

ADReCS-Target: target profiles for aiding drug safety research and application

Li-Hong Huang^{1,†}, Qiu-Shun He^{1,†}, Ke Liu¹, Jiao Cheng¹, Min-Dong Zhong¹,
Lin-Shan Chen¹, Li-Xia Yao² and Zhi-Liang Ji^{1,3,*}

¹State Key Laboratory of Stress Cell Biology, School of Life Sciences, Xiamen University, Xiamen, Fujian 361102, PR China, ²Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA and ³The Key Laboratory for Chemical Biology of Fujian Province, Xiamen University, Xiamen, Fujian 361005, PR China

Received July 24, 2017; Revised September 12, 2017; Editorial Decision September 22, 2017; Accepted October 04, 2017

ABSTRACT

Delivering safe and effective therapeutic treatment to patients is one of the grand challenges in modern medicine. However, drug safety research has been progressing slowly in recent years, compared to other fields such as biotechnologies and precision medicine, due to the mechanistic complexity of adverse drug reactions (ADRs). To fill up this gap, we develop a new database, the Adverse Drug Reaction Classification System-Target Profile (ADReCS-Target, <http://bioinf.xmu.edu.cn/ADReCS-Target>), which provides comprehensive information about ADRs caused by drug interaction with protein, gene and genetic variation. In total, ADReCS-Target includes 66,573 pairwise relations, among which 1710 are protein–ADR associations, 2613 are genetic variation–ADR associations, and 63,298 are gene–ADR associations. In a case study of exploring the mechanism of rash, we find that HLAs, C1QA and APOA1 are the key gene players and thus can be potential targets (or biomarkers) in monitoring or countermining rashes. In summary, ADReCS-Target can be a useful resource for the biomedical scientific community by serving researchers in the fields of drug development, clinical pharmacology, precision medicine, and from web lab to high-throughput computational platform. Particularly, it helps to identify drug with better ADR profile and design safer drug therapy regimen.

INTRODUCTION

Adverse drug reactions (ADRs) are harmful or unpleasant clinical events triggered by therapeutic treatment. Each year, ADRs cause ~3.8% of new hospital admissions in Europe between 2000 and 2014 (1). Besides, unsatisfied safety

profile is one of the major two causes of failure in new drug discovery and development and adds significantly to skyrocketing drug price (2). In recent years, the launch of precision medicine in the United States and similar projects in other countries call for right drug, right dosage and right time for right patient (3). This paradigm shift brings new challenges to current drug safety research.

Achieving better drug safety profile requires complete and accurate information of drug toxicity in patients. It is known that the on-target and off-target mechanisms explain most dose-dependent or idiosyncratic ADRs. Poor pharmacokinetics of drug can also lead to ADRs, by involving a bundle of enzymes, transporters and ion channels. Furthermore, the genetic variations are found to largely determine individual responses to drug therapy in terms of both efficacy and toxicity. For instance, the different genotypes of MHC HLA genes explains individual allergy responses (4). Using genetic to understand the variation in individual responses to drugs even established the relatively new field of pharmacogenomics (5). Regardless the exact molecular mechanisms behind most ADRs remain unclear, multiple genes, proteins, or genetic variations can be relevant (6). Therefore, integrated representation of information about ADR related genes, proteins and genetic variations is necessary for systematic investigation of ADR mechanisms, design of novel drug regimen, and rational drug refinement for better safety profile.

Many research efforts have been made to collect information regarding ADR related molecules. Previously, we developed Drug Adverse Reaction Target database (DART) (7) and Drug Induced Toxicity related Protein database (DITOP) (8) by manually extracting drug toxicity-protein relations from scientific literatures. Mattingly *et al.* developed and maintained the Comparative Toxicogenomics Database (CTD), which provides drug–gene–disease trilateral relations (9). PharmGKB (10), an international collaboration on pharmacokinetics, collects information of how genetic variations impact individual drug responses includ-

*To whom correspondence should be addressed. Tel: +86 592 2182897; Fax: +86 592 2182897; Email: appo@xmu.edu.cn

†These authors contribute equally to this work as first authors.

ing adverse reactions. More databases such as TTD (11), DrugBank (12), ClinVar (13) and dbGaP (14) also include some ADR related information. Unfortunately, none of the mentioned resources fulfill the need of systematic drug safety research from all three aspects—gene, protein and genetic variation by its own.

By recognizing the need of ontology and standardized controlled vocabulary to describe ADRs to enable large-scale automatic analysis (15), we further developed the Adverse Drug Reaction Classification System (ADReCS) for standardization and hierarchical classification of ADR terms (16). Built on top of the ADReCS framework, we introduce a new database, the Adverse Drug Reaction Classification System-Target Profile (ADReCS-Target) in this work. It provides comprehensive information about ADRs caused by drug interaction with protein, gene and genetic variation.

The rest of this paper is structured as following: In the database construction section, we explain how to collect, validate and standardize the data. In the data access section, we illustrate how users can query and download the data. Then, in the section of data statistics and database comparison, we make a statistic of the ADReCS-Target database and compare it with existing related resources. Following that, we demonstrate a case study, showing how to use the database to aid ADR mechanism study.

DATABASE CONSTRUCTION

Data extraction

ADReCS-Target mainly consists of three kinds of associations: protein-ADR, gene-ADR and genetic variation-ADR. We collected the protein-ADR associations from text-mining of the public scientific literatures. First, we mined relevant abstracts containing both keywords of protein-ADR pair from the local MEDLINE database (by August 2016) using a self-coded program. The protein name and synonyms were derived from the UniProt database. The ADR term and synonyms were derived from the ADReCS database. Then, we extracted the protein-ADR relations by reading the retrieved abstracts or full articles manually. In this study, only direct indications of protein-ADR interactions were collected. All the relations were double-checked.

The gene-ADR associations are selected results of the ADRAAlert project (<http://bioinf.xmu.edu.cn/ADRAAlert/gene/index.jsp>). The ADRAAlert is a mathematical model that determines the strength of gene-ADR associations by statistically solving complex drug-gene-ADR network in a large scale of 1156 ADRs and 8571 genes (17). In that project, we validated the gene-ADR associations performed well in drug safety evaluation using both internal data (marketed drugs) and external independent data (drugs in clinical trial). In ADReCS-Target, we only collected those gene-ADR pairs with association strength >0.05 for database presentation under two considerations: First, according to the distribution analysis of gene-ADR association strength, 0.05 is the proper cut-off for picking out those strong gene-ADR associations, which consists of $\sim 1.13\%$ of overall 2 443 256 gene-ADR pairs. Secondly, inclusion of all gene-ADR associations in database will

largely sacrifice the speed of database search and data presentation, especially visualization of the ADR-target interaction network. However, the user can download the full list of gene-ADR associations from the website of ADRAAlert.

We acquired and integrated the genetic variation-ADR associations mainly from public resources as well as text mining of scientific literatures. On one hand, we downloaded the drug relevant genetic variations from several public databases such as DrugBank, Allele Frequency Net Database, and GWAS Catalog. On the other hand, we mined relevant abstracts of drug-genetic variation interactions from the MEDLINE as what we did in retrieving protein-ADR interactions. Then, we extracted manually the reliable ADR-genetic variation relations by the criteria of having either multiple evidences or strong clinical significance. For instance, when selecting genetic variation-ADR associations from DrugBank, we excluded those inferred relations. From the GWAS Catalog database, we collected genetic variation-ADR relations with P -values $< 1.0 \times 10^{-5}$. From the Allele Frequency Net Database, we chose ADR associations determined by statistical evaluation of case-control study in patients. Before going further, all these associations were double-checked for information validity and integrity.

Data Standardization and ADR hierarchy

ADReCS-Target standardized ADR terms by adopting the exactly same protocol as ADReCS (16). ADReCS-Target also deployed same ADR hierarchy as ADReCS; the four levels in the hierarchy, from general term to specified term, are System Organ Class (SOC), High Level Group Term (HLGT), High Level Term (HLT), and Preferred Term (PT). Accordingly, each ADR term in ADReCS-Target has a typical ADReCS ID in xxx.xxx.xxx.xx format.

Besides, ADReCS-Target standardized gene name, protein name, and genetic variation by referring to the NCBI Entrez database (18), the UniProt database (19) and the dbSNP database (20), respectively.

Database construction

The ADReCS-Target database is constructed on a system architecture of Linux+Tomcat+Java, and at the same time adopts Oracle *11g* as background data management system. Of note, we coded ADReCS-Target using HTML5 technology to support mobile access. The ADReCS-Target database is freely accessible at: <http://bioinf.xmu.edu.cn/ADReCS-Target>.

DATA ACCESS

Data retrieval

The quick search mode. ADReCS-Target provides two data retrieval modes: the quick search mode (Figure 1A) and the browse mode (Figure 1B). In the homepage and all data retrieval or result webpages, ADReCS-Target embeds a foolproof quick search box for flexible data retrieval. The search box (Figure 1A) accepts both complete or partial keyword input in either aspects of ADR term (or synonym), ADReCS.ID (in xx.xx.xx.xxx format), gene symbol, gene Entrez.ID, protein name, SwissProt AC, genetic

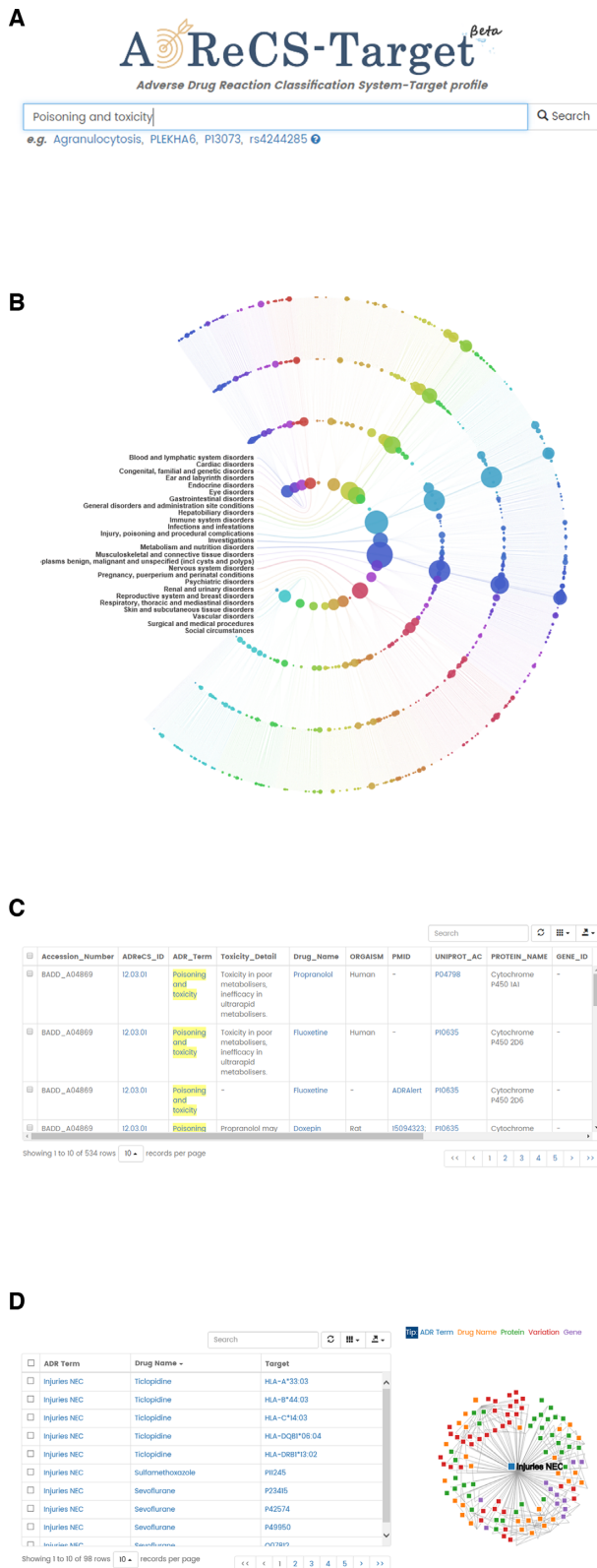


Figure 1. The interfaces of ADReCS-Target: the quick search interface (A), the browse interface of ADR hierarchy (B), the page of search results (C), and the page of detailed information (D).

variation_ID (dbSNP_ID in rsxxxx format), drug name (or synonym) and DrugBank_ID (DBxxxxx). The search engine will automatically recognize the input pattern for deciding either exact ID search or fuzzy keyword search. For instance, an input keyword starting with ‘DB’ and following by five digits will trigger DrugBank_ID search. Otherwise, the engine will search all supporting data types in a fuzzy way. Short keyword like less than three characters will take much longer searching time. Therefore, we strongly suggest using complete keyword, long keyword, or ID for best data retrieval experience. Current version of engine does not accept wild characters like ‘*’, ‘+’.

ADReCS-Target responds hits of keyword search in tabular format. The results are a list of ADR associations hitting the query keyword in any content of drug, gene, protein, genetic variation or ADR. Each association record contains information of ADReCS accession number, ADReCS_ID, ADR Term, toxicity detail, drug name and protein/gene/genetic variation properties when available (Figure 1C). The query keyword is highlighted in yellow, and the unavailable information is marked as ‘-’. Of note, the record shows the information of protein, gene, or genetic variation only if it is part of the association. For instance, the gene-ADR association record only shows gene properties but that of protein or genetic variation. Clicking on the ADReCS_ID will list all associations with this ADR in a new page of detailed information (Figure 1D).

The detailed information page presents ADR associations in two sections: a summarizing table and a dynamic interaction network. The table consists of three columns of ADR Term, Drug Name and Target (protein, gene or genetic variation). Clicking on the ADR Term will redirect to the ADR ontology information in ADReCS. ADReCS-Target also builds the crosslink of drug name to ADReCS for pharmacological and chemical information when available; otherwise, a link to PubChem is given. Moreover, the association record also gives several crosslinks to external resources, including ADReCS, DrugBank, UniProt, Entrez Gene and dbSNP. The interaction network is constructed upon the data presented in the summarizing table. In this network, different color nodes stand for different data types, for instances, blue for ADR term, orange for drug, green for protein, purple for gene, and red for genetic variation. The network works interactively that nodes in the network are draggable and selectable. Mouse-on the node will show the name and selecting a node will highlight connections with this node.

The browse mode. Other than the quick search mode, ADReCS-Target applies the Data-Driven Documents D3.js technology for direct and hierarchical visualization of ADR associations (Figure 1B). All ADR associations are categorized into a four-layer open concentric circle. From the inner to the outer, the circle layer represents the ADR hierarchy of SOC, HLT, HLT and PT, respectively. On each of these circle layers, there are many small transparent color circles. Each color circle stands for an ADR term in the ADR hierarchy. For instance, a circle on the HLT layer represents a HLT ADR term. The color of circle is just for differentiating ADR terms; however, the size of the circle is positively proportional to the association number with the ADR. The

bigger of the circle, the more associations with the ADR. The ADR circles are arranged by the order of ADR hierarchy that ADR circles of same or close ADR groups are allocated together. Parent ADR terms and children ADR terms are linked by lines. Mouse-on a circle will display the number and the type of associations for this ADR circle. Clicking on the circle will list all ADR associations in a new page (Figure 1D).

Data download

ADReCS-Target provides two ways for data retrieval. The user can download selected records via the embedded download function in six formats such as JSON, XML, CSV, TXT, SQL and MS-Excel. The database also supports batch data retrieval via the Download page (<http://bioinf.xmu.edu.cn/ADReCS-Target/download.jsp>). The data downloading is free.

DATA STATISTICS AND DATABASE COMPARISON

Overall, the ADReCS-Target deposits 66,573 association pairs with 2,257 standard ADR terms, including 11 SOCs, 17 HLGs, 554 HLTs and 1675 PTs. Of the 66 573 association pairs, 1710 are protein-ADR relations, 63 298 are potential drug-gene-ADR relations and 2613 are genetic variation-ADR relations. The database also includes 662 marketed drugs.

Here, we make a comparison of ADReCS-Target database to other similar and publicly available databases. DrugBank (12) and TTD (11) are the two mostly cited drug target databases, but they primarily focus on therapeutic targets for drugs on market or under research and development. Some of those therapeutic targets account for the dose-dependent ADRs, thus they are also ADR targets. According to our survey, TTD covers minimal information of ADR targets and DrugBank contains 112 non-inferred records of ADR related genetic variations as of January 2017. Eighty two of them are included in ADReCS-Target. PharmGKB (10) is an international project of identifying the relationships between genetic variations and drug responses, including ADRs. But the ADR target information is mostly embedded in unstructured descriptive text. Besides, several genetic variation databases such as dbSNP (20), ClinVar (13), dbGaP (14), Allele Frequency Net Database (21) and GWAS Catalog (22) also contain many genetic variations to clinical significance, including a small portion of ADR-associated variations. ADReCS-Target uses them as data sources for collecting genetic variation-ADR associations. However, ADReCS-Target further standardizes, integrates, and represents those genetic variation-ADR associations in structured relational database table. CTD (9) extracts chemical-gene-disease trilateral relations from literatures, although many of which are chemical, instead of drugs, induced diseases and symptoms. Again, CTD does not specify those relations in relational database table and make search and query impossible. ADReCS-Target builds on top of these databases and adds crosslinks to them, as well as UniProt and Entrez Gene to ensure included information both comprehensive and searchable. A more detailed comparison of ADReCS-Target with these databases is given in Table 1.

POTENTIAL APPLICATIONS

Essentially, ADReCS-Target is developed as a comprehensive data source for better understanding of ADR at a combo view of protein, gene and genetic variation. It also provides an optional way for systematic exploration of ADR. For instance, rashes, eruptions and exanthems NEC (REE, ADReCS.ID: 23.03.13) are a group of common skin ADRs whose mechanisms have yet been illustrated clearly. We searched the ADReCS-Target database by ADReCS.ID, resulting in one protein, 122 genetic variations and 274 genes associated with REE. These associations correspond to 281 nonredundant human genes. By mapping these genes into the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (23,24), we identified 173 REE-associated pathways. Of these pathways, the cellular senescence and the NOD-like receptor signaling pathway are the major two REE-associated pathways, which both pathways involve more than ten REE-associated genes. Furthermore, we constructed a weighted gene-gene interaction network upon the 137 REE-associated genes mapping to the 173 KEGG pathways. Analysis of this network identified 11 hub genes, which all these hub genes have a connectivity degree of 7 or above in the network. These hub genes are nine human leukocyte antigens (HLA-A, HLA-B, HLA-C, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQB1, HLA-DRB1 and HLA-F), complement C1q A chain (C1QA) and apolipoprotein A1 (APOA1). They are likely the major gene players in drug-induced rashes (Figure 2A). Literature surveillance of prior studies found some of these major gene players like HLA-A (25,26) and C1QA (27) associated with REE in different ethnic population. Hence, these hub genes can serve as potential targets (or biomarkers) in monitoring or counteracting REE.

Comparatively, using rash, eruption, and exanthem as keywords, we found one nonredundant REE-associated protein in DITOP, zero non-inferred REE-associated genetic variation in DrugBank, 15 significant and nonredundant REE-associated genetic variations in PharmGKB, and 28 REE-associated genes in CTD. In basis of these 44 nonredundant association pairs, we undertook the same pathway and network analyses. As the results, we identified 105 REE-associated KEGG pathways. The cellular senescence and the NOD-like receptor signaling pathway were still identified as the major two REE-associated pathways however involving only six REE-associated genes. As well, we detected C1QA, HLA-DBP1 and HLA-DRB1 as hub genes with the criteria of connectivity degree above seven (Figure 2B). Therefore, compared to the existing databases in mechanistic exploration of ADRs, ADReCS-Target shows substantial advantages in many aspects such as convenience in acquiring data, integrity in data types, consistence in data quality, and potential in automatic ADR research.

CONCLUSIONS

To our limited knowledge, ADReCS-Target is the most comprehensive target database for drug safety research so far. It builds on the other two databases DITOP (8) and

Table 1. Comparison of ADRCS-Target with several relevant databases

Resources	ADReCS-Target	DITOP 1.0	GWAS Catalog V1.0	CTD (by Jun 2017)	DrugBank (by July 2017)	Allele Frequencies (by July 2017)
Drug count	662	515	120	6448 ^a	9591 (90) ^d	30
ADR count	2257	539	129	5805 ^b	N.A. ^d	16
Variation-ADR	2613	N.A.	675	N.A.	2845	1245
Gene-ADR	63 298	N.A.	N.A.	21 160 629 ^c	N.A.	N.A.
Protein-ADR	1710	1008	N.A.	N.A.	N.A.	N.A.
All ADR Associations	66 573	1008	675	21 160 629 ^c	2845	1245
Partner	ADR	ADR	Trait	Disease	ADR	ADR
ADR Standardization	Yes	Yes	No	No	No	No

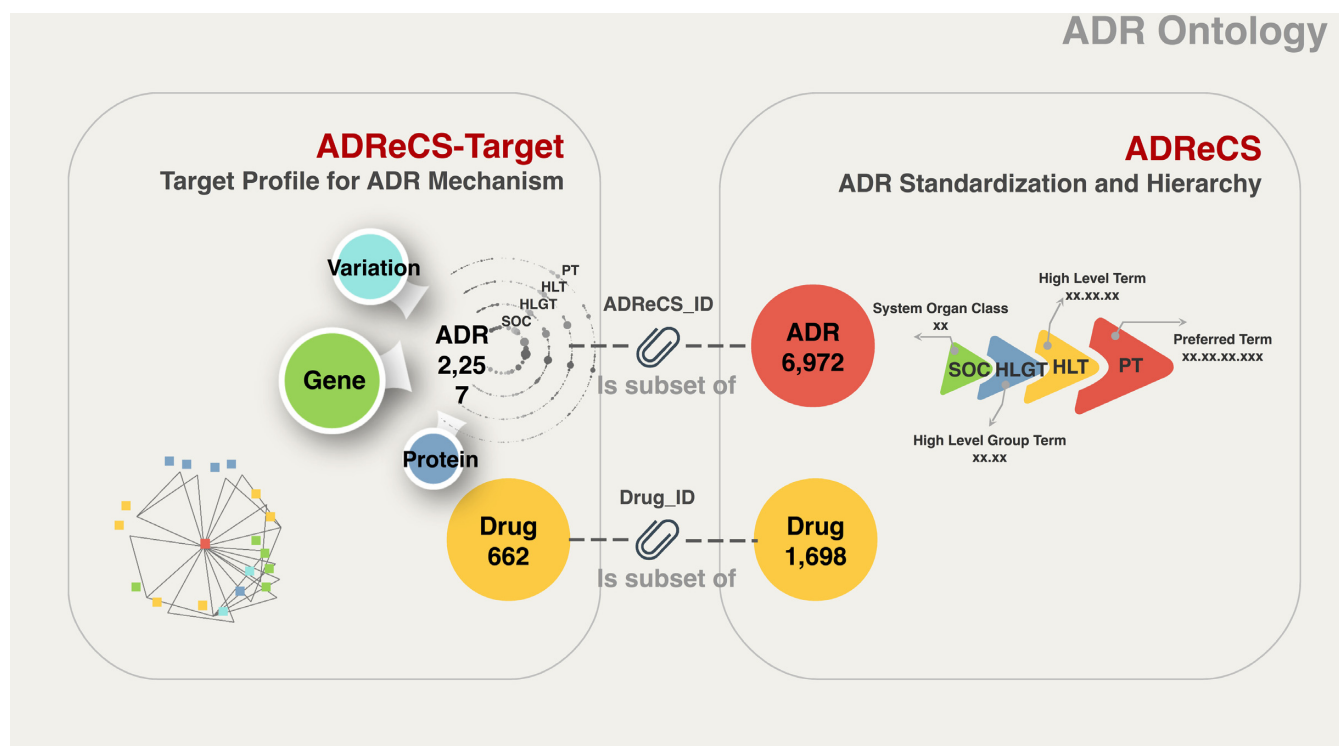
N.A. = not available.

^aOnly a part of chemicals in CTD are drugs, the exact number of drugs is not provided by the database.

^bThe associations in CTD are between genes and diseases.

^cIn CTD, 26 681 gene-disease associations have direct evidences.

^dOf 9591 drugs in DrugBank, 90 have associations with 103 non-redundant ADRs. The exact number of ADRs is not provided by database.

**Figure 3.** Construction of ADRCS-Target on top of ADRCS hierarchy.

DART (7) we developed previously. Compared to DITOP and DART, ADRCS-Target has four significant improvements: (i) ADRCS-Target implements ADRCS ontology. In other words, all ADR terms are standardized and characterized clearly (Figure 3), which enables sharing and communicating the information among people and computer, reuse and easy large-scale analysis of the knowledge. (ii) Data in ADRCS-Target are represented not in an isolated fashion, but in a connected network or systematic fashion. (iii) ADRCS-Target employs visualization technology for better data retrieval and representation. In particular, the database is implemented in HTML5, thus supports access on mobile devices. (iv) The data coverage of ADRCS-Target is about 66 times more than that of DITOP and DART. Especially, gene-ADR associations and

genetic variation-ADR associations are included for the first time.

In future, we plan to fully integrate ADRCS-Target and ADRCS, to enable users review all ADR relevant information, including ADR terms, ADR target profiles, and drug properties, in one-stop service. Secondly, we will develop various data analysis and visualization tools to improve user experience and to enable customized data mining and analysis tasks requested by users. Undoubtedly, we will keep expanding the database in terms of both record quantity and types. This kind of updates will be part of our maintenance plan. Our long-term goal is to keep ADRCS-Target informative, up-to-date and easy to use. In short, ADRCS-Target will serve as a major resource for drug discovery scientists in all sorts of drug safety studies. It will

also be of value to the communities of clinical pharmacology and precision medicine, especially in refining drugs with better ADR profiles and designing safer drug therapy regimens.

ACKNOWLEDGEMENTS

We would like to thank Professor Quan Zou, Tianjin University, Tianjin and the colleagues at Bioinformatics-Aided Drug Discovery Group, Xiamen for their contributions to the database.

FUNDING

Natural Science Foundation of China [NSFC#30873159, NSFC#31671362]; XMU-Xiamen Xianyue Hospital collaboration [#XDHT2016013C]. Funding for open access charge: Natural Science Foundation of China.

Conflict of interest statement. None declared.

REFERENCES

- Bouvy, J.C., De Bruin, M.L. and Koopmanschap, M.A. (2015) Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. *Drug Saf.*, **38**, 437–453.
- Jefferys, D.B., Leakey, D., Lewis, J.A., Payne, S. and Rawlins, M.D. (1998) New active substances authorized in the United Kingdom between 1972 and 1994. *Br. J. Clin. Pharmacol.*, **45**, 151–156.
- Collins, F.S. and Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.*, **372**, 793–795.
- Howell, W.M., Turner, S.J., Hourihane, J.O., Dean, T.P. and Warner, J.O. (1998) HLA class II DRB1, DQB1 and DPB1 genotypic associations with peanut allergy: evidence from a family-based and case-control study. *Clin. Exp. Allergy*, **28**, 156–162.
- Nebert, D.W. (1999) Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin. Genet.*, **56**, 247–258.
- Pirmohamed, M. and Park, B.K. (2001) Genetic susceptibility to adverse drug reactions. *Trends Pharmacol. Sci.*, **22**, 298–305.
- Ji, Z.L., Han, L.Y., Yap, C.W., Sun, L.Z., Chen, X. and Chen, Y.Z. (2003) Drug Adverse Reaction Target Database (DART). *Drug Saf.*, **26**, 685–690.
- Zhang, J.-X., Huang, W.-J., Zeng, J.-H., Huang, W.-H., Wang, Y., Zhao, R., Han, B.-C., Liu, Q.-F., Chen, Y.-Z. and Ji, Z.-L. (2007) DITOP: drug-induced toxicity related protein database. *Bioinformatics*, **23**, 1710–1712.
- Davis, A.P., Murphy, C.G., Saraceni-Richards, C.A., Rosenstein, M.C., Wieggers, T.C. and Mattingly, C.J. (2008) Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical–gene–disease networks. *Nucleic Acids Res.*, **37**, D786–D792.
- Thorn, C.F., Klein, T.E. and Altman, R.B. (2013) PharmGKB: the pharmacogenomics knowledge base. *Pharmacogenomics: Methods Protoc.*, 311–320.
- Chen, X., Ji, Z.L. and Chen, Y.Z. (2002) TTD: therapeutic target database. *Nucleic Acids Res.*, **30**, 412–415.
- Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
- Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. and Maglott, D.R. (2013) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.
- Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J. and Phan, L. (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
- Zhichkin, P.E., Athey, B.D., Avigan, M.I. and Abernethy, D.R. (2012) Needs for an expanded ontology-based classification of adverse drug reactions and related mechanisms. *Clin. Pharmacol. Ther.*, **91**, 963–965.
- Cai, M.-C., Xu, Q., Pan, Y.-J., Pan, W., Ji, N., Li, Y.-B., Jin, H.-J., Liu, K. and Ji, Z.-L. (2014) ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic Acids Res.*, **43**, D907–D913.
- Xiang, Y.P., Liu, K., Cheng, X.Y., Cheng, C., Gong, F., Pan, J.B. and Ji, Z.L. (2015) Rapid assessment of adverse drug reactions by statistical solution of gene association network. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, **12**, 844–850.
- Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R. et al. (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
- Consortium, T.U. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
- Ghattaoraya, G.S., Dundar, Y., Gonzalez-Galarza, F.F., Maia, M.H., Santos, E.J., da Silva, A.L., McCabe, A., Middleton, D., Alfirevic, A., Dickson, R. et al. (2016) A web resource for mining HLA associations with adverse drug reactions: HLA-ADR. *Database*, **2016**, baw069.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T. and Hindorf, L. (2013) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
- Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
- Yan, S., Chen, S.A., Zhang, W., Yang, F., Yang, Y., Zhu, Q., Zhu, H., Sun, X., Jiang, M., Su, Y. et al. (2016) HLA-A*02 alleles are associated with tetanus antitoxin-induced exanthematous drug eruptions in Chinese patients. *Pharmacogenet. Genomics*, **26**, 538–546.
- Fricke-Galindo, I., Martinez-Juarez, I.E., Monroy-Jaramillo, N., Jung-Cook, H., Falfan-Valencia, R., Ortega-Vazquez, A., Alonso-Vilatela, M.E. and Lopez-Lopez, M. (2014) HLA-A*02:01:01/-B*35:01:01/-C*04:01:01 haplotype associated with lamotrigine-induced maculopapular exanthema in Mexican Mestizo patients. *Pharmacogenomics*, **15**, 1881–1891.
- Sun-Tan, C., Ozgur, T.T., Kilinc, G., Topaloglu, R., Gokoz, O., Ersoy-Evans, S. and Sanal, O. (2010) Hereditary C1q deficiency: a new family with C1qA deficiency. *Turkish J. Pediatr.*, **52**, 184–186.