CancerResource—updated database of cancer-relevant proteins, mutations and interacting drugs

Bjoern-Oliver Gohlke[1,2], Janette Nickel[1,2], Raik Otto[1], Mathias Dunkel[1] and Robert Preissner[1,2,*]

[1]Charité – University Medicine Berlin, Structural Bioinformatics Group, Institute of Physiology & Experimental Clinical Research Center, Berlin 13125, Germany and [2]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

ABSTRACT

Here, we present an updated version of CancerResource, freely available without registration at http://bioinformatics.charite.de/care. With upcoming information on target expression and mutations in patients' tumors, the need for systems supporting decisions on individual therapy is growing. This knowledge is based on numerous, experimentally validated drug-target interactions and supporting analyses such as measuring changes in gene expression using microarrays and HTS-efforts on cell lines. To enable a better overview about similar drug-target data and supporting information, a series of novel information connections are established and made available as described in the following. CancerResource contains about 91 000 drug-target relations, more than 2000 cancer cell lines and drug sensitivity data for about 50 000 drugs. CancerResource enables the capability of uploading external expression and mutation data and comparing them to the database's cell lines. Target genes and compounds are projected onto cancer-related pathways to get a better overview about how drug-target interactions benefit the treatment of cancer. Features like cellular fingerprints comprising of mutations, expression values and drug-sensitivity data can promote the understanding of genotype to drug sensitivity associations. Ultimately, these profiles can also be used to determine the most effective drug treatment for a cancer cell line most similar to a patient's tumor cells.

INTRODUCTION

According to the World Health Organisation cancer is one of the most common causes for human death and has been responsible for about 8.2 million cases of death worldwide in the year 2012 (http://www.who.int/mediacentre/factsheets/fs310/en/index312.html). To overcome difficulties in cancer therapy and to develop new methods for cancer diagnosis and treatment a huge amount of information is generated in cancer research experiments like in drug-target assays, high-throughput screenings on cancer cell lines or large-scale cancer genomics projects including next-generation sequencing studies (1–3).

In 2002 after the sequencing of the human genome, Hopkins and Groom established the term 'druggable genome' which comprises proteins that are known (or predicted) to interact with drugs. In their study they reveal an amount of 3051 druggable targets (4). Since then, novel drug targets have been identified that are relevant for cancer and which could be bound by compounds to provoke an activating or inhibiting molecular reaction, e.g. Superoxide dismutase 1 (SOD1). The overexpression of SOD1 results in lung cancer cells' growth and reduces apoptosis (5,6). It could be shown that its enzymatic activity was inhibited by compounds in lung cancer cells leading to growth inhibition of the cancer cells suggesting it as a promising target for cancer therapy (5).

Application of microarray-based gene expression data for cancer research is a broadly used method for identifying significant differentially expressed genes, compared to normal tissue or other cancer tissues, or for profiling cancer signatures, which can be associated with clinical outcome (7–9). Microarray-based gene expression data can even be considered for identifying new therapeutic targets (10) or biomarkers for specific cancer types (11).

Nowadays, as a result of the establishment of next-generation sequencing technologies and improved bioinformatical evaluation a better understanding of the genomic

*To whom correspondence should be addressed. Tel: +49 30 450 540 755; Fax: +49 30 450 540 955; Email: robert.preissner@charite.de

foundation of cancer was achieved (12). Given that cancer is a genetic disease, mutational characteristics of a cancer type can vary from patient to patient even though if the patients are affected by the apparently same cancer type. These genomic alterations might affect an anti-cancer drug's efficacy on the tumor and influence the clinical response. For instance, the anti-cancer drug vemurafenib improves the overall survival rate of patients having the BRAF V600E mutation (13). Consideration of genomic alterations in patients is part of personalized medicine (14) and enables the opportunity of an improved cancer diagnosis and anti-cancer therapy. Nevertheless, the analysis of these data and the understanding of the genotype-phenotype relationship between genomic alterations and anti-cancer drug response remains a major challenge in cancer research (15). The Cancer Genome Atlas (TCGA) project focuses on generating large-scale cancer genomics data sets which are stored by the cBio Cancer Genomics Portal (cBioPortal) which also provides further analysis tools (16).

To support, promote and gain a better insight into these data the updated CancerResource database links gene expression values, gene mutations as well as drug-sensitivity data to cell lines related to cancer. From the inclusion of the data from the consortia 'catalogue of somatic mutations in cancer' (CoSMIC) (17), 'Cancer Cell Line Encyclopedia' (CCLE) (18) and the CellMiner database (17) an explorative data analysis is enabled helping to achieve a better understanding of specific drug response in cancer.

MATERIALS AND METHODS

Gene expression, mutation and drug sensitivity data

The cancer cell line expression, mutation and drug sensitivity data are provided by the CCLE (18), CoSMIC (19) and CellMiner (17) websites, respectively.

All expression data are based on the Affymetrix HG-U133 Plus 2.0 technology. In order to increase the comparability of the expression data from three different sources, all gene expression values were scaled and centered. To determine the similarity between cancer cell lines two similarity measurements are used: first, the Pearson correlation distance; second, the percentage of a categorical classification of genes based on the fold change between the genes of two cancer cell lines.

Only somatic mutations were included in the similarity analyses and germline variants such as single nucleotide polymorphisms (SNPs) were excluded. The mutational status of a gene was modeled boolean: gene is mutated, a one was set; in case a wild-type was present, a zero was set. Similarity of two cancer cell lines based on their mutations is the amount of shared mutations, divided by all observed mutations of both cancer cell lines. The similarity calculations for mutations are accordingly based on the Tanimoto coefficient calculation (20).

Drug sensitivity data provided by CellMiner were measured as growth inhibition (GI50). Data provided by CCLE and CosMIC are given by IC50 and EC50 values. In case the primary target of a drug is known, the consortia provide information about them. The activity data of the consortia were scaled and centered to obtain a uniform format of the data. The normalized data were used to create the cellular

fingerprint for every compound considering each consortium separately when a compound occurred in more than one consortium (21).

All cancer cell lines from one consortium have been compared to each other. The *P*-value calculations are based on a normal distribution to the observed similarity for all cancer cell lines of the same consortium. Cancer cell lines, whose observed similarity differs from the expected percentage of similarity in either direction with a *P*-value of less or equal to 0.05 were counted as significantly similar or dissimilar.

Compound mapping and target identification

Protein targets for the CancerResource update have been obtained from the ChEMBL database v. 19 (22). Those targets were filtered using the following criteria: First, all interactions with an activity comment 'inactive', 'inconclusive' or 'not active' were removed. Second, only 'homo sapiens' was considered as organism and third, only protein target types were extracted. The structures of the CancerResource compounds and the ChEMBL compounds have been standardized using JChem (Instant Jchem version 14 10.27.0, ChemAxon (http://www.chemaxon.com/)) for identifying CancerResource compounds in ChEMBL. The standardization steps included aromatization of the structures and addition of explicit hydrogens. Furthermore, solvents and salts were removed. Additionally, 3D structures were generated. For comparing the standardized CancerResource compounds with the equally standardized ChEMBL compounds InChiKeys were calculated and used for compound identification. Furthermore, drug-target information from CTD (23), TTD (24), PharmGKB (25) and DrugBank (26) has been added to the database. The final data set includes 91 000 interactions whereat 11 000 compounds and 3400 targets are involved. Where available, 3D structures are linked via Cancer3D (27).

Compound similarity

The structural similarity search for uploaded structures is based on extended-connectivity fingerprints (ECFP). For the computation of the circular topological fingerprints the diameter, which defines the circular neighborhood considered for each atom, was set to 4 (ECFP4). The calculation of these fingerprints was performed by the cheminformatics toolkit of ChemAxon (JChem compr (14.10.20.0), 201n (2014), ChemAxon (http://www.chemaxon.com/)). All other parameters provided by ChemAxon were used in default configuration. To determine the similarity between the compounds stored in the database, the Tanimoto coefficient is calculated. The Tanimoto calculation on the website is performed by MyChem (http://mychem.sourceforge.net/).

Pathways

To achieve a better understanding of drug-target interactions at molecular level, KEGG (signaling) pathways (28) were analyzed according to their relevance in cancer emergence and cancer development. This set comprises cancer-specific pathways, pathways related to cell-cycle regulation, replication, immune response and drug metabolism. Pathway maps are dynamically retrieved via Web service from
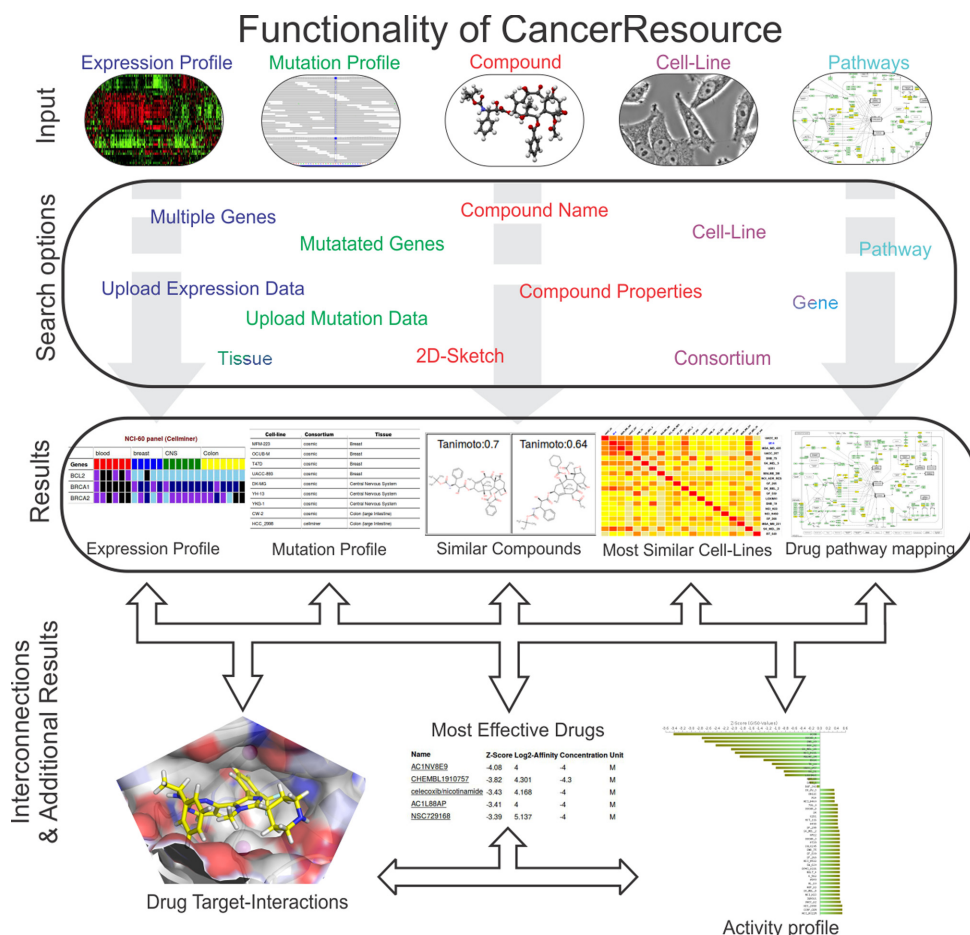
Figure 1. Overview of the functionality of CancerResource illustrating input variants, search options and result variants of the database.

KEGG facultative with highlighted expression data if gene expression is computed online before.

Server, database and system requirements

CancerResource is based on a relational MySQL database (http://www.mysql.com/). The database is normalized to the third normal form, for which large tables were split into smaller ones to minimize redundancy and dependency. The website of CancerResource is build using PHP (http://www.php.net/), JavaScript (http://www.java.com/), Ajax and web access is enabled via an Apache HTTP Server (http://apache.org/). For optimal usage we strongly recommend the latest version of Mozilla Firefox, Google Chrome or Safari browser and Internet Explorer, in descending order, with JavaScript option enabled.

RESULTS

CancerResource is comprised of about 50 000 compounds with detailed information like synonyms, structure identifier (SMILES, InChiKeys) as well as hydrogen bonds, molecular weight and log$P$ values. Additionally, the database provides links to PubChem. Furthermore, about 3400 protein targets could be identified for the compounds stored in the database. This results in about 91 000

compound-target interactions. By integrating CCLE and CoSMIC into CancerResource the total number of cancer cell lines now exceeds 2000. Mutation information for 19 834 genes, expression values for 23 016 genes and about 872 658 mutations from the consortia that were collected and included and are now available in CancerResource.

The CancerResource website provides different possibilities for the user to start using the database. Regardless of how the user begins a search, all results are interconnected via different joining's. An overview of the multiple search forms is displayed in Figure 1. These search options are described in more detail in the following sections.

Compound search for alternative, most effective drugs

CancerResource can easily be searched for compounds. For this purpose, two search categories are available. On the one hand, the database can be browsed by the compound's properties like name, molecular weight, number of atoms or log$P$ value. On the other hand, a connection to PubChem is established by which the user can search by compound name or SMILES. If no structures could be found via PubChem, a structure can be drawn by the molecule sketching tool. If the structure is available in the database detailed information about physicochemical properties, drug-target interactions, pathway information and an activity profile to all available

cancer cell lines are presented. Furthermore, the ten most similar compounds are listed for which detailed information can be displayed interactively.

Gene/target search

Targets can be found via different gene identifiers. Extended information about the target is displayed including expression profiles and drugs interacting with the target. Information about cancer cell lines, where the gene is mutated is also provided (mutation profile). If target-pathway mapping is possible, cancer relevant pathways are displayed.

Cell line/expression profile search

To search CancerResource for mRNA expression profiles of genes of interest different variants are prepared for the user. Besides searching for mRNA expression profiles of several selected genes between different cancer cell lines, the user can also upload an external expression profile to compare it to the database. As a result, the most similar cancer cell lines are presented to the user for which again detailed information can be displayed interactively. The results are accessible for at least one month by accessing the data via a bookmark of the web page.

A direct search for a specific cancer cell line or tissue type is also provided on the website. The results also embrace the most effective drugs against the cancer cell line as well as a tabular presentation of the similarity to other cancer cell lines based on a compound's activity-, mutation- and expression-fingerprint, respectively.

Mutation profile search

Searching for mutation profiles is enabled on the website. On the one hand, the user can search for gene mutations that occur in cancer-relevant genes. On the other hand, the user can search for gene mutations that occur in a cancer cell line of interest. Additionally, a tissue specific search is included. Furthermore, the opportunity to upload a mutation profile of user-provided tumor cells was implemented to compare them to well-characterized cancer cell lines and to identify the most similar cancer cell line based on mutations.

Pathway search

To provide a detailed insight into cancer relevant pathways a search by pathway names is provided. For this purpose pathway maps were extracted from the KEGG database. All targets of those maps for which compound-target interactions are available in CancerResource are highlighted. A mouse-over for those targets is made available displaying the binding compounds in a pop-up. Based on the chosen option, all cancer cell lines with certain mutated gene(s) or all mutations occurring in one specific cancer cell line are displayed.

From this site interactive browsing to detailed information, to cancer cell lines or genes is available to collect further information like most effective drugs or to compare the cancer cell line to others.

Similarity comparison

The user interface provides four ways to measure similarity of cancer cell lines. Similarity is calculated by the activity profile of compounds for two cancer cell lines, a Pearson distance correlation of expression values of genes, percentage similarity of known mutations of genes and a similarity measurement of a categorical classification of genes based on the fold changes between the genes of two cancer cell lines. Additionally, two options are provided to the user to calculate the similarity of own data to cancer cell line data from the three consortia. The results are presented in heat maps and differentially expressed genes are displayed.

Comparison to other databases

A detailed comparison to the original CancerResource (29), canSAR (30) and CancerDR (31) is represented in Table 1. The upload of external mutation data and expression values for cell lines or patient data to find the most similar cancer cell lines in the database is considered as a unique feature of the new version of CancerResource. Another feature of CancerResource is the mapping and annotation of cancer relevant protein targets to KEGG pathways. An additional extension of the updated CancerResource compared to other databases is the integration of mutation, expression and drug sensitivity data from the CCLE, CoSMIC and CellMiner consortia and to provide additionally a dynamic drug sensitivity comparison for external mutation or expression data. By this tool the user is enabled to create own hypotheses, which might possibly not have been developed by taking only one of the consortia into account.
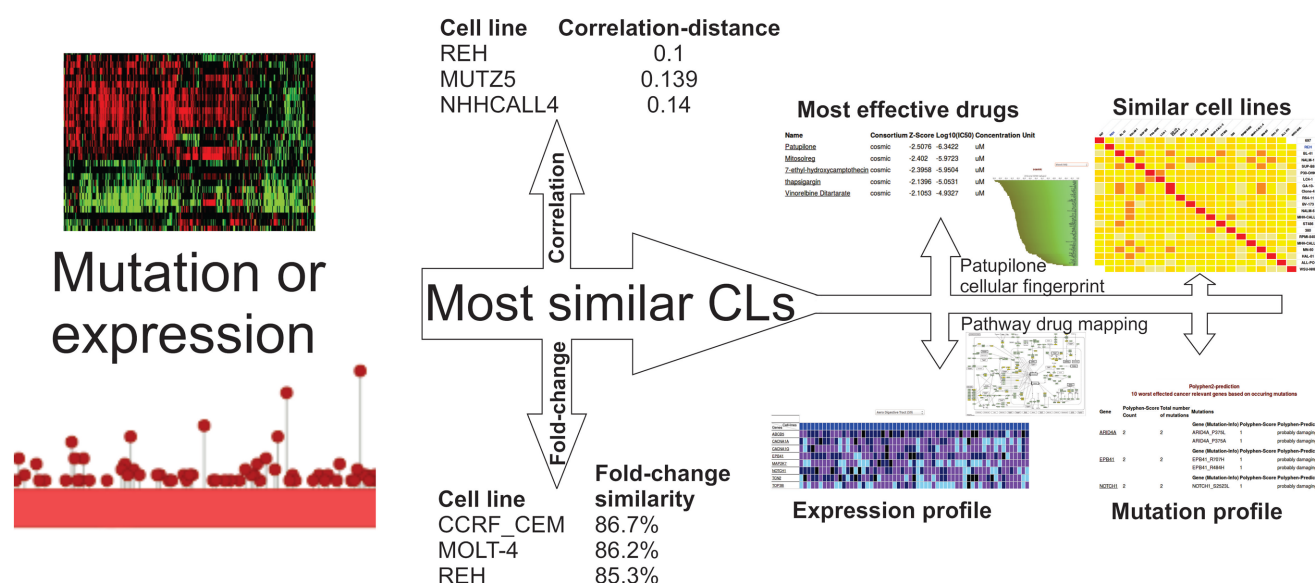
USE CASES

Upload of own gene expression or mutation data to find an alternative, most effective drug for a tumor similar to a cancer cell line

CancerResource can be used to identify the most similar cancer cell line of an external tissue sample by using either gene expression or mutation data. Therefore, normalized data from Human Genome U133A or Human Genome U133 Plus 2.0 microarray chips from Affymetrix can be queried either by Affymetrix probe set IDs or HGNC gene symbols. Based on the most similar cancer cell line the most effective drug for the external tissue is determined. To identify the most similar cancer cell line by applying gene expression data a Pearson correlation distance and additionally the percentage of the categorical classification of genes based on the fold change between the uploaded data and all cancer cell lines stored in CancerResource are calculated. Interconnections to most effected genes in the resulting cell line and additional similar cell lines are displayed as heatmaps. Alternatively, gene identifiers of mutated genes of the external tissue can be used to query the database for the most similar cancer cell line of the respective consortia. The next step gives an overview of the most effective drugs for the determined similar cancer cell line of the input data. A visualization of this use case is shown in Figure 2.

Table 1. Comparison of the updated CancerResource database with the original CancerResource, canSAR and CancerDR databases

	Update Cancer Resource	Cancer Resource (29)	canSAR (30)	CancerDR (31)
Expression, Mutation and Drug Sensitivity	All	Expression and drug sensitivity	All	Only mutation and drug sensitivity
Cellminer, CCLE and CoSMIC	All	Only CellMiner	All	Only CCLE and CoSMIC
Dynamic Drug Sensitivity Comparison	Yes	Yes	No	No
Pathways	Yes	Yes	Yes	Overview
Cell lines	2037	60	11 000	952
Drugs with Drug Sensitivity Data	48 404	≈40 000	16 000	148
Mutated Genes	19 799	No	Yes	Yes
No. of Genes	23 016	11 964	3466 studies	
Protein Targets	3387	2392	All	116
Integrated Similarity Measurements	Yes	Expression	No	No
Upload of external mutation data or expression values	Yes	Only expression	No	No

Figure 2. Use case—upload of external mutation or mRNA expression data to find similar cancer cell lines and alternative/most effective drugs for the external sample. Expression and mutation profiles for selected genes are available and in addition a mapping of genes to cancer relevant pathways is enabled.

Cancer cell line compound response

Almost 50 000 compounds that are stored in CancerResource have been screened against about 2000 cancer cell lines to determine their drug sensitivity. These data have been made available in the database. To identify which cancer cell line has a high or low sensitivity towards a compound, a similarity search is implemented within the database comparing a query compound to all compounds of CancerResource. If the query compound is found to be identical to a database compound, the user will directly be passed to the compound's details page. Otherwise, the ten most similar compounds found in the database are displayed. The 'similar property principle' (32) forms the bases for the similarity search and states that similar compounds might have similar properties in this case similar cancer cell line compound responses. By choosing one of the similar compounds the user will be forwarded to the compound's details page. On the details page of the identical or similar compound(s) the cellular fingerprint of the compound is displayed showing high and low responding cancer cell lines.

DISCUSSION AND OUTLOOK

Cancers, even from the same tissue, are extremely divergent in terms of gene alterations and therapy resistance. Therefore, individual therapy is required and will be made possible by understanding the entirety of single nucleotide polymorphisms (SNPs), complete or partial gene deletions, copy number variations, gene aberrations, gene fusions etc. All those issues may cause substantial dysfunctions or defected genes that have influence on gene regulation. The heterogeneity of tumors (33) and their reaction on chemotherapy (or other treatments) causes a further challenge, which will be addressed in a new release of CancerResource. Recently, tumor stratification not only based on somatic mutations but also on epigenetic changes like methylation has been

proven successful (34) and should become part of a future update to further support the development of personalized therapies. The data content of CancerResource is going to be updated in a yearly pattern based on the regular updates of the source data, which occur to be in a time period between three months and two years.

FUNDING

REFERENCES

1. Li,T., Ziniel,P.D., He,P.Q., Kommer,V.P., Crowther,G.J., He,M., Liu,Q., Van Voorhis,W.C., Williams,D.L. and Wang,M.W. (2015) High-throughput screening against thioredoxin glutathione reductase identifies novel inhibitors with potential therapeutic value for schistosomiasis. *Infect. Dis. Poverty*, **4**, 40.

2. Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, **6**, 813–823.

3. Tomczak,K., Czerwinska,P. and Wiznerowicz,M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contempor. Oncol. (Poznan, Poland)*, **19**, A68–A77.

4. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.

5. Somwar,R., Erdjument-Bromage,H., Larsson,E., Shum,D., Lockwood,W.W., Yang,G., Sander,C., Ouerfelli,O., Tempst,P.J., Djaballah,H. *et al.* (2011) Superoxide dismutase 1 (SOD1) is a target for a small molecule identified in a screen for inhibitors of the growth of lung adenocarcinoma cell lines. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16375–16380.

6. Papa,L., Manfredi,G. and Germain,D. (2014) SOD1, an unexpected novel target for cancer therapy. *Genes Cancer*, **5**, 15–21.

7. Axelsen,J.B., Lotem,J., Sachs,L. and Domany,E. (2007) Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 13122–13127.

8. Chibon,F. (2013) Cancer gene expression signatures—the rise and fall? *Eur. J. Cancer*, **49**, 2000–2009.

9. Sotiriou,C. and Pusztai,L. (2009) Gene-expression signatures in breast cancer. *N. Engl. J. Med.*, **360**, 790–800.

10. Sandhu,R., Parker,J.S., Jones,W.D., Livasy,C.A. and Coleman,W.B. (2010) Microarray-Based Gene Expression Profiling for Molecular Classification of Breast Cancer and Identification of New Targets for Therapy. *LabMedicine*, **41**, 364–372.

11. Kafetzopoulou,L.E., Boocock,D.J., Dhondalay,G.K., Powe,D.G. and Ball,G.R. (2013) Biomarker identification in breast cancer: Beta-adrenergic receptor signaling and pathways to therapeutic response. *Comput. Struct. Biotechnol. J.*, **6**, e201303003.

12. Meyerson,M., Gabriel,S. and Getz,G. (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.*, **11**, 685–696.

13. Chapman,P.B., Hauschild,A., Robert,C., Haanen,J.B., Ascierto,P., Larkin,J., Dummer,R., Garbe,C., Testori,A., Maio,M. *et al.* (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.*, **364**, 2507–2516.

14. Roychowdhury,S., Iyer,M.K., Robinson,D.R., Lonigro,R.J., Wu,Y.M., Cao,X., Kalyana-Sundaram,S., Sam,L., Balbin,O.A., Quist,M.J. *et al.* (2011) Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci. Transl. Med.*, **3**, 111–121.

15. Vuong,H., Cheng,F., Lin,C.C. and Zhao,Z. (2014) Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. *Genome Med.*, **6**, 81.

16. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.

17. Reinhold,W.C., Sunshine,M., Liu,H., Varma,S., Kohn,K.W., Morris,J., Doroshow,J. and Pommier,Y. (2012) CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.*, **72**, 3499–3511.

18. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehar,J., Kryukov,G.V., Sonkin,D. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.

19. Forbes,S.A., Beare,D., Gunasekaran,P., Leung,K., Bindal,N., Boutselakis,H., Ding,M., Bamford,S., Cole,C., Ward,S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.

20. Willett,P., Barnard,J.M. and Downs,G.M. (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, **38**, 983–996.

21. Fullbeck,M., Dunkel,M., Hossbach,J., Daniel,P.T. and Preissner,R. (2009) Cellular fingerprints: a novel approach using large-scale cancer cell line data for the identification of potential anticancer agents. *Chem. Biol. Drug Design*, **74**, 439–448.

22. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Kruger,F.A., Light,Y., Mak,L., McGlinchey,S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.

23. Davis,A.P., Grondin,C.J., Lennon-Hopkins,K., Saraceni-Richards,C., Sciaky,D., King,B.L., Wiegers,T.C. and Mattingly,C.J. (2015) The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.*, **43**, D914–D920.

24. Zhu,F., Shi,Z., Qin,C., Tao,L., Liu,X., Xu,F., Zhang,L., Song,Y., Liu,X., Zhang,J. *et al.* (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.*, **40**, D1128–D1136.

25. Whirl-Carrillo,M., McDonagh,E.M., Hebert,J.M., Gong,L., Sangkuhl,K., Thorn,C.F., Altman,R.B. and Klein,T.E. (2012) Pharmacogenomics knowledge for personalized medicine. *Clin. Pharm. Ther.*, **92**, 414–417.

26. Law,V., Knox,C., Djoumbou,Y., Jewison,T., Guo,A.C., Liu,Y., Maciejewski,A., Arndt,D., Wilson,M., Neveu,V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.

27. Porta-Pardo,E., Hrabe,T. and Godzik,A. (2015) Cancer3D: understanding cancer mutations through protein structures. *Nucleic Acids Res.*, **43**, D968–D973.

28. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.

29. Ahmed,J., Meinel,T., Dunkel,M., Murgueitio,M.S., Adams,R., Blasse,C., Eckert,A., Preissner,S. and Preissner,R. (2011) CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.*, **39**, D960–D967.

30. Bulusu,K.C., Tym,J.E., Coker,E.A., Schierz,A.C. and Al-Lazikani,B. (2014) canSAR: updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res.*, **42**, D1040–D1047.

31. Kumar,R., Chaudhary,K., Gupta,S., Singh,H., Kumar,S., Gautam,A., Kapoor,P. and Raghava,G.P. (2013) CancerDR: cancer drug resistance database. *Sci. Rep.*, **3**, 1445.

32. Johnson,M.A. and Maggiora,G.M. (1990) *Concepts and applications of molecular similarity*. John Willey & Sons, NY.

33. Marusyk,A. and Polyak,K. (2010) Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta*, **1805**, 105–117.

34. Cebrian,V., Fierro,M., Orenes-Pinero,E., Grau,L., Moya,P., Ecke,T., Alvarez,M., Gil,M., Algaba,F., Bellmunt,J. *et al.* (2011) KISS1 methylation and expression as tumor stratification biomarkers and clinical outcome prognosticators for bladder cancer patients. *Am. J. Pathol.*, **179**, 540–546.