'indirect' reactions and 'decompositions'. The first indicates the existence of black boxes along the pathway which may contain one or several as yet unidentified steps between the upstream and the downstream molecules (or genes) of such a reaction. Typically, indirect reactions are at the semantic level since obviously, no mechanistic details can reasonably be given for them. In contrast, virtual 'decompositions' are generally assigned to the pathway level and describe reactions occurring within complexes as was discussed earlier in detail (7). As a prerequisite, they require a complex formation reaction to be modeled in the database beforehand.

Pathways

Individual signaling reactions are grouped to chains and pathways. Chains are sequences of individual reactions that have been experimentally proven and reported to occur subsequently to each other. TRANSPATH® (public version 6.0) provides pathway information by (i) graphical overviews with click-able items that are available for the IL1 pathway (the reaction examples given above are part of this pathway), insulin pathway, p53 network, p53 phosphorylation map and p73 network; and (ii) using the PathwayBuilder™ (see below) to construct complex pathways and signaling networks out of the individual reactions stored.

The PathoSign module

The interference and dysfunction of intercellular signaling pathways, e.g. caused by gene mutations, are involved in the aetiology of a multitude of human diseases. To start coping with these data and as a basis for the modeling of pathological aberrations of signaling pathways we developed PathoSign, a relational database which collects information about pathologically relevant mutations of signal transduction components and their corresponding phenotypes. The database

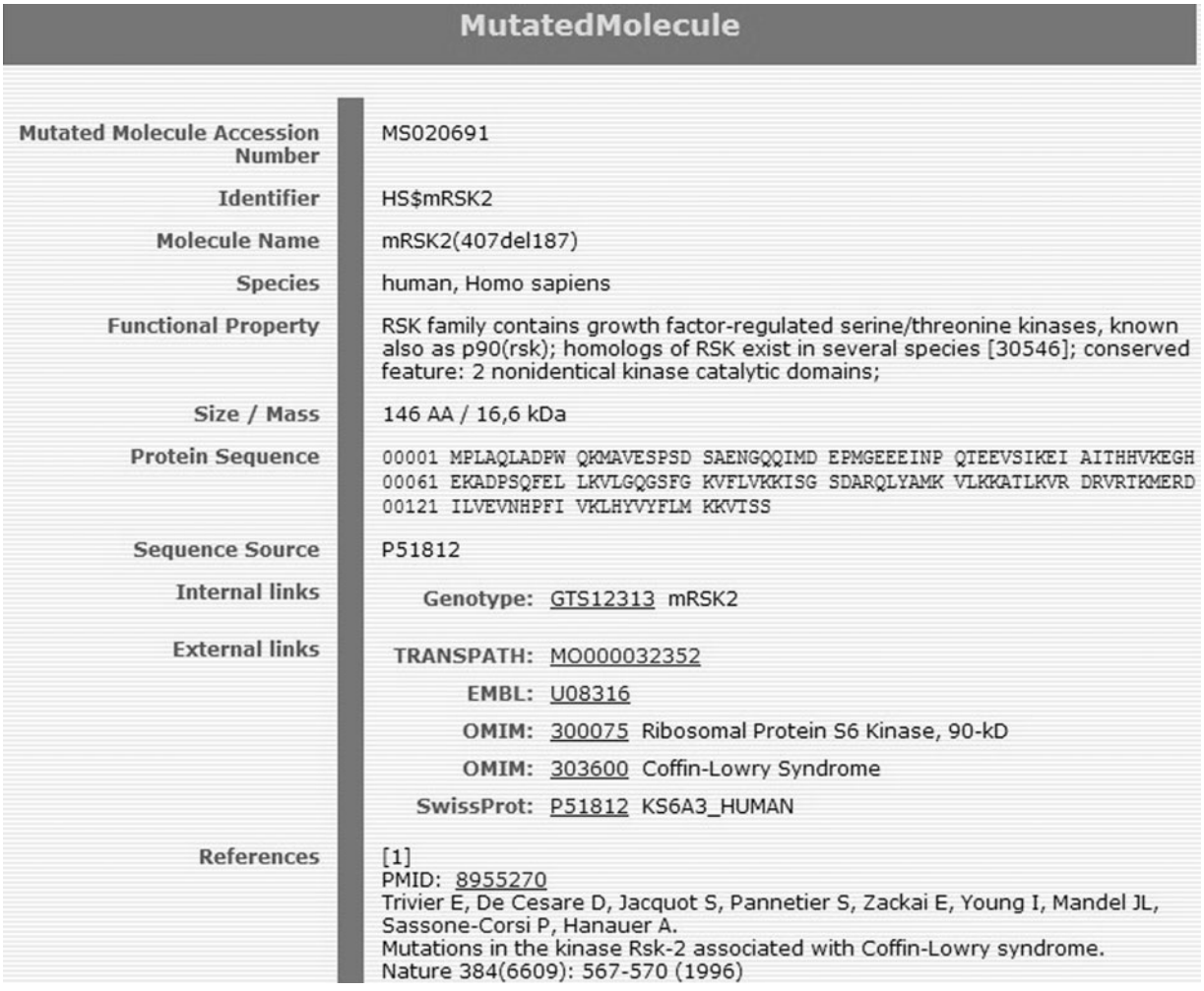

Figure 2. PathoSign mutated molecule view. Depicted is one of 14 mutations of the ribosomal S6 kinase 2 gene (RSK2) which are presently annotated. The Coffin–Lowry syndrome is caused by mutations in the RSK2 gene. Linked to the mutated molecule entry is the respective genotype (see 'Internal Links' in figure), wherefrom the phenotype entry referring to the Coffin–Lowry syndrome (PathoSign entry no. PTS00210) can be accessed.

comprises three main subjects, Mutated Molecules, Genotypes and Phenotypes. To showcase database functionality, PathoSign has been populated since its inception in March 2005 with data of ~214 signaling molecule mutations (Figure 2). Linked to these are tables that contain information about the corresponding phenotypes (40 entries) and the underlying genotypes (216 entries). The database is intended as a stand-alone data repository as well as a disease-related module of TRANSPATH®. Mutated variants of the molecules in PathoSign are linked to the corresponding normal molecules in TRANSPATH® and genes encoding TRANSPATH® molecules contain links to the corresponding disease-related variants in PathoSign. In future, this extension, in federation with additional resources, will encompass the modeling of pathological aberrations of cell signaling pathways.

Table 1. Number of entries in the TRANSPATH® 6.0 public release in comparison with the previous public release, 5.1, as well as with the current professional release

Table	TRANSPATH Public Release 5.1	TRANSPATH Public Release 6.0	TRANSPATH Professional Release 6.2
Molecule	16 894	20 155	28 779
Gene	4 509	7 603	11 157
Reaction	20 198	28 126	52 977
semantic	11 579	17 039	11 098
pathway step	1 343	1 930	2 111
molecular evidence	6 153	7 929	29 875
Clickable maps	2	5	61
Pathways + Chains	—	—	406

CONTENTS

All data in TRANSPATH® have been extracted from the full-text original publications by manual curation. While being maintained internally as a relational database, online access is granted to a system of flat files, comprising the text files Molecule, Reaction and Gene. The number of entries in each table is given in Table 1.

TRANSPATH® and PathoSign provide links to a number of commonly used databases: GO terms (10), EMBL/GenBank/DDBJ (11), ENSEMBL (12), RefSeq (13), UniProt (14),

EntrezGene (15), HGNC (16), MGI (17), RGD (18), InterPro (19), AFFYMETRIX, OMIM (20) and TRANSFAC® (21).

DATA ACCESS AND VISUALIZATION

The search interface has been significantly improved in comparison with the previous public release of TRANSPATH®. The new interface allows specifications by different operators like 'Entire words', 'Case sensitive', 'Fuzzy' along with Boolean operators for logical operations on the search terms. Users can define the output contents and results can be sorted by any field. Search results can be stored and subsequently refined.

The contents of the database can be browsed. For this, a comprehensive classification of signaling molecules has been developed which has been optimized for representing the function of signaling molecules along the information flow (7). The main classes are: adaptors, chaperones, co-factors, cytoskeletal proteins, enzymes, GTPase-controlling molecules, ligands, membrane-transducing components, transporters and transcription factors.

TRANSPATH® is associated with the visualization tool PathwayBuilder™. With the help of this tool the signaling network of TRANSPATH® can be traversed and visualized dynamically.

To obtain customized views on the pathways and networks of interest several parameters are available, such as network expansion, pathway direction (upstream or downstream relative to a given molecule), number of steps, complexity as well as different styles of the resulted view. A further functionality enables highlighting molecules according to freely adjustable filters; one can highlight, for instance, all adaptor proteins within a given network or all human proteins. The resulting visualizations can be saved and merged as necessary.

DISCUSSION

The TRANSPATH® database is used for different purposes, some of which may overlap with the application of protein–protein interaction networks, e.g. when interpreting gene expression data. It should be noted, however, that both types of networks exhibit some fundamental differences, mainly: (i) in contrast to the protein–protein interaction networks, signal transduction paths are directed, since the flow of information through the cell is directed; (ii) signal transduction pathways also comprise many non-proteinaceous molecules as essential items (e.g., steroids, phospholipids, inorganic ions, nucleic acids); (iii) by including gene regulatory events, they can be easily expanded to any number of 'cycles' of classical signal transduction pathways, describing the path of a signal from the membrane to, e.g. the nucleus. Therefore, protein–protein interaction data cannot cover all requirements of modeling functional processes in a cell, for which pathway-oriented databases are needed.

TRANSPATH® is one of the few databases in the field that aims at providing information about signal transduction pathways, and presenting this information on different abstraction levels of the reactions, and including all required molecular information about PTMs and complexes. We feel that this altogether may be a particular strength of TRANSPATH® compared with other sources like STKE (http://www.stke.org), AfCS (http://www.signaling-gateway.org/), Reactome [http://www.reactome.org/ (22)], BioCarta (http://www.biocarta.com) or the signaling part of KEGG [http://www.genome.ad.jp/kegg/ (23)], each of them having its own particular strengths and weaknesses.

We see a further advantage of our system in the synergistic potential between TRANSPATH® and PathoSign, which we just started to fully exploit. It is one of the aims of our future developments to make use of TRANSPATH® for modeling complex signaling processes, and to include PathoSign contents for exploring the (potential) far-reaching effects of mutations in signaling components onto the overall signaling network.

Another advantage of TRANSPATH® may be the sheer size of the available information. Since signal transduction mechanisms are under intense investigation by biochemists and molecular biologists since decades, there is an overwhelming body of data available in the literature. 'Complete' coverage of the sources is therefore nearly impossible to achieve, and even an attempt of picking up all relevant reactions with at least a few most important references attached is an enormous task. Therefore, any database in this field, including TRANSPATH®, is far from being 'complete' (in whatever sense). However, compared with other resources, TRANSPATH® seems to have a relatively good coverage of signaling reactions and pathways, which is rapidly expanding. Up to now, we refrained from including automatically retrieved information in order to keep quality and reliability of the data as high as possible. However, we will make increased use of the data retrieval pipeline established for the curation of the Proteome Databases [http://www.proteome.com (24)] in future to increase the efficiency of initial retrieval of relevant data, which we will continue processing manually afterwards.

AVAILABILITY

TRANSPATH® Public release 6.0, which has been described in this contribution, is freely accessible for users from non-profit organizations under http://www.gene-regulation.com/pub/databases.html#transpath, the PathoSign module is freely available under http://pathosign.bioinf.med.uni-goettingen.de/ or http://www.gene-regulation.com/pub/databases.html#pathosign.

ACKNOWLEDGEMENTS

Parts of the work were funded by grants of the German Ministry of Education and Research (BMBF) 'Intergenomics' (031U210B), collectively by BioRegioN GmbH and BMBF 'BioProfil' (0313092), by European Commission under FP6-'Life sciences, genomics and biotechnology for health' contract LSHG-CT-2004-503568 'COMBIO', by European Commission under 'Marie Curie research training networks' contract MRTN-CT-2004-512285 'TRANSISTOR'. Design of the PathoSign module was partially supported by Nationales Genomforschungsnetz (NGFN), German Ministry of Education and Research (BMBF), grant no. 01GR0480. Funding to pay the Open Access publication charges for this article was provided by Biobase GmbH.

Conflict of interest statement. None declared.

REFERENCES

1. Takai-Igarashi,T., Nadaoka,Y. and Kaminuma,T. (1998) A database for cell signaling networks. *J. Comput. Biol.*, **5**, 747–754.
2. Takai-Igarashi,T. and Kaminuma,T. (1999) A pathway finding system for the cell signaling networks database. *In Silico Biol.*, **1**, 129–146.
3. Schacherer,F., Choi,C., Götze,U., Krull,M., Pistor,S. and Wingender,E. (2001) The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, **17**, 1053–1057.
4. Demir,E., Babur,O., Dogrusoz,U., Gursoy,A., Nisanci,G., Cetin-Atalay,R. and Ozturk,M. (2002) PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, **18**, 996–1003.
5. Lemer,C., Antezana,E., Couche,F., Fays,F., Santolaria,X., Janky,R., Deville,Y., Richelle,J. and Wodak,S.J. (2004) The aMAZE LightBench: a web interface to a relational database of cellular processes. *Nucleic Acids Res.*, **32**, D443–D448.
6. Krull,M., Voss,N., Choi,C., Pistor,S., Potapov,A. and Wingender,E. (2003) TRANSPATH®: an integrated database on signal transduction and a tool for array analysis. *Nucleic Acids Res.*, **31**, 97–100.
7. Choi,C., Crass,T., Kel,A., Kel-Margoulis,O., Krull,M., Pistor,S., Potapov,A., Voss,N. and Wingender,E. (2004) Consistent re-modeling of signaling pathways and its implementation in the TRANSPATH database. *Genome Inform. Ser. Workshop Genome Inform.*, **15**, 244–254.
8. Kel-Margoulis,O., Matys,V., Choi,C., Reuter,I., Krull,M., Potapov,A.P., Voss,N., Liebich,I., Kel,A. and Wingender,E. (2005) Databases on Gene Regulation. In Bajic,V.B. and Tan,T.W. (eds), *Information Processing And Living Systems*. World Scientific Publishing Co., SingaporeVol. 2, pp. 709–727.
9. Wingender,E. (2003) TRANSFAC®, TRANSPATH® and CYTOMER® as starting points for an ontology of regulatory networks. *In Silico Biol.*, **4**, 0006.
10. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
11. Kanz,C., Aldebert,P., Althorpe,N., Baker,W., Baldwin,A., Bates,K., Browne,P., van den Broek,A., Castro,M., Cochrane,G. *et al.* (2005) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **33**, D29–D33.
12. Hubbard,T., Andrews,D., Caccamo,M., Cameron,G., Chen,Y., Clamp,M., Clarke,L., Coates,G., Cox,T., Cunningham,F. *et al.* (2005) Ensembl 2005. *Nucleic Acids Res.*, **33**, D447–D453.
13. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
14. Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R., Magrane,M. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
15. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
16. Wain,H.M., Lush,M.J., Ducluzeau,F., Khodiyar,V.K. and Povey,S. (2004) Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Res.*, **32**, D255–D257.
17. Bult,C.J., Blake,J.A., Richardson,J.E., Kadin,J.A., Eppig,J.T., Baldarelli,R.M., Barsanti,K., Baya,M., Beal,J.S., Boddy,W.J. *et al.* (2004) The Mouse Genome Database (MGD): integrating biology with the genome. *Nuceic Acids Res.*, **32**, D476–D481.
18. de la Cruz,N., Bromberg,S., Pasko,D., Shimoyama,M., Twigger,S., Chen,J., Chen,C.F., Fan,C., Foote,C., Gopinath,G.R. *et al.* (2005) The Rat Genome Database (RGD): developments towards a phenome database. *Nucleic Acids Res.*, **33**, D485–D491.
19. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bradley,P., Bork,P., Bucher,P., Cerutti,L. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.*, **33**, D201–D205.
20. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
21. Matys,V., Fricke,E., Geffers,R., Gossling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E., Kel-Margoulis,O.V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.
22. Joshi-Tope,G., Gillespie,M., Vastrik,I., D'Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R., Matthews,L. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
23. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resources for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
24. Hodges,P.E., Carrico,P.M., Hogan,J.D., O'Neill,K.E., Owen,J.J., Mangan,M., Davis,B.P., Brooks,J.E. and Garrels,J.I. (2002) Annotating the human proteome: the Human Proteome Survey Database (HumanPSD) and an in-depth target database for G protein-coupled receptors (GPCR-PD) from Incyte Genomics. *Nucleic Acids Res.*, **30**, 137–141.