Published online 8 May 2017 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* ***W215–W221***
doi: 10.1093/nar/gkx400

mTCTScan: a comprehensive platform for annotation and prioritization of mutations affecting drug sensitivity in cancers

Mulin Jun Li[1,2,†], Hongcheng Yao[2,3,†], Dandan Huang[1,†], Huanhuan Liu[1], Zipeng Liu[2], Hang Xu[2,3], Yiming Qin[2,3], Jeanette Prinz[4], Weiyi Xia[5], Panwen Wang[4], Bin Yan[2,3], Nhan L. Tran[6], Jean-Pierre Kocher[4], Pak C. Sham[2,7] and Junwen Wang[4,8,*]

[1]Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China, [2]Center for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, China, [3]School of Biomedical Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, China, [4]Department of Health Sciences Research, Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ 85259, USA, [5]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, China, [6]Department of Cancer Biology, Mayo Clinic, Scottsdale, AZ 85259, USA, [7]Departments of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, China and [8]Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ 85259, USA

Received February 27, 2017; Revised April 17, 2017; Editorial Decision April 26, 2017; Accepted April 27, 2017

ABSTRACT

Cancer therapies have experienced rapid progress in recent years, with a number of novel small-molecule kinase inhibitors and monoclonal antibodies now being widely used to treat various types of human cancers. During cancer treatments, mutations can have important effects on drug sensitivity. However, the relationship between tumor genomic profiles and the effectiveness of cancer drugs remains elusive. We introduce Mutation To Cancer Therapy Scan (mTCTScan) web server (http://jjwanglab.org/mTCTScan) that can systematically analyze mutations affecting cancer drug sensitivity based on individual genomic profiles. The platform was developed by leveraging the latest knowledge on mutation-cancer drug sensitivity associations and the results from large-scale chemical screening using human cancer cell lines. Using an evidence-based scoring scheme based on current integrative evidences, mTCTScan is able to prioritize mutations according to their associations with cancer drugs and preclinical compounds. It can also show related drugs/compounds with sensitivity classification by considering the context of the entire genomic profile. In addition, mTCTScan incorporates comprehensive filtering functions and cancer-related annotations to better interpret mutation effects and their association with cancer drugs. This platform will greatly benefit both researchers and clinicians for interrogating mechanisms of mutation-dependent drug response, which will have a significant impact on cancer precision medicine.

INTRODUCTION

An increasing number of cancer drugs have been developed to treat various types of human cancers. Studies have indicated that the genomic context of a tumor is a major factor affecting the effectiveness of cancer drugs (1). Tumor genomic alterations could either confer changes in drug sensitivity (2) or could be used to identify subsets of patients with a dramatic response (3). However, the association between tumor genomic profiles and drug effectiveness remains to be fully determined. To date, only a small proportion of cancer genomic alterations have been confirmed to be actionable with approved agents.

Recently, there have been unprecedented advancements in next generation sequencing technologies that now allow high-throughput tumor genomic profiling at relatively low cost. Along with the precision medicine initiative, large scale clinical trials, such as the NCI-Molecular Analysis for Therapy Choice (NCI-MATCH, also referred as NCT02465060 in ClinicalTrials.gov), are being conducted to investigate the association between drug response and somatic mutations (4). Meanwhile, rapidly accumulating data from chemical screening on cancer cell lines including the cancer cell line encyclopedia (5), Cancer Therapeutics Response Portal

*To whom correspondence should be addressed. Tel: +1 480 301 4644; Fax: +1 480 301 8387; Email: wang.junwen@mayo.edu
†These authors contributed equally to the paper as first authors.

(6,7) and Genomics of Drug Sensitivity in Cancer (GDSC) (8–10), can provide both drug sensitivity and cancer cell line genomic profiles. Both efforts are continuing to generate invaluable information for identifying new associations between mutations and drug/compound sensitivity. Following the emergence of these studies, several databases have been developed to collate and format these data. For example, Clinical Interpretations Of Variants In Cancer (CIViC) (11), My Cancer Genome (12), Gene Drug Knowledge Database (13) and MD Anderson Cancer Center's Personalized cancer therapy (14) focus on curating mutation associations with cancer drugs from peer-reviewed literature. Meanwhile, Cancer Driver Log (15) collects potentially actionable driver mutations with functional characterization or targeted by existing therapies from literature. However, these databases mainly focus on curating and formatting mutation association information, with data queries performed only at the single mutation level. Few tools could accept genomic profiles as input to interpret mutations associated with cancer drug sensitivity.

The integrating molecular profiles with actionable therapeutics (16) is an analysis pipeline that combines mutation calling from sequencing data with drug prioritization based on detected mutations. However, it utilizes a gene–drug interaction based strategy to make drug predictions and is limited to only U.S. Food and Drug Administration-approved drugs. Currently, there is no web-based tool that can automatically handle individual-level cancer genomic profiles and provide comprehensive mutation associations with cancer drugs. Moreover, for better analysis of mutations associated with drug sensitivity, it is necessary to integrate various types of information, such as mutation annotations, drug descriptions and related clinical trials. Therefore, an integrative platform which compiles all known mutation associations with cancer drugs and comprehensive annotations is urgently needed.

We developed a web server, called Mutation To Cancer Therapy Scan, or mTCTScan for short (http://jjwanglab.org/mTCTScan), that can analyze mutation-cancer drug associations based on given cancer genomic profiles. We first curated and compiled known mutation associations with cancer drugs from the literature and public resources, including CIViC and Gene Drug Knowledge Database. We also incorporated cell line level associations between cancer functional events and drug sensitivity from GDSC. Using an evidence-based scoring scheme, mTCTScan is able to prioritize mutations by incorporating all their associations with cancer drugs and preclinical compounds and to classify the drugs/compounds by considering the entire cancer genomic profile provided. In addition, comprehensive cancer-related mutation annotations and drug information, including mutation genomic features, mutation germline/somatic occurrence rates, known and predicted pathogenicity/deleteriousness of the mutation across different biological processes, conservation, drug descriptions and clinical trial information, were incorporated to better interpret the mutation effects on drug sensitivity.

METHODS AND PIPELINE

Data collection and processing

Mutation-drug sensitivity information. We collected and curated mutation-cancer drug associations from the late 2016 version of two publically available databases, CIViC and Gene Drug Knowledge Database, and from additional 245 publications from PubMed or conference abstracts. We then compiled these mutation-drug associations according to standardized criteria: (i) each association record included essential attributes such as drug name, gene name, mutation description, association direction, confidence level, related disease, genomic coordinates of the mutation and the reference; (ii) association directions were normalized to a value of either 'Increased sensitivity or response' or 'Reduced sensitivity or response'; (iii) confidence levels were normalized into five grades including 'Proven', 'Clinical trial stage', 'Case report stage', 'Preclinical stage' and 'Inferential stage'; and (iv) supporting references were provided as a PubMed ID or by URL links. To acquire structured disease terminology, we mapped the related disease onto Disease Ontology (17). We used DrugBank (18) and ClinicalTrials.gov (19) to obtain drug/compound descriptions and their corresponding clinical information. To associate drugs with their target genes, we integrated gene–drug interactions from DGIdb (20). Cell line-based cancer functional events and drug sensitivity associations were incorporated from GDSC, which employs ANNOVA analysis methods to identify significant associations (see Supplementary Table S1 for detailed information of the resources used).

Mutation annotation information. We used a series of base-wise annotations to help researchers interpret the underlying functions of the mutation in the cancer genomes, including mutation genomic features, germline and somatic occurrence frequency from population-scale projects, known and predicted pathogenicity/deleteriousness across different genomic areas, as well as sequence conservation. Gene-based annotations were retrieved from SnpEff (21) and SNVrap (22,23) for each mutation. Germline variant frequencies were integrated from 1000 Genomes Project phase3 (24) and Exome Aggregation Consortium (ExAC) (25). To acquire somatic mutation recurrence rates in cancer patients, aggregation of somatic mutations were constructed by integrating several cancer-specific resources including The Cancer Genome Atlas Data Coordinating Center (26), International Cancer Genome Consortium Data Portal (27) and Catalogue of Somatic Mutations in Cancer (28). Because the functional prediction of cancer mutations could facilitate in evaluating the biological activity of drug targets, we incorporated large-scale mutation functional prediction scores of three major biological domains: (i) non-coding single-nucleotide variant functional scores from dbNCFP (29), (ii) splicing single-nucleotide variant functional scores from dbscSNV (30) and (iii) non-synonymous single-nucleotide variants functional scores from dbNSFP (31). In addition, base-wise conservation score from CADD were also included (32) (see Supplementary Table S2 for detailed information of the annotations used).

Annotation and prioritization pipeline

Given an individual cancer genomic profile, mTCTScan uses an analytical pipeline to filter, annotate and prioritize mutations and cancer drugs. The overall workflow of mTCTScan is shown in Figure 1.

Mutation filtering and annotation at the transcript level. The Individual cancer genomic profile (usually The Variant Call Format (VCF) file generated by somatic mutation calling tools (33)) is first annotated by SnpEff and SNVrap to acquire mutation features, including mutation type and associated genomic effects. To efficiently convert the mutation descriptions between VCF format and Human Genome Variation Society (HGVS) format (DNA, RNA and protein levels) (34), we used TransVar (35) to perform forward (to translate a genomic variant to mRNA or protein changes) and reverse (to trace an mRNA or protein variant to all potential genomic origins) annotations considering all possible mRNA and protein isoforms. For cases where the input mutations contain known/germline alleles in human populations, we filtered them by an allele frequency <0.005 for human pan-populations using an aggregated allele frequency database. Furthermore, mutations could be further reduced by user-defined target genes, tumor types and specific mutation types at this step.

Identifying mutation-drug associations at different levels. For each filtered mutation at the gene transcript level, we used a series of matching strategies to link it to the collected mutation-cancer drug association records, including 'Exact Match', 'Partial Match', 'Small-scale Overlap' and 'Locus-based Overlap': (i) 'Exact Match' requires the genomic coordinates or protein effect of query mutations to be exactly the same as the collected records; (ii) 'Partial Match' indicates that the protein effect of the query mutation matches the amino acid change in the collected records (but the altered amino acid of the query mutation and the records are different), or the genomic coordinates or protein effect of the query mutation exactly match one or part of the record's mutations (some records require the concurrence of multiple mutations); (iii) 'Small-scale Overlap' means the genomic coordinates of query mutation overlaps with collected records at the small insertions or deletions (Indel) level; (iv) 'Locus-based Overlap' indicates that the genomic coordinates of query mutation overlaps with the collected records at the structure variations level. In addition, the web server would also search through cancer functional events and drug sensitivity associations from the cell line data by matching the query mutation to related cancer functional events. To retrieve all relevant data at the different levels, we mapped the filtered cancer mutations to each of above categories.

Evidence-based mutation prioritization. For all matched and overlapped mutation-drug associations for each query mutation, the web server then uses an evidence-based ensemble score to represent the matching degree, breadth and reliability of the retrieved association records according to their matching type and confidence level. The ensemble score is analogous to the 'Altmetric Attention Score' for measuring the degree of attention received by research outputs (36). The ensemble score is calculated as:

$$ES = \sum_{i}^{n} M_i \times C_i$$

where n is the total number of retrieved mutation-cancer drug association records for each mutation, M is the matching score for corresponding matching category ('Exact Match': 10; 'Partial Match': 3; 'Small-scale Overlap': 0.5; 'Locus-based Overlap': 0.1), C is the weighting factor for the corresponding confidence level ('Proven': 1; 'Clinical trial stage': 0.6; 'Case report stage': 0.4; 'Preclinical stage': 0.2; 'Inferential stage': 0.1). The same matching strategies and a fixed confidence level of 'Preclinical stage' are then applied to score the mutations matched with the cell line-based cancer functional events and drug sensitivity association records.

In the formula, the matching score describes the degree of agreement between the query mutation and retrieved association records. One mutation may retrieve multiple records by 'Locus-based Overlap' matching, especially if it falls on some popular cancer drug targets. Presumably, one 'Exact Match' should be assigned an equal or higher score than all 'Locus-based Overlap' matches together for each mutation. Considering that the highest number of possible 'Locus-based Overlap' matches for one query mutation is around 100 if it falls on ERBB2:104, PTEN:94, KRAS:86 or EGFR:83, we thus required a 100-fold difference between the 'Exact Match':10 and 'Locus-based Overlap':0.1 matching scores. The matching scores of 'Partial Match':2 and 'Small-scale Overlap':0.5 are then calculated as $10/100^{\frac{1}{3}}$ and $10/100^{\frac{2}{3}}$ respectively, and these two numbers are further adjusted to integers. The weighting factor describes the confidence level of the association records and the descending weighting factors from 'Proven' to 'Inferential stage' reflect the decreasing reliability of the association records. Taken together, this ensemble score reflects the relevance between query mutations and cancer drug responses based on the current evidence and can be used to prioritize the query mutations.

Cancer drug classification. Because different mutations may confer consistent or opposite effects on drug sensitivity when considering the entire cancer genomic profile, we classified the cancer drugs/compounds according to the matched mutation-drug associations in the above steps. For each drug/compound, the web server compares whether the mutations' impact on drug sensitivity is consistent or conflicting across all associated mutations at the 'Exact Match' and 'Partial Match' levels. The drugs/compounds are then divided into three categories based on the following criteria: (i) 'Increased Sensitivity' indicates the mutations consistently increase the drug's sensitivity or response; (ii) 'Reduced Sensitivity' indicates the mutations consistently decrease the drug's sensitivity or response; and (iii) 'Conflicting Sensitivity' indicates at least one mutation modulates the drug's sensitivity or response in the opposite direction compared with the other mutations. Finally, we used the number of associated mutations to order the drugs in each category.

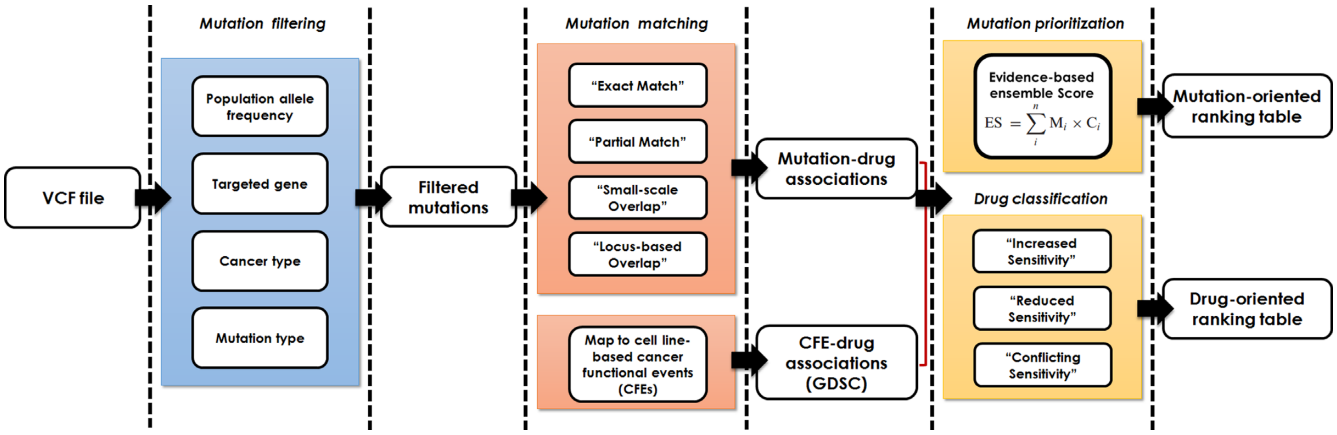

Figure 1. The workflow of mTCTScan (see the description for details of the pipeline).

WEB SERVER DESCRIPTION

Usage and interface

The mTCTScan web server accepts inputs as VCF text or as an uploaded VCF file. Users can either select single/multiple cancer type(s) to investigate disease-specific associations between mutations and drug sensitivity, or use a default pan-cancer analysis mode. Allele frequency filtering based on both 1000 Genomes Project and ExAC is well supported for filtering known variants. Users can also define a target gene panel to limit mutations to a list of genes of interest. To help users to identify mutations with certain genomic effects, mTCTScan allows users to select only mutations with specific features for further analysis. Once the job is submitted, users can track the job status from either an encrypted link or the job queue.

The mTCTScan platform displays the results in several user-friendly interfaces, such as the drug-oriented tab and the mutation-oriented tab with detailed mutation-drug sensitivity association information and interactive functions. The mutation-oriented tab provides a ranking table summarizing the integrative evidence for each mutation's impacts on cancer drug sensitivity according to the scoring scheme (Figure 2A). By clicking on the 'details' link for each mutation, users can examine the comprehensive mutation annotations at the gene transcript level including mutation genomic features, germline and somatic occurrence rates, mutation functional prediction scores and base pair level sequence conservation (Figure 2B). In addition to the detailed mutation-drug response associations and drug/compound information, mTCTScan also incorporates a lollipop-style mutation diagram (37) to visualize the variant position in the protein functional domain (Figure 2C). Furthermore, by mapping the mutation to its genomic locus, mTCTScan can also report on related drugs based on the gene–drug relationship. The drug-oriented tab classifies the mutation-related cancer drugs into three categories by combining all retrieved mutation-drug associations from the entire cancer genomic profile. Users can check all actionable mutations for each drug (Figure 2D). Last but not least, mTCTScan provides a query function that allows users to quickly search mutation-drug associations at the single mutation level.

Web server design

We implemented the mTCTScan web server using the Perl-based web framework 'Catalyst'. Annotation information is indexed and retrieved by Tabix (38). Oracle Grid Engine is used as the job management system for task submission and JQuery and related UI components are used to construct the interactive web pages.

Evaluation

We used several actionable mutations in protein tyrosine kinase domain of human epidermal growth factor receptor (EGFR) to evaluate the usefulness and effectiveness of our mTCTScan platform. One of the most common EGFR-activating mutations is an amino acid substitution in exon 21 (leucine to arginine at codon 858; L858R), which confers increased affinity in the adenosine triphosphate (ATP)-binding pocket for EGFR-tyrosine kinase inhibitor (EGFR-TKIs) compared to wild-type EGFR (39). We investigated this mutation using mTCTScan and found 45 mutation-drug association records, including 18 'Exact Match' records and 11 drugs/therapies with increased sensitivity or response. During the treatment of second-generation EGFR-TKIs, the major mechanism of acquired resistance can be explained by the occurrence of a secondary EGFR kinase domain mutation in exon 20 (threonine to methionine at codon 790; T790M) (40). When considering both L858R and T790M, mTCTScan identified more mutation-drug records and additional drug response categories, such as 'Reduced Sensitivity' and 'Conflicting Sensitivity'. For example, several third-generation EGFR-TKIs (such as Osimertinib and Rociletinib) show increased sensitivity, while, four drugs Afatinib, Dacomitinib, Gefitinib and Erlotinib were shown to have conflicting sensitivity when considering these two mutations simultaneously. Furthermore, recent clinical studies revealed that a 'tertiary' substitution mutation that changes EGFR cysteine 797 into serine (C797S) can affect covalent binding of drugs, which confers resistance to all third-generation EGFR-TKIs (41). When considering all three mutations (L858R, T790M and C797S), mTCTScan further revealed conflicting sensitivity for Osimertinib and EAI045, a promising fourth-generation EGFR-TKI with increased sensitivity in the preclinical

Figure 2. The web server result pages. (**A**) mutation-oriented tab; (**B**) mutation annotations at the transcript level; (**C**) lollipop-style mutation diagram; and (**D**) drug-oriented tab.

stage (42). Taken together, mTCTScan was able to accurately interpret mutation-dependent drug sensitivity based on the entire genomic context.

DISCUSSION

In current cancer treatment practices, only a small proportion of cancer genomic alterations have been confirmed actionable with approved therapies, which significantly limits the scope of cancer precision medicine. Recent large-scale clinical use of mutation-dependent treatments in cancer patients and chemical screening on human cancer cell lines have revealed many new associations between mutations and drug/compound sensitivity. Such association information is highly valuable for researchers and clinicians, but efforts to compile this data and to provide an integrative plat-

form for analyzing personalized genomic profiles are lacking. To address this, we developed mTCTScan to comprehensively analyze mutation-cancer drug associations considering an individual cancer genomic profile. Compared with existing software and databases, mTCTScan compiled more types of data and the latest information to build an integrative one-stop web server that allows clinicians and researchers to interpret the effects of mutations on cancer drug sensitivity.

Many drug sensitivities are associated with structure variations or combinations of mutations in tumor cells. However, exact matching of such types of genomic alteration between the query mutation and known mutation-drug associations is difficult, which could result in the underestimation of the effects of those mutations when using this web server. Furthermore, precise comparison of the pharmacogenomic effects of different mutations across existing anti-cancer agents is currently unfeasible. Therefore, our matching strategy serves more as an information retrieval method and should be interpreted with care, especially its biological significance. Even if cell line-based pharmacological screening can provide an effective way to interrogate the genetic effect of the mutations on drug response, there are still many other factors that could confound the independent mutation effect on drug sensitivity. Our prioritization pipeline uses an evidence-based ensemble score to measure the mutation's relevance to known mutation-cancer drug associations, as well as the breadth and confidence level of these associations, which provides an indicator of the actionability of individual mutations. In the future, we will continue curating and integrating more mutation-cancer drug response associations from both clinical and experimental studies. We also plan to introduce more accurate matching strategies and prioritization methods to better interpret the associations between mutations and drug sensitivity in cancers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Dr Maria P. Wong for her helpful comments and suggestions.

FUNDING

Y S and Christabel Lung Postgraduate Scholarship from the University of Hong Kong; Talent Excellence Program from Tianjin Medical University (to M.J.L); Research Grants Council, Hong Kong SAR, China [17121414M]; Mayo Clinic (Mayo Clinic Arizona and Center for Individualized Medicine); National Institute of Health [5R01CA170357, 2P30CA015083 to J.W.W, 1U54CA210180 to J.W.W, N.L.T]. Funding for open access charge: Mayo Clinic (Mayo Clinic Arizona and Center for Individualized Medicine).
Conflict of interest statement. None declared.

REFERENCES

1. Hu,X. and Zhang,Z. (2016) Understanding the genetic mechanisms of cancer drug resistance using genomic approaches. *Trends Genet.* , **32**, 127–137.
2. Schmitt,M.W., Loeb,L.A. and Salk,J.J. (2016) The influence of subclonal resistance mutations on targeted cancer therapy. *Nat. Rev. Clin. Oncol.*, **13**, 335–347.
3. Carr,T.H., McEwen,R., Dougherty,B., Johnson,J.H., Dry,J.R., Lai,Z., Ghazoui,Z., Laing,N.M., Hodgson,D.R., Cruzalegui,F. *et al.* (2016) Defining actionable mutations for oncology therapeutic development. *Nat. Rev. Cancer*, **16**, 319–329.
4. McNeil,C. (2015) NCI-MATCH launch highlights new trial design in precision-medicine era. *J. Natl. Cancer Inst.*, **107**, djv193.
5. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehar,J., Kryukov,G.V., Sonkin,D. *et al.* (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
6. 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.
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. 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.
9. 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.* (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
10. Iorio,F., Knijnenburg,T.A., Vis,D.J., Bignell,G.R., Menden,M.P., Schubert,M., Aben,N., Goncalves,E., Barthorpe,S., Lightfoot,H. *et al.* (2016) A Landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740–754.
11. Griffith,M., Spies,N.C., Krysiak,K., McMichael,J.F., Coffman,A.C., Danos,A.M., Ainscough,B.J., Ramirez,C.A., Rieke,D.T., Kujan,L. *et al.* (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.*, **49**, 170–174.
12. Yeh,P., Chen,H., Andrews,J., Naser,R., Pao,W. and Horn,L. (2013) DNA-Mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): a catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. *Clin. Cancer Res.*, **19**, 1894–1901.
13. Dienstmann,R., Jang,I.S., Bot,B., Friend,S. and Guinney,J. (2015) Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov.*, **5**, 118–123.
14. Johnson,A., Zeng,J., Bailey,A.M., Holla,V., Litzenburger,B., Lara-Guerra,H., Mills,G.B., Mendelsohn,J., Shaw,K.R. and Meric-Bernstam,F. (2015) The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov. Today*, **20**, 1433–1438.
15. Damodaran,S., Miya,J., Kautto,E., Zhu,E., Samorodnitsky,E., Datta,J., Reeser,J.W. and Roychowdhury,S. (2015) Cancer Driver Log (CanDL): catalog of potentially actionable cancer mutations. *J. Mol. Diagn.*, **17**, 554–559.
16. Hintzsche,J., Kim,J., Yadav,V., Amato,C., Robinson,S.E., Seelenfreund,E., Shellman,Y., Wisell,J., Applegate,A., McCarter,M. *et al.* (2016) IMPACT: a whole-exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples. *J. Am. Med. Inform. Assoc.*, **23**, 721–730.
17. Kibbe,W.A., Arze,C., Felix,V., Mitraka,E., Bolton,E., Fu,G., Mungall,C.J., Binder,J.X., Malone,J., Vasant,D. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
18. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a

comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
19. Zarin,D.A., Tse,T., Williams,R.J. and Carr,S. (2016) Trial reporting in ClinicalTrials.gov - the final rule. *N. Engl. J. Med.*, **375**, 1998–2004.
20. Wagner,A.H., Coffman,A.C., Ainscough,B.J., Spies,N.C., Skidmore,Z.L., Campbell,K.M., Krysiak,K., Pan,D., McMichael,J.F., Eldred,J.M. *et al.* (2015) DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.*, **44**, D1036–D1044.
21. Cingolani,P., Platts,A., Wang le,L., Coon,M., Nguyen,T., Wang,L., Land,S.J., Lu,X. and Ruden,D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.
22. Li,M.J. and Wang,J. (2015) Current trend of annotating single nucleotide variation in humans–A case study on SNVrap. *Methods*, **79-80**, 32–40.
23. Li,M.J., Sham,P.C. and Wang,J. (2012) Genetic variant representation, annotation and prioritization in the post-GWAS era. *Cell Res.*, **22**, 1505–1508.
24. Genomes Project,C., Auton,A., Brooks,L.D., Durbin,R.M., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
25. Lek,M., Karczewski,K.J., Minikel,E.V., Samocha,K.E., Banks,E., Fennell,T., O'Donnell-Luria,A.H., Ware,J.S., Hill,A.J., Cummings,B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
26. Cancer Genome Atlas Research, N., Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The cancer genome atlas pan-cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
27. Zhang,J., Baran,J., Cros,A., Guberman,J.M., Haider,S., Hsu,J., Liang,Y., Rivkin,E., Wang,J., Whitty,B. *et al.* (2011) International Cancer Genome Consortium Data Portal–a one-stop shop for cancer genomics data. *Database (Oxford)*, **2011**, bar026.
28. Forbes,S.A., Beare,D., Boutselakis,H., Bamford,S., Bindal,N., Tate,J., Cole,C.G., Ward,S., Dawson,E., Ponting,L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
29. Li,M.J., Pan,Z., Liu,Z., Wu,J., Wang,P., Zhu,Y., Xu,F., Xia,Z., Sham,P.C., Kocher,J.P. *et al.* (2016) Predicting regulatory variants with composite statistic. *Bioinformatics*, **32**, 2729–2736.
30. Jian,X., Boerwinkle,E. and Liu,X. (2014) In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.*, **42**, 13534–13544.
31. Liu,X., Wu,C., Li,C. and Boerwinkle,E. (2016) dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.*, **37**, 235–241.
32. Kircher,M., Witten,D.M., Jain,P., O'Roak,B.J., Cooper,G.M. and Shendure,J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
33. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
34. den Dunnen,J.T. and Antonarakis,S.E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.*, **15**, 7–12.
35. Zhou,W., Chen,T., Chong,Z., Rohrdanz,M.A., Melott,J.M., Wakefield,C., Zeng,J., Weinstein,J.N., Meric-Bernstam,F., Mills,G.B. *et al.* (2015) TransVar: a multilevel variant annotator for precision genomics. *Nat. Methods*, **12**, 1002–1003.
36. Thelwall,M., Haustein,S., Lariviere,V. and Sugimoto,C.R. (2013) Do altmetrics work? Twitter and ten other social web services. *PloS One*, **8**, e64841.
37. Jay,J.J. and Brouwer,C. (2016) Lollipops in the clinic: information dense mutation plots for precision medicine. *PloS One*, **11**, e0160519.
38. Li,H. (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*, **27**, 718–719.
39. Paez,J.G., Janne,P.A., Lee,J.C., Tracy,S., Greulich,H., Gabriel,S., Herman,P., Kaye,F.J., Lindeman,N., Boggon,T.J. *et al.* (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, **304**, 1497–1500.
40. Yun,C.H., Mengwasser,K.E., Toms,A.V., Woo,M.S., Greulich,H., Wong,K.K., Meyerson,M. and Eck,M.J. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 2070–2075.
41. Morgillo,F., Della Corte,C.M., Fasano,M. and Ciardiello,F. (2016) Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open*, **1**, e000060.
42. Jia,Y., Yun,C.H., Park,E., Ercan,D., Manuia,M., Juarez,J., Xu,C., Rhee,K., Chen,T., Zhang,H. *et al.* (2016) Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature*, **534**, 129–132.