Computational Methods for the Integrative Analysis of Genomics and Pharmacological Data

Jimmy Caroli, Martina Dori and Silvio Bicciato *

Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

Since the pioneering NCI-60 panel of the late'80's, several major screenings of genetic profiling and drug testing in cancer cell lines have been conducted to investigate how genetic backgrounds and transcriptional patterns shape cancer's response to therapy and to identify disease-specific genes associated with drug response. 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. The analysis of this enormous and heterogeneous amount of data required the development of computational methods for the integration of genomic profiles with drug responses across multiple screenings. Here, we will review the computational tools that have been developed to integrate cancer cell lines' genomic profiles and sensitivity to small molecule perturbations obtained from different screenings.

Keywords: genomics, pharmacogenomics, integration, bioinformatics, online databases

INTRODUCTION

Clinical responses to cancer treatment are strongly influenced by the patient's genomic landscape, pushing modern therapeutics toward a more personalized approach (1). To this end, despite their inability to reflect many aspects of a drug's behavior in the human body, cancer cell lines have been the most widely used models to explore the molecular basis of drug activity. Indeed, since the NCI-60 project, several major screenings of unite genetic profiling and drug testing have been created to investigate how genomic portraits can shape cancer response to therapy. These efforts required the definition of integrated frameworks that, leveraging on high-throughput technologies and computational methods, addressed the identification of genomic factors of cancer vulnerability associated with drug sensitivity. The NCI-60 project (https://dtp.cancer.gov/discovery_development/nci-60/) has been the first extensive screening of a massive number of chemical compounds (>50,000) on a well-defined set of cancer cell lines (60 across nine different tumoral tissues) (2, 3). Building on the NCI-60 approach, several other projects investigated the interplay between genomic backgrounds and responses to drug treatment in cancer cell lines (**Figure 1A**). All cancer cell line screenings basically adopt two approaches. In the first strategy, the molecular profiles of untreated cells and their response to various compounds are investigated in parallel to assess or predict how the molecular portraits determine intrinsic cell sensitivity and resistance to drugs or potential drugs. In the second, cell lines are profiled both before and after treatment to assess how their expression profiles respond to perturbation by the various agents tested. In particular, the Cancer Cell Line Encyclopedia (CCLE, https://portals.broadinstitute.org/ccle) project fully characterized the molecular profiles of more than 1,000 untreated cancer cell lines

along with their response to a panel of 24 Food and Drug Administration (FDA)-approved drugs (4–6). Similarly, the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org) and the Cancer Therapeutics Response Portal (CTRP, http://portals.broadinstitute.org/ctrp/) linked genomic features of more than 800 cancer cell lines to their sensitivity to hundreds of chemical compounds comprising FDA-approved drugs, clinical candidates, and small molecules (7–11). Conversely, the Connectivity Map (CMap) and its recent development, L1000 (CLUE, https://clue.io), profiled cancer cell lines before and after the treatment with several chemical compounds and genomic perturbagens, retrieving gene signatures directly associated to their administration (12–14). Although these screenings share a similar experimental pipeline, most of the produced data are heterogeneous and lack concordance in terms of investigated cell lines, tested compounds, and genomic information. In this review, we will describe some computational tools for the integrative analysis of data from different pharmacogenomics resources.

INTEGRATIVE ANALYSIS OF GENOMICS AND PHARMACOLOGICAL DATA

Inspired by the NCI-60 project, several collaborative efforts scaled up the number of cancer cell lines investigated in pharmacogenomics studies from the original 60 to more than 1,400, planning to reach over 10,000 publicly available cancer models in the near future (15). The massive amount of genomic and drug response data generated by these screenings are commonly collected in databases that, through dedicated web portals, provide direct insights into potential interactions between the analyzed cancer cell lines and the tested drugs. These databases are commonly equipped with computational resources specifically designed for the navigation and the analysis of the pharmacogenomics data, as for instance GDSCTools (16), CellMiner (17), Enrichr (18), L1000 Viewer (19), PharmacoGx, and PharmacoDB (20, 21), and the recently deployed RING (22). However, most of these tools are database specific and have limited capabilities in integrating data obtained from different screenings. This limitation is mostly due to the heterogeneity of data provided by the various studies, with drug tests not standardized across projects and genomic profiling not always available for the entire panel of cell lines. In addition, data are often unbalanced, with experiments comprising a high number of cell lines screened on few drugs (e.g., CCLE and GDSC) and, vice versa, screenings of large pools of chemical compounds performed on small cohorts of cancer cell lines (as in the NCI-60). Finally, while genomic data are rather homogeneous and can be easily integrated across studies after removing batch effects, pharmacological data derived from distinct experimental designs must be kept separate as they are profoundly different in terms of analytical assays, tested drug concentration, and retrieved inhibitory potential (23, 24). Despite these intrinsic limitations, several approaches have been proposed for the integrative analysis of genomics and pharmacological data collected from different screenings (**Figure 1B**). In particular, CellMinerCDB
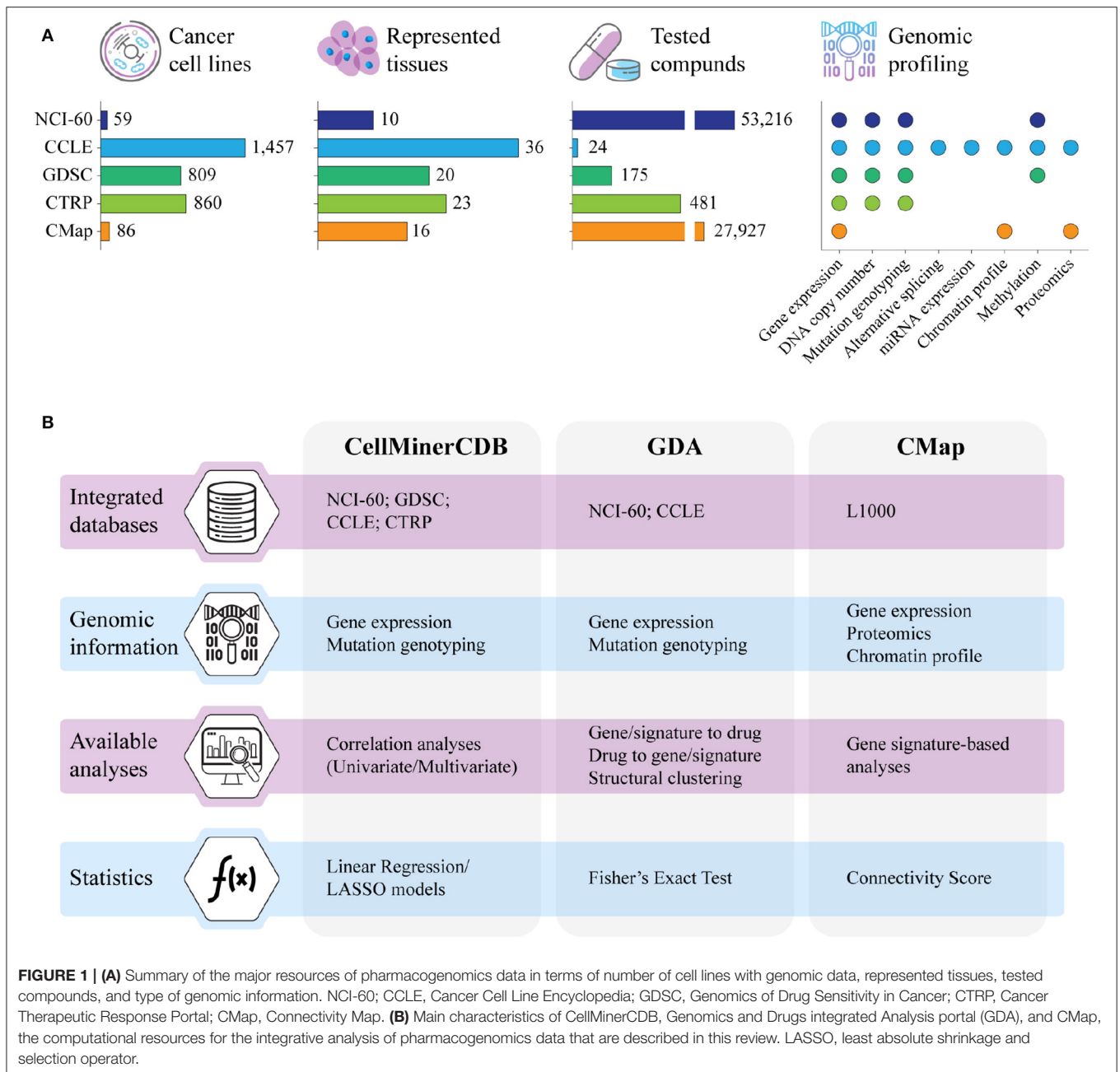
combines genomic profiles from NCI-60, CCLE, GDSC, and CTRP with the pharmacological data provided by the NCI-60 screening (25); the Genomics and Drugs integrated Analysis portal (GDA) integrates pharmacological data derived from the NCI-60 with the genomic information of NCI-60 and CCLE (26); and the CMap enables the investigation of the L1000 data through the correlation of gene lists and transcriptional signatures modulated by the drug treatment (12, 14, 27).

CellMinerCDB: Integrative Cross-Database Genomics and Pharmacogenomics Analyses

CellMinerCDB (https://discover.nci.nih.gov/cellminercdb/) expands the analysis power of CellMiner, the original NCI-60 analysis tool, with the integration of the cancer cell line data from the Sanger/Massachusetts General Hospital GDSC, the Broad/Novartis CCLE, and the Broad CTRP (25, 28). The integrated database comprises all molecular profiles of almost 1,400 different cancer cell lines, together with drug activity for more than 20,000 compounds. The guiding element, used to link pharmacological information to genomic data from different sources, is the set of common cancer cell lines between the NCI-60 and the other resources, with 55 NCI-60 lines shared with GDSC, 44 with CCLE, and 671 in common between CCLE and GDSC. CellMinerCDB performs correlation analyses to investigate and visualize relationships between the drug activity of a compound and the specific profile of a selected molecular feature across all the available cell lines (univariate analysis). In addition, linear regression methods are implemented for the integrative analysis of multiple identifiers (multivariate analysis). The confidence of the associations is assessed by statistical analyses conducted through a basic linear regression model or using least absolute shrinkage and selection operator (LASSO). An interesting feature of CellMinerCDB is the possibility to compare patterns associated to either drug activity or molecular data via the *Compare Pattern* function of the univariate analysis search. This analysis allows the identification of genomic determinants of drug response, as exemplified by the connection found between the expression of Schlafen 11 (SLFN11) and the response to several DNA-targeted anticancer drugs as platinum derivatives, topoisomerase inhibitors, and poly (ADP-ribose) polymerase (PARP) inhibitors (25).

Genomics and Drugs Integrated Analysis

GDA (gda.unimore.it/) is a web-based tool designed for the integrative analysis of drug response, mutations, and gene expression profiles derived from the NCI-60 consortium and the CCLE (26, 29). GDA comprises 73 cancer cell lines shared by NCI-60 and CCLE and treated with 50,816 compounds and integrates the drug response data from the NCI-60 screening with the mutations and genomic information derived from both CCLE and NCI-60. GDA allows four different types of analyses, namely, *from drug to gene, from gene to drug, from signature to drug*, and *from drug to signature*. Pharmacological and genomic data can be queried to identify drugs correlated to gene mutations (from gene to drug), gene mutations associated

FIGURE 1 | (A) Summary of the major resources of pharmacogenomics data in terms of number of cell lines with genomic data, represented tissues, tested compounds, and type of genomic information. NCI-60; CCLE, Cancer Cell Line Encyclopedia; GDSC, Genomics of Drug Sensitivity in Cancer; CTRP, Cancer Therapeutic Response Portal; CMap, Connectivity Map. **(B)** Main characteristics of CellMinerCDB, Genomics and Drugs integrated Analysis portal (GDA), and CMap, the computational resources for the integrative analysis of pharmacogenomics data that are described in this review. LASSO, least absolute shrinkage and selection operator.

to drug responses (from drug to gene), and drugs associated to active gene signatures (from signature to drug). Starting from a drug correlated to gene mutations, gene expression profiles can be used to identify genes differentially expressed in cell lines sensitive to the selected compound. The statistics behind GDA is based on drug response data. Basically, all pairs of cell lines and drugs are defined as responsive if the relative sensitivity is smaller than two standard deviations of the left tail of the distribution of all relative sensitivities, and non-responsive otherwise. Based on genomic data, cell lines are classified as mutant if treated with the compound and carrying the selected set of mutations and as wild type if treated with the compound but without the specific set of mutations. Given these classifications, compounds are ranked using a score defined by

the fraction of responsive in mutant multiplied by the fraction of non-responders in wild type. This score ranks each drug based on the enrichment of responsive in the mutant group. The statistical significance of this ranking is computed using a one-tailed Fisher's exact test for the enrichment of responsive in mutant as compared to non-responsive in wild type, given the number of non-responsive in mutant and responsive in wild type. Results are accessible through interactive graphical representations and tables and can be directly fed to external tools as Enrichr for functional annotation (18). When used to identify compounds able to inhibit the proliferation potential of cancer cell lines with aberrant nuclear YAP/TAZ activation, GDA retrieved imatinib analogs and statins as potentially active drugs. Following GDA indications, *in vitro* studies demonstrated that

the combination of statins with dasatinib, an imatinib analog enhances YAP/TAZ nuclear exclusion, is able to block YAP/TAZ transcriptional activity, and is much more active in inducing apoptosis in different tissues (29).

Connectivity Map and the CMap Linked User Environment

CMap (https://www.broadinstitute.org/connectivity-map-cmap) was one of the first computational resources developed for the investigation of connections between transcriptomics and drug-induced perturbations (12). As extensively reviewed in Musa et al. (30), the goal of CMap is to identify drug or disease-associated gene signatures correlating with transcriptomics changes induced by the administration of drugs or chemical compounds (31, 32). The original project comprised the gene expression profiling of three cancer cell lines before and after the treatment with 164 different small molecules, obtaining drug-associated gene signatures for each cell line. This initial version has been recently scaled up through the L1000 Assay Platform, a method to analyze the expression levels of 978 selected landmark transcripts (assayed with 1,058 probes, including 80 controls) that have been shown to be sufficient to recover more than 80% of the information relative to the full transcriptome (14). This new approach translated into the screening of 86 different cancer cell lines using 27,927 unique perturbagens, including 19,811 small molecules and 7,494 genetic perturbations (consisting of overexpression or knockdown of different genes associated with human diseases or biological pathways). This large-scale screening finally resulted in a collection of 476,251 gene expression signatures that can be analyzed through the CMap Linked User Environment (CLUE, https://clue.io). In CLUE, the Query tool allows to input a gene signature (i.e., a list of genes upregulated and downregulated) and search for perturbagens (chemical and/or genetic) that induce a similar (or opposite) expression profile in the treated cells. The statistical significance of the association is assessed through a connectivity score that takes into account the strength of the similarity between the query and the induced signature as compared to the enrichment of all other signatures in the database (14). This approach proved its efficacy in the identification of a novel inhibitor for the serine-threonine kinase CSNK1A, an enzyme essential in specific subtypes of myelodysplastic syndrome and acute myeloid leukemia. Starting from the loss of function signature of CSNK1A1, authors searched CMap for compounds mimicking the loss of this kinase and identified one compound (BRD-1868) with a high connectivity score relative to this signature. Further enzymatic assays confirmed both the binding between BRD-1868 and CSNK1A1 and its inhibitory effect on enzymatic activity (14). From its first publication, CLUE has been expanded to include also proteomics analysis ranging from expression arrays to histone modification signatures.

CONCLUDING REMARKS

Efforts to decipher the molecular mechanisms of cancer stimulated scientists to explore the interconnection between the genomic landscape of cancer models and their response to drug treatments. This resulted in large pharmacogenomics screenings that, with the advent of high-throughput technologies, generated large amounts of genomics and pharmacological data. However, the integration of these precious information is still challenging due to the variable type and number of drugs and cancer cell lines that have been screened by the various projects and the heterogeneous assays used for drug testing in the different studies (23, 24, 33–35). Despite these intrinsic difficulties, several computational approaches have been developed for the integrative analysis of genomics and pharmacological data. Their application allowed to discover several new connections between drug sensitivity and genomic backgrounds, enabling the potential repurposing of commercially available drugs to cancer treatment (36–38). However, these computational resources, although proven effective, still suffer the limitations of the original studies as the sparsity of the drug and cell interaction matrices, the effective impossibility to merge drug response data across different screenings, and the criticalities of cancer cell lines as a reliable cancer model (39–41). To this end, the project for a Patient-Derived Model Database (PDMB) launched in 2012 by the NCI might represent a potential breakthrough as genomic and drug response data directly collected from patients and patient-derived xenografts (PDXs) will reproduce more accurately the cancer disease and its environment than any cell line model (42). Furthermore, while novel experimental models are generating more accurate data, advanced computational methods are under development to enhance the analytical potential of existing algorithms. As recently discussed (43–45), artificial intelligence approaches as network-based models, deep-learning frameworks, and machine-learning techniques are increasingly applied to investigate pharmacogenomics connections and drug repositioning. These methods can be effective not only for data integration but also to predict new interactions and applications of already approved drugs (46–48). In summary, computational approaches for the integration of genomic and pharmacological data have the potential to become crucial for the systematic identification of new biomarkers of drug sensitivity and the discovery of novel anticancer drugs on the basis of specific genetic abnormalities, as long as reliable cellular models and highly curated data become available.

AUTHOR CONTRIBUTIONS

JC and SB conceived the project. JC, MD, and SB wrote and revised the manuscript.

FUNDING

REFERENCES

1. Roden DM, George AL Jr. The genetic basis of variability in drug responses. *Nat Rev Drug Discov.* (2002) 1:37–44. doi: 10.1038/nrd705
2. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer.* (2006) 6:813–23. doi: 10.1038/nrc1951
3. Abaan OD, Polley EC, Davis SR, Zhu YJ, Bilke S, Walker RL, et al. The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer Res.* (2013) 73:4372–82. doi: 10.1158/0008-5472.CAN-12-3342
4. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* (2012) 483:603–7. doi: 10.1038/nature11003
5. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature.* (2019) 569:503–8. doi: 10.1038/s41586-019-1186-3
6. Cancer Cell Line Encyclopedia Consortium and Genomics of Drugs Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. *Nature.* (2015) 528:84–7. doi: 10.1038/nature15736
7. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature.* (2012) 483:570–5. doi: 10.1038/nature11005
8. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell.* (2013) 154:1151–61. doi: 10.1016/j.cell.2013.08.003
9. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* (2013) 41:D955–61. doi: 10.1093/nar/gks1111
10. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* (2015) 5:1210–23. doi: 10.1158/2159-8290.CD-15-0235
11. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol.* (2016) 12:109–16. doi: 10.1038/nchembio.1986
12. 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) 313:1929–35. doi: 10.1126/science.1132939
13. Lamb J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer.* (2007) 7:54–60. doi: 10.1038/nrc2044
14. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell.* (2017) 171:1437–52 e1417. doi: 10.1016/j.cell.2017.10.049
15. Boehm JS, Golub TR. An ecosystem of cancer cell line factories to support a cancer dependency map. *Nat Rev Genet.* (2015) 16:373–4. doi: 10.1038/nrg3967
16. Cokelaer T, Chen E, Iorio F, Menden MP, Lightfoot H, Saez-Rodriguez J, et al. GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics.* (2018) 34:1226–8. doi: 10.1093/bioinformatics/btx744
17. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. 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.* (2012) 72:3499–511. doi: 10.1158/0008-5472.CAN-12-1370
18. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* (2013) 14:128. doi: 10.1186/1471-2105-14-128
19. Musa A, Tripathi S, Dehmer M, and Emmert-Streib, F. L1000 viewer: a search engine and web interface for the LINCS data repository. *Front Genet.* (2019) 10:557. doi: 10.3389/fgene.2019.00557
20. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics.* (2016) 32:1244–6. doi: 10.1093/bioinformatics/btv723
21. Smirnov P, Kofia V, Maru A, Freeman M, Ho C, El-Hachem N, et al. PharmacoDB: an integrative database for mining *in vitro* anticancer

22. drug screening studies. *Nucleic Acids Res.* (2018) 46:D994–1002. doi: 10.1093/nar/gkx911
22. Politano G, Di Carlo S, Benso A. "One DB to rule them all"—the RING: a Regulatory INteraction Graph combining TFs, genes/proteins, SNPs, diseases and drugs. *Database.* (2019) 2019:108. doi: 10.1093/database/baz108
23. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, et al. Inconsistency in large pharmacogenomic studies. *Nature.* (2013) 504:389–93. doi: 10.1038/nature12831
24. Weinstein JN, Lorenzi PL. Cancer: discrepancies in drug sensitivity. *Nature.* (2013) 504:381–3. doi: 10.1038/nature12839
25. Rajapakse VN, Luna A, Yamade M, Loman L, Varma S, Sunshine M, et al. CellMinerCDB for integrative cross-database genomics and pharmacogenomics analyses of cancer cell lines. *Science.* (2018) 10:247–64. doi: 10.1016/j.isci.2018.11.029
26. Caroli J, Sorrentino G, Forcato M, Del Sal G, Bicciato, S. GDA, a web-based tool for Genomics and Drugs integrated analysis. *Nucleic Acids Res.* (2018) 46:W148–56. doi: 10.1093/nar/gky434
27. Duan Q, Reid SP, Clark NR, Wang Z, Fernandez NF, Rouillard AD, et al. L1000CDS(2): LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl.* (2016) 2:15. doi: 10.1038/npjsba.2016.15
28. Polley E, Kunkel M, Evans D, Silvers T, Delosh R, Laudeman J, et al. Small cell lung cancer screen of oncology drugs, investigational agents, and gene and microRNA expression. *J Natl Cancer Inst.* (2016) 108. doi: 10.1093/jnci/djw122
29. Taccioli C, Sorrentino G, Zannini A, Caroli J, Beneventano D, Anderlucci L, et al. 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.* (2015) 6:38854–65. doi: 10.18632/oncotarget.5749
30. Musa A, Ghoraie LS, Zhang SD, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform.* (2018) 19:506–23. doi: 10.1093/bib/bbw112
31. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. *Cell.* (2000) 102:109–26. doi: 10.1016/S0092-8674(00)00015-5
32. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med.* (2011) 3:96ra76. doi: 10.1126/scitranslmed.3002648
33. Papillon-Cavanagh S, De Jay N, Hachem N, Olsen C, Bontempi G, Aerts HJ, et al. Comparison and validation of genomic predictors for anticancer drug sensitivity. *J Am Med Inform Assoc.* (2013) 20:597–602. doi: 10.1136/amiajnl-2012-001442
34. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. *Cell.* (2016) 166:740–54. doi: 10.1016/j.cell.2016.06.017
35. Safikhani Z, Smirnov P, Freeman M, El-Hachem N, She A, Rene Q, et al. Revisiting inconsistency in large pharmacogenomic studies. *F1000Res.* (2016) 5:2333. doi: 10.12688/f1000research.9611.1
36. Kim JH, Scialli AR. Thalidomide: the tragedy of birth defects and the effective treatment of disease. *Toxicol Sci.* (2011) 122:1–6. doi: 10.1093/toxsci/kfr088
37. Pritchard JLE, O'Mara TA, Glubb DM. Enhancing the promise of drug repositioning through genetics. *Front Pharmacol.* (2017) 8:896. doi: 10.3389/fphar.2017.00896
38. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med.* (2017) 376:1713–22. doi: 10.1056/NEJMoa1615664
39. Mullard A. Reliability of 'new drug target' claims called into question. *Nat Rev Drug Discov.* (2011) 10:643–4. doi: 10.1038/nrd3545
40. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature.* (2018) 560:325–30. doi: 10.1038/s41586-018-0409-3
41. Mullard A. Can you trust your cancer cell lines? *Nat Rev Drug Discov.* (2018) 17:613. doi: 10.1038/nrd.2018.154
42. Mer AS, Ba-Alawi W, Smirnov P, Wang YX, Brew B, Ortmann J, et al. Integrative pharmacogenomics analysis of patient-derived xenografts. *Cancer Res.* (2019) 79:4539–50. doi: 10.1158/0008-5472.CAN-19-0349

43. Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn-Feuer A, Dinov ID, et al. Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics.* (2018) 19:629–50. doi: 10.2217/pgs-2018-0008

44. Chiu YC, Chen HIH, Zhang T, Zhang S, Gorthi A, Wang LJ, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genom.* (2019) 12(Suppl.1):18–18. doi: 10.1186/s12920-018-0460-9

45. Sakellaropoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, et al. A deep learning framework for predicting response to therapy in cancer. *Cell Rep.* (2019) 29:3367–73.e3364. doi: 10.1016/j.celrep.2019.11.017

46. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE.* (2013) 8:e61318. doi: 10.1371/journal.pone.0061318

47. Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol.* (2014) 32:1202–12. doi: 10.1038/nbt.2877

48. Kalamara A, Tobalina L, Saez-Rodriguez J. How to find the right drug for each patient? *Adv Challenges Pharmacogenom.* (2018) 10:53–62. doi: 10.1016/j.coisb.2018.07.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.