Feature, Function, and Information of Drug Transporter Related Databases

Jiayi YIN¹, Fengcheng LI¹, Zhaorong LI², Lushan YU¹, Feng ZHU¹,²,³,* and Su ZENG¹, ⁴,*

¹ Institute of Drug Metabolism and Pharmaceutical Analysis, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

² Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China

³ Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Hangzhou 310018, China

⁴ Zhejiang Province Key Laboratory of Anti-Cancer Drug Research, Cancer Center of Zhejiang University, Hangzhou 310058, China

Running Title: Drug Transporter Related Databases

Corresponding Authors: Prof. Su ZENG, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China. Tel: +86-571-8820-8407, Email: zengsu@zju.edu.cn; Prof. Feng Zhu, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China. Tel: +86-189-8946-6518, Email: zhufeng@zju.edu.cn.

Number of Text Pages: 31
Number of Tables: 5
Number of Figures: 3
Number of References: 112
Number of Words in Abstract: 159
Number of Words in Introduction: 749
Number of Words in the Discussion: 504
Abbreviations:

Abstract

With the rapid progress in pharmaceutical experiments and clinical investigations, extensive knowledge of drug transporters (DTs) has accumulated, which is valuable data for the understanding of drug metabolism and disposition. However, such data is largely dispersed in the literature, which hampers its utility and significantly limits its possibility for comprehensive analysis. A variety of databases have, therefore, been constructed to provide DT-related data, and they were reviewed in this study. First, several knowledge bases providing data regarding clinically important drugs and their corresponding transporters were discussed, which constituted the most important resources of DT-centered data. Second, some databases describing the general transporters and their functional families were reviewed. Third, various databases offering transporter information as part of their entire data collection were described. Finally, customized database functions that are available to facilitate DT-related research were discussed. This review provided an overview of the whole collection of DT-related databases, which might facilitate research on precision medicine and rational drug use.

Significant Statement

A collection of well-established databases related to DTs were comprehensively reviewed, which were organized according to their importance in drug ADME research. These databases could collectively contribute to the research on rational drug use.
1. Introduction

Drug efficacy and safety are largely determined by multiple processes (absorption, distribution, metabolism, and excretion) that regulate pharmacokinetics (Terada et al., 2015). A variety of endogenous molecules (mostly proteins) are determinants of these processes (Giacomini et al., 2010). Typical examples of these molecules include drug-metabolizing enzymes that transform parent drugs to metabolites of very different physicochemical and pharmacological properties (Yu and Zhong, 2016; Yu et al., 2017; Hitchings and Kelly, 2019), and drug transporters (DTs) that mediate the uptake of endo/exogenous substances into cells as well as their efflux (Rodieux et al., 2016; To et al., 2017; Shu et al., 2019). Among these molecules, DTs are capable of (1) determining the pharmacokinetic profile of drugs by regulating their absorption, distribution and excretion or indirectly modifying their metabolism (DeGorter et al., 2012; Yang et al., 2019), (2) affecting drug pharmacodynamics by delivering them to proper target sites, controlling differential drug concentrations among tissues or altering their interactions with other molecules (Hu et al., 2015), (3) inducing drug toxicity through DT’s vulnerability to drug-drug interaction (DDI) or leading to drug resistance by reducing its concentration in targeted cells (Zhang and Hagenbuch, 2019), and (4) facilitating target discovery and rational use of the drug by revealing the mechanism of DDI, identifying the potential therapeutic target or improving the treatment of specific disease (Garibsingh and Schlessinger, 2019). Due to the essential role of DTs in drug efficacy and safety, it is necessary to acquire as much DT-centered knowledge as possible (Li et al., 2017; Zhu et al., 2019; Tang et al., 2020; Wang et al., 2021; Zhou et al., 2021).

With the advancement of experimental analysis (Li et al., 2018a) and clinical investigation (Stopfer et al., 2016), an extensive amount of DT knowledge has accumulated, and mainly involves five types of DT-centered information: 1) expression, distribution and function (Lin et al., 2015), 2) epigenetic modification (Hirota et al., 2017), 3) structural conformation and
variation (Zheng et al., 2018), 4) exogenous regulation (Li et al., 2018a), and 5) genetic polymorphism (Peng et al., 2016). Particularly, the data regarding **DT’s expression, distribution and function** demonstrate its disease-differential expression (Evers et al., 2018), organism-dependent abundance (Durmus et al., 2015), tissue-specific distribution (Nixon et al., 2016), transporting functional family (Shen et al., 2017), and so on; the data of **epigenetic modification on DT** describe the DNA/histone methylation & acetylation (Liu et al., 2016), non-coding RNA regulation (Yu et al., 2019), and so on; the data of **DT’s conformation and structural variations** involve species-specific evolution (Dias and Sa-Correia, 2014), the structures of the entire transporter (Penmatsa et al., 2013) and functional conserved/substrate-binding domain (Xue et al., 2016), and so on; the data of **DT’s exogenous regulation** discuss the clinical drug-drug interactions (Kosa et al., 2018), regulatory substrate, inhibitor & inducer (Muller et al., 2018), and so on; the data describing **DT’s genetic polymorphisms** provide the cytogenetic locations (Lewis and Girisha, 2020), disease indication induced by hereditary factor (Karimian et al., 2020), genetic variant & frequency (Veldic et al., 2019), and so on. The above knowledge is valuable for understanding the drug ADME process (Nigam, 2015; Ye et al., 2019), which is thus essential for current research on drug metabolism (Li et al., 2020b; Wang et al., 2020a; He et al., 2021) and disposition (Bai et al., 2016; Kawahara et al., 2020).

However, such valuable knowledge is largely dispersed in the literature, which hampers its utility and significantly limits its possibility for comprehensive analysis (Li et al., 2018b; Yang et al., 2020b). Therefore, a variety of databases have been constructed to offer DT-related data (Wang et al., 2020c; Yin et al., 2020; Saier et al., 2021). Some of them offer explicit information on drugs together with their corresponding transporters, and give special emphasis on DT variability (Yin et al., 2020); some others describe general transporters together with their (phylogenetic) classifications, and specifically highlight the ones of human origin.
(Elbourne et al., 2017); the remaining databases aim to provide general data on various transporters as a part of their data collection (UniProt, 2021). These databases guarantee the accessibility to DT-related knowledge, which is anticipated to be the key data resource for current ADME studies.

Therefore, a comprehensive review of these popular databases was conducted. First, the databases providing the data of clinically important drugs together with their corresponding transporters were discussed, which constituted the most important resources for DT-centered data. Second, several databases describing the general protein transporters and their functional families were reviewed, which were crucial for any study requiring transporter (especially DT) classification. Finally, the databases offering general transporter information as a part of their data collection were described, which could be adopted as essential complements to other available databases. The overview of these various types of databases were shown in Table 1.

2. Databases Providing Drugs and Their Corresponding Transporters

As estimated, approximately 10% (~2,000) of all proteins in the human genome are functionally associated with the transporting of endo or exogenous molecules (Hediger et al., 2013). However, regarding the transporting of clinically important drugs, the total number of involved DTs is still under debate (Yin et al., 2020). The valuable data of DTs together with their transporting drugs have been described in a variety of databases (Figure 1), and the characteristic groups of data covered by different databases were comprehensively reviewed and discussed in Table 2.

2.1 Pairing Data between Drug Transporter and Pharmaceutical Agent

Several databases are providing the pairing data between DTs and pharmaceutical agents. As the most widely used data resources of pharmaceutical agents, Therapeutic Target Database
DrugBank (https://go.drugbank.com/, (Wishart et al., 2018)), ChEMBL (https://www.ebi.ac.uk/chembl/, (Mendez et al., 2019)), IUPHAR/BPS Guide to Pharmacology (https://www.guidetopharmacology.org/, (Armstrong et al., 2020)), KEGG DRUG (https://www.kegg.jp/kegg/drug/, (Kanehisa et al., 2019)), and Metabolism and Transport Database (http://www-metrabase.ch.cam.ac.uk/, (Mak et al., 2015)) offered millions of molecules that were paired with the ADME-associated proteins. As shown in Figure 1, all these databases focused on the data of exogenous substances and their regulation (colored in green), especially the data of regulatory substances and the inhibitors/inducers of a studied DT (as shown in Table 2). Moreover, all the databases described the DTs of the approved/clinical trial drugs, and the number of DTs covered by these six databases varied greatly (from ~10 to over 100, as shown in Table 1). Among these databases, the Therapeutic Target Database and Metrabase were the only two describing the tissue-specific distribution of DTs, and such data could facilitate the critical analysis of distribution-induced adverse drug reactions (Yang et al., 2016).

PharmGKB (https://www.pharmgkb.org/, (Barbarino et al., 2018)) is a worldwide resource for pharmacogenomics knowledge that provides the alteration data of drug pharmacokinetics and pharmacodynamics that originate from genetic polymorphism. It focuses on the alterations in drug response and the effects on their clinical phenotypes and contains ~1,000 drugs related to the genetic variations on ~100 DTs. As shown in Figure 1, in addition to the exogenous regulation data, PharmGKB offered additional DT data on expression/distribution/function and genetic polymorphism. Compared with the databases above, PharmGKB offered many diverse groups of DT data (Table 2) by describing functional family, cytogenetic location, disease/phenotype induced by the hereditary factor, together with the genetic variant & frequency.
UCSF-FDA TransPortal (https://transportal.compbio.ucsf.edu/, (Morrissey et al., 2012)) and Transformer (http://bioinformatics.charite.de/transformer, (Hoffmann et al., 2014)) were two popular databases providing important drug transporters together with the exogenous substance and regulation, which contain 31 and 60 DTs for transporting approved or in clinical trial drugs, respectively. As demonstrated in Figure 1 and Table 2, UCSF-FDA TransPortal described the tissue-specific distribution information of DTs, while Transformer offered distinctive data on species-specific structural evolution and the 3D crystal structure of the entire transporter. Moreover, as shown in Table 2, these two databases are distinguished in covering the data of clinical drug-drug interaction (DDI) and are therefore applied to predict potential adverse drug reactions based on their DDI data (Cesar-Razquin et al., 2018; Carrascal-Laso et al., 2020). It is important to emphasize that the latest update of UCSF-FDA TransPortal was in 2012.

As a recently constructed pharmaceutical database, VARIDT (https://idrblab.org/varidt/, (Yin et al., 2020)) offered the most comprehensive set of DTs that were confirmed by the transporting drugs (either approved or clinical trial). As shown in Figure 1, this database contained the most diverse types of DT-related data (with the only exception of structure-based data). As described in Table 2, the data for all data groups under four different types were collected and provided. Moreover, the total number of DTs covered in this database was the largest compared with those knowledge bases in Figure 1. In particular, a comprehensive literature review of all drugs approved by the FDA and ~1,100 clinical trial drugs were first conducted. Then, a total of ~180 DTs were confirmed to transport approved drugs, and ~150 DTs were to transport clinical trial ones, which were substantially different from the relatively small numbers of DTs shown in available databases (Table 2). Due to the huge amount of accumulated DT data, the VARIDT is expected to provide strong support to the optimization of clinical treatment.

2.2 Diverse Data Illustrating Various Aspects of DT Variability
The variability data of DTs are essential for the determination of the inter-individual variations in drug response and side effects (Yee et al., 2018; Nie et al., 2020). Besides the variabilities in exogenous regulation and genetic polymorphism explicitly discussed in the 2.1 section, two additional aspects of variability (varied protein abundances & diverse epigenetic regulation) should be considered for DTs, due to their importance in bridging the preclinical investigations with clinical trials (Durmus et al., 2015) and leading to multidrug resistance in complex disease (Zhou et al., 2020), respectively. Therefore, current databases available for providing these two additional variability data were explicitly described in this review as follows.

The protein abundance of DTs plays an important role in several aspects of drug research, such as clinical toxicity analysis, clinical pharmacokinetics research and adverse reaction evaluation (Lin et al., 2015; Safar et al., 2019). There are three kinds of variability of DT abundances: (1) organism-specific expressions (Durmus et al., 2015), (2) tissue-differential distributions (Nixon et al., 2016) and (3) disease-dependent abundances (Evers et al., 2018). As provided in Table 2, the tissue-differential distribution data have been provided by multiple databases, such as TTD, PharmGKB, UCSF-FDA TransPortal, Metrabase and VARIDT, which further demonstrate the critical roles of such variability in drug disposition (Kawahara et al., 2020). For the remaining two kinds of variability data, VARIDT is the only knowledge base of such information, and the differential expression patterns are provided for 108 diseases and 3 model organisms.

Epigenetic regulation of DT genes has emerged as an important mechanism of individualized drug responses (Peng and Zhong, 2015; Hirota et al., 2017). Few epigenetic regulation data of DTs (Table 2) are provided by currently available knowledge bases, and the VARIDT is currently the only resource describing such variability. Particularly, it provided epigenetic regulation data on (1) epigenetic types (DNA methylation, ncRNA regulation, histone acetylation/methylation, etc.), (2) prevalence of occurrence, (3) locations, (4) description of the epigenetic phenomenon, (5) experimental methods, and (6) materials adopted for validating
each phenomenon. In total, over 20 thousand DNA methylations, ~100 histone modifications, and over 7 thousand non-coding RNA regulations were collected and comprehensively provided in the latest VARIDT.

3. Databases Describing Transporters and Transporter Families

Membrane transporters play critical roles in discovering new drugs and elucidating disease mechanisms (Nigam, 2018), which can be divided into channels, carriers, electron flow carriers, group translocators, and pumps for determining the molecular composition and energy state of cells (Cook et al., 2014; Saier et al., 2021). The identified drug uptake/efflux transporters may constitute only a small fraction of all these general transporters, but these general transporters are of great importance for elucidating drug metabolism (Wang et al., 2020b) and disposition (Kawahara et al., 2020). Thus, the databases available for describing general transporters and transporter families are valuable treasures for current DT research.

These databases include TCDB (Saier et al., 2021), METscout (Geffers et al., 2013), CFMD (Sosnay et al., 2011), iMusta4SLC (Higuchi et al., 2018), ABCMdb (Tordai et al., 2017), ALDinfo (Kemp et al., 2001), ABCA4 DataBase (Trezza et al., 2017), TransportDB (Elbourne et al., 2017), and SLC TABLES (Hediger et al., 2013)), which contain the data on classifying general transporters or describing a specific class of proteins. Such data provided the resources for the expansion of DT-related knowledge, which was described in many databases (as illustrated in Figure 2), and a variety of distinct data groups covered by these popular databases were also comprehensively reviewed and discussed in Table 3.

3.1 Systematic Classification of the General Protein Transporters

TCDB (https://www.tcdb.org/, (Saier et al., 2021)) is a well-established database for classifying general protein transporters. It contains transporter information of diverse species and describes
the general transporter in terms of structure, function, mechanism, evolution, disease/medicine, and related endogenous compounds (Figure 2 and Table 3). Collectively, this database covered over 20 thousand transporters classified into thousands of non-redundant function families, and it enables the assessments of the family members, the discovery of their evolutionary relationships, and inference of substrate and molecular functions. It has been adopted recently for optimizing experimental designs (Hong et al., 2020; Wang et al., 2020b; Yang et al., 2020a; Graf et al., 2021; Peng et al., 2021).

TransportDB (http://www.membranetransport.org/transportDB2/, (Elbourne et al., 2017)) is a functional annotation database containing information on a wide range of transporters derived from diverse species. Particularly, it annotates membrane transporters from ~2,000 genomes, and results in ~800 thousand transporters grouped into ~160 families. As shown in Figure 2, this database contained the data of expression, distribution, function and endogenous substrates for general transporter, which gives functional family, transporter phylogenetic classification, and structure of endogenous compounds (Table 3). TransportDB was suggested to be adopted for characterizing transporters of specific functions (Frioux et al., 2020; Bernstein et al., 2021).

3.2 Knowledge Bases Describing a Specific Transporter Family

SLC TABLES (http://slc.bioparadigms.org/, (Hediger et al., 2013)) is a database with specific data on ~400 absolute carriers (SLCs) classified into 52 functional families. This database is popular for describing detailed data on the functional family, endogenous substances, tissue differential distribution (Figure 2 and Table 3). Due to its specific scope of describing the SLC transporter family, it has been frequently used to facilitate SLC-related studies (Girardi et al., 2020). As another SLC-centered database, the METscout (http://metscout.mpg.de/, (Geffers et al., 2013)) provides the metabolic pathway and gene expression landscape, which describe wherein an organism each metabolic reaction takes place and where the SLCs transport the metabolites. It contains ~350 SLCs and covers almost all components within the mouse.
metabolic network. As illustrated in Figure 2, it provides the functional family, organism-specific abundance, tissue-differential distribution of transporters, together with the scRNA sequencing atlas (described in Table 3).

Other databases for a specific transporter family include: ABCMdb (http://abcm2.hegelab.org/, (Trezza et al., 2017)) & iMusta4SLC (http://cib.cf.ocha.ac.jp/slc/, (Higuchi et al., 2018)). Both knowledge bases focus on providing the genetic mutation information for a particular family of transporters (shown in Figure 2). ABCMdb has 45 ABC transcripts classified into 5 families and iMusta4SLC provides 573 SLC transcripts classified into 52 families. Both databases can predict potential disease liability based on the mutations in a general transporter and can help to develop extensive disease knowledge networks for improving disease management (Verkman et al., 2013; Tracewska et al., 2019; Chien et al., 2021).

3.3 Knowledge Bases Describing a Specific Transporter

A number of additional knowledge bases were constructed to describe genetic polymorphisms for specific transporter, which included CFMD (http://www.genet.sickkids.on.ca/cftr/, (Sosnay et al., 2011)), ALD Info (https://adrenoleukodystrophy.info/, (Kemp et al., 2001)), and ABCA4 DataBase (http://www.sbl.unisi.it/abca4/abcr_mainlist.php, (Trezza et al., 2017)). As indicated by their name, these knowledge bases contain information on only a single transporter, such as CFTR1, ABCD1, and ABCA4. Particularly, they provided comprehensive disease data that were associated with the sequential/structural polymorphism of a transporter. As illustrated in Figure 2 and Table 3, CFMD contains the data of cytogenetic locations and genetic variants & frequency, ALD Infor provides the data of genetically induced disease indications, and ABCA4 DataBase includes the data of genetically correlated phenotypes.

4. Databases Providing Transporters as Part of Their Data Collection
With the advancement of next-generation sequencing techniques (Lane et al., 2016; Yin et al., 2021), many protein-coding genes have been successfully characterized. In addition to the databases discussed above, some knowledge bases describing the whole protein atlas were also available. Within these databases, the transporters were described as a part of their entire data collection. Specifically, these databases included: dbSNP, EBI Expression Atlas, EncoMPASS, FINDbase, Human Protein Atlas, OMIM, PDB, PPTdb, MemProtDM, and UniProt. These databases focus on macroscopic and comprehensive descriptions of all proteins, and the information on all transporters is therefore provided in a generally described manner.

As shown in Figure 3, three types of transporter-related data were provided in these databases. For the data of **genetic polymorphism**, three databases were available, which included: dbSNP (https://www.ncbi.nlm.nih.gov/snp/, (Sherry et al., 2001)), FINDbase (http://www.findbase.org, (Kounelis et al., 2020)) and OMIM (https://omim.org/, (Amberger et al., 2019)). As described in Table 4, all three databases illustrated the genetic variant and frequency information of the studied transporters with OMIM being the only source for tissue-specific distribution, genetically correlated phenotype, and cytogenetic location. For the data of **expression, distribution and function**, three popular databases were available, which included: UniProt (https://www.uniprot.org/, (UniProt, 2021)), the Human Protein Atlas (https://www.proteinatlas.org/, (Uhlen et al., 2015)) together with EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home, (Papatheodorou et al., 2020)). Based on the descriptions in Table 4, all three databases focus on tissue-differential distribution and organism-specific abundance. UniProt provided the comprehensive information on functional families and transporter sequences, while EBI Expression Atlas places more emphasis on disease-varied expression. Among these three databases, the Human Protein Atlas has the widest coverage on the transporter data groups as described in Table 4. For the data of **structural conformation**, four reputable databases were available, which included: PDB (https://www.rcsb.org/, (Burley...
et al., 2021)), EncoMPASS (http://encompass.ninds.nih.gov, (Sarti et al., 2019)), MemProtDM (http://memprotmd.bioch.ox.ac.uk/, (Newport et al., 2019)) & PPTdb (http://pptdb.cgu.edu.tw, (Lee et al., 2019)). As described in Table 4, all four databases provided structural conformation for the entire transporter (if available). As the most comprehensive database for protein crystal structures, PDB also described the substrate binding domain of transporter with relatively high resolution. PPTdb had more data groups (Table 4) and gave the unique data on species-specific evolution and the structure of functionally conserved domain.

5. Customized Database Functions Facilitating DT-related Research

Besides the valuable DT-related data, a variety of customized functions facilitating DT-related research were also provided in several databases. These functions included sequence-based DT similarity screening (Elbourne et al., 2017), structure-based similarity screening for transported drug (Mendez et al., 2019), disease/tissue-specific differential expression analysis (Uhlen et al., 2015), interplay analysis among multiple DT variabilities (Yin et al., 2020), functional analysis based on signaling pathways (Kanehisa et al., 2019), DT annotation and classification (Saier et al., 2021), and so on. Based on these valuable functions together with their comprehensive DT-related information, the available databases provided much-enhanced power in the research of drug metabolism and disposition. As shown in Table 5, these functions facilitated the structure-based drug design/identification (Yu et al., 2016), discovery of target druggability based on DT sequence (Frioux et al., 2020), disease/tissue-specific differential expression analysis (Yu et al., 2020), structure similarity search by the transported drug (Sakai et al., 2021), interplay analysis among multiple DT variabilities (Wang et al., 2021), functional analysis based on the signaling pathways (Sakil et al., 2017), functional annotation and systematic classification of DTs (Peng et al., 2021), prediction of potential DDIs (Carrascal-Laso et al., 2020), drug safety assessment & toxicity discovery (Zhou et al., 2020), and identification of
potential drug resistance (Hlavac et al., 2020). Overall, these customized database functions are very diverse, which are capable of promoting DT-based research on the drug ADME process.

6. Summary and Prospect

Based on the above discussions, the available databases are useful for translating experimental results into clinical evidence, which can enable clinicians to formulate appropriate medications for a specific patient and provide qualified solutions for drug discovery. Recent studies showed that there is an increasing interest in the variability of DTs, which emphasized the importance of (1) epigenetic regulation and genetic polymorphism of DT, (2) species-, tissue- and disease-specific DT abundances, and (3) exogenous factors modulating DT activity (Yin et al., 2020). These data have been provided by some available databases, such as PharmaGKB (Barbarino et al., 2018), UCSF-FDA TransPortal (Morrissey et al., 2012), and Transformer (Hoffmann et al., 2014), and each database focuses on one particular aspect of DT variability.

Recent studies revealed the urgent necessity of conducting interplay analysis among multiple aspects of DT variability (Chen et al., 2016; Genovese et al., 2017; Ye et al., 2018; Yang et al., 2020b; Yang et al., 2020c). Taking the multidrug resistance as an example, the impaired uptake of organic cation transporter 1 (hOCT1) was found responsible for the chemoresistance of sorafenib in treating the cholangiocarcinoma (CCA). The decrease of the hOCT1 mRNA level was identified to be correlated with the hyper-methylation status of its promoter, and treatment of CCA cells with decitabine (a demethylating agent) was found to be able to restore hOCT1’s expression and increase the uptake of sorafenib (Lozano et al., 2019). This example explicitly demonstrates an interplay between (1) epigenetic regulation of DT and (2) exogenous regulation modulating DT activity.

Similar to hOCT1, organic anion transporter 2 (hOAT2) is another hotspot in current research, and its aberrant expression was reported to lead to insufficient intracellular drug accumulation,
which is responsible for the failure of chemotherapy in the patient with hepatocellular carcinoma (HCC). The transcriptional repression of hOAT2 is associated with histone deacetylations, and the activation of hOAT2’s transcription and enhanced uptake of the OAT2 substrate zidovudine can thus be achieved by histone deacetylases inhibitor Vorinostat (Wang et al., 2021). All in all, the joint consideration of multiple DT variabilities (in this situation, epigenetic regulation and exogenous chemicals) can help to discover potential chemo-sensitization strategies for treating cancers. Such valuable information has been provided in the latest version of VARIDT.

Finally, with the advent of the big-data era, the available pharmaceutical knowledge bases are expected to be fully connected to avoid the problem of “information isolated islands” (Fu et al., 2020). A careful review of all those discussed databases above-identified several reputable databases that were fully connected with other available knowledge bases, such as: ChEMBL, DrugBank, Therapeutic Target Database, PharmGKB, VARIDT, and so on. However, there are still some databases that are not fully linked to other information resources. To promote the publicity of these databases and truly demonstrate the capacity of big-data analysis, the global scientific community should strive together to build a comprehensive database that includes integral information on DTs and their variabilities. This comprehensive database may facilitate the discovery of the correlation between disease/adverse drug reaction and the studied DT.

Authorship Contributions

Participated in research design: Zeng and Zhu. Performed data analysis: Yin, Li and Yu. Wrote or contributed to the writing of the manuscript: Zeng, Zhu and Yin.

Footnotes
This work was funded by National Natural Science Foundation of China (81973394) & Natural Science Foundation of Zhejiang Province (LR21H300001). Supported by the Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Alibaba Cloud, and Information Technology Center of Zhejiang University.

Conflict of interest statement. No author has an actual or perceived conflict of interest with the contents of this article.
References


Genovese I, Ilari A, Assaraf YG, Fazi F, and Colotti G (2017) Not only P-glycoprotein: amplification of the ABCB1-containing chromosome region 7q21 confers multidrug resistance upon cancer cells by coordinated overexpression of an assortment of


Lewis SS and Girisha KM (2020) Whole exome sequencing identifies a novel pathogenic variation [p.(Gly194valfs*7)] in SLC45A2 in the homozygous state in multiple members of a family with oculocutaneous albinism in southern India. *Clin Exp Dermatol*


solute carrier. Drug Metab Dispos 45:228-236.


Yee SW, Brackman DJ, Ennis EA, Sugiyama Y, Kamdem LK, Blanchard R, Galetin A, Zhang


**Figure captions**

**Figure 1.** Ten representative databases providing the information of drugs together with the corresponding DTs. Five types of DT-related data were shown in the circles with various colors, including 3D structure, endogenous substrates, expression/distribution/function, polymorphism, and epigenetic-related data.

**Figure 2.** Nine representative databases describing the information of transporters and transporter families. Four types of transporter-related data were shown in the circles with various colors, including endogenous substrates, expression/distribution/function, structure, and polymorphism-related data.

**Figure 3.** Ten representative databases providing transporter information as part of their data collections. These databases included the transporter information of three types which were shown by the circles of different colors, including expression/distribution/function data, polymorphism data, and 3D structure.
Table 1. Summary of the data completeness of DT and its corresponding drugs, data availability, developer, updating frequency of the databases included in this review.

<table>
<thead>
<tr>
<th>Database</th>
<th>Year of Latest / First Release</th>
<th>No. of DTs (Drugs)</th>
<th>Developer</th>
<th>Updating Frequency</th>
<th>Data Batch Download</th>
<th>Official Website of the Database</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Databases with Its First Version Published in Recent Five Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCA4 database</td>
<td>2017 / 2017</td>
<td>1 (0)</td>
<td>Jack Brockhoff Foundation</td>
<td>(first version)</td>
<td>NO</td>
<td><a href="http://www.sbl.unisi.it/abca4/">http://www.sbl.unisi.it/abca4/</a></td>
</tr>
<tr>
<td>iMusta4SLC*</td>
<td>2018 / 2018</td>
<td>~220 (0)</td>
<td>BINDS</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="http://cib.cf.ocha.ac.jp/slc/">http://cib.cf.ocha.ac.jp/slc/</a></td>
</tr>
<tr>
<td>PPTdb*</td>
<td>2019 / 2019</td>
<td>~80 (0)</td>
<td>Chang Gung University</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="http://pptdb.cgu.edu.tw">http://pptdb.cgu.edu.tw</a></td>
</tr>
<tr>
<td>VARIDT</td>
<td>2020 / 2020</td>
<td>266 (886)</td>
<td>Zhejiang University</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="https://idrblab.org/varidt/">https://idrblab.org/varidt/</a></td>
</tr>
<tr>
<td><strong>Databases Published before and Keeping Update in Recent Five Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCMdb</td>
<td>2017 / 2012</td>
<td>36 (0)</td>
<td>Hungarian Academy of Science</td>
<td>5 Years</td>
<td>NO</td>
<td><a href="http://abcm2.hegelab.org/">http://abcm2.hegelab.org/</a></td>
</tr>
<tr>
<td>ChEMBL</td>
<td>2019 / 2012</td>
<td>124 (~800)</td>
<td>EMBL-EBI</td>
<td>2 Years</td>
<td>YES</td>
<td><a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a></td>
</tr>
<tr>
<td>DrugBank</td>
<td>2018 / 2006</td>
<td>136 (~800)</td>
<td>Genome Alberta</td>
<td>2 Years</td>
<td>YES (need registration)</td>
<td><a href="https://go.drugbank.com/">https://go.drugbank.com/</a></td>
</tr>
<tr>
<td>EBI Expression Atlas</td>
<td>2020 / 2010</td>
<td>~250 (0)</td>
<td>EMBL-EBI</td>
<td>4 Years</td>
<td>YES</td>
<td><a href="https://www.ebi.ac.uk/gxa/">https://www.ebi.ac.uk/gxa/</a></td>
</tr>
<tr>
<td>Database</td>
<td>Published</td>
<td>Entries</td>
<td>Institution</td>
<td>Life Span</td>
<td>Frequency</td>
<td>Registration Needed</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>---------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>FINDbase</td>
<td>2020 / 2007</td>
<td>25 (0)</td>
<td>GoldenHelix Fundtion</td>
<td>3 Years</td>
<td>NO</td>
<td><a href="http://www.findbase.org">http://www.findbase.org</a></td>
</tr>
<tr>
<td>IUPHAR/BPS</td>
<td>2020 / 2009</td>
<td>22 (~70)</td>
<td>NC-IUPHAR</td>
<td>2 Years</td>
<td>YES</td>
<td><a href="https://www.guidetopharmacology.org/">https://www.guidetopharmacology.org/</a></td>
</tr>
<tr>
<td>KEGG DRUG</td>
<td>2019 / 1999</td>
<td>14 (~150)</td>
<td>Kyoto University</td>
<td>1 Year</td>
<td>NO</td>
<td><a href="https://www.kegg.jp/kegg/drug/">https://www.kegg.jp/kegg/drug/</a></td>
</tr>
<tr>
<td>OMIM</td>
<td>2019 / 1995</td>
<td>~200 (0)</td>
<td>NHGRI</td>
<td>1 Year</td>
<td>YES (need registration)</td>
<td><a href="https://omim.org/">https://omim.org/</a></td>
</tr>
<tr>
<td>PDB</td>
<td>2021 / 2000</td>
<td>~60 (~20)</td>
<td>NSF/NIH</td>
<td>1 Year</td>
<td>YES</td>
<td><a href="https://www.rcsb.org/">https://www.rcsb.org/</a></td>
</tr>
<tr>
<td>PharmGKB</td>
<td>2018 / 2002</td>
<td>~100 (~800)</td>
<td>NIH/NHGRI/NICHD</td>
<td>6 Years</td>
<td>YES</td>
<td><a href="https://www.pharmgkb.org/">https://www.pharmgkb.org/</a></td>
</tr>
<tr>
<td>MemProtMD*</td>
<td>2019 / 2015</td>
<td>~20 (0)</td>
<td>University of Oxford</td>
<td>4 Years</td>
<td>YES</td>
<td><a href="http://memprotmd.bioch.ox.ac.uk/">http://memprotmd.bioch.ox.ac.uk/</a></td>
</tr>
<tr>
<td>TCDB</td>
<td>2021 / 2006</td>
<td>266 (0)</td>
<td>NIH</td>
<td>5 Years</td>
<td>YES</td>
<td><a href="https://www.tcdb.org/">https://www.tcdb.org/</a></td>
</tr>
<tr>
<td>TransportDB*</td>
<td>2017 / 2004</td>
<td>~250 (0)</td>
<td>Macquarie University</td>
<td>10 Years</td>
<td>YES</td>
<td><a href="http://www.membranetransport.org/">http://www.membranetransport.org/</a></td>
</tr>
<tr>
<td>TTD</td>
<td>2020 / 2002</td>
<td>~100 (~700)</td>
<td>Zhejiang University</td>
<td>2 Years</td>
<td>YES</td>
<td><a href="https://idrblab.org/ttd/">https://idrblab.org/ttd/</a></td>
</tr>
<tr>
<td>UniProt</td>
<td>2021 / 2004</td>
<td>266 (~100)</td>
<td>NIH</td>
<td>1 Year</td>
<td>YES</td>
<td><a href="https://www.uniprot.org/">https://www.uniprot.org/</a></td>
</tr>
<tr>
<td><strong>Databases Published before 2016 and without Any Update in Recent Five Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALD Info</td>
<td>2001 / 2001</td>
<td>1 (0)</td>
<td>University of Amsterdam</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="https://adrenoleukodystrophy.info/">https://adrenoleukodystrophy.info/</a></td>
</tr>
<tr>
<td>Database</td>
<td>Development Date</td>
<td>Data Size</td>
<td>Organization</td>
<td>Version</td>
<td>Manual/Computational</td>
<td>Website</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>------------</td>
<td>--------------------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CFMD</td>
<td>2011 / 2011</td>
<td>1 (0)</td>
<td>US CF Foundation</td>
<td>(first version)</td>
<td>NO</td>
<td><a href="http://www.genet.sickkids.on.ca/cftr/">http://www.genet.sickkids.on.ca/cftr/</a></td>
</tr>
<tr>
<td>dbSNP</td>
<td>2001 / 1999</td>
<td>~250 (0)</td>
<td>NLM</td>
<td>1 Year</td>
<td>YES</td>
<td><a href="https://www.ncbi.nlm.nih.gov/snp/">https://www.ncbi.nlm.nih.gov/snp/</a></td>
</tr>
<tr>
<td>Metrabase</td>
<td>2015 / 2015</td>
<td>20 (~500)</td>
<td>University of Cambridge</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="http://www-metrabase.ch.cam.ac.uk/">http://www-metrabase.ch.cam.ac.uk/</a></td>
</tr>
<tr>
<td>METscout</td>
<td>2013 / 2013</td>
<td>~200 (0)</td>
<td>Max Planck Society</td>
<td>(first version)</td>
<td>NO</td>
<td><a href="http://metscout.mpg.de/">http://metscout.mpg.de/</a></td>
</tr>
<tr>
<td>SLC TABLES</td>
<td>2013 / 2013</td>
<td>~220 (0)</td>
<td>University of Bern</td>
<td>(first version)</td>
<td>NO</td>
<td><a href="http://slc.bioparadigms.org/">http://slc.bioparadigms.org/</a></td>
</tr>
<tr>
<td>The Human Protein Atla</td>
<td>2015 / 2015</td>
<td>~250 (0)</td>
<td>Knut &amp; Alice Foundation</td>
<td>(first version)</td>
<td>YES</td>
<td><a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a></td>
</tr>
<tr>
<td>Transformer</td>
<td>2014 / 2010</td>
<td>60 (~250)</td>
<td>Universitätsmedizin Berlin</td>
<td>4 Years</td>
<td>NO</td>
<td><a href="http://bioinformatics.charite.de/transformer">http://bioinformatics.charite.de/transformer</a></td>
</tr>
<tr>
<td>UCSF-FDA TransPortal</td>
<td>2012 / 2012</td>
<td>31 (~480)</td>
<td>FDA Critical Path Initiative</td>
<td>(first version)</td>
<td>NO</td>
<td><a href="https://transportal.compbio.ucsf.edu/">https://transportal.compbio.ucsf.edu/</a></td>
</tr>
</tbody>
</table>

The “*” in the first column indicated that the data provided in the corresponding databases were primarily based on computational calculation or simulation, while the databases without “*” represented that their data were manually collected from scientific literatures that were based on experimental validations.

**Abbreviations:**

**BINDS**: Basis for supporting innovative drug discovery and life science research; **EMBL-EBI**: European bioinformatics institute; **NC-IUPHAR**: Nomenclature and standards committee of international union of clinical pharmacology; **NHGRI**: National human genome research institute; **NIH**: National institutes of health; **NICHD**: National institute of child health and human development; **NINDS**: National institute of neurological disorders and stroke; **NLM**: National library of medicine; **NSF**: National science foundation; **US CF Foundation**: Cystic fibrosis foundation.
Table 2. The availability of DT-related data types and groups among those ten databases (shown in Figure 1) that provided the information of drugs and their corresponding transporters (DTs).

<table>
<thead>
<tr>
<th></th>
<th>ChEMBL</th>
<th>DrugBank</th>
<th>TTD</th>
<th>IUPHAR BPS</th>
<th>PharmGKB</th>
<th>KEGG</th>
<th>UCSF-FDA TransPortal</th>
<th>TransFormer</th>
<th>Metabase</th>
<th>VARIDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Varied Expression</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Organism Specific Abundance</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Protein Functional Family</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Protein Sequence Information</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Tissue Differential Distribution</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cytogenetic Location Information</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Genetically Correlated Phenotype</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Genetic Variant &amp; Frequency</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Data Type</td>
<td>Genetically Induced Disease</td>
<td>Species-Specific Evolution</td>
<td>Whole Protein Structure</td>
<td>DNA/histone Methylation Acetylation</td>
<td>Non-coding RNA Regulation</td>
<td>Clinical Drug-drug Interaction</td>
<td>Exogenous Substances of DTs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td>-------------------------</td>
<td>-------------------------------------</td>
<td>---------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure-related Data of DTs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epigenetic Data of DTs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exogenous Substances of DTs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'√' indicated that the data type is available, while '−' denotes that the data type is not available.
Table 3. The availability of transporter-related data types and groups among those nine databases (shown in Figure 2) that described the information of transporters and transporter families.

<table>
<thead>
<tr>
<th>Expression, Distribution, and Function of DTs</th>
<th>TCDB</th>
<th>SLC TABLES</th>
<th>TransportDB</th>
<th>METscout</th>
<th>CFMD</th>
<th>ALD Info</th>
<th>ABCA4 Database</th>
<th>ABCMdb</th>
<th>iMusta4SLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Varied Expression</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
</tr>
<tr>
<td>Organism Specific Abundance</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Protein Functional Family</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Protein Sequence Information</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
</tr>
<tr>
<td>Phylogenetic Classification System</td>
<td>√</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Tissue Differential Distribution</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>scRNA Sequencing Atlas</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Genetic Polymorphism Data of DTs</td>
<td>Genetically Correlated Phenotype</td>
<td>Genetically Induced Disease</td>
<td>Genetically Variant &amp; Frequency</td>
<td>Species-Specific Evolution</td>
<td>Whole Protein Structure</td>
<td>Substrate Functional Family</td>
<td>Substrate Structure Information</td>
<td>Substrate Transporting Mechanism</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>Cytogenetic Location Information</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Genetic Variant &amp; Frequency</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Genetically Induced Disease</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>√</td>
<td>–</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

'√' indicated that the data type is available, while ‘−’ denotes that the data type is not available.
Table 4. The availability of data types and groups among the ten databases (shown in Figure 3) that provided the transporter-related information as part of their data collections.

<table>
<thead>
<tr>
<th></th>
<th>dbSNP</th>
<th>FINDbase</th>
<th>OMIM</th>
<th>UniProt</th>
<th>The Human Protein Atlas</th>
<th>EBI Expression Atlas</th>
<th>PDB</th>
<th>EncoMPASS</th>
<th>PPTdb</th>
<th>MemProt MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Varied Expression</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Organism Specific Abundance</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Protein Functional Family</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Protein Sequence Information</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Tissue Differential Distribution</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>scRNA Sequencing Atlas</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Polymorphism Data of DTs</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cytogenetic Location Information</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Genetically Correlated Phenotype</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Genetic Variant &amp; Frequency</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Genetically Induced Disease</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Functional Conserved Domains</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Species-Specific Evolution</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Substrate Binding Domain</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Whole Protein Structure</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

'✓' indicated that the data type is available, while '–' denotes that the data type is not available.
Table 5. The customized functions of all 29 databases discussed in this study and their representative applications. These functions fall into three classes: facilitating the rational use of drugs, discovering the potential therapeutic targets, and developing the new strategy for disease treatment.

<table>
<thead>
<tr>
<th>Class of Function</th>
<th>Customized Function of Each Analyzed Database</th>
<th>Typical Database(s)</th>
<th>Representative Applications of These Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovering the Potential Therapeutic Targets</td>
<td>Structure-based Drug Design or Identification</td>
<td>PDB, PPTdb, MemProt MD</td>
<td>PDB database was used to identify a novel AQP4 inhibitor binding deep inside this transporter based on the molecular dynamics using a high-resolution crystal AQP4 structure (Yu et al., 2016).</td>
</tr>
<tr>
<td></td>
<td>Sequence-based Discovery of Target Druggability</td>
<td>TransportDB, TTD, TCDB</td>
<td>TransportDB database was adopted for predicting transporters from the genome and providing a breakthrough for the functional annotation of a large number of transporters (Frioux et al., 2020).</td>
</tr>
<tr>
<td></td>
<td>Disease-specific Differential Expression Analysis</td>
<td>Human Protein Atlas, EBI Expression Atlas, VARIDT</td>
<td>Human Protein Atlas was used to extract the expression pattern of SLC16A1 and SLC16A3 for their clinical potential applications in the treatment of pancreatic adenocarcinoma (Yu et al., 2020).</td>
</tr>
<tr>
<td></td>
<td>Structure Similarity Search by Transported Drugs</td>
<td>ChEMBL, TTD, DrugBank</td>
<td>ChEMBL database was used to identify a new inhibitor of serotonin transporter with comparable affinity to the commercial drug by structure similarity search and virtual screening (Sakai et al., 2021).</td>
</tr>
<tr>
<td>Developing the New Strategy for Disease Treatment</td>
<td>Interplay Analysis among Multiple DT Variabilities</td>
<td>VARIDT</td>
<td>VARIDT database was used to facilitate the interplay analysis of OAT2 in hepatocellular carcinoma between its disease-specific differential expression and histone acetylation (Wang et al., 2021).</td>
</tr>
<tr>
<td></td>
<td>Functional Analysis Based on Signaling Pathways</td>
<td>KEGG, PharmGKB</td>
<td>KEGG database was applied to identify the key transporter pathways involving in the development of breast cancer (Sakil et al., 2017) and the microgravity effects in epidermal stem cells (Li et al., 2020a).</td>
</tr>
<tr>
<td></td>
<td>Functional Annotation and Systematic Classification</td>
<td>TCDB</td>
<td>TCDB database was adopted to facilitate the functional annotation and systematic classification of DT using its transporter automatic annotation pipeline (Graf et al., 2021; Peng et al., 2021).</td>
</tr>
</tbody>
</table>
Facilitating the Rational Use of Drugs

Prediction of DT-based Potential DDI

Transformer database was adopted to predict the potential DDIs for reducing the costs in novel drug development and optimizing the process of rational drug design (Carrascal-Laso et al., 2020).

Drug Safety Assessment and Toxicity Prediction

VARIDT was applied to reveal the biological mechanism of bile acids efflux using the tissue-specific expression of two subunits of organic solute transporter in ileum (Zhou et al., 2020).

Drug Resistance Identification of Potential OMMIM

PharmGKB database was used to predict the response of drugs in cancer treatment based on the pharmacogenomic analysis focusing on ATP binding cassette transporters (Hlavac et al., 2020).

Drug Resistance Potential DDI UCSF-FDA TransPortal

PharmGKB database was used to predict the response of drugs in cancer treatment based on the pharmacogenomic analysis focusing on ATP binding cassette transporters (Hlavac et al., 2020).

Drug Resistance Potential DDI UCSF-FDA TransPortal

PharmGKB database was used to predict the response of drugs in cancer treatment based on the pharmacogenomic analysis focusing on ATP binding cassette transporters (Hlavac et al., 2020).

Drug Resistance Potential DDI UCSF-FDA TransPortal

PharmGKB database was used to predict the response of drugs in cancer treatment based on the pharmacogenomic analysis focusing on ATP binding cassette transporters (Hlavac et al., 2020).

Drug Resistance Potential DDI UCSF-FDA TransPortal

PharmGKB database was used to predict the response of drugs in cancer treatment based on the pharmacogenomic analysis focusing on ATP binding cassette transporters (Hlavac et al., 2020).
Figure 1
Figure 2
Figure 3