OPEN

Citation: Cell Death and Disease (2014) 5, e1051; doi:10.1038/cddis.2014.9

www.nature.com/cddis

DRUGSURV: a resource for repositioning of approved and experimental drugs in oncology based on patient survival information

I Amelio[1], M Gostev[2], RA Knight[1], AE Willis[1], G Melino[1,3] and AV Antonov*[,1]

The use of existing drugs for new therapeutic applications, commonly referred to as drug repositioning, is a way for fast and cost-efficient drug discovery. Drug repositioning in oncology is commonly initiated by *in vitro* experimental evidence that a drug exhibits anticancer cytotoxicity. Any independent verification that the observed effects *in vitro* may be valid in a clinical setting, and that the drug could potentially affect patient survival *in vivo* is of paramount importance. Despite considerable recent efforts in computational drug repositioning, none of the studies have considered patient survival information in modelling the potential of existing/new drugs in the management of cancer. Therefore, we have developed DRUGSURV; this is the first computational tool to estimate the potential effects of a drug using patient survival information derived from clinical cancer expression data sets. DRUGSURV provides statistical evidence that a drug can affect survival outcome in particular clinical conditions to justify further investigation of the drug anticancer potential and to guide clinical trial design. DRUGSURV covers both approved drugs (~1700) as well as experimental drugs (~5000) and is freely available at http://www.bioprofiling.de/drugsurv.
Cell Death and Disease (2014) **5,** e1051; doi:10.1038/cddis.2014.9; published online 6 February 2014
Subject Category: Cancer

Drug repositioning (application of approved drugs to new therapeutic indications) is currently widely used because of the reduced development costs and simplicity of drug approval procedure. The availability of a vast amount of experimental data covering various diseases has stimulated computational efforts to identify novel potential indications for established drugs.[1–5] The computational principles of drug repositioning are based on a polypharmacology paradigm:[6] the drugs are considered in the context of all proteins (genes) affected upon treatment (i.e., the drug signature), and specific diseases are modelled by the multiple genes involved/perturbed in the disease state (i.e., disease signature). Significant similarity between drug and disease signatures is indicative of the potential application of the drug to treat the disease (Figure 1).[1–3]

Gene expression data were a primary source of information used by most computational approaches. The sets of genes that are up- and downregulated in a disease state compared with a normal state were used as a gene signature of a disease.[1,2,4] On the other hand, expression data from human cell lines treated with a broad range of approved drugs has been used to derive genes affected by the drugs. The linkage between a drug and a disease is computed as similarity between the drug and the disease gene signatures.[1,2] Different studies varied in principles to compute similarity. Some of them additionally incorporate gene pathway information.[2]

In oncology, effect on patient survival outcome is a key criterion of drug efficiency in clinical trials. However, none of the studies have considered patient survival information in modelling the potential of existing/new drugs in the management of cancer. Therefore, we have developed DRUGSURV; this is the first computational tool to estimate the potential effects of a drug using patient survival information derived from clinical cancer expression data sets. In contrast to other approaches, DRUGSURV uses genes significantly associated (P-value <0.01) with patient survival as a cancer signature specific for a cancer type or clinical condition studied in a particular data set (Figure 2b). At the moment, DRUGSURV covers 44 independent clinical cancer expression data sets (in most cases each data set contains >100 patients annotated with survival information).

DRUGSURV covers both FDA approved drugs (~1700) and experimental drugs (~5000). The coverage of drugs by DRUGSURV significantly exceeds any previous efforts in the field. Drug signature is defined based on known drug targets. This information is integrated from DrugBank[7] and Pubchem Bioassays[8] databases. The proteins that are known targets of a drug, or involved in the drug transport/metabolism, or have been reported to be inhibited by the drug in high-throughput screening chemical assays (Pubchem Bioassays) are referred to as direct drug targets (Figure 2a). We also use the term indirect drug targets to refer to the proteins that

[1]Medical Research Council Toxicology Unit, Leicester University, Leicester, UK; [2]Wellcome Trust Genome Campus, EBI, Hinxton, Cambridge, UK and [3]Department of Experimental Medicine and Surgery, University of Rome 'Tor Vergata', Rome, Italy
*Corresponding author: AV Antonov, Medical Research Council Toxicology Unit, Leicester University, Hodgkin Building, Lancaster Road, PO Box 138, Leicester LE1 9HN, UK. Tel: +44 (0)116 252 5562; Fax: +44 (0)116 252 5616; E-mail: aa668@le.ac.uk
Keywords: drug repositioning; clinical trial design; patient survival; thioridazine
Abbreviations: EGFR, epidermal growth factor receptor; DRUGSURV, computational tool to estimate the potential effects of a drug using patient survival information
Received 29.9.13; revised 11.12.13; accepted 19.12.13; Edited by M Piacentini

interact with the direct drug targets according to the IntAct database.[9]

DRUGSURV is incorporated in the bioprofiling.de analytical portal for high-throughput cell biology[10] and is freely available at http://www.bioprofiling.de/drugsurv. DRUGSURV provides multiple query options to explore systematically the effect of genes, which are known to be modulated upon drug treatment on survival in different cancers and clinical conditions. The user can query interested drug, specific cancer or explore any gene as a potential anticancer target. We demonstrate that DRUGSURV validates therapeutic indications for known cancer drugs. DRUGSURV also suggests that the antipsychotic agent, thioridazine, recently demonstrated *in vitro* to selectively target cancer stem cells,[11] could also be effective *in vivo*: there is a significant proportion of thioridazine targets associated with patient survival in several cancer expression data sets.

Results

Thioridazine: antipsychotic to anticancer agent. Originally thioridazine was positioned as a phenothiazine antipsychotic and has been used in the management of psychoses, including schizophrenia, and in the control of severely disturbed or agitated behaviour. It has been widely accepted that thioridazine blocks postsynaptic mesolimbic dopaminergic D1 and D2 receptors in the brain, blocks alpha-adrenergic effects, depresses the release of hypothalamic and hypophyseal hormones and is believed to depress the reticular activating system.[7]

Very recently, thioridazine was shown to selectively target cancer stem cells.[11] Thioridazine reduced the ability of human acute myeloid leukaemia samples to proliferate and to self-renew, as shown by a decrease in both the ability of the treated cells to form colonies *in vitro* and in the efficiency of transplantation into recipient mice.[12] The anticancer properties of thioridazine have also been shown in several other previous studies,[13] but thioridazine may become particularly important because the selective targeting of cancer stem cells offers promise for a new generation of therapeutics with anticancer potential.[12]

Thioridazine is known to act through dopamine receptors and this was a primary hypothesis while searching for a mechanism for thioridazine's anticancer activity.[11,12] Data from recent high-throughput screens indicate that thioridazine inhibits about 20 proteins, which are considered to be off target, including those that are known to be associated with tumour progression, such as EGFR. First, this suggests that thioridazine modulates more genes than previously considered. Second, DRUGSURV shows that a statistically significant proportion of these indirect targets affect patient survival in various expression data sets derived from various cancers (Table 1).

The results in Table 1 provide additional independent statistical evidence that thioridazine could have potential therapeutic effects in patients. For example, in the 'chronic lymphocytic leukaemia' data set, 86 (out of 502) indirect thioridazine targets are significantly associated with survival. In the 'multiple myeloma' data set, 55 (out of 502) indirect thioridazine targets are significantly associated with survival. DRUGSURV visualization of the 'drug-data set' model (Figure 3) simplifies our understanding of the potential anticancer mechanism of thioridazine and suggests that a major impact of thioridazine on cancer could be mediated by interaction with *EGFR* and *FYN* genes. Although expression

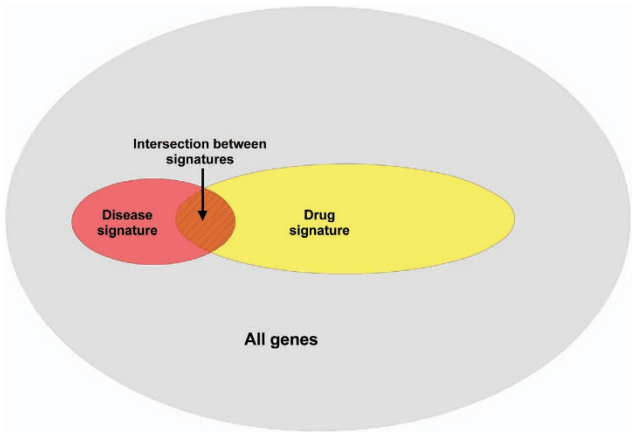

Figure 1 Computational principles of drug repositioning. Drugs are considered in the context of all proteins (genes) affected upon treatment (i.e., the drug signature). Disease is modelled by genes involved/perturbed in the disease state. Significant similarity (intersection between drug signature and disease signature) is indicative of the potential application of the drug to treat the disease

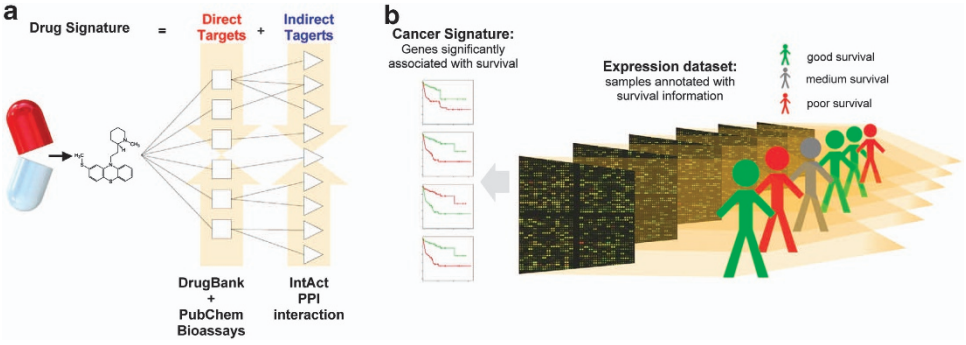

Figure 2 DRUGSURV data mining principles. (**a**) Drug signature is derived based on DrugBank, Pubchem BioAssays and IntAct databases. (**b**) Cancer signature (specific for each data set) is derived based on genes significantly (*P*-value <0.01) associated with survival in the data set. Each data set models specific for cancer type or clinical conditions (i.e. cancer stage, status)

Table 1 Cancer expression data sets significantly (FDR adjusted *P*-value <0.01) associated with thioridazine indirect targets

GEO Data set	Cancer type	*P*-value, FDR adjusted	Odds ratio	*k (I)*	*m (N)*
Prediction of survival in diffuse large B-cell lymphoma treated with chemotherapy plus rituximab	Diffuse large B-cell lymphoma	0.00019 (4.34e-06)	1.35	179 (502)	5432 (20 387)
Expression data from untreated CLL patients	Chronic lymphocytic leukaemia	0.00021 (9.56e-06)	1.61	86 (502)	2200 (20 386)
Molecular subclasses of high-grade glioma: prognosis, disease progression, and neurogenesis	High-grade glioma	0.0024 (0.00021)	1.64	58 (468)	1000 (12 940)
Subtype classification, grading, and outcome prediction of urothelial carcinomas by combined mRNA profiling and aCGH	Urothelial carcinomas	0.0024 (0.00021)	3.66	12 (340)	114 (10 911)
MAQC-II project: multiple myeloma data set	Multiple myeloma	0.0047 (0.00053)	1.60	55 (502)	1416 (20 387)
Validation cohort for genomic predictor of response and survival following neoadjuvant taxane-anthracycline chemotherapy in breast cancer	Breast cancer	0.00702 (0.00095)	1.61	49 (468)	858 (12 940)
Whole-transcript expression data for liposarcoma	Liposarcoma	0.0071 (0.0011)	1.38	90 (468)	1827 (12 940)
Experimentally derived metastasis gene expression profile predicts recurrence and death in colon cancer patients	Colon cancer	0.0098 (0.0017)	1.55	50 (502)	1326 (20 387)

Abbreviation: FDR, false discovery rate.
The last two columns (*k* (*I*), *m* (*N*)) report statistical details of association, *k* denotes the number of drug targets (genes) significantly associated (*P*-value <0.01) with survival in the data set, *I* denotes the overall number of indirect drug targets, *m* denotes the overall number of genes significantly associated with survival in the data set and *N* denotes the overall number of genes measured in the data set.

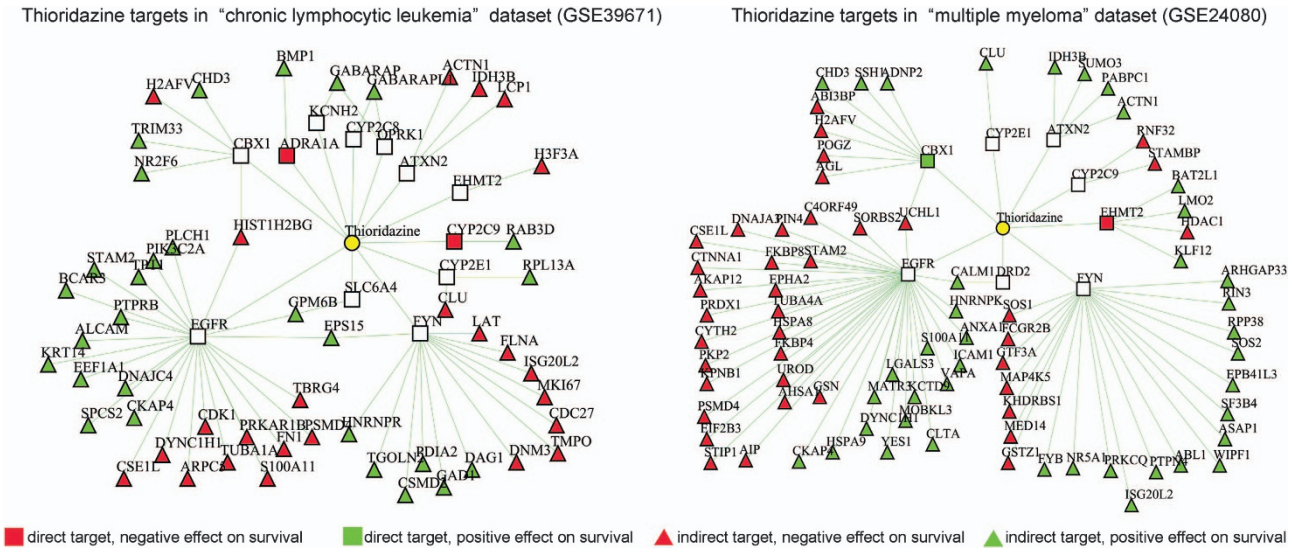

Figure 3 Visual output of DRUGSURV for 'drug-data set' models for thioridazine. Rectangles denote direct drug targets, triangles correspond to indirect targets. Colours indicate effect of gene overexpression on survival. In several available data sets, genes significantly associated with survival are overrepresented among thioridazine indirect targets

of *EGFR* and *FYN* genes are rarely associated with survival directly, both *EGFR* and *FYN* interact with multiple genes, which do affect survival in patients with chronic lymphocytic leukaemia and multiple myeloma.

DRUGSURV: validation therapeutic indications for known cancer drugs. Breast cancer is one of the most well-studied cancer types. DRUGSURV incorporates 17 independent clinical expression breast cancer data sets, which model various specific clinical conditions. We used breast cancer as an example to demonstrate that DRUGSURV validates therapeutic indications for well-established cancer drugs.

Among the top 10 drugs suggested by DRUGSURV (based on the indirect drug targets) to be potential breast cancer treatments, 6 are well-stablished anticancer drugs (Table 2). Tamoxifen and mitoxantrone are currently commonly used for the treatment of breast cancer, whereas danazol is used for the treatment of benign breast disorders (which are important risk factors for breast cancer[14]), and has been tested in clinical trials for the treatment of advanced breast cancer. It was concluded that danazol is an effective agent in patients with advanced breast cancer, but the response rate is inferior to that of other agents, such as tamoxifen.[15]

Sunitinib, erlotinib and sorafenib are tyrosine kinase inhibitors, which have been approved for the treatment of

Table 2 Drugs associated (FDR adjusted *P*-value <0.01) with at least with 10 independent breast cancer expression data sets ('*indirect drug targets*')

Drug	Drug type	Number of associated data sets	Known anticancer agent
Danazol	*Approved*	13	Yes
Sunitinib	*Approved*	12	Yes
Sorafenib	*Approved*	12	Yes
Mitoxantrone	*Approved*	10	Yes
Tamoxifen	*Approved*	10	Yes
Erlotinib	*Approved*	10	Yes
Bithionol	*Withdrawn*	10	No
Hexachlorophene	*Approved*	10	No
Vitamin A	*Approved*	10	No

Abbreviation: FDR, false discovery rate

different solid tumours. However, none of them has been approved for the treatment of breast cancer, although multiple preclinical studies have suggested their potential as likely breast cancer agents in human patients. For example, erlotinib was reported to inhibit tumour cell proliferation in hormone receptor-positive breast cancer and to induce breast cancer regression.[16,17] Sorafenib has been assessed in phase IIB trials with Capecitabine for locally advanced or metastatic human epidermal growth factor receptor 2 (HER2)-negative breast cancer. Addition of sorafenib to capecitabine improved progression-free survival in patients with HER2-negative advanced breast cancer, although with unacceptable toxicity for many patients.[18]

Sunitinib has demonstrated potential for the treatment of breast cancer in multiple preclinical studies, involving the human breast cancer MX-1 xenograft model, where in combination with docetaxel, doxorubicin or fluorouracil it enhanced the antitumour activity of the chemotherapeutic agents and increased survival.[19] Sunitinib also inhibited osteolysis and tumour growth in a mouse model of breast cancer metastatic to bone.[20] However, Sunitinib failed in a randomized phase III study, which investigated whether sunitinib plus docetaxel improved clinical outcomes for patients with (HER2)/neu-negative advanced breast cancer *versus* docetaxel alone.[21] Interestingly, DRUGSURV is able to predict this outcome. The only breast cancer data set where indirect targets of sunitinib were depleted among genes associated with survival is data set GSE3521, which investigated patients with distant metastases and poor outcomes. Patients in the data set were annotated with HER2 status and 72% of them were HER2-negative. Therefore, DRUGSURV indicates that clinical conditions modelled in the data set GSE3521 at the molecular level involve genes that are not modulated by sunitinib and, therefore, treatment with sunitinib is not expected to result in any benefit.

Finally, bithionol, hexachlorophene and vitamin A are three top-rated drugs by DRUGSURV, which have never been used as anticancer agents. Hexachlorophene is a chlorinated bisphenol antiseptic with a bacteriostatic action against Gram-positive organisms. Bithionol was shown to cause serious skin disorders and was withdrawn from the market in 1967. Both hexachlorophene and bithionol were reported to exhibit anticancer cell cytotoxicity[22,23] but have never been extensively studied for anticancer properties.

Discussion

In most studies, the novel anticancer therapeutic effect of new/established drugs is usually demonstrated *in vitro*, and there will always remain doubt whether the anticancer potential is still manifest *in vivo*. Clinical trials are very expensive and time consuming, but remain the only way to validate drug efficiency *in vivo*. Before embarking on the time and expense of a clinical trial, however, any additional, and more easily obtainable, evidence that the observed drug effect *in vitro* will also be observed (or not) *in vivo* would be of paramount importance. DRUGSURV is a tool, which is likely to provide such statistical evidence.

In contrast to other similar studies, DRUGSURV exploits patient survival information. In oncology, the effect on patient survival outcome is a key criterion of drug efficiency. From this standpoint, modelling cancer signatures with genes that are significantly associated with survival is more direct in comparison to previous approaches. Availability of data sets that model very specific clinical conditions provides a possibility to estimate drug efficiency in patients with specific cancer subtypes. For example, DRUGSURV would be able to predict the inefficiency of sunitinib in patients with (HER2)/neu-negative advanced breast cancer (see Results).

DRUGSURV implements as 'drug signature' known direct drug targets inferred from DrugBank and PubChem data. Previous studies have inferred drug signatures from databases containing gene expression data for cell lines treated with drugs (e.g., connectivity maps[24]). In this case, drug signatures are biased in relation to the cell cultures, which have been used in the experiments, and could contain multiple response genes, which are not drug specific.[25] In addition, multiple statistical issues exist as to how to determine precise estimates of statistical significance and false-positive rates.[24,25] Drug signatures implemented in DRUGSURV do not have these limitations, although for many drugs our current knowledge about targets is incomplete. Therefore, in these cases, the genes that are affected upon drug treatment are modelled only partially. Finally, the number of drugs covered by the connectivity map pilot project, for example, is only 164, whereas DRUGSURV covers both FDA approved drugs ($\sim$1700) and experimental drugs ($\sim$5000). We would like to emphasise that the coverage of drugs by DRUGSURV significantly exceeds any previous efforts in the field.

DRUGSURV provides multiple query options. The user can interrogate interested drug, specific cancer or explore any gene as a potential anticancer target. At present, DRUGSURV covers 44 independent clinical cancer expression data sets (in most cases each data set contains >100 patients annotated with survival information). DRUGSURV is regularly updated as new expression data sets become available[26] to cover novel cancer types or specific clinical conditions as well as to update information on drug targets.

Finally, we must caution that this kind of statistical inference (the limitation also applies to all previous and most probably to all future similar studies) is based on simplified assumptions that all genes from both signatures (drug and cancer) are

weighted equally (or could be weighted based on some data or assumptions). There might be cases when the modulation of one gene might be more important than modulation of many other genes.

Materials and Methods

Cancer expression data sets. Gene expression data sets were downloaded from the Gene Expression Omnibus repository.[26] To be selected, the data set must be a clinical (patients) microarray expression data set with at least 70 samples and annotated with patient survival data. At present, DRUGSURV covers more than 40 data sets.

Cancer survival gene signature. For each available data set, we computed the set of genes whose up/downregulation is associated (P-value < 0.01) with patient survival. Gene expression rank reflects relative mRNA expression level and is more consistent as it requires no normalization and thus introduces no normalization bias. For each gene in the data sets, samples were grouped with respect to expression rank of the gene.[27,28] The 'Low expression' and 'High expression' groups are those where the expression rank of the gene of interest is less or more than average expression rank across the data set, respectively. Standard statistical tests[29] were used to find any statistical differences in survival outcome between the 'Low expression' and 'High expression' patient groups. Genes those split patients in groups with significant differences (P-value < 0.01) in outcome were selected as a cancer gene signature specific for the clinical conditions studied in the data set.

Direct drug targets. The set of genes (derived based on the set of proteins) that are indicated in DrugBank[7] as drug target, drug transporter or drug-metabolizing enzyme is defined as direct drug targets. In addition, we used Pubchem Bioassay repository.[8] Reference to the Pubchem Bioassay repository means that the drug was tested in an HTS assay and was found to inhibit the activity of the tested protein.

Indirect drug targets. Indirect drug targets, along with direct drug targets, are proteins which interact with the direct drug targets based on the records of the IntAct database of protein–protein interactions.[9]

Linking statistically 'drug targets' with 'cancer survival gene signature'. Let us denote l to be the number of targets (either direct or indirect) for drug B and $k$of them associated with survival (P-value < 0.01) in the data set A. The rate k/l reflects the proportion of the drug B targets associated with survival. The rate $k/l$is compared with the rate m/N, where m is the total number of genes significantly associated with survival in the data set A and $N$is a number of all genes measured in the data set A. A standard Hypogeometric test (with parameters k, l, m, N) is applied to derive the P-value of enrichment. The same procedure is repeated across all available data sets. Finally, derived P-values (Hypogeometric) are adjusted for multiple testing using false discovery rate control procedure[30,31] (the number of hypotheses tested is equal to the number data sets available).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements. This work was supported by the UK Medical Research Council (MRC) and funding from Russian Federal grants 14.B37.21.1967 (to AA) and 11.G34.31.0069 (to GM).

1. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011; **3**: 96ra77.
2. Jin G, Fu C, Zhao H, Cui K, Chang J, Wong ST. A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. *Cancer Res* 2012; **72**: 33–44.
3. Li YY, Jones SJ. Drug repositioning for personalized medicine. *Genome Med* 2012; **4**: 27.
4. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 2010; **107**: 14621–14626.
5. Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, Slikker W Jr *et al. In silico* drug repositioning: what we need to know. *Drug Discov Today* 2013; **18**: 110–115.
6. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008; **4**: 682–690.
7. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 2011; **39**: D1035–D1041.
8. Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO *et al.* An overview of the PubChem BioAssay resource. *Nucleic Acids Res* 2010; **38**: D255–D266.
9. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012; **40**: D841–D846.
10. Antonov AV. BioProfiling.de: analytical web portal for high-throughput cell biology. *Nucleic Acids Res* 2011; **39**: W323–W327.
11. Sachlos E, Risueño RM, Laronde S, Shapovalova Z, Lee JH, Russell J *et al.* Identification of drugs including a dopamine receptor antagonist that selectively target cancer stem cells. *Cell* 2012; **149**: 1284–1297.
12. Burgess DJ. Stem cells. Antipsychotic to anticancer agent? *Nat Rev Cancer* 2012; **12**: 452–453.
13. Strobl JS, Peterson VA. Tamoxifen-resistant human breast cancer cell growth: inhibition by thioridazine, pimozide and the calmodulin antagonist, W-13. *J Pharmacol Exp Ther* 1992; **263**: 186–193.
14. Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K *et al.* Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005; **353**: 229–237.
15. Coombes RC, Perez D, Gazet JC, Ford HT, Powles TJ. Danazol treatment for advanced breast cancer. *Cancer Chemother Pharmacol* 1983; **10**: 194–195.
16. Catania C, De Pas TM, Pelosi G, Manzotti M, Adamoli L, Nolè F *et al.* Erlotinib-induced breast cancer regression. *Ann Pharmacother* 2006; **40**: 2043–2047.
17. Guix M, Granja Nde M, Meszoely I, Adkins TB, Wieman BM, Frierson KE *et al.* Short preoperative treatment with erlotinib inhibits tumor cell proliferation in hormone receptor-positive breast cancers. *J Clin Oncol* 2008; **26**: 897–906.
18. Baselga J, Segalla JG, Roché H, Del Giglio A, Pinczowski H, Ciruelos EM *et al.* Sorafenib in combination with capecitabine: an oral regimen for patients with HER2-negative locally advanced or metastatic breast cancer. *J Clin Oncol* 2012; **30**: 1484–1491.
19. Abrams TJ, Murray LJ, Pesenti E, Holway VW, Colombo T, Lee LB *et al.* Preclinical evaluation of the tyrosine kinase inhibitor SU11248 as a single agent and in combination with "standard of care" therapeutic agents for the treatment of breast cancer. *Mol Cancer Ther* 2003; **2**: 1011–1021.
20. Murray LJ, Abrams TJ, Long KR, Ngai TJ, Olson LM, Hong W *et al.* SU11248 inhibits tumor growth and CSF-1R-dependent osteolysis in an experimental breast cancer bone metastasis model. *Clin Exp Metastasis* 2003; **20**: 757–766.
21. Bergh J, Bondarenko IM, Lichinitser MR, Liljegren A, Greil R, Voytko NL *et al.* First-line treatment of advanced breast cancer with sunitinib in combination with docetaxel versus docetaxel alone: results of a prospective, randomized phase III study. *J Clin Oncol* 2012; **30**: 921–929.
22. Zhai L, Zamani A, Liu X, Wang Y, Maxuitenko Y, Alvarez R *et al.* Abstract 3727: Targeting autotaxin to reduce chemotherapy resistance in ovarian cancer. *Cancer Res* 2012; **72,** Supplement 1.
23. Park S, Gwak J, Cho M, Song T, Won J, Kim DE *et al.* Hexachlorophene inhibits Wnt/beta-catenin pathway by promoting Siah-mediated beta-catenin degradation. *Mol Pharmacol* 2006; **70**: 960–966.
24. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; **313**: 1929–1935.
25. Lamb J. The Connectivity Map: a new tool for biomedical research. *Nature reviews. Cancer* 2007; **7**: 54–60.
26. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF *et al.* NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res* 2011; **39**: D1005–D1010.
27. Antonov AV, Knight RA, Melino G, Barlev NA, Tsvetkov PO. MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ* 2013; **20**: 367.
28. Antonov AV, Krestyaninova M, Knight RA, Rodchenkov I, Melino G, Barlev NA. PPISURV: a novel bioinformatics tool for uncovering the hidden role of specific genes in cancer survival outcome. *Oncogene*; e-pub ahead of print 30 May 2013; doi:10.1038/onc.2013.119.
29. D H, T F. A class of rank test procedures for censored survival data. *Biometrika* 1982; **69**: 553–566.
30. Antonov AV, Dietmann S, Wong P, Mewes HW. TICL–a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. *FEBS J* 2009; **276**: 2084–2094.
31. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Statistics in medicine* 1990; **9**: 811–818.