***Supplementary Information for*:**

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,*], 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] State Key Laboratory for Diagnosis and Treatment of Infectious Disease, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang Provincial Key Laboratory for Drug Clinical Research and Evaluation, The First Affiliated Hospital, Zhejiang University, Hangzhou, Zhejiang 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;
Prof. Feng Zhu, Email: zhufeng@zju.edu.cn
Correspondence may also be addressed to Prof. Yuzong Chen. Email: chenyuzong@sz.tsinghua.edu.cn
Correspondence may also be addressed to Prof. Yunqing Qiu. Email: qiuyq@zju.edu.cn

† These authors contributed equally to this work as co-first authors.

Supplementary Table S1. A total of 96 transcriptomics datasets that were collected from NIH-NCBI *Gene Expression Omnibus* (GEO) and EMBL-EBI *Expression Atlas* (GXA). The analyses of these transcriptomic data were reported to be critical for the investigation of the variations in treatment responses, which provided a valuable opportunity to discover *predictive biomarkers* (PRDs) and unravel the molecular mechanisms underlying drug response/resistance and capable of reflecting a particular pattern of genes that could be associated with the therapy-induced toxicities, providing a sensitive and specific panel of *safety biomarkers* (SAFs) as well as insights in the mechanistic aspect of toxicity. A list of acronyms was provided at the end of this table, which were applied here to describe the combinatorial therapies involved in those provided transcriptomic datasets.

Involved Therapy	Dataset ID	Description on the Case Group	Description on the Control Group
List of Transcriptomic Datasets Used to Facilitate the Identification of *Predictive Biomarker* (PRD)			
Anti-PD1/PD-L1	GSE135222	6 samples responding to the studied therapy	21 samples not responding to the studied therapy
Azacitidine	E-MTAB-8208	18 samples responding to the studied therapy	21 samples not responding to the studied therapy
Bevacizumab	GSE19860	5 samples responding to the studied therapy	7 samples not responding to the studied therapy
BMS-536924	GSE18912	5 samples sensitive to the studied therapy	5 samples resistant to the studied therapy
Carboplatin	GSE15622	8 samples sensitive to the studied therapy	6 samples resistant to the studied therapy
Carfilzomib	GSE69078	6 samples sensitive to the studied therapy	6 samples resistant to the studied therapy
CB-5083	E-MTAB-7328	6 samples sensitive to the studied therapy	6 samples resistant to the studied therapy
Cisplatin	GSE23554	18 samples responding to the studied therapy	10 samples not responding to the studied therapy
Cisplatin	GSE21656	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy

Cyclophosphamide	GSE7114	5 samples sensitive to the studied therapy	5 samples resistant to the studied therapy
Dasatinib	GSE59357	9 samples sensitive to the studied therapy	9 samples resistant to the studied therapy
Dasatinib	GSE9633	11 samples sensitive to the studied therapy	5 samples resistant to the studied therapy
Decitabine	GSE84334	18 samples responding to the studied therapy	20 samples not responding to the studied therapy
Docetaxel	GSE33455	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Epirubicin	GSE16446	16 samples responding to the studied therapy	98 samples not responding to the studied therapy
Etoposide	GSE28413	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Fluorouracil	GSE196900	4 samples sensitive to the studied therapy	4 samples resistant to the studied therapy
Fulvestrant	GSE48905	15 samples responding to the studied therapy	2 samples not responding to the studied therapy
Gefitinib	GSE10696	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Glucocorticoid	GSE5820	13 samples sensitive to the studied therapy	16 samples resistant to the studied therapy
Ibrutinib	GSE93984	13 samples responding to the studied therapy	47 samples not responding to the studied therapy
Imatinib	E-MEXP-433	30 samples responding to the studied therapy	15 samples not responding to the studied therapy
Imatinib	GSE14671	41 samples responding to the studied therapy	18 samples not responding to the studied therapy
Insulin	GSE79636	6 samples sensitive to the studied therapy	6 samples resistant to the studied therapy

Irradiation	GSE9713	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Lapatinib	GSE16179	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Lapatinib	GSE38376	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Letrozole	GSE20181	37 samples responding to the studied therapy	15 samples not responding to the studied therapy
Melphalan	GSE19293	24 samples responding to the studied therapy	28 samples not responding to the studied therapy
Melphalan	GSE34599	8 samples responding to the studied therapy	7 samples not responding to the studied therapy
Methotrexate	GSE16085	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE81259	29 samples responding to the studied therapy	18 samples not responding to the studied therapy
Methotrexate	GSE11440	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE16066	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE16070	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE16080	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE16082	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE16089	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Methotrexate	GSE9412	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy

Neoadjuvant therapy	GSE191127	14 samples sensitive to the studied therapy	22 samples resistant to the studied therapy
Neoadjuvant therapy	GSE192341	24 samples responding to the studied therapy	61 samples not responding to the studied therapy
Nivolumab	GSE91061	10 samples responding to the studied therapy	39 samples not responding to the studied therapy
Pembrolizumab	GSE145996	2 samples responding to the studied therapy	2 samples not responding to the studied therapy
Pembrolizumab	GSE78220	15 samples responding to the studied therapy	13 samples not responding to the studied therapy
Platinum	GSE51373	16 samples sensitive to the studied therapy	12 samples resistant to the studied therapy
PLX4720	GSE34299	2 samples sensitive to the studied therapy	2 samples resistant to the studied therapy
Paclitaxel	GSE15622	13 samples sensitive to the studied therapy	7 samples resistant to the studied therapy
Radiation	GSE9712	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
SN38	E-MEXP-1171	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
Tamoxifen	GSE26459	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
TRAIL	GSE55859	4 samples sensitive to the studied therapy	4 samples resistant to the studied therapy
Trastuzumab	GSE37946	27 samples responding to the studied therapy	23 samples not responding to the studied therapy
Trastuzumab	GSE44272	29 samples responding to the studied therapy	16 samples not responding to the studied therapy
Vincristine	GSE7556	2 samples sensitive to the studied therapy	2 samples resistant to the studied therapy

Vismodegib	GSE58375	4 samples sensitive to the studied therapy	9 samples resistant to the studied therapy
WZ4002	GSE37699	3 samples sensitive to the studied therapy	3 samples resistant to the studied therapy
2-fluoropyrimidine	GSE122246	12 samples responding to the studied therapy	12 samples not responding to the studied therapy
AraC+IDA	GSE103424	18 samples responding to the studied therapy	18 samples not responding to the studied therapy
ATC+TAX	GSE25066	99 samples responding to the studied therapy	272 samples not responding to the studied therapy
BEV+CIS	GSE103668	7 samples responding to the studied therapy	14 samples not responding to the studied therapy
BTZ+DXM+THA	GSE68871	14 samples with complete response reported	42 samples with partial response reported only
CAP+CYC+DTX *etc.*	GSE42822	25 samples responding to the studied therapy	41 samples not responding to the studied therapy
CAP+DTX	GSE18728	8 samples responding to the studied therapy	13 samples not responding to the studied therapy
CBP+PTX	GSE14764	59 samples responding to the studied therapy	21 samples not responding to the studied therapy
CHB+RTX	GSE35935	57 samples responding to the studied therapy	5 samples not responding to the studied therapy
CIS+DOX+IFF *etc.*	GSE87437	10 samples responding to the studied therapy	11 samples not responding to the studied therapy
CYC+CAP+FU *etc.*	GSE23988	20 samples responding to the studied therapy	41 samples not responding to the studied therapy
CYC+DOX+FU	GSE20271	7 samples responding to the studied therapy	75 samples not responding to the studied therapy
CYC+DOX+FU *etc.*	GSE50948	22 samples responding to the studied therapy	71 samples not responding to the studied therapy

CYC+DOX+VCR *etc.*	GSE57611	10 samples responding to the studied therapy	5 samples not responding to the studied therapy
CYC+EPI+FU	GSE22093	29 samples responding to the studied therapy	68 samples not responding to the studied therapy
CYC+EPI+FU *etc.*	GSE140494	63 samples responding to the studied therapy	18 samples not responding to the studied therapy
CYC+EPI+PTX *etc.*	GSE32646	27 samples responding to the studied therapy	88samples not responding to the studied therapy
DOX+6-MP *etc.*	GSE14615	30 samples responding to the studied therapy	6 samples not responding to the studied therapy
DOX+CYC+FU *etc.*	GSE20194	46 samples responding to the studied therapy	165 samples not responding to the studied therapy
FU+IRT+LV	GSE62080	9 samples sensitive to the studied therapy	12 samples resistant to the studied therapy
FU+LV+OXA	GSE19860	9 samples responding to the studied therapy	20 samples not responding to the studied therapy
FU+LV+OXA	GSE28702	42 samples responding to the studied therapy	41 samples not responding to the studied therapy
FU+LV+OXA	GSE69657	7 samples responding to the studied therapy	9 samples not responding to the studied therapy
IRT+FU+LV	GSE72970	27 samples responding to the studied therapy	33 samples not responding to the studied therapy
IXA+CYC+DOX	GSE41998	105 samples responding to the studied therapy	31 samples not responding to the studied therapy
LAP+CYC+EPI *etc.*	GSE66305	8 samples responding to the studied therapy	23 samples not responding to the studied therapy
LAP+TRA+CYC *etc.*	GSE66305	13 samples responding to the studied therapy	21 samples not responding to the studied therapy
NIV+PEM	GSE126044	5 samples responding to the studied therapy	11 samples not responding to the studied therapy

OXA+BEV+FU *etc.*	GSE72970	10 samples responding to the studied therapy	16 samples not responding to the studied therapy
OXA+FU+LV	GSE72970	20 samples responding to the studied therapy	12 samples not responding to the studied therapy
PTX+CYC+DOX	GSE41998	90 samples responding to the studied therapy	34 samples not responding to the studied therapy
PTX+CYC+FU *etc.*	GSE20271	17 samples responding to the studied therapy	62 samples not responding to the studied therapy
TRA+CAP+EPI *etc.*	GSE42822	12 samples responding to the studied therapy	13 samples not responding to the studied therapy
TRA+CYC+MTX	GSE50948	31 samples responding to the studied therapy	32 samples not responding to the studied therapy
TRA+CYC+FU	GSE66305	6 samples responding to the studied therapy	17 samples not responding to the studied therapy
TRA+DOX+CYC	GSE20194	2 samples responding to the studied therapy	4 samples not responding to the studied therapy
TRA+VNR	GSE70233	3 samples responding to the studied therapy	19 samples not responding to the studied therapy
List of Transcriptomic Datasets Used to Facilitate the Identification of *Safety Biomarker* (SAF)			
Capecitabine	GSE171468	39 samples with reported hand-foot syndrome	18 samples with no toxicity reported
Checkpoint inhibitors	GSE186143	43 samples with reported adverse events	17 samples with no toxicity reported
Irradiation	GSE178708	10 samples with reported radiotherapy toxicity	10 samples with no toxicity reported

List of Acronyms Shown in This Table

AraC: cytarabine; ATC: anthracycline

BEV: bevacizumab; BTZ: bortezomib

CAP: capecitabine; CBP: carboplatin; CHB: chlorambucil; CIS: cisplatin; CYC: cyclophosphamide

DOX: doxorubicin; DTX: docetaxel; DXM: examethasone

EPI: epirubicin

FU: 5-fluorouracil

IDA: idarubicin; IFF: ifosfamide; IRT: irinotecan; IXA: ixabepilone

LAP: lapatinib; LV: leucovorin

MTX: methotrexate

NIV: nivolumab

OXA: oxaliplatin

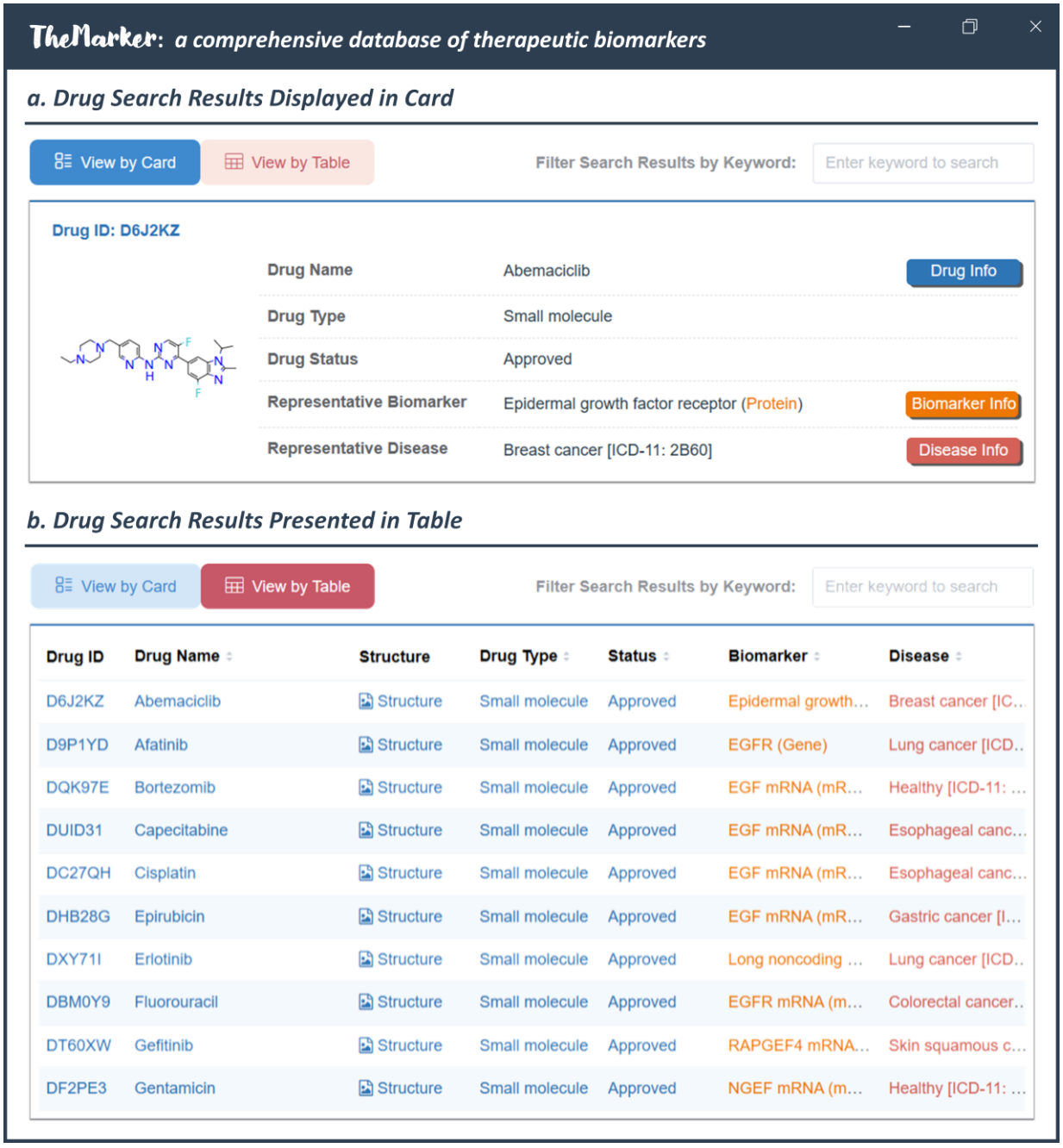
PDN: prednisone; PEM: pembrolizumab; PEX: pemetrexed; PTX: paclitaxel

RTX: rituximab

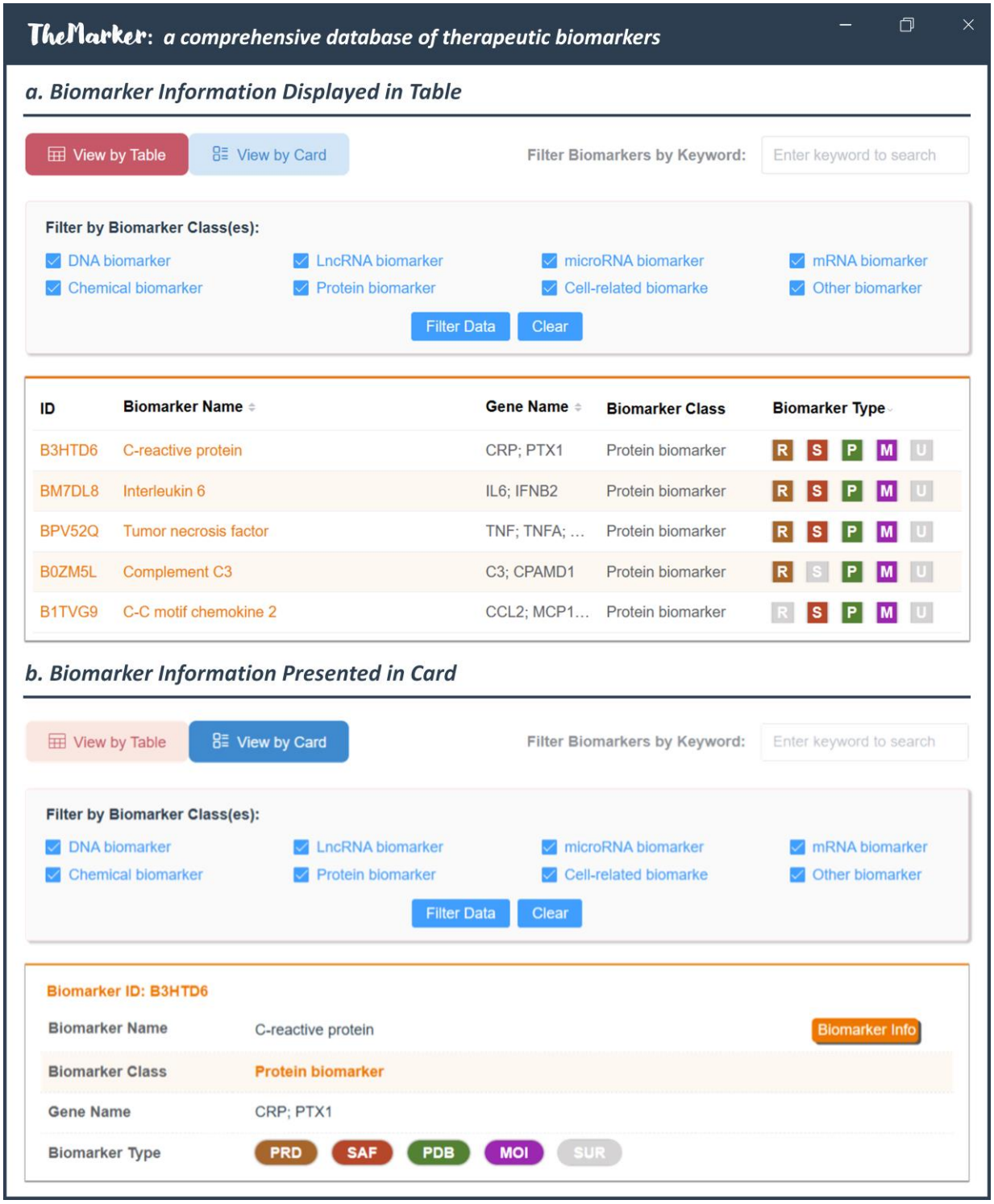
TAX: taxane; THA: thalidomide; TRA: trastuzumab

VCR: vincristine; VNR: vinorelbine

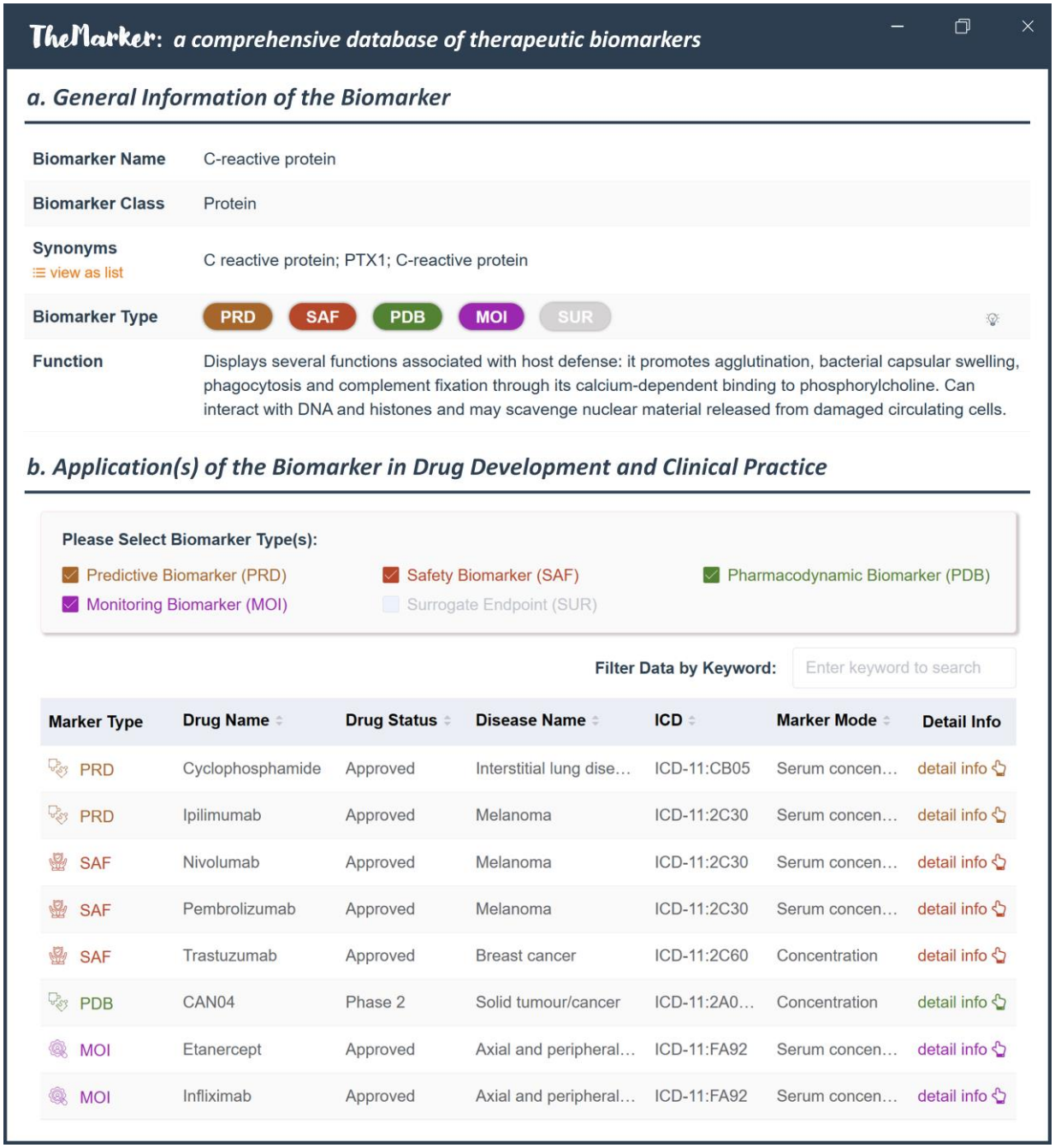
6-MP: mercaptopurine



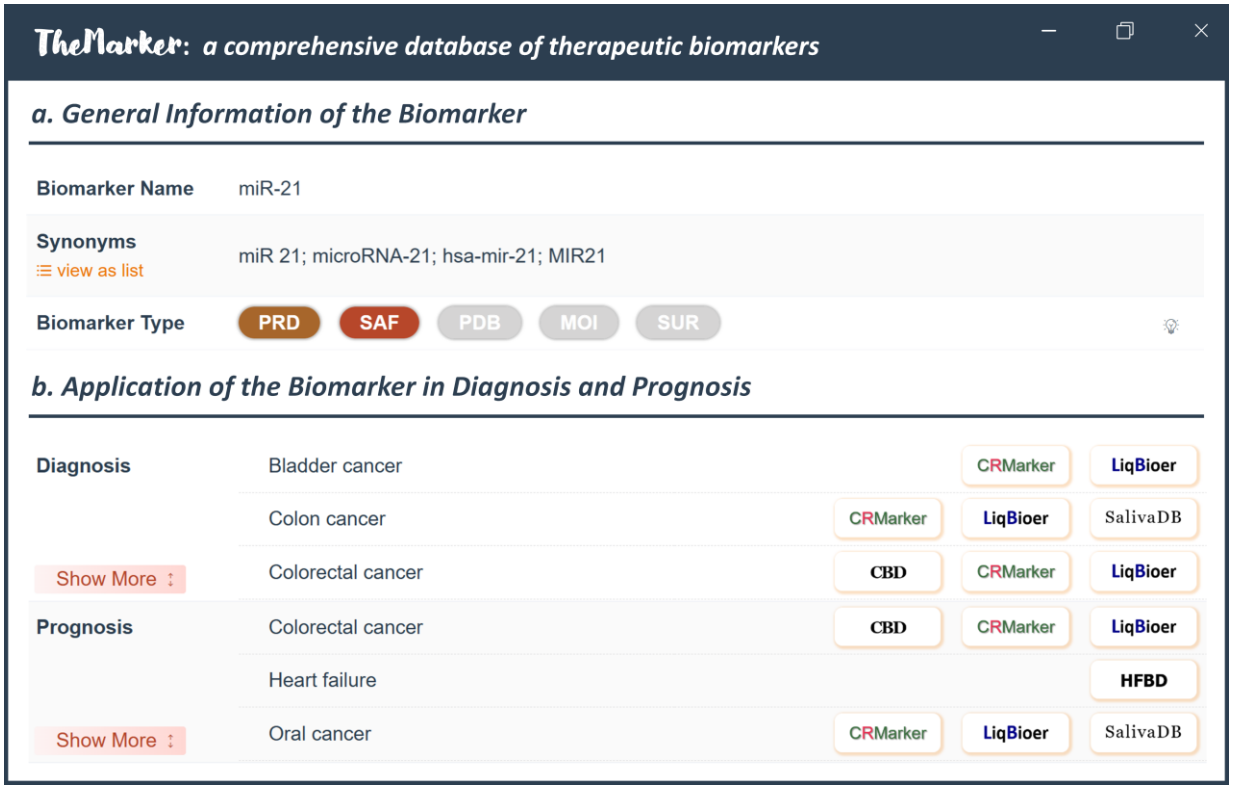
Supplementary Figure S1. Two views of the searching results page. (***a***) search results displayed in *table*; (***b***) search results presented in *card*. User can select their preferred view by clicking the buttons of '*View by Table*' and '*View by Card*'. In both views, all drug data are filtered according to the entering of the particular keywords.



Supplementary Figure S2. Two views of browse page in TheMarker. (***a***) browse data displayed in *table*; (***b***) browse data presented in *card*. User can choose their preferred view by clicking the buttons of '*View by Table*' and '*View by Card*'. In both views, all *therapeutic biomarker* data can be filtered by selecting specific biomarker classes or by entering particular keywords.



Supplementary Figure S3. Table view of the application(s) of a biomarker in drug development and clinical practice. Users can filter the data by choosing specific biomarker classes or entering the particular keywords.



Supplementary Figure S4. A typical TheMarker page showing related information of diagnosis and prognosis biomarker. For any therapeutic biomarker (ThMAR), the data of its application in diagnosis and prognosis were provided in the bottom of the biomarker page. Users can find more detailed data by clicking the database logos on the right side. Those databases used to collect the diagnosis/prognosis information included: CBD, CMBD, CRC-EBD, CRMarker, CancerLivER, GlioMarker, HFBD, LBD, LiqBioer, Lnc2Cancer, MarkerDB, and SalivaDB.