



Contents lists available at ScienceDirect

## Methods

journal homepage: [www.elsevier.com/locate/ymeth](http://www.elsevier.com/locate/ymeth)

# Improving drug safety: From adverse drug reaction knowledge discovery to clinical implementation

Yuxiang Tan<sup>a,1</sup>, Yong Hu<sup>a,1</sup>, Xiaoxiao Liu<sup>a,1</sup>, Zhinan Yin<sup>a</sup>, Xue-wen Chen<sup>b</sup>, Mei Liu<sup>c,\*</sup><sup>a</sup> Big Data Decision Institute, The First Affiliated Hospital, International Immunology Center, The Biomedical Translational Research Institute, Jinan University, Guangzhou, Guangdong, China<sup>b</sup> Department of Computer Science, Wayne State University, Detroit, USA<sup>c</sup> Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, USA

## ARTICLE INFO

## Article history:

Received 15 April 2016

Received in revised form 13 July 2016

Accepted 30 July 2016

Available online xxx

## Keywords:

Drug–target interaction

Drug safety

Adverse drug reaction

Pharmacogenomic

## ABSTRACT

Adverse drug reactions (ADRs) are a major public health concern, causing over 100,000 fatalities in the United States every year with an annual cost of \$136 billion. Early detection and accurate prediction of ADRs is thus vital for drug development and patient safety. Multiple scientific disciplines, namely pharmacology, pharmacovigilance, and pharmacoinformatics, have been addressing the ADR problem from different perspectives. With the same goal of improving drug safety, this article summarizes and links the research efforts in the multiple disciplines into a single framework from comprehensive understanding of the interactions between drugs and biological system and the identification of genetic and phenotypic predispositions of patients susceptible to higher ADR risks and finally to the current state of implementation of medication-related decision support systems. We start by describing available computational resources for building drug–target interaction networks with biological annotations, which provides a fundamental knowledge for ADR prediction. Databases are classified by functions to help users in selection. Post-marketing surveillance is then introduced where data-driven approach can not only enhance the prediction accuracy of ADRs but also enables the discovery of genetic and phenotypic risk factors of ADRs. Understanding genetic risk factors for ADR requires well organized patient genetics information and analysis by pharmacogenomic approaches. Finally, current state of clinical decision support systems is presented and described how clinicians can be assisted with the integrated knowledgebase to minimize the risk of ADR. This review ends with a discussion of existing challenges in each of disciplines with potential solutions and future directions.

© 2016 Elsevier Inc. All rights reserved.

## Contents

1. Introduction	00
2. Deriving pharmacological knowledge of ADR from drug–target interactions	00
2.1. Existing databases for building drug–target interactions	00
2.2. Predicting drug–target interaction network with biological annotations	00
2.3. Predicting ADR from drug–target interaction network	00
3. Enhancing ADR knowledge with post-marketing surveillance	00
3.1. Detecting drug–event associations – data driven approach	00
3.2. Identifying genetic risk factors of ADR – pharmacogenomic approach	00
3.2.1. Candidate gene approach	00
3.2.2. Genome-wide association studies	00
3.2.3. Next-generation sequencing	00
3.2.4. Comparison of the approaches	00

\* Corresponding author.

E-mail address: [meiliu@kumc.edu](mailto:meiliu@kumc.edu) (M. Liu).<sup>1</sup> These authors contributed equally to this work.<http://dx.doi.org/10.1016/j.ymeth.2016.07.023>

1046-2023/© 2016 Elsevier Inc. All rights reserved.

3.3.	Phenotypic risk factors of ADR .....	00
3.3.1.	Age .....	00
3.3.2.	Kidney and liver function .....	00
3.3.3.	Comorbidity and polypharmacy .....	00
4.	ADR prevention in healthcare with clinical decision support systems .....	00
4.1.	Basic medication-related decision support .....	00
4.1.1.	Drug-allergy checking .....	00
4.1.2.	Basic dosing guidance .....	00
4.1.3.	Drug-drug interaction checking .....	00
4.2.	Advanced medication-related decision support .....	00
4.2.1.	Advanced dosing guidance .....	00
4.2.2.	Advanced guidance for medication-associated laboratory testing .....	00
4.2.3.	Advanced drug-disease interaction checking .....	00
5.	Challenges and future research directions .....	00
5.1.	Challenges in predicting ADRs using drug-target interaction networks .....	00
5.2.	Future of predicting ADRs using drug-target interaction networks .....	00
5.3.	Challenges in pharmacovigilance .....	00
5.4.	Future of pharmacovigilance .....	00
5.5.	Challenges in pharmacogenomics .....	00
5.6.	Future of pharmacogenomics .....	00
5.7.	Challenges in medication-related CDSS .....	00
5.8.	Future of CDSS .....	00
	Acknowledgements .....	00
	References .....	00

## 1. Introduction

Clinical benefits of prescription drugs do not only depend on their efficacy in treating diseases but also their safety and tolerability in patients. Each dispensed prescription carries its own risks for causing adverse drug reactions (ADRs), ranging the full spectrum of severity from cosmetic to severe morbidity and mortality [1,2]. ADR is estimated to cause over 2 million hospitalizations and more than 100,000 fatalities each year in the United States alone [3,4], with an estimated annual cost of \$136 billion [5,6].

Between 1976 and 2005, severe ADRs have caused 28 drugs to be withdrawn from the United States market [7], with the top drug-induced toxicities being hepatotoxicity (21%, 6), nephrotoxicity (7%, 2), cardiotoxicity (7%, 2), torsades (21%, 6), and rhabdomyolysis (7%, 2). For example, Vioxx is approved by the Food and Drug

Administration (FDA) in 1999 and gained widespread acceptance among physicians (prescribed to over 80 million people worldwide) in treating patients with arthritis and other conditions causing chronic or acute pain. Five years later, it was pulled off the market due to significantly increased risk of heart attack and stroke, becoming one of the most widely used drugs ever to be withdrawn [8,9].

Thus, early identification as well as precise prediction of ADRs is crucial for drug discovery and development and patient safety. ADRs are often classified as Type A and Type B where Type A reactions are typically dose-related, expressing an extended therapeutic effect of a drug, for example, hypotension with anti-hypertensive therapy and bleeding events with warfarin; Type B reactions are 'idiosyncratic', occurring only in susceptible individuals [10]. The etiology of variable drug responses is multifactorial, including both genetic (e.g., severe haemolytic anaemia can occur

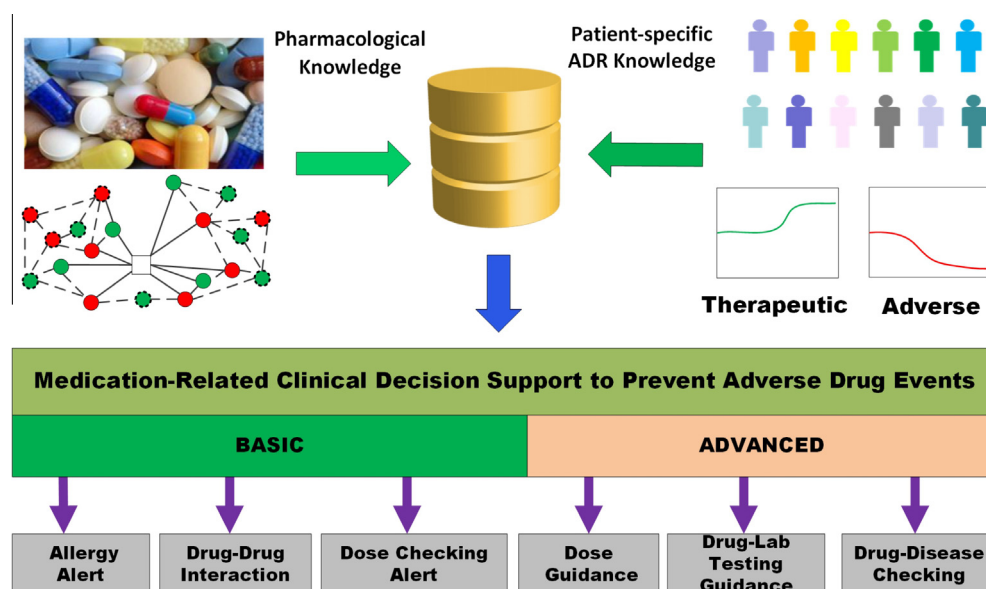


Fig. 1. Overview of the effort in ADR prevention from knowledgebase construction to clinical implementation.

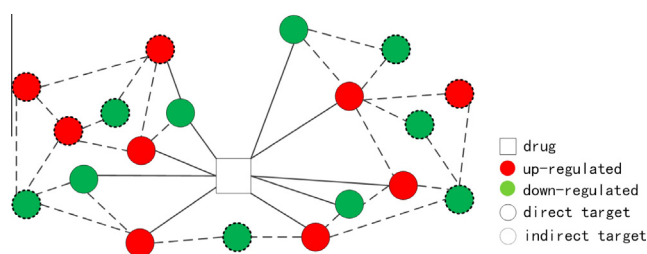
in glucose-6-phosphate dehydrogenase (G6PD) deficient patients after the ingestion of fava beans and the use of certain anti-malaria drugs of sulfonamides [11]) and phenotypic (e.g., age, race, comorbidities) risk factors. Better understanding of the ADR risk factors, particularly for Type B reactions, could significantly improve the drug development process and decrease the healthcare cost through precision medication delivery to the right patients with the right dose at the right time.

Patient safety has always been a top priority of regulatory agencies and healthcare industries. The US FDA has established multiple departments and checkpoint systems such as the Center for Drug Evaluation and Research (CDER), the New Drug Application (NDA), and Post-marketing Surveillance (PMS) program to assess and monitor drug safety risks at each stage of the drug discovery and development process and throughout its market life. In the healthcare sector, hospitals have implemented electronic medical records (EMRs) with computerized medication-related clinical decision support systems to improve patient safety and quality of care.

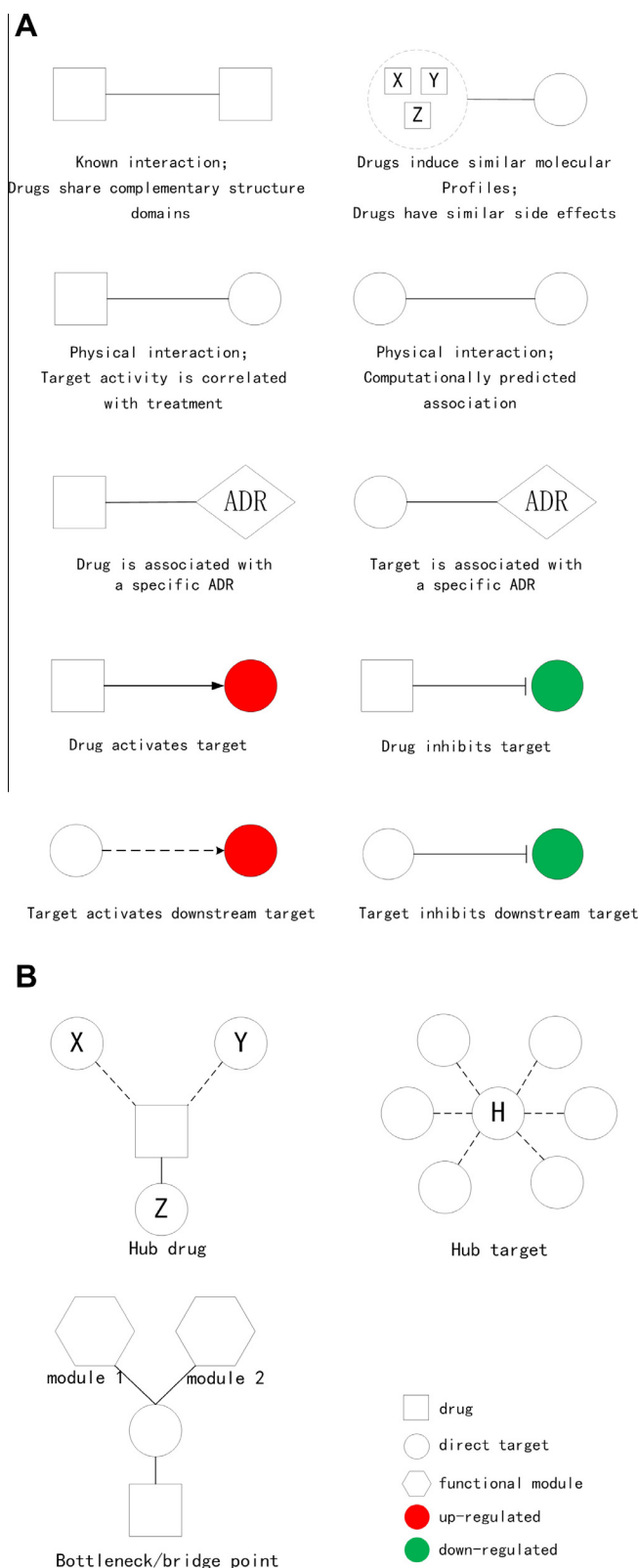
This paper summarizes the current research efforts in preventing ADR from knowledge discovery through a comprehensive understanding of the interactions between drugs and biological systems (Section 2) and the identification of genetic and phenotypic predispositions of patients susceptible to higher ADR risks (Section 3) to knowledge delivery in healthcare systems (Section 4) through implementation of medication-related decision support systems (Fig. 1). Pharmacology, pharmacovigilance, and pharmacoinformatics seem to be unrelated scientific disciplines, but they share a core goal of improving patient medication safety. Cooperation between the disciplines will accelerate knowledge discovery and knowledge translation to clinical practice to achieve real impact in healthcare.

## 2. Deriving pharmacological knowledge of ADR from drug-target interactions

Many risk factors may contribute to ADRs, and one of the most important factors is the interaction between drug and its targets, which include the therapeutic targets and/or other cellular macro-molecules [12]. Drugs function through interacting with molecular targets. Although designed to bind specific targets, unexpected binding of drugs to other proteins, namely off-target effects, is also frequently observed in drug discovery and clinical practice, resulting in ADRs [13]. Moreover, besides the desired pathway, inhibition of a target may also affect other pathways that the target is involved in, and generate ADRs [14]. As a result, studying drug-target interaction networks can provide better understanding of the polypharmacology of drugs (Figs. 2 and 3A).



**Fig. 2.** Drug-target interaction network model. Square stands for the drug as the center of this network. Nodes represents target molecules with red means up-regulated and green means down-regulated. Edges are the physical interactions or other biological associations. Solid-line means direct interaction with the drug, while dashed-line means interaction is indirect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Key components of drug-target interaction networks. (A) Examples of key types of single interactions: their visualization and sources. Red means a target is up-regulated while green means down-regulated. (B) Interaction-module examples: (1) Hub drug with multiple targets, which is easier to have ADRs; (2) Hub target with multiple downstream targets. If a drug uses it as a target, the chance of having ADRs will increase; (3) Bottleneck target is the bridge to connect functional modules. Using it as a drug target can disturb the whole function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Information extracted from drug–target interaction networks can be translated into the knowledge of ADRs. Henceforth, it will facilitate the process of drug discovery in the early stage, reduce the cost of *in vitro* and *in vivo* experiments, and ultimately benefit patients who are under various treatments.

### 2.1. Existing databases for building drug–target interactions

The most common drug targets are proteins, however, some nucleic acids can also serve as drug targets [15]. Advances in whole genome sequencing and the mass spectrometry characterizations of human proteome provide new opportunities and make the identification of drug–target interactions more achievable.

Currently, several databases of interaction networks of chemicals and proteins are available (Table 1). For instance, DrugBank is the most comprehensive bioinformatics/cheminformatics database that comprises detailed drug and target information [16]. It is applied in drug target discovery, drug design, docking, metabolism prediction, and interaction prediction. STITCH (v4.0), which stands for “search tool for interactions of chemicals”, is another commonly used database. It integrates information about structures of targets, binding experiments and drug–target relationships, which are derived from experiments, databases and literature. It is an informative resource for the interactions of drug-like chemicals and proteins [17,18]. ChEMBL [19] is also a

well-established database for getting compound information. It contains binding, functional and absorption, distribution, metabolism, and excretion – toxicity (ADMET) information for millions of drug-like bioactive compounds and thousands of biologics. Additionally, PubChem Compound Database [20] covers validated chemical depiction information, with specialization in pre-cluster and cross-reference by identity and similarity. Other similar drug–drug interaction (DDI) databases include Manually Annotated Targets and Drugs Online Resource (MATADOR) [21], Psychoactive Drug Screening Program (PDSP) [22] and IntAct [23]. Methods used for constructing these databases are summarized in Table 1, and most databases integrate information from others.

Since gene expression profiling has been routinely performed due to the significant price-drop, some databases, such as ConnectivityMap (cmap v02) [24] and Mode of Action by NeTwoRk Analysis (MANTRA) [25], exploit the gene expression changes upon treatment to build the drug–target network and group compounds with similar profile together. At present, a microarray platform, called L1000 technology, examines only 1000 genes to represent the whole transcriptome, which reduces the price per sample to only a few dollars. With the low cost, tons of compounds are under investigation, and the data is saved in cmap by the LINCS project [26]. Using numerous associations between compounds and expression profiles, researchers can potentially build a more comprehensive drug–target interaction network with indirect edges, which were previously neglected.

**Table 1**  
Databases used in building drug–target interaction network with adverse drug reaction (ADR) knowledge.

Database	Website	Version	Methods
<b>Databases of primary data source</b>			
<i>Molecular</i>			
ChEMBL	<a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a>	v21	E, M,
MATADOR	<a href="http://matador.embl.de">http://matador.embl.de</a>	NA	M,
PDSP	<a href="http://pdsp.med.unc.edu/">http://pdsp.med.unc.edu/</a>	NA	E,
IntAct	<a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a>	NA	E, M
PDB	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>	NA	E,
<i>Gene Expression</i>			
GEO	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	NA	E,
<i>Pathways and gene annotation</i>			
KEGG	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>	NA	M
GO	<a href="http://geneontology.org/page/go-database">http://geneontology.org/page/go-database</a>	NA	M
<i>ADR databases</i>			
FAERS	<a href="https://open.fda.gov/data/faers/">https://open.fda.gov/data/faers/</a>	NA	M
EMA	<a href="http://www.ema.europa.eu">http://www.ema.europa.eu</a>	NA	M
JAPIC	<a href="http://www.japic.or.jp/">http://www.japic.or.jp/</a>	NA	M
VAERS	<a href="https://vaers.hhs.gov/">https://vaers.hhs.gov/</a>	NA	M
<i>Protein–protein interaction databases</i>			
TTD	<a href="http://bidd.nus.edu.sg/group/cjttd/">http://bidd.nus.edu.sg/group/cjttd/</a>	v4.3	D, E
BRENDA	<a href="http://www.brenda-enzymes.org/">http://www.brenda-enzymes.org/</a>	NA	D, E, M
GLIDA	<a href="http://pharminfo.pharm.kyoto-u.ac.jp/services/gliida/">http://pharminfo.pharm.kyoto-u.ac.jp/services/gliida/</a>	v2.04	D, E, M
Pfam	<a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a>	v29.0	D, M, HMM
STRING	<a href="http://string-db.org/">http://string-db.org/</a>	v10	D, PM
MINT	<a href="http://mint.bio.uniroma2.it/mint/">http://mint.bio.uniroma2.it/mint/</a>	NA	D, E, M
HAPPI	<a href="http://bio.informatics.iupui.edu/HAPPI/">http://bio.informatics.iupui.edu/HAPPI/</a>	v2.0	D, O
<i>Gene expression profile network databases</i>			
ConnectivityMap	<a href="http://www.broadinstitute.org/cmap/">http://www.broadinstitute.org/cmap/</a>	v02	E, PM
MANTRA	<a href="http://mantra.tigem.it">http://mantra.tigem.it</a>	NA	E, PM
<i>Well-integrated interaction network databases (including ADR information)</i>			
DrugBank	<a href="http://www.drugbank.ca/">http://www.drugbank.ca/</a>	v4.3	D, M, P, T
STITCH	<a href="http://stitch.embl.de/">http://stitch.embl.de/</a>	v4.0	D, M, T
PubChem	<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>	NA	D, E, P, B
SuperTarget	<a href="http://bioinf-apache.charite.de/supertarget_v2/">http://bioinf-apache.charite.de/supertarget_v2/</a>	v2	D, M
SIDER	<a href="http://sideeffects.embl.de/">http://sideeffects.embl.de/</a>	v4.1	D, M, T
My-DTome	<a href="http://www.my-dtome.lu">www.my-dtome.lu</a>	NA	D, O
DR.PRODIS	<a href="http://cssb.biology.gatech.edu/dr.prodis/">http://cssb.biology.gatech.edu/dr.prodis/</a>	NA	D, FINDSITE <sup>comb</sup>
PharmGKB	<a href="https://www.pharmgkb.org/">https://www.pharmgkb.org/</a>	NA	D, M, T

Methods used to extract information: E: Experiment result; D: Databases integration; M: Manual curation; P: Protein annotation; T: Text mining.

Methods used to construct network: B: Bipartite Network; PM: Pattern-Matching algorithms; HMM: Hidden Markov Models; O: Other network methods.

NA: not available.

Meanwhile, because proteins are the main drug targets, protein-protein interaction (PPI) is an important component of the drug-target network. Thus many protein databases are excellent sources of drug-target interactions. The Protein Data Bank (PDB) [27] is an archive for 3D structures of proteins, nucleic acid, and complex assemblies, and commonly used for docking. Therapeutic Target Database (TTD) [28] focuses on therapeutic proteins and nucleic acids as drug targets. BRENDA [29] is an enzyme database, which is considered as the database of drug target candidates because a lot of drugs target the enzymes. Moreover, Guanosine-binding Protein Coupled Receptor (GPCR) Ligand Database (GLIDA) [30] is an example of databases that specialize in protein superfamilies with a key biological function. Other PPI databases include Pfam [31], Human Annotated and Predicted Protein Interaction (HAPPI) [32], STRING (v10) [33], and MINT [34].

## 2.2. Predicting drug-target interaction network with biological annotations

Information necessary for predicting drug-target interactions network can be derived from validated experimental data, literature search and computational modeling [35]. First, a large amount of data (e.g., from high through-put screening and cell line profiling) is available in public domain. Literature text mining of the Medline database with optimized algorithm can investigate various drug-target interaction problems [36]. Another widely used method is docking simulations, which are based on *in silico* modeling of compound-protein interactions [37]. LaBute et al. suggested that using high performance computers, molecular docking provided reliable prediction of interaction network of a drug candidate, which can be used to identify clinically important ADRs [38]. In their study, ADRs were predicted by combining the results of molecular docking (VinaLC) and the known ADR information from DrugBank and SIDER [39], based on L1-regularized logistic regression models. However, the drawback of this approach is the demand of long computation time and the low accuracy of molecular docking. Structures of targets alone provide a poor docking platform, and only high-resolution drug-target complex structures can be used in precise docking. However, the high quality complex structures are inadequate, particularly for GPCRs, the most common drug targets. Skolnick's group developed threading/structure-based FINDSITE-based approaches, which is feasible for searching the entire proteomes against many compounds [40]. This approach does not require known interactions for a drug or protein target nor high-resolution protein structures. It is more efficient than traditional docking methods, and most importantly, has better accuracy for ranking drug-target interactions than traditional docking methods. It is worth to mention that, most drug-target interaction network are predicted by various mathematical models or machine learning methods. However, although the number of validated interactions is increasing, there is no reliable negative control. As a matter of fact, it is becoming critical to build a systematic method of screening drug and target with no interaction at all [41].

Despite the fact that drug-target interaction network can be predicted, to derive further biological insights, studies must include biological databases that contain information about pathways, ontology terms and expression profiles. For example, Kyoto Encyclopedia of Genes and Genomes (KEGG) [42] is a extensively used database including information of genes, pathways and diseases, etc. The Gene Ontology (GO) [43] database annotates genes with GO terms, which are used to group target genes by functional similarities. Similar to MATADOR, but with extensive information such as pathways, ontologies and Cyp450s, SuperTarget [21] integrates drug related annotations for medical indication. Another valuable database is Gene expression Omnibus (GEO), a repository

for high-throughput data (microarray and next-generation sequencing) on functional genomic, which can be used for preclinical validation of drug candidates [44].

## 2.3. Predicting ADR from drug-target interaction network

Although drug-target interaction network can be used to generate multi-dimensional knowledge such as novel drug target discovery or drug repositioning, this paper focuses on ADR prediction. Human body is a complex system, in which a myriad of biological interactions entangle into a network. Actually, hitting one node in the network by a drug can cause diverse reactions including ADRs. Due to the development of computational-biomedical technologies (analytic methods, computational power and related databases, etc.), researchers can now extend existing knowledge to represent this network by adding edges between nodes and to predict more and more reactions. Generally, ADR are caused by the interaction of undesired targets. This is likely to happen when a drug is the hub drug or it targets at a hub target (Fig. 3B).

To reliably predict ADR, it requires well-annotated databases to connect drugs to side effects. One of the key players is SIDER [45] (v4.1), a database containing marketed medicines and their recorded ADRs. It is characterized by the information of side effect frequency, classifications of drug and side effects. Other governmental side-effect databases include FDA Adverse Event Reporting System (FAERS) [46], European Medicines Agency (EMA) [47] and the Japan Pharmaceutical Information Center (JAPIC) [48] database. Tatonetti et al. [49] reported an improved performance of their off-label and DDI side-effect database (OFFSIDES and TWOSIDES), which is reported to be superior to SIDER. Using the databases, they predicted adverse cardiovascular events from co-prescription of thiazides and selective serotonin reuptake inhibitors (SSRIs). Moreover, there are specialized drug-ADR databases such as Vaccine Adverse Event Reporting System (VAERS) [50], a specialized database focusing on vaccine ADRs, and Myocardial Infarction Drug-Target Interactome network (My-DTome) [47], a database specifically aimed at myocardial infarction (MI). Recently, well-integrated database across multiple -omics has attracted much interest. DR.PRODIS integrated DRugome, PROteome, DISeaseome into a webserver, providing querying service as well as virtual target screening of new compounds with side-effect predictions [51]. The Pharmacogenetics Knowledge Base (PharmGKB) [52] is another example of integrated database with genomic, phenotypic and clinical information collected from ongoing pharmacogenetic studies. It has a knowledge pyramid showing what information of different levels is and how it is collected and to be used.

Based on these resources and the structure of the interaction network, researchers can predict ADRs by examining (1) structure similarities of compounds and proteins; (2) drug-drug interaction (DDI). The first class of methods is mainly used for predicting ADRs of new drugs according to the existing ADRs of drugs, while the second class requires comprehensive drug-target interaction networks. Nowadays, the second class of methods is attracting more attentions because of the popularity of "cocktail" therapies (co-prescriptions).

In the first class of methods, by using SIDER and drug structure information, Liu et al. [53] developed a causality analysis model based on structure learning (CASTLE) to identify contributing factors of ADRs, which was critical because it showed causal relations. They also showed that the CASTLE model performed better than baseline support vector machine (SVM) trained on full and randomly selected feature set and against LASSO logistic regression. Furthermore, by integrating the phenotypic characteristics of drugs and ADRs from SIDER, the same group successfully predicted



ADRs associated with the withdrawal of *rofecoxib* and *cerivastatin*. Huang and coworkers [54] reported a feature selection approach using both the drug structure and protein-protein interaction (PPI) information, as well as SIDER that accurately predicted ADRs of 25 out of 27 withdrawn drugs. In the myocardial infarction field, with the help of My-DTome, cardiovascular ADRs of non-cardiovascular drugs have been identified, and some of them have been investigated by *in vitro/vivo* animal models and even post-marketing studies [47,55].

In the second class of methods, Hennessy and Flockhart [56] raised attentions and questions about ADRs caused by drug-drug interaction. Generally, if co-dosed drugs have no interaction, the side effects will be just their own ones. However, if drugs have synergistic and/or antagonistic effects, using them together may cause serious, undesirable effects, and one of the reasons is the change of dose effect.

Although most ADR prediction studies used drug-target networks, there are also reports that did not use any of them. Kim et al. [57] presented an early example of a cohort-oriented study on drug-disease association using only their record of treatments for chronic heart failure. Recently, by using commercial databases utilized by hospitals, Cami et al. [58] applied predictive pharmacosafety networks (PPNs) in predicting likely unknown ADRs of known drugs through building drug-ADR association directly across medical domains. This method showed low sensitivity but high specificity, which is a reliable way to identify unknown ADRs of existing drugs.

### 3. Enhancing ADR knowledge with post-marketing surveillance

Pharmaceutical manufacturers of a new drug, in their own and in FDA-regulated pre-marketing trials, typically conduct testing on no more than a few thousands of volunteers, many of whom are healthier than the patients receiving the drug post-approval. For any potentially rare ADR with the occurrence rate less than 0.1%, it will be extremely difficult for the pre-marketing trials to identify the related ADRs [59]. Moreover, the highest risk patients (e.g., the most ill or debilitated) are normally excluded from these trials [60]. For example, elderly patients may have multiple comorbidities and receive more than 10 concomitant medications, creating unique situations not covered by pre-market testing. Therefore, whenever drug prescribing patterns diverge from those used in pre-marketing clinical trials, severe ADRs may occur, making post-marketing surveillance a critical part of the drug safety research.

#### 3.1. Detecting drug-event associations – data driven approach

Pharmacovigilance effort worldwide have largely focused on data-driven surveillance by creating spontaneous reporting systems (SRSs) to collect post-market drug safety concerns from clinicians, patients, and pharmaceutical companies. The collected spontaneous reports are then mined for new ADR signals. The most established ADR detection methods aim to quantify the degree to which a drug-and-ADR pair co-occurs “disproportionally” in the SRS reports compared with what would be expected if there was no association [61–76]. Early pharmacovigilance efforts have cross-correlated (“data mined”) the Food and Drug Administration (FDA) and World Health Organization (WHO) adverse event databases looking for unanticipated patterns [65,77–81], yielding a number of previously unrecognized adverse drug events. However, significant limitations of the SRSs are widely known which include highly biased populations selected for representation, lack of adequate control groups, and biased nature of selecting events to report [82,83].

With the recent FDA’s Sentinel Initiative, an active surveillance system congressionally mandated to begin in 2012 to mine 100 million electronic medical records (EMRs) for drug safety signals, pharmacovigilance has begun to shift its focus toward EMR-based drug safety surveillance [84–95]. The advantage of using EMRs to detect iatrogenic harm of drugs retrospectively is well recognized. It is more objective and allows active surveillance and discovery of novel ADR signals and is not dependent on self-reporting. There exists a number of studies that aim to detect/discover ADRs using either structured [84,85,88–90,94–96] or unstructured EMRs [91,92,97–99], demonstrating the promise of EMR for studying drug effects. Methods employed are mostly based on statistical and machine learning algorithms such as odds ratio [90,100], Bayesian network model [101], and decision tree model [92]. However, these models can merely narrow down the range of associations (correlated relations) between a drug and condition without validation of their causal relations. In fact, indirect association stemming from another event, e.g., confounder, is a major issue in pharmacovigilance. Since relationships between medications, symptoms, and diseases are rarely mentioned in EMRs, computationally detected ADR signals from EMRs are likely to be confounded by co-medication, indication, comorbidity, or any of the combinations. It is thus critical for observational studies utilizing EMRs to identify and remove possible confounders. Furthermore, it is important to note that the models described in this section are not intended to examine ADR risk factors.

#### 3.2. Identifying genetic risk factors of ADR – pharmacogenomic approach

Genetic predisposition is one important factor contributing to ADR. Classen et al. [102] estimated 50% of ADRs are likely to be related to genetic factors. To examine genetic risk factors of ADR in the post-marketing phase, various pharmacogenetics or pharmacogenomics (abbreviated here as PGx) approaches have been developed to link inherited differences to variable drug responses.

##### 3.2.1. Candidate gene approach

A number of important pharmacogenetics associations were discovered using the candidate gene approach and were proven to be useful in clinical practice. One successful example is the discovery that a relative or absolute deficiency of thiopurine s-methyltransferase (TPMT) can increase the risk of severe bone marrow suppression [103], leading to TPMT level checking prior to the initiation of *azathioprine* in clinics. A more recent example is achieving more accurate and safe dosing of *warfarin* by following genotyping at *CYP2C9* and *VKORC1* genes [104,105]. Furthermore, *abacavir* hypersensitivity has been linked to the human leukocyte antigen (HLA) region where the association of *HLA-B\*57:01* and *abacavir* hypersensitivity is perhaps the best example of the translation of PGx findings into clinical practice [106].

Although the candidate gene approach have resulted in important PGx findings, many biomarkers discovered through candidate gene studies have failed to be replicated, thus hindering the overall success in clinical implementation. There are many reasons for this which includes poor sample size and poor phenotyping strategies. Another major limitation is that the candidate gene approach assumes that we have a complete understanding of the underlying mechanisms of how a drug acts in terms of both efficacy and safety, which is not true.

##### 3.2.2. Genome-wide association studies

In contrast to the candidate gene approach above, genome-wide association studies (GWAS) is a hypothesis free approach where all genes can be scanned for common genetic variations. GWAS is particularly valuable in identifying novel mechanisms beyond the

current known mechanisms that cause drug response variability and detecting small effects. These studies compare alleles at thousands of loci of individuals having different phenotypes. GWAS typically looks for associations between single-nucleotide polymorphisms (SNPs) and a disease status or trait through regression analyses. If a SNP is more frequent in people with a trait, the variant is said to be associated with the trait. The SNPs with the most significant associations identified at the 'discovery phase' are then retested in a new sample in the 'replication phase' [107]. An example of the use of GWAS in PGx research is the assessment of the variation in cytochrome P450 2C19 genotype associated antiplatelet effect with clopidogrel [108].

The main issue with GWAS is also the sample size. For most ADR studies, since genetic risk factors of ADRs tend to be drug specific, the number of cases available is small, which limits the power to detect significant effects, particularly for rare ADRs. A potential approach in combating the problem is meta-analysis of multiple GWAS undertaken by different groups on the same phenotype. However, meta-analysis using imputed genotypic data across different platforms may be limited by the bias introduced into such derived data due to the weaknesses of various imputation methods [109,110].

### 3.2.3. Next-generation sequencing

A relative new PGx approach is to utilize next-generation sequencing, a high-throughput parallel-sequencing technology that can produce thousands or millions of sequences concurrently, enabling examination of the whole genome at a low cost. As the cost continues to fall, it will enable individuals with rare or novel pharmacogenomic variants to be discovered, which would have been undetectable previously using common variant analysis technologies [111]. For example, 250 variants in 231 ADME (Absorption, Distribution, Metabolism, and Excretion) related genes in an individual were identified by conventional screening assays, compared to 17,733 variants identified using next-generation sequencing, and of which 861 variants were thought to be functionally significant [111,112].

Next-generation sequencing technologies have already made a significant impact on the identification of drug response predictors. In cancer research, sequencing of somatic genome has resulted in the identification of driver mutations that can be treated by drug. In addition, sequencing of germline genome may enable the identification of sub-phenotypes of complex diseases with variable responses to different drugs [113]. There are already examples of rare variants identified to be associated with serious ADRs, for example, association between butyrylcholinesterase deficiency and prolonged apnea from the treatment with *suxamethonium*. Furthermore, Behr and Roden [114] found at least 10% of drug-induced torsades de pointes episodes may be due to rare mutations in the congenital long QT syndrome genes.

### 3.2.4. Comparison of the approaches

Each of the abovementioned approaches has its own advantages and application scenarios that each is most suitable for. The candidate gene approach makes use of known targets with small searching range, thus it is generally cheaper than the other approaches. However, it assumes these candidate genes are dominant, which neglects the connection among genes. On the other hand, GWAS focus on known SNP locations, which is much more systematic than the candidate gene approach but it is still challenging to identify unknown rare variants. Finally, next-generation sequencing is the most de novo approach that can find variants in unexpected or even unknown regions. However, its biggest drawback is cost and its requirement for intensive data analysis effort.

### 3.3. Phenotypic risk factors of ADR

In addition to the genetic risk factors contributing to ADR, response to a drug also depends partly on the patient's phenotypic characteristics, e.g., differences in metabolism due to age, sex, race, or hepatic/renal insufficiency, coexistence of other disorders, and use of other drugs. This section discusses some important phenotypic risk factors of ADR that can be discovered through data analysis.

#### 3.3.1. Age

Patient age has been found to be an independent risk factor for ADRs because a variety of age-related physiological changes can affect the pharmacokinetic and pharmacodynamics of drugs [115]. Infants and very young children are at much higher risk for ADRs because their metabolism is not fully developed yet. For example, the antibiotic chloramphenicol given to newborns can cause a serious and often fatal reaction called gray baby syndrome due to their inability to metabolize and eliminate the drug [116]. Moreover, if infants and young children are given another antibiotic tetracycline during their teeth forming period, it may permanently discolor their teeth enamel [117].

Older adults are also at high risk of having ADRs. Elderly are more likely to have complex diseases, thus may be taking multiple drugs to manage their health. The risk of adverse drug interactions increases exponentially with each new drug added to an individual's existing regime [118]. Also, as people age, their renal and liver reserve will decrease, thus increasing the risk for delayed renal and hepatic clearance of drugs [119].

#### 3.3.2. Kidney and liver function

Kidney and liver are two principal organs to eliminate drugs and their metabolites from the human body. Since drugs are mostly lipid soluble, they must undergo biotransformation to more polar metabolites before it can be efficiently removed from the blood circulation by renal excretory mechanisms. Renal dysfunction not only alters the renal excretion mechanism but also leads to modifications in the distribution, transport, and biotransformation of drugs, and affects the pharmacodynamics actions of drugs [120]. Similarly, hepatic dysfunction also has complex effects on drug clearance, biotransformation, and pharmacokinetics. It may alter intestinal absorption, plasma protein binding, hepatic extraction ratio, liver blood flow, biliary excretion, and etc., resulting in increased levels of bioavailable drug [121]. This alteration may cause normal drug doses to have toxic effects. For example, sensitivity of opioids and sedatives is often enhanced in patients with chronic liver disease where a small dose intake may precipitate encephalopathy.

Therefore, patients with acute or chronic kidney/liver disease must have drug dosage adjustment. For patients with renal impairment, it is a standard of practice to adjust dosages of drugs according to creatinine clearance or glomerular filtration rate (GFR). For patients with hepatic impairment, there are no general rules for adjusting drug dosages because levels and effects of individual drug do not correlate well with the type of liver injury or severity. Current guidelines recommend that before administering drugs that are largely cleared by hepatic mechanisms, the drugs' potential therapeutic benefits must be carefully counterbalanced with their risk of serious ADRs [121]. If these drugs are required by patients with severe hepatic insufficiency, it is recommended that they should be started at a low dose and subsequently titrated to its desired therapeutic effect.

#### 3.3.3. Comorbidity and polypharmacy

Many studies have also found that the risk of ADR increases with the disease burden and the number of medications a patient

is concurrently on [122]. In other words, ADRs can be caused by drug-disease interaction and drug-drug interaction/polypharmacy. For instance, hypertensive patients co-administering drugs that can cause sodium retention will increase their blood pressure, resulting in serious ADRs. Other examples include the use of aspirin and beta adrenoceptor blockers may worsen asthma, the use of oral contraceptives can exacerbate thrombo-embolic disorders [123], and cyclo-oxygenase has been linked to death in cardiac patients [124].

Another important risk factor for serious ADR is drug-drug interaction. For the past 15 years, the percentage of Americans consuming multiple medications at the same time has increased by 20%, with more than 76% of Americans over 60-years-old taking multiple drugs at the same time [125]. Drug-drug interaction can increase the magnitude of drug toxicities [126]. Specifically, one drug may alter the absorption, distribution, metabolism or elimination of another drug, which may result in higher serum concentration of the agents and lead to excessive response or toxicity. For example, *cerivastatin* caused 31 cases of fatal rhabdomyolysis prior to its withdrawal in June 2001; the combination *cerivastatin* – *gemfibrozil* was implicated in 12 of the 31 deaths [127] because *gemfibrozil* caused elevated blood levels of the statin resulting in a higher risk of myopathy and rhabdomyolysis. DDIs may also cause altered toxicities. For example, lipid-lowering agent *pravastatin* and the antidepressant *paroxetine* taken together can lead to unexpected increases in blood glucose levels, while no such effect is observed when either drug is administered alone [128].

#### 4. ADR prevention in healthcare with clinical decision support systems

To influence real world medical decisions, knowledge discovered and validated from previous sections must to be translated to clinical practice. In effort to prevent adverse drug events, most hospitals have implemented clinical decision support systems (CDSS) based on the accumulated knowledge in their computerized physician order entry (CPOE). The CDSSs encompass a variety of computerized tools aimed at reducing medication error, especially at the prescribing stage, which include computerized reminders and guidance regarding drug selection, dosage, interaction, and allergies. Functionality of the current CDSS varies across hospitals, providing basic or advanced guidance to the prescribers.

##### 4.1. Basic medication-related decision support

Basic medication-related decision support systems typically include drug-allergy, drug-drug interaction and basic dosage checking.

###### 4.1.1. Drug-allergy checking

The drug-allergy checking functionality in CDSS will present an alert to the provider if a patient has an electronic documentation of a specific drug allergy and the medication is ordered. Although extremely valuable, current drug-allergy checking is inconsistent across applications and major shortcomings make accurate allergy checking nearly impossible; for example, in some CPOE applications, allergens and allergic reactions are not recorded as structured or coded entry. In addition, many applications do not distinguish between drug allergy and drug sensitivity, generating a large amount of irrelevant alerts that is highly disruptive of the clinical workflow [129]. For example, a drug-allergy alert is triggered on a *cephalosporin* order for patients allergic to only *penicillin*. A major reason behind the over-alerting is due to inadequate drug prescribing knowledgebase. Another reason that

hinders optimal drug-allergy alerting is the poor quality of allergy data recorded in the EMR.

###### 4.1.2. Basic dosing guidance

Medications with narrow margins of safety, such as oncologic agents, sedatives, and narcotics, are sensitive to dosing mistakes. CDSS can improve medication dosing through multiple mechanisms. One simple method is to provide the clinician a list of patient-appropriate dosing parameters for each medication with default selection of the most appropriate initial dose [130] or present to clinicians a list of complete order sentences (e.g., include dose, dose form, route, etc.) for selection [131]. A more complex method is to review each medication order through algorithmic analysis and only alert clinicians when reasonable dosing parameters have been exceeded [132].

###### 4.1.3. Drug-drug interaction checking

Automated drug-drug interaction checking has been implemented at point-of-care based on knowledgebase supplied by many sources including commercial vendors. Despite its promise to reduce harm to patients, it can generate large numbers of clinically insignificant alerts that clinicians typically ignore. The override rate for high-severity drug-drug interaction alerts can be as high as 89% [133]. One major reason behind the large overridden rate is again inadequate knowledgebase that include minor interactions of low clinical relevance, and is not treated differently from the clinical significant ones. Another related reason is that the available drug-drug interaction knowledge is still very limited, especially lacking information on susceptible sub-populations.

##### 4.2. Advanced medication-related decision support

Advanced medication-related decision support in preventing adverse drug events include advanced dosing guidance, drug-associated lab testing, and drug-disease interaction checking.

###### 4.2.1. Advanced dosing guidance

As described in Section 3, pediatric and geriatric populations and patients with renal or hepatic insufficiency are especially susceptible to serious ADRs and require appropriate dosing adjustments. However, the basic dose checking CDSS sometimes assume that patients are non-geriatric adults and have normal renal and hepatic functions. Thus, advanced medication dosing decision support tools have been implemented to consider the complex patient variations [134,135]. Many factors may affect the safe dosage for a particular patient, which include patient characteristics such as age, weight, height, comorbidities, and physiologic status (e.g., renal and hepatic function), indication of the drug, concurrent medications, and patient's previous response to drug.

###### 4.2.2. Advanced guidance for medication-associated laboratory testing

A number of drug classes including anticoagulants, digoxin, and antiepileptic drugs have narrow therapeutic window, thus requiring close monitoring of drug levels and possible adverse effects with laboratory tests prior to the initial administration as well as at regular intervals during treatment. For certain medications like digoxin and antiepileptic drugs, the monitoring is primarily for tracking drug levels. For other medications like angiotensin-converting enzyme inhibitors that can raise serum potassium and creatinine levels, laboratory monitoring is mainly for tracking any possible adverse effects. Finally, medications such as aminoglycoside would require monitoring of both drug levels and physiological status, i.e., creatinine levels. Implementation of reminders in CPOE can help physicians comply with medication monitoring recommendations. Three important prerequisites exist for implementing a robust medication-lab reminder: (1) patient's



previous laboratory results must be readily accessible; (2) must inform providers when a patient is due or overdue for a monitoring test; (3) monitoring recommendations should be evidence-based.

#### 4.2.3. Advanced drug-disease interaction checking

There exist many contraindications that clinicians should avoid during prescribing based on patients' pre-existing comorbidities. For example, non-selective beta-blocking drugs should not be prescribed to patients with asthma as it may adversely affect pulmonary function, and many drugs are contraindicated with myasthenia gravis. CDSS can assist physicians to avoid contraindicated medication by alerting them about patients' underlying conditions. To enable such checking, one must have an adequate knowledgebase of contraindications, in which many vendors and healthcare organizations have already developed from package inserts. Contraindication checking also requires patients' diagnoses and conditions to be accurately and comprehensively documented as coded entry in EMR.

### 5. Challenges and future research directions

#### 5.1. Challenges in predicting ADRs using drug-target interaction networks

- There are dozens of databases providing information on different levels (Table 1). Part of the information is complementary to each other, but the rest is not. For example, DrugBank, DR.PRODIS, SIDER, STITCH and SuperTarget all have predicted ADRs for drugs, some of the predictions are concordant, some are complementary, but some are rated differently or even conflictive. However, depending on the credible level, the number of predicted ADRs for drugs in databases can be huge and experimental validations on all these predictions are not feasible. One of the solutions is to generate a database on the top of them to clean up all these predictions and give a score to represent the reliability of each prediction (by counting the time a prediction show up in different databases or using more comprehensive statistical models). The predictions with low score or with conflicts must be highlighted by question mark until experimental validation is done. However, this activity requires a huge collaboration and investment from the drug development community, but the benefit is not that obvious, which make it challenging to come true.
- The reliability of databases is different because of the various resolutions of the data. For instance, a reported ADR from a patient in the SRS should be more reliable than an indication from an EMR record. Also, how representative a validation on animal model or cell-lines can be? Therefore, researchers should consider this factor when they are doing predictions using multiple data-sources.
- Balancing sensitivity and specificity is always a problem of mathematical models. Here, it is the question of how to find the meaningful targets from numerous interactions in the network. Besides of refining models, narrowing down the search space to sub-categories by introducing appropriate factors may be beneficial.

#### 5.2. Future of predicting ADRs using drug-target interaction networks

- It is known that lots of drugs (especial antibiotics) will change the microbiome in a human body. Recently, studies reported the associations between the microbiome in gut and intestinal and chronic diseases [136,137]. However, the association between drug and microbiome is not well considered in the ADR prediction models yet. As the knowledge among drug, microbiome and disease increases, using drug-microbiome

interaction network in ADR prediction will become important, especially in the field of preventive medicine.

- Recently, deep-learning has made some breakthrough and AlphaGo beat Lee Sedol in the board game Go. Also, IBM Watson has shown a good example of using deep-learning in the medical filed. Deep-learning can solve more complex problems compared to regular machine learning methods. And the structure of drug-target interaction networks is similar to the neural network, which fits the concept of deep-learning well. Therefore, using deep-learning to study the network and predict ADRs will be a trend.

#### 5.3. Challenges in pharmacovigilance

- Although EMR has become a mainstream data resource for pharmacovigilance, there are a number of challenges in using it for such studies. Clinical data are highly heterogeneous, and their heterogeneity includes variation in both structured and unstructured data. Much relevant clinical information such as drug exposure and responses is embedded in narrative clinical text and is not immediately accessible for analysis. Additionally, compared with randomized clinical trial data, observational data from EMRs typically are less robust for drawing scientific conclusions due to multiple factors such as missing data and confounders. Hence, more effective and efficient algorithms for data extraction and preparation are highly desirable.

#### 5.4. Future of pharmacovigilance

- With the current advances in clinical natural language processing (NLP) systems [138], much of the buried information in clinical notes can be accurately extracted. As substantial portion of the medical information accumulated are textual (e.g., EMRs, medical case reports, research articles, patents, etc.), NLP or text mining will be an essential part of pharmacovigilance.
- To ascertain causal relationships between drugs and adverse events, indirect associations due to confounders must be effectively identified and eliminated. In the past, causal inference in biomedical research has primarily taken the pragmatic approach with randomized controlled trials; however conducting clinical trials have many well-known limitations such as ethical and practical constraints. Recently in the computational world, advances in causal Bayesian network theory have enabled computational modeling and learning of causal structures from observational data [139,140]. Thus, causal discovery based on large observational datasets is a promising direction for pharmacovigilance.

#### 5.5. Challenges in pharmacogenomics

- The major challenge and criticism for existing pharmacogenomics studies regardless of approach (i.e., candidate gene, GWAS, or next-generation sequencing) is sample size. Evaluation of genetic risk factors of any phenotype requires the interrogation of large patient databases containing well-characterized cases and carefully selected controls, especially when the effect sizes of individual variants are small. Partnership between healthcare systems, academia, regulatory agencies and industry are needed to form large data networks.
- Another major challenge in identifying genetic risk factors for ADRs is the non-uniform or non-standardized clinical phenotype definitions. Phenotype definition standardization is clearly required so that well-characterized cases and controls can be collected; this is also important in facilitating replication studies. To develop such standardization, it requires discussion among scientists, clinicians, and regulators.

### 5.6. Future of pharmacogenomics

- Currently, the most advanced pharmacogenomic tests implemented in clinical care are only ordered at the initiation of a treatment regimen. The genetic tests are not ordered at the optimal time point, where the treatment regimen must be recalled if changes are determined to be needed after retrieval of the genetic tests results. Ideally, the genetic test result is to guide drug selection or dosing and the delay is undesirable. To achieve genetic-based prescription, two objectives must be satisfied: (1) known polymorphisms associated with ADRs are clinically relevant and can significantly reduce ADR occurrence; (2) we need more economic and easy pharmacogenomics tests that can be retrieved in a short period of time for clinicians.

### 5.7. Challenges in medication-related CDSS

- The main challenge for medication-related CDSS is over-alerting with clinically irrelevant alerts, causing alert-fatigue for physicians. Alert-fatigue may result in serious consequences because important alerts can be missed by physicians leading to adverse outcomes of patients. The major cause of over-alerting is that the drug prescribing knowledgebase is often inadequate, containing information of questionable clinical value.

### 5.8. Future of CDSS

- To realize the benefits of CDSS, multiple stakeholders such as application and knowledge-base vendors, healthcare providers, and government agencies will need to collaborate. Clinical information system or EMR vendors need to thoughtfully implement as many proved and effective medication-related decision support features as possible. Alerts must be concise and specific with sufficient information for clinicians to understand the rationale for the interruption. Knowledgebase vendors need to create clinically meaningful and pragmatic knowledge that is evidence-based. Finally, government agencies need to encourage organizations that have been successfully custom building their CDSS (usually large medical centers like academic institutions) to share knowledge and expertise.

## Acknowledgements

The authors would like to thank the support from a number of research grants. Y.T. is supported by Guangdong Innovative and Entrepreneurial Research Team Program (No. 201301S0105240297) and Major International Joint Research Program of China (Grant 31420103901). Y.H. is supported by the National Science Foundation of China (71271061, 70801020), Science and Technology Planning Project of Guangdong Province, China (2010B010600034, 2012B091100192). X.C. is supported by National Science Foundation (Award OIA-1028098).

## References

- [1] International Drug Monitoring: The Role of the Hospital, Technical Report Series No. 425, World Health Organization, Geneva, Switzerland, 1966.
- [2] M. Pirmohamed, A.M. Breckenridge, N.R. Kitteringham, B.K. Park, Adverse drug reactions, *BMJ* 316 (7140) (1998) 1295–1298.
- [3] J. Lazarou, B.H. Pomeranz, P.N. Corey, Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies, *JAMA* 279 (15) (1998) 1200–1205.
- [4] A.K. Jha, G.J. Kuperman, E. Rittenberg, J.M. Teich, D.W. Bates, Identifying hospital admissions due to adverse drug events using a computer-based monitor, *Pharmacoevidenciol. Drug Saf.* 10 (2) (2001) 113–119.
- [5] R. Leone, L. Sottosanti, M. Luisa Iorio, C. Santuccio, A. Conforti, V. Sabatini, U. Moretti, M. Venegoni, Drug-related deaths: an analysis of the Italian spontaneous reporting database, *Drug Saf.* 31 (8) (2008) 703–713.
- [6] C.S. van der Hooft, M.C. Sturkenboom, K. van Groothoest, H.J. Kingma, B.H. Stricker, Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands, *Drug Saf.* 29 (2) (2006) 161–168.
- [7] R.A. Wilke, D.W. Lin, D.M. Roden, P.B. Watkins, D. Flockhart, I. Zineh, K.M. Giacomini, R.M. Krauss, Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges, *Nat. Rev. Drug Discov.* 6 (11) (2007) 904–916.
- [8] T.J. Moore, M.R. Cohen, C.D. Furberg, Serious adverse drug events reported to the Food and Drug Administration, 1998–2005, *Arch. Intern. Med.* 167 (16) (2007) 1752–1759.
- [9] K.M. Giacomini, R.M. Krauss, D.M. Roden, M. Eichelbaum, M.R. Hayden, Y. Nakamura, When good drugs go bad, *Nature* 446 (7139) (2007) 975–977.
- [10] M. Pirmohamed, B.K. Park, Genetic susceptibility to adverse drug reactions, *Trends Pharmacol. Sci.* 22 (6) (2001) 298–305.
- [11] D.J. Weatherall, Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases, *BMJ* 321 (7269) (2000) 1117–1120.
- [12] S.F. Lin, K.T. Xiao, Y.T. Huang, C.C. Chiu, V.W. Soo, Analysis of adverse drug reactions using drug and drug target interactions and graph-based methods, *Artif. Intell. Med.* 48 (2–3) (2010) 161–166.
- [13] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces, *Bioinformatics* 24 (13) (2008) i232–i240.
- [14] G.K. Dy, A.A. Adjei, Understanding, recognizing, and managing toxicities of targeted anticancer therapies, *CA Cancer J. Clin.* 63 (4) (2013) 249–279.
- [15] J.P. Overington, B. Al-Lazikani, A.L. Hopkins, How many drug targets are there?, *Nat. Rev. Drug Discov.* 5 (12) (2006) 993–996.
- [16] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res.* 34 (Database issue) (2006) D668–D672.
- [17] M. Kuhn, C. von Mering, M. Campillos, L.J. Jensen, P. Bork, STITCH: interaction networks of chemicals and proteins, *Nucleic Acids Res.* 36 (Database issue) (2008) D684–D688.
- [18] M. Kuhn, D. Szklarczyk, A. Franceschini, C. von Mering, L.J. Jensen, P. Bork, STITCH 3: zooming in on protein-chemical interactions, *Nucleic Acids Res.* 40 (Database issue) (2012) D876–D880.
- [19] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.* 40 (D1) (2012) D1100–D1107.
- [20] E. Bolton, Y. Wang, P.A. Thiessen, S.H. Bryant, PubChem: Integrated Platform of Small Molecules and Biological Activities, Chapter 12 in *Annual Reports in Computational Chemistry*, American Chemical Society, Washington, DC, 2008.
- [21] S. Gunther, M. Kuhn, M. Dunkel, M. Campillos, C. Senger, E. Petsalaki, J. Ahmed, E.G. Urdiales, A. Gewiss, L.J. Jensen, R. Schneider, R. Skoblo, R.B. Russell, P.E. Bourne, P. Bork, R. Preissner, SuperTarget and Matador: resources for exploring drug-target relationships, *Nucleic Acids Res.* 36 (Database issue) (2008) D919–D922.
- [22] B. Roth, E. Lopez, S. Patel, W. Kroeze, The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches?, *Neuroscientist* 6 (2000) 252–262.
- [23] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R.C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeifferberger, P. Porras, A. Raghunath, B. Roecher, S. Orchard, H. Hermjakob, The IntAct molecular interaction database in 2012, *Nucleic Acids Res.* 40 (Database issue) (2012) D841–D846.
- [24] J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A. Armstrong, S.J. Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, T.R. Golub, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease, *Science* 313 (5795) (2006) 1929–1935.
- [25] F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaokara, R. Ferriero, L. Murinob, R. Tagliaferri, N. Brunetti-Pierri, A. Isacchi, D.D. Bernardoa, Discovery of drug mode of action and drug repositioning from transcriptional responses, *PNAS* 107 (33) (2010) 14621–14626.
- [26] Q. Duan, C. Flynn, M. Niepel, M. Hafner, J.L. Muhlich, N.F. Fernandez, A.D. Rouillard, C.M. Tan, E.Y. Chen, T.R. Golub, P.K. Sorger, A. Subramanian, A. Ma'ayan, LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures, *Nucleic Acids Res.* 42 (Web Server issue) (2014) W449–W460.
- [27] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein data bank, *Nucleic Acids Res.* 28 (1) (2000) 235–242.
- [28] F. Zhu, Z. Shi, C. Qin, L. Tao, X. Liu, F. Xu, L. Zhang, Y. Song, J. Zhang, B. Han, P. Zhang, Y. Chen, Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery, *Nucleic Acids Res.* 40 (Database issue) (2012) D1128–D1136.
- [29] M. Scheer, A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Sohngen, M. Stelzer, J. Thiele, D. Schomburg, BRENDA, the enzyme information system in 2011, *Nucleic Acids Res.* 39 (Database issue) (2011) D670–D676.
- [30] Y. Okuno, A. Tamon, H. Yabuuchi, S. Nijima, Y. Minowa, K. Tonomura, R. Kunimoto, C. Feng, GLIDA: GPCR ligand database for chemical genomics drug discovery database and tools update, *Nucleic Acids Res.* 36 (Database) (2007) D907–D912.

- [31] R.D. Finn, P. Coghill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.* 44 (D1) (2016) D279–D285.
- [32] J.Y. Chen, S. Mamidipalli, T. Huan, HAPPI: an online database of comprehensive human annotated and predicted protein interactions, *BMC Genomics* 10 (Suppl. 1) (2009) S16.
- [33] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.* 43 (Database issue) (2015) D447–D452.
- [34] A. Ceol, A. Chatr Aryamontri, L. Licata, D. Peluso, L. Briganti, L. Perfetto, L. Castagnoli, G. Cesareni, MINT, the molecular interaction database: 2009 update, *Nucleic Acids Res.* 38 (Database issue) (2010) D532–D539.
- [35] C.L. Myers, D. Robson, A. Wible, M.A. Hibbs, C. Chiriac, C.L. Theesfeld, K. Dolinski, O.G. Troyanskaya, Discovery of biological networks from diverse functional genomic data, *Genome Biol.* 6 (13) (2005) R114.
- [36] S.F. Zhu, Y. Okuno, G. Tsujimoto, H. Mamitsuka, A probabilistic model for mining implicit 'chemical compound-gene' relations from literature, *Bioinformatics* 21 (2005) 245–251.
- [37] A.C. Cheng, R.G. Coleman, K.T. Smyth, Q. Cao, P. Souillard, D.R. Caffrey, A.C. Salzberg, E.S. Huang, Structure-based maximal affinity model predicts small-molecule druggability, *Nat. Biotechnol.* 25 (1) (2007) 71–75.
- [38] M.X. LaBute, X.H. Zhang, J. Lenderman, B.J. Bennion, S.E. Wong, F.C. Lightstone, Adverse drug reaction prediction using scores produced by large-scale drug-protein target docking on high-performance computing machines, *PLoS One* 9 (9) (2014).
- [39] M. Kuhn, I. Letunic, L.J. Jensen, P. Bork, The SIDER database of drugs and side effects, *Nucleic Acids Res.* 44 (D1) (2016) D1075–D1079.
- [40] H.Y. Zhou, J. Skolnick, FINDSITEcomb: a threading/structure-based, proteomic-scale virtual ligand screening approach, *J. Chem. Inf. Model.* 53 (1) (2013) 230–240.
- [41] H. Liu, J. Sun, J. Guan, J. Zheng, S. Zhou, Improving compound-protein interaction prediction by building up highly credible negative samples, *Bioinformatics* 31 (12) (2015) i221–i229.
- [42] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori, The KEGG resource for deciphering the genome, *Nucleic Acids Res.* 32 (Database) (2004) 277–280.
- [43] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The gene ontology consortium, *Nat. Genet.* 25 (1) (2000) 25–29.
- [44] M. Sirota, J.T. Dudley, J. Kim, A.P. Chiang, A.A. Morgan, A. Sweet-Cordero, J. Sage, A.J. Butte, Discovery and preclinical validation of drug indications using compendia of public gene expression data, *Sci. Transl. Med.* 3 (96) (2011). 96ra77.
- [45] M. Kuhn, M. Campillos, I. Letunic, L.J. Jensen, P. Bork, A side effect resource to capture phenotypic effects of drugs, *Mol. Syst. Biol.* 6 (2010) 343.
- [46] J.S. Floyd, R. Kaspera, K.D. Marciano, N.S. Weiss, S.R. Heckbert, T. Lumley, K.L. Wiggins, B. Tamraz, P.Y. Kwok, R.A. Totah, B.M. Psaty, A screening study of drug-drug interactions in rivastatin users: an adverse effect of clopidogrel, *Clin. Pharmacol. Ther.* 91 (5) (2012) 896–904.
- [47] F.J. Azuaje, L. Zhang, Y. Devaux, D.R. Wagner, Drug-target network in myocardial infarction reveals multiple side effects of unrelated drugs, *Sci. Rep.* 1 (2011) 52.
- [48] M. Takarabe, M. Kotera, Y. Nishimura, S. Goto, Y. Yamanishi, Drug target prediction using adverse event report systems: a pharmacogenomic approach, *Bioinformatics* 28 (18) (2012) i611–i618.
- [49] N.P. Tatonetti, P.P. Ye, R. Daneshjoui, R.B. Altman, Data-driven prediction of drug effects and interactions, *Sci. Transl. Med.* 4 (125) (2012). 125ra31.
- [50] R. Ball, T. Botsis, Can network analysis improve pattern recognition among adverse events following immunization reported to VAERS?, *Clin Pharmacol. Ther.* 90 (2) (2011) 271–278.
- [51] H. Zhou, M. Gao, J. Skolnick, Comprehensive prediction of drug-protein interactions and side effects for the human proteome, *Sci. Rep.* 5 (2015) 11090.
- [52] E.M.M.W.Y.G.R.A.T. Klein, From pharmacogenomic knowledge acquisition to clinical applications: the Pharm GKB as a clinical pharmacogenomic biomarker resource, *Biomarkers Med.* 5 (2011).
- [53] M. Liu, Y. Wu, Y. Chen, J. Sun, Z. Zhao, X.W. Chen, M.E. Matheny, H. Xu, Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs, *J. Am. Med. Assoc. JAMA* 19 (e1) (2012) e28–e35.
- [54] L.C. Huang, X. Wu, J.Y. Chen, Predicting adverse drug reaction profiles by integrating protein interaction networks with drug structures, *Proteomics* 13 (2) (2013) 313–324.
- [55] F.J. Azuaje, Y. Devaux, D.R. Wagner, Prediction of adverse cardiovascular events of noncardiovascular drugs through drug-target interaction networks, *Clin. Transl. Sci.* 5 (1) (2012) 111.
- [56] S. Hennessy, D.A. Flockhart, The need for translational research on drug-drug interactions, *Clin. Pharmacol. Ther.* 91 (5) (2012) 771–773.
- [57] J. Kim et al., A novel data mining approach to the identification of effective drugs or combinations for targeted endpoints – application to chronic heart failure as a new form of evidence-based medicine, *Cardiovasc. Drugs Ther.* (2004).
- [58] A. Cami, A. Arnold, S. Manzi, B. Reis, Predicting adverse drug events using pharmacological network models, *Sci. Transl. Med.* 3 (114) (2011).
- [59] R.S. Ledley, L.B. Lusted, Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason, *Science* 130 (3366) (1959) 9–21.
- [60] M.R. Griffin, C.M. Stein, W.A. Ray, Postmarketing surveillance for drug safety: surely we can do better, *Clin. Pharmacol. Ther.* 75 (6) (2004) 491–494.
- [61] J.L. Montastruc, A. Sommet, H. Bagheri, M. Lapeyre-Mestre, Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database, *Br. J. Clin. Pharmacol.* 72 (6) (2011) 905–908.
- [62] A. Bate, S.J. Evans, Quantitative signal detection using spontaneous ADR reporting, *Pharmacoepidemiol. Drug Saf.* 18 (6) (2009) 427–436.
- [63] M. Hauben, D. Madigan, C.M. Gerrits, L. Walsh, E.P. Van Puijenbroek, The role of data mining in pharmacovigilance, *Exp. Opin. Drug Saf.* 4 (5) (2005) 929–948.
- [64] S.J. Evans, P.C. Waller, S. Davis, Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports, *Pharmacoepidemiol. Drug Saf.* 10 (6) (2001) 483–486.
- [65] A. Szafrman, S.G. Machado, R.T. O'Neill, Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database, *Drug Saf.* 25 (6) (2002) 381–392.
- [66] I. Ahmed, F. Haramburu, A. Fourrier-Reglat, F. Thiessard, C. Kreft-Jais, G. Miremont-Salame, B. Begaud, P. Tubert-Bitter, Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting, *Stat. Med.* 28 (13) (2009) 1774–1792.
- [67] W. DuMouchel, Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system, *Am. Statist.* 53 (3) (1999) 177–202.
- [68] J.S. Almenoff, E.N. Pattishall, T.G. Gibbs, W. DuMouchel, S.J. Evans, N. Yuen, Novel statistical tools for monitoring the safety of marketed drugs, *Clin. Pharmacol. Ther.* 82 (2) (2007) 157–166.
- [69] W. DuMouchel, E.T. Smith, R. Beasley, H. Nelson, X. Yang, D. Fram, J.S. Almenoff, Association of asthma therapy and Churg-Strauss syndrome: an analysis of postmarketing surveillance data, *Clin. Ther.* 26 (7) (2004) 1092–1104.
- [70] A.L. Gould, Accounting for multiplicity in the evaluation of "signals" obtained by data mining from spontaneous report adverse event databases, *Biomed. J.* 49 (1) (2007) 151–165.
- [71] A. Bate, M. Lindquist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner, R.M. De Freitas, A Bayesian neural network method for adverse drug reaction signal generation, *Eur. J. Clin. Pharmacol.* 54 (4) (1998) 315–321.
- [72] M. Lindquist, I.R. Edwards, A. Bate, H. Fucik, A.M. Nunes, M. Stahl, From association to alert—a revised approach to international signal analysis, *Pharmacoepidemiol. Drug Saf.* 8 (Suppl. 1) (1999) S15–S25.
- [73] M. Lindquist, M. Stahl, A. Bate, I.R. Edwards, R.H. Meyboom, A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database, *Drug Saf.* 23 (6) (2000) 533–542.
- [74] L. An, K.Y. Fung, D. Krewski, Mining pharmacovigilance data using Bayesian logistic regression with James-Stein type shrinkage estimation, *J. Biopharm. Stat.* 20 (5) (2010) 998–1012.
- [75] I. Ahmed, C. Dalmasso, F. Haramburu, F. Thiessard, P. Broet, P. Tubert-Bitter, False discovery rate estimation for frequentist pharmacovigilance signal detection methods, *Biometrics* 66 (1) (2010) 301–309.
- [76] I. Ahmed, F. Thiessard, G. Miremont-Salame, B. Begaud, P. Tubert-Bitter, Pharmacovigilance data mining with methods based on false discovery rates: a comparative simulation study, *Clin. Pharmacol. Ther.* 88 (4) (2010) 492–498.
- [77] M. Lindquist, I.R. Edwards, The WHO programme for international drug monitoring, its database, and the technical support of the Uppsala monitoring center, *J. Rheumatol.* 28 (5) (2001) 1180–1187.
- [78] E. Björnsson, R. Olsson, Suspected drug-induced liver fatalities reported to the WHO database, *Dig. Liver Dis.* 38 (1) (2006) 33–38.
- [79] K.A. Chan, M. Hauben, Signal detection in pharmacovigilance: empirical evaluation of data mining tools, *Pharmacoepidemiol. Drug Saf.* 14 (9) (2005) 597–599.
- [80] M. Hauben, L. Reich, C.M. Gerrits, Reports of hyperkalemia after publication of RALES—a pharmacovigilance study, *Pharmacoepidemiol. Drug Saf.* 15 (11) (2006) 775–783.
- [81] M. Hauben, L. Reich, S. Chung, Postmarketing surveillance of potentially fatal reactions to oncology drugs: potential utility of two signal-detection algorithms, *Eur. J. Clin. Pharmacol.* 60 (10) (2004) 747–750.
- [82] M. Hauben, V. Patadia, C. Gerrits, L. Walsh, L. Reich, Data mining in pharmacovigilance: the need for a balanced perspective, *Drug Saf.* 28 (10) (2005) 835–842.
- [83] M. Hauben, Application of an empiric Bayesian data mining algorithm to reports of pancreatitis associated with atypical antipsychotics, *Pharmacotherapy* 24 (9) (2004) 1122–1129.
- [84] J.S. Brown, M. Kuldorff, K.A. Chan, R.L. Davis, D. Graham, P.T. Pettus, S.E. Andrade, M.A. Raebel, L. Herrinton, D. Roblin, D. Boudreau, D. Smith, J.H. Gurwitz, M.J. Gunter, R. Platt, Early detection of adverse drug events within population-based health networks: application of sequential testing methods, *Pharmacoepidemiol. Drug Saf.* 16 (12) (2007) 1275–1284.



- [85] D.R. Berlowitz, D.R. Miller, S.A. Oliveria, F. Cunningham, A. Gomez-Camirero, J.A. Rothendler, Differential associations of beta-blockers with hemorrhagic events for chronic heart failure patients on warfarin, *Pharmacoevidemiol. Drug Saf.* 15 (11) (2006) 799–807.
- [86] H.D. Jin, J. Chen, H.X. He, G.J. Williams, C. Kelman, C.M. O'Keefe, Mining unexpected temporal associations: applications in detecting adverse drug reactions, *IEEE Trans. Inf. Technol. Biomed.* 12 (4) (2008) 488–500.
- [87] Y.Q. Ji, H. Ying, P. Dews, A. Mansour, J. Tran, R.E. Miller, R.M. Massanari, A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance, *IEEE Trans. Inf. Technol. Biomed.* 15 (3) (2011) 428–437.
- [88] I. Zorych, D. Madigan, P. Ryan, A. Bate, Disproportionality methods for pharmacovigilance in longitudinal observational databases, *Stat. Methods Med. Res.* (2011).
- [89] D. Yoon, M.Y. Park, N.K. Choi, B.J. Park, J.H. Kim, R.W. Park, Detection of adverse drug reaction signals using an electronic health records database: Comparison of the Laboratory Extreme Abnormality Ratio (CLEAR) algorithm, *Clin. Pharmacol. Ther.* 91 (3) (2012) 467–474.
- [90] M. Liu, E.R. McPeck Hinz, M.E. Matheny, J.C. Denny, J.S. Schildcrout, R.A. Miller, H. Xu, Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records, *J. Am. Med. Inform. Assoc.* (2012).
- [91] X. Wang, G. Hripcsak, M. Markatou, C. Friedman, Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study, *J. Am. Med. Inform. Assoc.* 16 (3) (2009) 328–337.
- [92] S. Sohn, J.P. Kocher, C.G. Chute, G.K. Savova, Drug side effect extraction from clinical narratives of psychiatry and psychology patients, *J. Am. Med. Inform. Assoc.* 18 (Suppl. 1) (2011) i144–i149.
- [93] P. Warrar, E.H. Hansen, L. Juhl-Jensen, L. Aagaard, Using text-mining techniques in electronic patient records to identify ADRs from medicine use, *Br. J. Clin. Pharmacol.* 73 (5) (2012) 674–684.
- [94] H.W. Jin, J. Chen, H. He, G.J. Williams, C. Kelman, C.M. O'Keefe, Mining unexpected temporal associations: applications in detecting adverse drug reactions, *IEEE Trans. Inf. Technol. Biomed.* 12 (4) (2008) 488–500.
- [95] Y. Ji, H. Ying, P. Dews, A. Mansour, J. Tran, R.E. Miller, R.M. Massanari, A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance, *IEEE Trans. Inf. Technol. Biomed.* 15 (3) (2011) 428–437.
- [96] M.E. Matheny, R.A. Miller, T.A. Ikizler, L.R. Waitman, J.C. Denny, J.S. Schildcrout, R.S. Dittus, J.F. Peterson, Development of inpatient risk stratification models of acute kidney injury for use in electronic health records, *Med. Decis. Making* 30 (6) (2010) 639–650.
- [97] X. Wang, G. Hripcsak, C. Friedman, Characterizing environmental and phenotypic associations using information theory and electronic health records, *BMC Bioinformatics* 10 (Suppl. 9) (2009) S13.
- [98] X. Wang, H. Chase, M. Markatou, G. Hripcsak, C. Friedman, Selecting information in electronic health records for knowledge acquisition, *J. Biomed. Inform.* 43 (4) (2010) 595–601.
- [99] P. LePendu, S.V. Iyer, A. Bauer-Mehren, R. Harpaz, J.M. Mortensen, T. Podchyska, T.A. Ferris, N.H. Shah, Pharmacovigilance using clinical notes, *Clin. Pharmacol. Ther.* 93 (6) (2013) 547–555.
- [100] D. Yoon, M.Y. Park, N.K. Choi, B.J. Park, J.H. Kim, R.W. Park, Detection of adverse drug reaction signals using an electronic health records database: comparison of the laboratory extreme abnormality ratio (CLEAR) algorithm, *Clin. Pharmacol. Ther.* (2012).
- [101] L. Duan, M. Kohoshneshin, W. Street, M. Liu, Adverse drug effect detection, *IEEE J. Biomed. Health Inform.* 17 (2) (2013) 305–311.
- [102] D.C. Classen, S.L. Pestotnik, R.S. Evans, J.F. Lloyd, J.P. Burke, Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality, *JAMA* 277 (4) (1997) 301–306.
- [103] L. Lennard, J.A. Van Loon, R.M. Weinshilboum, Pharmacogenetics of acute azathioprine toxicity: relationship to thiopurine methyltransferase genetic polymorphism, *Clin. Pharmacol. Ther.* 46 (2) (1989) 149–154.
- [104] M. Pirmohamed, G. Burnside, N. Eriksson, A.L. Jorgensen, C.H. Toh, T. Nicholson, P. Kesteven, C. Christensen, B. Wahlstrom, C. Staffberg, J.E. Zhang, J.B. Leathart, H. Kohnke, A.H. Maitland-van der Zee, P.R. Williamson, A.K. Daly, P. Avery, F. Kamali, M. Wadelius, E.-P. Group, A randomized trial of genotype-guided dosing of warfarin, *N. Eng. J. Med.* 369 (24) (2013) 2294–2303.
- [105] H.Z. Ring, D.L. Kroetz, Candidate gene approach for pharmacogenetic studies, *Pharmacogenomics* 3 (1) (2002) 47–56.
- [106] R. Pavlos, S. Mallal, E. Phillips, HLA and pharmacogenetics of drug hypersensitivity, *Pharmacogenomics* 13 (11) (2012) 1285–1306.
- [107] B.E. Stranger, E.A. Stahl, T. Raj, Progress and promise of genome-wide association studies for human complex trait genetics, *Genetics* 187 (2) (2011) 367–383.
- [108] A.R. Shuldiner, J.R. O'Connell, K.P. Bliden, A. Gandhi, K. Ryan, R.B. Horenstein, C.M. Damcott, R. Pakyz, U.S. Tantry, Q. Gibson, T.I. Pollin, W. Post, A. Parsa, B. D. Mitchell, N. Faraday, W. Herzog, P.A. Gurbel, Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy, *JAMA* 302 (8) (2009) 849–857.
- [109] J. Marchini, B. Howie, Genotype imputation for genome-wide association studies, *Nat. Rev. Genet.* 11 (7) (2010) 499–511.
- [110] B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G.R. Abecasis, Fast and accurate genotype imputation in genome-wide association studies through pre-phasing, *Nat. Genet.* 44 (8) (2012) 955–959.
- [111] G. Potamias, K. Lakiotaki, T. Katsila, M.T. Lee, S. Topouzis, D.N. Cooper, G.P. Patrinos, Deciphering next-generation pharmacogenomics: an information technology perspective, *Open Biol.* 4 (7) (2014).
- [112] C. Mizzi, B. Peters, C. Mitropoulou, K. Mitropoulos, T. Katsila, M.R. Agarwal, R. H. van Schaik, R. Drmanac, J. Borg, G.P. Patrinos, Personalized pharmacogenomics profiling using whole-genome sequencing, *Pharmacogenomics* 15 (9) (2014) 1223–1234.
- [113] T.J. Urban, Whole-genome sequencing in pharmacogenetics, *Pharmacogenomics* 14 (4) (2013) 345–348.
- [114] E.R. Behr, D. Roden, Drug-induced arrhythmia: pharmacogenomic prescribing?, *Eur Heart J.* 34 (2) (2013) 89–95.
- [115] K.C. Sikdar, J. Dowden, R. Alaghebandan, D. MacDonald, P. Peter, V. Gadag, Adverse drug reactions in elderly hospitalized patients: a 12-year population-based retrospective cohort study, *Ann. Pharmacother.* 46 (7–8) (2012) 960–971.
- [116] J. McIntyre, I. Choonara, Drug toxicity in the neonate, *Biol. Neonate* 86 (4) (2004) 218–221.
- [117] A.R. Sanchez, R.S. Rogers 3rd, P.J. Sheridan, Tetracycline and other tetracycline-derivative staining of the teeth and oral cavity, *Int. J. Dermatol.* 43 (10) (2004) 709–715.
- [118] C. Zhan, R. Correa-de-Araujo, A.S. Bierman, J. Sangl, M.R. Miller, S.W. Wickizer, D. Stryer, Suboptimal prescribing in elderly outpatients: potentially harmful drug-drug and drug-disease combinations, *J. Am. Geriatr. Soc.* 53 (2) (2005) 262–267.
- [119] P. Carbonin, M. Pahor, R. Bernabei, A. Sgadari, Is age an independent risk factor of adverse drug reactions in hospitalized medical patients?, *J. Am. Geriatr. Soc.* 39 (11) (1991) 1093–1099.
- [120] R.K. Verbeeck, F.T. Musuamba, Pharmacokinetics and dosage adjustment in patients with renal dysfunction, *Eur. J. Clin. Pharmacol.* 65 (8) (2009) 757–773.
- [121] R.K. Verbeeck, Pharmacokinetics and dosage adjustment in patients with hepatic dysfunction, *Eur. J. Clin. Pharmacol.* 64 (12) (2008) 1147–1161.
- [122] C.A. Naranjo, U. Busto, E.M. Sellers, P. Sandor, I. Ruiz, E.A. Roberts, E. Janecsek, C. Domecq, D.J. Greenblatt, A method for estimating the probability of adverse drug reactions, *Clin. Pharmacol. Ther.* 30 (2) (1981) 239–245.
- [123] D. Kurnik, R. Loebstein, Z. Farfel, D. Ezra, H. Halkin, D. Olchovsky, Complex drug-drug-disease interactions between amiodarone, warfarin, and the thyroid gland, *Medicine* 83 (2) (2004) 107–113.
- [124] E.C. Davies, C.F. Green, D.R. Mottram, M. Pirmohamed, Adverse drug reactions in hospitals: a narrative review, *Curr. Drug Saf.* 2 (1) (2007) 79–87.
- [125] Q. Gu, C.F. Dillon, V.L. Burt, Prescription drug use continues to increase: U.S. prescription drug data for 2007–2008, National Center for Health Statistics (2010).
- [126] J. Kuhlmann, W. Muck, Clinical-pharmacological strategies to assess drug interaction potential during drug development, *Drug Saf.* 24 (10) (2001) 715–725.
- [127] J.A. Staffa, J. Chang, L. Green, Cerivastatin and reports of fatal rhabdomyolysis, *N. Engl. J. Med.* 346 (7) (2002) 539–540.
- [128] N.P. Tatonetti, J.C. Denny, S.N. Murphy, G.H. Fernald, G. Krishnan, V. Castro, P. Yue, P.S. Tsao, I. Kohane, D.M. Roden, R.B. Altman, Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels, *Clin. Pharmacol. Ther.* 90 (1) (2011) 133–142.
- [129] G.J. Kuperman, T.K. Gandhi, D.W. Bates, Effective drug-allergy checking: methodological and operational issues, *J. Biomed. Inform.* 36 (1–2) (2003) 70–79.
- [130] J.M. Teich, P.R. Merchia, J.L. Schmitz, G.J. Kuperman, C.D. Spurr, D.W. Bates, Effects of computerized physician order entry on prescribing practices, *Arch. Intern. Med.* 160 (18) (2000) 2741–2747.
- [131] G.J. Kuperman, J.M. Teich, T.K. Gandhi, D.W. Bates, Patient safety and computerized medication ordering at Brigham and Women's Hospital, *Jt. Comm. J. Qual. Improv.* 27 (10) (2001) 509–521.
- [132] M.I. Oppenheim, C. Vidal, F.T. Velasco, A.G. Boyer, M.R. Cooper, J.G. Hayes, W. W. Frayer, Impact of a computerized alert during physician order entry on medication dosing in patients with renal impairment, *Proceedings/AMIA... Annual Symposium. AMIA Symposium* (2002) 577–581.
- [133] J.R. Spina, P.A. Glassman, P. Belperio, R. Cader, S. Asch, V.A.L.A.H.S. Primary Care Investigative Group, Clinical relevance of automated drug alerts from the perspective of medical providers, *Am. J. Med. Qual.* 20 (1) (2005) 7–14.
- [134] J.F. Peterson, G.J. Kuperman, C. Shek, M. Patel, J. Avorn, D.W. Bates, Guided prescription of psychotropic medications for geriatric inpatients, *Arch. Intern. Med.* 165 (7) (2005) 802–807.
- [135] G.M. Chertow, J. Lee, G.J. Kuperman, E. Burdick, J. Horsky, D.L. Seger, R. Lee, A. Mekala, J. Song, A.L. Komaroff, D.W. Bates, Guided medication dosing for inpatients with renal insufficiency, *JAMA* 286 (22) (2001) 2839–2844.
- [136] J. Aron-Wisnewsky, K. Clement, The gut microbiome, diet, and links to cardiometabolic and chronic disorders, *Nat. Rev. Nephrol.* 12 (3) (2016) 169–181.
- [137] M. Knip, H. Siljander, The role of the intestinal microbiota in type 1 diabetes mellitus, *Nat. Rev. Endocrinol.* 12 (3) (2016) 154–167.
- [138] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, *J. Am. Med. Inform. Assoc.* 17 (5) (2010) 507–513.
- [139] J.P. Pellet, Using Markov blankets for causal structure learning, *J. Mach. Learn. Res.* 9 (2008) 1295–1342.
- [140] C.F. Aliferis, A. Statnikov, I. Tsanardinou, S. Mani, X.D. Koutsoukos, Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation, *J. Mach. Learn. Res.* 11 (2010) 171–234.