W148–W156 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* *Published online 25 May 2018*
doi: 10.1093/nar/gky434

GDA, a web-based tool for Genomics and Drugs integrated analysis

Jimmy Caroli[1], Giovanni Sorrentino[2], Mattia Forcato[1], Giannino Del Sal[3,4] and Silvio Bicciato[1,*]

[1]Dept. of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi, 287, 41125 Modena, Italy, [2]Laboratory of Metabolic Signaling, Ecole Polytechnique Fédérale de Lausanne EPFL, Lausanne, Switzerland, [3]Laboratorio Nazionale CIB, Area Science Park Padriciano, Trieste, Italy and [4]Dept. of Life Sciences, University of Trieste, Trieste, Italy

Received January 30, 2018; Revised May 03, 2018; Editorial Decision May 03, 2018; Accepted May 08, 2018

ABSTRACT

Several major screenings of genetic profiling and drug testing in cancer cell lines proved that the integration of genomic portraits and compound activities is effective in discovering new genetic markers of drug sensitivity and clinically relevant anticancer compounds. Despite most genetic and drug response data are publicly available, the availability of user-friendly tools for their integrative analysis remains limited, thus hampering an effective exploitation of this information. Here, we present GDA, a web-based tool for Genomics and Drugs integrated Analysis that combines drug response data for >50 800 compounds with mutations and gene expression profiles across 73 cancer cell lines. Genomic and pharmacological data are integrated through a modular architecture that allows users to identify compounds active towards cancer cell lines bearing a specific genomic background and, conversely, the mutational or transcriptional status of cells responding or not-responding to a specific compound. Results are presented through intuitive graphical representations and supplemented with information obtained from public repositories. As both personalized targeted therapies and drug-repurposing are gaining increasing attention, GDA represents a resource to formulate hypotheses on the interplay between genomic traits and drug response in cancer. GDA is freely available at http://gda.unimore.it/.

INTRODUCTION

Since the pioneering NCI-60 panel (1,2), diverse cancer cell lines have been used in large-scale screenings of combined small-molecule sensitivity and genomic profiling to investigate how genetic backgrounds and transcriptional portraits shape cancer response to therapy and to identify disease-specific genes associated with drug response (3–7). Historically, pharmacogenomics screenings have been largely heterogeneous in terms of investigated cell lines, assay technologies, number of compounds, type and quality of genomic data and methods for their computational analysis. In particular, the NCI-60 study screened >50 800 compounds on a panel of about 60 cell lines characterized with genetic, transcriptional and phenotypic data (2,8–10). Differently, the Cancer Cell Line Encyclopedia (CCLE; (4)) and the Genomics of Drug Sensitivity in Cancer (GDSC; (3,11)) projects profiled the genome of >1000 cancer cell lines, but probed both collections against small sets of anticancer drugs (24 and 265, respectively). Finally, more recent studies selected CCLE subsets for screening hundreds of small molecules and compounds (5,6). The analysis of this enormous and heterogeneous amount of data requires computational methods that, through common data formats, robust statistics, and user-friendly interfaces, allow the integration of genomic profiles with drug responses across multiple screenings.

Here, we present GDA (Genomics and Drugs integrated Analysis), a web-based tool for the integrative analysis of drug response data, mutations, and gene expression profiles in a panel of 73 cancer cell lines treated with 50 816 compounds. GDA builds on our previously published Mutation and Drug Portal (MDP; (12)) that was developed to match response data of the NCI-60 DTP drug screening with mutations from the CCLE and NCI-60 profiling. Briefly, MDP offered the possibility to overcome the limited number of molecules investigated in the CCLE study by correlating CCLE genomic data to the NCI-60 DTP large panel of drug responses. In its original version, MDP could only be queried for discovering associations between gene mutations and drug families with growth-inhibitory effects on cancer cell lines bearing those mutations or to identify the mutational background of cancer cell lines responsive (or non-responsive) to a given compound. Both types of queries

*To whom correspondence should be addressed. Tel: +39 59 2055 219; Fax: +39 59 2055 410; Email: silvio.bicciato@unimore.it

could be performed using the variant data for 1651 oncogenes from CCLE or the whole-exome sequencing of 15 000 human genes from the NCI-60 repository. Although MDP proved its efficacy in retrieving both known and novel pharmacogenomics associations between gene mutations and responses of mutated cell lines towards precise compounds, still the absence of gene expression data represented a major limitation to identify multiple levels of interactions between drug responses and genomic determinants. To overcome this limitation, we developed GDA, a web server that adds the integrative analysis of transcriptional profiles and drug response data to MDP original functionalities. Specifically, starting form a list of genes, GDA can be queried to identify drugs showing activity towards cells with a defined transcriptional portrait. Vice versa, starting from a compound, GDA retrieves gene signatures that differentiate responsive from non-responsive cell lines. These gene signatures can be directly functionally annotated using Enrichr (13), compared to results from the Library of Integrated Network-Based Cellular Signatures (LINCS) L1000 project (14,15), or used to identify drugs with growth-inhibitory effects. Furthermore, to support the design of novel anticancer molecules, we implemented a structural similarity analysis to verify the existence of a shared, common structure among compounds active in cells with a specific genomic background. All analysis modules are accessible through a user-friendly graphical interface that does not require any programming skill and results are returned as intuitive graphical representations and downloadable tables. In this manuscript, we summarize the modules for the identification of drugs correlated to gene mutations (*from gene to drug*) and of gene mutations associated to drug response (*from drug to gene*), already implemented in MDP, and present the major novelties introduced by GDA, i.e. the *from signature to drug* and the *from drug to signature* modules, the structural clustering of significant drugs, and the differential analysis of gene expression levels in cell lines responsive or non-responsive to a given compound. A step-by-step guide to a complete analysis in GDA is reported in the Supplementary Information.

WEBSERVER CONTENT AND ARCHITECTURE

The GDA webserver comprises drug response, exome-sequencing, and gene expression profiling data from the NCI-60 and the CCLE datasets, 2D compound structures from PubChem, a modular web interface for querying data and visualizing results, and a backend based on Python and R for the statistical analyses.

Database content

Drug response data of 115 cancer cell lines treated with 50 816 compounds were derived from the NCI-60 GI_{50} file (https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition+Data; September 2014 release) and transformed into relative sensitivities (RS) as described in (16). Briefly, the GI_{50} value, i.e. the drug concentration required for 50% growth inhibition in vitro as named by NCI-60 DTP (https://dtp.cancer.gov/databases_tools/docs/compare/compare_methodology.htm), was transformed into relative sensitivities by mean centering, in logarithmic scale, the GI_{50} of each compound in each cell line (i.e. $RS = \log_2 GI_{50} - \text{average}(\log_2 GI_{50})$) where the average is taken across all cell lines). Based on the RS values, each combination of drug and cell line was classified as responsive (or non-responsive) if the RS was lower (or higher) than two standard deviations of the distribution of all RS in the given cell line (16,17). Mutation calls for the CCLE panel were retrieved from the CCLE portal (https://portals.broadinstitute.org/ccle/data) and exome sequencing data for the NCI-60 cell lines were obtained from CellMiner (https://discover.nci.nih.gov/cellminer/loadDownload.do). Raw gene expression data of CCLE and NCI-60 cell lines were downloaded from Gene Expression Omnibus (GEO) series GSE36133 and GSE32474, respectively. Prior to analysis, we merged the two series in a compendium of 231 samples representing the 73 cell lines that had at least one sample in at least one series. The compendium comprises at least four replicate samples for >80% of the different cell lines. Expression values were generated from intensity signals using the multi-array average procedure (RMA) and a custom CDF based on Entrez genes for Affymetrix HG-U133 Plus2 arrays (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp).
Drug structures were retrieved from the PubChem FTP site (ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/CURRENT-Full/SDF/) in SMILES (Simplified Molecular-Input Line-Entry) format and matched to NCI-60 drug names using the CID (PubChem Compound Identification) name format. The central objects of the database, used to link NCI-60 drug responses to CCLE and NCI-60 genomic data, are the cancer cell lines in common between the NCI-60 drug screening and the CCLE and NCI-60 genomic repositories (Figure 1).

Analysis modules

GDA is composed of four main analysis modules that allow identifying (i) drugs active in cancer cell lines bearing specific gene mutations (*from gene to drug*); (ii) gene mutations characterizing cancer cell lines that are responsive to a selected compound (*from drug to gene*); iii) drugs active in cancer cell lines bearing the activation of a specific gene signature (*from signature to drug*); and iv) up- and down-regulated genes in cancer cell lines that respond to a specific compound (*from drug to signature*) (Figure 1). Queries are performed through drop-down menus and either checkboxes or radio buttons, depending on the type of input. In the *from gene to drug* and *from drug to gene* modules, genes and compound are selected via a drop-down menu that auto-completes based on the gene mutations and drugs present in the database. In the *from signature to drug*, gene lists can be pasted into a dedicated input text box or uploaded as a text file using HUGO symbols for the gene names.

In the *from gene to drug* analysis, given a set of mutations, compounds are ranked based on an enrichment score given by the fraction of cell lines bearing the set of mutations that are responsive multiplied by the fraction of non-mutated cell lines that are non-responsive (12). The statisti-

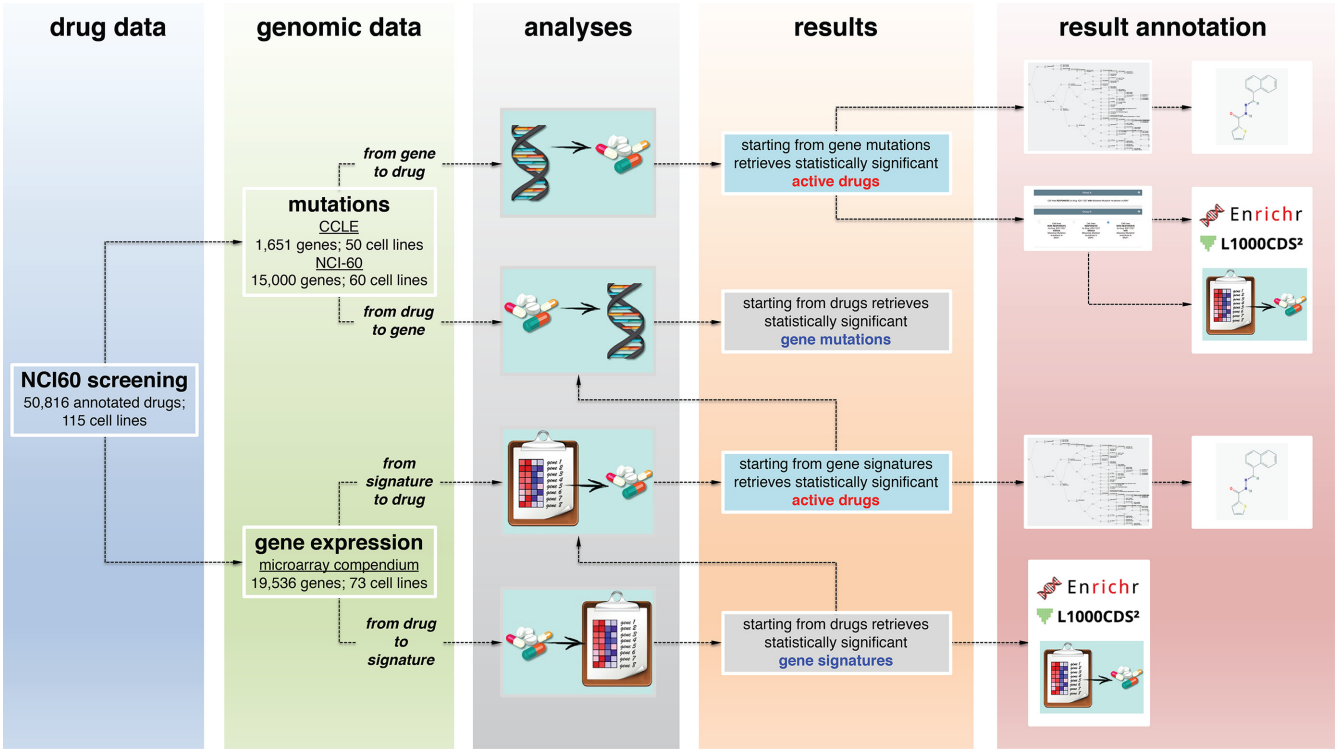

Figure 1. Overview of GDA data analysis workflow. GDA is based on the pharmacological data obtained from the NCI-60 screening for a total of 50,816 compounds on 115 cancer cell lines and on the genomic and transcriptional profiles of the CCLE and NCI-60 studies. GDA can be interrogated through four main modules to identify drugs active in cancer cell lines bearing specific gene mutations (*from gene to drug*); gene mutations characterizing cancer cell lines that are responsive to a selected compound (*from drug to gene*); drugs active in cancer cell lines bearing the activation of a specific gene signature (*from signature to drug*); and up- and down-regulated genes in cancer cell lines that respond to a specific compound (*from drug to signature*). Results from the analyses can be fed into additional GDA modules (as the drug clustering, the Maximum Common Structure, and the differential gene expression analyses) or sent to external web services (as Enrichr, L1000CDS^2 and PubChem) for functional annotation and comparison.

cal significance (*P*-value) is computed, for each drug, using a one-tailed Fisher's exact test for the enrichment of mutant and responsive as compared to wild-type and non-responsive cell lines, given the number of non-responses and responses in mutant and wild-type cells, respectively (12). The same ranking function and statistical test is used to identify the most enriched mutations starting from a drug in the *from drug to gene* module. In this case, the RS values of the queried compound are used to partition cells lines in responsive and non-responsive and then each gene mutation in each cell line is ranked based on the enrichment score and the statistical significance defined in the *from gene to drug* module (12). Both *from gene to drug* and *from drug to gene* queries can be run either on the CCLE genomic data, containing variant calls for a limited set of 1651 oncogenes, or on the NCI-60 exome sequencing data for 15 000 genes.

The *from signature to drug* and *from drug to signature* modules exploit the transcriptional data of the gene expression compendium. In the *from signature to drug*, cell lines are first split into two groups (i.e. high- and low-signature) based on the sign of a signature score quantified as the sum of the standardized expression values of genes composing the input gene list (18). Then, similarly to the mutation and drug modules, each drug is ranked according to the score and statistical significance of the enrichment of responses in the group of cell lines with either high or low signature (depending if the gene signature is considered up- or down-regulated). In the *from drug to signature* module, first cell lines are split into responsive and non-responsive based on the RS of the queried compound and then, if each group contains at least three samples, the transcriptional profiles of responsive and non-responsive cell lines can be analyzed using either SAM (Significance Analysis of Microarray; (19)) or a *t*-test to determine the list of up- and down-regulated genes in responsive and non-responsive cells.

From the result page of the *from gene to drug*, it is possible to access two additional modules, i.e. the drugs clustering and, once a drug is selected, the differential expression analysis. The drug clustering (available also in the *from signature to drug* result page) groups active compounds based on the similarity of their chemical structures. Briefly, first the SMILES structures of statistically significant compounds (e.g. with enrichment score ≥ 0.3 and *P*-value ≤ 0.05) are converted into structural coordinates (coded in the structural data information files); then, drugs are grouped based on their structural coordinates using a hierarchical agglomerative clustering with Pearson correlation as distance metric and average agglomeration method (as implemented in the function *hclust* of the *R stats* package). Once clustered, a group of compounds can be analyzed to identify a shared molecular scaffold using the Maximum Common Structure (MCS) algorithm of the *fmcs* package (https://bitbucket.org/dalke/fmcs; (20)). The differential ex-

pression module allows comparing the gene expression levels in the cancer cell lines responsive to a statistically significant drug and bearing a specific gene mutation (Group A) against the gene expression levels of cell lines (Group B) that are either (i) non-responsive to the selected drug and lacking the mutation in the selected gene; (ii) responsive to the selected drug even if lacking the mutation in the selected gene or (iii) non-responsive to the selected drug although bearing the mutation in the selected gene. As in the *from drug to signature* module, differential expression analysis can be run using SAM (Significance Analysis of Microarray; (19)) or a parametric test (*t*-test).

Output description

The various modules generate different result pages. Specifically, the *from gene to drug* analysis returns the list of all drugs that are active on the analyzed cell lines bearing the selected set of mutations and the list of drug families that are significantly enriched, given the set of active drugs (Figure 2A). In the table of active drugs, molecules are identified in terms of compound ID (linking to PubChem), name, drug family, mechanism of action (MoA), score and statistical significance. Results can be downloaded in tabular form for storage and external analyses, as well as visualized using different graphical representations. The plots display (i) the score and *P*-value of statistically significant compounds; (ii) the distribution of compound scores grouped by drug family; (iii) the distribution of relative sensitivity in mutant/responsive and wild-type/non-responsive cell lines and (iv) the gene expression level of the selected gene in mutant/responsive and wild-type/non-responsive cell lines (Figure 2B). The *from signature to drug* module returns the same outputs of the *from gene to drug* module listing all drugs that are active on cell lines with the up- or down-regulation of the input gene signature. The result page of the *from drug to gene* module reports the chemical structure of the input compound with the number of mutations found statistically correlated to response and no-response; an interactive volcano plot showing score and *P*-value of gene mutations present in cancer cell lines that are responsive (non-responsive) to the selected compound; the distribution of mutations in cell lines, tissues, and variant types; the list of all mutations that have been found present in cancer cell lines responsive (non-responsive) to the action of the selected compound, along with their functional and structural annotations. The output page of the *from drug to signature* analysis provides the lists of genes over-expressed in responsive (Group A) and non-responsive (Group B) cancer cell lines. These lists can be functionally annotated using Enrichr (13), compared to results of the LINCS project through the LINCS L1000 characteristic direction signatures search engine (L1000CDS^2; (14)), or directly used to generate gene signatures for the *from signature to drug* module of GDA (Figure 3A). Finally, the output page of the *drug clustering* (accessible from the *from gene to drug* and *from signature to drug* modules) provides an interactive clustering tree of all significant drugs grouped by structural similarity (Figure 3B). Once selected, each node of the tree returns a list of molecules that can be further used in the Maximum Common Structure analysis to retrieve a common scaffold shared by all compounds belonging to the group.

Web implementation

GDA is based on the lightweight and flexible Twig template engine for PHP (https://twig.symfony.com/) for a fast crosstalk between web pages and analysis modules. The web interface is structured with PHP 5.0, HTML 5.0, and a completely customized CSS for buttons and animations. Statistical analyses are performed using Python 2.7 cgi scripts that have been optimized for fast fetching and data delivery and are structured to interact with R (version 3.1.2). For the graphical representations of results, we used the jQuery JavaScript library 2.1.1 (https://jquery.com/), based on the structure of the latest Google APIs release (2016), and the Plotly JavaScript library 1.35.2 (https://plot.ly/javascript/). The connections to Enrichr and L1000CDS^2 have been coded using the javascript APIs provided by Enrichr (http://amp.pharm.mssm.edu/Enrichr/help#api) and a Python 2.7 cgi script (http://amp.pharm.mssm.edu/L1000CDS2/help/#python), respectively. GDA is hosted on a Linux server with 512 GB RAM and 64 processors, can work on any recently updated browser, and is designed to record all performed analyses for a fast fetching of results (i.e. results are immediately loaded if the query of interest has been submitted before).

CASE STUDY RESULTS

The tutorial, available as Supplementary Information and on the GDA website, provides representative case studies for each of the four analysis modules, including sample inputs and working examples. Here, we will briefly present how GDA can integrate drug responses, mutations, and gene expression profiles to formulate hypotheses on the mechanisms connected to drug resistance or to elucidate the interplay between the transcriptional activation of signaling pathways and drug response.

Identification of no-response mechanisms to BRAF inhibitors

Cancers bearing somatic BRAF mutations, and in particular the V600E mutation within the kinase domain, are treated with BRAF inhibitors like Dabrafenib (21) and Vemurafenib (22), but although these therapies improved overall and disease-free survival, nonetheless almost 20% of the patients do not respond to the treatment. To discover putative mechanisms of drug resistance in cancers with BRAF mutations, we first identified drugs active in cancer cell lines with BRAF missense mutations using the *from gene to drug* module. The analysis returned, among the most active compounds, the Imatinib analogs Dabrafenib (21), Vemurafenib (22), and Selumetinib (Figure 2A). Then, to search for genes likely interfering with the anticancer action of e.g. Vemurafenib, we investigated (using SAM differential expression analysis and responsive cells with the mutation as control group; Group A) the gene expression profiles of cell lines that, although bearing BRAF missense mutations, do not respond to the treatment with Vemurafenib (Group B). The comparison indicated that cells

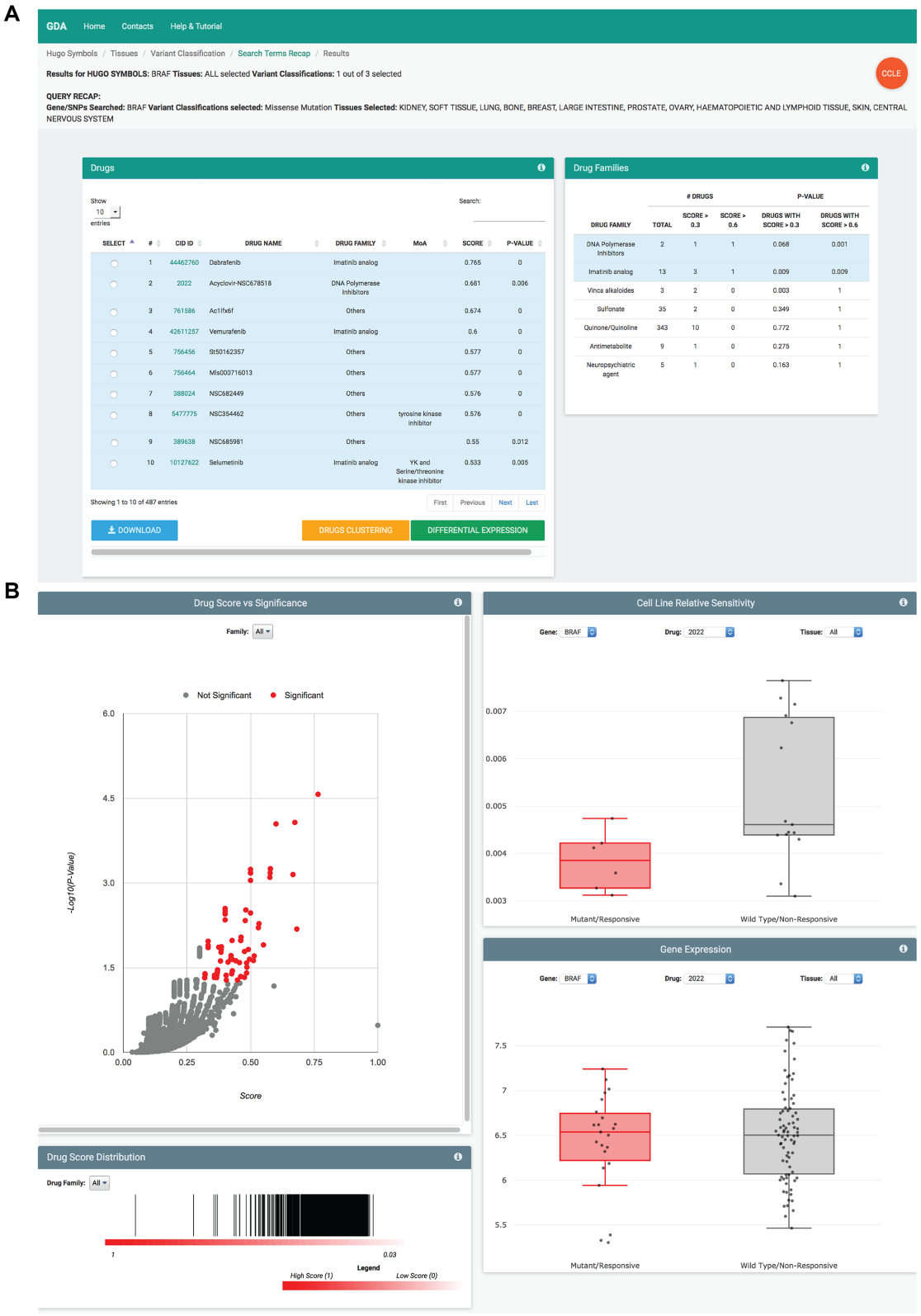

Figure 2. Result pages of the *from gene to drug* and *from signature to drug* modules. (**A**) The result pages of the *from gene to drug* and *from signature to drug* analyses list of all drugs that are active on cell lines bearing the selected genomic background. In the table, drugs are identified in terms of compound ID (linking to the PubChem), name, drug family, mechanism of action (MoA), score and statistical significance. (**B**) Results can be visualized in terms of (from top left to bottom right): score and *P*-value of statistically significant compounds; distribution of compound scores grouped by drug family; distribution of relative sensitivity in mutant/responsive and wild-type/non-responsive cell lines; gene expression level of the selected gene in mutant/responsive and wild-type/non-responsive cell lines.

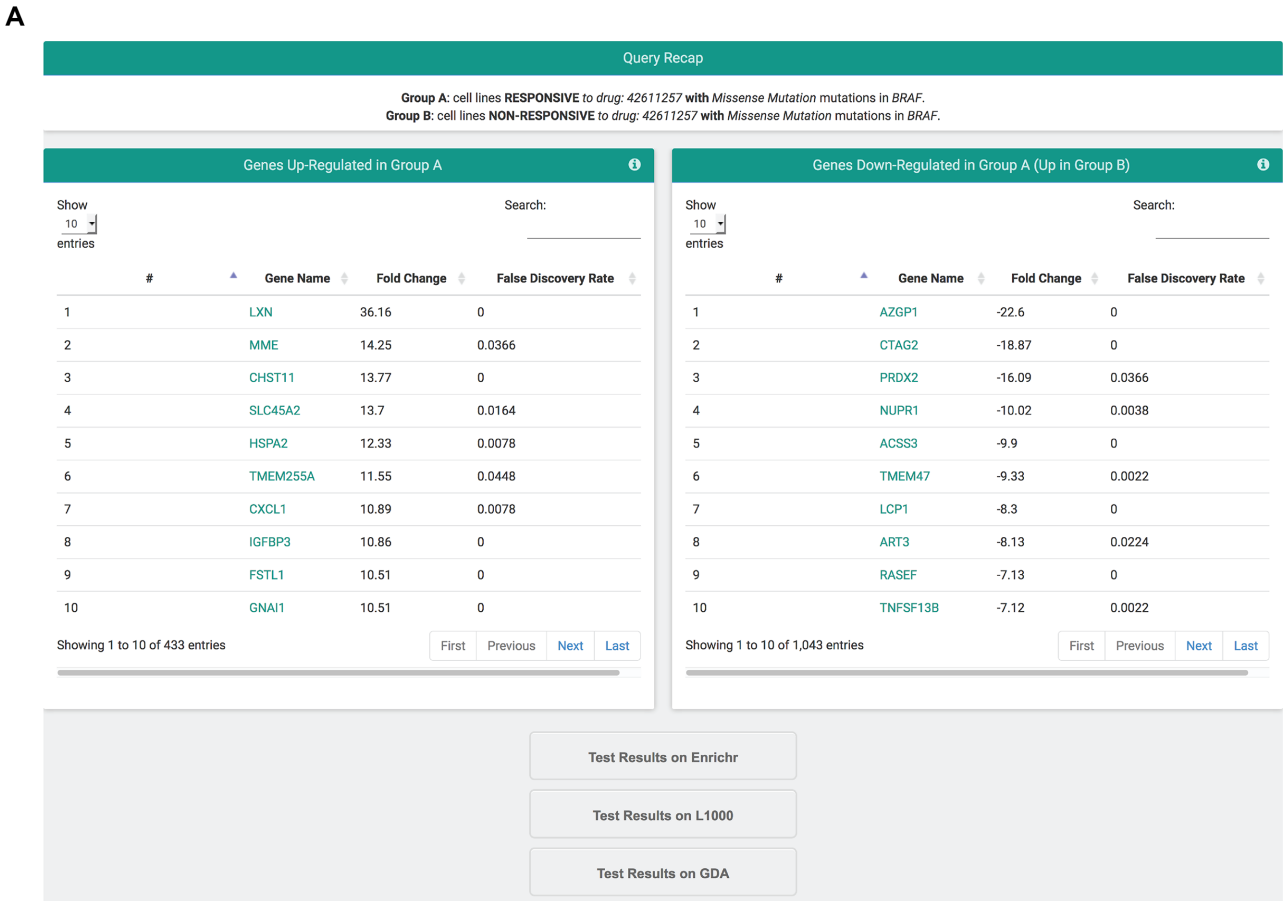

Figure 3. Result pages of the *from drug to signature* and *drug clustering* modules. (**A**) The output page of the *from drug to signature* analysis provides the lists of genes over-expressed in responsive (Group A) and non-responsive (Group B) cancer cell lines. These lists can be linked to external web services or directly used to generate gene signatures for the *from signature to drug* module of GDA. (**B**) The output page of the drug clustering returns an interactive clustering tree of all significant drugs grouped by structural similarity.

non-responsive to Vemurafenib over-express (at FDR $\leq$ 0.05) a total of 1043 genes and that those with a fold change $\geq$2 (316 genes) are functionally enriched in the KEGG pathway of insulin resistance, ultimately mediated by activation of the PI3K/Akt pathway. Intriguingly, this result supports previous evidences of a mechanistic link between insulin, the PI3K/Akt pathway, and attenuated therapeutic efficacy of BRAF inhibitors and suggests that the defective response to Vemurafenib might be overcome by the concomitant use of PI3K inhibitors (23).

Interplay between transcriptional activation of signaling pathways and drug sensitivity

The phosphatase and tensin homolog (PTEN) protein is a phosphatase that inhibits the phosphoinositol-3-kinase (PI3K)/AKT signaling pathway and suppresses cell survival as well as cell proliferation. Loss of function mutations in the PTEN gene have been associated with a wide range of human tumors, making PTEN the second most frequently mutated gene in human cancers after TP53. Transcriptional signatures of PTEN have been used to investigate the molecular consequences of PTEN loss in cancer. We took advantage of a loss-of-function PTEN signature (PTEN_DN.V1_UP gene set of the Molecular Signatures Database; http://software.broadinstitute.org/gsea/msigdb/index.jsp) obtained from isogenic tumor cell line pairs in which PTEN was inactivated through RNA interference (24) to identify drugs with selective cytotoxicity for cancer cells showing PTEN inactivation. To this aim, we used the *from signature to drug analysis* and, among the compounds predicted to be active on all tissues, we found several mitochondrial inhibitors (as Leucinostatin-A, Ossamycin, Cytovaricin, Oligomycin-A, Oligomycin-B, and Oligomycin-C; Figure 4A) that have been recently demonstrated to be selective cytotoxic drugs for PTEN-null cancer cells (25). Conversely, we used the *from drug to signature* module to identify signaling pathways active in cells responsive to mitochondrial inhibitors. The functional enrichment of over-expressed genes in cells responsive to Oligomycins or Trichopolyn-B (*Up in Group A* at FDR $<$ 0.05) indicated that sensitivity to mitochondrial inhibitors is accompanied by the activation of oxidative phosphorylation, respiratory electron transport chain, and ATP synthesis pathways (Figure 4B), thus suggesting a possible mechanism of action of these compounds in cancer cells (25).

COMPARISON TO OTHER SIMILAR TOOLS

GDA can be compared to some other tools for the integrative analysis of pharmacological and genomic data in cancer cell lines, as for instance CellMiner and CellMiner Cross Database (CellMinerCDB) (8,9); CancerResource (26); CancerDR (27); the Sanger Genomics of Drug Sensitivity in Cancer (GDSC) data portal (11) and the Cancer Therapeutics Response Portal (5–7). However, among these tools, only CellMinerCDB and CancerResource, as GDA, integrate drug activity and molecular data obtained from both CCLE and NCI-60 DTP consortia thus allowing to cross-interrogate, e.g. the extensive pharmacological collection of the NCI-60 on the genomic data of the CCLE profiling. Although lacking the data of the Sanger GDSC project, included in both CellMinerCDB and CancerResource, GDA still present some advantages over both tools in terms of analysis options, usage simplicity, and interactive visualization of results. CellMinerCDB (https://discover.nci.nih.gov/cellminercdb/) is still in a developmental phase and the analyses are limited to the visualization of univariate associations and regressions between predefined entities (e.g. one gene and one compound). CancerResource (http://data-analysis.charite.de/care/index.php) contains no module for a direct integrative analysis of genomic features and pharmacological data. Instead, mutations and gene signatures are used to identify a set of cell lines that share similar genomic traits and the most effective drugs are then listed for any single cell lines, thus missing the direct association between a specific genomic background and the response to a drug, irrespective of the tissue type. Differently from similar tools, GDA (i) directly integrates genomic and pharmacological data from the two largest screenings of combined small-molecule sensitivity and genomic profiling in cancer cell lines; (ii) allows both the identification of compounds active towards cancer cell lines with given mutational or transcriptional traits and the mutational or transcriptional portraits of cells responsive or non-responsive to a specific compound; (iii) has a completely interlaced modular architecture that allows feeding results from one analysis as input to other modules without any external, manual rearrangement of the data and (iv) returns results through a set of intuitive graphical representations and tabular formats, supplemented with direct link to public repositories. A unique feature of GDA is the gene expression compendium. Given the high concordance between CCLE and NCI-60 gene expression measurements (28,29), we generated gene expression data merging the two transcriptional studies, thus obtaining a number of replicates per single cell line that allows statistically robust differential analyses between any subgroup defined, e.g. on drug response or signature activation. As compared to MDP (12), GDA presents several major enhancements and updates including: (i) the addition of the gene expression data and of all related modules; (ii) the drug clustering and Maximum Common Structure analyses to test the existence of a shared, common structure among active compounds and (iii) a re-designed interface comprising entirely new graphical representation and annotation of results.

CONCLUSIONS AND OUTLOOK

Large-scale screenings of genomic profiles and pharmacological responses provide a unique resource to formulate hypotheses on the interplay between genomic traits and drug sensitivity or resistance in cancer. However, to be extensively exploited by the scientific community, these data require computational tools that combine robust integrative analyses with easy-to-use, user-friendly interfaces. GDA is a webserver designed to facilitate the integrative analysis of genomic and drug response data available from two major cancer cell line screens (i.e. the NCI-60 and the Cancer Cell Line Encyclopedia) for the systematic identification of new biomarkers of drug sensitivity and the selection of putative therapeutic options for patients that, based on their

A

Drugs Table

Show 10 entries

Search:

SELECT	#	COMPOUND ID	NAME	DRUG FAMILY	MoA	SCORE	P-VALUE
○	1	76455	3055-84-3	Others	--	0.63	0.001
○	2	5477807	Leucinostatin-A	Others	--	0.578	0
○	3	5351598	Ossamycin	Others	--	0.534	0.004
○	4	5471806	NSC713799	Others	--	0.534	0.004
○	5	5477715	Cytovaricin	Others	--	0.534	0.004
○	6	54608021	NSC606307	Others	--	0.534	0.004
○	7	382162	NSC669610	Others	--	0.526	0.002
○	8	5472285	Oligomycin-A	Others	--	0.512	0.007
○	9	5472286	Oligomycin-B	Others	--	0.512	0.007
○	10	5472287	Oligomycin-C	Others	--	0.512	0.007

Showing 1 to 10 of 456 entries

First Previous Next Last

B

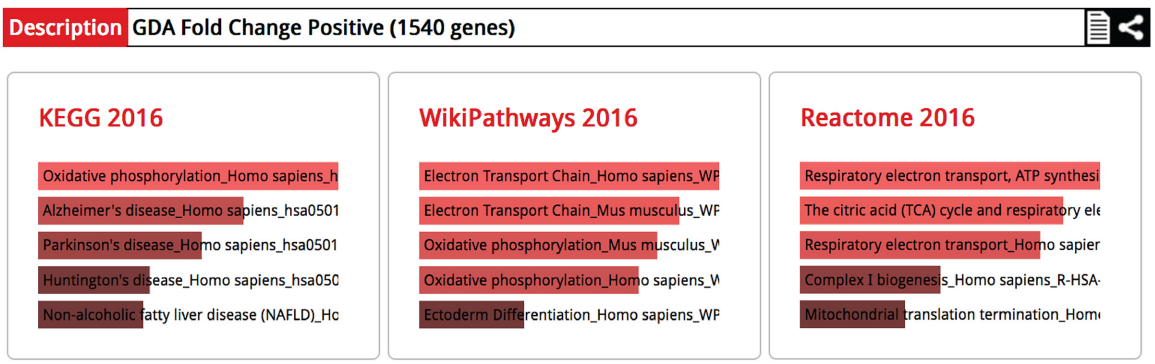

Figure 4. Transcriptional activation of signaling pathways and drug sensitivity. (**A**) The *from signature to drug* analysis indicates that cancer cells showing PTEN inactivation are responsive to mitochondrial inhibitors. (**B**) Genes identified by the *from drug to signature* module as up-regulated in cells responsive to mitochondrial inhibitors (as e.g. Oligomycin-A at FDR < 0.05) are functionally enriched in the activation of oxidative phosphorylation, respiratory electron transport chain and ATP synthesis pathways.

genomic background, fail to respond to standard therapies. A key feature of the webserver is its user-friendly web interface that does not require any bioinformatics expertise nor manual formatting of the data to input the various, intertwined analysis modules. Given some reported discrepancies between drug sensitivity measurements in the various studies (28,29), the current version of GDA incorporates only the pharmacological data of the NCI-60 project. Nonetheless, its modular architecture can, in principle, host genomic and pharmacological information from any other screening as, for instance, epigenomic profiles and sensitivities to specific drugs or combinations of drugs of patient-derived models (as cell cultures, xenografts and organoids) directly interrogated as *in vitro* proxy of human tumors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

AIRC Special Program Molecular Clinical Oncology '5 per mille' [10016 to G.D.S., S.B.]; AIRC IG [17659 to G.D.S.]; Regione FVG [ex art. 15 L.R. 17/2014 (RiFT, TNBC-neo, PerMiD) and ex art. 20 L.R. 20/2015 (TuMaGiDo) to G.D.S.]; Italian Ministry of Education, University and Research and the National Research Council grant Italian Epigenomics Flagship Project (Epigen) (to S.B.); Italian Ministry of Education, University and Research [PRIN-2015-8KZKE3 to G.D.S.]; 'RAN-translation of normal and expanded nucleotide repeat containing transcripts to neurotoxic polypetides in neurodegenerative diseases' from Fondazione Cariplo call 2014; European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program [670126-DENOVOSTEM to S.B.]. Funding for open access charge: Fondazione Cariplo call 2014 (to S.B.).

Conflict of interest statement. None declared.

REFERENCES

1. Weinstein,J.N., Myers,T.G., O'Connor,P.M., Friend,S.H., Fornace,A.J., Kohn,K.W., Fojo,T., Bates,S.E., Rubinstein,L. V, Anderson,N.L. *et al.* (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science*, **275**, 343–349.
2. Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, **6**, 813–823.
3. Garnett,M.J., Edelman,E.J., Heidorn,S.J., Greenman,C.D., Dastur,A., Lau,K.W., Greninger,P., Thompson,I.R., Luo,X., Soares,J. *et al.* (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, **483**, 570–575.
4. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehár,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.
5. Basu,A., Bodycombe,N.E., Cheah,J.H., Price,E. V., Liu,K., Schaefer,G.I., Ebright,R.Y., Stewart,M.L., Ito,D., Wang,S. *et al.* (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, **154**, 1151–1161.
6. Rees,M.G., Seashore-Ludlow,B., Cheah,J.H., Adams,D.J., Price,E. V, Gill,S., Javaid,S., Coletti,M.E., Jones,V.L., Bodycombe,N.E. *et al.* (2016) Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.*, **12**, 109–116.
7. Seashore-Ludlow,B., Rees,M.G., Cheah,J.H., Cokol,M., Price,E. V, Coletti,M.E., Jones,V., Bodycombe,N.E., Soule,C.K., Gould,J. *et al.* (2015) Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.*, **5**, 1210–1223.
8. Shankavaram,U.T., Varma,S., Kane,D., Sunshine,M., Chary,K.K., Reinhold,W.C., Pommier,Y. and Weinstein,J.N. (2009) CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics*, **10**, 277.
9. 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.
10. Reinhold,W.C., Sunshine,M., Varma,S., Doroshow,J.H. and Pommier,Y. (2015) Using CellMiner 1.6 for systems pharmacology and genomic analysis of the NCI-60. *Clin. Cancer Res.*, **21**, 3841–3852.
11. Yang,W., Soares,J., Greninger,P., Edelman,E.J., Lightfoot,H., Forbes,S., Bindal,N., Beare,D., Smith,J.A., Thompson,I.R. *et al.* (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
12. Taccioli,C., Sorrentino,G., Zannini,A., Caroli,J., Beneventano,D., Anderlucci,L., Lolli,M., Bicciato,S., Del Sal,G., Taccioli,C. *et al.* (2015) MDP, a database linking drug response data to genomic information, identifies dasatinib and statins as a combinatorial strategy to inhibit YAP/TAZ in cancer cells. *Oncotarget*, **6**, 38854–38865.
13. Kuleshov,M.V., Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
14. Duan,Q., Reid,S.P., Clark,N.R., Wang,Z., Fernandez,N.F., Rouillard,A.D., Readhead,B., Tritsch,S.R., Hodos,R., Hafner,M. *et al.* (2016) L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *NPJ Syst. Biol. Appl.*, **2**, 16015.
15. Subramanian,A., Narayan,R., Corsello,S.M., Peck,D.D., Natoli,T.E., Lu,X., Gould,J., Davis,J.F., Tubelli,A.A., Asiedu,J.K. *et al.* (2017) A next generation connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, **171**, 1437–1452.
16. Yu,X., Vazquez,A., Levine,A.J. and Carpizo,D.R. (2012) Allele-specific p53 mutant reactivation. *Cancer Cell*, **21**, 614–625.
17. Vazquez,A. (2009) Optimal drug combinations and minimal hitting sets. *BMC Syst. Biol.*, **3**, 81.
18. Adorno,M., Cordenonsi,M., Montagner,M., Dupont,S., Wong,C., Hann,B., Solari,A., Bobisse,S., Rondina,M.B., Guzzardo,V. *et al.* (2009) A Mutant-p53/Smad complex opposes p63 to empower TGFβ-Induced metastasis. *Cell*, **137**, 87–98.
19. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
20. Dalke,A. and Hastings,J. (2013) FMCS: a novel algorithm for the multiple MCS problem. *J. Cheminform.*, **5**, O6.
21. Gibney,G.T. and Zager,J.S. (2013) Clinical development of dabrafenib in BRAF mutant melanoma and other malignancies. *Expert Opin. Drug Metab. Toxicol.*, **9**, 893–899.
22. 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.
23. Chi,M., Ye,Y., Zhang,X.D. and Chen,J. (2014) Insulin induces drug resistance in melanoma through activation of the PI3K/Akt pathway. *Drug Des. Devel. Ther.*, **8**, 255–262.
24. Vivanco,I., Palaskas,N., Tran,C., Finn,S.P., Getz,G., Kennedy,N.J., Jiao,J., Rose,J., Xie,W., Loda,M. *et al.* (2007) Identification of the JNK signaling pathway as a functional target of the tumor suppressor PTEN. *Cancer Cell*, **11**, 555–569.
25. Naguib,A., Mathew,G., Reczek,C.R., Watrud,K., Ambrico,A., Herzka,T., Salas,I.C., Lee,M.F., El-Amine,N., Zheng,W. *et al.* (2018) Mitochondrial complex I inhibitors expose a vulnerability for selective killing of Pten-Null cells. *Cell Rep.*, **23**, 58–67.
26. Gohlke,B.-O., Nickel,J., Otto,R., Dunkel,M. and Preissner,R. (2016) CancerResource—updated database of cancer-relevant proteins, mutations and interacting drugs. *Nucleic Acids Res.*, **44**, D932–D937.
27. Kumar,R., Chaudhary,K., Gupta,S., Singh,H., Kumar,S., Gautam,A., Kapoor,P. and Raghava,G.P.S. (2013) CancerDR: cancer drug resistance database. *Sci. Rep.*, **3**, 1445.
28. Haibe-Kains,B., El-Hachem,N., Birkbak,N.J., Jin,A.C., Beck,A.H., Aerts,H.J.W.L. and Quackenbush,J. (2013) Inconsistency in large pharmacogenomic studies. *Nature*, **504**, 389–393.
29. Weinstein,J.N. and Lorenzi,P.L. (2013) Discrepancies in drug sensitivity. *Nature*, **504**, 381–383.