D1118–D1123 *Nucleic Acids Research, 2014, Vol. 42, Database issue* *Published online 21 November 2013*
doi:10.1093/nar/gkt1129

Therapeutic target database update 2014: a resource for targeted therapeutics

Chu Qin[1,2,3], Cheng Zhang[1,4,5,6], Feng Zhu[7], Feng Xu[4,5], Shang Ying Chen[1], Peng Zhang[1], Ying Hong Li[7], Sheng Yong Yang[2], Yu Quan Wei[2], Lin Tao[1,3,*] and Yu Zong Chen[1,2,4,5,*]

[1]Bioinformatics and Drug Design Group, Department of Pharmacy, and Center for Computational Science and Engineering, National University of Singapore, 117543 Singapore, [2]Molecular Medicine Research Center, State Key Laboratory of Biotherapy, West China Hospital, West China School of Medicine, Sichuan University, Chengdu, China, [3]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 117456, Singapore, [4]College of Pharmacy and Tianjin Key Laboratory of Molecular Drug Research, Nankai University, Tianjin 300071, People's Republic of China, [5]State Key Laboratory of Medicinal Chemistry & Biology, Tianjin International Joint Academy of Biotechnology & Medicine, Tianjin 300457, People's Republic of China, [6]Computation and Systems Biology, Singapore-MIT Alliance, National University of Singapore, Singapore and [7]Innovative Drug Research Centre and College of Chemistry and Chemical Engineering, Chongqing University, Chongqing, People's Republic of China

Received September 15, 2013; Revised October 18, 2013; Accepted October 24, 2013

ABSTRACT

Here we describe an update of the Therapeutic Target Database (http://bidd.nus.edu.sg/group/ttd/ttd.asp) for better serving the bench-to-clinic communities and for enabling more convenient data access, processing and exchange. Extensive efforts from the research, industry, clinical, regulatory and management communities have been collectively directed at the discovery, investigation, application, monitoring and management of targeted therapeutics. Increasing efforts have been directed at the development of stratified and personalized medicines. These efforts may be facilitated by the knowledge of the efficacy targets and biomarkers of targeted therapeutics. Therefore, we added search tools for using the International Classification of Disease ICD-10-CM and ICD-9-CM codes to retrieve the target, biomarker and drug information (currently enabling the search of almost 900 targets, 1800 biomarkers and 6000 drugs related to 900 disease conditions). We added information of almost 1800 biomarkers for 300 disease conditions and 200 drug scaffolds for 700 drugs. We significantly expanded Therapeutic Target Database data contents to cover >2300 targets (388 successful and 461 clinical trial targets), 20 600 drugs (2003 approved and 3147 clinical trial drugs), 20 000 multitarget agents against almost 400 target-pairs and the activity data of 1400 agents against 300 cell lines.

INTRODUCTION

Modern drug development has been primarily focused on targeted therapeutics (1–3) with increasing movement toward stratified and personalized medicines (4–6). Extensive efforts from the research, industry, clinical, regulatory and management communities and the chemistry, biology, pharmaceutics and medicine disciplines have been collectively directed at the discovery, investigation, application, monitoring and management of targeted therapeutics and biomarkers (4,7–10). The knowledge of the efficacy targets and biomarkers is useful not only for the discovery and development of targeted therapeutics (11,12) but also for facilitating the development and practice of stratified and personalized medicines (4,13,14).

In particular, the information of targeted therapeutics and biomarkers may be potentially incorporated into the widely used disease classification systems for more refined classification of disease subclasses and patient subpopulations responsive to a particular treatment so as to better facilitate the diagnosis, prescription,

*To whom correspondence should be addressed. Tel: +65 6516 6877; Fax: +65 6774 6756; Email: phacyz@nus.edu.sg
Correspondence may also be addressed to L. Tao. Tel: +65 9105 6863; Fax: +65 6774 6756; Email: g0901898@nus.edu.sg

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

monitoring and management of patient care in stratified and personalized medicines. Although the information about targeted therapeutics and biomarkers can be obtained from the established drug (15), efficacy target (16) and biomarker (17–19) databases, the data retrieval tools of these databases are not specifically designed for optimally supporting such tasks. There is a need to enable data retrieval by using the widely used International Classification of Diseases (ICD) codes (20,21) for facilitating broader, more convenient and automatic data access, processing and exchange by the bench-to-clinic communities, particularly non-domain experts.

To better serve the multiple bench-to-clinic communities and to facilitate the development and practice of stratified and personalized medicines, we made several major improvements to the Therapeutic Target Database (TTD, http://bidd.nus.edu.sg/group/ttd/ttd.asp). First, we added information and search tools based on the ICD codes (22,23) for searching the targets, biomarkers, drugs and other TTD data related to various disease conditions. For more extensive coverage of potential biomarkers and for enabling their convenient access by the ICD codes, we added a significantly higher number (1755) of literature-reported biomarkers for more variety of disease conditions (365) than those in the existing biomarker databases that primarily focus on molecular biomarkers of specific disease classes (17,19) or clinically prioritized sets (18). We also added information and enabled the search of TTD data via drug scaffold names (227 scaffolds for 736 drugs and leads) for facilitating the search of the drugs, targets and diseases related to specific molecular scaffolds. Moreover, we added the Anatomical Therapeutic Chemical (ATC) Classification System codes for 1521 approved drugs for supporting the convenient and automated access of clinical drug data (24).

By using the literature search methods described in our earlier article (16), we also significantly expanded TTD contents to include 388 successful, 461 clinical trial and 1467 research targets; 2003 approved (1008 nature product derived), 3147 clinical trial, 498 discontinued clinical trial and 14 856 experimental drugs, 20 818 multitarget agents against 385 target-pairs and the activity data of 1436 drugs against 274 cell lines. These are compared with the 364 successful, 286 clinical trial and 1331 research targets; 1540 approved (939 natural product derived), 1423 clinical trial, 345 discontinued clinical trial and 14 853 experimental drugs, and 3681 multitarget agents active against 108 target pairs in our last update (16). The statistics of our updated data is summarized in Table 1.

International classification of diseases

ICD has been developed by the World Health Organization (WHO), sponsored by the United Nations, adopted by >110 countries and used by physicians, researchers, nurses, health workers, health information managers, policy makers, insurers and health program managers for defining and studying diseases, monitoring and managing health care and allocating resources (20,21). ICD codes have been regularly revised to the current version ICD-10 (20). But the previous version ICD-9 is still used by some organizations while proceeding with the transition to ICD-10 (the expected completion date for the transition to ICD-10 in the United States is October 1, 2014) (25). ICD-10 is composed of 68 000 alphanumeric codes as compared with the 13 000 numeric codes in ICD-9, thus offering more comprehensive coverage and better representation of medical conditions (20). A number of nations have developed their own adaptations of the ICD codes. For instance, the United States have developed ICD-9 and ICD-10 clinical modification ICD-9-CM (17 000 codes) and ICD-10-CM (155 000 codes) for covering additional morbidity details (26),

Table 1. Statistics of the drug targets, drugs and their structure and potency data in 2014 version of TTD database

Category	Item	2014 Update	2012 Update
Statistics of drug targets	Number of all targets	2360	2025
	Number of successful targets	388	364
	Number of clinical trial targets	461	286
	Number of research targets	1467	1331
Statistics of drugs	Number of all drugs	20 667	17 816
	Number of approved drugs (no of natural product derived drugs)	2003 (1008)	1540 (939)
	Number of clinical trial drugs (no of natural product derived drugs)	3147 (369)	1423 (369)
	Number of discontinued drugs	498	345
	Number of pre-clinical drugs	163	165
	Number of experimental drugs	14 856	14 853
	Number of multitarget agents	20 818	3681
	Number of drug combinations	115	115
Statistics of drugs with available structure or sequence data	Number of small molecular drugs with available structure	17 012	14 170
	Number of antisense drugs with available sequence data	652	652
Statistics of drugs with activity data or structure-activity relationship	Number of agents with potency data against target	11 810	11 810
	Number of agents with potency data against a disease model such as a cell-line, *ex vivo*, *in vivo* model	1753	497
	Number of quantitative structure-activity relationship qsar models (no of chemical types)	841 (228)	841 (228)

which were used in TTD because of their more comprehensive coverage.

The ICD-9-CM and ICD-10-CM codes were matched to the TTD target, drug and biomarker entries by the following procedure. First, automated word match was conducted for matching the disease name or names of each TTD target, drug or biomarker entry with the disease descriptions of each ICD codes. Second, each of the fully or partially matched TTD entry was manually checked to either validate the match or to find the right ICD codes. Third, manual search was conducted for every non-matched TTD entries. So far, we were able to find the ICD codes for 785 targets and 3080 drugs related to 732 disease conditions. From the TTD 'Search drugs and targets by disease or ICD identifier' field, users can search TTD target and drug entries related to a specific disease or an ICD-9-CM or ICD-10-CM code. The TTD biomarker entries may also be searched by selecting an ICD-9-CM or ICD-10-CM code from the 'Search for biomarkers' field. Users may also download from the TTD download page the lists of TTD target, drug and biomarker entries with the corresponding ICD-9-CM and ICD-10-CM codes.

A new ICD version ICD-11 is in development and scheduled for endorsement by WHO in 2015 (WHO. The International Classification of Diseases 11th revision is due by 2015. Retrieved from http://www.who.int/classifications/icd/revision/en/), which offers more refined disease classifications based on more recent scientific understanding of the disease mechanisms. For instance, small cell lung cancer, which represents ~13% of all lung cancer diagnoses (27), is not explicitly classified in the ICD-10 and earlier ICD versions but is now explicitly represented in the ICD-11 beta draft. Therefore, ICD-11 is expected to be more useful for developing a more refined disease classification system for stratified and personalized medicine. Effort will be made to upgrade TTD to the ICD-11 version on its official release.

Biomarkers

Biomarkers have been developed as non-invasive tests for early detection and indication of disease risks, monitoring of disease progression and recurrence and classification of disease subtypes and patient subpopulations for providing the most appropriate treatments (28–30). As many therapies have been found to elicit markedly different clinical responses in individual patients (31,32), there is a particular need for more biomarkers capable of predicting drug response in individual patients, which has led to intensive efforts in the discovery of such biomarkers (4,33). Table 2 gives examples of the approved and clinically tested biomarkers for facilitating the prescription of a particular drug to specific patient subpopulation. Moreover, there are considerable interests in adopting the multimarker strategy for parallel evaluation of multiple existing and novel biomarkers in the diagnosis and prognostics of diseases and treatment responses in individual patients (34,35). These efforts may be facilitated by significantly expanding biomarker coverage in the biomarker databases. We, therefore, searched literature-reported biomarkers, mapped them to the ICD-9-CM and ICD-10-CM codes and added the relevant information and ICD code search tools in TTD.

To broadly cover various types of biomarkers, we conducted comprehensive literature search in the PubMed database (36) by using combination of keywords 'biomarker', 'clinical', 'patient', 'disease', 'drug' and specific disease names. Additional sources such as the FDA website and the abstracts of the American society of clinical oncology were also systematically searched. Overall we collected 1755 biomarkers for 365 disease conditions, which include both process biomarkers (genetic mutations or alterations, gene amplification and levels of proteins, gene expression, microRNAs, small molecules, or metabolites that capture a molecular/biochemical aspect of disease pathogenesis and the biological responses to the disease process and/or treatment) and global biomarkers (such as tumor sizes, brain structures in neurodegeneration and shape of cells in anemia). These biomarkers may be searched in the 'Search for biomarkers' field by using keywords or by selecting an ICD-9-CM or ICD-10-CM code.

Based on the literature descriptions, our collected biomarkers were classified into one or more of the following 12 classes: associative (disease correlation), antecedent (pre-illness risk identification), detective (disease early stage detection), classification (disease categorization and patient assignment for differential treatment), differentiative (differentiation of related diseases), diagnostic (recognition of overt diseases), monitoring (monitoring of disease state or treatment response), pharmacodynamic (examination of the biological basis of drug response variations), pharmacogenomic (genomics-based prediction of drug response, adverse drug reaction and appropriate drug dose), prognostic (prediction of future disease course and response to therapy), surrogate (substitute of a clinical end point for predicting therapeutic benefit) and theragnostic (identification and monitoring of biochemical effects or mode of action of drug and downstream processes) classes.

Apart from the literature-reported biomarkers, the profiles of various known drug resistance mutations (37–39) and drug response regulators (e.g. the genes promoting drug bypass signaling (40,41) or hindering drug actions (42) have been studied for predicting drug resistance, which may be potentially explored as drug response biomarkers (43). Potential biomarkers, particularly multimarkers, have also been predicted from the genetic and gene expression data of patients by using such computational methods as the principal components analysis feature selection method (44), weighted voting classification feature selection method (45), hierarchical clustering feature selection method (46), differentially expressed genes method (47,48) and machine learning feature selection methods (49,50). These potential biomarkers may also be included in TTD and other biomarker databases for facilitating their future exploration.

More refined classification of patient subpopulations for targeted therapeutics

From the examples of the approved and clinically tested drug response biomarkers in Table 2, it seems feasible to

Table 2. Examples of the approved and clinically tested biomarkers for facilitating the prescription of a particular drug to specific patient subpopulation

Disease	Therapeutic target	Biomarker for the targeted therapeutics	Patient subpopulation likely responsive to targeted therapeutics	Drug therapy specific for patient subpopulation
Acute promyelocytic leukemia (APL)	PML–RAR	PML–RAR (gene translocation)	APL with PML–RARα t(15:17) translocation	Arsenic Trioxide
Alzheimer's	PPAR	apolipoprotein E and TOMM40 genotypes and age	Mild cognitive impairment due to Alzheimer's disease	Pioglitazone
Breast cancer	HER2	HER2 (gene amplification)	HER2 amplified and/or over-expressed breast cancer	Trastuzumab
	Estrogen receptor	Estrogen receptor (protein expression)	ER overexpressed breast cancer	Tamoxifen
	PARP	BRCA1/2 (mutation)	Breast cancer defective in BRCA1 or BRCA2	Olaparib, veliparib
Cystic fibrosis	CFTR	CFTR G551D mutation	Cystic fibrosis patients with CFTR G551D mutation	Ivacaftor
Hepatitis C infection	HCV non-structural protein 3	IL28B rs12979860 genotype	HCV infected patients with IL28B rs12979860 genotype	Boceprevir
Melanoma	BRAF	BRAF V600E (mutation)	Melanoma with RAF V600E mutation	Vemurafenib, Dabrafenib
	MEK	BRAF mutations	Melanoma with RAF mutations	Trametinib
Post-menopausal osteoporosis	RANK ligand	Post-menopausal women with persistent total hip, femoral neck, or lumbar spine BMD T-scores −1.8 to −4.0, or clinical fracture	Post-menopausal osteoporosis at high risk for fractures	Denosumab

incorporate target and biomarker codes into the ICD codes for more refined classification of patient subpopulations responsive to a particular targeted therapy. However, many of the existing biomarkers are based on the profile of a single gene. For highly heterogenetic diseases such as cancers, single-gene biomarkers are highly limited in their coverage of drug escape mechanisms, and multimarkers may be needed for more sufficient coverage of drug escape mechanisms and for more accurate classification of patient subpopulations in stratified and personal medicines. For instance, $BRAF^{V600E}$ inhibitor dabrafenib has shown improved therapeutic effect in $BRAF^{V600E}$ metastatic melanoma patients (51) due in part to its specificity to $BRAF^{V600E}$ tumors with a greater therapeutic window (52). However, drug resistance still emerges (51) partly due to tumor activation of several BRAF inhibitor escape pathways (52–54). Therefore, the use of a single-gene biomarker, $BRAF^{V600E}$ mutation, is insufficient for predicting long-term drug response to dabrafenib therapy, and multimarkers are needed for adequately covering these and other active drug escape mechanisms.

Drug scaffolds

The approved and clinical trial drugs are composed of a limited number of molecular scaffolds (55–57) in contrast to the high number of bioactive molecular scaffolds (58,59). For instance, many drugs have been derived from individual scaffold groups such as macrocycles (60), and 12 FDA-approved anticancer kinase inhibitor drugs (61,62) are grouped into three scaffold groups (63). Investigation and exploration of these highly privileged drug scaffolds are important for discovering new drug-like scaffolds, molecular analogs and drugs. To support the relevant efforts, we searched the literatures for the molecular scaffolds of the approved and clinical trial drugs or their drug leads. By using the combination of keywords drug name or alternative name, 'scaffold', 'discovery', 'synthesis' to search the Pubchem database (36), we found 210 scaffolds for 714 drugs and drug leads. Users can search the TTD drug and target entries related to a drug scaffold by keyword search or by selecting from the list of drug scaffold names in the 'Search for drug scaffolds' field.

Remarks

The efforts in the discovery and application of targeted therapeutics increasingly involve collective efforts from multiple bench-to-clinic communities (1–3) and these efforts are increasingly directed at the development of stratified and personalized medicines (4–6). The drug, target, biomarker and other relevant chemical, biological, pharmaceutical and clinical data need to be more integrated and be made easily accessible by the multiple bench-to-clinic communities. These efforts may be partly facilitated by introducing into the relevant databases the ICD code-based data retrieval tools coupled to the other domain knowledge codes such as the codes of drugs (e.g. ATC codes), targets and biomarkers. Continuous efforts will be made to expand the linkage of the ICD and ATC codes to more complete sets of drugs, efficacy targets and biomarkers and to provide the latest and comprehensive information about the drugs, efficacy targets and biomarkers for better serving the multiple bench-to-clinic

communities in their collective efforts for the discovery, investigation, application, monitoring and management of targeted therapeutics.

FUNDING

Singapore Academic Research Fund [R-148-000-181-112]; National Natural Science Foundation of China [81202459]; Chongqing Natural Science Foundation [cstc2012jjA10116]; Fundamental Research Funds for the Central Universities [CQDXWL-2012-Z003]; Start up founding of Youth 100 Talents program of Chongqing University [0903005203176]; The Major State Basic Research Development Program of China [2013CB967204]. Funding for open access charge: National Natural Science Foundation of China [81202459].

Conflict of interest statement. None declared.

REFERENCES

1. Zheng,C.J., Han,L.Y., Yap,C.W., Ji,Z.L., Cao,Z.W. and Chen,Y.Z. (2006) Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol. Rev.*, **58**, 259–279.
2. Overington,J.P., Al-Lazikani,B. and Hopkins,A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.
3. Rask-Andersen,M., Almen,M.S. and Schioth,H.B. (2011) Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.*, **10**, 579–590.
4. La Thangue,N.B. and Kerr,D.J. (2011) Predictive biomarkers: a paradigm shift towards personalized cancer medicine. *Nat. Rev. Clin. Oncol.*, **8**, 587–596.
5. Trusheim,M.R., Burgess,B., Hu,S.X., Long,T., Averbuch,S.D., Flynn,A.A., Lieftucht,A., Mazumder,A., Milloy,J., Shaw,P.M. *et al.* (2011) Quantifying factors for the success of stratified medicine. *Nat. Rev. Drug Discov.*, **10**, 817–833.
6. Volzke,H., Schmidt,C.O., Baumeister,S.E., Ittermann,T., Fung,G., Krafczyk-Korth,J., Hoffmann,W., Schwab,M., Meyer zu Schwabedissen,H.E., Dorr,M. *et al.* (2013) Personalized cardiovascular medicine: concepts and methodological considerations. *Nat. Rev. Cardiol.*, **10**, 308–316.
7. Kneller,R. (2010) The importance of new companies for drug discovery: origins of a decade of new drugs. *Nat. Rev. Drug Discov.*, **9**, 867–882.
8. Bunnage,M.E. (2011) Getting pharmaceutical R&D back on target. *Nat. Chem. Biol.*, **7**, 335–339.
9. Chataway,J., Fry,C., Marjanovic,S. and Yaqub,O. (2012) Public-private collaborations and partnerships in stratified medicine: making sense of new interactions. *N. Biotechnol.*, **29**, 732–740.
10. Maliepaard,M., Nofziger,C., Papaluca,M., Zineh,I., Uyama,Y., Prasad,K., Grimstein,C., Pacanowski,M., Ehmann,F., Dossena,S. *et al.* (2013) Pharmacogenetics in the evaluation of new drugs: a multiregional regulatory perspective. *Nat. Rev. Drug Discov.*, **12**, 103–115.
11. Engelman,J.A., Zejnullahu,K., Mitsudomi,T., Song,Y., Hyland,C., Park,J.O., Lindeman,N., Gale,C.M., Zhao,X., Christensen,J. *et al.* (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*, **316**, 1039–1043.
12. Chen,Z., Cheng,K., Walton,Z., Wang,Y., Ebi,H., Shimamura,T., Liu,Y., Tupper,T., Ouyang,J., Li,J. *et al.* (2012) A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature*, **483**, 613–617.
13. Hall,I.P. (2013) Stratified medicine: drugs meet genetics. *Eur. Respir. Rev.*, **22**, 53–57.
14. Fugel,H.J., Nuijten,M. and Postma,M. (2012) Stratified medicine and reimbursement issues. *Front Pharmacol.*, **3**, 181.
15. Knox,C., Law,V., Jewison,T., Liu,P., Ly,S., Frolkis,A., Pon,A., Banco,K., Mak,C., Neveu,V. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
16. Zhu,F., Shi,Z., Qin,C., Tao,L., Liu,X., Xu,F., Zhang,L., Song,Y., Zhang,J., Han,B. *et al.* (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.*, **40**, D1128–D1136.
17. Yang,I.S., Ryu,C., Cho,K.J., Kim,J.K., Ong,S.H., Mitchell,W.P., Kim,B.S., Oh,H.B. and Kim,K.H. (2008) IDBD: infectious disease biomarker database. *Nucleic Acids Res.*, **36**, D455–D460.
18. Srivastava,S. (2013) The early detection research network: 10-year outlook. *Clin. Chem.*, **59**, 60–67.
19. 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.
20. Bramer,G.R. (1988) International statistical classification of diseases and related health problems. Tenth revision. *World Health Stat. Q.*, **41**, 32–36.
21. Wood,P.H. (1990) Applications of the International Classification of Diseases. *World Health Stat. Q.*, **43**, 263–268.
22. (1986) ICD-9-CM. International Classification of Diseases, 9th edn, Clinical Modification: update. Official authorized addendum, effective 1 October 1986. *J. Am. Med. Rec. Assoc.*, **57**, S1–S32.
23. Steindel,S.J. (2010) International classification of diseases, 10th edn, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. *J. Am. Med. Inform. Assoc.*, **17**, 274–282.
24. Lim,C.M., Aryani Md Yusof,F., Selvarajah,S. and Lim,T.O. (2011) Use of ATC to describe duplicate medications in primary care prescriptions. *Eur. J. Clin. Pharmacol.*, **67**, 1035–1044.
25. Averill,R. and Bowman,S. (2012) There are critical reasons for not further delaying the implementation of the new ICD-10 coding system. *J. AHIMA*, **83**, 42–48.
26. Topaz,M., Shafran-Topaz,L. and Bowles,K.H. (2013) ICD-9 to ICD-10: evolution, revolution, and current debates in the United States. *Perspect Health Inf. Manag.*, **10**, 1d.
27. Califano,R., Abidin,A.Z., Peck,R., Faivre-Finn,C. and Lorigan,P. (2012) Management of small cell lung cancer: recent developments for optimal care. *Drugs*, **72**, 471–490.
28. Robinson,W.H., Lindstrom,T.M., Cheung,R.K. and Sokolove,J. (2013) Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat. Rev. Rheumatol.*, **9**, 267–276.
29. Ludwig,J.A. and Weinstein,J.N. (2005) Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer*, **5**, 845–856.
30. Walsh,P., Elsabbagh,M., Bolton,P. and Singh,I. (2011) In search of biomarkers for autism: scientific, social and ethical challenges. *Nat. Rev. Neurosci.*, **12**, 603–612.
31. Ho,C. and Laskin,J. (2009) EGFR-directed therapies to treat non-small-cell lung cancer. *Expert. Opin. Investig. Drugs*, **18**, 1133–1145.
32. Linardou,H., Dahabreh,I.J., Bafaloukos,D., Kosmidis,P. and Murray,S. (2009) Somatic EGFR mutations and efficacy of tyrosine kinase inhibitors in NSCLC. *Nat. Rev. Clin. Oncol.*, **6**, 352–366.
33. Blennow,K., Hampel,H. and Zetterberg,H. (2013) Biomarkers in amyloid-beta immunotherapy trials in alzheimer's disease. *Neuropsychopharmacology*, **39**, 189–201.
34. Ortiz,A., Massy,Z.A., Fliser,D., Lindholm,B., Wiecek,A., Martinez-Castelao,A., Covic,A., Goldsmith,D., Suleymanlar,G., London,G.M. *et al.* (2012) Clinical usefulness of novel prognostic biomarkers in patients on hemodialysis. *Nat. Rev. Nephrol.*, **8**, 141–150.
35. Ahmad,T., Fiuzat,M., Felker,G.M. and O'Connor,C. (2012) Novel biomarkers in chronic heart failure. *Nat. Rev. Cardiol.*, **9**, 347–359.
36. NCBI Resource Coordinators. (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **41**, D8–D20.
37. Cools,J., Mentens,N., Furet,P., Fabbro,D., Clark,J.J., Griffin,J.D., Marynen,P. and Gilliland,D.G. (2004) Prediction of resistance to

small molecule FLT3 inhibitors: implications for molecularly targeted therapy of acute leukemia. *Cancer Res.*, **64**, 6385–6389.
38. Rizvi,N.A., Rusch,V., Pao,W., Chaft,J.E., Ladanyi,M., Miller,V.A., Krug,L.M., Azzoli,C.G., Bains,M., Downey,R. *et al.* (2011) Molecular characteristics predict clinical outcomes: prospective trial correlating response to the EGFR tyrosine kinase inhibitor gefitinib with the presence of sensitizing mutations in the tyrosine binding domain of the EGFR gene. *Clin. Cancer Res.*, **17**, 3500–3506.
39. Molinari,F., Felicioni,L., Buscarino,M., De Dosso,S., Buttitta,F., Malatesta,S., Movilia,A., Luoni,M., Boldorini,R., Alabiso,O. *et al.* (2011) Increased detection sensitivity for KRAS mutations enhances the prediction of anti-EGFR monoclonal antibody resistance in metastatic colorectal cancer. *Clin. Cancer Res.*, **17**, 4901–4914.
40. Wang,W., Cassidy,J., O'Brien,V., Ryan,K.M. and Collie-Duguid,E. (2004) Mechanistic and predictive profiling of 5-Fluorouracil resistance in human cancer cells. *Cancer Res.*, **64**, 8167–8176.
41. Dai,Z., Barbacioru,C., Huang,Y. and Sadee,W. (2006) Prediction of anticancer drug potency from expression of genes involved in growth factor signaling. *Pharm. Res.*, **23**, 336–349.
42. Brase,J.C., Schmidt,M., Fischbach,T., Sultmann,H., Bojar,H., Koelbl,H., Hellwig,B., Rahnenfuhrer,J., Hengstler,J.G. and Gehrmann,M.C. (2010) ERBB2 and TOP2A in breast cancer: a comprehensive analysis of gene amplification, RNA levels, and protein expression and their influence on prognosis and prediction. *Clin. Cancer Res.*, **16**, 2391–2401.
43. Zhang,J., Jia,J., Zhu,F., Ma,X., Han,B., Wei,X., Tan,C., Jiang,Y. and Chen,Y. (2012) Analysis of bypass signaling in EGFR pathway and profiling of bypass genes for predicting response to anticancer EGFR tyrosine kinase inhibitors. *Mol. Biosyst.*, **8**, 2645–2656.
44. Hilsenbeck,S.G., Friedrichs,W.E., Schiff,R., O'Connell,P., Hansen,R.K., Osborne,C.K. and Fuqua,S.A. (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl Cancer Inst.*, **91**, 453–459.
45. Staunton,J.E., Slonim,D.K., Coller,H.A., Tamayo,P., Angelo,M.J., Park,J., Scherf,U., Lee,J.K., Reinhold,W.O., Weinstein,J.N. *et al.* (2001) Chemosensitivity prediction by transcriptional profiling. *Proc. Natl Acad. Sci. USA*, **98**, 10787–10792.
46. Rosenwald,A., Wright,G., Chan,W.C., Connors,J.M., Campo,E., Fisher,R.I., Gascoyne,R.D., Muller-Hermelink,H.K., Smeland,E.B., Giltnane,J.M. *et al.* (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.*, **346**, 1937–1947.
47. Takata,R., Katagiri,T., Kanehira,M., Tsunoda,T., Shuin,T., Miki,T., Namiki,M., Kohri,K., Matsushita,Y., Fujioka,T. *et al.* (2005) Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. *Clin. Cancer Res.*, **11**, 2625–2636.
48. Balko,J.M., Potti,A., Saunders,C., Stromberg,A., Haura,E.B. and Black,E.P. (2006) Gene expression patterns that predict sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer cell lines and human lung tumors. *BMC Genomics*, **7**, 289.
49. Okano,T., Kondo,T., Fujii,K., Nishimura,T., Takano,T., Ohe,Y., Tsuta,K., Matsuno,Y., Gemma,A., Kato,H. *et al.* (2007) Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor tyrosine kinase inhibitor in lung adenocarcinoma. *Clin. Cancer Res.*, **13**, 799–805.
50. Ma,Y., Ding,Z., Qian,Y., Shi,X., Castranova,V., Harner,E.J. and Guo,L. (2006) Predicting cancer drug response by proteomic profiling. *Clin. Cancer Res.*, **12**, 4583–4589.
51. Hauschild,A., Grob,J.J., Demidov,L.V., Jouary,T., Gutzmer,R., Millward,M., Rutkowski,P., Blank,C.U., Miller,W.H. Jr, Kaempgen,E. *et al.* (2012) Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*, **380**, 358–365.
52. Cichowski,K. and Janne,P.A. (2010) Drug discovery: inhibitors that activate. *Nature*, **464**, 358–359.
53. Nazarian,R., Shi,H., Wang,Q., Kong,X., Koya,R.C., Lee,H., Chen,Z., Lee,M.K., Attar,N., Sazegar,H. *et al.* (2010) Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*, **468**, 973–977.
54. Li,Q.L., Gu,F.M., Wang,Z., Jiang,J.H., Yao,L.Q., Tan,C.J., Huang,X.Y., Ke,A.W., Dai,Z., Fan,J. *et al.* (2012) Activation of PI3K/AKT and MAPK pathway through a PDGFRbeta-dependent feedback loop is involved in rapamycin resistance in hepatocellular carcinoma. *PLoS One*, **7**, e33379.
55. Bemis,G.W. and Murcko,M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.*, **39**, 2887–2893.
56. Wang,J. and Hou,T. (2010) Drug and drug candidate building block analysis. *J. Chem. Inf. Model*, **50**, 55–67.
57. Duarte,C.D., Barreiro,E.J. and Fraga,C.A. (2007) Privileged structures: a useful concept for the rational design of new lead drug candidates. *Mini Rev. Med. Chem.*, **7**, 1108–1119.
58. Koch,M.A., Schuffenhauer,A., Scheck,M., Wetzel,S., Casaulta,M., Odermatt,A., Ertl,P. and Waldmann,H. (2005) Charting biologically relevant chemical space: a structural classification of natural products (SCONP). *Proc. Natl Acad. Sci. USA*, **102**, 17272–17277.
59. Kong,D.X., Jiang,Y.Y. and Zhang,H.Y. (2010) Marine natural products as sources of novel scaffolds: achievement and concern. *Drug Discov. Today*, **15**, 884–886.
60. Mallinson,J. and Collins,I. (2012) Macrocycles in new drug discovery. *Future Med. Chem.*, **4**, 1409–1438.
61. Newman,D.J. and Cragg,G.M. (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.*, **70**, 461–477.
62. Butler,M.S. (2008) Natural products to drugs: natural product-derived compounds in clinical trials. *Nat. Prod. Rep.*, **25**, 475–516.
63. Zhu,F., Qin,C., Tao,L., Liu,X., Shi,Z., Ma,X., Jia,J., Tan,Y., Cui,C., Lin,J. *et al.* (2011) Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. *Proc. Natl Acad. Sci. USA*, **108**, 12943–12948.