Future Directions of Pharmacovigilance Studies Using Electronic Medical Recording and Human Genetic Databases

Young Hee Choi[1,†], Chang Yeob Han[2,†], Kwi Suk Kim[3] and Sang Geon Kim[3,4]

[1]*College of Pharmacy, Dongguk University_Seoul, Goyang, Korea*
[2]*Department of Pharmacology, School of Medicine, Wonkwang University, Iksan, Korea*
[3]*Department of Pharmacy, Seoul National University Hospital, Seoul, Korea*
[4]*College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, Korea*

Abstract

Adverse drug reactions (ADRs) constitute key factors in determining successful medication therapy in clinical situations. Integrative analysis of electronic medical record (EMR) data and use of proper analytical tools are requisite to conduct retrospective surveillance of clinical decisions on medications. Thus, we suggest that electronic medical recording and human genetic databases are considered together in future directions of pharmacovigilance. We analyzed EMR-based ADR studies indexed on PubMed during the period from 2005 to 2017 and retrospectively acquired 1161 (29.6%) articles describing drug-induced adverse reactions (e.g., liver, kidney, nervous system, immune system, and inflammatory responses). Of them, only 102 (8.79%) articles contained useful information to detect or predict ADRs in the context of clinical medication alerts. Since insufficiency of EMR datasets and their improper analyses may provide false warnings on clinical decision, efforts should be made to overcome possible problems on data-mining, analysis, statistics, and standardization. Thus, we address the characteristics and limitations on retrospective EMR database studies in hospital settings. Since gene expression and genetic variations among individuals impact ADRs, pharmacokinetics, and pharmacodynamics, appropriate paths for pharmacovigilance may be optimized using suitable databases available in public domain (e.g., genome-wide association studies (GWAS), non-coding RNAs, microRNAs, proteomics, and genetic variations), novel targets, and biomarkers. These efforts with new validated biomarker analyses would be of help to repurpose clinical and translational research infrastructure and ultimately future personalized therapy considering ADRs.

Key words: Adverse drug reactions, Electronic medical record, Retrospective surveillance

INTRODUCTION

Large electronic medical datasets that include patients' medical records have already proven useful in clinical research and have become essential for the analysis of patient medication in the era of big data in healthcare (1). Various medication-related decision processes have been implemented to ensure the efficiency and safety of medicines, such as dose guidance, drug allergy, and drug-drug interactions (2). Various parameters, including disease, age, the liver and kidney functions of patients, effects of excipients, additives, or preservatives, and food-drug interactions, should also be considered for the proper administration of medications.

Several studies have shown the value of pharmacovigilance research by using electronic medical record (EMR) data for use as decision support tools. EMR data may include passive and active referential information, remind-

Correspondence to: Sang Geon Kim, Department of Pharmacy, Seoul National University Hospital, College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: sgk@snu.ac.kr

[†]The first two authors contributed equally to this work.

ers, alerts, and guidelines related to adverse drug reactions (ADRs). Thus, EMR data may have great potential in pharmacovigilance research and could enable the rapid identification of patients for inclusion in interventional and observational studies. The use of EMR data and data-mining processes may enable us to produce effective decision support tools for the prediction of ADRs (2), thereby having the potential to repurpose clinical and translational research infrastructure.

ADRs are commonly used as an important factor to determine the success of a therapy. In particular, as clinical data of ADRs are now documented electronically, efforts to compile information on ADRs have received widespread support (2). However, few studies using EMRs have observed significant benefits on patient outcomes (3), perhaps owing to the small sample sizes or short investigational times, which did not allow clinically important events to be revealed. The exploitation of efficient medication depends on the accuracy of information in the EMR database (4). For example, duplication of medications, contra-

indication of patient's condition, and changes in the efficacy and toxicity of drugs with respect to their pharmacokinetic and pharmacogenomic characteristics may often be proficiently assessed by clinicians and pharmacists by using EMR data. As their use becomes more widespread, it is increasingly important to have better ways of analyzing EMR data to ensure the validity of the studies (5).

PAST APPROACHES FOR RETROSPECTIVE ANALYSIS

Examples of ADR articles published in PubMed. In our analysis of articles indexed in the PubMed databases, we assessed a set of reports published between 2005 and 2017 by using the following key words: EMRs, drug therapy, drug-related side effects, adverse reactions, drug, medication, pharmacoepidemiology, or pharmacology. The number of papers describing ADRs of major organs in patients is listed in Fig. 1. Drug-induced liver injury (DILI) constituted the greatest percentage of ADRs, as
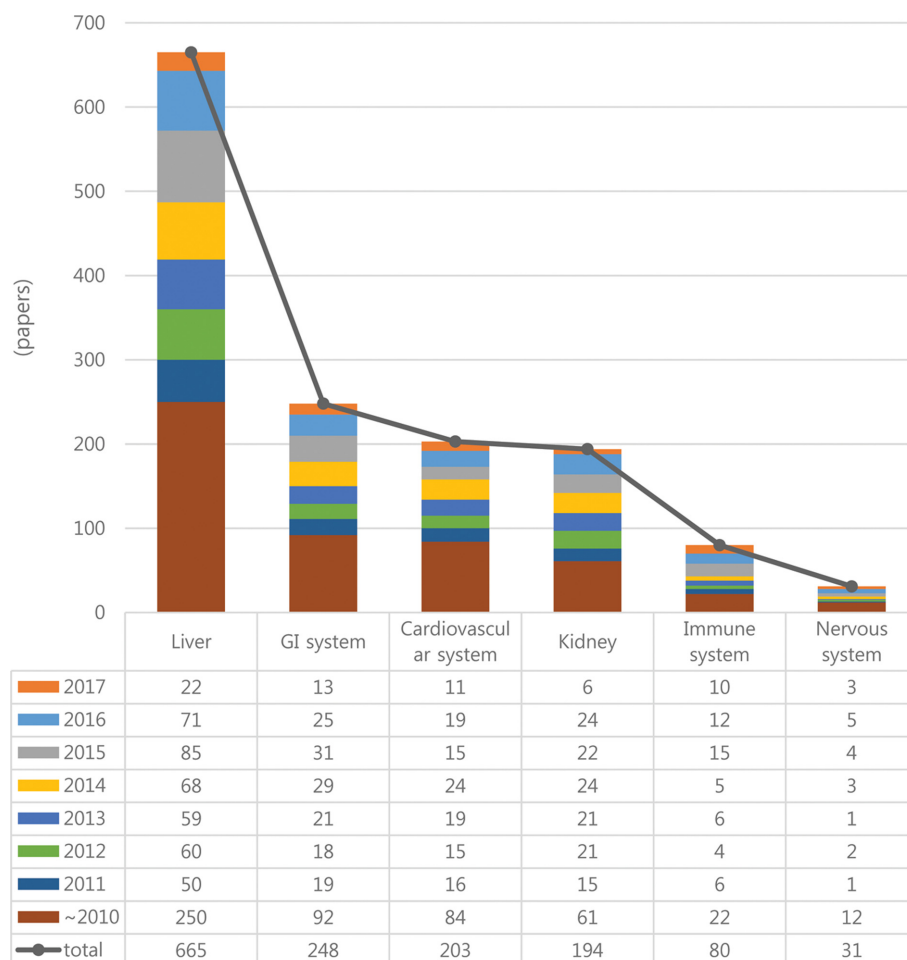

	Liver	GI system	Cardiovascular system	Kidney	Immune system	Nervous system
■ 2017	22	13	11	6	10	3
■ 2016	71	25	19	24	12	5
■ 2015	85	31	15	22	15	4
■ 2014	68	29	24	24	5	3
■ 2013	59	21	19	21	6	1
■ 2012	60	18	15	21	4	2
■ 2011	50	19	16	15	6	1
■ ~2010	250	92	84	61	22	12
●— total	665	248	203	194	80	31

Fig. 1. Number of articles retrospectively identifying adverse drug reactions (ADRs) by using electronic medical record (EMR) data. Articles archived in PubMed from 2005 through 2017 were sub-grouped by ADR occurrences in different organs.

shown in Fig. 1; the gastrointestinal tract, cardiovascular system, and kidney were the next most vulnerable locations to ADRs. There were no significant annual differences in the total number of ADRs. To assess the target organ, we counted the number of papers archived for different organs. From a total of 1421 potentially relevant publications, 1161 retrospective full-text publications were obtained after screening titles and abstracts using keywords presenting ADR events and target organs; 513 reported occurrences of ADRs only, 908 reported both efficacy and ADRs, and 648 (45.6%) focused on ADR reports based on EMR database mining. In addition, 86, 892, and 443 papers showed ADRs in subjects who were healthy, had disease(s) directly associated with the ADRs, and had disease(s) indirectly associated with ADRs, respectively. This review process was conducted based on the PRISMA guidelines (Fig. 2). In Fig. 2, ADR reports from EMR-based quantitative analyses were ~56% of all ADR studies, suggesting that EMR data-mining is still widely used for ADR assessment.

The papers referring to ADRs currently have several shortcomings for determining a conclusion on incidence rates or the severity of ADRs in sub-groups. For example, it was difficult to make a sophisticated decision on 'drug-induced liver injury' and 'no drug-induced liver injury' and to calculate the mismatched occurrence rate of DILI (6). Especially, when researchers try to predict ADR occurrence rates and severity of liver injury depending on sub-groups after combining the results from retrospective and prospective studies, the total number of patients exposed to target drugs cannot be calculated exactly. In addition, information on the characteristics of patients in each sub-group was not easy to obtain from retrospective EMR studies (6,7). Moreover, ADRs were further categorized to
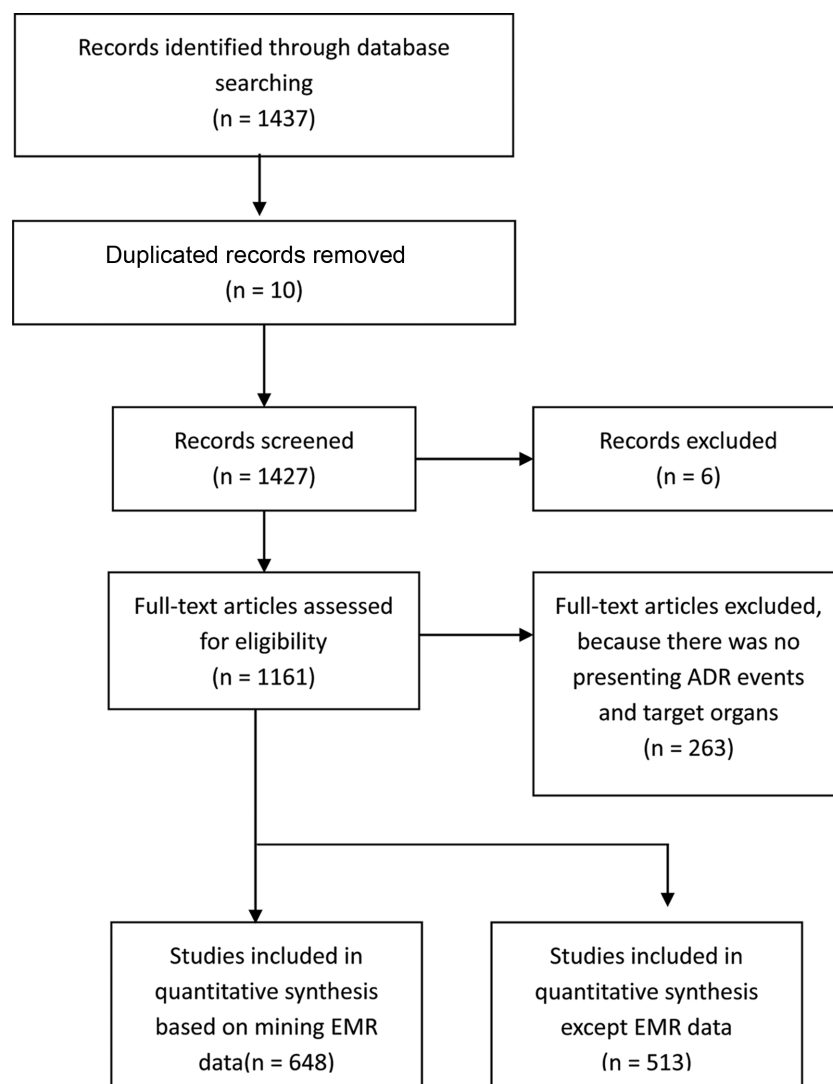

Fig. 2. Review process of full-text articles presenting adverse drug reaction (ADR) events.

sub-groups, such as DILI and gastrointestinal tract injury. In these cases, various incidence rates of ADRs can be calculated. As reports of ADRs emerge for new drugs or for various clinical cases (8), the analytical methodology to combine larger scale reports of ADRs is a prerequisite.

Routine analytical procedures for ADR prediction from EMR databases. ADRs have become a clinical issue and a concern to public health system; they are responsible for 6.5% of all hospital admissions (e.g., approximately one quarter of the patients had a risk of death) (9,10). As knowledge of the occurrence and the content of these ADR reports has increased (11,12), drug safety evaluation, including ADR monitoring, has become an import-

ant issue in pharmacotherapy (13). Thus, signal detection for ADR alerts in hospitals is currently regarded as active drug safety surveillance. To reduce ADRs, it is essential to identify the causal relationship between drug medications and the incidence of ADRs (14). Data acquisition using laboratory signals from patients' EMRs is the first step in the identification of ADRs and other conditions in patients (15). Reviews of EMR data may be used to assess the incidence of ADRs (16). Three major methods are often used for ADR reports: 1) retrospective chart reviews; 2) ADR reviews based on patients' EMRs (17); and 3) spontaneous reports by clinicians or patients, which can be focused on for additional chart reviews.
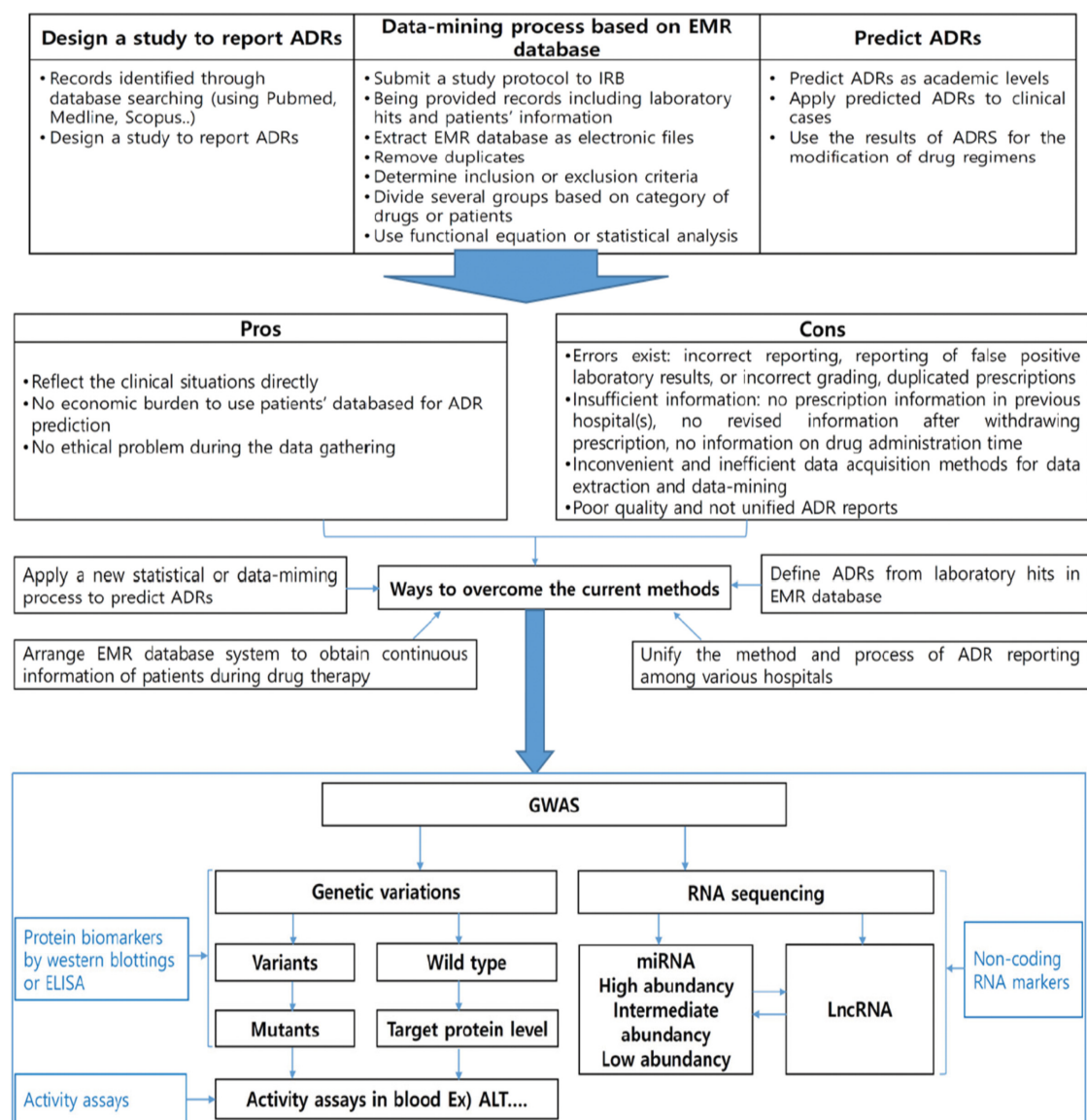
The analytical processes of the EMR database can be

Fig. 3. A schematic illustrating new biomarkers and methodologies necessary to improve electronic medical record (EMR)-based pharmacovigilance studies.

summarized into four steps: Institutional Review Board (IRB) approval; data extraction from the EMR database; data-mining of extracted dataset; and statistical analysis. The procedures for data extraction, analysis, and evaluation of potential ADRs are prerequisites to generate drug safety information. Clinical outcomes from hospitals and efficacy/safety data submitted by pharmaceutical companies can be adjusted to define the proper information on ADRs (Fig. 3).

Here, we focus on how to identify ADRs by using hospital data. First, the study design should be approved by the IRB of the respective hospital. The protocol submitted to the IRB includes the purpose of study, the researchers' certificate numbers taken from the IRB educational program, and a detailed description of the study. The descriptions should mention the following factors: administered drugs; drug administration period; patient information, such as age, gender, diseases, and exclusion criteria; required laboratory signals of the study; and analytical methods using acquired dataset. Secondly, after IRB approval, data extraction is conducted; data collection process means that the required dataset is extracted from EMR database in electro-medical team in hospitals. If necessary, ADR signals defined by World Health Organization (WHO), as well as potential signals in the EMR database, can be used. In particular, specific ADR hits and signals may raise hypotheses on safety information, which may affect regulatory decisions (12). In our data extraction from Seoul National University Hospital (SNUH), twenty laboratory signals, such as phosphate, glucose, hemoglobin A1c (HbA1c), blood urea nitrogen (BUN), serum creatinine (Scr), cholesterol, protein, albumin, uric acid, total bilirubin, aspartate transaminase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT), C-reactive protein (CRP), alpha-fetoprotein (AFP), white blood cells (WBC), absolute neutrophil count (ANC), hemoglobin test (Hb), hematocrit (Hct), platelet count (PLT), international normalized ratio (INR), sodium (Na^+), potassium (K^+), and calcium (Ca^{+2}), were retrieved. In addition, patients' disease, co-administered drugs, prescription patterns (e.g., dose, period, and route of administration), age, or gender can be extracted as candidates of general and potential signals from EMR database (12).

To assess the quality of dataset, the reporting number, rate, and percentage of serious adverse events should be considered (12). Using the extracted dataset, pharmacists and/or researchers initiate data-mining to predict ADRs. The data-mining and acquisition process exemplified in this article may be important for clinicians to understand ADRs. Moreover, the data processing and scoring systems are required to assess the ADR alert system, through which clinicians optimize medication schedules at prescription. The procedures of ADR data-mining and the acquisition methods using laboratory signal hits are also applicable to assess other effects. The scoring method is more advanced if other parameters, such as comorbidity, polypharmacy, gender, and age are combined, which allows us to obtain more reliable information on the severity, onset time and/or duration of ADRs, and incidence rate. The prevention of ADRs may ultimately contribute to the reduction of unnecessary healthcare costs.

To apply ADR reports to prevent medication errors and to improve the quality of pharmaceutical care, causality assessments of ADRs and the development of criteria for ADR reports are required. To accomplish this, the validation of the ADR reporting process and the provision of feedback to medical teams on the potential harmful effects of the prescriptions are required. However, there are several points to overcome when using EMR data for the identification of proper ADRs.

First, there may be intrinsic errors, such as incorrect reporting, reporting of false positive laboratory results, or incorrect grading (18). Patient compliance may also be an issue (e.g., gap between patients' drug administration and prescription) in the following situations: 1) duplicated prescriptions ordered from different departments, even within a single hospital; 2) no prescription information in previous hospital(s); 3) no revised information after prescription withdrawal; and/or 4) no information on drug administration time (e.g., exact time for administration of different drugs). In addition, laboratory signals may be affected by the duration time of drug treatment, although it is not easy to retrieve information on the exact treatment time or duration. In particular, treatment time and duration are critical factors for laboratory signals: a generalized hypothesis for the treatment or duration may cause confusion regarding the occurrence of drug-induced adverse effects. Incorrect reporting, reporting of false-positive laboratory results, and/or incorrect grading result in erroneous use of the EMR database. An example of incorrect reporting would be a description error, such as hypokalemia vs hyperkalemia. Matching laboratory data to a description can be a solution; this is not easy, may be time consuming, and sometimes impossible. The reporting of false-positive laboratory results may occur owing to improper specimen collection, inaccurate normalized values of laboratory report, and/or the occurrence of concurrent opposite symptoms. The errors should be identified and re-checked for the addition of information, as well as extracting and mining EMR datasets.

Second, the application of a data acquisition method using laboratory signals in patients' EMRs allows us to rapidly and conveniently monitor ADRs and other patient conditions (16,19). To create a data-acquisition method by using EMRs and its systematical application, such as patient laboratory signals, age, organ functions, pathological factors, co-administered drugs, and lifestyles should be

considered (15,20-22). Because this method is retrospective, laboratory signal hits have been used for analysis. However, there are limitations to the process of extracting the EMR database because of difficulty in using commercial database programs for data-mining (i.e., the necessity of customized programs). Despite calls for greater transparency (20-23), code lists have seldom been reported in published papers (21,22). Thus, one of possible reasons for the poor quality of reporting is the varied and inconsistent clinical coding. EMR studies adopt definitions of clinical entities. When patient information, including laboratory signals is provided from medical staff in hospitals, the lists of clinical codes are generally extracted and converted into ".csv" or "Excel" files. However, drug medication data vary and cannot be perfectly extracted by automated systems. Moreover, descriptions by clinicians are sometimes important factors to observe patient conditions during medication and should be extracted after communications between clinicians and medical informatics staff. For example, inconsistency in code selection and the resulting small errors in the analysis of laboratory signals in EMR data-mining process may cause large numbers of misclassified patients and a large degree of potential inaccuracy in ADR prediction, causing biased outcomes and errors affecting conclusions in unpredictable ways (24). In particular, clinical definitions may change over time during the observation period, which may necessitate changes in the code lists in patients' information.

To improve ADR prediction from the EMR database by using new analytical approaches, "prescription sequence symmetry analysis" (PSSA) may be used as a signal detection method (25). This method employs a simple algorithm, which is computationally rapid, and requires a minimal dataset of only three factors, such as drug name, date of supply, and a patient identifier. Another approach is the simultaneous consideration of the rate of drug prescription and ADR occurrence for ADR prediction (26). As the disproportionality of drug prescription and ADR reports has limitations for the prediction of sensitivity, it would be useful to create priority-based listings for signal detections in databases (26). In this method, receiver-operating characteristics curves are used, including the specificity and sensitivity of ADRs (Fig. 2).

LACK OF SENSITIVE AND SELECTIVE BIOMARKERS IN ROUTINE CHECK-UPS

Emerging evidence from numerous biomedical studies and advancements in scientific technology in recent decades have suggested the potential value of novel biomarkers in the prediction and/or diagnosis of ADRs. As ADRs affect many organs in humans and vary broadly in severity, diverse biological events (e.g., gene expression and signaling pathway activation) could be changed (27,28). This concept proposes the importance of the identification of novel targets and/or biomarkers based on elucidation of complex molecular mechanisms and their utilization in coordination with the EMR database analysis, to allow applications in clinical situations (29). The integration and understanding of the factors may enable us to provide directions for pharmacovigilance studies, as suggested in Table 1.

One of the problems in the use of routine EMR data is the limitation of traditional biomarkers for the proper prediction of ADRs. Circulating protein biomarkers have been used for the diagnosis and/or prognosis of ADRs, which associate with specific tissue damage, irrespective of the

Table 1. Potential new targets and/or biomarkers for pharmacovigilance studies

Category	Targets and/or biomarkers	Pharmacological effects and/or ADRs	References
Genetic variations (traditional)	CYP2D6	Drug metabolism (rapid or slow metabolizers)	(75,76)
	SLCO1B1	Statins (myopathy)	(77)
	VKORC1	Warfarin (anti-coagulant effect)	(78,79)
miRNAs (novel)	miR-122	Acetaminophen (hepatotoxicity)	(57)
	miR-27b, miR-298	Drug metabolism [possibly effects on drugs metabolized by CYP3A4 (e.g., benzodiazepines, antivirals and steroids)]	(50,51)
	miR-378	Drug metabolism [possibly effects on drugs metabolized by CYP2E1 (e.g., acetaminophen, isoniazid)]	(52,53)
	miR-122a, miR-422a	Bile acid synthesis (CYP7A1); possibly effects on ADRs affecting liver and biliary system, and/or responses to statins)	(54,82)
	miR-125b	Vitamin D3 metabolism (CYP24A1); possibly effects on ADRs affecting cancer susceptibility and/or calcium homeostasis	(55,83)
	miR-124, miR-18a-5p	Skin blistering reactions (SJS/TEN)	(60,61)
Secretory proteins (novel)	HMGB1	Tissue injury, immune response, acetaminophen (hepatotoxicity)	(35,38)

etiology of the diseases and/or drugs (30-34). For example, the measurement of ALT and AST activities for liver injury is an example. Although the traditional biomarkers of tissue injury have been continuously and widely used, several limitations exist: the markers may become elevated only when the tissues are significantly damaged (i.e., low sensitivity), and the markers may be produced by various organs and/or toxic stimuli (i.e., low specificity). In addition, the markers are incomplete for determining the precise mechanism of ADRs and/or the specific cell types affected. ALT often shows greater specificity than AST (31), but it has disadvantages, such as low sensitivity and possible alterations by other comorbid conditions. To overcome these limitations, the identification and utilization of novel biomarkers is necessary for future pharmacovigilance studies. Indeed, recent improvements in the discovery of new biomarkers based on genetic variations [i.e., genome-wide association studies (GWAS), non-coding RNAs, microRNAs], proteomics, gene network, and signal pathways have enabled us to understand the benefits of new ADR indicators (8,30).

In addition, the mechanistic approaches to the discovery of biomarkers can certainly be applied to various tissues and organs commonly affected by ADRs, such as the liver, in which toxicity may be specific to certain cell types (30). High-mobility group box 1 (HMGB1) has been considered as a prognostic biomarker for ADRs. It is released from necrotic hepatocytes and activated immune cells, which is an important link between cell death, inflammation, and the disease progression (30,35). It has been shown that HMGB1 isoforms were more sensitive than ALT for the prediction of DILI development and adverse reactions caused by hepatotoxicants, such as acetaminophen overdose (35). Therefore, despite the limitations [i.e., time-consuming diagnostic assays (MS/MS) or inability to distinguish between different acetyl and redox isoforms of HMGB1 (ELISA)] (35), HMGB1 is a promising candidate novel biomarkers. ADRs possess strong genetic predisposition, although it is difficult to discern the genetic components underlying any particular ADR. For example, the human leukocyte antigen (HLA) alleles are highly polymorphic and are associated with different types of ADRs (36). Others include genes encoding for drug metabolizing enzymes and drug transporters (37,38). Thus, the identification of genetic factors and the implementation of new genetic approaches contribute to the safer use of drugs, as shown in some clinical practice (37,38).

POTENTIAL BIOMARKERS FROM GENETICS APPLICATIONS

Transcriptomics from genome-wide association studies. Genome-wide association studies (GWAS) continue to be used to offer a more comprehensive view of drug responses and ADRs (39). GWAS for ADRs are characterized by smaller sample sizes than GWAS for common diseases; often, only dozens of cases and hundreds of controls are used, in comparison to GWAS for common diseases, which usually need thousands of cases and controls (39,40). ADRs are often related to immunological features, as many drugs or metabolites can interact with major histocompatibility (MHC) molecules, and these associations have been detected in the early GWAS of ADRs (39,41). Despite various adverse reactions of drug hypersensitivity, recent studies have reported the links of ADRs with loci outside of the MHC region, such as human leukocyte antigen (HLA) alleles (e.g., *NUDT15*, *EPHA5*, *RBMS3*, and *TCL1A*), which suggest that many ADRs might be involved in other reactions, as well as immunological phenomena (39).

Emerging roles of non-coding RNAs. The human genome encodes RNAs that do not translate to proteins, known as non-coding RNAs (ncRNAs). These comprise of microRNAs (miRNAs, a type of small non-coding RNAs), intronic RNAs, repetitive RNAs, and long noncoding RNAs (lncRNAs). Their functions are related to the control of the transcription, stability or translation of transcripts for the *de novo* synthesis of proteins, and affect the expression of other genes. Thus, the dysregulation of ncRNAs may facilitate the progression of various diseases (42,43). Emerging evidence have suggested that ncRNAs provide various regulatory functions in the cell, and that the RNA networks may be critical for the coordination of genomics and proteomics (44). Notably, GWAS of common diseases have shown that the most (> 90%) of the disease-associated genomic variants are noncoding variants (41,45), which indicates the potential of dysregulated ncRNAs as regulatory mediators causing the development and/or progression of various diseases. It is also of note that a majority (82%) of ADR-associated genomic variants exist in noncoding regions of the genome (39), which also supports the inclusion of ncRNAs in pharmacovigilance studies.

The most well-studied ncRNAs are miRNAs, but many other types of ncRNAs with various lengths and characteristics may also have roles in the regulation of cellular homeostasis and disease progression (42,46). The functional impacts of ncRNAs on human disease have been well described from the research to discover the abnormal expression patterns of miRNAs (46). For example, certain miRNAs can control various processes of tumorigenesis, neurodegenerative diseases, or cardiovascular disorders (46,47). In addition, a small nuclear RNA (snoRNA) and lncRNA functions are also impaired (46).

MicroRNAs as new potential biomarkers. MiRNAs control diverse biological processes through the post-transcriptional regulation of their target genes. Specifically,

miRNAs may modulate the expression of proteins accountable for the regulation of pharmacokinetics, which involves P450 metabolizing enzymes, and ABC or SLC transporters (48-53). Therefore, drug metabolism and disposition can be affected by miRNA-dependent alterations in gene expression and the consequent changes in biological functions. Human CYP3A4 is the most abundant in organs such as the liver and small intestine and metabolizes > 50% of drugs, such as benzodiazepines, antivirals, and steroids (50). miR-27b and miR-298 act directly on the 3′ untranslated region of CYP3A4 mRNA (51). Other types of cytochrome P450 genes are also regulated by certain miRNAs [e.g., the inhibition of CYP7A1 by miR-122a and miR-422a (54), and CYP24A1 by miR-125b (55)]. Therefore, miRNAs may become valuable biomarkers for the detection of ADRs, as well as targets for drug discovery.

Many types of toxicants alter the expression of miRNAs in target organs. miRNAs are considered to be stable in the plasma; thus several miRNAs originated from various tissues (e.g., the liver) can be released into the bloodstream (56). Thus, the dysregulation of miRNAs can be found in easily obtainable biological fluids. The advantages of using miRNAs, in combination with the early responses to toxic challenges and their stability, allow the molecules to act as novel biomarkers for drug safety assessment (57-61). Recent developments in the discovery of biomarkers for DILI support the view that the newly identified biomarkers have enabled us to overcome the limitations of traditional markers for the diagnosis and comprehension of the etiology of ADRs (30). Specifically, miR-122, as a liver-enriched miRNA, is a promising target. Circulating miR-122 has been shown to be specific for acute hepatocyte injury in acetaminophen overdose and is more sensitive for the early detection of liver injury than traditional tests (58,59). Further research is necessary to validate the utilization of miR-122 as a diagnostic or prognostic indicator of late-onset idiosyncratic DILI (30). Therefore, the precise understanding of the basis underlying miRNA biology in the context of drug responses would provide an opportunity to gain insight into the specific tissue damage and the pathogenesis of injury (30). This would also provide valuable information for drug development during preclinical testing and early phase human trials.

Biomarker discovery from proteomics applications.
Given the functional roles of proteins in most of cellular processes and the diverse proteomic patterns in response to environmental and/or chemical stress, including drugs, proteomics may also represent a new model for ADR assessment. A recent study applying large-scale proteomics of blood samples and pathway analysis discovered the unknown effects of torcetrapib on the immune and inflammatory functions, in addition to the changes in the endo-crine systems, indicating the improved assessment of drug safety through proteomic analysis (62). A comprehensive proteome scale approach was developed to predict drug-protein interactions, providing the information on ADRs as well as drug repositioning (63). In addition, a method was also suggested for the expectation of ADRs through the integration of protein-protein interaction (PPI) networks with drug structures, which showed that the integration significantly improved the prediction of ADRs (64). Compared with genomics, proteomic techniques may convey different information and offer the advantage of early detection of ADRs; thus, these approaches complement each other (65,66).

New information from gene network and signaling pathways.
Many complex traits for drug responses are associated with alterations in various biological pathways rather than single gene changes (67). A predictive framework was presented, in which gene expression data were captured into activity states of signal transduction circuits as sub-pathways connecting receptor proteins to the ultimate effectors activating reactions in the cell (67). These mechanism-based biomarkers may provide insight into the molecular basis of drug actions (67).

Considerations of genetic variations.
Advancements in genetic studies have allowed the mapping of individual genetic variations based on human genome sequencing (68). A number of reports suggested that genetic differences might be related to the progression of disease, responses to drugs, and ADRs (68,69). Recently, the use of the pharmacogenomic database has been considered to explain variations in biological events among individuals in the context of pharmacokinetics, pharmacodynamics, and adverse reactions (70,71). Indeed, the findings from pharmacogenetic studies have already been utilized in clinical applications for the purpose of future customized therapies (70,72). In addition, technical advances in laboratory works and bioinformatics have allowed genetic research outcomes to be applied to more complex genetic diseases (69), suggesting their relevance to the evaluation of drug responses and ADRs, especially in the areas of oncology, neurodegenerative and cardiovascular diseases, and metabolic disorders.

Although the research data showing the impact of pharmacogenomic biomarkers on drug responses have rapidly expanded, the clinical applications of evidence-based findings are small and thus only a small fraction of the studies identifying predictive novel biomarkers of ADRs has been translated to clinical practice (30,71). It is important to utilize the information on novel biomarkers for pharmacovigilance studies and to coordinate EMR data with multiple biomarker panels (genomic/non-genomic), which can be integrated through the application of different

'-omics' technologies (e.g., transcriptomics, proteomics, and metabolomics) and thereby provide more detailed information on drug responses (30).

Given the roles of the molecules in biological functions, genetic alterations can significantly affect pharmacokinetics, drug actions, and other factors involving therapeutic outcomes and ADRs (71,73). Of note, unpredictable and individual ADRs are considered a major risk factor for safe and successful therapy (69,74). Therefore, the discovery of genetic factors affecting ADRs would be of great help to reduce the medical problems and mortality for a subset of the population (69,75). Among drug-metabolizing enzymes, genetic variations in cytochrome P450s (CYPs) have been well studied. One good example is the CYP2D6 subtype, which belongs to the main enzymes responsible for drug biotransformation (i.e., 20%-25% of clinical drugs) (76). According to the differences in gene copy numbers of *CYP2D6*, patients are categorized as rapid or slow metabolizers (69,76). Compared with rapid metabolizers, slow metabolizers who have insufficient functional *CYP2D6* genes may be exposed to more frequent occurrence of ADRs (69,75). Polymorphisms in genes that encode transporters and drug receptors also contribute to therapeutic effects and ADRs (71,77-79). A genome-wide study identified that a single-nucleotide polymorphism (SNP) in *SLCO1B1*, which belongs to the solute carrier organic anion transporter family members, may be related to statin-induced myopathy (77). Therefore, the FDA in USA has revised many drug labels to provide information on pharmacogenetic biomarkers (73).

CONCLUSIONS

Advancements in technology have allowed us to expand the application of novel biomarkers into the EMR database. For example, the ultimate tool for the identification of ADRs in patients with different phases of diseases would require a combination of specific biomarkers of each disease and traditionally monitored EMR parameters (80). Key factors for the successful clinical applications of pharmacogenomic data include the development of clinical guidelines to guarantee consistent interpretation and prescribing practices, in addition to evidence-based information databases and the appropriate educational programs for decision making (71,76).

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

All authors declare that there is no conflict of interest.

REFERENCES

1. Wang, S.D. (2013) Opportunities and challenges of clinical research in the bigdataera: from RCT to BCT. *J. Thorac. Dis.*, **5**, 721-723.
2. Liu, M., Melton, B.L., Ator, G. and Waitman, L.R. (2017) Integrating medication alert data into a clinical data repository to enable retrospective study of drug interaction alerts in clinical practice. *AMIA Jt. Summits Transl. Sci. Proc.*, **2017**, 213-220.
3. Jaspers, M.W., Smeulers, M., Vermeulen, H. and Peute, L.W. (2011) Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J. Am. Med. Inform. Assoc.*, **18**, 327-334.
4. Murdoch, T. and Detsky, A. (2013) The inevitable application of big data to healthcare. *JAMA*, **309**, 1351-1352.
5. Springate, D.A., Kontopantelis, E., Ashcroft, D.M., Olier, I., Rosa Parisi, R., Chamapiwa, E. and Reeves, D. (2014) Clinical codes: an online clinical codes repository toimprove the validity and reproducibility of researchusing electronic medical records. *PLoS ONE*, **9**, e99825.
6. Cheetham, T.C., Lee, J., Hunt, C.M., Niu, F., Reisinger, S., Murray, R., Powell, G. and Papay, J. (2014) An automated causality assessment algorithm to detect drug-induced liver injury in electronic medical record data. *Pharmacoepidemiol. Drug Saf.*, **23**, 601-608.
7. Björnsson, E.S. (2016) Hepatotoxicity by drugs: the most common implicated agents. *Int. J. Mol. Sci.*, **17**, 224.
8. Kullak-Ublick, G.A., Andrade, R.J., Merz, M., End, P. and Benesic, A. (2017) Drug-induced liver injury: recent advances in diagnosis and risk assessment. *Gut*, **66**, 1154-1164.
9. Pirmohamed, M., James, S. and Meakin, S. (2004) Adverse drug reaction as cause of admission to hospital: prospective analysis of 18820 patients. *BMJ*, **329**, 15-19.
10. Lazarous, J., Pomeranz, B. and Corey, P.N. (1998) Incidence of adverse drug reactions in hospitalized patients. A meta-analysis of prospective studies. *JAMA*, **279**, 1200-1205.
11. Kohn, L.T., Corrgan, J.M. and Donaldson, M.S. (2000) To Error in Human; Building a Safer Health System, National Academy Press, Washington, DC.
12. Shin, J., Hunt, D.M., Suzuki, A., Papay, J.I., Beach, K.J. and Cheetham, T.C. (2013) Characterizing phenotypes and outcomes of drug-associated liver injury using electronic medical record data. *Pharmacoepidemiol. Drug Saf.*, **22**, 190-198.
13. Lin, K.J. and Schneeweiss, S. (2016) Considerations for the analysis of longitudinal electronic health records linked to claims data to study the effectiveness and safety of drugs. *Clin. Pharmacol. Ther.*, **100**, 147-157.
14. Andreea, F. and Marius, B. (2009) Adverse drug reactions in clinical practice: a causality assessment of a case of drug-induced pancreatitis. *J. Gastrointest. Liver Dis.*, **18**, 353-358.
15. Jha, A.K., Kuperman, G.J., Teich, J.M., Leape, L., SHea, B.,

Rittenberg, E., Burdick, E., Seger, D.L., Vander Vliet, M., Bates, D.W. (1998) Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J. Am. Med. Inform. Assoc.*, **5**, 305-314.

16. Tisdale, J.E. and Miller, D.A. (2005) Drug-Induced Diseases: Prevention, Detection and Management, American Society of Health-System Pharmacist Press, Bethesda, Maryland, pp. 1004-1007.

17. Lee, Y.H., Kang, U.G. and Park, R.W. (2008) Development of adverse drug event surveillance system using BI technology. *Int. J. Contents*, **9**, 106-113.

18. Miller, T.P., Ki, Y., Getz, K.D., Dudley, J., Burrows, E., Pennington, J., Ibrahimova, A., Fisher, B.T., Baqatell, R., Seif, A.E., Grundmeier, R. and Aplenc, R. (2017) Using electronic medical record data to report laboratory adverse events. *Br. J. Haematol.*, **177**, 283-286.

19. Pandit, A., Sachdeva, T. and Bafna, P. (2012) Drug-induced hepatotoxicity: a review. *J. Appl. Pharm. Sci.*, **02**, 233-243.

20. Mauben, M., Madigan, D. and Gerritis, C.M. (2007) The role of data mining in pharmacovigilance. *Exp. Opin. Drug Saf.*, **14**, 929-948.

21. Phansalkar, S., Hoffman, J.M., Nebeker, J.R. and Hurdle, J.F. (2007) Pharmacists versus nonpharmacists in adverse drug event detection: a meta-analysis and systematic review. *Am. J. Health Syst. Pharm.*, **64**, 842-849.

22. Phansalkar, S., Hoffman, J.M., Hurdle, J.F. and Patel, V.L. (2009) Understanding pharmacist decision making for adverse drug event (ADE) detection. *J. Eval. Clin. Pract.*, **15**, 266-275.

23. Tojios, S. and Fontana, R.J. (2011) Mechanisms of drug-induced liver injury: from bedside to bench. *Nat. Rev. Gastroenterol. Hepatol.*, **8**, 202-211.

24. Manuel, D.G., Rosella, L.C. and Stukel, T.A. (2010) Importance of accurately identifying disease in studies using electronic health records. *BMJ*, **341**, c4226.

25. Pratt, N., Cham, E.W., Choi, N.K., Kimura, M., Kimura, T., Kubota, K., Lai, E.C., Man, L.L., Ooba, N., Park, B.J., Sato, T., Shin, J.Y., Wong, I.C., Kao Yang, Y.H. and Roughead, E.E. (2015) Prescription sequence symmetry analysis: assessing risk, temporality, and consistency for adverse drug reactions across datasets in five countries. *Pharmacoepidemiol. Drug Saf.*, **24**, 858-864.

26. Scholl, J.H.G., van Hunsel, F.P.A.M., Hak, E. and van Puijenbroek, E.P. (2017) A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol. Drug Saf.*, **27**, 199-205.

27. Chen, X., Wang, Y., Wang, P., Lian, B., Li, C., Wamg, J., Li, X. and Jiang, W. (2015) Systematic analysis of the associations between adverse drug reactions and pathways. *Biomed. Res. Int.*, **2015**, 670949.

28. Kohonen, P., Parkkinen, J.A., Willinghaegen, E.L., Ceder, R., Wennerbern, K., Kaski, S. and Grafstrom, R.C. (2017) A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nat. Commun.*, **8**, 15932.

29. Koutkias, V.G. and Jaulent, M.-C. (2015) Computational approaches for pharmacovigilance signaldetection: toward integrated and semantically-enriched frameworks. *Drug Saf.*, **38**, 219-232.

30. Carr, D.F. and Pirmohamed, M. (2017) Biomarkers of adverse drug reactions. *Exp. Biol. Med. (Maywood)*, **243**, 291-299.

31. Xu, Q., Higgins, T. and Cembrowski, G.S. (2015) Limiting the testing of AST: a diagnostically nonspecific enzyme. *Am. J. Clin. Pathol.*, **144**, 423-426.

32. Alfirevic, A., Neely, D., Armitage, J., Chinoy, H., Cooper, R.G., Laaksonen, R., Carr, D.F., Bloch, K.M., Fahy, J., Hanson, A., Yue, Q.Y., Wadelius, M., Maitland-van Der Zee, A.H., Voora, D., Psaty, B.M., Paimer, C.N. and Pirmohamed, M. (2014) Phenotype standardization for statin-induced myotoxicity. *Clin. Pharmacol. Ther.*, **96**, 470-476.

33. Duma, R.J. and Siegel, A.L. (1965) Serum creatinine phohpkinase in acute myocardial infarction: diagnostic value. *Arch. Intern. Med.*, **115**, 443-451.

34. Kindermann, W. (2016) Creatine kinase levels after exercise. *Dtsch. Arztebl. Int.*, **113**, 344.

35. Wilke, R.A., Lin, D.W., Roden, D.M., Watkins, P.B., Flockhart, D., Zineh, I., Giacomini, K.M. and Krauss, R.M. (2007) Identifying genetic risk factors for serious adverse drug reactions:current progress and challenges. *Nat. Rev. Drug Discov.*, **6**, 904-916.

36. Redwood, A.J., Pavlos, R.K., White, K.D. and Phillips, E.J. (2018) Human leukocyte antigens: key regulators of T-cell mediated drug hypersensitivity. *HLA*, **91**, 3-16.

37. Wang, C.-W., Chung, W.-H. and Hung, S.-I. (2017) Genetics of adverse drug reactions. *eLS*, 1-10.

38. Lea, J.D., Clarke, J.I., McGuire, N. and Antoine, D.J. (2016) Redox-dependent HMGB1 isoforms as pivotal co-ordinators of drug-induced liver injury: mechanistic biomarkers and therapeutic targets. *Antioxid. Redox. Signal.*, **24**, 652-665.

39. Chan, S.L., Jin, S., Loh, M. and Brunham, L.R. (2015) Progress in understanding the genomic basis for adverse drug reactions: a comprehensive review and focus on the role of ethnicity. *Pharmacogenomics*, **16**, 1161-1178.

40. Nelson, M.R., Bacanu, S.A., Mosteller, M., Li, L., Bowman, C.E., Roses, A.D., Lai, E.H. and Ehm, M.G. (2009) Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. *Pharmacogenomics J.*, **9**, 23-33.

41. Pichler, W.J. (2003) Delayed drug hypersensitivity reactions. *Ann. Intern. Med.*, **139**, 683-693.

42. Ling, H., Fabbri, M. and Calin, G.A. (2013) MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.*, **12**, 847-865.

43. Matsui, M. and Corey, D.R. (2017) Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.*, **16**, 167-179.

44. Fu, X.D. (2014) Non-coding RNA: a new frontier in regulatory biology. *Natl. Sci. Rev.*, **1**, 190-204.

45. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R. and Stamatoyannopoulos, J.A. (2012) Systematic localization of

common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.

46. Esteller, M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861-874.

47. Croce, C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704-714.

48. Nakajima, M. and Yokoi, T. (2011) MicroRNAs from biology to future pharmacotherapy: regulation of cytochrome P450s and nuclear receptors. *Pharmacol. Ther.*, **131**, 330-337.

49. Yu, A.M. (2009) Role of microRNAs in the regulation of drug metabolism and disposition. *Exp. Opin. Drug Metab. Toxicol.*, **5**, 1513-1528.

50. Zanger, U.M. and Schwab, M. (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.*, **138**, 103-141.

51. Pan, Y.Z., Gao, W. and Yu, A.M. (2009) MicroRNAs regulate CYP3A4 expression via direct and indirect targeting. *Drug Metab. Dispos.*, **37**, 2112-2117.

52. Lu, Y. and Cederbaum, A.I. (2008) CYP2E1 and oxidative liver injury by alcohol. *Free Radic. Biol. Med.*, **44**, 723-738.

53. Mohri, T., Nakajima, M., Fukami, T., Takamiya, M., Aoki, Y. and Yokoi, T. (2010) Human CYP2E1 is regulated by miR-378. *Biochem. Pharmacol.*, **79**, 1045-1052.

54. Song, K.H., Li, T., Owsley, E. and Chiang, J.Y. (2010) A putative role of micro RNA in regulation of cholesterol 7alpha-hydroxylase expression in human hepatocytes. *J. Lipid Res.*, **51**, 2223-2233.

55. Komagata, S., Nakajima, M., Takagi, S., Mohri, T., Taniya, T. and Yokoi, T. (2009) Human CYP24 catalyzing the inactivation of calcitriol is post-transcriptionally regulated by miR-125b. *Mol. Pharmacol.*, **76**, 702-709.

56. Lin, H., Ewing, L.E., Koturbash, I., Gurley, B.J. and Miousse, I.R. (2017) MicroRNAs as biomarkers for liver injury: current knowledge, challenges and future prospects. *Food Chem. Toxicol.*, **110**, 229-239.

57. Marrone, A.K., Beland, F.A. and Pogribny, I.P. (2015) The role for microRNAs in drug toxicity and in safety assessment. *Exp. Opin. Drug Metab. Toxicol.*, **11**, 601-611.

58. Antoine, D.J., Dear, J.W., Lewis, P.S., Platt, V., Coyle, J., Masson, M., Thanacoody, R.H., Gray, A.J., Webb, D.J., Moggs, J.G., Bateman, D.N., Goldring, C.E. and Park, B.K. (2013) Mechanistic biomarkers provide early and sensitive detection of acetaminophen-induced acute liver injury at first presentation to hospital. *Hepatology*, **58**, 777-787.

59. Howell, L.S., Ireland, L., Park, B.K. and Goldring, C.E. (2018) MiR-122 and other microRNAs as potential circulating biomarkers of drug-induced liver injury. *Exp. Rev. Mol. Diagn.*, **18**, 47-54.

60. Sato, S., Ichihara, A., Jinnin, M., Izuno, Y., Fukushima, S. and Ihn, H. (2015) Serum miR-124 up-regulation as a disease marker of toxic epidermal necrolysis. *Eur. J. Dermatol.*, **25**, 457-462.

61. Ichihara, A., Wang, Z., Jinnin, M., Izuno, Y., Shimozono, N., Yamane, K., Fujisawa, A., Moriya, C., Fukushima, S., Inoue, Y., Shimozono, N., Yamane, K., Fujisawa, A., Moriya, C., Fukushima, S., Inoue, Y. and Ihn, H. (2014) Upregulation of miR-18a-5p contributes to epidermal necrolysis in severe

drug eruptions. *J. Allergy Clin. Immunol.*, **133**, 1065-1074.

62. Williams, S.A., Murthy, A.C., DeLisle, R.K., Hyde, C., Malarstig, A., Ostroff, R., Weiss, S.J., Segal, M.R. and Ganz, P. (2018) Improving assessment of drug safety through proteomics: early detection and mechanistic characterization of the unforeseen harmful effects of torcetrapib. *Circulation*, **137**, 999-1010.

63. Zhou, H., Gao, M. and Skolnick, J. (2015) Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.*, **5**, 11090.

64. Huang, L.C., Wu, X. and Chen, J.Y. (2013) Predicting adverse drug reaction profiles by integrating protein interaction networks with drug structures. *Proteomics*, **13**, 313-324.

65. Ge, F. and He, Q.Y. (2009) Genomic and proteomic approaches for predicting toxicity and adverse drug reactions. *Exp. Opin. Drug Metab. Toxicol.*, **5**, 29-37.

66. Wilkins, M.R. (2002) What do we want from proteomics in the detection and avoidance of adverse drug reactions. *Toxicol. Lett.*, **127**, 245-249.

67. Amadoz, A., Sebastian-Leon, P., Vidal, E., Salavert, F. and Dopazo, J. (2015) Using activation status of signaling pathways as mechanism-based biomarkers to predict drug sensitivity. *Sci. Rep.*, **5**, 18494.

68. Frueh, F.W. (2010) Real-world clinical effectiveness, regulatory transparency and payer coverage: three ingredients for translating pharmacogenomics into clinical practice. *Pharmacogenomics*, **11**, 657-660.

69. Ventola, C.L. (2011) Pharmacogenomics in clinical practice: reality and expectations. *P T*, **36**, 412-450.

70. Madian, A.G., Wheeler, H.E., Jones, R.B. and Dolan, M.E. (2012) Relating human genetic variation to variation in drug responses. *Trends Genet.*, **28**, 487-495.

71. Ventola, C.L. (2013) Role of pharmacogenomic biomarkers in predicting and improving drug response: part 1: the clinical significance of pharmacogenetic variants. *P T*, **38**, 545-560.

72. Trent, R.J. (2010) Pathology practice and pharmacogenomics. *Pharmacogenomics*, **11**, 105-111.

73. Ma, J.D., Lee, K.C. and Kuo, G.M. (2012) Clinical application of pharmacogenomics. *J. Pharm. Pract.*, **25**, 417-427.

74. Squassina, A., Manchia, M., Manolopoulos, V.G., Artac, M., Lappa-Manakou, C., Karkabouna, S., Mitropoulos, K., Del Zompo, M. and Patrinos, G.P. (2010) Realities and expectations of pharmacogenomics and personalized medicine: impact of translating genetic knowledge into clinical practice. *Pharmacogenomics*, **11**, 1149-1167.

75. Vogenberg, F.R., Isaacson Barash, C. and Pursel, M. (2010) Personalized medicine: part 1: evolution and development into theranostics. *P T*, **35**, 560-576.

76. Gervasini, G., Benitez, J. and Carrillo, J.A. (2010) Pharmacogenetic testing and therapeutic drug monitoring are complementary tools for optimal individualization of drug therapy. *Eur. J. Clin. Pharmacol.*, **66**, 755-774.

77. Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M. and Collins, R. (2008) SLCO1B1 variants and statin-induced myopathy--a genome-wide study. *N. Engl. J. Med.*, **359**, 789-799.

78. Flockhart, D.A., O'Kane, D., Williams, M.S., Watson, M.S., Flockhart, D.A., Gage, B., Gandolfi, R., King, R., Lyon, E.,

Nussbaum, R., O'Kane, D., Schulman, K., Veenstra, D., Williams, M.S., Watson, M.S.; ACMG Working Group on Pharmacogenetic Testing of CYP2C9, VKORC1 Alleles for Warfarin Use (2008) Pharmacogenetic testing of CYP2C9 and VKORC1 alleles for warfarin. *Genet. Med.*, **10**, 139-150.

79. Scott, S.A. (2011) Personalizing medicine with clinical pharmacogenetics. *Genet. Med.*, **13**, 987-995.

80. Taneja, I., Reddy, B. and Damhorst, G. (2017) Combining biomarkers with EMR data to identify patients in different phases of sepsis. *Sci. Rep.*, **7**, 10800.

81. Crews, K.R., Hicks, J.K., Pui, C.H., Relling, M.V. and Evans, W.E. (2012) Pharmacogenomics and individualized medicine: translating science into practice. *Clin. Pharmacol. Ther.*, **92**, 467-475.

82. Jiang, X.Y., Zhang, Q., Chen, P., Li, S.Y., Zhang, N.N., Chen, X.D., Wang, G.C., Wang, H.B., Zhuang, M.Q. and Lu, M. (2012) CYP7A1 polymorphism influences the LDL cholesterol-lowering response to atorvastatin. *J. Clin. Pharm. Ther.*, **37**, 719-723.

83. Zhu, M., Qiu, S., Zhang, X., Wang, Y., Souraka, T.D.M., Wen, X., Liang, C. and Tu, J. (2018) The associations between CYP24A1 polymorphisms and cancer susceptibility: a meta-analysis and trial sequential analysis. *Pathol. Res. Pract.*, **214**, 53-63.