CYCLONET—an integrated database on cell cycle regulation and carcinogenesis

Fedor Kolpakov[1,2], Vladimir Poroikov[3], Ruslan Sharipov[1,2,4,*],
Yury Kondrakhin[1,2], Alexey Zakharov[3], Alexey Lagunin[3],
Luciano Milanesi[5] and Alexander Kel[6]

[1]Institute of Systems Biology, 15, Detskiy proezd, Novosibirsk 630090, Russia, [2]Design Technological Institute of Digital Techniques, Siberian Branch of Russian Academy of Sciences, 6, Institutskaya, Novosibirsk 630090, Russia, [3]Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, 10, Pogodinskaya Street, Moscow 119121, Russia, [4]Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, 10, Lavrentyev Avenue, Novosibirsk 630090, Russia, [5]CNR-Institute of Biomedical Technologies, 93, Via Fratelli Cervi, Segrate (MI) 20090, Italy and [6]BIOBASE GmbH, 33, Halchtersche Strasse, Wolfenbuettel 38304, Germany

ABSTRACT

Computational modelling of mammalian cell cycle regulation is a challenging task, which requires comprehensive knowledge on many interrelated processes in the cell. We have developed a web-based integrated database on cell cycle regulation in mammals in normal and pathological states (Cyclonet database). It integrates data obtained by 'omics' sciences and chemoinformatics on the basis of systems biology approach. Cyclonet is a specialized resource, which enables researchers working in the field of anticancer drug discovery to analyze the wealth of currently available information in a systematic way. Cyclonet contains information on relevant genes and molecules; diagrams and models of cell cycle regulation and results of their simulation; microarray data on cell cycle and on various types of cancer, information on drug targets and their ligands, as well as extensive bibliography on modelling of cell cycle and cancer-related gene expression data. The Cyclonet database is also accessible through the BioUML workbench, which allows flexible querying, analyzing and editing the data by means of visual modelling. Cyclonet aims to predict promising anticancer targets and their agents by application of Prediction of Activity Spectra for Substances. The Cyclonet database is available at http://cyclonet.biouml.org.

INTRODUCTION

The main goal of the Cyclonet database is to integrate information from genomics, proteomics, chemoinformatics and systems biology on mammalian cell cycle regulation in normal and pathological states. This will help molecular biologists working in the field of anticancer drug development to analyze systematically all these data and generate experimentally testable hypotheses (Figure 1).

Cyclonet incorporates data on various carcinogenesis related topics, such as: cell cycle control in mammals (Figure 2), cell survival programs (e.g. NF-κB pathway), regulation of covalent histone modifications and chromatin remodelling in cell cycle, DNA methylation and other epigenetic mechanisms of cell growth and differentiation. Biological pathways, computer models of cell cycle, microarray data coming from studies of cell cycle and analysis of cancer-related materials are also systematically collected in this database (1) (http://www.impb.ru/~rcdl2004/cgi/get_paper_pdf.cgi?pid=30).

Cyclonet supports discovery of novel drug targets and development of effective anticancer therapies by collecting all available data related to the control of cell cycle in normal and pathological states and providing a system biology platform for knowledge-based anticancer drug discovery.

Novel software technologies were used for the database development:

(i) the BioUML workbench [http://www.biouml.org, (2,3)] was used for formal description and visual modelling of biological pathways and processes related to the cell cycle regulation and cancer (Figure 2). It also allows to

---

simulate the behaviour of the described systems using Java or MATLAB simulation engines;

(ii) BeanExplorer Enterprise Edition (http://www.beanexplorer.com) was used to develop web interface for the Cyclonet database (Figure 3).
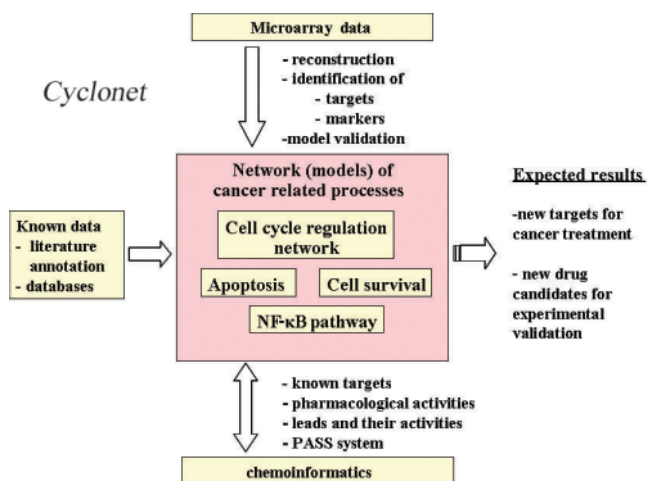

Figure 1. Diagrams and models of carcinogenesis related processes as the basis for information integration in the Cyclonet database.

THE CYCLONET DATABASE STRUCTURE AND CONTENT

The Cyclonet database consists of three main components (see Table 1):

(i) diagrams and models of biological pathways (metabolic pathways, signal transduction pathways and gene networks) involved in cell cycle regulation and carcinogenesis;

(ii) microarray original data and results of their statistical analysis;

(iii) chemoinformatics data—drug targets, ligands and pharmacological activities for cancer treatment.

The Cyclonet database is organized as a relational database (MySQL DBMS). All sections contain a number of tables that are highly interconnected through crossreferences. Such elaborated relational schema enables complex queries combining various types of information.

Data in Cyclonet are compiled mainly by manual literature annotation. Links to the public databases, such as, Gene-Ontology (4), RefSeq (5) and Ensembl (6) are provided from genes, proteins and other respective entries. Cyclonet also contains a vast body of literature references that are arranged by categories.
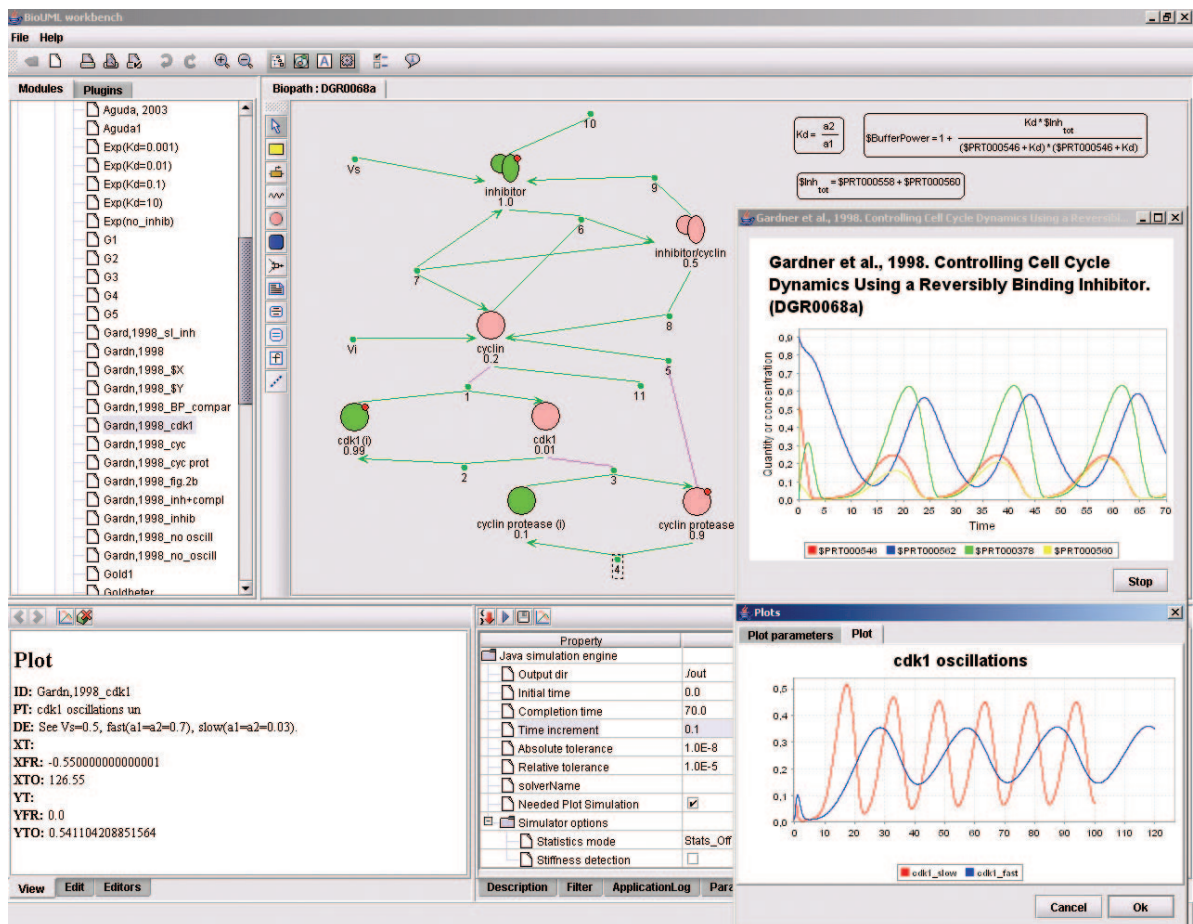

Figure 2. An example of cell cycle model visualization and simulation by the BioUML workbench (the diagram DGR0068a of the Cyclonet database).

Figure 3. Web interface of the Cyclonet database generated by BeanExplorer technology. Top screen displays fragment of microarray series classification in the Cyclonet database, bottom left screen demonstrates a fragment of the list of pharmacological activities for anticancer therapy, bottom right—examples of chemical structure of two tubulin antagonists.

Biological pathways

We use BioUML for formal description of signal transduction pathways and gene networks involved in cell cycle regulation and carcinogenesis (2,3,7). Cyclonet pathways section allows to store the detailed description of biological pathways, their components, models as well as the results of simulation.

We are using several diagram types of BioUML to describe cell cycle regulation and carcinogenesis:

(i) semantic networks describing relationships between the main concepts (for example, G1 phase, G1/S transition, mitotic checkpoints) and components of cell cycle regulation;
(ii) pathways describing structure of cell cycle regulatory networks as compartmentalized graphs. We have classified the annotated networks into a number of categories that describe different parts of cell cycle regulatory networks in details (for example, a network that provides G1/S transition, NF-κB signal transduction pathway and its influence on apoptosis and others).

Models

The BioUML technology was also used for visual modelling of cell cycle regulation. Known cell cycle models were imported from SBML (8) and CellML (9) model repositories. We added into Cyclonet several new recent models by manual annotation of respective literature sources. We also created our own novel model of regulation of G1/S transition of cell cycle.

Currently, Cyclonet contains 37 models of cell cycle regulation. All models can be classified into two groups: (i) general models that simulate behaviour of rather small systems including abstract objects that reflect real biological components in the cell; (ii) 'portrait' models that try to simulate different sub-processes in cell cycle and include real genes, proteins and other cellular components. We validated each model by using the BioUML simulation engine and comparing the results with the published results. The results of such simulations were then stored in the Cyclonet database. These data can be displayed as graphs by the BioUML workbench (Figure 2) or web isnterface generated by BeanExplorer EE.

Table 1. Number of entries for the main blocks (tables, sections) of the Cyclonet database

Sections of the Cyclonet database	Number of Entries
Biological pathways	
Compartments: cellular and organism compartments, for example, cytoplasm, liver, blood	120
Concepts: biological processes, states, functions etc. For example, G1 phase, mitosis	987
Genes: list of genes involved in regulation of cell cycle and related processes	156
Proteins	1942
Reactions	2406
Relations: list of semantic relations (for example, is-a, part-of, activate) between the pathway components	6137
RNA	30
Substances	485
Diagrams	179
Models	37
Microarray data and result of their analysis	
Genes: complete list of human genes compiled as union of data from HGNC and UniGene databases	45 759
Gene crossreferences	181 949
Clones: list of human cDNA clones used in cDNA microarray experiments. It was compiled from UniGene database	2 455 666
Information resources: information resources related with microarray data. For example, SMD, GEO, Oncomine	41
Microarray experiments: categorized list of microarray experiments. Each entry contains brief description of the experiment and links to original article and primary data	354
Cell cycle	24
Cancer	310
Uploaded microarray experiments: number of cDNA experiments which primary data were uploaded into the Cyclonet database	10
Samples	223
Observations	3 407 624
Gene lists (results of analyses of microarray data)	33
Genes in the lists: total number of genes that were included in these lists. Note—one gene can be included into several lists.	17 318
Chemoinformatics	
Activities: names of molecular mechanisms of action (e.g. Topoisomerase II inhibitor) and pharmacotherapeutic action (e.g. Antimetabolite)	201
Structures: structural formulas of drug-like molecules interacting with the key elements of the cell cycle	4205
Targets: molecular targets interacting with chemical compounds mentioned above	182

Microarray data

Cyclonet contains a comprehensive list of human genes which is composed from the genes described in HGNC (10) and UniGene (11) databases. Cyclonet also contains all assignment of cDNA clones to the corresponding human genes.

We analysed 41 microarray resources [mainly, Standford Microarray Database (12), GEO (13), Oncomine (14) and published articles, for example, (15)] and obtained 354 links to microarray experimental data related to the cell cycle and cancer. These links to microarray data were classified according to cancer types.

Currently data for five microarray experiments related to breast cancer and five experiments with cell cycle time series were loaded into the Cyclonet database and analysed. We did a statistical analysis as well as meta-analysis of the data (see Supplementary Data) and obtained 33 gene lists (IDs GL0001–GL0033 in 'Microarray data and results' of Cyclonet) that belong to several categories:

(i) lists of genes periodically expressed during cell cycle (GL0007, GL0020 and GL0021) (15,16);

(ii) lists of genes whose expression is changing monotonically during cell cycle (GL0022) (15,16);

(iii) breast cancer gene lists:

    (a) up- and down-regulated genes in each of the five experiments (GL0001–GL0006, GL0008–GL0018) (17–21);

    (b) up- and down-regulated genes revealed on the basis of meta-analysis (GL0019, GL0023–GL0033) (22);

(iv) lists of genes obtained by other authors during microarray analysis of breast cancer (18–22) and pancreatic cancer (23).

Such lists of differentially expressed genes are very good resources for selecting cancer biomarkers as well as perspective targets for further experimental and bioinformatic analysis. Statistical methods used in this analysis are described in Supplementary Data.

Chemoinformatics data

Chemoinformatics section summarizes the current knowledge about known anticancer targets, anticancer agents, mechanisms of their action and conditions where those compounds are applied. For this purpose we are collecting the following information as it is represented in Supplementary Figure 1S:

(i) names of anticancer agents (generic name, brand name) and its synonyms;

(ii) chemical name;

(iii) CAS number;

(iv) structural formulae;

(v) class (activity)—includes information about molecular mechanisms of action (e.g. Topoisomerase II inhibitor) and pharmacotherapeutic action (e.g. Antimetabolite);

(vi) literature references where the data were obtained for the respective anticancer agent.

Semantic networks provide a reasonable formalism to describe the relationships between the anticancer agents and their targets, activities and cancer types (or other conditions) where these agents are generally applied (Figure 4). Summary statistics of the chemoinformatics section is shown in Table 1.

INTEGRATION BETWEEN COMPONENTS OF CYCLONET

Integration between all three components of the Cyclonet database, namely, biological pathways and models, microarray data and chemoinformatics data, is provided by the following mechanisms:

(i) All data are stored in the same relational database. This allows us to develop the complex SQL queries to
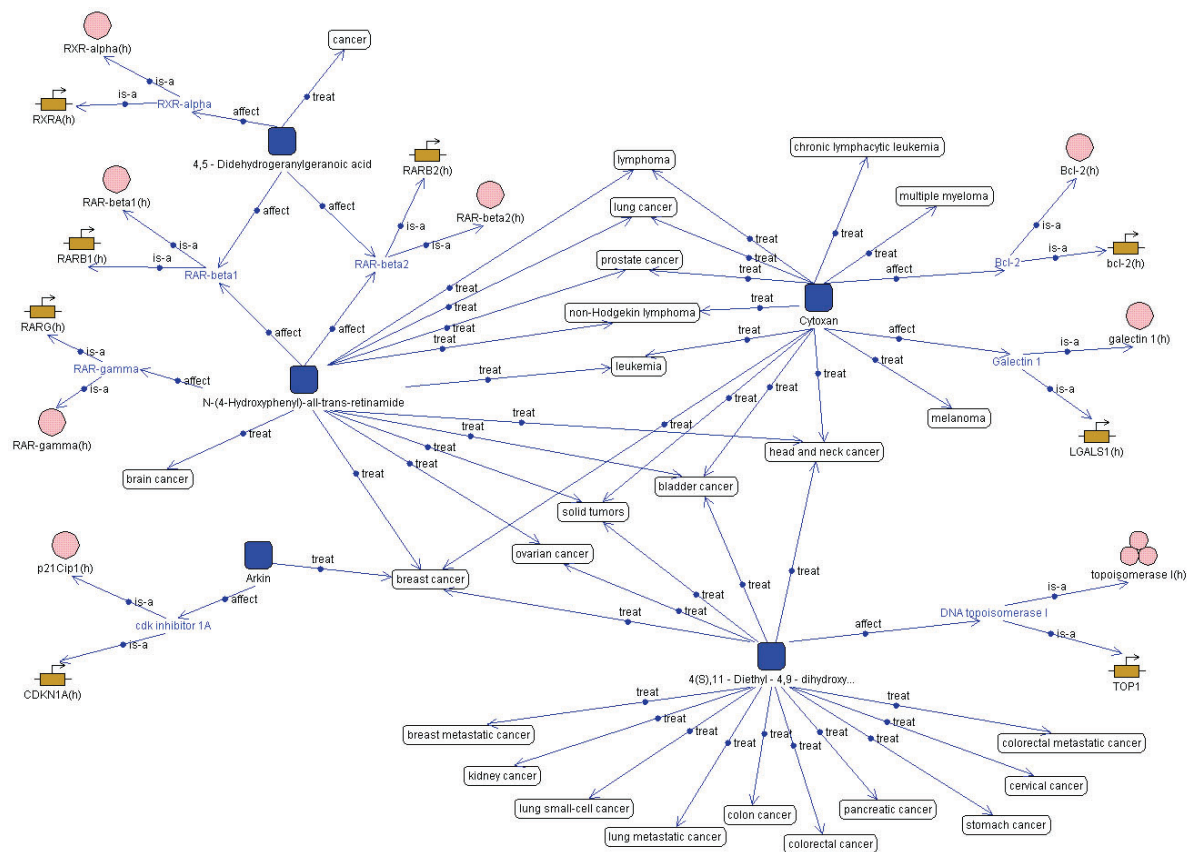
Figure 4. A fragment of semantic network that describes influences of several leads on common targets taking into account different cancer types (conditions). The diagram DGR0277a in the Cyclonet database.

integrate data from different components. A number of predefined SQL queries are provided through the web interface for the Cyclonet database.

(ii) The web interface provides detailed representation (view) of components of biological pathways, microarray and chemoinformatics data with a number of crossreferences between the components. For example, a view for an anticancer agent contains links to its activities, cancer types, conditions of its application for anticancer therapy, components of biological pathways (genes and proteins) that are targets for this agent. These targets, in turn, can be linked to diagrams and dynamic models of cell cycle. Another example is a gene view that contains links to cDNA clones used for this gene in microarray experiments, microarray experiments where expression level of this gene was measured, gene lists where this gene was revealed as result of microarray analyses, anticancer agents for which this gene is a target, diagrams and models where this gene participates.

(iii) The BioUML search engine allows to find the relationships between the anticancer agent and biological pathway components and display these results as an editable graph. As a starting point user can select the anticancer agent (small molecule), concept, gene or protein.

APPLICATION OF THE CYCLONET DATABASE

Prediction of new anticancer agents for known targets/mechanisms of action

All anticancer agents are grouped in the Cyclonet database according to their targets/mechanisms of action and chemical structure. This information is used for the training of computer program PASS (Prediction of Activity Spectra for Substances) (24). As a result of the training procedure, PASS can predict if new molecules from databases of commercially available samples may have activities related to the regulation of cell cycle. Three commercially available chemical compounds' sample databases were analysed, provided by ASINEX, ChemBridge and InterBioScreen (IBS). They contain totally the structures of 1 445 018 compounds. We predicted a number of compounds as potential cell cycle regulators using probability threshold $Pa > 70\%$. By increasing the Pa threshold, e.g. to 90%, one can select highly specific compounds only. The results of this analysis are stored in the Cyclonet database (see the statistics in Table 2). One may conclude that commercially available chemical compounds databases contain a plethora of ligands acting on different targets related to the cell cycle regulation.

Table 2. Potential cell cycle regulating agents in ASINEX, ChemBridge and IBS databases

Supplier	Number of compounds	Number of selected activities	Number of selected compounds
AsInEx	366 170	138	137 447
CHEMBRIDGE	734 140	142	363 057
IBS	344 708	151	115 959

Application of Cyclonet to model the cell cycle

Computer simulation methods have been applied to study the dynamics of gene networks regulating the cell cycle of vertebrates. The data on the regulation of the key genes obtained from the Cyclonet database have been used as a basis to construct gene networks of different degrees of complexity controlling the G1/S transition, one of the most important stages of the cell cycle. The behaviour dynamics of the model has been analysed. Two qualitatively different functional modes of the system have been obtained. It has been shown that the transition between these modes depends on the duration of the proliferation signal. It has also been demonstrated that the additional feedback from factor E2F to genes *c-fos* and *c-jun*, which was predicted earlier based on the computer analysis of promoters (25), plays an important role in the transition of the cell to the S phase (see Supplementary Figure 2S) as it is documented in gene expression databases TRANSFAC (26) and TRANSPATH (27).

Application of Cyclonet for searching of new targets for anticancer therapy

The Cyclonet database can be applied for searching of new targets for anticancer therapy. For this purpose we have revealed genes whose expression are significantly deregulated during breast cancer and created a set of diagrams in the Cyclonet database (diagrams DGR0228–DGR0240) and mapped information about gene expression into the diagrams. An example of gene expression data mapping is shown in Supplementary Figure 3S for a fragment of a diagram of the proapoptotic network (DGR240).

FURTHER DEVELOPMENT

Now we are developing a set of plug-ins in the BioUML workbench for visual modelling of integration between the biological pathways and microarray data that will provide: coloring of diagrams for biological pathways to display data on gene expression levels, reconstruction of gene networks and fitting the model parameters in accordance with the microarray data. Also, a new information arising from both 'omic'-sciences and chemoinformatics is added periodically to the Cyclonet database, to update its content.

SUPPLEMENTARY DATA

Supplementary data are available at NAR online.

ACKNOWLEDGEMENTS

REFERENCES

1. Kolpakov,F.A., Deineko,I., Zhatchenko,S.A. and Kel,A.E. (2004) Cyclonet—a database on cell cycle regulation. In *Proceedings of the 6th Russian Conference on Digital Libraries RCDL2004*, September 29–October 1, 2004, Pushchino, Russia, pp. 4–9.
2. Kolpakov,F.A. (2004) BioUML—open source extensible workbench for systems biology. In *Proceedings of The Fourth International Conference on Bioinformatics of Genome Regulation and Structure*, July 25–30, 2004, Novosibirsk, Russia, v.2, pp. 77–80.
3. Kolpakov,F., Puzanov,M. and Koshukov,A. (2006) BioUML: visual modeling, automated code generation and simulation of biological systems. In *Proceedings of The Fifth International Conference on Bioinformatics of Genome Regulation and Structure*, July 16–22, 2006, Novosibirsk, Russia, v.3, pp. 281–285.
4. Gene Ontology Consortium. (2001) Creating the gene ontology resource: design and implementation. *Genome Res.*, **11**, 1425–1433.
5. Wheeler,D.L., Chappey,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **28**, 10–14.
6. Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T. *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.
7. Kolpakov,F., Sharipov,R., Cheremushkina,E. and Kalashnikova,E. (2006) Biopath—a new approach to formalized description and simulation of biological systems. In *Proceedings of The Fifth International Conference on Bioinformatics of Genome Regulation and Structure,* July 16–22, 2006, Novosibirsk, Russia, v.3, pp. 96–100.
8. Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D., Cornish-Bowden,A. *et al.* (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
9. Lloyd,C.M., Halstead,M.D. and Nielsen,P.F. (2004) CellML: its future, present and past. *Prog. Biophys. Mol. Biol.*, **85**, 433–450.
10. Eyre,T.A., Ducluzeau,F., Sneddon,T.P., Povey,S., Bruford,E.A. and Lush,M.J. (2006) The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res.*, **34**, D319–D21.
11. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatusova,T.A. *et al.* (2003) Database Resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
12. Ball,C.A., Awad,I.A., Demeter,J., Gollub,J., Hebert,J.M., Hernandez-Boussard,T., Jin,H., Matese,J.C., Nitzberg,M., Wymore,F. *et al.* (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res.*, **33**, D580–D582.
13. Barrett,T., Suzek,T.O., Troup,D.B., Wilhite,S.E., Ngau,W.C., Ledoux,P., Rudnev,D., Lash,A.E., Fujibuchi,W. and Edgar,R. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res.*, **33**, D562–D566.
14. Rhodes,D.R., Yu,J., Shanker,K., Deshpande,N., Varambally,R., Ghosh,D., Barrette,T., Pandey,A. and Chinnaiyan,A.M. (2004)

ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*, **6**, 1–6.

15. Whitfield,M.L., Sherlock,G., Saldanha,A.J., Murray,J.I., Ball,C.A., Alexander,K.A., Matese,J.C., Perou,C.M., Hurt,M.M., Brown,P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumours. *Mol. Biol. Cell*, **13**, 1977–2000.

16. Kondrakhin,Y.V., Kel,A.E., Sharipov,R.N. and Kolpakov,F.A. (2006) Identification of binding site patterns in regulatory regions of human cell cycle genes. In *The 7th International Conference on Systems Biology*, Yokohama Japan, 9–11 October, 2006 (ICSB-2006), poster ID BC02.

17. Hedenfalk,I., Duggan,D., Chen,Y., Radmacher,M., Bittner,M., Simon,R., Meltzer,P., Gusterson,B., Esteller,M., Kallioniemi,O.-P. *et al.* (2001) Gene expression profiles in hereditary breast cancer. *N. Engl. J. Med.*, **344**, 539–548.

18. Ma,X.-J., Salunga,R., Tuggle,J.T., Gaudet,J., Enright,E., McQuary,P., Payette,T., Pistone,M., Stecker,K., Zhang,B.M. *et al.* (2003) Gene expression profiles of human breast cancer progression. *Proc. Natl Acad. Sci. USA*, **100**, 5974–5979.

19. Sorlie,T., Perou,C.M., Tibshirani,R., Aas,T., Geisler,S., Johnsen,H., Hastie,T., Eisen,M.B., van de Rijn,M., Jeffrey,S.S. *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. *Proc. Natl Acad. Sci. USA*, **98**, 10869–10874.

20. Perou,C.M., Jeffrey,S.S., van de Rijn,M., Rees,C.A., Eisen,M.B., Ross,D.T., Pergamenschikov,A., Williams,C.F., Zhu,S.X., Lee,J.C.F. *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA*, **96**, 9212–9217.

21. Zhao,H., Langerød,A., Ji,Y., Nowels,K.W., Nesland,J.M., Tibshirani,R., Bukholm,I.K., Kåresen,R., Botstein,D., Børresen-Dale,A.-L. *et al.* (2004) Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell*, **15**, 2523–2536.

22. Kondrakhin,Y.V., Poroikov,V.V., Sharipov,R.N., Kel,A.E. and Kolpakov,F.A. (2006) Meta-analysis of breast cancer microarray data: reliable identification of up- and down-regulated genes. In *The 7th International Conference on Systems Biology*, Yokohama Japan, 9–11 October, 2006 (ICSB-2006), poster ID MC08.

23. Grutzmann,R., Boriss,H., Ammerpohl,O., Luttges,J., Kalthoff,H., Schackert,H.K., Kloppel,G., Saeger,H.D. and Pilarsky,C. (2005) Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene*, **24**, 5079–88.

24. Poroikov,V. and Filimonov,D. (2005) PASS: prediction of biological activity spectra for substances. In Helma,C. (ed.), *Predictive Toxicology*. Taylor & Francis, pp. 459–478.

25. Kel,A., Deineko,I., Kel-Margoulis,O.V., Wingender,E. and Ratner,V. (2000) Modelling of gene regulatory network of cell cycle control. Role of E2F feedback loops. In *Proceedings of the German Conference on Bioinformatics (GCB 2000)*, pp. 107–114.

26. Matys,V., Kel-Margoulis,O., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC(r) and its module TRANSCompel(r): transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.

27. Krull,M., Pistor,S., Voss,N., Kel,A., Reuter,I., Kronenberg,D., Michael,H., Schwarzer,K., Potapov,A., Choi,C. *et al.* (2006) TRANSPATH(r): an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.*, **3**, D546–D551.