Nucleic Acids Research, 2024, **52**, D1450–D1464
https://doi.org/10.1093/nar/gkad862
Advance access publication date: 18 October 2023
Database issue

TheMarker: a comprehensive database of *therapeutic biomarkers*

Yintao Zhang[1,2,†], Ying Zhou[1,3,†], Yuan Zhou[1], Xinyuan Yu[1], Xinyi Shen[4], Yanfeng Hong[5], Yuxin Zhang[1], Shanshan Wang[6], Minjie Mou[1], Jinsong Zhang[1], Lin Tao[5], Jianqing Gao[1], Yunqing Qiu[1,3,*], Yuzong Chen[7,8,*] and Feng Zhu[1,2,*]

[1]College of Pharmaceutical Sciences, The First Affiliated Hospital, The Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310058, China
[2]Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China
[3]National Key Laboratory of Diagnosis and Treatment of Severe Infectious Disease, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang Provincial Key Laboratory for Drug Clinical Research and Evaluation, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, 310000, China
[4]Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven 06510, USA
[5]Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, School of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China
[6]Qian Xuesen Collaborative Research Center of Astrochemistry and Space Life Sciences, Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China
[7]State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, The Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
[8]Institute of Biomedical Health Technology and Engineering, Shenzhen Bay Laboratory, Shenzhen 518000, China

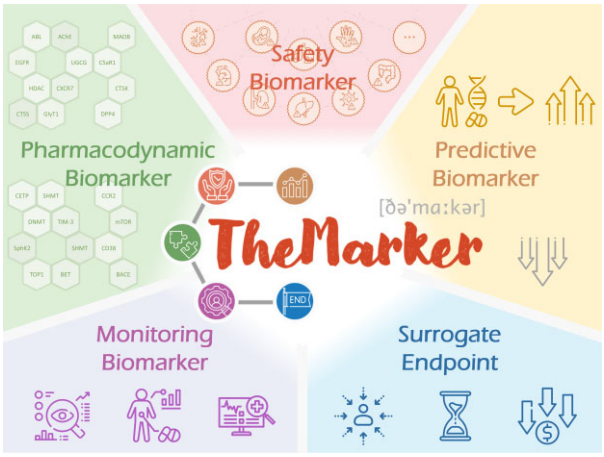
[*]To whom correspondence should be addressed. Tel: +86 189 8946 6518; Fax: +86 571 8820 8444; Email: zhufeng@zju.edu.cn
Correspondence may also be addressed to Yuzong Chen. Email: chenyuzong@sz.tsinghua.edu.cn
Correspondence may also be addressed to Yunqing Qiu. Email: qiuyq@zju.edu.cn
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Distinct from the traditional diagnostic/prognostic biomarker (adopted as the indicator of disease state/process), the *therapeutic biomarker* (ThMAR) has emerged to be very crucial in the clinical development and clinical practice of all therapies. There are five types of ThMAR that have been found to play indispensable roles in various stages of drug discovery, such as: *Pharmacodynamic Biomarker* essential for guaranteeing the pharmacological effects of a therapy, *Safety Biomarker* critical for assessing the extent or likelihood of therapy-induced toxicity, *Monitoring Biomarker* indispensable for guiding clinical management by serially measuring patients' status, *Predictive Biomarker* crucial for maximizing the clinical outcome of a therapy for specific individuals, and *Surrogate Endpoint* fundamental for accelerating the approval of a therapy. However, these data of ThMARs has not been comprehensively described by any of the existing databases. Herein, a database, named 'TheMarker', was therefore constructed to (a) systematically offer all five types of ThMAR used at different stages of drug development, (b) comprehensively describe ThMAR information for the largest number of drugs among available databases, (c) extensively cover the widest disease classes by not just focusing on anticancer therapies. These data in TheMarker are expected to have great implication and significant impact on drug discovery and clinical practice, and it is freely accessible without any login requirement at: https://idrblab.org/themarker.

Received: July 19, 2023. **Revised:** September 21, 2023. **Editorial Decision:** September 25, 2023. **Accepted:** September 29, 2023

Graphical abstract



Introduction

Distinct from the traditional diagnostic/prognostic biomarker (adopted as the indicator of disease state/process), the *therapeutic biomarker* (ThMAR) has emerged to be very crucial in the clinical development and clinical practice of all therapies (1,2). There are five types of ThMAR that are deeply involved in various stages of drug discovery (as illustrated in Figure 1), which are defined by the BEST category of U.S. Food & Drug Administration (3) as: *pharmacodynamic biomarker* (PDY), *safety biomarker* (SAF), *monitoring biomarker* (MOI), *predictive biomarker* (PRD), and *surrogate endpoint* (SUR). For every therapy (shown in Table 1), the PDYs, SAFs, MOIs, PRDs, and SURs are reported to be crucial for guaranteeing pharmacological effect using its targets (4), fundamental for assessing the extent or likelihood of therapy-induced toxicity (5), indispensable for guiding clinical management by serially measuring patient status (6), critical for maximizing the clinical outcome of a therapy for particular group of patients (7), and valuable for accelerating the approval of a therapy using smaller patient number and shorter trial period (8). With the rapid accumulation of ThMAR data in recent years, it is highly demanded to have a database providing the information of five types of ThMAR, which should be collectively assessed considering the extremely high-level of interplay among different stages of drug development (9,10).

So far, a variety of biomarker-relevant databases have been constructed, most of which focus on giving *diagnostic/prognostic biomarkers*, such as MarkerDB (11), Lnc2Cancer (12), Exposome-Explorer (13), BioMuta & BioXpress (14) and several other databases (15–33). These databases have accumulated great research interests from worldwide audience, but they do not provide any ThMAR data. Two databases have been available for providing ThMAR information: CTR-DB (34) and ResMarkerDB (35). However, these databases mainly focus on providing the *predictive biomarker* (PRD, one of the five ThMAR types shown in Figure 1) for anticancer therapy, which make them unable to assess the interplay among different discovery stages (10). Moreover, most (>60%) of ThMARs are not for anticancer therapy, which limits the use of the available databases (9). Therefore, it is urgently needed to construct a database for all five ThMAR types.

In this study, a comprehensive database of therapeutic biomarkers entitled 'TheMarker' was thus constructed. It was the first knowledge base covering all five types of ThMAR, which allowed a collective consideration among different stages of drug development. TheMarker contained: (a) 218 *pharmacodynamic biomarkers* indicating the clinical efficacies of 115 drug classes (such as: AChE inhibitors, MetAP2 inhibitors and LPA1 antagonists) for the treatments of 112 classes of disease defined by the WHO ICD-11 (such as *Alzheimer* disease, obesity and systemic sclerosis); (b) 624 *safety biomarkers* that monitored the clinical toxicity (such as gastrointestinal, hepatic, and hematological) of 263 drugs treating 106 disease classes (such as thrombocytopenia, seizure and *Parkinson*); (c) 104 *monitoring biomarkers* that helped to guide the clinical management of a therapy through serially measuring patient status for 60 drugs treating 33 disease classes (such as hemophilia and melanoma); (d) 15 893 *predictive biomarkers* that facilitated the identification of individuals who are more likely to experience favorable or unfavorable effect from 352 drugs for treating 95 diseases (such as hepatitis and hypercholesterolemia); (e) 103 *surrogate endpoints* that provided the clinical outcomes of 435 approved drugs (including 193 accelerated approvals) for treating 102 diseases (such as tuberculosis, muscular dystrophy and *Fabry* disease).

In sum, TheMarker systematically provided five types of ThMAR, which described ThMAR data for the largest number of drugs among all those available databases, and covered the widest range of disease classes, which provided the most diverse pathological data among available databases by not just focusing on anticancer therapies. Due to the rapid application of *Artificial Intelligence* in biomedical studies (36–39), the comprehensive data provided in this database may be valuable for both drug development and clinical practice. TheMarker is now freely accessible without any login requirement at: https://idrblab.org/themarker.

Factual content and data retrieval

Systematic collection of the information of therapeutic biomarkers

The *therapeutic biomarkers* (ThMARs) and their applications in the drug development & clinical practice were

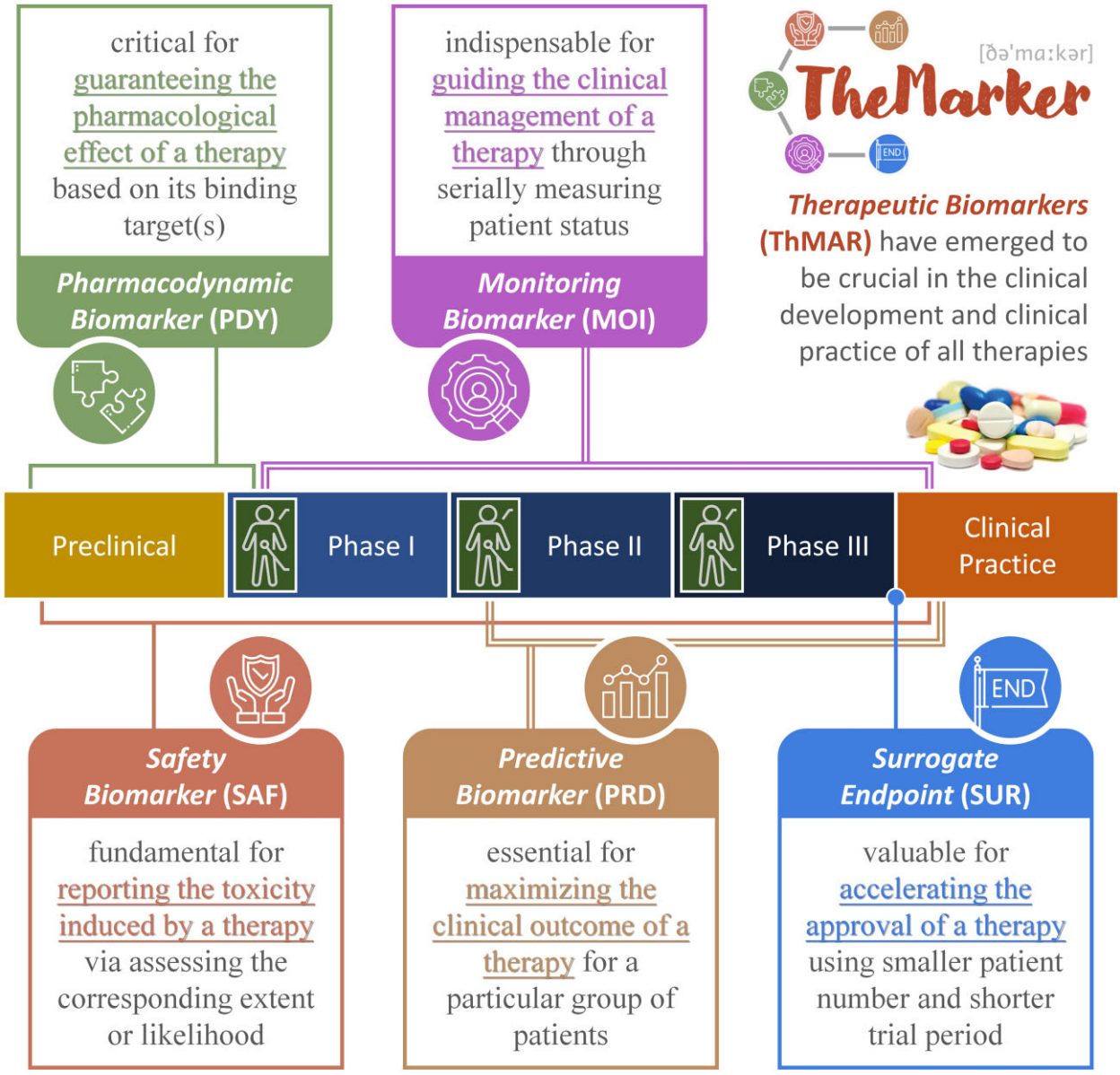


Figure 1. Five distinct types of *therapeutic biomarker* (ThMAR) and their key role in the clinical development and clinical practice of all therapies, which included: *pharmacodynamic biomarker* (PDY), *safety biomarker* (SAF), *monitoring biomarker* (MOI), *predictive biomarker* (PRD), and *surrogate endpoint* (SUR). The ThMARs were deeply involved in every stages of drug discovery and known to be essential for guaranteeing the pharmacological effect, fundamental for assessing the extent/likelihood of therapy-induced toxicity, indispensable for guiding clinical management, critical for maximizing the clinical outcome, and valuable for accelerating therapy approval.

collected based on the following procedure. *First*, comprehensive literature review was conducted using such keywords/combinations as 'therapeutic biomarker + drug', 'treatment response + biomarker', 'pharmacodynamic biomarker', 'target engagement biomarker', 'drug safety biomarker', and 'surrogate endpoint'. Retrieved literatures were then carefully reviewed, and those reported ThMARs together with their corresponding therapies were recorded. *Second*, the valuable data of pharmacogenomic biomarker officially provided by the U.S. FDA-approved drug labels were extracted, which were scattered throughout the different sections of these labels such as *indications and usage*, *adverse reactions*, and *use in specific populations*. According to the roles of these biomarkers played, the types of these biomarkers were manually labelled. *Third*, the surrogate endpoints that have been applied to facilitate drug approval were comprehensively collected from the official website of U.S. FDA, and the drugs approved based on these surrogate endpoints were also identified. In addition, detailed descriptions on the applications of ThMARs in clinical development and clinical practice were also identified, which included biomarker class (such as protein, DNA, and miRNA), biomarker mode (such as expression level, mutation, urine concentration, and polymorphisms), biomarker source (such as tumor tissue, plasma, and urine), experimental testing method (such as ELISA and RT-PCR), reported biomarker variation, and so on. Moreover, one of the most widely applied strategies for categorizing biomarkers, also known as the 'BEST' category officially provided by U.S. Food & Drug Administration (3), was adopted by the study for

classifying all collected ThMARs into five types: *pharmacodynamic biomarker* (PDY), *safety biomarker* (SAF), *monitoring biomarker* (MOI), *predictive biomarker* (PRD) and *surrogate endpoint* (SUR). The definition and importance of these ThMAR types were explicitly described in Table 1, and a detailed discussion on five ThMAR types was provided below.

Pharmacodynamic biomarker guaranteeing the clinical efficacy of a therapy

Measuring the binding of drug to its targets and determining the association of drug efficacy with target engagement are essential steps in target validation and drug discovery, which heavily relies on the utilization of *pharmacodynamic biomarker* (PDY), especially in human trials (40). PDYs are considered as critical tool for initially assessing the beneficial therapeutic activity, supporting clinical translation from animal to human, and providing valuable data on mechanisms of action, dose response, and drug efficacy (4). A retrospective study revealed that ∼20% of the failures in *Phase 2* clinical trial was due to inadequate target exposure, emphasizing the importance of PDY (40), and it was also reported that the clinical proof of the target binding mechanism using PDYs could substantially increase the probability of a project advancing to *Phase 2* by 25% (10).

Herein, the PDYs were systematically collected based on literature review, and an exemplar PDY for *Alzheimer* disease were described in Figure 2. General information of a studied disease was provided in the upper section, which included disease name, disease class and ICD-11. All PDYs that have potential application in the disease was categorized by biomarker class (such as protein and chemical). For each PDY, the corresponding drug and its therapeutic class, analyzed species, marker source, and testing method were provided. As a result, a total of 218 PDYs indicating the efficacy of 115 therapeutic classes for treating 112 diseases were provided in TheMarker.

Safety biomarker evaluating the likelihood/extent of therapy-induced toxicity

Drug safety was widely and persistently considered during the process of drug development and clinical practice (41–44). In the preclinical and clinical phases of drug development, safety issue remained one of the critical reasons of drug attrition, accounting for over 30% of all drug failures (45,46). One of the effective ways to prevent/mitigate the drug-induced toxicity was to use *safety biomarker* (SAF) in the early drug development, which offered drug developer with guidance to optimize drug candidate and then increase the likelihood of success (5,46). Moreover, it is known that numerous adverse drug reactions were not directly observed in clinical trial but identified in post-marketing surveillance (47,48). Thus, the discovery of SAF can shed light on the underlying molecular mechanism of drug toxicity, differentiate compounds with less toxicity to advance into clinical trial, and lead to safer treatment with much lower morbidity and mortality (49–54).

Herein, an exemplar SAF and its applications for reporting therapy-induced toxicity were shown in Figure 3. General information of the SAF was offered, which included SAF name, synonyms, and so on, and the drugs whose safety could be assessed using the SAF were grouped using their clinical status. For a drug, reported biomarker variation, therapy-induced toxicity, studied disease, experimental species, biomarker source and testing method were explicitly described. The SAFs collected here covered the very diverse types of toxicity, such as gastrointestinal, cardiovascular, hepatic, and hematological toxicity. Moreover, it was reported that transcriptomic analysis would reflect a particular pattern of genes that could be associated with drug-induced toxicity, providing a more sensitive and specific panel of SAFs as well as insight in the mechanistic aspect of toxicity (55). Thus, various transcriptomic datasets were collected and statistically analyzed in this study. *First*, a comprehensive search in GEO was conducted based on such keywords as 'drug toxicity', 'adverse drug reaction', and 'side effect', which resulted in ∼150 datasets by limiting the dataset 'Organism' to *Homo sapiens* and the dataset 'Type' to *expression profiling by array* & *expression profiling by high throughput sequencing*. *Second*, these retrieved datasets were manually checked to guarantee that all those analyzed samples were from patients (such as disease tissue, peripheral blood, blood plasma, and urine) and all patients had been exposed to a therapy (both control and case sample groups), which resulted in three datasets: GSE186143 (60 melanoma patients treated with checkpoint inhibitor), GSE171468 (57 colorectal cancer patients treated with capecitabine), and GSE178708 (20 breast cancer patients treated with a radiation therapy). *Third*, original CEL files for microarray data and raw read counts for RNA-seq data were processed using *oligo* (56) and *DESeq2* (57), respectively. All those genes with fold change >1.5 and adjusted P-value <0.05 between control and case groups were collected. As described in Supplementary Table S1, these three datasets were provided at the bottom, which described two groups of patients administrated with the same therapy (one with observed toxic event, while the other with *non-toxic* event). As a result, 624 SAFs indicating the extent/likelihood of therapy-induced toxicity were collected.

Monitoring biomarker for optimizing the clinical management of a therapy

The *monitoring biomarker* (MOI) was a group of indicators serially measured for assessing status of disease or medical condition or for evidence of exposure to (or effect of) a therapy (58), which was very essential for guiding the clinical managements of the corresponding therapy (6). It was measured during one or more periods of patient's clinical course, such as following the diagnosis of disease and prior to the intervention, during the period in which the therapy is being delivered, and after the delivery of a therapy has been completed (3). Taking the mRNA level of BCR-ABL as an example, it was serially monitored to identify the patients of *chronic myelogenous leukemia* being treated with nilotinib who may be the candidates for clinical treatment discontinuation (59). The monitor of a studied MOI included the evaluations of its magnitude, its magnitude of change, its rate of change over time, its relation of changes to the patients, and so on (10).

Herein, MOI data were systematically collected (as illustrated in Figure 4). General information of a MOI was provided, such as MOI name, MOI class, gene name, synonyms, function, external links, and so on. The drugs to which the MOI was applied were grouped by drugs' clinical status (such as approved and in clinical trial). For each drug, the studied disease, tested species, testing method, biomarker source, and disease ICD-11 were extensively provided. As a

Table 1. Five types of *therapeutic biomarker* (ThMAR) defined by the official 'BEST' category of the U.S. Food & Drug Administration (3), which deeply involved in various stages of drug discovery

ThMAR type	Definition and importance of the corresponding ThMAR type	Typical example
Pharmacodynamic biomarker (PDY)	A group of indicators of drug effect on its target in a studied organism (90), which is essential for guaranteeing pharmacological effects, establishing proof-of-concept, assisting dose selection, and measuring the response to specific therapy (4).	DKK3 is a PDY indicating the inhibition of HTRA1 in patients using anti-HTRA1 antibody (91)
Safety biomarker (SAF)	A group of indicators denoting the likelihood, presence, or extent of therapy-induced toxicity as adverse drug reaction (92), which identifies patients for whom particular therapies should not be initiated or continued due to significant safety risk (5).	Urinary KIM1 and NGAL are two typical SAFs for detecting acute drug-induced nephrotoxicity (93)
Monitoring biomarker (MOI)	A group of indicators serially measured for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a studied therapy (58), which is indispensable for guiding the clinical management of this medication (6).	HCV-RNA is used as a MOI for measuring & guiding the usage of antiretroviral therapies (94)
Predictive biomarker (PRD)	A group of indicators identifying the individuals who are more likely to experience a favorable/unfavorable effect from the exposure to a therapy (60), which is essential for maximizing the clinical outcome for particular group of individuals (7).	PD-L1 is an extensively studied PDY predicting the response to immune checkpoint inhibitor (95)
Surrogate endpoint (SUR)	A group of indicators used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives (96), which predicts the clinical benefit or harm based on epidemiologic, therapeutic, and pathophysiologic evidences (8).	The reduction of hemoglobin A1C is a SUR facilitating the drug approval for diabetes mellitus (97)

There were five ThMAR types: pharmacodynamic biomarker (PDY), safety biomarker (SAF), monitoring biomarker (MOI), predictive biomarker (PRD) and surrogate endpoint (SUR). The definition and importance of each ThMAR type were explicitly described, and the typical example was also provided for each ThMAR type. DKK3: Dickkopf-related protein 3; HTRA1: high-temperature requirement A serine peptidase 1; KIM1: kidney injury molecule 1; NGAL: neutrophil gelatinase associated lipocalin; HCV: hepatitis C virus.

result, a total of 104 MOIs guiding the clinical management of 60 drugs for the treatment of 33 classes of disease (such as obesity, hemophilia and melanoma) were collected and provided in TheMarker.

Predictive biomarker for maximizing the clinical outcomes of studied patients

The *predictive biomarker* (PRD) promoted the identification of individuals who were more likely to experience favorable/unfavorable effect from the exposure to studied therapy (60), which was essential for maximizing the clinical outcome for particular group of individuals (7). Particularly, treatment responses of a drug varied considerably among individuals, resulting in only a fraction of patients benefiting from the studied therapy (61–65). As reported, >60% of depression patients failed to recover after drug treatment, and 20% of them even did not respond to any intervention (66). One of the effective ways to address this issue was the discovery of the reliable and sensitive PRDs that could optimize the clinical outcome for a particular group of individuals (67–70). Such PRDs could eventually facilitate a precision medication (71–74). In addition, PRDs would greatly contribute to drug/target discovery (66). In other words, the collected PRDs were highly expected to revolutionize the ways of both drug administration and drug development (75–79).

Herein, the experimentally/clinically identified PRDs for a great number of drugs were collected. An exemplar PRD for a therapy was shown in Figure 5. General drug information was provided, which included synonyms, indication, structure, drug properties, external links, and so on. Those literature-reported PRDs were categorized based on their molecular classes (such as microRNA and protein biomarker), and the application of each PRD in the corresponding drug was explicitly demonstrated, including reported biomarker variations, disease, experimental species, biomarker source, experimental testing approach, and so on. Moreover, the accumulation of transcriptomic data investigating the differences in treatment responses also provided a valuable opportunity to discover PRD and unravel the molecular mechanism underlying drug response/resistance (80,81). Therefore, such valuable transcriptomic data were also incorporated to the database. Particularly, comprehensive search in GEO (82) and Expression Atlas (83) was *first* conducted based on such keywords/combinations as 'drug response', 'drug resistance', and 'treatment response'; *second*, all eligible datasets were statistically analyzed in the same way as that for *safety biomarker* (SAF). As illustrated in Supplementary Table S1, a total of 93 transcriptomics datasets were collected, analyzed and provided, which were fully covered and described by current version of TheMarker. As illustrated in Figure 5c, detailed information of each dataset (such as the studied drug, disease, and sample) was given and the gene expression between control and case was presented in scatter plots. Up- and down-regulated PRDs were colored in red and blue respectively, and fold changes and adjusted *P*-values were also provided. As a result, a total of 15 893 PRDs that facilitated the identification of individuals who were more likely to experience a favorable or unfavorable effect from 352 drugs for treating 95 diseases were systematically provided in TheMarker.

Surrogate endpoint for substantially accelerating the approval of a therapy

In an effort to expedite approval of drug for treating the disease of unmet medical need, FDA has long been positive to the proper applications of *surrogate endpoint* (SUR) in drug discovery (8). SUR referred to the biomarker that was used in clinical trials for predicting clinical benefit/harm, rather than to directly measure clinical outcome (whether the patient feels or functions better, or lives longer) (84). As assessing clinical outcome (such as all-cause mortality) often required large sample size and long follow-up time, the use of SUR was an ideal way to elevate the efficacy of clinical trials with smaller

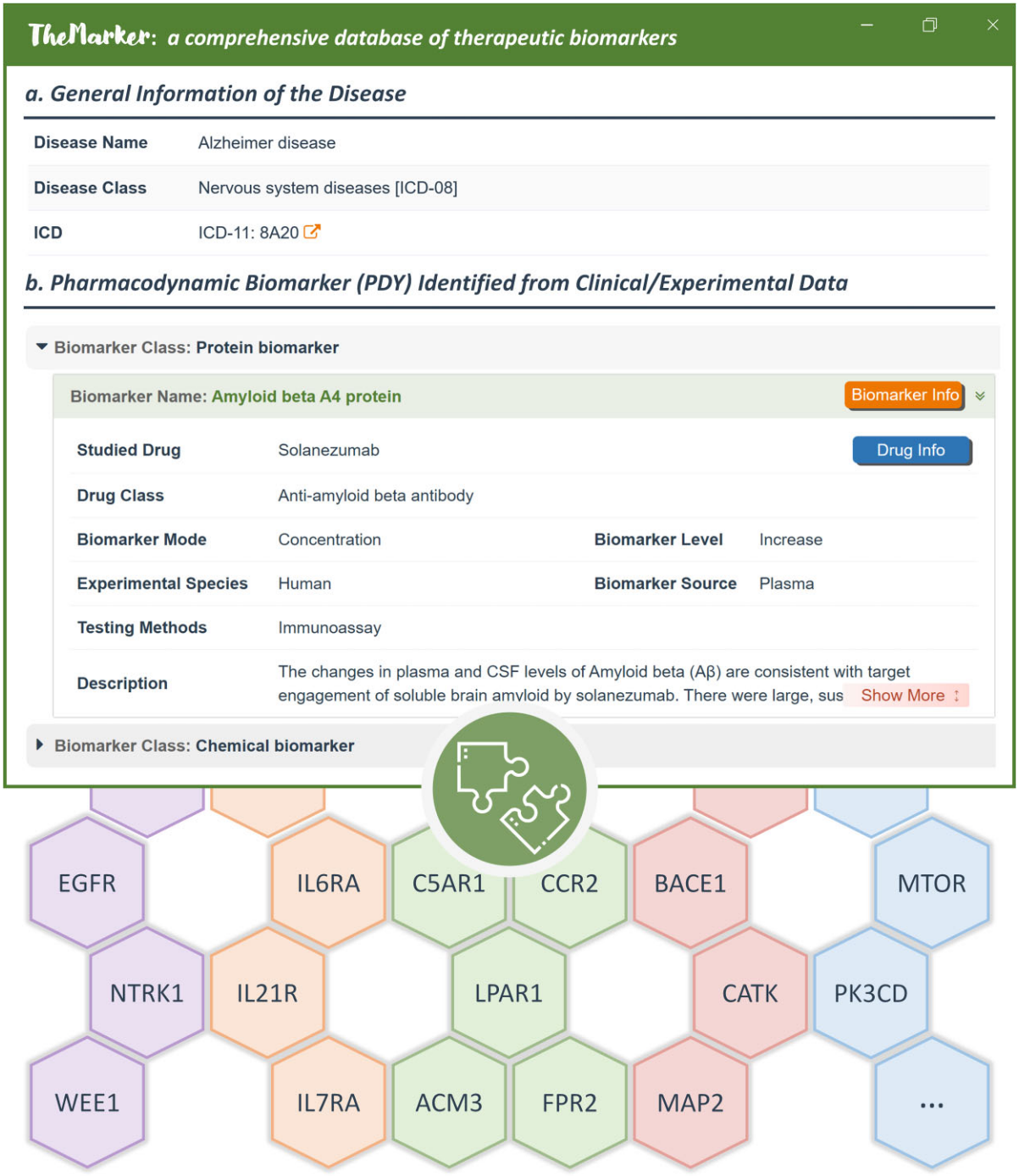


Figure 2. A typical disease page showing the application of *pharmacodynamic biomarker* (PDY) for diseases and the representative engagement targets of the PDY provided in TheMarker. In the upper section, the typical page of PDYs applied to a disease in TheMarker was described. (**a**) the general disease information such as disease name, disease class, and ICD-11. (**b**) the PDYs used in the disease categorized by the biomarker class (such as protein & compound). For a PDY, drug class, biomarker mode/level, experimental species, and testing method were provided.

patient number and shorter trial duration (85). Depending on the level of clinical validation and whether there was enough evidence to support the prediction of specific clinical benefit, SURs were grouped to '*validated SUR*', '*reasonably likely SUR*' and '*candidate SUR*' (86). Particularly, *validated SUR* indicated the biomarker that had clear causal/mechanistic rationale of the disease processes and had obtained clinical data relevant to clinical benefit, while *reasonably likely SUR* was supported by strong mechanistic data and/or epidemiologic rationale, but existing clinical data were still insufficient to demonstrate their capacity in predicting clinical benefit. Such systematical collection of SURs would be able to promote the identification of new SURs for drug developers when designing their drug development programs (84).

In this study, reported SURs were extensively collected and systematically presented. As shown in Figure 6, the general information of SURs was provided in the upper section, which included name, biomarker class, biomarker status (such as

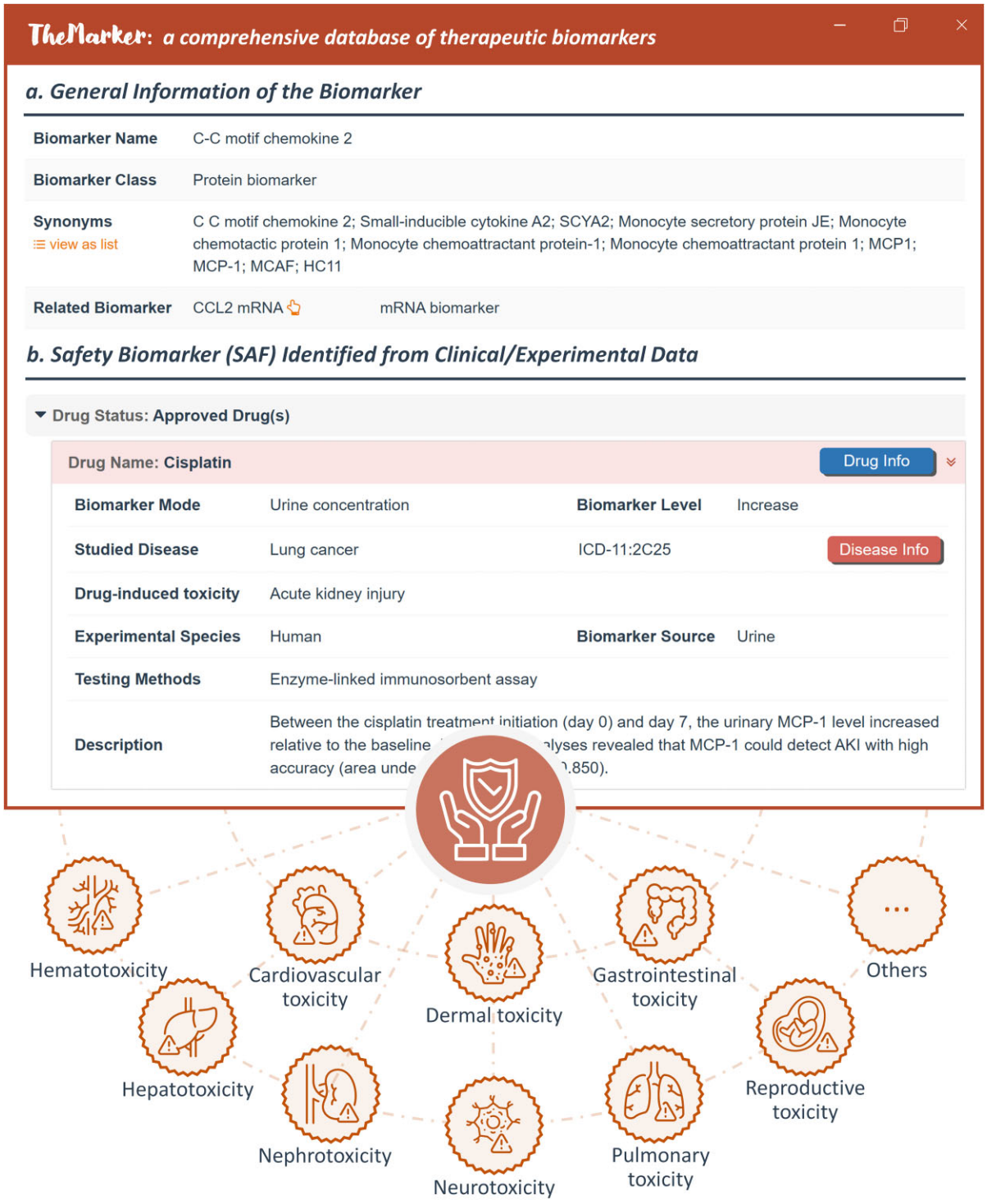


Figure 3. A typical biomarker page describing *safety biomarker* (SAF) and representative types of toxicity offered by TheMarker. In the upper section, SAF's applications to predict/monitor the drug safety were shown. (**a**) *general information*. The basic data of the SAF included: biomarker name, class, synonyms, and so on. (**b**) *SAF identified from clinical/experimental data*. The drugs whose safety could be predicted/monitored by this SAF were categorized using drug status. For any drug, biomarker mode & level, induced toxicity, disease indication, tested species, biomarker source, and testing method were explicitly described. User can find detailed data of the drug and disease by clicking '*Drug Info*' and '*Disease Info*'. In the bottom section, the representative types of toxicity indicated by SAF were listed, such as cardiovascular toxicity and neurotoxicity.

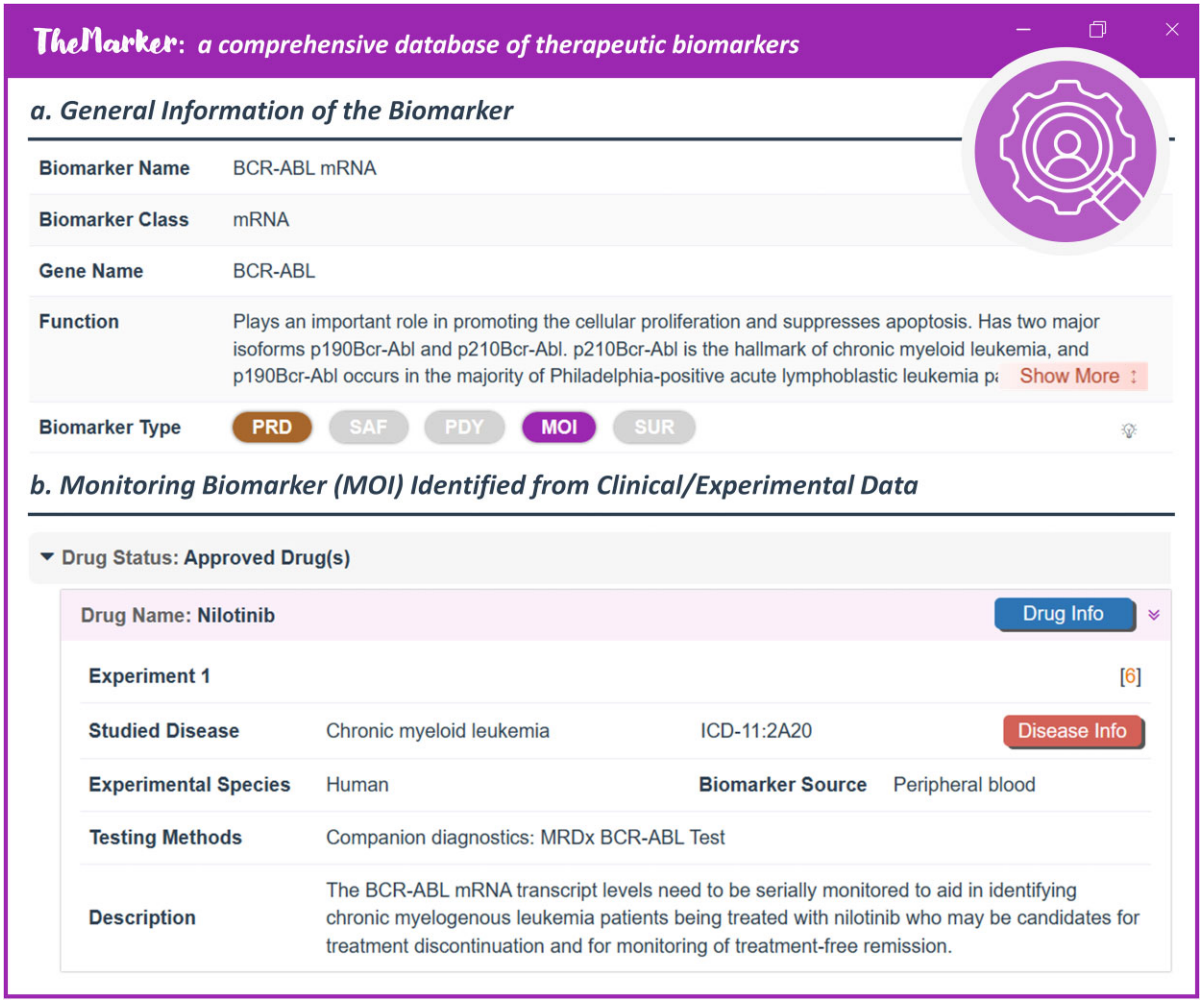


Figure 4. A typical ThMAR page showing *monitoring biomarker* (MOI). (**a**) *general information of ThMAR*. Such general information included MOI name, MOI class & MOI function. (**b**) *MOI identified from clinical/experimental data*. The drugs were categorized using drug status. For any drug, disease indication, tested species, biomarker source, and experimental testing method were explicitly described. Users can find detailed data of the drug and disease by clicking '*Drug Info*' and '*Disease Info*', respectively.

validated and *reasonably likely SUR*), approval types (*accelerated* and *traditional*), disease indications, and the corresponding patient population. In the lower section of Figure 6, the drugs approved based on SURs were shown and categorized using disease class. For each drug, the status of approval program (such as verified, ongoing and withdrawn) and the appropriate patient age were also described. All in all, a total of 103 *surrogate endpoints* that indicated the clinical outcomes of 435 approved drugs (including 193 accelerated approvals) for the treatment of 102 disease classes were finally provided in TheMarker.

Tissue- and disease-specific expression of the therapeutic biomarkers

Tissue-specific expressions of the biomarkers in TheMarker were collected as follows. *First*, the expression data for 182 transcriptomic and 201 proteomic samples across 32 major human tissues were collected (87). *Second*, RNA expression data were processed with *logarithm* transformation at base 2 and the protein abundances were normalized using *Z*-scores. *Third*, for each therapeutic biomarker, both RNA expression and protein abundance (if available) were given in the form of boxplot (shown in Figure 7a). Detailed procedure for collecting the disease-specific expressions of the biomarkers was described as follows: Affymetrix HG-U133 plus 2.0 microarrays datasets were *first* retrieved from GEO (82), and studied disease and tissue of each dataset were manually annotated, which led to a total of 612 datasets covering 97 disease classes and 59 tissues; *second*, the samples of the same disease from the same tissue were combined and the raw expression data were processed using RMA function of *affy* package (88); *third*, the median expression intensity array was selected as the baseline and all arrays were normalized (89); *finally*, fold changes and *t*-test were used to identify the differential expression of each biomarker among different groups, such as disease cases, healthy controls, and adjacent tissue (if available). As shown in Figure 7, the gene expression for each biomarker in a disease was provided in the format of violin plot. In addition, detailed information such as disease indication, tissue, significance test (*P*-value & fold change) was also provided in the website of TheMarker. As a result, a total of 18 890 therapeutic biomarkers in TheMarker were provided with tissue- and disease-specific expression data.

Figure 5. A typical drug page describing the *predictive biomarker* (PRD) of a drug in TheMarker. (**a**) *general information of drug*. Such information included: drug name, disease indication, drug structure, and drug-like property. (**b**) *PRD identified from clinical/experimental data*. Literature-reported PRDs for studied drugs were categorized based on biomarker class (such as microRNA and protein), and the applications of each PRD were explicitly provided (such as biomarker mode, disease, tested species, biomarker source, and experimental testing approach). User can click the button of '*Biomarker Info*' & '*Disease Info*' to retrieve detailed information on the corresponding biomarker & disease. (**c**) *PRD discovered from transcriptomic data*. Detailed information of each transcriptomic dataset (such as experimental drug, disease and sample) was given and the genes with fold change >1.5 & adjusted *P*-value <0.05 between controls and cases were considered as potential PRDs. The up- and down-regulated PRDs were colored in red and blue, respectively.

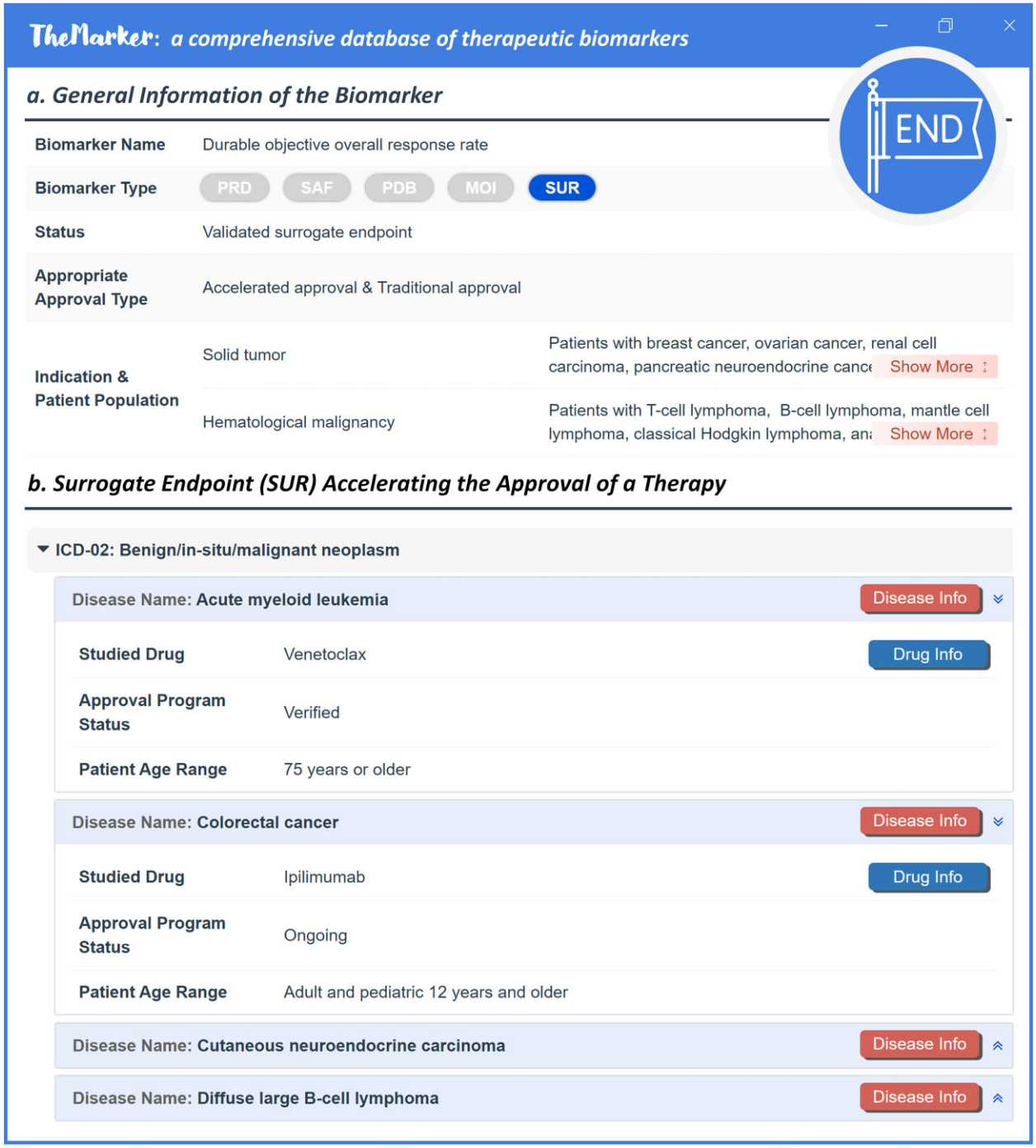


Figure 6. A typical ThMAR page describing *surrogate endpoint* (SUR). (**a**) *general information of ThMAR*. Such general information included the SUR status (such as validated and reasonably likely surrogate endpoint), approval type (such as accelerated approval and traditional approval), disease indication and the corresponding patient population. (**b**) *SUR accelerating drug approval*. Drugs approved based on this SUR were also provided, which were categorized based on disease classes. For each drug, the status of approval programs (such as verified, ongoing and withdrawn) and the patient age range were described. The detailed information of the drug and disease could be retrieved by clicking the buttons of '*Drug Info*' and '*Disease Info*', respectively.

Data standardization, access, retrieval and visualization

To make the access of TheMarker data convenient for users, the collected raw data were cleaned up and then standardized, which included the standardizations of disease indication, external IDs, 2D & 3D structures, drug-like properties, and so on. Moreover, two additional functions enabling data visualization were provided, which included the *filter function* for searching results and the *browse function* for entire ThMARs. As shown in Supplementary Figure S1, filter function for searching results were described, and the readers could try out this new function by accessing an exemplar weblink (http://themarker.idrblab.cn/search-drug?api=fullText&keyword=EGFR). As illustrated in Supplementary Figure S2, the browse functions for entire ThMAR

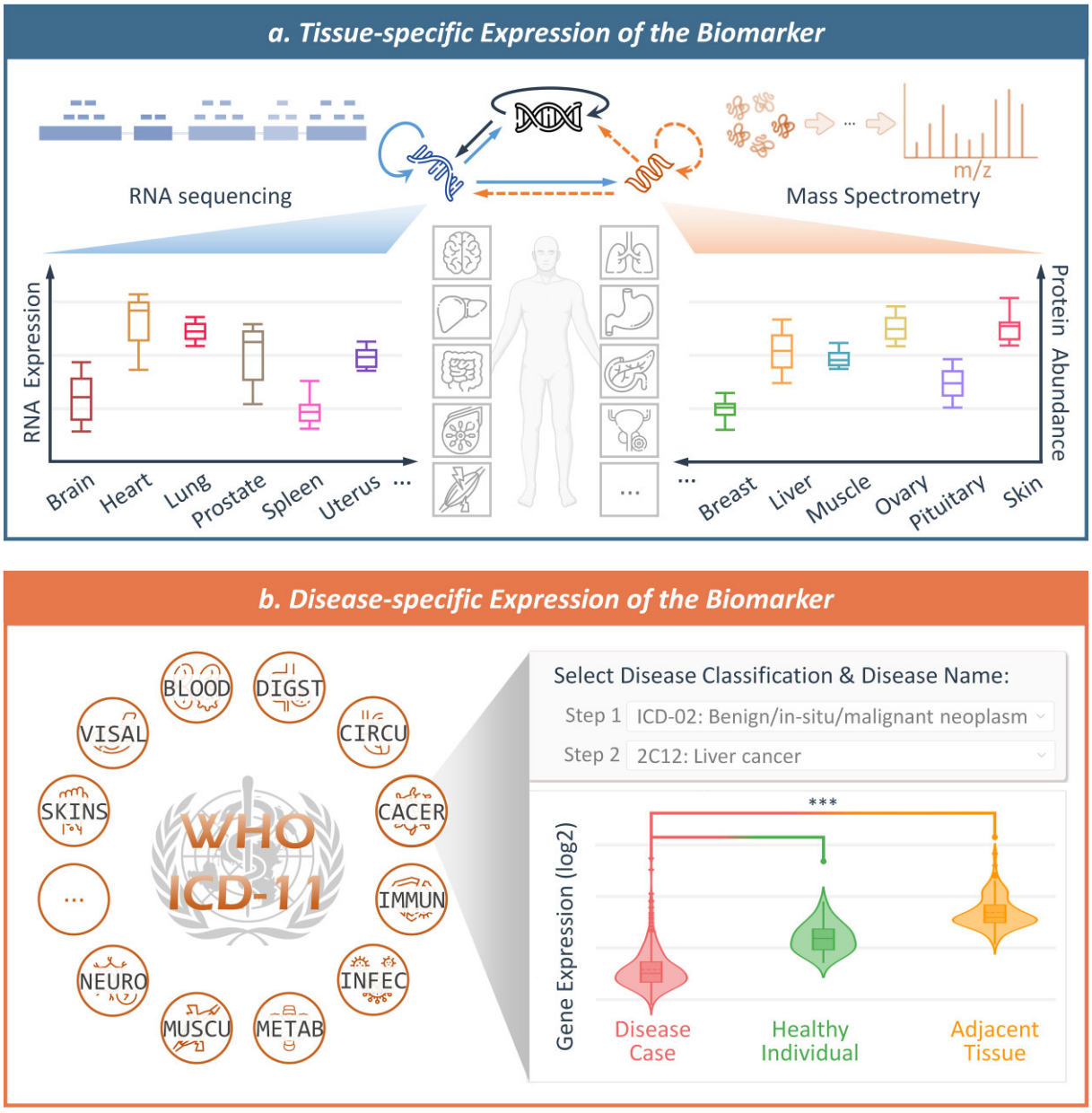


Figure 7. Tissue- and disease-specific expression information of *therapeutic biomarkers* covered in TheMarker. (**a**) *Tissue-specific expression of the biomarker*. Both RNA expression and protein abundance across 32 human tissues/organs were shown in the format of 'boxplot', which covered those major organs of human body such as brain, heart, lung, liver and so on. (**b**) *Disease-specific expression of the biomarker*. A total of 97 disease classes (defined by ICD-11) were covered by TheMarker, and the expression patterns of a biomarker among disease cases, healthy individuals, and adjacent tissues were described using 'violin plot' for each disease (fold change and *P*-value were explicitly provided). BLOOD: disease of blood/blood-forming organs; CACER: neoplasm; CIRCU: disease of circulatory systems; DIGEST: disease of digestive systems; IMMUN: disease of immune systems; INFEC: infectious/parasitic disease; METAB: endocrine/metabolic disease; MUSCU: musculoskeletal/connective-tissue disease; NEURO: nervous system disease; SKINS: skin disease; VISAL: visual system disease.

were provided, and the readers could try out this function by visiting: http://themarker.idrblab.cn/browse.

The layouts of TheMarker were organized by presenting ThMAR data in a tabular format, and a filter function was also provided on both the main page (*biomarker*, *drug* and *disease* pages) and the searching results page. Taking the ThMAR page (Supplementary Figure S3) as an example, the readers can view the data by *table*, and a try out page was also provided for the user to access (http://themarker.idrblab.cn/data/marker?id=B3HTD6). Furthermore, systematical review on the databases offering diagnostic/prognostic biomarkers was conducted, which led to > 10 databases, such as MarkerDB (11), Lnc2Cancer (12) and CRMarker (18). The correspondence between the ThMARs and diagnostic/prognostic ones were *then* established, which were *finally* provided at the bottom of the ThMAR page (Supplementary Figure S4).

Conclusion and perspectives

Therapeutic biomarker (ThMAR), distinct from traditional diagnostic and prognostic biomarker, has emerged to be critical in drug development and clinical practice of

therapies. In this study, a database titled 'TheMarker' was constructed to provide the comprehensive data on ThMARs. It is unique in systematically providing five types of ThMAR to realize the collective consideration among various stages of drug discovery, comprehensively describing ThMAR data for the largest number of drugs among existing databases, and extensively covering the widest range of diseases. However, due to the complexity of the ThMAR data in newly published papers, it was unrealistic to automatically update our database. Therefore, we would like to update it in an annual/biennial manner. All in all, the data provided in TheMarker are highly expected to have great implications and significant impacts on both drug development and clinical practice.

Data availability

TheMarker is freely accessible without any login requirement at: https://idrblab.org/themarker.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Natural Science Foundation of China [82373790, 22220102001, U1909208, 81872798, 81971982]; Natural Science Foundation of Zhejiang Province [LR21H300001]; National Key R&D Program of China [2022YFC3400501]; Leading Talent of the 'Ten Thousand Plan' National High-Level Talents Special Support Plan of China; The Double Top-Class Universities [181201*194232101]; Fundamental Research Funds for Central Universities [2018QNA7023]; Key R&D Program of Zhejiang Province [2020C03010]; Westlake Laboratory (Westlake Laboratory of Life Science & Biomedicine); Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; The Information Technology Center of Zhejiang University. Funding for open access charge: Natural Science Foundation of Zhejiang Province [LR21H300001].

Conflict of interest statement

None declared.

References

1. Mollenhauer,B. and von Arnim,C.A.F. (2022) Toward preventing Parkinson's disease. *Science*, **377**, 818–819.
2. Meric-Bernstam,F., Larkin,J., Tabernero,J. and Bonini,C. (2021) Enhancing anti-tumour efficacy with immunotherapy combinations. *Lancet*, **397**, 1010–1022.
3. Cagney,D.N., Sul,J., Huang,R.Y., Ligon,K.L., Wen,P.Y. and Alexander,B.M. (2018) The FDA NIH biomarkers, endpoints, and other tools (BEST) resource in neuro-oncology. *Neuro. Oncol.*, **20**, 1162–1172.
4. Emmerich,C.H., Gamboa,L.M., Hofmann,M.C.J., Bonin-Andresen,M., Arbach,O., Schendel,P., Gerlach,B., Hempel,K., Bespalov,A., Dirnagl,U., *et al.* (2021) Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat. Rev. Drug Discov.*, **20**, 64–81.
5. Pognan,F., Beilmann,M., Boonen,H.C.M., Czich,A., Dear,G., Hewitt,P., Mow,T., Oinonen,T., Roth,A., Steger-Hartmann,T., *et al.* (2023) The evolving role of investigative toxicology in the pharmaceutical industry. *Nat. Rev. Drug Discov.*, **22**, 317–335.
6. Argollo,M., Kotze,P.G., Kakkadasam,P. and D'Haens,G. (2020) Optimizing biologic therapy in IBD: how essential is therapeutic drug monitoring?*Nat. Rev. Gastroenterol. Hepatol.*, **17**, 702–710.
7. Pilie,P.G., George,A. and Yap,T.A. (2020) Patient selection biomarker strategies for PARP inhibitor therapy. *Ann. Oncol.*, **31**, 1603–1605.
8. Ciani,O., Manyara,A.M. and Taylor,R.S. (2022) Surrogate endpoints in trials-a call for better reporting. *BMJ*, **378**, o1912.
9. Brown,M.A., Li,Z. and Cao,K.L. (2020) Biomarker development for axial spondyloarthritis. *Nat Rev Rheumatol*, **16**, 448–463.
10. Davis,K.D., Aghaeepour,N., Ahn,A.H., Angst,M.S., Borsook,D., Brenton,A., Burczynski,M.E., Crean,C., Edwards,R., Gaudilliere,B., *et al.* (2020) Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nat. Rev. Neurol.*, **16**, 381–400.
11. Wishart,D.S., Bartok,B., Oler,E., Liang,K.Y.H., Budinski,Z., Berjanskii,M., Guo,A., Cao,X. and Wilson,M. (2021) MarkerDB: an online database of molecular biomarkers. *Nucleic Acids Res.*, **49**, D1259–D1267.
12. Gao,Y., Shang,S., Guo,S., Li,X., Zhou,H., Liu,H., Sun,Y., Wang,J., Wang,P., Zhi,H., *et al.* (2021) Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic Acids Res.*, **49**, D1251–D1258.
13. Neveu,V., Nicolas,G., Salek,R.M., Wishart,D.S. and Scalbert,A. (2020) Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Res.*, **48**, D908–D912.
14. Dingerdissen,H.M., Torcivia-Rodriguez,J., Hu,Y., Chang,T.C., Mazumder,R. and Kahsay,R. (2018) BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. *Nucleic Acids Res.*, **46**, D1128–D1136.
15. Cheng,X., Liu,Y., Wang,J., Chen,Y., Robertson,A.G., Zhang,X., Jones,S.J.M. and Taubert,S. (2022) cSurvival: a web resource for biomarker interactions in cancer outcomes and in cell lines. *Brief Bioinform*, **23**, bbac090.
16. Wang,X., Chai,Z., Pan,G., Hao,Y., Li,B., Ye,T., Li,Y., Long,F., Xia,L. and Liu,M. (2021) ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer. *Brief Bioinform*, **22**, bbaa088.
17. Yerlikaya,S., Broger,T., MacLean,E., Pai,M. and Denkinger,C.M. (2017) A tuberculosis biomarker database: the key to novel TB diagnostics. *Int. J. Infect. Dis.*, **56**, 253–257.
18. Zhang,J., Yan,S., Li,R., Wang,G., Kang,S., Wang,Y., Hou,W., Wang,C. and Tian,W. (2021) CRMarker: a manually curated comprehensive resource of cancer RNA markers. *Int. J. Biol. Macromol.*, **174**, 263–269.
19. He,H., Shi,M., Lin,Y., Zhan,C., Wu,R., Bi,C., Liu,X., Ren,S. and Shen,B. (2021) HFBD: a biomarker knowledge database for heart failure heterogeneity and personalized applications. *Bioinformatics*, **37**, 4534–4539.
20. Wu,W.S., Yang,T.H., Chen,K.D., Lin,P.H., Chen,G.R. and Kuo,H.C. (2022) KDmarkers: a biomarker database for investigating epigenetic methylation and gene expression levels in Kawasaki disease. *Comput. Struct. Biotechnol. J.*, **20**, 1295–1305.
21. Ran,Z., Yang,J., Liu,Y., Chen,X., Ma,Z., Wu,S., Huang,Y., Song,Y., Gu,Y., Zhao,S., *et al.* (2022) GlioMarker: an integrated database for knowledge exploration of diagnostic biomarkers in gliomas. *Front. Oncol.*, **12**, 792055.
22. Liu,X., Zhang,X., Chen,J., Ye,B., Ren,S., Lin,Y., Sun,X.F., Zhang,H. and Shen,B. (2020) CRC-EBD: epigenetic biomarker database for colorectal cancer. *Front. Genet.*, **11**, 907.
23. Yip,S.H., Fujiwara,N., Burke,J., Shetler,A., Peralta,C., Qian,T., Hoshida,H., Zhu,S. and Hoshida,Y. (2019) MPIC: molecular prognostic indicators in cirrhosis database for clinical context-specific in silico prognostic biomarker validation. *Front. Genet.*, **10**, 830.

24. Arora,A., Kaur,D., Patiyal,S., Kaur,D., Tomer,R. and Raghava,G.P.S. (2023) SalivaDB-a comprehensive database for salivary biomarkers in humans. *Database*, **2023**, baad002.
25. Tan,B., Xin,S., Hu,Y., Feng,C. and Chen,M. (2022) LBD: a manually curated database of experimentally validated lymphoma biomarkers. *Database*, **2022**, baac051.
26. Chen,J., Liu,X., Shen,L., Lin,Y. and Shen,B. (2021) CMBD: a manually curated cancer metabolic biomarker knowledge database. *Database*, **2021**, baaa094.
27. Domingo-Fernandez,D., Provost,A., Kodamullil,A.T., Marin-Llao,J., Lasseter,H., Diaz,K., Daskalakis,N.P., Lancashire,L., Hofmann-Apitius,M. and Haas,M. (2019) PTSD biomarker database: deep dive metadatabase for PTSD biomarkers, visualizations and analysis tools. *Database*, **2019**, baz081.
28. Li,F., Yin,J., Lu,M., Mou,M., Li,Z., Zeng,Z., Tan,Y., Wang,S., Chu,X., Dai,H., *et al.* (2023) DrugMAP: molecular atlas and pharma-information of all drugs. *Nucleic Acids Res.*, **51**, D1288–D1299.
29. Fu,T., Li,F., Zhang,Y., Yin,J., Qiu,W., Li,X., Liu,X., Xin,W., Wang,C., Yu,L., *et al.* (2022) VARIDT 2.0: structural variability of drug transporter. *Nucleic Acids Res.*, **50**, D1417–D1431.
30. Amahong,K., Zhang,W., Zhou,Y., Zhang,S., Yin,J., Li,F., Xu,H., Yan,T., Yue,Z., Liu,Y., *et al.* (2023) CovInter: interaction data between coronavirus RNAs and host proteins. *Nucleic Acids Res.*, **51**, D546–D556.
31. Sun,X., Zhang,Y., Li,H., Zhou,Y., Shi,S., Chen,Z., He,X., Zhang,H., Li,F., Yin,J., *et al.* (2023) DRESIS: the first comprehensive landscape of drug resistance information. *Nucleic Acids Res.*, **51**, D1263–D1275.
32. Wang,X., Li,F., Qiu,W., Xu,B., Li,Y., Lian,X., Yu,H., Zhang,Z., Wang,J., Li,Z., *et al.* (2022) SYNBIP: synthetic binding proteins for research, diagnosis and therapy. *Nucleic Acids Res.*, **50**, D560–D570.
33. Yin,J., Li,F., Zhou,Y., Mou,M., Lu,Y., Chen,K., Xue,J., Luo,Y., Fu,J., He,X., *et al.* (2021) INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res.*, **49**, D1233–D1243.
34. Liu,Z., Liu,J., Liu,X., Wang,X., Xie,Q., Zhang,X., Kong,X., He,M., Yang,Y., Deng,X., *et al.* (2022) CTR-DB, an omnibus for patient-derived gene expression signatures correlated with cancer drug response. *Nucleic Acids Res.*, **50**, D1184–D1199.
35. Perez-Granado,J., Pinero,J. and Furlong,L.I. (2019) ResMarkerDB: a database of biomarkers of response to antibody therapy in breast and colorectal cancer. *Database*, **2019**, baz060.
36. Shen,W.X., Zeng,X., Zhu,F., Wang,Y.L., Qin,C., Tan,Y., Jiang,Y.Y. and Chen,Y.Z. (2021) Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations. *Nat. Mach. Intell.*, **3**, 334–343.
37. Xia,W., Zheng,L., Fang,J., Li,F., Zhou,Y., Zeng,Z., Zhang,B., Li,Z., Li,H. and Zhu,F. (2022) PFmulDL: a novel strategy enabling multi-class and multi-label protein function annotation by integrating diverse deep learning methods. *Comput. Biol. Med.*, **145**, 105465.
38. Fu,J., Zhang,Y., Wang,Y., Zhang,H., Liu,J., Tang,J., Yang,Q., Sun,H., Qiu,W., Ma,Y., *et al.* (2022) Optimization of metabolomic data processing using NOREVA. *Nat. Protoc.*, **17**, 129–151.
39. Li,F., Zhou,Y., Zhang,Y., Yin,J., Qiu,Y., Gao,J. and Zhu,F. (2022) POSREG: proteomic signature discovered by simultaneously optimizing its reproducibility and generalizability. *Brief. Bioinform.*, **23**, bbac040.
40. Stefaniak,J. and Huber,K.V.M. (2020) Importance of quantifying drug-target engagement in cells. *ACS Med. Chem. Lett.*, **11**, 403–406.
41. Yin,J., Sun,W., Li,F., Hong,J., Li,X., Zhou,Y., Lu,Y., Liu,M., Zhang,X., Chen,N., *et al.* (2020) VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res.*, **48**, D1042–D1050.
42. Wang,T., Sun,J. and Zhao,Q. (2023) Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism. *Comput. Biol. Med.*, **153**, 106464.
43. Shan,Y., Cheung,L., Zhou,Y., Huang,Y. and Huang,R.S. (2023) A systematic review on sex differences in adverse drug reactions related to psychotropic, cardiovascular, and analgesic medications. *Front. Pharmacol.*, **14**, 1096366.
44. Xue,W., Fu,T., Deng,S., Yang,F., Yang,J. and Zhu,F. (2022) Molecular mechanism for the allosteric inhibition of the human serotonin transporter by antidepressant escitalopram. *ACS Chem. Neurosci.*, **13**, 340–351.
45. Tosca,E.M., Bartolucci,R., Magni,P. and Poggesi,I. (2021) Modeling approaches for reducing safety-related attrition in drug discovery and development: a review on myelotoxicity, immunotoxicity, cardiovascular toxicity, and liver toxicity. *Expert Opin. Drug Discov.*, **16**, 1365–1390.
46. Ai,H., Chen,W., Zhang,L., Huang,L., Yin,Z., Hu,H., Zhao,Q., Zhao,J. and Liu,H. (2018) Predicting drug-induced liver injury using ensemble learning methods and molecular fingerprints. *Toxicol. Sci.*, **165**, 100–107.
47. Basile,A.O., Yahi,A. and Tatonetti,N.P. (2019) Artificial intelligence for drug toxicity and safety. *Trends Pharmacol. Sci.*, **40**, 624–635.
48. Ding,R.F., Yu,Q., Liu,K., Du,J., Yin,H.J. and Ji,Z.L. (2021) Gene network analyses unveil possible molecular basis underlying drug-induced glaucoma. *BMC Med Genomics*, **14**, 109.
49. Sullivan,R.J. and Weber,J.S. (2022) Immune-related toxicities of checkpoint inhibitors: mechanisms and mitigation strategies. *Nat. Rev. Drug Discov.*, **21**, 495–508.
50. Aulin,L.B.S., de Lange,D.W., Saleh,M.A.A., van der Graaf,P.H., Voller,S. and van Hasselt,J.G.C. (2021) Biomarker-guided individualization of antibiotic therapy. *Clin. Pharmacol. Ther.*, **110**, 346–360.
51. Hasan Ali,O., Bomze,D., Ring,S.S., Berner,F., Fassler,M., Diem,S., Abdou,M.T., Hammers,C., Emtenani,S., Braun,A., *et al.* (2020) BP180-specific IgG is associated with skin adverse events, therapy response, and overall survival in non-small cell lung cancer patients treated with checkpoint inhibitors. *J. Am. Acad. Dermatol.*, **82**, 854–861.
52. Groves,A.M., Williams,J.P., Hernady,E., Reed,C., Fenton,B., Love,T., Finkelstein,J.N. and Johnston,C.J. (2018) A potential biomarker for predicting the risk of radiation-induced fibrosis in the lung. *Radiat. Res.*, **190**, 513–525.
53. Yang,Q., Li,B., Chen,S., Tang,J., Li,Y., Li,Y., Zhang,S., Shi,C., Zhang,Y., Mou,M., *et al.* (2021) MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis. *J. Proteomics*, **232**, 104023.
54. Yang,Q., Wang,Y., Zhang,Y., Li,F., Xia,W., Zhou,Y., Qiu,Y., Li,H. and Zhu,F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.*, **48**, W436–W448.
55. Rodrigues,D., Souza,T., Jennen,D.G.J., Lemmens,L., Kleinjans,J.C.S. and de Kok,T.M. (2019) Drug-induced gene expression profile changes in relation to intestinal toxicity: state-of-the-art and new approaches. *Cancer Treat. Rev.*, **77**, 57–66.
56. Carvalho,B.S. and Irizarry,R.A. (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, **26**, 2363–2367.
57. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
58. Riviere-Cazaux,C., Lacey,J.M., Carlstrom,L.P., Laxen,W.J., Munoz-Casabella,A., Hoplin,M.D., Ikram,S., Zubair,A.B., Andersen,K.M., Warrington,A.E., *et al.* (2023) Cerebrospinal fluid 2-hydroxyglutarate as a monitoring biomarker for IDH-mutant gliomas. *Neurooncol. Adv.*, **5**, vdad061.
59. Jorgensen,J.T. (2021) The current landscape of the FDA approved companion diagnostics. *Transl Oncol*, **14**, 101063.
60. Wallington-Beddoe,C.T. and Mynott,R.L. (2021) Prognostic and predictive biomarker developments in multiple myeloma. *J. Hematol. Oncol.*, **14**, 151.
61. Sundstrom,J., Lind,L., Nowrouzi,S., Hagstrom,E., Held,C., Lytsy,P., Neal,B., Marttala,K. and Ostlund,O. (2023) Heterogeneity in

blood pressure response to 4 antihypertensive drugs: a randomized clinical trial. *JAMA*, **329**, 1160–1169.
62. Yurchenko,A.A., Pop,O.T., Ighilahriz,M., Padioleau,I., Rajabi,F., Sharpe,H.J., Poulalhon,N., Dreno,B., Khammari,A., Delord,M., *et al.* (2022) Frequency and genomic aspects of intrinsic resistance to vismodegib in locally advanced basal cell carcinoma. *Clin. Cancer Res.*, **28**, 1422–1432.
63. Gruener,R.F., Ling,A., Chang,Y.F., Morrison,G., Geeleher,P., Greene,G.L. and Huang,R.S. (2021) Facilitating drug discovery in breast cancer by virtually screening patients using in vitro drug response modeling. *Cancers*, **13**, 885.
64. Li,Y.H., Li,X.X., Hong,J.J., Wang,Y.X., Fu,J.B., Yang,H., Yu,C.Y., Li,F.C., Hu,J., Xue,W.W., *et al.* (2020) Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. *Brief. Bioinform*, **21**, 649–662.
65. Xue,W., Yang,F., Wang,P., Zheng,G., Chen,Y., Yao,X. and Zhu,F. (2018) What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.*, **9**, 1128–1140.
66. Mora,C., Zonca,V., Riva,M.A. and Cattaneo,A. (2018) Blood biomarkers and treatment response in major depression. *Expert Rev. Mol. Diagn.*, **18**, 513–529.
67. Hostallero,E., D.,W., L.,W., L.,C. and Emad,A. (2023) Preclinical-to-clinical anti-cancer drug response prediction and biomarker identification using TINDL. *Genom Proteom Bioinform*, https://doi.org/10.1016/j.gpb.2023.01.006.
68. Klumper,N., Saal,J., Berner,F., Lichtensteiger,C., Wyss,N., Heine,A., Bauernfeind,F.G., Ellinger,J., Brossart,P., Diem,S., *et al.* (2022) C reactive protein flare predicts response to checkpoint inhibitor treatment in non-small cell lung cancer. *J. Immunother. Cancer*, **10**, e004024.
69. Huang,E.W., Bhope,A., Lim,J., Sinha,S. and Emad,A. (2020) Tissue-guided LASSO for prediction of clinical drug response using preclinical samples. *PLoS Comput. Biol.*, **16**, e1007607.
70. Tang,J., Fu,J., Wang,Y., Li,B., Li,Y., Yang,Q., Cui,X., Hong,J., Li,X., Chen,Y., *et al.* (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief. Bioinform.*, **21**, 621–636.
71. Ashina,M., Terwindt,G.M., Al-Karagholi,M.A., de Boer,I., Lee,M.J., Hay,D.L., Schulte,L.H., Hadjikhani,N., Sinclair,A.J., Ashina,H., *et al.* (2021) Migraine: disease characterisation, biomarkers, and precision medicine. *Lancet*, **397**, 1496–1504.
72. Nath,A., Lau,E.Y.T., Lee,A.M., Geeleher,P., Cho,W.C.S. and Huang,R.S. (2019) Discovering long noncoding RNA predictors of anticancer drug sensitivity beyond protein-coding genes. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 22020–22029.
73. Emad,A., Cairns,J., Kalari,K.R., Wang,L. and Sinha,S. (2017) Knowledge-guided gene prioritization reveals new insights into the mechanisms of chemoresistance. *Genome Biol.*, **18**, 153.
74. Li,B., Tang,J., Yang,Q., Li,S., Cui,X., Li,Y., Chen,Y., Xue,W., Li,X. and Zhu,F. (2017) NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.*, **45**, W162–W170.
75. Hua,H., Zhang,H., Chen,J., Wang,J., Liu,J. and Jiang,Y. (2021) Targeting AKT in cancer for precision therapy. *J. Hematol. Oncol.*, **14**, 128.
76. Fassler,M., Diem,S., Mangana,J., Hasan Ali,O., Berner,F., Bomze,D., Ring,S., Niederer,R., Del,C.G., Cruz,C., *et al.* (2019) Antibodies as biomarker candidates for response and survival to checkpoint inhibitors in melanoma patients. *J. Immunother. Cancer*, **7**, 50.
77. Zhang,Y., Sun,H., Lian,X., Tang,J. and Zhu,F. (2023) ANPELA: significantly enhanced quantification tool for cytometry-based single-cell proteomics. *Adv. Sci. (Weinh.)*, **10**, e2207061.
78. Zhang,S., Amahong,K., Zhang,C., Li,F., Gao,J., Qiu,Y. and Zhu,F. (2021) RNA-RNA interactions between SARS-CoV-2 and host benefit viral development and evolution during COVID-19 infection. *Brief. Bioinform.*, **23**, bbab397.
79. Yang,Q., Li,B., Tang,J., Cui,X., Wang,Y., Li,X., Hu,J., Chen,Y., Xue,W., Lou,Y., *et al.* (2020) Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief. Bioinform.*, **21**, 1058–1068.
80. Bareche,Y., Kelly,D., Abbas-Aghababazadeh,F., Nakano,M., Esfahani,P.N., Tkachuk,D., Mohammad,H., Samstein,R., Lee,C.H., Morris,L.G.T., *et al.* (2022) Leveraging big data of immune checkpoint blockade response identifies novel potential targets. *Ann. Oncol.*, **33**, 1304–1317.
81. Zhang,W., Lee,A.M., Jena,S., Huang,Y., Ho,Y., Tietz,K.T., Miller,C.R., Su,M.C., Mentzer,J., Ling,A.L., *et al.* (2023) Computational drug discovery for castration-resistant prostate cancers through in vitro drug response modeling. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2218522120.
82. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M., *et al.* (2013) NCBI GEO: archive for functional genomics data sets update. *Nucleic Acids Res.*, **41**, D991–D995.
83. Papatheodorou,I., Moreno,P., Manning,J., Fuentes,A.M., George,N., Fexova,S., Fonseca,N.A., Fullgrabe,A., Green,M., Huang,N., *et al.* (2020) Expression atlas update: from tissues to single cells. *Nucleic Acids Res.*, **48**, D77–D83.
84. Agyeman,A.S., Siegel,J.N. and Leptak,C. (2022) Establishing a public resource for acceptable surrogate endpoints to support FDA marketing applications. *Front. Med.*, **9**, 820990.
85. Kumar,S. and Rajkumar,S.V. (2019) Surrogate endpoints in randomised controlled trials: a reality check. *Lancet*, **394**, 281–283.
86. Schuster Bruce,C., Brhlikova,P., Heath,J. and McGettigan,P. (2019) The use of validated and nonvalidated surrogate endpoints in two european medicines agency expedited approval pathways: a cross-sectional study of products authorised 2011-2018. *PLoS Med.*, **16**, e1002873.
87. Jiang,L., Wang,M., Lin,S., Jian,R., Li,X., Chan,J., Dong,G., Fang,H., Robinson,A.E., Consortium,G.T., *et al.* (2020) A quantitative proteome map of the human body. *Cell*, **183**, 269–283.
88. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) Affy-analysis of affymetrix genechip data at the probe level. *Bioinformatics*, **20**, 307–315.
89. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
90. Rabiner,E.A., Uz,T., Mansur,A., Brown,T., Chen,G., Wu,J., Atienza,J., Schwarz,A.J., Yin,W., Lewis,Y., *et al.* (2022) Endogenous dopamine release in the human brain as a pharmacodynamic biomarker: evaluation of the new GPR139 agonist TAK-041 with ((11)C)PHNO PET. *Neuropsychopharmacology*, **47**, 1405–1412.
91. Tom,I., Pham,V.C., Katschke,K.J., Li,W., Liang,W.C., Gutierrez,J., Ah Young,A., Figueroa,I., Eshghi,S.T., Lee,C.V., *et al.* (2020) Development of a therapeutic anti-HTRA1 antibody and the identification of DKK3 as a pharmacodynamic biomarker in geographic atrophy. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 9952–9963.
92. Schomaker,S., Ramaiah,S., Khan,N. and Burkhardt,J. (2019) Safety biomarker applications in drug development. *J. Toxicol. Sci.*, **44**, 225–235.
93. Soo,J.Y., Jansen,J., Masereeuw,R. and Little,M.H. (2018) Advances in predictive in vitro models of drug-induced nephrotoxicity. *Nat. Rev. Nephrol.*, **14**, 378–393.
94. Ennishi,D., Terui,Y., Yokoyama,M., Mishima,Y., Takahashi,S., Takeuchi,K., Okamoto,H., Tanimoto,M. and Hatake,K. (2008) Monitoring serum hepatitis C virus RNA in patients with HCV-infected CD20-positive B-cell lymphoma undergoing rituximab combination chemotherapy. *Am. J. Hematol.*, **83**, 59–62.

95. Shum,B., Larkin,J. and Turajlic,S. (2022) Predictive biomarkers for response to immune checkpoint inhibition. *Semin. Cancer Biol.*, **79**, 4–17.
96. Heerspink,H.J.L., Greene,T., Tighiouart,H., Gansevoort,R.T., Coresh,J., Simon,A.L., Chan,T.M., Hou,F.F., Lewis,J.B., Locatelli,F., *et al.* (2019) Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *Lancet. Diabetes Endocrinol.*, **7**, 128–139.
97. Lipska,K.J. and Krumholz,H.M. (2017) Is hemoglobin A1C the right outcome for studies of diabetes?*JAMA*, **317**, 1017–1018.

Received: July 19, 2023. **Revised:** September 21, 2023. **Editorial Decision:** September 25, 2023. **Accepted:** September 29, 2023