Epigenetics and Complex Disease: From Etiology to New Therapeutics

Carolyn Ptak and Arturas Petronis

The Krembil Family Epigenetics Laboratory, Center for Addiction and Mental Health, Toronto, Ontario, Canada M5T 1R8; email: arturas_petronis@camh.net

Annu. Rev. Pharmacol. Toxicol. 2008. 48:257–76

First published online as a Review in Advance on September 17, 2007

The *Annual Review of Pharmacology and Toxicology* is online at http://pharmtox.annualreviews.org

This article's doi:
10.1146/annurev.pharmtox.48.113006.094731

0362-1642/08/0210-0257$20.00

Key Words

DNA methylation, histone modification, epigenetic

Abstract

Epigenetics is a new development in complex non-Mendelian disease, which may not only uncover etiologic and pathogenic mechanisms but may also provide the basis for the development of medications that would target the primary epigenetic causes of such diseases. Such epigenetic drugs would be novel, potentially possessing substantially higher therapeutic potential and a much lower rate of adverse effects in comparison to current symptomatic treatments. A collection of epigenetic drugs already exist at various stages of development and, although their effectiveness has yet to be maximized, they show great promise in the treatment of cancer, psychiatric disorders, and other complex diseases. Here we present a review of the epigenetic theory of complex disease and an evaluation of current epigenetic therapies, as well as predictions of the future directions in this expanding field.

RELEVANCE OF EPIGENETICS TO COMPLEX DISEASE

Complex disease: conditions caused by a combination of genetic, epigenetic, and environmental factors that do not follow Mendelian/monogenic inheritance

DNMT: DNA methyltransferase

Histone: basic proteins around which DNA is wrapped in higher-order chromatin structures

HAT: histone acetyltransferase

HDAC: histone deacetylase

MBD: methyl-binding domain

MeCP2: methyl-CpG-binding protein 2

Epigenetics refers to the regulation of various genomic functions, including gene expression, that are brought about by heritable, but potentially reversible, changes in DNA methylation and various modifications of histones (acetylation, methylation, phosphorylation, etc.) (1). The two epigenetic mechanisms work in concert, with alterations in DNA modification affecting chromatin conformation and vice versa. In humans and animals, methylation of DNA occurs at the C^5 position of cytosines, primarily within cytosine/guanine dinucleotides (CpG), which are established and maintained by the DNA methyltransferase (DNMT) family of enzymes. DNA is wrapped around octamers of basic histone proteins (H2A, H2B, H3, and H4), forming higher-order nucleosome structures. Modification of these proteins, such as acetylation, methylation, phosphorylation, ubiquitination, etc., control chromatin states, which can be open (transcriptionally active) or closed (inactive). Among numerous other histone modification enzymes, histone acetyltransferases (HATs) acetylate lysine residues on the N-terminal tail of histone proteins. This neutralizes the positive charge of the protein, thereby decreasing its affinity for DNA and leading to a looser interaction (2) that creates an open chromatin structure and increases accessibility for the transcription machinery. In contrast, histone deacetylases (HDACs) remove acetyl groups, which results in condensed chromatin and gene inactivation (3). Proteins with N-terminal methyl-CpG-binding domains (MBD), such as methyl-CpG-binding protein 2 (MeCP2), can bind to methylated sites on DNA and complex with HDACs and the corepressor Sin3a. This leads to histone deacetylation and the silencing of genes downstream from the methylated CpG site. The effects of histone methylation depend on the specific lysine or arginine that is modified, and can also result in either gene activation or repression (4).

Epigenetic studies of various species—from *Escherichia coli* and yeast to animals and humans—demonstrated that epigenetic regulation is critically important for the normal functioning of genomes. Cells can operate normally only if both DNA sequence and epigenetic components of the genome function properly; epigenetically misregulated genes, despite impeccable DNA sequences, can be harmful and cause disease. To date, the role of epigenetic factors has been primarily investigated in rare pediatric syndromes (5) and malignant transformation of cells in cancer (6, 7). More importantly, epigenetics can be highly relevant to various complex non-Mendelian diseases, as epigenetic mechanisms allow for integrating a variety of apparently unrelated clinical, epidemiological, and molecular data into a new theoretical framework.

There are three fundamental points that enable us to consider epigenetic factors as etiological candidates in complex disease. First, the epigenetic status of genes is more dynamic in comparison to DNA sequence and can be altered by developmental programs and the environment of the organism (8); furthermore, epigenetic changes may occur even in the absence of evident environmental differences, i.e., owing to stochastic reasons (9). Second, some epigenetic signals can be transmitted along with DNA sequence across the germline generations, i.e., such signals exhibit partial meiotic stability (10). Third, epigenetic regulation is critical for normal genomic function, such

as segregation of chromosomes in mitosis, inactivation of parasitic DNA elements, and regulation of gene activity (34). Partial epigenetic stability, or metastability, and the primary role in controlling activities of DNA sequences can shed new light on non-Mendelian irregularities of complex diseases.

Monozygotic (MZ) twins: genetically identical twins arising from a single zygote

Discordance of identical [monozygotic (MZ)] twins has been one of the hallmarks of complex non-Mendelian disease. Concordance of MZ twins reaches only ~15% in breast cancer, 20% in ulcerative colitis, 25%–30% in multiple sclerosis, 25%–45% in diabetes, 50% in schizophrenia, and 40%–70% for Alzheimer's disease (11). Discordance of MZ twins traditionally has been explained by the differential effect of environmental factors, which supposedly produce disease in one of the two genetically predisposed co-twins (12). Identification of such factors has been very difficult, and thus far only a couple environmentally derived disease risk factors have been identified (e.g., smoking in lung cancer, diet in cardiovascular diseases). The epigenetic explanation of MZ twin discordance is that due to the partial stability of epigenetic factors, a substantial degree of disease-relevant epigenetic dissimilarity can be accumulated in such twins (13, 14). Epigenetic differences in identical twins may reflect differential exposure to a wide variety of environmental factors. For example, intake of folic acid affects both the global methylation level in the genome and regulation of imprinted genes (15, 16). During pregnancy, maternal dietary methyl supplements increase DNA methylation and change methylation-dependent epigenetic phenotypes in mammalian offspring (17, 18). There could be numerous environmental effects, including even maternal behavior (8), that cause some epigenetic "trace."

One of the important peculiarities of complex disease is sexual dimorphism—differential susceptibility to a disease in males and females. Multiple sclerosis, rheumatoid arthritis, Crohn's disease, panic disorder, structural heart disease, and hyperthyroidism are more common in females, whereas males are more often affected with autism, Hirschsprung's disease, ulcerative colitis, Parkinson's disease, alcoholism, allergies, and asthma (especially at a young age) (19). In psychiatric diseases, such as Alzheimer's disease, schizophrenia, alcoholism, and mood and anxiety disorders, psychopathology exhibits a number of differences between the sexes in rates of illness as well as the course of disease (20). It is important to note that effects of gender in complex diseases cannot be explained by sex chromosome-linked genes; in fact, effects of gender quite often have been detected in genetic linkage and association studies on autosomal chromosomes (21). Although sex hormones have been the usual "culprit" in the explanation of gender effects in complex diseases, there are no specific mechanisms proposed as to how such hormones predispose to or protect from a disease. The gender-specific effects in genetic linkage and association studies suggest that chromosomes and individual genes can be the target of sex hormones. Although such hormones cannot change DNA sequence, they can be potent modifiers of epigenetic status, which controls gene expression and various other genomic activities. It is known that hormones, including sex hormones, can control gene expression via epigenetic modifications, and therefore it can be hypothesized that differential susceptibility to complex disease in males and females is mediated by sex hormone–induced differences in epigenetic regulation of genes (21).

Epimutation: epigenetic changes that cause or predispose an organism to a disease

In some complex diseases, the risk to offspring depends on the sex of the affected parent. For example, asthma, bipolar disorder, and epilepsy are more often transmitted from the mother, whereas type 1 diabetes seems to be more often transmitted from the affected father (11). Parent-of-origin-dependent clinical differences have also been detected in schizophrenia (22). Molecular genetic studies, although rarely performed in a gender-specific fashion, also reveal parental origin effects in a wide variety of phenotypes, such as obesity (23), Alzheimer's disease (24), atopy and asthma (25), autism (26), and major psychosis (27). One of the most common mechanisms of parent-of-origin effects is genomic imprinting (28). The essence of genomic imprinting consists of the differential epigenetic modification of genes depending on their parental origin (29). Disruption of the normal imprinting pattern often causes diseases that affect cell growth, development, and behavior (30).

Whereas explanation of all the above non-Mendelian features is based on the partial epigenetic stability in somatic cells, there is also an interesting perspective on the role and outcomes of the partial epigenetic stability during the maturation of the germline. The meiotic epigenetic metastability allows for rethinking of the issue of familiality (minor proportion of all cases) and sporadicity (overwhelming majority of the cases) in complex disease. Extrapolating from the intergenerational dynamics of epigenetic regulation in mammals (31), it can be hypothesized that disease epimutations may develop in two possible ways: (*a*) they regress toward the norm in the germline of an affected individual and his/her offspring will not be affected (such cases are treated as sporadic) or (*b*) they persist across generations and present as quasi-Mendelian familial cases (32).

The epigenetic model of complex disease can be imagined as a chain of aberrant epigenetic events that begins with a pre-epimutation, a primary epigenetic problem that takes place during the maturation of the germline; pre-epimutation increases the risk for the disease but is not necessarily sufficient to cause the disease. The misregulation can be tolerated to some extent, and age of disease onset may depend on the effects of tissue differentiation, stochastic factors, hormones, and probably some external environmental factors (nutrition, infections, medications, addictions, etc.) (32–34). It may take decades to reach a critical threshold beyond which the genome, cell, or tissue is no longer able to function normally. Only some predisposed individuals will reach the threshold of epigenetic misregulation and acquire phenotypic changes that meet the diagnostic criteria for a clinical disorder. Severity of epigenetic misregulation may fluctuate over time, which in clinical terms is called remission and relapse. In some cases, "aging" epimutations may slowly regress back to the norm. For example, in major psychosis, this is seen as fading psychopathology or even partial recovery, which is consistent with age-dependent epigenetic changes in the genome (35). The same applies to asthma (36) and attention deficit and hyperactivity syndrome (37).

To date, epigenetic factors in complex disease have been poorly investigated, with the exception of cancer. The epigenomes of cancer cells commonly undergo large-scale alterations in DNA methylation, with global hypomethylation and promoter-specific hypermethylation (38) linked to aberrant patterns of histone modification (39). Genes involved in various cellular pathways may become misregulated, but epigenetic silencing of tumor suppressor genes, such as the gene encoding the cell

cycle inhibitor, $p16^{INK4a}$, and the DNA repair genes, *BRCA1* and *hMLH1*, which are the two most extensively studied examples of suppressor genes in cancer (40). Current estimates suggest that the average tumor will contain approximately 100–400 hypermethylated promoter regions (40). In addition to hypermethylation, global hypomethylation is often observed in cancer cells (41, 42); it is believed to cause a decrease in genomic stability and the formation of abnormal chromosome structures. Not surprisingly, in addition to aberrant DNA methylation changes, histone modification changes have been detected in malignant cells (39, 43). Despite the consistency in epigenetic changes in cancer, it is not clear which epimutations are primary causes and hallmarks of early-stage malignant transformation and which ones represent effects of the primary causes (44, 45). Without such information, etiological treatment of cancers that target such genes is not possible.

In addition to cancer, epigenetic studies of psychiatric diseases are now underway. The maintenance DNMT, DNMT1, has been shown to be upregulated in GABAergic medium spiny neurons in layers I and II of the cerebral prefrontal cortex in schizophrenia and bipolar disorder patients with psychosis (46). An increase in DNMT1 levels, along with a decrease in reelin and glutamic acid decarboxylase 67 (GAD67), also occurs in GABAergic medium spiny neurons of the caudate nucleus and putamen in schizophrenia patients (46). In autism studies, a substantial proportion of postmortem brain samples from autistic individuals revealed monoallelic or highly skewed allelic expression of $GABA_A$ receptor subunit (GABR) genes, whereas such genes were biallelically expressed in control brain samples (47). Rett syndrome, an X-linked neurodevelopmental disorder, has been shown to result from a mutation to the MeCP2 gene, the protein product of which represses gene transcription by binding to 5-methylcytosine residues (48). Fragile X syndrome has been linked to epigenetic silencing and loss of expression of the fragile X mental retardation 1 (FMR1) gene owing to expansion of a CGG repeat in its 5′-untranslated region (49). Although the underlying epimutations remain unknown in most complex diseases, many epigenetic therapeutic agents have already been developed. Several of these compounds are progressing through clinical trials or have become approved treatments for particular conditions, as described below.

EPIGENETIC THERAPEUTICS: CLASSES AND MECHANISMS OF EPIGENETIC DRUGS

DNA Methylation Inhibitors

Two classes of DNA methylation inhibitors exist, nucleoside analogs and non-nucleoside analogues, both of which aim to reactivate aberrantly silenced genes. Nucleoside analogues consist of either a ribose or deoxyribose moiety fused to a modified cytosine ring. Kinases phosphorylate the nucleoside into a nucleotide and, once it is incorporated into DNA, it prevents methylation. Under normal circumstances DNMTs flip cytosine rings out from newly synthesized DNA, forming an intermediate complex and incorporating a methyl group from *S*-adenosyl-L-methionine (Ado-Met) onto the C^5 position; the enzyme is then released (50). The

presence of a modification at this position causes the enzyme to become trapped in a DNA-protein adduct, preventing further methylation of progeny DNA by depleting genomic DNMT stores (51).

Some of the earlier nucleoside analogues that were synthesized in the 1960s, namely 5-azacytidine (5-aza-CR) and 5-aza-2′-deoxycytidine (5-aza-CdR, decitabine), have been extensively studied and can induce cellular differentiation/inhibit DNA methylation at micromolar doses (52). Despite showing significant toxicity at high doses, low-dose decitabine demonstrated a 49% overall response rate and a 20% complete response rate in elderly, high-risk myelodysplastic syndrome (MDS) patients (53). A recent study has shown treatment with decitabine to be superior to supportive care for global health status, physical functioning, fatigue, and dyspnea in elderly patients with MDS (54). Two disadvantages of the original analogues are their instability in aqueous solutions and their susceptibility to deactivation by cytidine deaminase; these issues do not apply to the recently discovered methylation inhibitor, Zebularine [1-β-$_D$-ribofuranosyl-2(1H)-pyrimidinone]. Zebularine inhibits cytidine deaminase, is stable in aqueous solution, can be administered orally, and preferentially targets cancer cells. Its clinical development is limited by its higher dosage (55, 56) and poor bioavailability, as observed in rats, mice, and monkeys (57). One question facing all nucleoside analogues is whether their effects are due to demethylation or the cytotoxic effects caused by depletion of DNMTs. Further study is required to resolve this concern and to reduce the toxicity associated with this drug class.

Non-nucleoside analogues are small molecules that either bind the active site of DNMTs or prevent expression of the enzymes without incorporating into DNA; they are somewhat less toxic than nucleoside analogues. The antisense oligonucleotide inhibitor of human DNMT1, MG98, inhibits the translation of DNMT1 mRNA by targeting its 3′ untranslated region. A partial response to MG98 was documented in a patient with renal cell carcinoma in a Phase I trial (58); however, a Phase II trial was less successful (59).

Another small molecule, RG108, binds the catalytic pocket of human DNMT1 and directly inhibits its activity. Despite showing concentration-dependent demethylation of genomic DNA with little cytotoxicity (59), RG108 has not entered clinical trials. Similarly, other non-nucleoside analogues exist, such as hydralazine, the anesthetic procaine, and the antiarrhythmic procainamide. Few have been studied beyond Phase I, with the exception of hydralazine, which has proven to be a well-tolerated, effective demethylator in a Phase I study. It reactivates tumor suppressor genes in patients with untreated cervical cancer without affecting global DNA methylation (60).

Some promising methylation inhibitors occur naturally, for example, the psammaplins, which are bromotyrosine derivatives extracted from the sponge *Pseudoceratina purpurea*, and which can inhibit both histone deacetylases and DNMTs. Psammaplin A, in particular, has demonstrated potent histone deacetylase inhibitory activity (61). The green tea extract EGCG [(-) epigallocatechin-3-gallate] is a competitive inhibitor of DNMT, with potential as an anticancer treatment (62); neither psammaplins nor EGCG have entered clinical trials.

Histone Deacetylase Inhibitors

The histone deacytylase inhibitors (HDACi) comprise a large, diverse class of drugs. Each compound contains a different functional group with HDAC-inhibitory activity, leading to restoration of gene function or acetylation of transcription factors/tumor suppressors. Although their mechanisms of action are not entirely clear, HDACi have been shown to act via selective induction of apoptosis in tumor cells (63), elevation of reactive oxygen species production (64), the mitochondrial (intrinsic) pathway (65), the death receptor (extrinsic) pathway (66), activation of BH3-only proteins (Bcl-2 family) (67), or by causing cell cycle arrest, usually at G1/S (68). Many HDACi are also believed to have radiosensitizing properties (69). There are four main classes of HDACi: short-chain fatty acids (SCFAs), hydroxamic acids, benzamides, and cyclic tetrapeptides, as well as a few miscellaneous drugs, most of which are analogues of the cyclic tetrapeptides.

SCFA: short-chain fatty acid

HDACi: histone deacetylase inhibitor

Short-chain fatty acids. These drugs are already used to treat a variety of diseases. The original SCFAs are valproate, an antiepileptic and mood stabilizer, and butyrate, which is naturally produced in the colon and considered the most potent antiproliferative drug in this class (70). The SCFAs, in general, are relatively weak inhibitors of HDACs, as their effective doses are in the millimolar range. They also show pleiotropic effects on other enzymes, have low bioavailability, and low specificity (71). Despite these qualities, valproate is often used in combination with other drugs as part of cancer therapy regimes. Valproate has been shown to potently augment the tumoricidal activity of gemtuzumab ozogamicin toward acute myeloid leukemic (AML) cell lines and primary AML blasts (72); preclinical and Phase I and II studies of valproate in combination with decitabine have supported this approach for treatment of advanced leukemia (73); Phase II studies have shown that the combination of valproate and all-*trans* retinoic acid is useful in the treatment of myelodysplastic syndrome and relapsed or refractory AML (74). Tributyrin, an analogue of sodium butyrate, has recently been synthesized and shown to inhibit proliferation and induce differentiation and apoptosis of leukemia cells in vitro (75). As their mechanisms are elucidated, novel SCFAs will likely emerge and may be useful, alone or in combination, in the treatment of various leukemias and other conditions.

Hydroxamic acids. Hydroxamates are promising, potent HDACi that show pharmacologic activity at micromolar concentrations or lower. Key examples are Trichostatin A (TSA), a *Streptomyces hydroscopicus*–derived antibiotic, and suberoylanilide hydroxamic acid (SAHA), a synthetic anticancer agent that is undergoing clinical trials; in both cases, their HDACi properties were not discovered immediately. The anticancer properties of TSA were originally attributed to its action as an inducer of cell differentiation (76), and its ability to cause cell cycle arrest at low concentrations; it was later discovered that the effects of TSA were due to its potent and specific inhibition of HDACs (77). TSA radiosensitizes tumor cells (69) and has recently been shown to induce global and gene-specific DNA demethylation in human bladder and breast carcinoma cell lines (78). In preclinical studies, TSA produced a significant,

concentration-dependent antiproliferative effect in an endometrial stromal cell line. This effect, combined with its potential to restore gene dysregulation via chromatin remodeling, suggests that TSA is a promising potential treatment for endometriosis (79).

SAHA was developed as a more potent (i.e., induces murine erythroleukaemia cell differentiation at micromolar concentrations), yet less toxic, derivative of hexamethylene bisacetamide (HMBA), an acid capable of inducing transformed cell growth arrest and cell death (80). It inhibits HDACs by binding to a zinc ion within their catalytic domain (81); it has also been proposed that the drug induces cell death through cleavage of Bid, a BH3-only Bcl-2 protein (64). SAHA has demonstrated anticancer activity in Phase I studies of refractory hematologic and solid tumors (82). Recently, a Phase II trial in refractory cutaneous T cell lymphoma (CTCL) showed that SAHA can cause a partial response, as well as relief from the associated pruritis (83); in 2006, an FDA approval was issued for the use of SAHA (Vorinostat, Zolinza) to treat persistent skin lesions arising from CTCL. Other mechanisms responsible for the actions of SAHA have been proposed. In addition to its HDACi effects, SAHA may act as a protein acetylase and interfere with the function of chaperones and their client proteins via disruption of tubulin deacetylases (such as HDAC6) (84). Further study of the mechanisms of SAHA may allow for higher tumor-type selectivity and better design of combination therapy strategies.

Although SAHA is becoming well established, some newer hydroxamic acid-containing HDACi are progressing toward clinical usage; the agents that are farthest along in clinical trials are LAQ824 and PDX-101. PDX-101 is a sulfonate that inhibits HDACs at low-nanomolar concentrations, and it has entered Phase I clinical trials (85). LAQ824 works well in combination with 13-*cis*-retinoic acid, showing antitumor activity in human malignant melanoma (86). It causes hyperacetylation, induces apoptosis, and has entered clinical trial (87).

Benzamides. These drugs are amide derivatives of benzoic acid, two of which have reached the clinical trial phase. CI-994 (N-acetyldinaline) can be administered orally at micromolar drug concentrations, and it has shown in vitro and in vivo antitumor activity; its mechanism of action is thought to involve inhibition of histone deacetylation and cell cycle arrest (88). CI-994 is often used in combination with other treatments, such as paclitaxel and carboplatin, and has shown antitumor activity in several Phase I studies (89, 90). However, a recent Phase II study revealed that CI-994 in combination with gemcitabine showed no advantage over gemcitabine alone in the treatment of advanced pancreatic carcinoma (91). Considering the promising results of Phase I trials and the fact that CI-994 has synergy and antitumor activity with gemcitabine preclinically, Phase II testing should continue on different tumor types.

Another benzamide, MS-275, acts by binding the zinc ion within the HDAC catalytic domain and has demonstrated radiosensitizing properties on human prostate carcinoma and glioma cell lines (92). In Phase I trials, orally administered MS-275 effectively inhibited HDACs in patients with advanced myeloid leukemias (93), and was well tolerated in patients with advanced and refractory solid tumors or lymphoma, causing histone deacetylase inhibition in peripheral-blood mononuclear cells (94). It

is also capable of activating the retinoic acid receptor β2, which produces antitumor effects in cases of prostate cancer and renal cell carcinoma (95). Similar to CI-994, MS-275 is now being studied in Phase II clinical trials.

Cyclic tetrapeptides. Cyclic tetrapeptides contain an α-epoxyketone group that allows them to alkylate, thus deactivating, the HDAC catalytic pocket. The epoxyketone moiety can be substituted, as analogs of this class contain similar functional groups, such as pentafluoroethyl and trifluoromethyl ketones, sulfydryls, or retrohydroxamate, which also target the catalytic pocket. HDAC-inhibitory activity of these electrophilic ketones depends on the degree of hydration of the ketone, which chelates the zinc ion (96). The most famous cyclic tetrapeptide, depsipeptide, has completed several Phase I studies; it has shown antitumor activity in patients with refractory neoplasms (97), cutaneous and peripheral T cell lymphoma (98), chronic lymphocytic leukemia, and acute myeloid leukemia (99), and it is well tolerated in children with recurrent or refractory solid tumors (100). In a recent Phase II trial, depsipeptide was ineffective in cases of metastatic renal cell cancer (101); however, it is also undergoing Phase II studies for the treatment of T cell lymphoma (98).

Aplidine, a cyclic depsipeptide isolated from the marine tunicate *Aplidium albicans*, was well tolerated in Phase I studies involving patients with refractory tumors, and it is now entering Phase II trials in hematological malignancies and solid tumors (102). Trapoxins A and B, derived from the fungus *Helicoma ambiens*, are hydrophobic cyclotetrapeptides that cause irreversible HDAC inhibition by binding covalently to the HDAC via its epoxide group. A low concentration (<10 nM) is required for this effect (103), although both compounds are toxic and instable, thus they have not entered clinical trials.

Several analogs of cyclotetrapeptides exist, for example, apicidin (OSI-2040), which lacks the terminal α-epoxyketone. The potent antifungal, antiprotozoal metabolite was isolated from two *Fusarium* species, ATCC 74289 and ATCC 74322 (104). It has demonstrated antiproliferative activity against various cancer cell lines via selective induction of genes involved in the cell cycle and cell morphology, namely $p21^{WAF1/Cip1}$ (105). Hybrids of trichostatin and trapoxin, the cyclic hydroxamic acid–containing peptides (CHAPs), contain a hydroxamate side chain attached to the cyclic peptide core in place of the epoxyketone. These compounds reversibly inhibit HDACs at low-nanomolar concentrations of drug, and by changing the methylene chain length, many different CHAPs can be created, each with its own properties (106).

Psychotropic drugs with epigenetic actions. Several psychotropics that are currently in clinical use exhibit epigenetic effects in addition to their commonly understood modes of action. These epigenetic mechanisms provide insight into drug characteristics, such as the necessary duration of treatment, dependence, and tolerance.

The SCFA valproate is commonly utilized as a mood stabilizer for the treatment of bipolar disorder. Valproate acts epigenetically to induce gene expression via inhibition of HDACs and its ability to perform DNA demethylation (108). The mechanisms

underlying numerous other actions of valproate are currently unknown, but several hypotheses exist. It has been reported to block voltage-dependent Na^+ and T-type Ca^{2+} channels, elevate whole-brain GABA levels, decrease aspartate levels in the brain, and reduce the turnover of dopamine (107).

Fluoxetine, a selective serotonin reuptake inhibitor (SSRI) antidepressant, has been shown to induce genes encoding the MBDs MeCP2 and MBD1 by continuously activating the serotonergic (5-HT) system in the adult rat brain, suggesting that gene expression is repressed in the presence of fluoxetine. Induction of HDAC2 mRNA accompanied the protein increase, and decreased amounts of histone H3 were detected in three serotonin projection areas: the caudate-putamen, the frontal cortex, and the dentate gyrus of the hippocampus (109). Taken together, it appears that increased HDAC2 expression and recruitment to DNA plays a role in the regulation of histone acetylation and repression of gene expression in response to fluoxetine.

Imipramine, a tricyclic antidepressant, can also affect chromatin remodeling and gene expression. In the mouse hippocampus, chronic imipramine treatment reverses the downregulation of *Bdnf* transcript III and IV promoter regions that results from chronic social defeat stress. It also causes hyperacetylation of histone H3 at *Bdnf* promoters P3 and P4, which was associated with a downregulation of HDAC5. Interestingly, it was also discovered that chronic stress is associated with a histone dimethylation, a marker of repressive histone modification, at these same promoters that persists for a month after stress cessation (110). This indicates that stress may leave a long-lasting mark that is not easily reversible; such a situation calls for the development of more effective demethylating agents.

LIMITATIONS OF EPIGENETIC DRUGS

Despite the promising results produced by many novel epigenetic treatments, considerable room for improvement still exists. One major concern about epigenetic drugs is their general lack of target specificity. The DNA methylation inhibitors are employed when specific hypermethylation is implicated in cancer etiology and pathogenesis, but the resulting global demethylation can potentially activate oncogenes (111), which would aggravate the situation. Similarly, the HDACi show pleiotropic effects on multiple isoforms of HDAC and nonhistone proteins. As a consequence of their promiscuity, their effects differ based on cell type, or even within a cell type; compounds with lower HDAC isoform specificity, such as SAHA, will demonstrate a wider range of biological activities than the more selective inhibitors (112). To maximize efficiency of epigenetic drugs, their mechanisms and targets must be more clearly defined.

Although some epigenetic drugs are effective at low doses and have relatively low toxicity profiles, their effects are transient and the aberrant epigenetic patterns return once the treatment period terminates. Currently, many epigenetic treatments have to be used in combination with other therapies, for example, to sensitize tumors to cytotoxic agents or radiation (113). However, very few are used alone as first-line therapies; this is also attributed to their still somewhat low level of success, relative to established medications. Epigenetic drugs, however, are quite promising

chemopreventive agents in cases when epimutations increase the risk of developing a disease in individuals who do not yet show signs of malignancy (114). Before they can maximize their predictive strength, comprehensive knowledge of the epigenome must first be accumulated so as to identify key biomarkers. Overall, epigenetic medications are on the brink of becoming effective treatments for cancer and other conditions, but further progress is required, in particular, in the development of compounds with higher specificity and greater efficacy.

FUTURE DIRECTIONS

Because nearly all epigenetic therapeutic agents are rather nonspecific, a requirement for new epigenetic medications is a substantially higher specificity toward a particular histone-modifying enzyme. High-throughput screening of combinatorial chemistry libraries of small organic molecules (115) can provide a variety of new opportunities in identification of target-specific ligands. This is well illustrated by a recent study that used a library of 125,000 small molecules to screen for specific inhibitors against histone lysine methyltransferases (HMTases) using recombinant G9a as the target enzyme. A highly specific inhibitor, BIX-01294 (diazepinquinazolin-amine derivative), for the euchromatic G9a HMTase was identified (116). In cellular assays, transient incubation of several cell lines with BIX-01294 lowered the H3K9me2 levels and these were restored upon removal of the inhibitor, whereas in the mouse embryonic stem cells, the new ligand significantly reduced promoter-proximal H3K9me2 marks (116). Similar strategies can be used for all DNA- and histone-modification enzymes, and hopefully, specific ligands can be identified. Such ligands should exhibit higher therapeutic efficacy in comparison to nonspecific agents, such as SAHA and TSA. An increasing number of DNA and histone-modifying enzymes have been subjected to X-ray crystallography and nuclear magnetic resonance (NMR) studies, which reveal their three-dimensional structures (117). Such information enables the use of virtual screening technologies, which become even more powerful when paired with the ligand motif-based libraries and fragment-based drug designs (118).

Developments that target etiological disease epimutations, rather than DNA- and histone-modification enzymes, may be even more promising. Technological advancements, such as large-scale DNA methylation profiling (119) and chromatin immunoprecipitation experiments using microarrays (120), will soon allow for genome-wide detection of methylation patterns and histone modifications, which should make a significant contribution to the overall understanding of the epigenetic basis of human diseases. Fine mapping of epimutations and uncovering the mechanisms of their origin will provide the basis for etiological treatment. As demonstrated in the ligand search for histone-modifying enzymes, small-molecule libraries can be screened for their agonistic and antagonistic effects on the protein of epigenetically misregulated genes.

Two other technologies can be useful in the development of future epigenetic therapeutics. The first is based on a feature of zinc-finger proteins (ZFPs) to specifically recognize DNA sequences and bind to short stretches of DNA (typically 9–18 base pairs) (121). ZFPs can be used to carry out a variety of cellular activities when they are combined with different domains. In theory, an epigenetic problem can be resolved

if an epigenetically misregulated gene is treated by a corresponding histone or DNA modifying enzyme attached to a gene-specific ZFP. The ZFP will specifically bind to the epimutation locus, while the modifying enzyme permanently fixes the epigenetic misregulation.

Another promising technology is based on RNA interference (RNAi). RNAi involves double-stranded RNA–induced destruction of homologous mRNA, thus disabling protein production; small interfering RNAs (siRNAs) are endogenously produced and incorporated into an RNA-induced silencing complex (RISC), which then targets and cleaves mRNA transcripts (122). It is known that RNA interference may have an impact on the local chromatin structure, heterochromatin assembly, and gene silencing, although mechanistic details as to how the RNA and chromatin connect remain unclear (123). Several siRNAs have recently been created, including ones to knock down beta-secretase (BACE1) in Huntington's and Alzheimer's disease, and SCA1 gene in spinocerebellar ataxia (124). A clinical trial has been submitted to the FDA proposing the use of siRNA against vascular endothelial growth factor in cases of age-related macular degeneration (125). Great potential exists for the therapeutic use of siRNA to knock down mutated proteins in various disease states, although issues such as nonspecific silencing of partially homologous genes, safe delivery, and inhibition of microRNA (miRNA) must first be resolved.

In summary, although further information is required, there is a very good chance that epigenetics will contribute to the development of etiological and individualized treatment of human diseases within the next decade.

SUMMARY POINTS

1. Epigenetics is critical for normal functioning of the genome, and epimutation can be viewed as the first (etiological) step in the pathogenesis of complex diseases, which offers an explanation for their non-Mendelian characteristics.
2. Medications that target epimutations show great potential in the treatment of a variety of complex diseases, such as cancer and psychiatric disorders, although such medications are still in the early stages of development.
3. As techniques advance in the profiling of DNA methylation and histone modification patterns, disease epimutations may be detected and novel epigenetic drugs will emerge that have the potential to significantly improve the treatment of complex diseases.

FUTURE ISSUES

1. The specificity of epigenetic drugs must be greatly increased; this can be accomplished by better understanding the structure and function of target molecules, as well as elucidating the interactions between the drugs and epigenetic enzymes.

2. The epigenome must be fully mapped in normal and disease states to properly identify etiological epimutations that can also serve as biomarkers of a disease.
3. The screening of small-molecule libraries will assist in the identification of target-specific ligands, which will hopefully lead to increased drug specificity and efficacy.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Mental Health (R01 MH074127), Canadian Institutes for Health and Research (CIHR), Ontario Mental Health Foundation (OMHF), NARSAD, and the Stanley Foundation to A.P.

LITERATURE CITED

1. Henikoff S, Matzke MA. 1997. Exploring and explaining epigenetic effects. *Trends Genet.* 13:293–95
2. Margueron R, Trojer P, Reinberg D. 2005. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* 15:163–76
3. Thiagalingam S, Cheng KH, Lee HJ, Mineva N, Thiagalingam A, Ponte JF. 2003. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann. N.Y. Acad. Sci.* 983:84–100
4. Klose RJ, Zhang Y. 2007. Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* 8(4):307–18
5. Weksberg R, Smith AC, Squire J, Sadowski P. 2003. Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development. *Hum. Mol. Genet.* 12(Spec. No. 1):R61–68
6. Jones PA, Laird PW. 1999. Cancer epigenetics comes of age. *Nat. Genet.* 21:163–67
7. Baylin SB, Herman JG. 2000. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* 16:168–74
8. Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, et al. 2004. Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7(8):847–54
9. Riggs A, Xiong Z, Wang L, JM L. 1998. Methylation dynamics, epigenetic fidelity and X chromosome structure. In *Epigenetics*, ed. A Wolffe, pp. 214–27. Chichester: John Wiley & Sons
10. Richards EJ. 2006. Inherited epigenetic variation—revisiting soft inheritance. *Nat. Rev. Genet.* 7:395–401
11. Petronis A. 2001. Human morbid genetics revisited: relevance of epigenetics. *Trends Genet.* 17:142–46

12. Reiss D, Plomin R, Hetherington EM. 1991. Genetics and psychiatry: an unheralded window on the environment. *Am. J. Psychiatry* 148:283–91
13. Petronis A, Gottesman II, Kan P, Kennedy JL, Basile VS, et al. 2003. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr. Bull.* 29:169–78
14. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA* 102:10604–9
15. Wolff GL, Kodell RL, Moore SR, Cooney CA. 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J.* 12:949–57
16. Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, et al. 2003. Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. *Lancet* 361:1693–99
17. Waterland RA, Jirtle RL. 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell Biol.* 23:5293–300
18. Cooney CA, Dave AA, Wolff GL. 2002. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J. Nutr.* 132:2393S–400
19. Ostrer H. 1999. Sex-based differences in gene transmission and gene expression. *Lupus* 8:365–69
20. Seeman MV. 1997. Psychopathology in women and men: focus on female hormones. *Am. J. Psychiatry* 154:1641–47
21. Kaminsky Z, Wang SC, Petronis A. 2006. Complex disease, gender and epigenetics. *Ann. Med.* 38:530–44
22. Ohara K, Xu HD, Mori N, Suzuki Y, Xu DS, Wang ZC. 1997. Anticipation and imprinting in schizophrenia. *Biol. Psychiatry* 42:760–66
23. Guo YF, Shen H, Liu YJ, Wang W, Xiong DH, et al. 2006. Assessment of genetic linkage and parent-of-origin effects on obesity. *J. Clin. Endocrinol. Metab.* 91:4001–5
24. Bassett SS, Avramopoulos D, Fallin D. 2002. Evidence for parent of origin effect in late-onset Alzheimer disease. *Am. J. Med. Genet.* 114:679–86
25. Demenais F, Chaudru V, Martinez M. 2001. Detection of parent-of-origin effects for atopy by model-free and model-based linkage analyses. *Genet. Epidemiol.* 21(Suppl. 1):S186–91
26. Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, et al. 2005. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J. Med. Genet.* 42:132–37
27. Schulze TG, Chen YS, Badner JA, McInnis MG, DePaulo JR Jr, McMahon FJ. 2003. Additional, physically ordered markers increase linkage signal for bipolar disorder on chromosome 18q22. *Biol. Psychiatry* 53:239–43
28. Hall JG. 1990. Genomic imprinting: review and relevance to human diseases. *Am. J. Hum. Genet.* 46:857–73
29. Barlow DP. 1995. Gametic imprinting in mammals. *Science* 270:1610–13
30. Delaval K, Wagschal A, Feil R. 2006. Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *Bioessays* 28:453–59

31. Allen ND, Norris ML, Surani MA. 1990. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* 61:853–61
32. Petronis A. 2004. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol. Psychiatry* 55:965–70
33. Sutherland JE, Costa M. 2003. Epigenetics and the environment. *Ann. N.Y. Acad. Sci.* 983:151–60
34. Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33(Suppl.): 245–54
35. Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, et al. 2004. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann. Hum. Genet.* 68:196–204
36. van den Toorn LM, Overbeek SE, Prins JB, Hoogsteden HC, de Jongste JC. 2003. Asthma remission: Does it exist? *Curr. Opin. Pulm. Med.* 9:15–20
37. Faraone SV, Biederman J, Mick E. 2006. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychol. Med.* 36:159–65
38. Esteller M. 2005. DNA methylation and cancer therapy: new developments and expectations. *Curr. Opin. Oncol.* 17:55–60
39. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, et al. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* 37:391–400
40. Esteller M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 8:286–98
41. Ting Hsiung D, Marsit CJ, Houseman EA, Eddy K, Furniss CS, et al. 2007. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 16:108–14
42. Ehrlich M. 2002. DNA methylation in cancer: too much, but also too little. *Oncogene* 21:5400–13
43. Pogribny IP, Ross SA, Tryndyak VP, Pogribna M, Poirier LA, Karpinets TV. 2006. Histone H3 lysine 9 and H4 lysine 20 trimethylation and the expression of Suv4-20h2 and Suv-39h1 histone methyltransferases in hepatocarcinogenesis induced by methyl deficiency in rats. *Carcinogenesis* 27:1180–86
44. Laird PW. 2005. Cancer epigenetics. *Hum. Mol. Genet.* 14(Spec. No. 1): R65–76
45. Baylin S, Bestor TH. 2002. Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer Cell* 1:299–305
46. Veldic M, Kadriu B, Maloku E, Agis-Balboa RC, Guidotti A, et al. 2007. Epigenetic mechanisms expressed in basal ganglia GABAergic neurons differentiate schizophrenia from bipolar disorder. *Schizophr. Res.* 91:51–61
47. Hogart A, Nagarajan RP, Patzel KA, Yasui DH, Lasalle JM. 2007. 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. *Hum. Mol. Genet.* 16:691–703
48. Deng V, Matagne V, Banine F, Frerking M, Ohliger P, et al. 2007. FXYD1 is a MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Hum. Mol. Genet.* 16(6):640–50

49. Hagerman RJ, Ono MY, Hagerman PJ. 2005. Recent advances in fragile X: a model for autism and neurodegeneration. *Curr. Opin. Psychiatry* 18:490–96
50. Jeltsch A. 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* 3:274–93
51. Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP. 2002. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* 321:591–99
52. Jones PA, Taylor SM. 1980. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20:85–93
53. Wijermans P, Lubbert M, Verhoef G, Bosly A, Ravoet C, et al. 2000. Low-dose 5-aza-2′-deoxycytidine, a DNA hypomethylating agent, for the treatment of high-risk myelodysplastic syndrome: a multicenter phase II study in elderly patients. *J. Clin. Oncol.* 18:956–62
54. Kantarjian H, Issa JP, Rosenfeld CS, Bennett JM, Albitar M, et al. 2006. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer* 106:1794–803
55. Cheng JC, Matsen CB, Gonzales FA, Ye W, Greer S, et al. 2003. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. *J. Natl. Cancer Inst.* 95:399–409
56. Cheng JC, Yoo CB, Weisenberger DJ, Chuang J, Wozniak C, et al. 2004. Preferential response of cancer cells to zebularine. *Cancer Cell* 6:151–58
57. Holleran JL, Parise RA, Joseph E, Eiseman JL, Covey JM, et al. 2005. Plasma pharmacokinetics, oral bioavailability, and interspecies scaling of the DNA methyltransferase inhibitor, zebularine. *Clin. Cancer Res.* 11:3862–68
58. Stewart DJ, Donehower RC, Eisenhauer EA, Wainman N, Shah AK, et al. 2003. A phase I pharmacokinetic and pharmacodynamic study of the DNA methyltransferase 1 inhibitor MG98 administered twice weekly. *Ann. Oncol.* 14:766–74
59. Brueckner B, Boy RG, Siedlecki P, Musch T, Kliem HC, et al. 2005. Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. *Cancer Res.* 65:6305–11
60. Zambrano P, Segura-Pacheco B, Perez-Cardenas E, Cetina L, Revilla-Vazquez A, et al. 2005. A phase I study of hydralazine to demethylate and reactivate the expression of tumor suppressor genes. *BMC Cancer* 5:44
61. Kim DH, Shin J, Kwon HJ. 2007. Psammaplin A is a natural prodrug that inhibits class I histone deacetylase. *Exp. Mol. Med.* 39:47–55
62. Fang M, Chen D, Yang CS. 2007. Dietary polyphenols may affect DNA methylation. *J. Nutr.* 137:223S–28
63. Dokmanovic M, Marks PA. 2005. Prospects: histone deacetylase inhibitors. *J. Cell Biochem.* 96:293–304
64. Ruefli AA, Ausserlechner MJ, Bernhard D, Sutton VR, Tainton KM, et al. 2001. The histone deacetylase inhibitor and chemotherapeutic agent suberoylanilide hydroxamic acid (SAHA) induces a cell-death pathway characterized by cleavage of Bid and production of reactive oxygen species. *Proc. Natl. Acad. Sci. USA* 98:10833–38

65. Rosato RR, Almenara JA, Dai Y, Grant S. 2003. Simultaneous activation of the intrinsic and extrinsic pathways by histone deacetylase (HDAC) inhibitors and tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) synergistically induces mitochondrial damage and apoptosis in human leukemia cells. *Mol Cancer Ther.* 2:1273–84
66. Nebbioso A, Clarke N, Voltz E, Germain E, Ambrosino C, et al. 2005. Tumor-selective action of HDAC inhibitors involves TRAIL induction in acute myeloid leukemia cells. *Nat. Med.* 11:77–84
67. Tan J, Zhuang L, Jiang X, Yang KK, Karuturi KM, Yu Q. 2006. Apoptosis signal-regulating kinase 1 is a direct target of E2F1 and contributes to histone deacetylase inhibitor-induced apoptosis through positive feedback regulation of E2F1 apoptotic activity. *J. Biol. Chem.* 281:10508–15
68. Chen Z, Clark S, Birkeland M, Sung CM, Lago A, et al. 2002. Induction and superinduction of growth arrest and DNA damage gene 45 (GADD45) alpha and beta messenger RNAs by histone deacetylase inhibitors trichostatin A (TSA) and butyrate in SW620 human colon carcinoma cells. *Cancer Lett.* 188:127–40
69. Biade S, Stobbe CC, Boyd JT, Chapman JD. 2001. Chemical agents that promote chromatin compaction radiosensitize tumour cells. *Int. J. Radiat. Biol.* 77:1033–42
70. Louis M, Rosato RR, Brault L, Osbild S, Battaglia E, et al. 2004. The histone deacetylase inhibitor sodium butyrate induces breast cancer cell apoptosis through diverse cytotoxic actions including glutathione depletion and oxidative stress. *Int. J. Oncol.* 25:1701–11
71. Yoo CB, Jones PA. 2006. Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.* 5:37–50
72. ten Cate B, Samplonius DF, Bijma T, de Leij LF, Helfrich W, Bremer E. 2007. The histone deacetylase inhibitor valproic acid potently augments gemtuzumab ozogamicin-induced apoptosis in acute myeloid leukemic cells. *Leukemia* 21:248–52
73. Garcia-Manero G, Kantarjian HM, Sanchez-Gonzalez B, Yang H, Rosner G, et al. 2006. Phase 1/2 study of the combination of 5-aza-2′-deoxycytidine with valproic acid in patients with leukemia. *Blood* 108:3271–79
74. Kuendgen A, Knipp S, Fox F, Strupp C, Hildebrandt B, et al. 2005. Results of a phase 2 study of valproic acid alone or in combination with all-trans retinoic acid in 75 patients with myelodysplastic syndrome and relapsed or refractory acute myeloid leukemia. *Ann. Hematol.* 84(Suppl. 13):61–66
75. Yin H, Chen ZX, Cen JN, Wang W, Duan WM, Yao L. 2005. Mechanism of differentiation and apoptosis in leukemia cells induced by tributyrin. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* 13:774–77
76. Yoshida M, Nomura S, Beppu T. 1987. Effects of trichostatins on differentiation of murine erythroleukemia cells. *Cancer Res.* 47:3688–91
77. Yoshida M, Kijima M, Akita M, Beppu T. 1990. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J. Biol. Chem.* 265:17174–79

78. Ou JN, Torrisani J, Unterberger A, Provencal N, Shikimi K, et al. 2007. Histone deacetylase inhibitor trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. *Biochem. Pharmacol.* 73(9):1297–307
79. Wu Y, Guo SW. 2006. Inhibition of proliferation of endometrial stromal cells by trichostatin A, RU486, CDB-2914, N-acetylcysteine, and ICI 182780. *Gynecol. Obstet. Invest.* 62:193–205
80. Richon VM, Webb Y, Merger R, Sheppard T, Jursic B, et al. 1996. Second generation hybrid polar compounds are potent inducers of transformed cell differentiation. *Proc. Natl. Acad. Sci. USA* 93:5705–8
81. Finnin MS, Donigian JR, Cohen A, Richon VM, Rifkind RA, et al. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* 401:188–93
82. Marks PA, Breslow R. 2007. Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* 25:84–90
83. Duvic M, Talpur R, Ni X, Zhang C, Hazarika P, et al. 2007. Phase 2 trial of oral vorinostat (suberoylanilide hydroxamic acid, SAHA) for refractory cutaneous T-cell lymphoma (CTCL). *Blood* 109:31–39
84. Grant S, Easley C, Kirkpatrick P. 2007. Vorinostat. *Nat. Rev. Drug Discov.* 6:21–22
85. McLaughtin F, Nielsen A, Thougard A, Sehested M, Buhl Jensen P, et al. 2003. *Molecular Targets and Cancer Therapeutics*. Presented at *AACR-NCI-EORTC Int. Conf.*, Boston, MA
86. Kato Y, Salumbides BC, Wang XF, Qian DZ, Williams S, et al. 2007. Antitumor effect of the histone deacetylase inhibitor LAQ824 in combination with 13-*cis*-retinoic acid in human malignant melanoma. *Mol. Cancer Ther.* 6:70–81
87. Remiszewski SW, Sambucetti LC, Bair KW, Bontempo J, Cesarz D, et al. 2003. N-hydroxy-3-phenyl-2-propenamides as novel inhibitors of human histone deacetylase with in vivo antitumor activity: discovery of (2E)-N-hydroxy-3-[4-[[(2-hydroxyethyl)[2-(1H-indol-3-yl)ethyl]amino]methyl]phenyl]-2-propenamide (NVP-LAQ824). *J. Med. Chem.* 46:4609–24
88. Loprevite M, Tiseo M, Grossi F, Scolaro T, Semino C, et al. 2005. In vitro study of CI-994, a histone deacetylase inhibitor, in non-small cell lung cancer cell lines. *Oncol. Res.* 15:39–48
89. Pauer LR, Olivares J, Cunningham C, Williams A, Grove W, et al. 2004. Phase I study of oral CI-994 in combination with carboplatin and paclitaxel in the treatment of patients with advanced solid tumors. *Cancer Invest.* 22:886–96
90. Prakash S, Foster BJ, Meyer M, Wozniak A, Heilbrun LK, et al. 2001. Chronic oral administration of CI-994: a phase 1 study. *Invest. New Drugs* 19:1–11
91. Richards DA, Boehm KA, Waterhouse DM, Wagener DJ, Krishnamurthi SS, et al. 2006. Gemcitabine plus CI-994 offers no advantage over gemcitabine alone in the treatment of patients with advanced pancreatic cancer: results of a phase II randomized, double-blind, placebo-controlled, multicenter study. *Ann. Oncol.* 17:1096–102
92. Camphausen K, Burgan W, Cerra M, Oswald KA, Trepel JB, et al. 2004. Enhanced radiation-induced cell killing and prolongation of gammaH2AX foci expression by the histone deacetylase inhibitor MS-275. *Cancer Res.* 64:316–21

93. Gojo I, Jiemjit A, Trepel JB, Sparreboom A, Figg WD, et al. 2006. Phase 1 and pharmacological study of MS-275, a histone deacetylase inhibitor, in adults with refractory and relapsed acute leukemias. *Blood* 109(7):2781–90
94. Ryan QC, Headlee D, Acharya M, Sparreboom A, Trepel JB, et al. 2005. Phase I and pharmacokinetic study of MS-275, a histone deacetylase inhibitor, in patients with advanced and refractory solid tumors or lymphoma. *J. Clin. Oncol.* 23:3912–22
95. Wang XF, Qian DZ, Ren M, Kato Y, Wei Y, et al. 2005. Epigenetic modulation of retinoic acid receptor beta2 by the histone deacetylase inhibitor MS-275 in human renal cell carcinoma. *Clin. Cancer Res.* 11:3535–42
96. Jose B, Oniki Y, Kato T, Nishino N, Sumida Y, Yoshida M. 2004. Novel histone deacetylase inhibitors: cyclic tetrapeptide with trifluoromethyl and pentafluoroethyl ketones. *Bioorg. Med. Chem. Lett.* 14:5343–46
97. Sandor V, Bakke S, Robey RW, Kang MH, Blagosklonny MV, et al. 2002. Phase I trial of the histone deacetylase inhibitor, depsipeptide (FR901228, NSC 630176), in patients with refractory neoplasms. *Clin. Cancer Res.* 8:718–28
98. Piekarz RL, Robey R, Sandor V, Bakke S, Wilson WH, et al. 2001. Inhibitor of histone deacetylation, depsipeptide (FR901228), in the treatment of peripheral and cutaneous T-cell lymphoma: a case report. *Blood* 98:2865–68
99. Byrd JC, Marcucci G, Parthun MR, Xiao JJ, Klisovic RB, et al. 2005. A phase 1 and pharmacodynamic study of depsipeptide (FK228) in chronic lymphocytic leukemia and acute myeloid leukemia. *Blood* 105:959–67
100. Fouladi M, Furman WL, Chin T, Freeman BB 3rd, Dudkin L, et al. 2006. Phase I study of depsipeptide in pediatric patients with refractory solid tumors: a Children's Oncology Group report. *J. Clin. Oncol.* 24:3678–85
101. Stadler WM, Margolin K, Ferber S, McCulloch W, Thompson JA. 2006. A phase II study of depsipeptide in refractory metastatic renal cell cancer. *Clin. Genitourin. Cancer* 5:57–60
102. Maroun JA, Belanger K, Seymour L, Matthews S, Roach J, et al. 2006. Phase I study of Aplidine in a dailyx5 one-hour infusion every 3 weeks in patients with solid tumors refractory to standard therapy. A National Cancer Institute of Canada Clinical Trials Group study: NCIC CTG IND 115. *Ann. Oncol.* 17:1371–78
103. Kijima M, Yoshida M, Sugita K, Horinouchi S, Beppu T. 1993. Trapoxin, an antitumor cyclic tetrapeptide, is an irreversible inhibitor of mammalian histone deacetylase. *J. Biol. Chem.* 268:22429–35
104. Darkin-Rattray SJ, Gurnett AM, Myers RW, Dulski PM, Crumley TM, et al. 1996. Apicidin: a novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc. Natl. Acad. Sci. USA* 93:13143–47
105. Han JW, Ahn SH, Park SH, Wang SY, Bae GU, et al. 2000. Apicidin, a histone deacetylase inhibitor, inhibits proliferation of tumor cells via induction of p21WAF1/Cip1 and gelsolin. *Cancer Res.* 60:6068–74
106. Furumai R, Komatsu Y, Nishino N, Khochbin S, Yoshida M, Horinouchi S. 2001. Potent histone deacetylase inhibitors built from trichostatin A and cyclic tetrapeptide antibiotics including trapoxin. *Proc. Natl. Acad. Sci. USA* 98:87–92

107. Bourin M, Prica C. 2007. The role of mood stabilisers in the treatment of the depressive facet of bipolar disorders. *Neurosci. Biobehav. Rev.* In press
108. Milutinovic S, D'Alessio AC, Detich N, Szyf M. 2007. Valproate induces widespread epigenetic reprogramming which involves demethylation of specific genes. *Carcinogenesis* 28:560–71
109. Cassel S, Carouge D, Gensburger C, Anglard P, Burgun C, et al. 2006. Fluoxetine and cocaine induce the epigenetic factors MeCP2 and MBD1 in adult rat brain. *Mol. Pharmacol.* 70:487–92
110. Tsankova NM, Berton O, Renthal W, Kumar A, Neve RL, Nestler EJ. 2006. Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat. Neurosci.* 9:519–25
111. Eden A, Gaudet F, Waghmare A, Jaenisch R. 2003. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300:455
112. Bolden JE, Peart MJ, Johnstone RW. 2006. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* 5:769–84
113. Conley BA, Wright JJ, Kummar S. 2006. Targeting epigenetic abnormalities with histone deacetylase inhibitors. *Cancer* 107(4):832–40
114. Laird PW. 2003. The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3:253–66
115. Strausberg RL, Schreiber SL. 2003. From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science* 300:294–95
116. Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, et al. 2007. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol. Cell* 25:473–81
117. Nielsen TK, Hildmann C, Dickmanns A, Schwienhorst A, Ficner R. 2005. Crystal structure of a bacterial class 2 histone deacetylase homologue. *J. Mol. Biol.* 354:107–20
118. Hajduk PJ, Greer J. 2007. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* 6:211–19
119. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, et al. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet.* 37(8):853–62
120. Kim TH, Ren B. 2006. Genome-wide analysis of protein-DNA interactions. *Annu. Rev. Genomics Hum. Genet.* 7:81–102
121. Jamieson AC, Miller JC, Pabo CO. 2003. Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug Discov.* 2:361–68
122. Waggoner D. 2007. Mechanisms of disease: epigenesis. *Semin. Pediatr. Neurol.* 14:7–14
123. Bernstein E, Allis CD. 2005. RNA meets chromatin. *Genes Dev.* 19:1635–55
124. Pushparaj PN, Melendez AJ. 2006. Short interfering RNA (siRNA) as a novel therapeutic. *Clin. Exp. Pharmacol. Physiol.* 33:504–10
125. Reich SJ, Fosnot J, Kuroki A, Tang W, Yang X, et al. 2003. Small interfering RNA (siRNA) targeting VEGF effectively inhibits ocular neovascularization in a mouse model. *Mol. Vis.* 9:210–16

Annual Review of Pharmacology and Toxicology

Volume 48, 2008

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml

Systems Pharmacology: Network Analysis to Identify Multiscale Mechanisms of Drug Action

Shan Zhao and Ravi Iyengar

Department of Pharmacology and Systems Therapeutics, and Systems Biology Center New York, Mount Sinai School of Medicine, New York, New York 10029; email: ravi.iyengar@mssm.edu

Annu. Rev. Pharmacol. Toxicol. 2012. 52:505–21

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-010611-134520

0362-1642/12/0210-0505$20.00

Keywords

enhanced pharmacodynamics, genomics and personalized therapy, drug discovery, adverse event predictions

Abstract

Systems approaches have long been used in pharmacology to understand drug action at the organ and organismal levels. The application of computational and experimental systems biology approaches to pharmacology allows us to expand the definition of systems pharmacology to include network analyses at multiple scales of biological organization and to explain both therapeutic and adverse effects of drugs. Systems pharmacology analyses rely on experimental "omics" technologies that are capable of measuring changes in large numbers of variables, often at a genome-wide level, to build networks for analyzing drug action. A major use of omics technologies is to relate the genomic status of an individual to the therapeutic efficacy of a drug of interest. Combining pathway and network analyses, pharmacokinetic and pharmacodynamic models, and a knowledge of polymorphisms in the genome will enable the development of predictive models of therapeutic efficacy. Network analyses based on publicly available databases such as the U.S. Food and Drug Administration's Adverse Event Reporting System allow us to develop an initial understanding of the context within which molecular-level drug-target interactions can lead to distal effectors in a process that results in adverse phenotypes at the organ and organismal levels. The current state of systems pharmacology allows us to formulate a set of questions that could drive future research in the field. The long-term goal of such research is to develop polypharmacology for complex diseases and predict therapeutic efficacy and adverse event risk for individuals prior to commencement of therapy.

INTRODUCTION

Over the past 60 years, drug therapy for many complex noncommunicable diseases has been quite successful. Drugs are now routinely used to control hypertension and treat peptic ulcers, asthma, and many types of cancers. In spite of these successes over past decades, it has become clear that the drug discovery process has slowed down as the costs of bringing a drug to market have gone up tremendously (1). New targets are most often identified by linking individual cellular components to an organismal- or tissue/organ-level phenotype. Such distal correlations, although a good starting point, often do not work for drug development because the cellular and tissue/organ-level systems are treated as black boxes. This leads to a lack of mechanistic understanding of how drug interactions at the molecular level manifest themselves as alterations in tissue/organ-level function. This lack of understanding, in turn, leads to confounding situations at various stages during the drug discovery process. Drugs that are promising in cell-based assays often do not work in vivo, and even when they do, they show variable efficacy. Most new drugs often fail in Phase II and Phase III trials. Another limitation of the black-box approach is the inability to predict adverse events when the drug is brought to market and used by the population at large. The occurrence of serious and sometimes fatal adverse events has led to the withdrawal of or tight restrictions on the use of drugs that are beneficial to the majority of the population. Such regulatory caution is warranted because of our lack of ability to predict who among the target users of a drug will have a serious adverse event such that the risk of adverse events outweighs the therapeutic advantage. These problems have led to calls for different approaches to drug discovery and therapeutics (2).

The advances in mammalian genomics, biochemistry, molecular and cell biology, and physiology have allowed us initial glimpses of the complexity of human and mammalian biology at multiple scales that involve organization at the atomic/molecular, cellular and tissue, organ, and organismal levels. Although this understanding is still far from complete, the emerging picture indicates that network analysis, wherein we study the organization (i.e., topology) of interactions among components of a system, can provide a useful approach for multiscale understanding. Network analyses allow us to define the relationship between emergent functions and topology at each level of organization (atomic/molecular, cellular/tissue, organ, and organismal) and connections between levels that give rise to organ- and organismal-level functions. Understanding the explicit relationships between scales (i.e., levels) of organization allows us to appreciate how drugs that interact with molecular components and have their first effects at the cellular level are able to produce organ- and organismal-level effects, both therapeutic and adverse. In this approach, which focuses on what we term multiscale mechanisms, black-box assumptions are purposefully avoided and qualitative relationships or quantitative parameters are directly related to molecular interactions or cellular functions.

The term systems pharmacology now describes a field of study that uses experimental and computational approaches to provide us with a broad view of drug action rooted in molecular interactions between the drug and its targets in the context of such targets interacting with and regulating other cellular components. This newer definition expands the older usage, in which systems pharmacology was used to describe drug action in a specific organ system such as ocular pharmacology or reproductive pharmacology (3). Here we review the current status of this new field and describe how, in our view, integrating network analysis with dynamical quantitative approaches in pharmacokinetic and pharmacodynamic models can help us develop a mechanistic understanding of drug action in the context of an individual's genomic status and environmental exposure.

TYPES OF NETWORKS USED IN ANALYZING DRUG ACTION

A network is defined as a series of entities connected to one another on the basis of a defined criterion. The entities in a network are named nodes, which represent different types of objects such as genes (4), proteins (5), drugs (6), and disease (7, 8). Nodes in a network can also be used to specify the state of a system. Such specifications can be computed using Boolean dynamics (9, 10), in which each node has a chance to exist in two states (inactive or active), or using concentrations of the nodes with dynamical models based on ordinary differential equations (9, 11). The latter approach is most commonly used in pharmacokinetic-pharmacodynamic models.

The connections between the nodes are termed edges and can be specified using criteria of interest. In the context of studying drug action, edges may represent protein-protein interactions (5), drug-target interactions (12), or transcriptional regulation (13). Edges can also be defined on the basis of similarities between two nodes such as the chemical or therapeutic similarity (6), the similarities between proteins based on the shared number of diseases (14), or the similarities of diseases based on the shared number of genes (14). These types of complex definitions for edges allow the networks to transcend multiple scales of interactions from the atomic- and molecular-level drug-target interactions to the coordinated functional outputs of multiple organs, i.e., phenotype. Such multiscale organization is featured in **Figure 1**, which shows networks at two levels: the cellular/tissue level and the organ level. There is a possibility that tissue-level networks are distinct from cellular-level networks. This distinction will have to be specified on a case-by-case basis. Nevertheless, developing networks at various scales allows us to explicitly track drug effects from atomic-level interactions to organismal physiology.

The edges of a network can be directed, in which the source node causes an effect on the target node, and the relationship is valid in only one direction. An example of directed edges is the protein kinase activation of a transcription factor and the regulation of a target gene by that

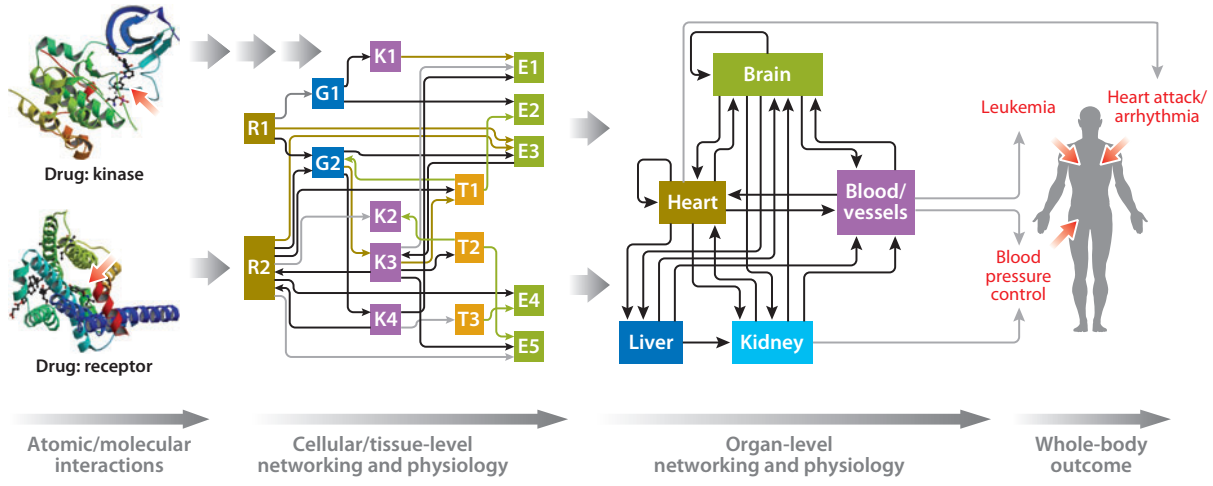


Figure 1

A schematic representation of the multiscale networks needed to understand and predict drug action. Atomic interactions between drug and target lead to alterations in the function of cellular regulatory networks, which lead to changes in cellular- and tissue-level physiology, which, in turn, lead to alterations in organ-level networking, which lead to changes in whole-body functions. Networks at both the cellular/tissue level and organ level are needed to understand the mechanism of drug action and to predict therapeutic efficacy and adverse event probability. The drug-protein structures are taken from structures deposited in the Protein Data Bank (**http://www.pdb.org**) with PDB IDs of 3QC4 and 2Y03 (88, 89), with the authors' permission.

transcription factor (13). Alternatively, the edges can be undirected, in which interactions can occur in both directions. Examples of undirected edges include the interactions between a protein and its scaffold. The edges can also be given weight based on the strength of their association. These weights can be derived from numerous criteria, ranging from statistical correlations for distal relationships (such as gene-disease relationships) to kinetic rate constants for direct physical interactions (such as hormone or drug binding to receptors).

In analyzing drug actions, one can use a variety of networks based on different types of nodes and edges. The simplest network has a directed edge connecting a drug node to its target protein node (12). The target protein node is then connected to other proteins that physically interact with the drug-target protein, and these proteins can be linked to additional proteins using the same criterion (15). In this network, all edges have the same weight, implying that they have the same extent of connectivity. This simplifying assumption is not always true, and we need to be careful in ascertaining when the network as depicted is a reasonable representation of the system. These types of networks are known as interaction networks. Interaction networks allow us to quickly determine the potential downstream and upstream interactors of a particular node (16), which can be useful in identifying pathways for signal flow and regulatory motifs such as feed-forward and feedback loops that have information processing capability.

Interaction networks are also the foundation of more highly specified networks that are typically used to study biological systems such as dynamical, Boolean, and stochastic networks. However, the interaction networks require the least amount of knowledge regarding nodes and edges, allowing them to be easily constructed and applied to a variety of problems (6, 17–19). The simplicity of interaction networks allows networks at different scales of organization to be combined on the basis of the notion that they "interact"; i.e., they are related. A protein-protein interaction network can be expanded by connecting proteins to their physiological function (20), and, in doing so, one can analyze drug-to-physiological function through combining edges that represent drug-protein, protein-protein, and protein-physiology interactions. The types of network used to analyze drug action depend on the type of action of interest. These networks can be constructed, taken apart, and reorganized from different data sets that define methodologies for identifying edges that connect the nodes. Performing the analyses correctly requires knowledge about the edges' limitations and meanings.

EXPERIMENTAL APPROACHES IN SYSTEMS PHARMACOLOGY

The data sets needed to build networks require simultaneous measurements on a large number of variables in response to a perturbation, such as a pathophysiological state or drug treatment. Systems biology uses "omics" experiments, in which a large number of output variables is measured in response to one or more perturbations. Typically, such experiments fall into one of three categories: genomics, proteomics, or metabolomics.

Genomics Analyses

Genomics analyses involve the sequencing or characterization of many genes, typically the whole genome simultaneously. At the DNA level, genomics involves sequencing of the genome to identify variations and to determine transcriptional binding sites and epigenetic status. At the mRNA level, genome-wide profiling is largely focused on characterizing gene expression patterns in a disease state or before and after drug treatment. This type of profiling was accomplished mostly through the use of microarrays, but in the past few years, direct sequencing, termed mRNA seq, has become more widely used.

As the sequences of many organisms were determined, it became apparent that positional variations exist within the DNA sequences of the individuals in a single species and even between the sequences of pairs of chromosomes. These variations are named single-nucleotide polymorphisms (SNPs). Sometimes SNPs fall within the coding regions of genes, leading to changes in protein primary sequences and activity (21). Although coding-region SNPs are infrequent, they are important in drug action because genes for several drug-metabolizing enzymes have coding-region SNPs, such as rs1799853 in cytochrome P450–2C9 (CYP2C9*2) for warfarin dosing (22). Consequently, genomics analysis that is focused on DNA-sequencing methodologies has become important in understanding drug action. Sequencing technologies are rapidly changing, and the cost of sequencing is decreasing (23). These developments indicate that whole-genome sequencing is likely to play an important role in systems pharmacology.

An interesting approach for characterizing drug action has been the use of gene expression patterns as a means to connect drugs with disease states. Combining profiling experiments with pattern-matching software, Lamb et al. (24) have created a library of gene expression signatures from adding drugs—both U.S. Food and Drug Administration (FDA)-approved drugs and non-therapeutic small molecules used in laboratory research—to human cell lines in vitro to obtain whole-genome gene expression patterns. These studies indicate that structurally different compounds that converge on common targets can yield the same gene expression signature. As the field progresses from cell-based to tissue-based analyses of drug action, such gene expression signatures can be useful in understanding drug action in multiple tissues and organs.

Proteomics Analyses

Proteomics involves the study of changes in the levels or states of large numbers of proteins in a sample of interest such as a cell extract, the plasma, or a tissue sample. Typically, the measurement of proteins is by mass spectrometry, although sometimes protein arrays are also used. In contrast to genomics approaches, the use of proteomics in drug discovery and study of drug action has been limited. A major issue is the difficulty in obtaining tissue biopsies sufficient for proteomics analyses to correlate changes in target tissues and organs with drug action in humans. Most proteomics studies have focused on human cancer cell lines and can be used for target profiling (25) and mechanism-based classification of potential drugs (26).

Metabolomics Analyses

Metabolomics focuses on measuring changes in a large number of metabolites simultaneously (27). The method of choice for identification of metabolites is mass spectrometry, generally preceded by chromatographic resolution. The most readily available source for metabolic profiling in humans is plasma. Several studies have shown identifiable metabolic signatures associated with drug treatment. A study on 50 patients with schizophrenia being treated with antidepressants showed identifiable changes in lipid patterns after treatment (28). These observations raise the possibility that metabolic signatures of drug treatment could be an additional tool in assessing drug therapy in patients. A recent study (29) on patients with major depressive disorders has shown an interesting relationship between genomics and metabolomics in predicting drug action. Metabolomics was used to characterize levels of amino acids in plasma. Patients who were nonresponsive to therapy with the serotonin reuptake inhibitor citalopram showed higher baseline levels of glycine, which remained unaltered after treatment. Genomics analyses indicated that in nonresponders, the SNP rs10975641 in the glycine dehydrogenase gene was associated with treatment outcome. The authors of the study (29) did not elucidate the mechanisms by which the SNP in the glycine

dehydrogenase gene can result in elevated plasma levels of glycine, but the study is important because it provides an approach for developing mechanistic multiscale studies wherein genomic changes can be correlated with the biochemical profile of the plasma, which, in turn, can be correlated with treatment outcomes. Such multilevel correlation can be used to specify mechanisms that operate within cells and tissues.

PHARMACOGENOMICS AS A SUCCESSFUL MODEL FOR SYSTEMS PHARMACOLOGY

A success story in explicitly connecting genomic status and drug action has come from the field of pharmacogenomics. This type of linkage is important for understanding drug action and effects at an individual level. Genomics is thought to account for a significant part of interindividual variability (30) in the drug effect, while eliciting consistent intraindividual responses (31). Many FDA-approved drugs now contain pharmacogenomic information within their labels. Pharmacogenomic specification is used for a wide range of drugs from antiasthmatics (e.g., montelukast) (32) to cancer therapeutics (e.g., cetuximab) (33). The study of genomics variations in altering drug responses can be divided into three major areas: pharmacokinetics, pharmacodynamics, and responsiveness to therapy. A list of select pharmacogenomic biomarkers in the various areas for FDA-approved drugs is given in **Table 1**.

A major use of genomic information is in relating genomic status to drug dosage and metabolism (pharmacokinetics) because drug-metabolizing enzymes play a large role in converting prodrugs into active metabolites (e.g., codeine) (34) or active drugs into inactive drugs or toxic metabolites (e.g., nortriptyline) (35). Sometimes, the same metabolizing enzyme can do both: Both codeine and nortriptyline are metabolized by the cytochrome P450 isoform CYP2D6 (35). Codeine is converted from the prodrug to the active drug, whereas nortriptyline is converted from the active

Table 1 Various types of pharmacogenomic effects in drug action

Drug	Gene	Effect
Pharmacokinetics		
Codeine	*CYP2D6* (34)	Increase in the amount of active drug by variants
Clopidogrel	*CYP2C19* (80)	Increase in the amount of active drug by variants
Warfarin	*CYP2C9* (81)	Changes in drug levels in blood by variants
Pharmacodynamics		
Warfarin	*VKORC1* (21)	Increase or decrease of effectiveness of drug
Capecitabine	*DPD* (82)	Decrease in breakdown of 5-FU metabolite
Responsiveness		
Panitumumab	*k-RAS* (83)	Requirement of wild-type *k-RAS* for drug efficacy
Imatinib	*c-KIT* (84)	Requirement of wild-type *c-KIT* for drug efficacy
Tretinoin	*PML/RARα* translocation (85)	Increased drug responsiveness
Unknown mechanisms		
Carbamazepine	*HLA-B*1502* (86)	Increased risk of Stevens-Johnson syndrome and toxic epidermal necrolysis
Abacavir	*HLA-B*5701* (87)	Multiorgan systemic hypersensitivity, which may lead to death

Abbreviations: 5-FU, fluorouracil; CYP, cytochrome P450; DPD, dihydropyrimidine dehydrogenase; HLA, human leukocyte antigen; PML, promyelocytic leukemia; RAR, retinoic acid receptor; VKOR, vitamin K epoxide reductase.

form to the inactive form. In both cases, the relationship between the intake dose of the drug and the active drug in the plasma is related to the activity of CYP2D6. These findings have led to tests for drug-metabolizing proteins in individuals (36, 37) so that intake dosage can be set at a safe level for each patient. Other drug-binding proteins with polymorphisms include drug transporters such as the ATP-binding cassette family and solute carrier organic anion transporter family of membrane transporters. These transporters affect plasma concentrations of non-nucleoside reverse transcriptase inhibitors sunitinib (a tyrosine kinase inhibitor) and methotrexate (a dihydrofolate reductase enzyme inhibitor) (29, 38, 39).

The cytochrome P450 protein CYP2C9 regulates the metabolism of warfarin, which is used as an anticoagulant. CYP2C9 has two polymorphisms that reduce the level of enzyme activity; consequently, increased levels of warfarin in the blood result in an increased risk of bleeding (22). Thus, knowing if a patient has CYP2C9 polymorphisms allows the physician to titrate the dosage of warfarin to optimize therapy while reducing the risk of bleeding as a serious adverse event. Testing for warfarin metabolism has become a common approach to titrating warfarin dosage in clinical practice.

The relationship between genomic status and pharmacodynamics can also be important. An example is the polymorphisms in vitamin K epoxide reductase (VKOR) that alter an individual's sensitivity to warfarin (21). VKOR activity is required for the γ-carboxylation of multiple clotting factors, and warfarin exerts its action by inhibiting VKOR and thus reducing clotting. Polymorphisms in the VKOR genes result in variants that have lower or higher sensitivity to warfarin; thus, the dosage of warfarin needs to be adjusted in patients with different polymorphisms (21). In the case of warfarin, both pharmacokinetic and pharmacodynamic properties are regulated by SNPs, so both aspects need to be considered in the dosing of individual patients.

Genomic status can also predict responsiveness in therapy. This type of relationship has been used mainly in cancer therapy—the presence of certain mutant oncogenes is an indicator for lack of responsiveness to targeted therapy. *KIT* oncogene mutations reduce the responsiveness of gastrointestinal stromal tumors to imatinib (40); *k-RAS* oncogene mutations in colorectal cancer reduce responsiveness to cetuximab (33); and epidermal growth factor receptor mutations in non-small-cell lung cancer alter responsiveness to gefitinib or erlotinib (41, 42). Other examples in which a genotype–drug response phenotype is known but the underlying mechanism is not fully understood include patients with variants in major histocompatibility complex class IB (HLA-B*5701) who show hypersensitivity to the antiviral drug abacavir (43) and musculoskeletal toxicity from aromatase inhibitors used to treat breast and ovarian cancers (44).

Several tests are commercially available for identifying polymorphisms associated with drug responses. Those approved by the FDA include the microarray-based Roche AmpliChip® for cytochrome P450 genotyping. Because warfarin effects are regulated by polymorphisms in both cytochrome P450 enzymes and VKOR, the FDA recommends testing both CYP2C9 and VKORC1 polymorphisms for warfarin (45). The FDA has also approved the Invader® UGT1A1 Molecular Assay for the polymorphisms that increase the risk of neutropenia associated with the colon cancer drug irinotecan, which inhibits DNA topoisomerase. The FDA recommends the testing of HLA-B*5701 for abacavir; low-density lipoprotein receptor variants for atorvastatin; thiopurine *S*-methyltransferase for azathioprine; HLA-B*1502 for carbamazepine; epidermal growth factor receptor for cetuximab; cytochrome P450 protein CYP2C19 for clopidogrel; breakpoint cluster region–Abl tyrosine kinase translocation for dasatinib and imatinib; UDP glucuronosyltransferase 1 family, polypeptide A1 for irinotecan; *k-RAS* oncogene for panitumumab; glucose-6-phosphate dehydrogenase for rasburicase; erythroblastic leukemia viral oncogene homolog 2 for trastuzumab; and carbamoyl-phosphate synthase 1 for valproic acid. However, no specific tests for these polymorphisms have been approved by the FDA.

NEED FOR DETAILED PHARMACODYNAMIC MODELS THAT INCLUDE CELLULAR AND TISSUE MECHANISMS

Although in the case of warfarin one can explicitly relate drug interactions with metabolizing enzymes and target proteins to physiological events such as thrombosis and bleeding at the organismal level, explicitly identifying such multiscale relationships for most drugs is difficult. A major challenge for systems pharmacology is the development of a mechanistic understanding of how different cellular- and tissue-level regulatory networks control variability in drug response at the organismal level. For this, we need enhanced pharmacodynamic models that couple detailed models of cellular regulatory networks with measurable pharmacokinetic and pharmacodynamic parameters. A recent pharmacodynamic modeling study (46) on antagonists of the calcium-sensing receptor and its relationship to parathyroid hormone (PTH) secretion has shown how negative allosteric modulators of the calcium-sensing receptor, coupled with negative feedback, can explain the observed homeostatic relationship among plasma calcium, PTH, and calcium absorption. The mechanisms of the PTH-driven negative feedback are not known; this demonstrates the need to build detailed models of intra- and intercellular signal networks so that the molecular and cellular bases of homeostasis can be explicitly described and understood. It has long been known that intracellular signaling networks form regulatory motifs such as positive feedback loops that can function as bistable switches (47) that are involved in cellular state change such as synaptic plasticity (48). Large signaling networks are full of regulatory motifs such as feedback and feed-forward loops that can process information as signal flows through (16). Characterizing the topology of cellular regulatory networks and understanding the dynamic capability of the topology can help explain both therapeutic and adverse drug action. In the calcium-sensing receptor and PTH, identifying the molecular components that participate in the negative feedback loop and how they can be modulated can help us design better antagonists at the calcium-sensing receptor or develop polypharmacology for treatment of osteoporosis.

Detailed studies of intracellular pathways, such as metabolic or signaling pathways, can be useful for understanding the efficacy of drug action in humans. Panetta et al. (49) built models of methotrexate metabolism and action for the treatment of childhood acute lymphoblastic leukemia. The model incorporated the production of methotrexate polyglutamate metabolites and the regulation of the folate pathway enzymes by both methotrexate and its metabolites in T and B cells. The development of the detailed model allowed for simulations that were able to explain how changing dosage and duration of infusion affected efficacy of treatment. The simulations also took into account how SNPs in folate metabolism genes affect drug responses. This combination of pathway simulations and pharmacokinetic and pharmacodynamic patient data with knowledge of genomic status can be useful in predicting drug efficacy in individual patients.

NETWORK ANALYSIS FOR DISCOVERY OF NEW DRUG TARGETS

Drug discovery for complex diseases requires the identification of therapeutic targets that can be used to achieve the desired therapeutic effect while reducing the risk of adverse events. Network analysis methods provide computational tools for pharmacologists and physiologists to identify and rank potential targets, which can then be used to develop drugs. The tasks of simultaneously identifying the appropriate target, determining efficacy for the therapeutic effect, and predicting adverse events constitute a problem that cannot be solved through the use of high-throughput experimental techniques alone owing to the high dimensional size of the problem. Determination of efficacy requires detailed dynamical models based on the biochemical kinetic parameters of the target and other proteins involved in the phenotypic responses. Network analysis can be used, in

an unbiased way, to define the system that needs to be dynamically modeled and identify targets that, in theory, would enable one to have a higher specificity for the selected drug targets. Such combined network and dynamical analysis should both increase therapeutic efficacy and decrease the adverse events (15).

The mammalian signaling and regulatory network is complex, and even perturbations on the intended targets may lead to adverse events owing to propagation of signal to distal effectors in multiple cell types or tissues. Certain proteins, such as GRB2 or MYC, may directly interact with several hundred different proteins, according to a PubMed search in March 2011. Most protein kinases have 10 or more substrates. Such multiplicity of connections poses serious challenges in designing drugs that affect single pathways that lead to the desired therapeutic effects. To identify new drug targets, it is important to know how specificity of signal flow is achieved within pathways. Network analysis can be used (*a*) to identify how many different proteins will be affected by the targeting of a particular protein or (*b*) to identify if a protein participates in a motif (16). This approach involves the use of what is known as a path discovery. The simplest and oldest is Dijkstra's (50) shortest-path algorithm: When given a seed node, the algorithm finds the shortest path between that node and potentially every other node. By looking for a path—that is, a series of edges from a starting node to an ending node—one can identify relationships and topology for the network. This algorithm can be expanded by setting requirements for the path; for example, one can require a path to start at a receptor, go through specific types of intracellular proteins, and eventually reach a known transcription factor. However, imposing such requirements would require a longer run time that increases with their complexity. A protein that has a high connectivity is commonly lethal, whereas diseased genes with lower connectivity are nonlethal but lead to a diseased phenotype (7). This type of analysis can produce initial ranked lists of potential targets that can be further vetted by additional criteria.

Both positive and negative effects can be predicted for a potential target on the basis of network analysis methods. For example, a network algorithm utilizing mean first passage time (MFPT) can, on the basis of a set of known genes that cause a prolonged long QT interval, identify drugs that may lead to a long QT interval event (15). In this method, Berger et al. (15) used the MFPT as a distance measure to assess how "close" a protein is to another protein that is known to be related to long QT intervals. Other methods of measuring functional distance can also be used. The nearest-neighbor method assesses distance by measuring the shortest length of a path between a protein of interest and any of the known long QT–causing proteins. It is important to know the scoring metric used in the algorithm; for example, two methods looking at protein-protein interactions may measure them differently according to the problem of interest (6, 51). Network analysis methods can be used to determine the association of a protein with a physiological phenotype of interest on the basis of a defined set of proteins related to the phenotype. Therefore, proteins can be selected as potential targets on the basis of their high proximity to the phenotype of interest and their distance from proteins involved in adverse events. The potential of finding one target that can be optimized for a variety of criteria simultaneously is likely to be low. Computational network analysis allows us to explore combinatorial targets and thus increases our chances of finding useful therapeutic agents for complex diseases.

Network Analysis to Define the Context of Targets Involved in Therapeutic and Adverse Action

Network analysis can be used to identify physiologically relevant targets and the neighborhood within which these targets have their action. The network-building techniques require the selection of a seed list of proteins that are related to the physiology or pathophysiology of interest.

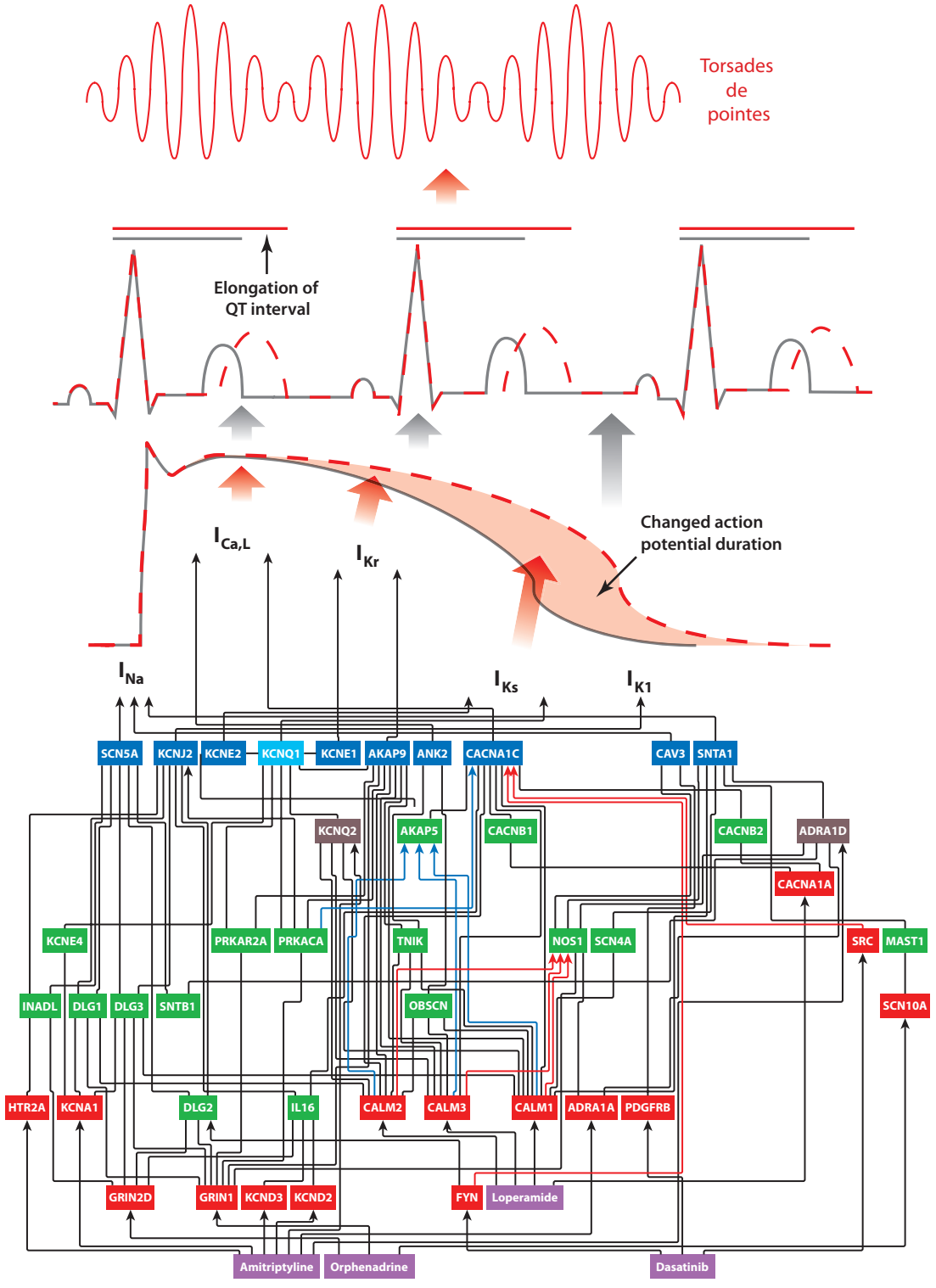
A seed list is a set of related nodes in a network based on some predefined characterization; for example, the list of gene mutations related to the congenital prolonged QT interval used by Berger et al. (15) is a typical seed list. Other examples include lists of genes identified through microarray experiments or genes associated with a particular disease phenotype in the Online Mendelian Inheritance in Man® catalog (OMIM®). Characterization of the seed list can be based on key phrases that describe an organismal-level physiological event such as water retention or clotting. The list of relevant genes/proteins involved in these processes can be identified through the use of databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (52), the Gene Ontology project (53), and Reactome (54), or the list may be manually curated on the basis of the physiology of interest (16, 55). These seed lists serve as inputs to network-building algorithms (e.g., MFPT, nearest-neighbor, or other network distance metrics) that identify proteins within the neighborhood associated with the phenotype of interest. To identify candidates that are more easily druggable, researchers can further filter identified proteins for ontological classifications such as receptors, plasma membrane proteins, or cytoplasmic protein kinases. Thus, both potential drug targets and their functional neighborhood can be identified.

A similar approach can be used with a known drug: One can explore functional distances that are downstream of known drug targets to understand their therapeutic and adverse effects. For example, starting with rosiglitazone as the drug and peroxisome proliferator-activated receptor γ (PPARγ) as the target, we can identify a series of potentially important PPARγ-regulated effectors such as PTGS2 (prostaglandin 2 synthase), SERPINE1 (plasminogen activator inhibitor), VEGFA (vascular endothelial growth factor A), APOB (apolipoprotein B), TSPO (mitochondrial translocator protein), MMP9 (metalloproteinase 9), IL-6 (interleukin-6), CASP3 (caspase 3), and CA2/4 (carbonic anhydrase 2/4). Several of these proteins are associated with cardiac function and myocardial infarction (56–63) and may in part be responsible for the observed association between rosiglitazone and myocardial infarction (64). Thus, using information from known drug targeting and building networks of related proteins can allow us to understand how drugs can have varied effects, some beneficial and others detrimental.

Such network building can also be used to understand how drug-induced effects percolate through multiple layers of organization. This is shown in **Figure 2** for a few drugs that induce long QT syndrome as an adverse event. Targets of these drugs are part of the cellular networks that directly regulate the ion channels whose activity shapes the cardiac myocyte action potential. Changes in this cell physiological phenotype (myocyte action potential) lead to the observed organ-level phenotype: the prolonged QT interval in the electrocardiogram. Such prolongation can lead

Figure 2

An intracellular network to explain how drug-induced adverse events can propagate across scales of organization. Drug interaction with the target leads through the network to an alteration of channel activity, which leads to a change in duration of the myocyte action potential, which leads to prolongation of the QT interval as seen in the electrocardiogram. This can result in fatal arrhythmias such as torsades de pointes. This network explains how drugs used to treat very different pathophysiologies such as diarrhea (loperamide) and cancer (dasatanib) can lead to long QT syndrome as an adverse event. However, this network does not explain why only some people show drug-induced long QT syndrome and why only some patients with long QT syndrome develop fatal arrhythmias. Additional information on genomic status and tissue- and multiorgan-level networks may be needed to explain individual susceptibility. The drugs are shown in purple boxes. Red boxes are drug targets, green boxes are intermediate nodes, and blue boxes are channels responsible for the various phases of the myocyte action potential. The light blue box represents a node that is an intermediate and also a channel involved in myocyte action potential, and each brown box represents a node that is both an intermediate and also a drug target. Black arrows indicate edges that are either undirected or directed with an unknown effect type (inhibition or activation), red arrows indicate edges that are activating, and blue arrows indicate edges that are inhibitory. Abbreviation: I, current that arises from the functioning of the channel protein. Adapted from Berger et al. (15) with permission.



to arrhythmias that can result in life-threatening events such as torsades de pointes. Although this diagram implies that we understand the multiscale mechanism by which these drugs cause long QT syndrome and fatal arrhythmias, the intracellular network does not explain why such adverse events are observed in only a few patients. The answer may lie in building tissue-level networks to explain how changes in myocyte action potential may or may not lead to changes in electrocardiogram profiles. Building such multicellular networks that are anchored in molecular interaction networks will be the required next step to address multiscale biological problems of this kind.

Network Analyses of the FDA Adverse Events Reporting System Database

Building organ-level and organismal-level networks to identify concurrence of therapeutic and adverse phenotypes requires that one define the loci identifying relationships between drug-target interactions and phenotypes in humans. The FDA Adverse Events Reporting System (AERS) is a publicly available database that records drug-induced adverse events in people using one or more drugs. The FDA AERS database, in combination with other databases such as DrugBank, enables one to examine drug target relationships to phenotypes that are adverse events in an unbiased way in the context of multitherapeutic systems without control subjects. This allows us to associate drugs with possible adverse effects without having to conduct specialized trials. AERS can be used to determine relative adverse event profiles. For example, using AERS to select for patients who are being treated for schizophrenia and who experienced tardive dyskinesia as an adverse event, we identify haloperidol, promazine, risperidone, quetiapine, ziprasidone, and clozapine. All these drugs are associated with this adverse event in a clinical trial (65, 66).

Network analysis can be used to associate distal drug targets with particular adverse events. Using these drug targets as seed nodes, we can use methods such as MFPT to look for genes that are closely connected to them. This approach would help us understand and identify potentially useful targets for combination therapies that might mitigate the adverse events or help us predict drug targets that should be avoided because of their potential to lead to adverse events. A major limitation of using the FDA AERS database, however, is that it does not have data on the total number of people using a drug of interest. This information is critical in determining the prevalence of incidents and reporting bias. These problems can be ameliorated in the future through the use of electronic medical records, as they become more commonplace in hospitals and clinics. Mining data from electronic medical records to identify unknown drug interactions and adverse events is likely to be useful.

RELATING TARGETS OF A DRUG TO ITS STRUCTURE

The network analyses described above focus on drug targets as macromolecules that function within the context of cellular regulatory systems. Such an analysis does not take into account atomic-level interactions between the drug and its targets. As we understand more about cellular-level and tissue/organ-level networks, potential drug targets will have to be filtered by structural criteria to determine the ability of the target to be regulated by drugs. Interactions between a drug and its target depend on structural determinants both in the drug and in the target. This is a well-studied area in medicinal chemistry and structural pharmacology. The binding pockets of targets are often analyzed to understand how the drugs fit and the types of conformational changes they induce. Agonists and antagonists interact with the same binding pocket, and the differences in detailed interactions lead to either activation or inhibition of the receptor. The structure of a

drug can be used to determine the targeting of a drug through various computational algorithms (67–72) and for ADMET (absorption, distribution, metabolism, excretion, and toxicity) predictions (73). The identification of targets that bind or metabolize structural variants of a drug differently can serve as the starting point for network analysis in the identification of off-target physiological events. However, there are no tools to conduct such scalable computation easily. To build such integrated networks, we require knowledge of the structures of both the drug and the targets of the drug. Often, target structures are not readily available. This lack of structural knowledge can lead to false-positive drug-target interactions. The potential of false positives may be decreased through the application of orthogonal experimental techniques such as high-content screening (74), in vitro ADMET (75), medicinal chemistry techniques (76, 77), binding screens (78), and gene regulation (79). Methods based on binding screens and gene regulation may also be used to develop drug-specific target interactions as this information is made available through the differential binding interactions and gene regulation with and without the drug.

PERSPECTIVE

Although in its infancy, the field of systems pharmacology has enormous potential to impact both drug development and drug usage in the future. Currently, drug development is largely focused on noncommunicable diseases such as cancers, metabolic diseases, type 2 diabetes, psychiatric disorders, and immune disorders, as well as communicable diseases wherein the pathology arises from complex host-pathogen interactions and is not susceptible to simple treatment approaches such as the use of antibiotics. Developing drugs for these diseases through classical empirical methods has not proven to be productive. Many molecules that show good therapeutic effects in cellular or animal models fail to be efficacious in humans. These failures arise from our lack of understanding of human biology as defined by the multiscale mechanisms that underlie the propagation of effects from molecular-level drug-target interactions to organismal-level phenotypes. In addition to understanding and predicting efficacy of therapy, it is becoming increasingly imperative that treatments are personalized so that the risks associated with the drug therapy are understood before treatment commences. These considerations lead to a set of key questions (see sidebar, Questions in Systems Pharmacology) that need to be addressed by research in the field of systems pharmacology. As such research progresses, the pace of drug discovery and therapy should become more proportional to the pace of discovery in basic biomedical sciences.

QUESTIONS IN SYSTEMS PHARMACOLOGY

1. What are the characteristics of diseases for which drugs at a single target may not be therapeutically efficacious?
2. How does intracellular and intercellular networking give rise to adverse events?
3. How do we relate the efficacy of (poly)pharmacology to the genomic status of the individual, and how does genomic status interact with environment and behavior to control (poly)pharmacology efficacy?
4. How do we determine what combinations of targets are most likely to be effective for polypharmacology of complex diseases?
5. Can we use the human interactome and the genomic status of the individual to predict therapeutic efficacy and adverse event probability prior to commencement of therapy?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors' research is supported by grant GM54508 and Systems Biology Center New York grant P50-GM071558, both from the NIH's National Institute of General Medical Sciences (NIGMS). S.Z. is currently supported by a predoctoral training grant in pharmacological sciences (T32-GM062754) and is a trainee in the NIGMS Medical Scientist Training Program.

LITERATURE CITED

1. Li JW-H, Vederas JC. 2009. Drug discovery and natural products: end of an era or an endless frontier? *Science* 325:161–65
2. Fitzgerald G. 2010. Drug development needs a new brand of science. *Nature* 468:869
3. Brunton LL, Chabner BA, Knollman BC, eds. 2010. *Goodman & Gilman's The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill. 1808 pp. 12th ed.
4. Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature* 468:911–20
5. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. 2011. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* 39:D698–704
6. Zhao S, Li S. 2010. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS ONE* 5:e11764
7. Barabási A-L, Gulbahce N, Loscalzo J. 2011. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12:56–68
8. Vidal M, Cusick ME, Barabási A-L. 2011. Interactome networks and human disease. *Cell* 144:986–98
9. Ferrell JE, Tsai TY-C, Yang Q. 2011. Modeling the cell cycle: Why do certain circuits oscillate? *Cell* 144:874–85
10. Singhania R, Sramkoski RM, Jacobberger JW, Tyson JJ. 2011. A hybrid model of mammalian cell cycle regulation. *PLoS Comput. Biol.* 7:e1001077
11. Neves SR, Tsokas P, Sarkar A, Grace EA, Rangamani P, et al. 2008. Cell shape and negative links in regulatory motifs together control spatial information flow in signaling networks. *Cell* 133:666–80
12. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36:D901–6
13. Zhao F, Xuan Z, Liu L, Zhang MQ. 2005. TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.* 33:D103–7
14. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. 2007. The human disease network. *Proc. Natl. Acad. Sci. USA* 104:8685–90
15. Berger SI, Ma'ayan A, Iyengar R. 2010. Systems pharmacology of arrhythmias. *Sci. Signal.* 3:ra30
16. Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, et al. 2005. Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* 309:1078–83
17. Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, et al. 2009. VirusMINT: a viral protein interaction database. *Nucleic Acids Res.* 37:D669–73
18. Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S. 2009. MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics* 25:690–91
19. Sharan R, Ulitsky I, Shamir R. 2007. Network-based prediction of protein function. *Mol. Syst. Biol.* 3:88
20. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, et al. 2009. PID: the Pathway Interaction Database. *Nucleic Acids Res.* 37:D674–79
21. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K, et al. 2004. Mutations in *VKORC1* cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 427:537–41

22. Aithal GP, Day CP, Kesteven PJ, Daly AK. 1999. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 353:717–19
23. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, et al. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327:78–81
24. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–35
25. Rix U, Remsing Rix LL, Terker AS, Fernbach NV, Hantschel O, et al. 2010. A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells. *Leukemia* 24:44–50
26. Muroi M, Kazami S, Noda K, Kondo H, Takayama H, et al. 2010. Application of proteomic profiling based on 2D-DIGE for classification of compounds according to the mechanism of action. *Chem. Biol.* 17:460–70
27. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. 2008. Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* 48:653–83
28. Kaddurah-Daouk R, McEvoy J, Baillie RA, Lee D, Yao JK, et al. 2007. Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Mol. Psychiatry* 12:934–45
29. Ji Y, Hebbring S, Zhu H, Jenkins GD, Biernacka J, et al. 2010. Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. *Clin. Pharmacol. Ther.* 89:97–104
30. Evans WE, McLeod HL. 2003. Pharmacogenomics—drug disposition, drug targets, and side effects. *N. Engl. J. Med.* 348:538–49
31. Drazen JM, Silverman EK, Lee TH. 2000. Heterogeneity of therapeutic responses in asthma. *Br. Med. Bull.* 56:1054–70
32. Mougey EB, Feng H, Castro M, Irvin CG, Lima JJ. 2009. Absorption of montelukast is transporter mediated: a common variant of OATP2B1 is associated with reduced plasma concentrations and poor response. *Pharmacogenet. Genomics* 19:129–38
33. De Roock W, Jonker DJ, Di Nicolantonio F, Sartore-Bianchi A, Tu D, et al. 2010. Association of *KRAS* p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA* 304:1812–20
34. Ciszkowski C, Madadi P, Phillips MS, Lauwers AE, Koren G. 2009. Codeine, ultrarapid-metabolism genotype, and postoperative death. *N. Engl. J. Med.* 361:827–28
35. Dalén P, Dahl ML, Bernal Ruiz ML, Nordin J, Bertilsson L. 1998. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional *CYP2D6* genes. *Clin. Pharmacol. Ther.* 63:444–52
36. Beer B, Erb R, Pitterl F, Niederstätter H, Maroñas O, et al. 2011. CYP2D6 genotyping by liquid chromatography-electrospray ionization mass spectrometry. *Anal. Bioanal. Chem.* 400(8):2361–70
37. Guengerich FP. 2002. Cytochrome P450 enzymes in the generation of commercial products. *Nat. Rev. Drug Discov.* 1:359–66
38. Treviño LR, Shimasaki N, Yang W, Panetta JC, Cheng C, et al. 2009. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J. Clin. Oncol.* 27:5972–78
39. van Erp NP, Eechoute K, van der Veldt AA, Haanen JB, Reyners AKL, et al. 2009. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J. Clin. Oncol.* 27:4406–12
40. Heinrich MC, Corless CL, Demetri GD, Blanke CD, von Mehren M, et al. 2003. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J. Clin. Oncol.* 21:4342–49
41. Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, et al. 2005. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 352:786–92
42. Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, et al. 2005. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* 2:e73
43. Mallal S, Phillips E, Carosi G, Molina J-M, Workman C, et al. 2008. HLA-B*5701 screening for hypersensitivity to abacavir. *N. Engl. J. Med.* 358:568–79

44. Ingle JN, Schaid DJ, Goss PE, Liu M, Mushiroda T, et al. 2010. Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *J. Clin. Oncol.* 28:4674–82
45. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, et al. 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J.* 1:167–70
46. Abraham AK, Maurer TS, Kalgutkar AS, Gao X, Li M, et al. 2011. Pharmacodynamic model of parathyroid hormone modulation by a negative allosteric modulator of the calcium-sensing receptor. *AAPS J.* 13:265–73
47. Bhalla US, Iyengar R. 1999. Emergent properties of networks of biological signaling pathways. *Science* 283:381–87
48. Tanaka K, Augustine GJ. 2008. A positive feedback signal transduction loop determines timing of cerebellar long-term depression. *Neuron* 59:608–20
49. Panetta JC, Sparreboom A, Pui C-H, Relling MV, Evans WE. 2010. Modeling mechanisms of in vivo variability in methotrexate accumulation and folate pathway inhibition in acute lymphoblastic leukemia cells. *PLoS Comput. Biol.* 6:e1001019
50. Dijkstra EW. 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1:269–71
51. Huang T, Wang P, Ye Z-Q, Xu H, He Z, et al. 2010. Prediction of deleterious non-synonymous SNPs based on protein interaction network and hybrid properties. *PLoS ONE* 5:e11900
52. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38:D355–60
53. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25:25–29
54. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, et al. 2009. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 37:D619–22
55. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GSS, et al. 2010. NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* 11:R3
56. Holly TA, Drincic A, Byun Y, Nakamura S, Harris K, et al. 1999. Caspase inhibition reduces myocyte cell death induced by myocardial ischemia and reperfusion in vivo. *J. Mol. Cell. Cardiol.* 31:1709–15
57. Schaller S, Paradis S, Ngoh GA, Assaly R, Buisson B, et al. 2010. TRO40303, a new cardioprotective compound, inhibits mitochondrial permeability transition. *J. Pharmacol. Exp. Ther.* 333:696–706
58. Abbott JD, Huang Y, Liu D, Hickey R, Krause DS, Giordano FJ. 2004. Stromal cell-derived factor-1α plays a critical role in stem cell recruitment to the heart after myocardial infarction but is not sufficient to induce homing in the absence of injury. *Circulation* 110:3300–5
59. Diedrich M, Tadic J, Mao L, Wacker MA, Nebrich G, et al. 2007. Heart protein expression related to age and sex in mice and humans. *Int. J. Mol. Med.* 20:865–74
60. Morange PE, Saut N, Alessi MC, Yudkin JS, Margaglione M, et al. 2007. Association of plasminogen activator inhibitor (PAI)-1 (SERPINE1) SNPs with myocardial infarction, plasma PAI-1, and metabolic parameters: the HIFMECH study. *Arterioscler. Thromb. Vasc. Biol.* 27:2250–57
61. Shintani S, Murohara T, Ikeda H, Ueno T, Honma T, et al. 2001. Mobilization of endothelial progenitor cells in patients with acute myocardial infarction. *Circulation* 103:2776–79
62. Mukherjee D, Nissen SE, Topol EJ. 2011. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA* 286:954–59
63. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, et al. 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 364:937–52
64. Nissen SE, Wolski K. 2007. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N. Engl. J. Med.* 356:2457–71
65. Caroff SN, Hurford I, Lybrand J, Campbell EC. 2011. Movement disorders induced by antipsychotic drugs: implications of the CATIE schizophrenia trial. *Neurol. Clin.* 29:127–48
66. Saltz BL, Woerner MG, Kane JM, Lieberman JA, Alvir JM, et al. 1991. Prospective study of tardive dyskinesia incidence in the elderly. *JAMA* 266:2402–6

67. Schmidtke P, Le Guilloux V, Maupetit J, Tufféry P. 2010. Fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.* 38(Suppl.):W582–89
68. Ballester PJ, Mitchell JBO. 2010. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 26:1169–75
69. Thangudu RR, Tyagi M, Shoemaker BA, Bryant SH, Panchenko AR, Madej T. 2010. Knowledge-based annotation of small molecule binding sites in proteins. *BMC Bioinformatics* 11:365
70. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. 2010. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* 6:e1000938
71. Kitchen DB, Decornez H, Furr JR, Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3:935–49
72. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. 2007. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25:197–206
73. Schneider G, Fechner U. 2005. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* 4:649–63
74. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, et al. 2011. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10:188–95
75. Gleeson MP, Hersey A, Montanari D, Overington J. 2011. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* 10:197–208
76. Wilson ID. 2011. High-performance liquid chromatography-mass spectrometry (HPLC-MS)-based drug metabolite profiling. *Methods Mol. Biol.* 708:173–90
77. Lenz EM. 2011. Metabolic profiling. *Methods Mol. Biol.* 708:299–319
78. Clemons PA, Bodycombe NE, Carrinski HA, Wilson JA, Shamji AF, et al. 2010. Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *Proc. Natl. Acad. Sci. USA* 107:18787–92
79. Iskar M, Campillos M, Kuhn M, Jensen LJ, van Noort V, Bork P. 2010. Drug-induced regulation of target expression. *PLoS Comput. Biol.* 6:e1000925
80. Kim KA, Park PW, Hong SJ, Park J-Y. 2008. The effect of CYP2C19 polymorphism on the pharmacokinetics and pharmacodynamics of clopidogrel: a possible mechanism for clopidogrel resistance. *Clin. Pharmacol. Ther.* 84:236–42
81. Wadelius M, Sörlin K, Wallerman O, Karlsson J, Yue Q-Y, et al. 2004. Warfarin sensitivity related to CYP2C9, CYP3A5, ABCB1 (MDR1) and other factors. *Pharmacogenomics J.* 4:40–48
82. Ishikawa T, Sekiguchi F, Fukase Y, Sawada N, Ishitsuka H. 1998. Positive correlation between the efficacy of capecitabine and doxifluridine and the ratio of thymidine phosphorylase to dihydropyrimidine dehydrogenase activities in tumors in human cancer xenografts. *Cancer Res.* 58:685–90
83. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, et al. 2008. Wild-type *KRAS* is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol.* 26:1626–34
84. Debiec-Rychter M, Dumez H, Judson I, Wasag B, Verweij J, et al. 2004. Use of *c-KIT/PDGFRA* mutational analysis to predict the clinical response to imatinib in patients with advanced gastrointestinal stromal tumours entered on phase I and II studies of the EORTC Soft Tissue and Bone Sarcoma Group. *Eur. J. Cancer* 40:689–95
85. Kakizuka A, Miller WH, Umesono K, Warrell RP, Frankel SR, et al. 1991. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RARα with a novel putative transcription factor, PML. *Cell* 66:663–74
86. Chung W-H, Hung S-I, Hong H-S, Hsih M-S, Yang L-C, et al. 2004. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* 428:486
87. Mallal S, Nolan D, Witt C, Masel G, Martin A, et al. 2002. Association between presence of *HLA-B*5701*, *HLA-DR7*, and *HLA-DQ3* and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* 359:727–32
88. Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, et al. 2011. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* 469:241–44
89. Erlanson DA, Arndt JW, Cancilla MT, Cao K, Elling RA, et al. 2011. Discovery of a potent and highly selective PDK1 inhibitor via fragment-based drug discovery. *Bioorg. Med. Chem. Lett.* 21:3083–78

Annual Review of Pharmacology and Toxicology

Volume 52, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml

Using Genome-Wide Association Studies to Identify Genes Important in Serious Adverse Drug Reactions

Ann K. Daly

Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom; email: a.k.daly@ncl.ac.uk

Annu. Rev. Pharmacol. Toxicol. 2012. 52:21–35

First published online as a Review in Advance on August 1, 2011

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-010611-134743

0362-1642/12/0210-0021$20.00

Keywords

adverse drug reaction, liver toxicity, hypersensitivity, myotoxicity, cardiotoxicity, HLA

Abstract

Genome-wide association (GWA) studies have detected novel associations for serious, idiosyncratic, adverse drug reactions including liver toxicity, hypersensitivity, skin rash, and myotoxicity. Human leukocyte antigen (HLA) genotype has been established as an important predictor of susceptibility to drug-induced liver injury, including injury with some drugs where immune-related toxicity was not suspected previously. Similarly, GWA studies have shown a key role for HLA genotype in susceptibility to carbamazepine-related skin rash and hypersensitivity. HLA genotype is not a risk factor for all forms of drug-induced liver injury or for myotoxicity or cardiotoxicity. For simvastatin-related myotoxicity, a strong association with *SLCO1B1*, which encodes the hepatic statin uptake transporter, has been detected. Genome-wide studies have not yet found clear associations for drug-induced cardiotoxicity, but for bisphosphonate-induced necrosis of the jaw, polymorphisms in the cytochrome P450 CYP2C8 may predict susceptibility. Larger GWA studies and whole-genome sequencing may provide additional insights into all these toxicities.

INTRODUCTION

A serious adverse drug reaction is defined as an undesirable experience concerned with a particular drug and that leads to any of the following: death or life-threatening event, hospitalization, disability or permanent damage, congenital abnormality or birth defect (1, 2). Such events may occur during drug development or may emerge only when the drug has been licensed. Most serious adverse drug reactions can be classified as either type A, where the underlying mechanism is dose dependent, or type B or idiosyncratic, where the event is not predictable from the normal pharmacology of the drug and is generally independent of dose (3, 4). Idiosyncratic adverse events are generally rarer than type A events, and this lower frequency means that they usually emerge late in the drug development process or after licensing. In the period from 1976 to 2005, 28 different drugs were withdrawn from the market in the United States as a result of idiosyncratic serious adverse reactions (5, 6). In particular, the types of reaction involved cardiotoxicity (including torsades de pointes) (28% of withdrawals), hepatotoxicity (21% of withdrawals), nephrotoxicity (7% of withdrawals), and rhabdomyolysis (7% of withdrawals). Additional, more minor, contributions from other toxicities included skin rash and hemolytic anemia.

In addition to these high-profile withdrawals of otherwise valuable drugs, there are a number of examples of licensed drugs that give rise, though only on rare occasions, to serious adverse reactions. Usually, the label includes a warning regarding the possibility of an adverse reaction. These drugs continue to be prescribed, often because of an absence of effective alternatives. There is also increasing interest in developing screening tests that would enable researchers to predict which patients are at risk of suffering adverse drug reactions. Such a development may potentially allow for the reintroduction of some valuable drugs that had been withdrawn previously; this development may also help avoid some of the current serious adverse reactions seen with licensed drugs.

The possibility that genetic tests could be developed for certain types of serious adverse reactions has been investigated for at least the past 20 years (6). There is now considerable evidence that genetic factors contribute to susceptibility to these reactions (6, 7), although it is important to stress that other factors may also contribute. The development of genome-wide association (GWA) studies and their successful application to the identification of novel susceptibility genes for several complex polygenic diseases (8) resulted in an interest in applying GWA studies to the area of serious adverse drug reactions. There are specific problems in doing this: For example, the rarity of serious adverse reactions means that multicenter studies, often involving several countries, may be needed to recruit adequate numbers of cases. Furthermore, a uniform phenotype including clear causality must also be ensured. Nonetheless, a number of GWA studies on serious adverse drug reactions have now been completed and are discussed in detail in this article. Most of these studies relate to either drug-induced hepatotoxicity or hypersensitivity reactions including skin rash, but there are also single studies on drug-induced myotoxicity, cardiotoxicity, and osteonecrosis of the jaw (see **Table 1**). After a general introduction to GWA studies, this review focuses on each type of serious adverse drug reaction for which GWA data are currently available.

GENOME-WIDE ASSOCIATION STUDIES TO IDENTIFY SUSCEPTIBILITY GENES

Until 2004, the main approach to the study of complex disease genetics was the use of candidate-gene case-control studies, with additional family studies possible for some, but not the majority of, diseases. Overall, these studies led to the reproducible identification of only a small number of disease genes (9), because most studies involved relatively small numbers of patients and the polymorphisms chosen for study in candidate genes were often selected arbitrarily. Since 2005,

Table 1 Summary of published genome-wide association studies on serious adverse drug reactions

Type of toxicity	Number of published studies	Drugs involved	Genes implicated	Highest level of significance (lowest p value)	Reference(s)
Liver	4	Ximelagatran, Flucloxacillin, Lumiracoxib, Amoxicillin-clavulanate	HLA classes I and II	8.7×10^{-33}	37, 39, 40, 42
Skin and hypersensitivity	3	Carbamazepine plus miscellaneous agents	*HLA-A*	1.2×10^{-13}	58, 59, 62
Myotoxicity	1	Simvastatin	*SLCO1B1*	4.0×10^{-9}	68
QT prolongation	1	Iloperidone	*CERKL*	2.8×10^{-6}	78
Osteonecrosis of the jaw	1	Pamidronate, zoledronic acid	*CYP2C8*	6.2×10^{-6}	80

the availability of comprehensive data on variability in human genes from the HapMap (10) together with the development of methods allowing very large numbers of single-nucleotide polymorphisms (SNPs) to be genotyped simultaneously (11) have revolutionized the study of complex disease genetics. Using the HapMap information, researchers have covered the majority of common (>5%) variability in the human genome by genotyping sets of cases and controls for large numbers (typically 500,000 to 1,000,000) of different SNPs in GWA studies. A particular advantage of GWA studies over candidate-gene studies is their open nature because the genotyping is performed for polymorphisms in all genes, not simply those that are obvious candidates for effects on the disease of interest. Because large numbers of polymorphisms are being genotyped, it is necessary to analyze the results using statistical tests that correct appropriately for the large number of assays being performed, but these methods are now well established. Manolio et al. (12) and Hardy & Singleton (13) provide more detailed description of all aspects of GWA.

In GWA studies on common complex genetic diseases, typically at least 1,000 cases and 1,000 controls need to be studied to detect the relatively low odds ratios seen for most SNPs affecting disease susceptibility, as found in, for example, type II diabetes and breast cancer (14, 15). GWA studies involving similarly large numbers of cases have also been successfully performed in relation to drug response, including responses to warfarin (16), acenocoumarol (17), and clopidogrel (18). However, for studies on serious idiosyncratic adverse drug reactions, which may typically occur at frequencies of 1 in 10,000 or 100,000 patients treated (3), finding sufficient cases for GWA studies is more difficult. Though investigators have set up several large networks to enable large numbers of cases with serious adverse drug reactions to be recruited (19–21), the GWA studies reported to date include several hundred cases at most; thus, the power to detect small effects is limited. GWA findings are normally confirmed using a replication cohort (13), usually of similar size to or greater size than the original group of cases studied, which is particularly difficult given the challenges of finding sufficient cases for the initial study. In spite of these limitations, some progress in understanding genetic susceptibility to serious adverse drug reactions has been made (**Table 1**).

LIVER-RELATED ADVERSE DRUG REACTIONS

Idiosyncratic hepatotoxicity relating to drug exposure is usually referred to as drug-induced liver injury (DILI), a rare but clinically important problem. Drugs that give rise to this toxicity are

structurally diverse and belong to a number of different therapeutic classes, but certain antimicrobial agents and nonsteroidal anti-inflammatory drugs are among the most common causes of idiosyncratic DILI (22–24). One U.S.-based survey suggested that DILI accounts for 20% of all hospital admissions due to severe liver injury and 50% of acute liver failure cases, 75% of which require liver transplantation (25). DILI is also the most common reason that clinical trials of new therapeutic agents are terminated (26). Many different drugs can cause DILI, with the precise pattern of injury varying between drugs. Typically, DILI reactions are classified as hepatocellular when the injury is focused on the hepatocyte and cholestatic when the damage occurs at the hepatocyte canalicular membrane or further downstream in the biliary tree (27). The underlying mechanism by which DILI develops is likely to be complex but may involve both direct toxic effects by the drug, for example, involving oxidative stress or cellular damage, and, for some drugs, formation of reactive intermediates resulting in either direct toxicity or an inappropriate immune response (28).

The first study on a possible genetic association for DILI susceptibility appeared more than 20 years ago as a report showing an increased incidence in frequency of certain human leukocyte antigen (HLA) class II serotypes among DILI cases compared with controls (29). These cases included DILI induced by several drugs. A number of further reports of associations with particular HLA serotypes and genotypes followed, including, in particular, two independent reports suggesting that the HLA class II allele *DRB1*1501* was a risk factor for DILI induced by the antimicrobial agent, amoxicillin-clavulanate (30, 31). This form of DILI has been suggested to relate predominantly to the clavulanic acid component of the drug (32), though this has still not been demonstrated directly. Candidate-gene association studies have also led to the detection of several other associations with non-HLA genes, either for DILI due to individual drugs (33, 34) or for cases of this adverse drug reaction linked to a range of different drugs (35, 36).

GWA approaches have now been used to investigate susceptibility to hepatotoxicity in four studies, each involving different drugs associated with DILI (see **Table 2** for a summary). Slightly unexpectedly, all four studies have shown statistically significant associations with particular HLA class I or II alleles, suggesting that T cell responses contribute to the toxicity, but no significant

Table 2 Genome-wide association studies on drug-induced liver injury

Drug	Number of cases	SNP(s)[a] showing lowest p value	p value[b]	Odds ratio (95% CI)[b]	Gene and allele tagged by SNP	Reference
Ximelagatran	74	rs2858869	6.0×10^{-6}	Not done	*HLA-DRB1*0701-DQA1*0201*	37
Flucloxacillin	51	rs2395029	8.7×10^{-33}	45 (19.4–105)	*HLA-B*5701*	39
Lumiracoxib	41	rs9270986	2.8×10^{-10}	5.3 (3.0–9.2)	*HLA-DRB1*1501-DQB1*0602*	40
Amoxicillin-clavulanate	201	rs9274407	4.8×10^{-14}	3.1 (2.3–4.2)	None	42
		rs9267992	6.8×10^{-13}	3.1 (2.3–4.2)	None	
		rs3135388	3.5×10^{-11}	2.8 (2.1–3.8)	*HLA-DRB1*1501-DQB1*0602*	
		rs2523822	1.8×10^{-10}	2.3 (1.8–2.9)	*HLA-A*0201*	

[a]SNP, single-nucleotide polymorphism.
[b]Based on allele frequency for SNP.

genome-wide non-HLA associations were detected in any of the studies. The association with HLA is not unexpected for amoxicillin-clavulanate, as this had been demonstrated previously by two smaller candidate-gene association studies (30, 31).

The earliest of the four GWA studies relating to DILI focused on the direct thrombin inhibitor ximelagatran, which was developed as a potential replacement for warfarin and other coumarin anticoagulants but was withdrawn by the manufacturers in 2006 (37). This drug was associated with raised alanine aminotransferase (ALT) levels (transaminitis) among some patients, though this elevation was relatively small in most, but not all, affected individuals. A GWA study performed on 74 cases of transaminitis linked to ximelagatran was the earliest reported GWA on any serious adverse reaction. This study involved a set of 266,000 SNPs, whereas the three more recent studies involved between 700,000 and 1,000,000 SNPs. The study on ximelagatran also involved genotyping for a range of additional candidate genes to increase coverage of key SNPs and separate HLA genotyping. Both GWA and candidate SNPs were tested for association with raised ALT rather than a control group. The GWA study failed to detect any significant GWAs, but a relatively low p value (6.0×10^{-6}) was obtained for a SNP in the HLA class II *DRB1* gene. Further direct HLA typing confirmed a significant association between the level of ALT increase and *HLA-DRB1*0701* ($p = 4 \times 10^{-5}$) (37).

This first slightly limited, though interesting, study on a form of DILI involving GWA was followed by a study on flucloxacillin-related DILI. Flucloxacillin is an example of a drug prescribed commonly in a number of countries worldwide but which is associated occasionally (in <1 in 10,000 patients prescribed the drug) with DILI (38). The DNA samples studied were from 51 patients of Northern European ethnic origin who had suffered moderate to severe DILI with a clear causal link to flucloxacillin (39). The GWA study involved genotyping of these cases and matched populations controls for approximately 900,000 SNPs. A number of SNPs in the major histocompatibility complex (MHC) region of chromosome 6 where HLA genes are located showed genome-wide significant p values, with the top SNP in the *HCP5* gene showing a p value of 8.7×10^{-33}. This SNP is in strong linkage disequilibrium with the class I HLA allele *B*5701*. Direct HLA typing confirmed that carriage of *B*5701* was a strong risk factor for flucloxacillin-related DILI {odds ratio 80.6 [95% confidence interval (CI) 23–285] based on carriage of at least one *B*5701* allele}; this finding was also duplicated in a replication cohort of 16 further cases. Despite the strong association, only 1 in 500 to 1,000 *B*5701*-positive individuals prescribed flucloxacillin are predicted to develop DILI, so genotyping for this allele prior to flucloxacillin prescription is unlikely to be a useful pharmacogenetic test. It does, however, have potential as a diagnostic in suspected DILI cases (39).

A more recent, slightly smaller GWA study involved 41 cases of DILI collected during a phase III clinical trial of the nonsteroidal anti-inflammatory drug lumiracoxib, which has recently been withdrawn from the market in a number of countries owing to a relatively high incidence (approximately 2.6%) of raised ALT levels in users (40). The 41 DILI cases, together with lumiracoxib-exposed controls who had not suffered DILI, were genotyped for approximately 700,000 SNPs. A number of SNPs in the MHC region showed genome-wide significance, with the lowest p value at 4.4×10^{-12}. As with the entirely separate flucloxacillin DILI study, further HLA typing was performed, and a clear association with the HLA class II allele *DRB1*1501* was detected (odds ratio 5.0, 95% CI 3.6 –7.0). Though the overall HLA association for lumiracoxib DILI was less strong than that for flucloxacillin DILI, typing for a particular HLA allele (*DQA1*0102*) in linkage disequilibrium with *DRB1*1501* prior to prescription may help prevent doctors from prescribing lumiracoxib to the 34% of Europeans positive for this allele. This interesting approach is being considered as a possible means of reintroducing lumiracoxib to the market worldwide. By genotyping, the incidence of DILI in lumiracoxib users could drop to less than 1% by excluding those positive for *DQA1*0102* (40, 41).

The largest GWA study on DILI so far reported relates to the drug amoxicillin-clavulanate (mentioned above). Amoxicillin-clavulanate-related DILI has a number of features in common with flucloxacillin-related DILI in terms of frequency and type of liver injury. It also provides another example of a widely prescribed drug showing occasional liver toxicity. In a study of 201 DILI cases of European ethnic origin with ethnically matched population controls, representing the largest GWA study on DILI reported to date, genome-wide significance was again seen in the MHC region (42). The most significant SNPs (lowest p value $= 4.8 \times 10^{-14}$) localized to both the HLA class I and class II regions, whereas the previous GWA studies had found associations within only either class I or class II. Detailed HLA typing provided evidence that both *DRB1*1501-DQB1*0602* and *A*0201* were risk factors for development of DILI [odds ratios 3.3 (95% CI 2.0–5.7) and 2.2 (95% CI 1.6–3.2), respectively] and that there was a statistically significant genetic interaction between these alleles increasing the risk of DILI in individuals positive for both. The *DRB1*1501-DQB1*0602* association was in agreement with previous reports (30, 31), but the *A*0201* association was novel (42). These data are significant, but as with the findings for flucloxacillin, overall positive predictive value is low and the only possible use of genotyping for the risk factors described here would be for diagnostic purposes.

Though some DILI cases show features such as rash and/or eosinophilia that may indicate an HLA or other immune system association, only a minority of cases overall show such features. One particular feature of the DILI GWA study findings is that apparent associations with the HLA haplotype *DRB1*1501-DQB1*0602-DQA1*0102* have been detected for both amoxicillin-clavulanate and lumiracoxib-related DILI (40, 42). These compounds showing similar HLA associations for DILI are not obviously structurally similar (see **Figure 1**). In addition, there are phenotypic differences in the pattern of liver injury observed with the two drugs (40, 42). The association between *HLA-B*5701* and flucloxacillin-induced DILI is also seen for abacavir-induced hypersensitivity reactions that normally do not affect the liver (43), but the positive predictive value for *B*5701* in abacavir hypersensitivity is considerably higher than that for flucloxacillin DILI (39). There is also no obvious structural similarity between flucloxacillin and abacavir (**Figure 1**).

All four GWA studies on DILI involved smaller numbers of cases (between 41 and 201) than have been studied in GWA investigations on complex diseases. This has still enabled the detection of the relatively strong HLA associations seen for these drug reactions, but the power to detect weaker associations, perhaps involving non-HLA genes, is more limited. In addition, in a recent study on DILI relating to the anticancer drug lapatanib, an initial GWA study failed to detect any genome-wide significant associations, but in further candidate-gene analysis, a significant association was found with the HLA class II allele *DQA1*0201* (44), which forms part of the haplotype also associated with ximelagatran transaminitis (37). As with the other examples of HLA alleles associated with serious adverse drug reactions, these two compounds are not structurally related (**Figure 1**).

The mechanism underlying the HLA associations seen in DILI remains unclear. GWA studies have covered a wide range of genetic markers across the MHC region, and the strongest associations have been localized to specific class I and II HLA genes. However, although there is still no direct evidence that these gene products are causal, the parent drug or a metabolite either may interact directly with specific HLA class I or II proteins in an antigen presentation reaction to T cells or may form a covalent complex with intracellular proteins that is then cleaved, recognized specifically by certain HLA molecules, and presented to T cells.

Most of the drugs shown in **Figure 1** are subject to metabolism, though data on whether clavulanic acid undergoes enzyme-mediated metabolism are limited (45, 46). However, for DILI related to the drugs on which GWA studies have been performed, no association between DILI and genes concerned with drug disposition has been detected, despite the excellent representation

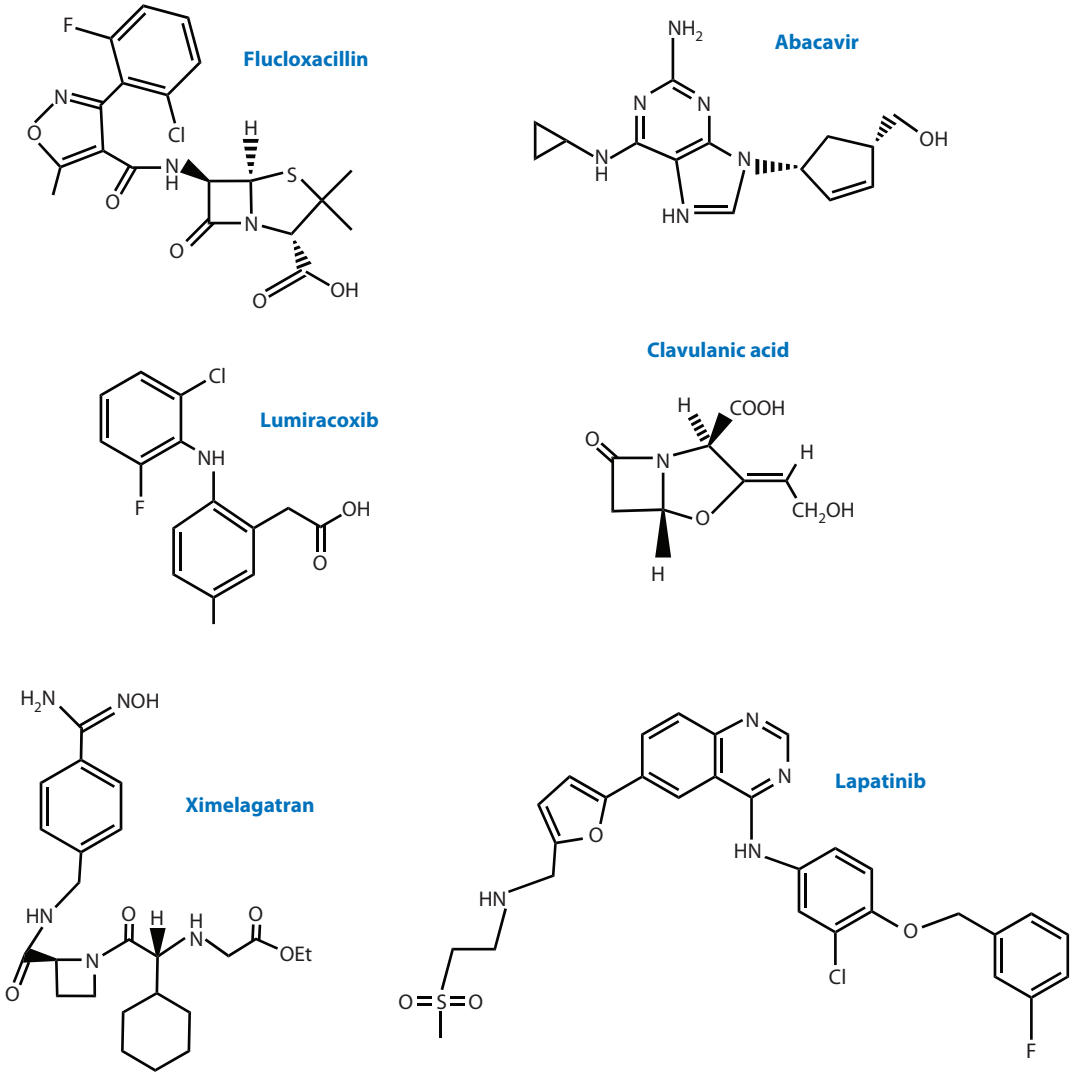


Figure 1

Chemical structures of drugs showing human leukocyte antigen (HLA) associations with drug-induced liver injury. The structures of ximelagatran, flucloxacillin, lumiracoxib, clavulanic acid, and lapatanib are indicated. Also shown is the structure of abacavir, which does not give rise to liver injury but does give rise to hypersensitivity reactions that, similar to flucloxacillin-induced liver injury, are also associated with *HLA-B*5701*.

of relevant SNPs on the platforms used for genotyping. It remains possible that only rare variants in these genes are relevant or that there is inadequate power to detect associations with more common variants given the small effect sizes and the relatively small numbers of cases included in the GWA studies.

Though HLA-related SNPs were the only SNPs showing genome-wide significance in the published GWA studies, two of the studies detected some additional SNPs showing possible associations. In particular for flucloxacillin DILI, the analysis suggested that a gene that has a possible role in B cell immune responses and is expressed in the liver, ST6 β-galactosamide

α-2,6-sialyltranferase 1 (*ST6GAL1*) (47), also contributed to flucloxacillin toxicity. To observe genome-wide significance for the *ST6GAL1* SNP, it was necessary to exclude any samples that were negative for *B***5701* (39). In the GWA study on amoxicillin-clavulanate DILI (42), additional analyses were performed to assess the contribution of SNPs relevant to drug disposition and to autoimmune disease. As discussed above, no significant associations with the SNPs in genes relevant to drug disposition were detected, but when the SNPs relevant to autoimmune disease were examined, two SNPs in *PTPN22*, which encodes the lymphoid-specific protein tyrosine phosphatase, nonreceptor type 22 involved in T-cell-receptor signaling, showed relatively low p values that remained statistically significant after correction for multiple testing.

To date, GWA studies on DILI have focused either on drugs that are very widely prescribed and occasionally give rise to DILI (flucloxacillin and amoxicillin-clavulanate) or are either newly licensed or still in development with a liability to give rise to DILI detected during clinical trials (ximelagatran and lumiracoxib). Assembling suitably sized groups of DILI cases due to other drugs occasionally associated with DILI such as specific NSAIDs, statins, and certain other antimicrobials to enable further GWA studies to be performed with adequate statistical power is more challenging, though achievable via international collaboration. Whether the HLA genotype will be the strongest risk factor for DILI linked to these other drugs is still unclear. There is evidence from candidate-gene association studies that drug-metabolism genes contribute to susceptibility to some forms of DILI, for example, *NAT2* in the case of isoniazid-related DILI (reviewed in Reference 48). It would be of value if such findings could be confirmed by GWA studies.

HYPERSENSITIVITY AND SKIN REACTIONS

A hypersensitivity reaction is an inappropriate immune reaction to an otherwise nontoxic agent. The manifestations of hypersensitivity reactions are broad. Certain forms of DILI, as discussed above, can be regarded as hypersensitivity reactions, for example, flucloxacillin and coamoxiclav-induced liver injury. Skin reactions, which may also involve other organs such as the liver, lungs, or kidneys, are the most common type of drug-induced hypersensitivity reactions (49).

The antiretroviral drug abacavir is associated with hypersensitivity: Up to 8% of patients prescribed this drug suffer symptoms including fever, malaise, gastrointestinal symptoms, and internal organ involvement (49). Skin rash is also seen in many affected individuals. As discussed in more detail in Liver-Related Adverse Drug Reactions (section above), almost all individuals who suffer abacavir hypersensitivity are positive for *HLA-B***5701*, the HLA allele also associated with flucloxacillin-induced liver injury. However, the association between abacavir-induced hypersensitivity and *B***5701* was established by candidate-gene association analysis, not by a GWA study (43).

Carbamazepine, a widely used anticonvulsant, causes skin rash in up to 10% of patients, and occasionally, this may progress to a hypersensitivity syndrome (50, 51) that can include rare blistering skin reactions such as Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) and also hepatitis (52). Using HLA as a candidate gene, a study in patients from Taiwan has shown a very strong association between *HLA-B***1502* and carbamazepine-induced SJS (53). However, in Europeans (54, 55) and Japanese (56), the allele frequency of *HLA-B***1502* is lower, and no association for development of Stevens-Johnson syndrome has been shown for this allele. A range of other drugs including other anticonvulsants such as phenytoin and lamotrigine, allopurinol and sulfonamide antimicrobial agents may also rarely give rise to hypersensitivity reactions involving skin and other tissues. In the case of allopurinol, a highly significant association between *HLA-B***5801* and both hypersensitivity and SJS/TEN reactions induced by this drug has been detected by candidate-gene studies in a range of ethnic groups (55, 57).

Recently, three GWA studies on drug-induced skin rash have added to the existing data obtained from candidate-gene association studies discussed above. The first involved 53 Japanese cases of carbamazepine-induced skin rash, including cases of Stevens-Johnson syndrome, toxic epidermal necrolysis, and drug-induced hypersensitivity syndrome (58). The strongest association was seen with a SNP in strong linkage disequilibrium with *HLA-A*3101*. Detailed HLA typing confirmed the association with *HLA-A*3101* and the expected lack of association with *HLA-B*1502* in this population. The *HLA-A*3101* association was also confirmed in a replication cohort of 61 cases (58); the allele is a risk factor for both SJS/TEN and hypersensitivity reactions, unlike the association between *HLA-B*1502* and carbamazepine toxicity in Chinese, which is more specific to SJS/TEN. In a second GWA study of patients of European ethnic origin, a similar *HLA-A*3101* association for carbamazepine hypersensitivity was detected, which again related to both hypersensitivity and SJS/TEN (59). The European GWA study involved 65 cases and was replicated in a further 145 cases. Both studies showed odds ratios in excess of 10 for hypersensitivity reactions and 25 for SJS/TEN. In addition, relatively high specificity and sensitivity values were estimated, suggesting that genotyping for *HLA-A*3101* prior to carbamazepine prescription would be cost effective in preventing hypersensitivity reactions (59), as already is the case for *HLA-B*1502* typing in Han Chinese and some other ethnicities (60).

These findings are also of considerable interest in terms of the biological basis for hypersensitivity reactions, indicating that more than one HLA class I gene product is able to present an antigen complex including carbamazepine or a metabolite to cytotoxic T cells, possibly owing to overlap in peptide-binding specificity (61).

The third GWA study on drug-induced skin injury was in a European population consisting of 96 cases of SJS or TEN induced by a variety of different drugs including the anticonvulsant lamotrigine and the antimicrobial cotrimoxazole (62). Only 3 cases related to carbamazepine exposure. For this heterogeneous group, no genome-wide significance was obtained for any individual SNP marker. When subgroup analysis was performed for lamotrigine and cotrimoxazole, a signal with a relatively low, though not genome-wide significant, p value was seen for lamotrigine in a SNP adjacent to the *ADAM22* gene, but no obvious possible associations were seen for cotrimoxazole. As *ADAM22* was suggested previously to be a susceptibility gene for epilepsy, the association in the lamotrigine cases may relate to an increased frequency in epilepsy cases compared with population controls rather than an actual drug-induced skin injury association. The failure to see any overall statistical significance in this GWA study suggests that genetic risk factors for drug-induced skin injury may be drug specific.

DRUG-INDUCED MYOPATHY

A number of different drugs are associated with myopathy, which usually involves subacute manifestation of myopathic symptoms such as muscle weakness, myalgia, creatine phosphokinase (CPK) elevation, or myoglobinuria. The precise disease phenotype is somewhat dependent on the individual drug (63). Most cases are not serious and are readily reversible by drug withdrawal, but a more severe form of disease resulting in rhabdomyolysis followed by death also occurs rarely.

Despite being very effective drugs that are used successfully worldwide, statins can cause muscle toxicity. As with other myotoxic drugs, this manifests as an asymptomatic rise in CPK but can be more serious on rare occasions. The mechanism by which statins give rise to toxicity is still not completely clear, but increasing evidence indicates an induction of expression of the protein atrogin-1 in affected muscle tissue leading to muscular atrophy, possibly because of inhibition of geranylgeranyl isoprene unit transfer by statins (64). Drug interactions seem to be an important contributor to statin-induced myopathy, but there is also increasing evidence for a role for genetic

polymorphisms relevant to their metabolism and transport in susceptibility to toxicity (65) and more limited evidence that genes encoding proteins relevant to muscular function may contribute (66, 67). Unlike the case with DILI, there is currently no evidence for a role for the immune system in susceptibility to myopathy induced by either statins or other drugs.

Understanding the genetic basis of susceptibility to simvastatin-induced myopathy was greatly increased by a GWA study of 85 cases of myopathy and 90 simvastatin-exposed controls without evidence of myopathy (68). The cases and controls were all of European ethnic origin. A highly significant association ($p = 4 \times 10^{-9}$) was seen for a single SNP in *SLCO1B1* with an odds ratio of 4.5 per copy of the variant allele. This effect was confirmed in 21 cases of myopathy from a separate replication cohort. *SLCO1B1* encodes an anionic drug transporter located on the sinusoidal face of the hepatocyte, which is the main inward transporter for a number of different statins (65). The significant SNP was in strong linkage disequilibrium with a nonsynonymous SNP in the *SLCO1B1*15* allele (also present in the rarer *SLCO1B1*5* allele) that is associated with higher plasma levels of statins owing to impaired transport (69). This association is, therefore, very biologically plausible. The significant polymorphism is common with a variant allele frequency of 0.13 in European populations, but possession of the variant allele explains only approximately 18% of attributable risk, with substantial numbers of myopathy cases homozygous wild type for *SLCO1B1*. The study had limited power to detect variants with a smaller effect than that of *SLCO1B1*, but no suggestion of significant effects for a list of candidate genes studied in more detail was obtained. There is a need for further larger studies with power to detect smaller effects to explain a higher proportion of risk for this toxicity. The association of muscle injury with *SLCO1B1*15* has recently been confirmed for milder toxicity and several different statins in a candidate-gene study (70). In addition, in a large study of individuals with type II diabetes who were receiving statins, carriage of the *SLCO1B1*15* allele was associated with a significantly increased risk of "statin intolerance," which was defined by either biochemical abnormalities and a change in prescription including discontinuation or a change in prescription alone (71).

DRUG-INDUCED LONG QT SYNDROME

As discussed in the Introduction (section above), cardiotoxicity is currently the most common reason for withdrawal of licensed drugs from the market. Examples of such drugs come from a variety of different classes including antipsychotics, antihistaminics, and antimicrobials. In susceptible individuals, these drugs are associated with delay of cardiac repolarization, which can be detected by prolongation of the QT interval on an electrocardiogram (ECG), and onset of a form of ventricular tachycardia called torsades de pointes (also detectable by ECG), which can lead to ventricular fibrillation and death (for a detailed review, see Reference 72). QT interval prolongation is an imperfect marker for the arrhythmic potential of a drug, given than many drugs prolong the QT interval but do not progress to arrhythmia, but it is currently the only available measure. QT prolongation can either be an inherited congenital disease or an acquired form that is triggered by exposure to environmental factors including certain drugs. There is considerable evidence to suggest that drugs associated with QT prolongation affect cardiac ion channels. In addition, the causative mutations associated with rare congenital long QT syndromes are often found in genes encoding ion channels (72). To date, GWA studies have focused on factors affecting QT length in populations, not drug-induced long QT (73–75). However, findings from these studies have resulted in the identification of SNPs in more than 10 different genes including the nitric oxide synthase 1 (NOS1) regulator *NOS1AP*, a range of sodium and potassium channel genes including *SCN5A* and *KCNJ2*, and miscellaneous other genes as important genetic predictors. As a result, these findings may contribute to an increased understanding of genetic

factors underlying susceptibility to drug-induced long QT syndrome. Unusually for a serious adverse drug reaction, there is evidence from family studies that susceptibility to drug-induced QT prolongation is, in part, genetically determined (76). Support for the relevance of GWA studies on population variation in QT length to drug-induced QT prolongation has been given by a finding for verapamil where increased QT prolongation was seen in patients positive for the variant in *NOS1AP* associated with longer QT interval in the general population (77).

The only GWA study on drug-induced QT prolongation thus far reported involved a phase III clinical trial of the antipsychotic drug iloperidone in which 183 patients had QT measurements performed 14 days after the start of drug treatment (78). No genome-wide significant signals were detected, but relatively low p values were obtained for several loci including the *CERKL* gene, which encodes a protein involved in the ceramide pathway, a regulator of currents conducted by potassium channels, and *SLCO3A1*, which may contribute to prostaglandin translocation. No trends toward significance with the SNPs in either ion channels or *NOS1AP* relevant to QT length in the previous studies were detected. There may be value in further GWA studies on patients who have suffered drug-induced QT prolongation from currently prescribed drugs to enable clearer identification of "at risk" genotypes.

BISPHOSPHONATE-INDUCED OSTEONECROSIS OF THE JAW

Bisphosphonates such as pamidronate and zoledronic acid are used widely in the treatment of cancer and osteoporosis to limit loss of bone mass. In some patients, their use is associated with necrotic damage to bone tissue in the jaw, which is often triggered by dental problems (79). In a GWA study involving 25 cases, four SNPs in the cytochrome P450 CYP2C8 gene showed altered frequencies compared with controls that narrowly escaped genome-wide significance (lowest p value was 6.22×10^{-6}) (80). One of the associated SNPs is located in the CYP2C8 promoter region and forms part of a haplotype seen at a frequency of 9.9% in Europeans. The other three associated SNPs are also found within this haplotype as well as in a second, slightly rarer, haplotype. There is some preliminary evidence that this haplotype is associated with lower than average CYP2C8 activity (81, 82). CYP2C8 plays a major role in the metabolism of a small number of drugs (83), but there is no evidence that bisphosphonates undergo metabolism by this or any other cytochrome P450. However, because a more general biological role for CYP2C8 in metabolism of inflammatory mediators (84) and a key role for inflammation in the bisphosphonate-induced osteonecrosis have been proposed (85), this finding is of interest but needs follow-up in a larger number of cases.

CONCLUDING REMARKS

GWA studies have facilitated progress in understanding genetic susceptibility to a range of serious adverse drug reactions. To date, most associations detected are strong and highly significant. Though this seems to be a particular feature of the GWA studies on adverse drug reactions reported so far, it also seems likely that only a proportion of genetic susceptibility is being explained by these strong associations and that additional studies, either larger GWA studies with better power to detect smaller effect sizes or whole-genome sequencing to detect rare genetic variants, will identify additional factors that contribute to risk. In particular, for examples such as flucloxacillin-induced liver injury or statin-induced myopathy, though strong associations have already been detected, the predictive value of the genotypes involved in these associations appears insufficient for their incorporation into routine decisions on prescribing. However, genotyping for the risk factors may be of value in diagnosing these adverse drug reactions. By contrast, the recent findings

of an association between *HLA-A*3101* and carbamazepine hypersensitivity may be translated more generally to the clinic (59), and it has also been proposed that a HLA-typing test could be incorporated into treatment with lumiracoxib (41).

The current findings from GWA studies should also be of value in the design of better model systems to detect idiosyncratic adverse drug reactions during drug development. For example, the potential importance of T cell responses in DILI is now much clearer, even though the underlying biology is still not understood, but multicellular systems involving both hepatocytes and immune cells may be helpful in identifying potentially hepatotoxic drugs. Similar approaches may also be applicable to other immune-related drug toxicities such as skin rash.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. World Health Organization. 2004. *Pharmacovigilance: Ensuring the Safe Use of Medicines*. Geneva: WHO
2. US Food and Drug Administration. 2011. *MedWatch: the FDA safety information and adverse event reporting program*. **http://www.fda.gov/Safety/MedWatch/default.htm**
3. Pirmohamed M, Breckenridge AM, Kitteringham NR, Park BK. 1998. Adverse drug reactions. *Br. Med. J.* 316:1295–98
4. Aronson JK, Ferner RE. 2005. Clarification of terminology in drug safety. *Drug Saf.* 28:851–70
5. Smith DA, Schmid EF. 2006. Drug withdrawals and the lessons within. *Curr. Opin. Drug Discov. Dev.* 9:38–46
6. Wilke RA, Lin DW, Roden DM, Watkins PB, Flockhart D, et al. 2007. Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. *Nat. Rev. Drug Discov.* 6:904–16
7. Becquemont L. 2009. Pharmacogenomics of adverse drug reactions: practical applications and perspectives. *Pharmacogenomics* 10:961–69
8. Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78
9. Hirschhorn JN. 2009. Genomewide association studies—illuminating biologic pathways. *N. Engl. J. Med.* 360:1699–701
10. International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
11. Sawcer SJ, Maranian M, Singlehurst S, Yeo T, Compston A, et al. 2004. Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Hum. Mol. Genet.* 13:1943–49
12. Manolio TA, Brooks LD, Collins FS. 2008. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Investig.* 118:1590–605
13. Hardy J, Singleton A. 2009. Genomewide association studies and human disease. *N. Engl. J. Med.* 360:1759–68
14. Bonnefond A, Froguel P, Vaxillaire M. 2010. The emerging genetics of type 2 diabetes. *Trends Mol. Med.* 16:407–16
15. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. 2010. Genetic susceptibility to breast cancer. *Mol. Oncol.* 4:174–91
16. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, et al. 2009. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* 5:e1000433
17. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, et al. 2009. A genome-wide association study of acenocoumarol maintenance dosage. *Hum. Mol. Genet.* 18:3758–68
18. Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, et al. 2009. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* 302:849–57

19. Hoofnagle JH. 2004. Drug-induced liver injury network (DILIN). *Hepatology* 40:773
20. Molokhia M, McKeigue P. 2006. EUDRAGENE: European collaboration to establish a case-control DNA collection for studying the genetic basis of adverse drug reactions. *Pharmacogenomics* 7:633–38
21. Lonjou C, Thomas L, Borot N, Ledger N, de Toma C, et al. 2006. A marker for Stevens-Johnson syndrome: ethnicity matters. *Pharmacogenomics J.* 6:265–68
22. Andrade RJ, Lucena MI, Fernandez MC, Pelaez G, Pachkoria K, et al. 2005. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. *Gastroenterology* 129:512–21
23. Bjornsson E. 2009. The natural history of drug-induced liver injury. *Semin. Liver Dis.* 29:357–63
24. Russmann S, Jetter A, Kullak-Ublick GA. 2010. Pharmacogenetics of drug-induced liver injury. *Hepatology* 52:748–61
25. Lee WM. 2003. Medical progress: drug-induced hepatotoxicity. *N. Engl. J. Med.* 349:474–85
26. Spriet-Pourra C, Auriche M. 1994. *Drug Withdrawal from Sales*. Richmond, VA: PJB Publ.
27. Assis DN, Navarro VJ. 2009. Human drug hepatotoxicity: a contemporary clinical perspective. *Expert Opin. Drug Metab. Toxicol.* 5:463–73
28. Park BK, Kitteringham NR, Maggs JL, Pirmohamed M, Williams DP. 2005. The role of metabolic activation in drug-induced hepatotoxicity. *Annu. Rev. Pharmacol. Toxicol.* 45:177–202
29. Stricker BH, Blok AP, Claas FH, Van Parys GE, Desmet VJ. 1988. Hepatic injury associated with the use of nitrofurans: a clinicopathological study of 52 reported cases. *Hepatology* 8:599–606
30. Hautekeete ML, Horsmans Y, van Waeyenberge C, Demanet C, Henrion J, et al. 1999. HLA association of amoxicillin-clavulanate-induced hepatitis. *Gastroenterology* 117:1181–86
31. O'Donohue J, Oien KA, Donaldson P, Underhill J, Clare M, et al. 2000. Co-amoxiclav jaundice: clinical and histological features and HLA class II association. *Gut* 47:717–20
32. Stricker BH, Van den Broek JW, Keuning J, Eberhardt W, Houben HG, et al. 1989. Cholestatic hepatitis due to antibacterial combination of amoxicillin and clavulanic acid (augmentin). *Dig. Dis. Sci.* 34:1576–80
33. Watanabe I, Tomita A, Shimizu M, Sugawara M, Yasumo H, et al. 2003. A study to survey susceptible genetic factors responsible for troglitazone-associated hepatotoxicity in Japanese patients with type 2 diabetes mellitus. *Clin. Pharmacol. Ther.* 73:435–55
34. Daly AK, Aithal GP, Leathart JB, Swainsbury RA, Dang TS, Day CP. 2007. Genetic susceptibility to diclofenac-induced hepatotoxicity: contribution of UGT2B7, CYP2C8, and ABCC2 genotypes. *Gastroenterology* 132:272–81
35. Choi JH, Ahn BM, Yi J, Lee JH, Nam SW, et al. 2007. MRP2 haplotypes confer differential susceptibility to toxic liver injury. *Pharmacogenet. Genomics* 17:403–15
36. Lucena MI, Andrade RJ, Martinez C, Ulzurrun E, Garcia-Martin E, et al. 2008. Glutathione S-transferase m1 and t1 null genotypes increase susceptibility to idiosyncratic drug-induced liver injury. *Hepatology* 48:588–96
37. Kindmark A, Jawaid A, Harbron CG, Barratt BJ, Bengtsson OF, et al. 2008. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J.* 8:186–95
38. Russmann S, Kaye JA, Jick SS, Jick H. 2005. Risk of cholestatic liver disease associated with flucloxacillin and flucloxacillin prescribing habits in the UK: cohort study using data from the UK General Practice Research Database. *Br. J. Clin. Pharmacol.* 60:76–82
39. Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, et al. 2009. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* 41:816–19
40. Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, et al. 2010. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat. Genet.* 42:711–14
41. Aithal GP, Daly AK. 2010. Preempting and preventing drug-induced liver injury. *Nat. Genet.* 42:650–51
42. Lucena MI, Molokhia M, Shen Y, Urban TJ, Aithal GP, et al. 2011. Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and class II alleles. *Gastroenterology* 141:338–47
43. Mallal S, Nolan D, Witt C, Masel G, Martin AM, et al. 2002. Association between presence of *HLA-B*5701*, *HLA-DR7*, and *HLA-DQ3* and hypersensitivity to HIV-1 reverse transcriptase inhibitor abacavir. *Lancet* 359:727–32

44. Spraggs CF, Budde LR, Briley LP, Bing N, Cox CJ, et al. 2011. HLA-DQA1*02:01 is a major risk factor for lapatinib-induced hepatotoxicity in women with advanced breast cancer. *J. Clin. Oncol.* 29:667–73
45. Bolton GC, Allen GD, Davies BE, Filer CW, Jeffery DJ. 1986. The disposition of clavulanic acid in man. *Xenobiotica* 16:853–63
46. Testa B, Kramer SD. 2007. The biochemistry of drug metabolism—an introduction: part 3. Reactions of hydrolysis and their enzymes. *Chem. Biodivers.* 4:2031–122
47. Nasirikenari M, Segal BH, Ostberg JR, Urbasic A, Lau JT. 2006. Altered granulopoietic profile and exaggerated acute neutrophilic inflammation in mice with targeted deficiency in the sialyltransferase ST6Gal I. *Blood* 108:3397–405
48. Daly AK. 2010. Drug-induced liver injury: past, present and future. *Pharmacogenomics* 11:607–11
49. Phillips EJ, Mallal SA. 2010. Pharmacogenetics of drug hypersensitivity. *Pharmacogenomics* 11:973–87
50. Vittorio CC, Muglia JJ. 1995. Anticonvulsant hypersensitivity syndrome. *Arch. Intern. Med.* 155:2285–90
51. Leeder JS. 1998. Mechanisms of idiosyncratic hypersensitivity reactions to antiepileptic drugs. *Epilepsia* 39(Suppl. 7):S8–16
52. Rzany B, Correia O, Kelly JP, Naldi L, Auquier A, Stern R. 1999. Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis during first weeks of antiepileptic therapy: a case-control study. Study Group of the International Case Control Study on Severe Cutaneous Adverse Reactions. *Lancet* 353:2190–94
53. Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, et al. 2004. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* 428:486
54. Alfirevic A, Mills T, Harrington P, Pinel T, Sherwood J, et al. 2005. Association between serious carbamazepine hypersensitivity reactions and the HSP70 gene cluster. *Toxicology* 213:264–65
55. Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, et al. 2008. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenet. Genomics* 18:99–107
56. Kaniwa N, Saito Y, Aihara M, Matsunaga K, Tohkin M, et al. 2008. HLA-B locus in Japanese patients with anti-epileptics and allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis. *Pharmacogenomics* 9:1617–22
57. Hung SI, Chung WH, Liou LB, Chu CC, Lin M, et al. 2005. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc. Natl. Acad. Sci. USA* 102:4134–39
58. Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, et al. 2011. Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. *Hum. Mol. Genet.* 20:1034–41
59. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, et al. 2011. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *N. Engl. J. Med.* 364:1134–43
60. Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, et al. 2011. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. *N. Engl. J. Med.* 364:1126–33
61. Chelvanayagam G. 1996. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. *Immunogenetics* 45:15–26
62. Shen Y, Nicoletti P, Floratos A, Pirmohamed M, Molokhia M, et al. 2011. Genome-wide association study of serious blistering skin rash caused by drugs. *Pharmacogenomics J.* In press, doi:10.1038/tpj.2010.84
63. Dalakas MC. 2009. Toxic and drug-induced myopathies. *J. Neurol. Neurosurg. Psychiatry* 80:832–38
64. Cao P, Hanai JI, Tanksale P, Imamura S, Sukhatme VP, Lecker SH. 2009. Statin-induced muscle damage and atrogin-1 induction is the result of a geranylgeranylation defect. *FASEB J.* 23:2844–54
65. Niemi M, Pasanen MK, Neuvonen PJ. 2011. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol. Rev.* 63:157–81
66. Vladutiu GD, Simmons Z, Isackson PJ, Tarnopolsky M, Peltier WL, et al. 2006. Genetic risk factors associated with lipid-lowering drug-induced myopathies. *Muscle Nerve* 34:153–62
67. Oh J, Ban MR, Miskie BA, Pollex RL, Hegele RA. 2007. Genetic determinants of statin intolerance. *Lipids Health Dis.* 6:7
68. Link E, Parish S, Armitage J, Bowman L, Heath S, et al. 2008. SLCO1B1 variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* 359:789–99
69. Niemi M. 2010. Transporter pharmacogenetics and statin toxicity. *Clin. Pharmacol. Ther.* 87:130–33

70. Voora D, Shah SH, Spasojevic I, Ali S, Reed CR, et al. 2009. The SLCO1B1*5 genetic variant is associated with statin-induced side effects. *J. Am. Coll. Cardiol.* 54:1609–16
71. Donnelly LA, Doney AS, Tavendale R, Lang CC, Pearson ER, et al. 2011. Common nonsynonymous substitutions in SLCO1B1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: a go-DARTS study. *Clin. Pharmacol. Ther.* 89:210–16
72. Kannankeril P, Roden DM, Darbar D. 2010. Drug-induced long QT syndrome. *Pharmacol. Rev.* 62:760–81
73. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, et al. 2006. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat. Genet.* 38:644–51
74. Newton-Cheh C, Eijgelsheim M, Rice KM, de Bakker PI, Yin X, et al. 2009. Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* 41:399–406
75. Nolte IM, Wallace C, Newhouse SJ, Waggott D, Fu J, et al. 2009. Common genetic variation near the phospholamban gene is associated with cardiac repolarisation: meta-analysis of three genome-wide association studies. *PLoS ONE* 4:e6138
76. Kannankeril PJ, Roden DM, Norris KJ, Whalen SP, George AL Jr, Murray KT. 2005. Genetic susceptibility to acquired long QT syndrome: pharmacologic challenge in first-degree relatives. *Heart Rhythm* 2:134–40
77. van Noord C, Aarnoudse AJ, Eijgelsheim M, Sturkenboom MC, Straus SM, et al. 2009. Calcium channel blockers, NOS1AP, and heart-rate-corrected QT prolongation. *Pharmacogenet. Genomics* 19:260–66
78. Volpi S, Heaton C, Mack K, Hamilton JB, Lannan R, et al. 2009. Whole-genome association study identifies polymorphisms associated with QT prolongation during iloperidone treatment of schizophrenia. *Mol. Psychiatry* 14:1024–31
79. Khan AA, Sandor GK, Dore E, Morrison AD, Alsahli M, et al. 2009. Bisphosphonate associated osteonecrosis of the jaw. *J. Rheumatol.* 36:478–90
80. Sarasquete ME, Garcia-Sanz R, Marin L, Alcoceba M, Chillon MC, et al. 2008. Bisphosphonate-related osteonecrosis of the jaw is associated with polymorphisms of the cytochrome P450 CYP2C8 in multiple myeloma: a genome-wide single nucleotide polymorphism analysis. *Blood* 112:2709–12
81. Bahadur N, Leathart JBS, Mutch E, Steimel-Crespi D, Dunn SA, et al. 2002. CYP2C8 polymorphisms in Caucasians and their relationship with paclitaxel 6 alpha-hydroxylase activity in human liver microsomes. *Biochem. Pharmacol.* 64:1579–89
82. Rodriguez-Antona C, Niemi M, Backman JT, Kajosaari LI, Neuvonen PJ, et al. 2008. Characterization of novel CYP2C8 haplotypes and their contribution to paclitaxel and repaglinide metabolism. *Pharmacogenomics J.* 8:268–77
83. Totah RA, Rettie AE. 2005. Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance. *Clin. Pharmacol. Ther.* 77:341–52
84. Fleming I. 2004. Cytochrome P450 epoxygenases as EDHF synthase(s). *Pharmacol. Res.* 49:525–33
85. Lesclous P, Abi Najm S, Carrel JP, Baroukh B, Lombardi T, et al. 2009. Bisphosphonate-associated osteonecrosis of the jaw: a key role of inflammation? *Bone* 45:843–52

Annual Review of Pharmacology and Toxicology

Volume 52, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml

Review in Advance first posted online on October 6, 2014. (Changes may still occur before final publication online and in print.)

Predicting Toxicities of Reactive Metabolite–Positive Drug Candidates

Amit S. Kalgutkar[1] and Deepak Dalvie[2]

Pharmacokinetics, Dynamics and Metabolism Department, Pfizer Worldwide Research and Development, [1]Cambridge, Massachusetts 02139 and [2]San Diego, California 92121; email: amit.kalgutkar@pfizer.com, deepak.dalvie@pfizer.com

Annu. Rev. Pharmacol. Toxicol. 2015. 55:12.1–12.20

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-010814-124720

Keywords

adverse drug reactions, ADRs, bioactivation, covalent binding, cytochrome P450, electrophile, glutathione, hepatotoxicity, idiosyncratic, liver microsomes, structural alert, precision medicine, prediction in pharmacology

Abstract

Because of the inability to predict and quantify the risk of idiosyncratic adverse drug reactions (IADRs) and because reactive metabolites (RMs) are thought to be responsible for the pathogenesis of some IADRs, the potential for RM formation within new chemical entities is routinely examined with the ultimate goal of eliminating or reducing the liability through iterative design. Likewise, avoidance of structural alerts is almost a standard practice in drug design. However, the perceived safety concerns associated with the use of structural alerts and/or RM screening tools as standalone predictors of toxicity risks may be overexaggerated. Numerous marketed drugs form RMs but do not cause idiosyncratic toxicity. In this review article, we present a critique of the structural alert/RM concept as applied in drug discovery and evaluate the evidence linking structural alerts and RMs to observed toxic effects. Pragmatic risk mitigation strategies to aid the advancement of drug candidates that carry a RM liability are also discussed.

INTRODUCTION

The formation of electrophilic reactive metabolites (RMs) is considered to be an undesirable feature in drug candidates. This notion arises from evidence linking RM liability with mechanism-based inactivation of cytochrome P450 (CYP) isoforms, which can result in clinical drug-drug interactions (DDIs) (1) and/or covalent modification of DNA that results in mutagenicity (2). Furthermore, it is now widely accepted that RMs, as opposed to the parent molecules from which they are derived, can also be responsible for the etiology of some idiosyncratic adverse drug reactions (IADRs) (3–6). The term idiosyncratic simply implies that little is known about the underlying mechanism and that the toxicity is unpredictable. IADRs can manifest in drug-treated patients as rare and sometimes life-threatening reactions that cannot be explained by the primary pharmacology of the drug. For instance, felbamate is used to treat convulsions but can cause aplastic anemia and hepatotoxicity. Many IADRs are immune mediated and occur in very low frequency in a small subset of patients either acutely or as a delayed response. The observations that certain IADRs (e.g., hypersensitivity associated with the antiretroviral agent abacavir, hepatotoxicity associated with the nonsteroidal anti-inflammatory drug lumiracoxib and the antiretroviral drug nevirapine) are linked to specific human leukocyte antigen (HLA) genes (7–9) provide compelling evidence for the immune-mediated nature of these toxicities. The precise mechanisms of IADRs remain unclear; however, the vast majority may be caused by immunogenic conjugates formed via the covalent interaction of a RM with cellular proteins, resulting in cellular dysfunction or an immune response via the formation of a hapten (10). The link between RM formation and drug toxicity first became evident from studies on the hepatotoxic anti-inflammatory agent acetaminophen. Mechanistic studies, which have served as a gold standard for drug toxicity assessment over the decades (11), established the CYP-mediated oxidation of acetaminophen to a reactive quinone-imine species that could deplete levels of the endogenous antioxidant glutathione (GSH) and/or bind covalently to liver biomacromolecules, leading to hepatotoxicity. Idiosyncratic toxicities are, by definition, difficult to reproduce in the human population, and there are few, if any, generally applicable animal models for examining them (12). Consequently, reliably predicting the occurrence of IADRs with new drug candidates represents a significant challenge in preclinical drug discovery and development. Under the basic premise that a molecule devoid of RM formation could mitigate idiosyncratic toxicity risks, in vitro screens [e.g., reactive metabolite trapping with GSH and/or protein covalent binding in NADPH-supplemented human liver microsomes (HLM)] have been implemented to assess CYP-mediated RM formation for new molecular entities with the ultimate goal of minimizing or eliminating this liability through iterative medicinal chemistry (13, 14).

THE STRUCTURAL ALERT/REACTIVE METABOLITE CONCEPT IN DRUG DESIGN: WHAT HAS RETROSPECTIVE STRUCTURE TOXICITY ANALYSIS TAUGHT US?

Because the link between RM formation and idiosyncratic toxicity is not well understood, one tactic frequently adopted in drug discovery is that of avoidance. The term avoidance refers to a philosophical argument wherein certain functional groups (known as structural alerts or toxicophores) must be excluded from drug design, irrespective of whether these substituents would offer pharmacologic (e.g., improved intrinsic potency), pharmacokinetic (e.g., low plasma clearance), and/or biopharmaceutical (e.g., improved aqueous solubility) advantages. The concept of structural alerts (extensively reviewed in Reference 15) originated from studies that characterized the mechanism of RM formation within numerous drugs associated with idiosyncratic toxicity. An analysis of 68 drugs recalled or associated with a black box warning (BBW) for

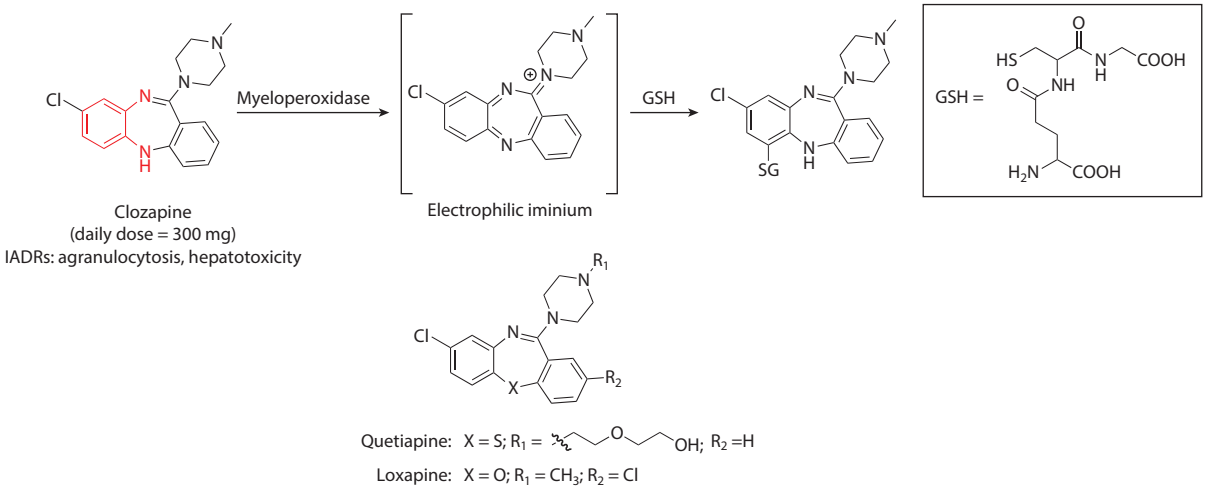


Figure 1

Structure-toxicity relationships for the dibenzodiazepine derivatives: clozapine versus loxapine and quetiapine. Abbreviations: GSH, glutathione; IADR, idiosyncratic adverse drug reaction.

idiosyncratic toxicity indicated that 55 (80.8%) contained one or more structural alerts, and evidence for RM formation [characterization of adducts with biological nucleophiles such as GSH and/or covalent binding to target organ tissue (e.g., liver microsomes)] was provided for 36 out of the 55 drugs (65%) (16). A prominent structural alert in the analysis was the aniline/anilide motif, which was present in ~30 out of the 68 (44%) toxic drugs. Among all known structural alerts, the aniline/anilide motif is perhaps most notorious for its association with mutagenicity, direct organ toxicity, methemoglobinemia, and immunogenic allergenic toxicity (17).

A compelling argument for chemotype-specific toxicity is also evident from structure-activity relationship (SAR) studies, wherein the absence of RM liability is consistent with the improved safety profile of successor drugs. For instance, whereas clozapine use is limited by a high incidence of agranulocytosis and hepatotoxicity, quetiapine and loxapine are not associated with these adverse events. Clozapine exhibits covalent binding to human neutrophils in vitro via the myeloperoxidase-catalyzed oxidation of the dibenzodiazepine ring to a reactive iminium ion, which covalently binds to target tissue and GSH (**Figure 1**) (18, 19). Proteins covalently modified with clozapine have been detected in neutrophils of patients being treated with clozapine; this finding reaffirms the relevance of the in vitro studies (19). In the cases of quetiapine and loxapine, the bridging nitrogen atom is replaced with a sulfur or oxygen atom (**Figure 1**); consequently, these drugs cannot form a reactive iminium species (20). Although anecdotal for the most part, the structure-toxicity relationships suggest that avoiding structural alerts in drug design would lead to therapeutic agents that do not cause IADRs. In fact, knowledge-based systems such as Derek for Windows that are used to predict the toxicity of a chemical from its structure have evolved from such findings. Predictions from knowledge-based systems, however, can be misleading at times. For example, 2-aminopyridine and 2-aminopyrimidine are not predicted to be structural alerts in Derek and are commonly utilized in drug discovery as aniline replacements. Certainly, the 2-aminopyrimidine scaffold found in the anxiolytic agent buspirone is not metabolized to a RM, unlike the aniline derivative and RM-positive hepatotoxin nefazodone (**Figure 2**) (21).

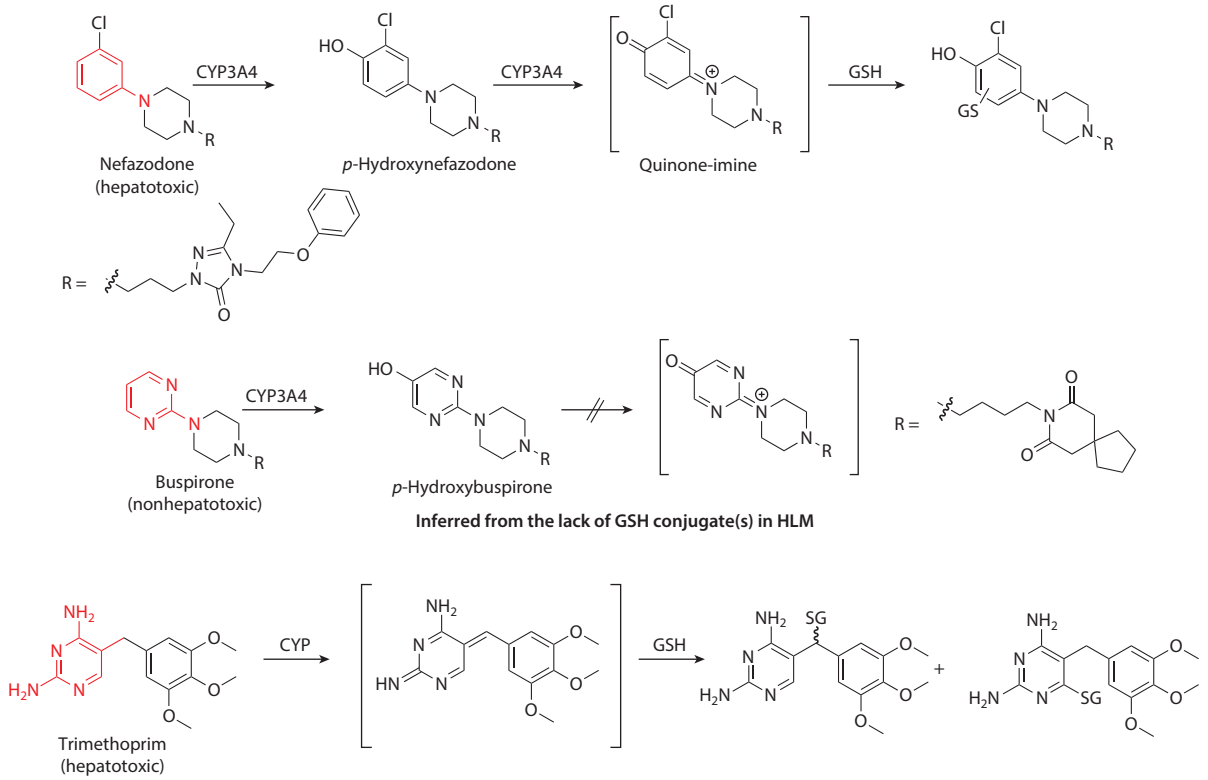


Figure 2

RM formation within aniline and aminopyrimidine structural alerts. The aniline (nefazodone) and aminopyrimidine (buspirone and trimethoprim) alerts are highlighted in red. Abbreviations: CYP, cytochrome P450; GSH, glutathione; HLM, human liver microsomes; RM, reactive metabolite.

However, an exception to the rule is the two-electron oxidation of the 2-aminopyridine group in trimethoprim to an electrophilic imine-methide species, which is a causative factor in the idiosyncratic toxicity associated with this antibacterial drug (**Figure 2**) (22). As novel (and proprietary) functional groups are continuously sought in drug design, unanticipated bioactivated pathways leading to RMs may emerge and thus expand the existing knowledge on structural alerts.

As such, the application of the structural alert concept in drug design has several shortcomings. First, not all compounds possessing structural alerts are metabolized to RMs. The likelihood of RM formation depends on the binding pose of the compound in the catalytic site of the drug-metabolizing enzyme (e.g., CYP) and the subsequent positioning of the structural alert toward metabolism to a RM. Metabolism may occur at a site other than the structural alert and lead to nonreactive products. For example, both sudoxicam and meloxicam (**Figure 3**) contain the 2-aminothiazole structural alert, but only sudoxicam forms the reactive acylthiourea, which appears to be responsible for its hepatotoxicity (23). Although the introduction of a methyl group at the C-5 position on the thiazole ring in meloxicam is the only structural difference, the change dramatically alters the metabolic profile. Oxidation of the C-5 methyl group to the alcohol (and carboxylic acid) metabolites constitutes the major metabolic fate of meloxicam in humans (24). Additionally, alternative routes of drug clearance could also influence the metabolism of structural alerts, as in the cases of ranitidine and pramipexole (**Figure 3**). Both drugs are eliminated by urinary

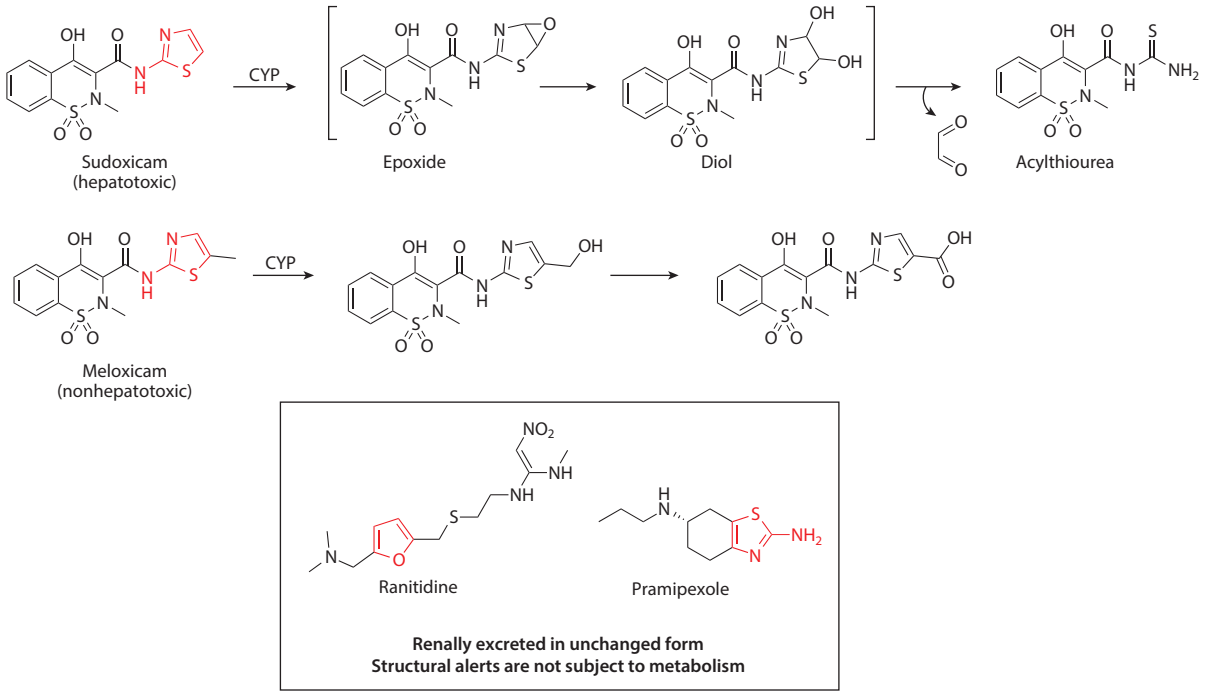


Figure 3

Illustrations of drugs that contain structural alerts but do not form RMs. Structural alerts are highlighted in red. Abbreviations: CYP, cytochrome P450; RM, reactive metabolite.

excretion primarily in unchanged parent form (25, 26), and no RM formation is seen on the furan and 2-aminothiazole structural alerts. Notably, ranitidine and pramipexole are marketed as agents for the treatment of peptic ulcers and Parkinson's disease and generally do not cause IADRs.

Second, because the categorization of structural alerts is knowledge based, avoiding as-yet-unknown structures that can form RMs is not possible. For example, the mechanism for RM formation within the anticonvulsant felbamate, which is associated with aplastic anemia and hepatotoxicity, involves the formation of the electrophilic α,β-unsaturated 2-phenylpropenal via an uncharacteristic multistep process (**Figure 4**) (27). Evidence for the occurrence of this pathway in vivo has arisen from the characterization of urinary mercapturic acid conjugates following felbamate administration to humans (28). As seen in **Figure 4**, felbamate is devoid of prototypical structural alerts.

Third, and more importantly, structural alerts fall into one of two categories: ones that form RMs versus all others. There is no clear distinction as to when a particular functional group is viewed as a structural alert. The vast majority of marketed drugs possess a phenyl ring, which is a structural alert because its biotransformation to the corresponding phenol metabolite proceeds through a reactive epoxide intermediate, which can be trapped with GSH in some cases (15, 16). Removing simple phenyl rings from the repertoire of substituents in drug design is practically impossible, and mankind would be deprived of countless useful therapies if phenyl-containing drugs had not been developed because of the phenyl ring's status as a structural alert.

Finally, the simplistic notion that the absence of a structural alert and/or RM liability in a drug candidate serves as a guarantee of its safety is not necessarily true. There is no evidence

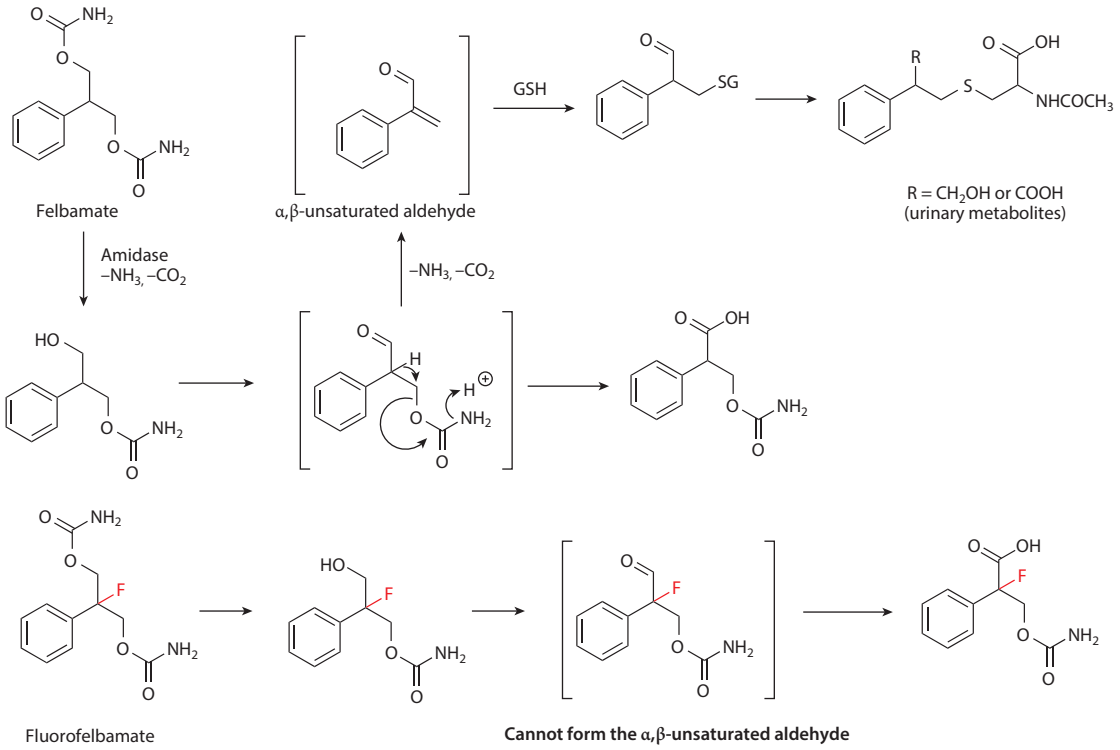


Figure 4

Mechanism of RM formation within the anticonvulsant felbamate, which led to the design of fluorofelbamate. Abbreviations: GSH, glutathione; RM, reactive metabolite.

that the idiosyncratic hepatotoxicity associated with the recalled thrombin inhibitor ximelagatran (**Figure 5**) is associated with RM formation, and the drug does not exhibit any alerts in its chemical structure (29). Likewise, there are no structural alerts or evidence for RM formation within drugs such as chlormezanone, isoxicam, pemoline, and flecainide (**Figure 5**), which have been withdrawn owing to idiosyncratic toxicity (16). The widely prescribed antihyperlipidemic agent niacin (**Figure 5**) does not contain conventional structural alerts but possesses the highest potential for hepatotoxicity when administered in the sustained-release form. The hepatotoxic effects of niacin are related to a high-affinity, low-capacity metabolic pathway that affords nicotinamide and *N*-methyl-2- and *N*-methyl-4-pyridone-5-carboxamide metabolites; thus, the sustained-release formulation can lead to higher levels of toxic metabolites (30). The alternative competing metabolic pathway is a low-affinity, high-capacity conjugation pathway (involving the formation of a glycine amide metabolite) that leads to prostaglandin-mediated vasodilation and subsequent cutaneous flushing (31). The immediate-release formulation overwhelms the higher-affinity oxidation pathway, and the majority of the niacin dose is metabolized via the high-capacity glycine conjugation pathway, leading to a much lower rate of hepatotoxicity (30).

Although structural alerts must be used with caution, particularly at the lead optimization/candidate selection stage in drug discovery, it is imperative to demonstrate experimentally whether structural alerts, if present in a molecule of interest, actually are prone to RM formation. In the case of RM-positive drugs, identifying the biochemical mechanism and the enzymes responsible

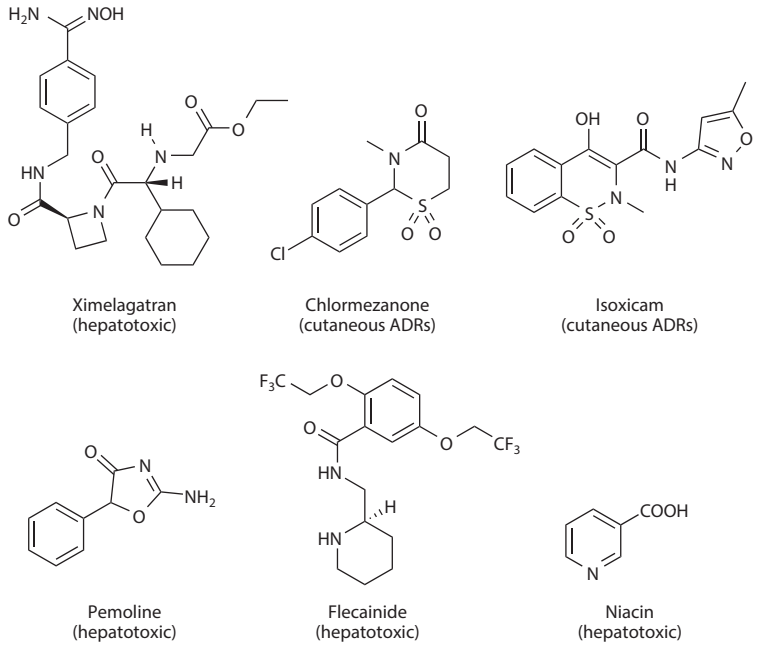


Figure 5

Illustrations of drugs associated with IADRs that do not contain structural alerts and do not form RMs. Abbreviations: ADR, adverse drug reaction; IADR, idiosyncratic adverse drug reaction; RM, reactive metabolite.

for RM formation is necessary. The information can then be used, as appropriate, to modify the structure of the RM-positive drugs in order to eliminate the liability. For instance, fluorofelbamate was specifically designed to eliminate the RM liability of felbamate on the basis of the bioactivation mechanism depicted in **Figure 4**. The strategic placement of the fluorine atom on the benzylic position prevents the β-elimination process that affords the α,β-unsaturated aldehyde (32). Successful case studies involving metabolism-guided design to circumvent RM formation in drug discovery are abundant in the medicinal chemistry/chemical toxicology literature (33–36). For instance, in the course of efforts leading to the discovery of taranabant, a selective and potent inhibitor of the cannabinoid-1 receptor and a Phase III clinical candidate for the treatment of obesity, the lead compound **1** depicted in **Figure 6** revealed a high level of covalent binding to HLM in a NADPH-dependent fashion, consistent with RM formation. Elucidation of the structure of the GSH conjugate suggested that the RM was an arene oxide intermediate derived from epoxidation of the electron-rich phenoxy ring (37). Replacement of the phenoxy ring with the trifluoromethylpyridyl ring afforded taranabant, which was devoid of RM formation yet retained potency and selectivity against the cannabinoid-1 receptor.

Another example pertains to the discovery of the first glucokinase activator, piragliatin (**Figure 6**), which has shown efficacy (e.g., lowering of pre- and postprandial glucose levels, improvements in insulin secretory profile) in Phase II clinical trials in patients with type 2 diabetes (38). The prototype candidate RO0281675 (**Figure 6**) was withdrawn from Phase I clinical trials owing to its narrow safety margin in preclinical toxicology studies. In chronic toxicology studies in rats and dogs, RO0281675 caused reversible hepatic lipidosis, which was believed to occur via the metabolism of the 2-aminothiazole motif to a thiourea metabolite. The hypothesis

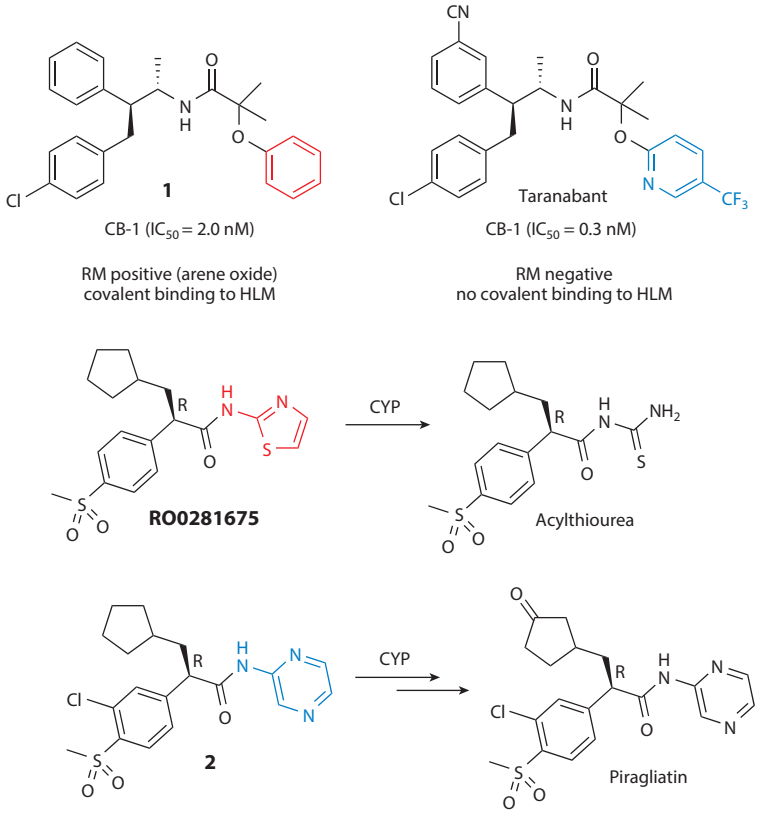


Figure 6

Medicinal chemistry tactics to eliminate RM liability on the basis of established pathways of bioactivation. Structural alerts are highlighted in red, and functional groups that are not structural alerts are highlighted in blue. Abbreviations: CB-1, cannabinoid-1; CYP, cytochrome P450; HLM, human liver microsome; RM, reactive metabolite.

was further substantiated on the basis of two observations: (*a*) the thiourea derivative was formed as a metabolite upon incubating RO0281675 in NADPH-supplemented liver microsomes from preclinical species and humans, and (*b*) five-day toxicity studies in rats with an authentic standard of the thiourea metabolite led to hepatic lipidosis in a manner similar to that noted with RO0281675. Subsequent SAR studies seeking thiazole ring replacements led to the identification of a pyrazine-based lead analog, labeled compound **2** in **Figure 6**. In vitro metabolite identification revealed several oxidative metabolites on the cyclopentyl ring of compound **2**, which were synthesized and shown to possess pharmacological activity comparable with that of the compound itself. Additional profiling of in vitro and in vivo safety and efficacy of the oxidative metabolites led to the selection of piragliatin as the clinical candidate. Subchronic and chronic toxicology studies with piragliatin in rats and dogs revealed no evidence of hepatic lipidosis. Furthermore, piragliatin is relatively less lipophilic than compound **2** (clog P of compound **2** = 2.69 versus clog P of piragliatin = 0.47) and exhibits superior oral absorption (lower plasma clearance leading to increased oral absorption) in preclinical species and humans. In practice, however, the exercise of eliminating or reducing RM formation is not trivial; medicinal chemistry tactics to eliminate RM

formation could confer a detrimental effect on pharmacology (e.g., changes in agonist/antagonist behavior, subtype selectivity for target receptor or enzyme) and/or pharmacokinetic attributes. In our experience, if the structural alert can be readily replaced with an alternative moiety without significant loss of desired pharmacology/pharmacokinetic attributes, then doing so is advisable. As a result, the need for additional risk assessment beyond the standard drug safety package as well as further internal debate on this topic for the remainder of the development program could be avoided.

Although the strategy of circumventing IADR risks via removing RM liability represents a pragmatic starting point in drug design, there is growing concern that the perceived safety hazards associated with structural alerts and RM-positive compounds may be overexaggerated. Several blockbuster drugs contain structural alerts and form RMs but do not cause idiosyncratic toxicity, suggesting that the structural alert concept and RM screening tools in drug discovery may be too stringent and thus could halt the advancement of novel medicines. A survey of 108 structurally distinct drugs that were among the most prescribed in 2009 revealed that 58 (53%) contained structural alerts, and evidence for RM formation has been provided in 24 out of the 58 (41%) cases (16). Likewise, 13 out of the 15 small-molecule drugs, which constitute the top drugs based on total sales in 2009, possess structural alerts. In vitro and/or in vivo experimental evidence for RM formation has been presented for 10 out of the 13 drugs (16). Overall, the analysis indicates that the percentage of structural alert–positive and/or RM-positive drugs in the most-prescribed or total-sales drug category is largely similar to that noted for drugs recalled or associated with a BBW. The alerts are fairly diverse in nature and include aniline/anilide, thiophene, olefin, and quinone precursors found in the toxic drugs. In the case of the top-ranked drug (on the basis of dispensed prescriptions and sales) atorvastatin (Lipitor®) (**Figure 7**), CYP3A4-catalyzed monohydroxylation on the acetanilide structural alert leads to the formation of the active *ortho*- and *para*-hydroxyacetanilide metabolites (39), which can be oxidized to reactive quinone-imine species in a manner similar to that noted for acetaminophen. The observation that atorvastatin covalently binds to HLM in a NADPH-dependent fashion partially validates the hypothesis (40). Interestingly, atorvastatin was ranked number one in terms of dispensed prescriptions and total sales for 2009.

In some cases, RM formation is essential to the pharmacological activity of a drug. In the case of the blockbuster cardiovascular drug and $P2Y_{12}$ purinoreceptor antagonist clopidogrel, the thiophene structural alert is metabolized by a CYP enzyme or enzymes to a pharmacologically active RM (speculated to be an electrophilic sulfenic acid). This RM forms a covalent disulfide linkage with a cysteinyl residue on the $P2Y_{12}$ receptor in platelets, leading to inhibition of platelet aggregation (**Figure 7**) (41, 42). Similar to clopidogrel, the antidepressant duloxetine contains a pendant thiophene ring, which can be oxidized by CYP to RMs. Indeed, incubation of duloxetine in NADPH- and GSH-supplemented HLM indicated the presence of several GSH conjugates (43). Interestingly, structural characterization of these conjugates reveals that GSH adduction occurs on the naphthalene ring rather than on the thiophene ring and likely proceeds via a reactive epoxide intermediate (**Figure 7**).

RMs can be the bases for DDIs. In humans, the antidepressant paroxetine is metabolized by CYP2D6 on the 1,3-benzodioxole structural alert to a catechol intermediate (44). The process also leads to the mechanism-based inactivation of the CYP isozyme and DDIs with CYP2D6 substrates in the clinic (45). In vitro studies with [^{3}H]-paroxetine have demonstrated the NADPH-dependent covalent binding to human liver microsomal and S9 proteins and have also demonstrated the characterization of GSH conjugates of reactive quinone metabolites (**Figure 8**) (46). The selective estrogen receptor modulator raloxifene is metabolized by CYP3A4 on the phenolic structural alerts to yield reactive quinone species that can be trapped with GSH (47). The process is also

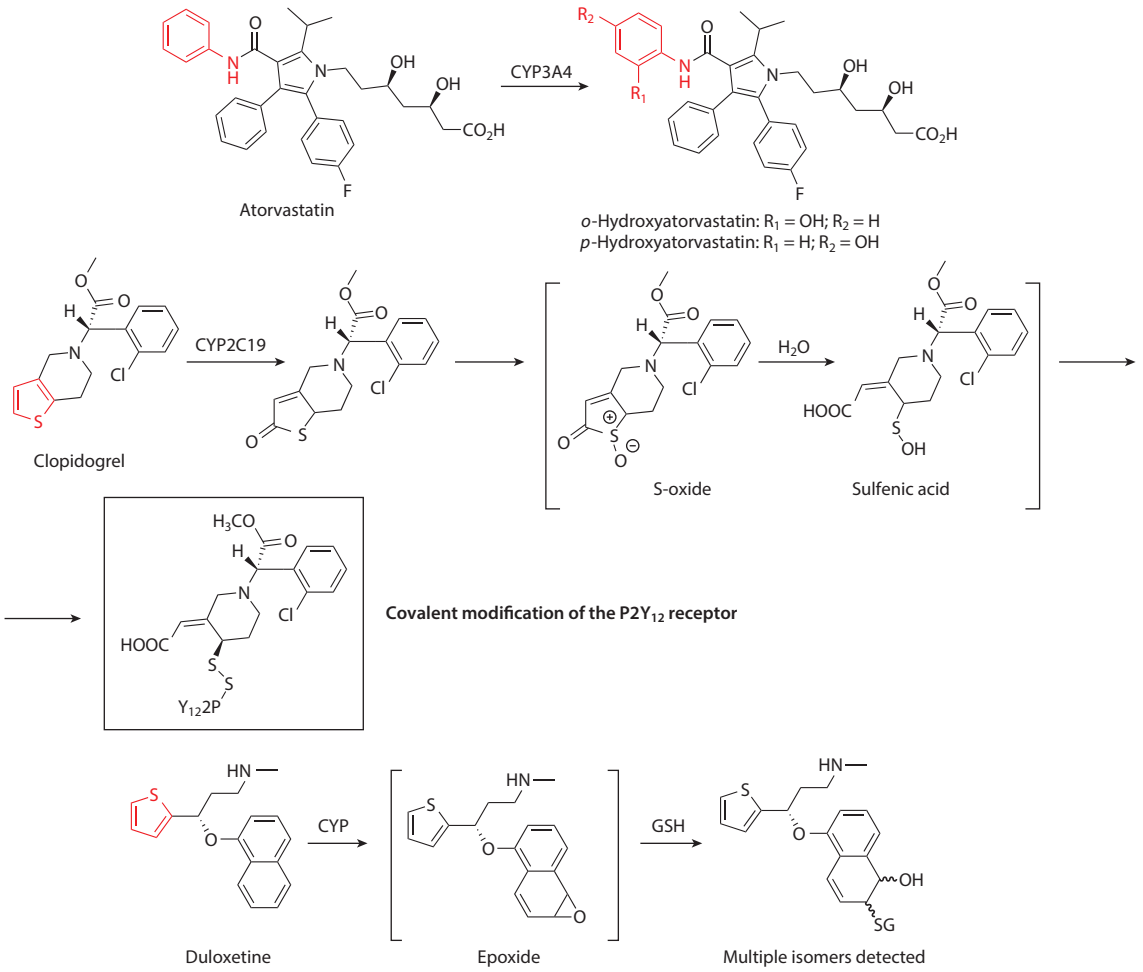


Figure 7

Examples of commercial blockbuster drugs that form RMs. Structural alerts are highlighted in red. Abbreviations: CYP, cytochrome P450; GSH, glutathione; RM, reactive metabolite.

accompanied by the mechanism-based inactivation of CYP3A4. Cyclobenzaprine is a skeletal muscle relaxant, which is metabolized on the olefin structural alert to yield the corresponding dihydrodiol metabolite in significant quantities in human urine (**Figure 8**) (48). The formation of the dihydrodiol metabolite is consistent with olefin epoxidation as a rate-limiting step.

In addition to the above analysis, examination of the structural trends for recently approved drugs (2009–present) revealed the presence of structural alerts in several cases (**Figure 9**). Foremost among these is the thiophene structural alert in the sodium glucose cotransporter 2 inhibitor and antidiabetic agent canagliflozin. However, its principal elimination mechanism in humans involves glucuronidation (on the sugar moiety) by UGT (uridine diphosphate glucose glucuronosyltransferase) enzymes (49). The lack of GSH conjugate formation in canagliflozin incubations in HLM suggests that the thiophene ring is latent to CYP metabolism (A.S. Kalgutkar, unpublished observations). The selective direct factor Xa inhibitors apixaban and rivaroxaban are new oral anticoagulants that contain structural alerts (*para*-methoxyaniline and *bis*-anilide motifs in

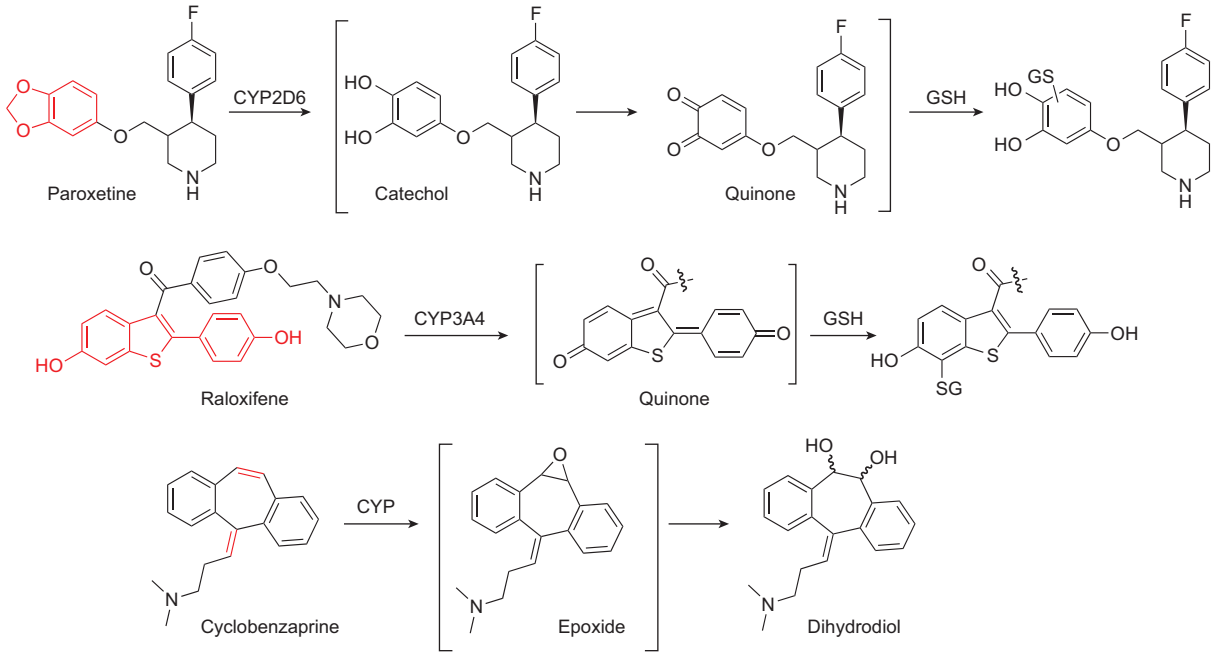


Figure 8

Additional examples of successfully marketed drugs that form RMs. Structural alerts are highlighted in red. Abbreviations: CYP, cytochrome P450; GSH, glutathione; RM, reactive metabolite.

apixaban; chlorothiophene and *bis*-anilide motifs in rivaroxaban). Human mass balance studies using [^{14}C]-apixaban and rivaroxaban indicate that the alerts in those compounds are not subject to metabolism and/or RM formation (50, 51). In the case of rivaroxaban, the pendant chlorothiophene motif is essential for pharmacology and cannot be replaced. The aniline structural alert is also present in the oral direct thrombin inhibitor dabigatran. However, dabigatran is not subject to oxidative metabolism by CYP enzymes in humans (52).

The remainder of drugs flagged for structural alert presence (depicted in **Figure 9**) have been approved for various oncology indications. Enzalutamide is an androgen receptor antagonist used in the treatment of metastatic castration-resistant prostate cancer. Radiolabeled mass balance studies in humans with [^{14}C]-enzalutamide revealed *N*-dealkylation and amide bond hydrolysis as the principal routes of metabolism (53). The thiourea structural alert in enzalutamide is not subject to metabolism. Pomalidomide is a thalidomide derivative and has been approved for the treatment of relapsed and refractory multiple myeloma. In humans, a significant proportion of the metabolism occurs on the aniline structural alert (via the catalytic action of CYP1A2 and CYP3A4) to yield the corresponding *ortho*- and *para*-hydroxyaniline derivatives as stable metabolites (54). However, GSH conjugates of the quinone-imine species (the two-electron oxidation product of the hydroxyaniline metabolites of pomalidomide) have not been observed in HLM incubations (55). The *ortho*-hydroxyaniline metabolite of pomalidomide is subject to glucuronidation and is one of the major metabolites in human excreta.

Bosutinib, cabozantinib, pazopanib, ponatinib, crizotinib, and vandetanib are tyrosine kinase inhibitors. In humans, bosutinib is metabolized primarily by CYP3A4 to yield *N*-desmethyl bosutinib and oxydechlorinated bosutinib as major circulating metabolites. The oxydechlorinated

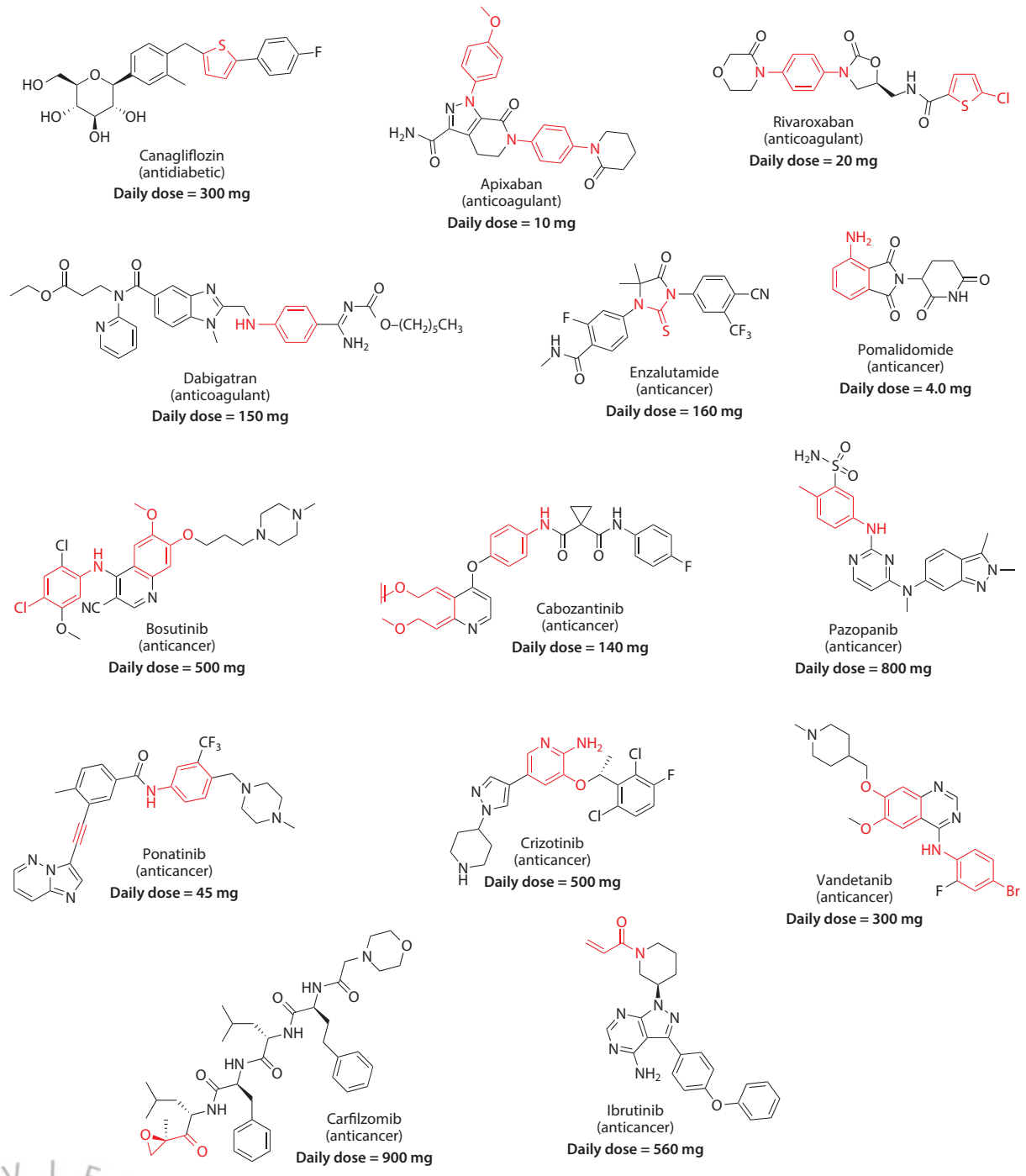


Figure 9

Recently approved drugs (2009–present) that contain structural alerts. Structural alerts are highlighted in red.

bosutinib metabolite is an *ortho*-hydroxyaniline derivative with the potential to form a quinone-imine species. However, there are no literature reports that provide evidence for the formation of electrophilic quinone species in the course of bosutinib metabolism. Cabozantinib was granted orphan drug status by the United States Food and Drug Administration (FDA) in 2011 and approved in 2012 for the treatment of medullary thyroid cancer. Cabozantinib contains two structural alerts: anilide and dialkoxyether. However, the primary pathway of cabozantinib clearance in humans involves *N*-oxidation by CYP3A4 (56) and does not appear to involve metabolism of either of the two structural alerts. Pazopanib and ponatinib have been approved for the treatment of renal cell/soft tissue carcinoma and chronic myeloid leukemia, respectively. The two drugs are associated with severe and sometimes fatal hepatotoxicity in clinical studies, which resulted in BBW labels (57, 58). A GSH conjugate has been detected in HLM incubations of pazopanib (59). Whether the RM is generated from a two-electron oxidation of the *para*-methylaniline structural alert in pazopanib is not clear. The metabolism of ponatinib proceeds via CYP-mediated *N*-demethylation and *N*-oxidation pathways. In addition, amide bond hydrolysis affords the corresponding carboxylic acid and amine (aniline) derivatives as metabolites (60). There are no reports on the oxidation of the aniline metabolite to an RM or RMs as a causative factor to account for the hepatotoxicity associated with ponatinib. Several structural alerts are also found in crizotinib (*ortho*-alkoxyaniline) and vandetanib (aniline, bromobenzene, and dialkoxyaromatic). Oxidative metabolism plays a significant role in the elimination of crizotinib in humans. Metabolic profiling demonstrated that crizotinib and a lactam metabolite (formed via oxidation of the piperidine ring) were the principal circulating components in plasma. Other metabolites, representing <10% of circulating radioactivity individually, included glucuronide and sulfate conjugates of *O*-desalkyl crizotinib and *O*-desalkyl crizotinib lactam. The *O*-desalkyl crizotinib metabolite is a hydroxyaniline derivative and possible precursor of electrophilic quinone-imine species. In the case of vandetanib, there is no evidence for metabolism on any of the structural alerts. Unchanged vandetanib, vandetanib *N*-oxide, and *N*-desmethyl vandetanib are the principal components detected in plasma and excreta following oral administration of vandetanib to humans (61).

The last two illustrations of recently approved drugs that contain structural alerts, carfilzomib and ibrutinib, are intrinsically electrophilic in nature. Carfilzomib is a selective and irreversible proteasome inhibitor that has been approved for the treatment of multiple myeloma. Carfilzomib is intrinsically electrophilic in nature owing to the presence of the epoxyketone group that irreversibly binds to the 20S proteasome (62). Despite its electrophilic nature and a high daily intravenous dose (900 mg), carfilzomib has demonstrated a favorable safety profile and significant antitumor activity in patients with relapsed and refractory multiple myeloma. In humans, carfilzomib has a short half-life ($t_{1/2} \sim 30$ min) and is cleared largely extrahepatically via peptidase cleavage and hydrolysis of the electrophilic epoxide (63). In vitro, carfilzomib demonstrates time- and concentration-dependent inhibition of human CYP3A4, consistent with mechanism-based inactivation (63). Direct alkylation of the enzyme by the epoxyketone moiety can be ruled out because enzyme inhibition requires NADPH cofactor. This observation suggests that carfilzomib is metabolized by CYP3A4 to a reactive species that covalently adducts to the CYP protein. RM trapping studies have not been performed with carfilzomib. In early 2013, the FDA granted ibrutinib breakthrough-therapy designations as a monotherapy for patients with two B cell malignancies: patients with relapsed or refractory mantle cell lymphoma who have received prior therapy and patients with Waldenström's macroglobulinemia. In November 2013, ibrutinib was approved for the treatment of mantle cell lymphoma. Ibrutinib is an orally administered selective covalent inhibitor of Bruton's tyrosine kinase. Like carfilzomib, ibrutinib is intrinsically electrophilic owing

to the presence of the acrylamide substituent that covalently binds to a cysteine residue in the tyrosine kinase (64).

DEALING WITH REACTIVE METABOLITES IN DRUG DISCOVERY—THE PATH FORWARD

The mere presence of structural alerts cannot in itself predict the type, severity, or incidence of IADRs associated with drugs. Likewise, RM screening tools (exogenous trapping with nucleophiles and/or protein covalent binding in HLM) are not intended to predict toxicity; rather, they are meant merely to detect the formation of RMs, some of which may carry a toxic liability. Experiments that unambiguously define a 1:1 relationship between RM formation (e.g., the in vitro and in vivo characterization of GSH conjugates) and toxicity in humans are extremely rare. Although GSH adducts and/or downstream mercapturic acid metabolites that are measured in vivo represent short-term exposure to RMs, protein adducts reflect the internal exposure of cells to RMs in vivo, which is more relevant for risk assessment purposes. Whether covalent binding measures in vivo are likely to be more informative about the in vivo safety risk than covalent binding studies in vitro remains to be established. This uncertainty arises from a paucity of data on absolute levels of in vivo covalent binding that could lead to a toxic outcome versus levels of binding that are safe. At the present time, there is no consensus on a preclinical discovery strategy to investigate safety hazards and risks posed by RM formation for a particular drug candidate in humans. Ultimately, only studies in humans can currently be used to unearth mechanisms of serious IADRs, to determine cause and effect with respect to RM formation in humans, and subsequently to determine clinical outcome. Although reducing exposure to RMs is viewed as a pragmatic approach to minimize IADR risks during drug development, these strategies should not rely solely on structural alert/RM information, as overall metabolic fate and other considerations (discussed below) provide additional valuable information that can be used in a weight-of-evidence approach toward risk assessment and management. For example, when a structural alert prone to RM formation is essential for intrinsic pharmacologic potency and cannot be replaced, additional information needs to be considered.

The fact that certain classes of drugs, such as the thienopyridine antithrombotics (e.g., clopidogrel), rely on RMs for pharmacological action underscores the concept that bioactivation per se need not equate to a toxicological response. Overall, this conundrum raises a fundamental question: Why are some RM-positive drugs safe, whereas others are not? A limitation of the in vitro RM screens is that they are typically conducted in HLM (in the presence and absence of NADPH cofactor); thus, they examine only CYP-catalyzed RM formation. In some instances, RM formation may be observed in microsomes in a CYP-dependent fashion, but in vivo, the compound may undergo a distinctly different and perhaps more facile metabolic fate that circumvents RM formation. For instance, both paroxetine and raloxifene form GSH conjugates and covalently bind to HLM in a CYP-dependent fashion; however, in vivo, the quinone precursors are metabolized principally via competing *O*-methylation and glucuronidation pathways, respectively (44, 65). It is tempting to speculate that in the modern drug discovery paradigm, paroxetine and raloxifene would likely not be considered as candidates for clinical development because of the high degree of microsomal covalent binding and GSH adduct formation (66). Minimizing false positives requires that initial RM assessments in HLM be followed by more detailed studies in fully integrated in vitro biological matrices such as hepatocytes and/or liver S9 fractions from both human and animal species (67). Establishing a clear understanding of the in vivo clearance mechanisms in animals and how they relate to RM formation in in vitro matrices would lead to data-driven decision making with regard to compound selection.

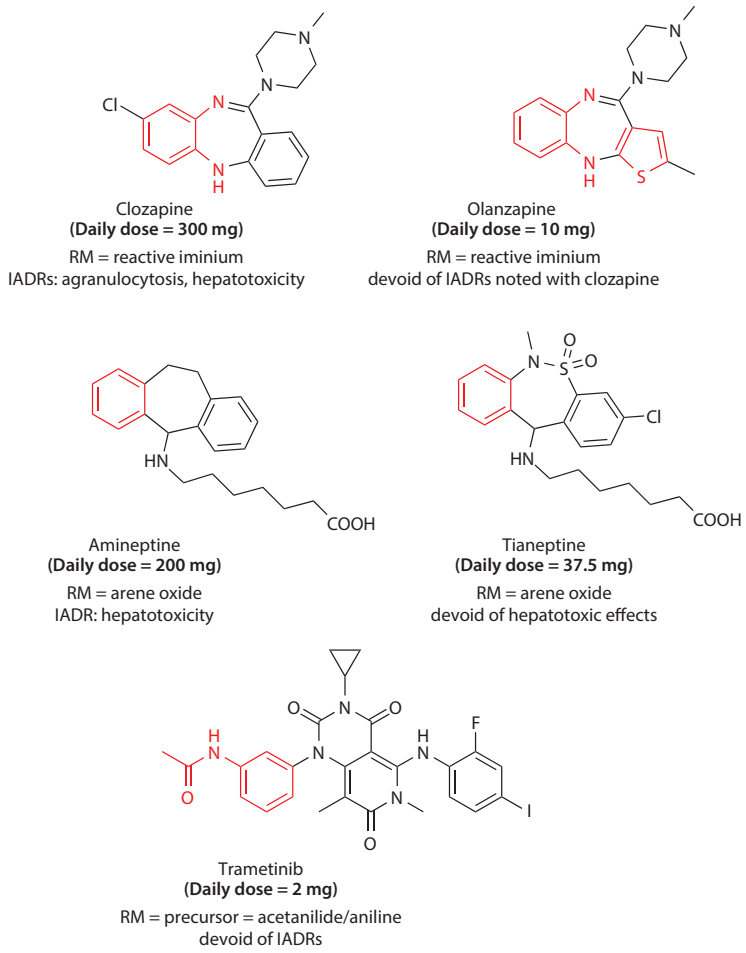


Figure 10

Low-daily-dose drugs do not exhibit the IADR liability associated with high-daily-dose drugs. Structural alerts are highlighted in red. Abbreviations: IADR, idiosyncratic adverse drug reaction; RM, reactive metabolite.

Comparison of the daily dosing regimen of toxic versus nontoxic drugs indicates that high-dose drugs (>100 mg) tend to be the ones that most frequently cause IADRs, whereas low-dose drugs (<50 mg) rarely are problematic in this regard (whether or not these agents are prone to RM formation) (16). The vast majority of structural alert–positive and/or RM-positive drugs in the top 200 list (in terms of dispensed prescriptions and total sales) are low-daily-dose drugs. The improved safety of low-dose drugs could arise from a marked reduction in the total body burden of RM exposure via efficient scavenging by GSH (and other competing metabolic pathways), such that the reactive species are unlikely to exceed the safety threshold needed for toxicity. For example, olanzapine (**Figure 10**) forms a reactive iminium metabolite analogous to the one observed with clozapine, yet olanzapine is not associated with a significant incidence of agranulocytosis (19). One difference between the two drugs is the daily dose; clozapine is given at a dose of >300 mg/day, whereas the maximum recommended daily dose of olanzapine is 10 mg/day. Another example becomes evident upon comparison of the tricyclic antidepressants amineptine

and tianeptine (**Figure 10**). Both drugs form reactive arene oxide species, but only amineptine is hepatotoxic (16, 68, 69). The improved tolerance of tianeptine in the clinic likely arises from the ~5–6-fold lower recommended dose relative to that of amineptine (daily doses of amineptine and tianeptine are 200 mg and 37.5 mg, respectively). Likewise, in the case of clopidogrel, the majority (>70%) of its daily dose of 75 mg is rapidly hydrolyzed by human carboxylesterases to the inactive carboxylic acid metabolite (~80–85% of circulating metabolites) (70), which means that only a small percentage of the parent drug (20 mg or less) is theoretically available for conversion to the active RM. Indeed, covalent binding to platelets accounts for only 2% of radiolabeled clopidogrel in human mass balance studies (71). Finally, trametinib (**Figure 10**), a kinase inhibitor recently approved for the treatment of patients with unresectable or metastatic melanoma, also falls into the category of a structural alert–positive agent but a low-daily-dose drug (72). Despite hydrolysis on the acetanilide alert to the corresponding aniline derivative as a primary metabolic pathway, no idiosyncratic hepatotoxicity has been noted thus far. This absence of hepatotoxicity could be ascribed to the very low recommended dosage (2 mg daily). Recent advances in risk assessment methodologies—such as the estimate of total daily body burden of covalent binding in hepatocytes and the zone classification, which takes the clinical dose into consideration—are positive steps toward quantitative prediction of IADR risks with drug candidates (40, 73, 74). Given this general trend in which low daily dose is a key factor in reducing IADR risks, optimization of lead compounds in drug discovery programs should focus on improving intrinsic pharmacologic potency and optimizing pharmacokinetics as a means of decreasing the projected clinically efficacious plasma concentrations (and hence the dose) and the associated body burden of a parent drug and its metabolites. However, there will be classes of drugs (e.g., antibacterials, antiretrovirals) for which this goal will be difficult to achieve.

Numerous drugs form RMs and cause idiosyncratic toxicity, yet they remain on the market and are widely prescribed because of favorable benefit-risk considerations; for example, Stepan et al. (16) found that aniline sulfonamide antibacterial agents such as sulfamethoxazole and sulfadoxine, which carry a BBW for IADRs including skin rashes and hepatotoxicity, are linked to RM formation. As noted in this work, the aniline/anilide motif is widely used in de novo kinase inhibitor design, as is evident with the recently approved tyrosine kinase inhibitors (see **Figure 9**). Some of them (e.g., pazopanib, ponatinib, and sunitinib) are even associated with cases of idiosyncratic hepatotoxicity, and available circumstantial evidence points toward RM liability as a potential causative factor in some instances. Lapatinib presents an interesting example that also illustrates the weight of unmet medical need over the risk of hepatotoxicity. Lapatinib is used in combination with capecitabine for the treatment of advanced or metastatic breast cancer and is associated with several cases of hepatotoxicity (some resulting in fatalities). Not only is the drug bioactivated to a quinone-imine species, resulting in covalent modification of the CYP3A4 isozyme (75), but its recommended daily dose is 1.25 grams. These observations suggest that the level of risk (e.g., idiosyncratic toxicity, DDI risk due to CYP inhibition) that would be deemed acceptable for drug candidates intended to treat major unmet medical needs, life-threatening diseases, and/or orphan diseases is significantly higher than the acceptable risk level associated with the treatment of chronic nondebilitating conditions for which alternate treatment options are already available. This issue also raises for debate a philosophical question regarding medicinal chemistry investments in removing structural alerts such as the aniline motif, which is widely utilized in kinase inhibitor programs and is challenging to mimic by isosteric replacement. The argument also applies to unprecedented pharmacologic targets or molecules that carry a significant risk with regard to predicted human pharmacokinetics, whose primary goal is to first demonstrate early signs of efficacy in the clinic and/or adequate systemic exposure. For such programs, RM-positive molecules can be advanced into first-in-human studies as probes to address pharmacokinetics

and proof of mechanism, provided they are deemed safe in standard preclinical toxicology studies. While proof of mechanism is being obtained, additional efforts can be invested in the identification of backup molecules that are devoid of RM formation. An illustration of such tactics is evident in our work on the RM-positive 5-trifluoromethylpyrido[4,3-*d*]pyrimidin-4(3*H*)-one short-acting series of calcium receptor antagonists. In these efforts, achieving a narrow window of human pharmacokinetics for safety and efficacy was a primary driver of success (76, 77).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Orr ST, Ripp SL, Ballard TE, Henderson JL, Scott DO, et al. 2012. Mechanism-based inactivation (MBI) of cytochrome P450 enzymes: structure-activity relationships and discovery strategies to mitigate drug-drug interaction risks. *J. Med. Chem.* 55:4896–933
2. Dobo KL, Obach RS, Luffer-Atlas D, Bercu JP. 2009. A strategy for the risk assessment of human genotoxic metabolites. *Chem. Res. Toxicol.* 22:348–56
3. Evans DC, Baillie TA. 2005. Minimizing the potential for metabolic activation as an integral part of drug design. *Curr. Opin. Drug Discov. Dev.* 8:44–50
4. Guengerich FP, MacDonald JS. 2007. Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.* 20:344–69
5. Park BK, Boobis A, Clarke S, Goldring CE, Jones D, et al. 2011. Managing the challenge of chemically reactive metabolites in drug development. *Nat. Rev. Drug Discov.* 10:292–306
6. Stachulski AV, Baillie TA, Park BK, Obach RS, Dalvie DK, et al. 2013. The generation, detection, and effects of reactive drug metabolites. *Med. Res. Rev.* 33:985–1080
7. Lucas A, Nolan D, Mallal S. 2007. HLA-B*5701 screening for susceptibility to abacavir hypersensitivity. *J. Antimicrob. Chemother.* 59:591–93
8. Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, et al. 2010. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat. Genet.* 42:711–14
9. Haas DW, Mootsikapun P, Ruxrungtham K, Podzamczer D, Nevirapine Toxicogenomics Study Team. 2012. Clinical perspectives on human genetic screening to prevent nevirapine toxicity. *Pers. Med.* 9:773–82
10. Uetrecht J, Naisbitt DJ. 2013. Idiosyncratic adverse drug reactions: current concepts. *Pharmacol. Rev.* 65:779–808
11. Park BK, Kitteringham NR, Maggs JL, Pirmohamed M, Williams DP. 2005. The role of metabolic activation in drug-induced hepatotoxicity. *Annu. Rev. Pharmacol. Toxicol.* 45:177–202
12. Ng W, Lobach AR, Zhu X, Chen X, Liu F, et al. 2012. Animal models of idiosyncratic drug reactions. *Adv. Pharmacol.* 63:81–135
13. Evans DC, Watt AP, Nicoll-Griffith DA, Baillie TA. 2004. Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.* 17:3–16
14. Argikar UA, Mangold JB, Harriman SP. 2011. Strategies and chemical design approaches to reduce the potential for formation of reactive metabolic species. *Curr. Top. Med. Chem.* 11:419–49
15. Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, et al. 2005. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.* 6:161–225
16. Stepan AF, Walker DP, Bauman J, Price DA, Baillie TA, et al. 2011. Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. *Chem. Res. Toxicol.* 24:1345–410
17. Shamovsky I, Ripa L, Blomberg N, Eriksson LA, Hansen P, et al. 2012. Theoretical studies of chemical reactivity of metabolically activated forms of aromatic amines toward DNA. *Chem. Res. Toxicol.* 25:2236–52

18. Liu ZC, Uetrecht JP. 1995. Clozapine is oxidized by activated human neutrophils to a reactive nitrenium ion that irreversibly binds to the cells. *J. Pharmacol. Exp. Ther.* 275:1476–83
19. Gardner I, Leeder JS, Chin T, Zahid N, Uetrecht JP. 1998. A comparison of the covalent binding of clozapine and olanzapine to human neutrophils in vitro and in vivo. *Mol. Pharmacol.* 53:999–1008
20. Uetrecht J, Zahid N, Tehim A, Fu JM, Rakhit S. 1997. Structural features associated with reactive metabolite formation in clozapine analogues. *Chem. Biol. Interact.* 104:117–29
21. Kalgutkar AS, Vaz AD, Lame ME, Henne KR, Soglia J, et al. 2005. Bioactivation of the nontricyclic antidepressant nefazodone to a reactive quinone-imine species in human liver microsomes and recombinant cytochrome P450 3A4. *Drug Metab. Dispos.* 33:243–53
22. Damsten MC, de Vlieger JS, Niessen WM, Irth H, Vermeulen NP, Commandeur JN. 2008. Trimethoprim: novel reactive intermediates and bioactivation pathways by cytochrome P450s. *Chem. Res. Toxicol.* 21:2181–87
23. Obach RS, Kalgutkar AS, Ryder TF, Walker GS. 2008. In vitro metabolism and covalent binding of enol-carboxamide derivatives and anti-inflammatory agents sudoxicam and meloxicam: insights into the hepatotoxicity of sudoxicam. *Chem. Res. Toxicol.* 21:1890–99
24. Davies NM, Skjodt NM. 1999. Clinical pharmacokinetics of meloxicam: a cyclo-oxygenase-2 preferential nonsteroidal anti-inflammatory drug. *Clin. Pharmacokinet.* 36:115–26
25. Bell JA, Dallas FA, Jenner WN, Martin LE. 1980. The metabolism of ranitidine in animals and man. *Biochem. Soc. Trans.* 8:93
26. Diao L, Shu Y, Polli JE. 2010. Uptake of pramipexole by human organic cation transporters. *Mol. Pharm.* 7:1342–47
27. Dieckhaus CM, Thompson CD, Roller SG, Macdonald TL. 2002. Mechanisms of idiosyncratic drug reactions: the case of felbamate. *Chem. Biol. Interact.* 142:99–117
28. Thompson CD, Barthen MT, Hopper DW, Miller TA, Quigg M, et al. 1999. Quantification in patient urine samples of felbamate and three metabolites: acid carbamate and two mercapturic acids. *Epilepsia* 40:769–76
29. Testa L, Bhindi R, Agostoni P, Abbate A, Zoccai GG, van Gaal WJ. 2007. The direct thrombin inhibitor ximelagatran/melagatran: a systematic review on clinical applications and an evidence based assessment of risk benefit profile. *Exp. Opin. Drug Saf.* 6:397–406
30. McKenney JM, Proctor JD, Harris S, Chinchili VM. 1994. A comparison of the efficacy and toxic effects of sustained- versus immediate-release niacin in hypercholesterolemic patients. *JAMA* 271:672–77
31. Pieper JA. 2003. Overview of niacin formulations: differences in pharmacokinetics, efficacy, and safety. *Am. J. Health Syst. Pharm.* 60:S9–14
32. Roecklein BA, Sacks HJ, Mortko H, Stables J. 2007. Fluorofelbamate. *Neurotherapeutics* 4:97–101
33. Hartz RA, Ahuja VT, Zhuo X, Mattson RJ, Denhart DJ, et al. 2009. A strategy to minimize reactive metabolite formation: discovery of (*S*)-4-(1-cyclopropyl-2-methoxyethyl)-6-[6-(difluoromethoxy)-2,5-dimethylpyridin-3-ylamino]-5-oxo-4,5-dihydropyrazine-2-carbonitrile as a potent, orally bioavailable corticotropin-releasing factor-1 receptor antagonist. *J. Med. Chem.* 52:7653–68
34. Subramanian R, Lee MR, Allen JG, Bourbeau MP, Fotsch C, et al. 2010. Cytochrome P450–mediated epoxidation of 2-aminothiazole-based AKT inhibitors: identification of novel GSH adducts and reduction of metabolic activation through structural changes guided by in silico and in vitro screening. *Chem. Res. Toxicol.* 23:653–63
35. Kalgutkar AS, Mascitti V, Sharma R, Walker GW, Ryder T, et al. 2011. Intrinsic electrophilicity of a 4-substituted-5-cyano-6-(2-methylpyridin-3-yloxy)pyrimidine derivative: structural characterization of glutathione conjugates in vitro. *Chem. Res. Toxicol.* 24:269–78
36. Kalgutkar AR, Ryder TF, Walker GS, Orr ST, Cabral S, et al. 2013. Reactive metabolite trapping studies on imidazo- and 2-methylimidazo[2,1-*b*]thiazole-based inverse agonists of the ghrelin receptor. *Drug Metab. Dispos.* 41:1375–88
37. Hagmann WK. 2008. The discovery of taranabant, a selective cannabinoid-1 receptor inverse agonist for the treatment of obesity. *Arch. Pharm.* 341:405–11
38. Sarabu R, Bizzarro FT, Corbett WL, Dvorozniak MT, Geng W, et al. 2012. Discovery of piragliatin—first glucokinase activator studied in type 2 diabetic patients. *J. Med. Chem.* 55:7021–36

39. Lennernas H. 2003. Clinical pharmacokinetics of atorvastatin. *Clin. Pharmacokinet.* 42:1141–60
40. Nakayama S, Atsumi R, Takakusa H, Kobayashi Y, Kurihara A, et al. 2009. A zone classification system for risk assessment of idiosyncratic drug toxicity using daily dose and covalent binding. *Drug Metab. Dispos.* 37:1970–77
41. Dansette PM, Libraire J, Bertho G, Mansuy D. 2009. Metabolic oxidative cleavage of thioesters: evidence for the formation of sulfenic acid intermediates in the bioactivation of the antithrombotic prodrugs ticlopidine and clopidogrel. *Chem. Res. Toxicol.* 22:369–73
42. Savi P, Zachayus JL, Delesque-Touchard N, Labouret C, Hervé C, et al. 2006. The active metabolite of clopidogrel disrupts P2Y12 receptor oligomers and partitions them out of lipid rafts. *Proc. Natl. Acad. Sci. USA* 103:11069–74
43. Wu G, Vashishtha SC, Erve JC. 2010. Characterization of glutathione conjugates of duloxetine by mass spectrometry and evaluation of in silico approaches to rationalize the site of conjugation for thiophene containing drugs. *Chem. Res. Toxicol.* 23:1393–404
44. Haddock RE, Johnson AM, Langley PF, Nelson DR, Pope JA, et al. 1989. Metabolic pathways of paroxetine in animals and man and the comparative pharmacological properties of the metabolites. *Acta Psychiatr. Scand. Suppl.* 350:24–26
45. Venkatakrishnan K, Obach RS. 2005. In vitro–in vivo extrapolation of CYP2D6 inactivation by paroxetine: prediction of nonstationary pharmacokinetics and drug interaction magnitude. *Drug Metab. Dispos.* 33:845–52
46. Zhao SX, Dalvie DK, Kelly JM, Soglia JR, Frederick KS, et al. 2007. NADPH-dependent covalent binding of [^{3}H]paroxetine to human liver microsomes and S-9 fractions: identification of an electrophilic quinone metabolite of paroxetine. *Chem. Res. Toxicol.* 20:1649–57
47. Chen Q, Ngui JS, Doss GA, Wang RW, Cai X, et al. 2002. Cytochrome P450 3A4–mediated bioactivation of raloxifene: irreversible enzyme inhibition and thiol adduct formation. *Chem. Res. Toxicol.* 15:907–14
48. Hucker HB, Stauffer SC, Balletto AJ, White SD, Zacchei AG, Arison BH. 1978. Physiological disposition and metabolism of cyclobenzaprine in the rat, dog, rhesus monkey, and man. *Drug Metab. Dispos.* 6:659–72
49. Devineni D, Curtin CR, Polidori D, Gutierrez MJ, Murphy J, et al. 2013. Pharmacokinetics and pharmacodynamics of canagliflozin, a sodium glucose co-transporter 2 inhibitor, in subjects with type 2 diabetes mellitus. *J. Clin. Pharmacol.* 53:601–10
50. Zhang D, He K, Raghavan N, Wang L, Mitroka J, et al. 2009. Comparative metabolism of ^{14}C-labeled apixaban in mice, rats, rabbits, dogs and humans. *Drug Metab. Dispos.* 37:1738–48
51. Weinz C, Schwarz T, Kubiza D, Mueck W, Lang D. 2009. Metabolism and excretion of rivaroxaban, an oral, direct factor Xa inhibitor, in rats, dogs, and humans. *Drug Metab. Dispos.* 37:1056–64
52. Stangier J. 2008. Clinical pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor dabigatran etexilate. *Clin. Pharmacokinet.* 47:285–95
53. Xtandi® [package insert]. Northbrook, IL: Astellas Pharma Inc.; 2012
54. Hoffmann M, Kasserra C, Reyes J, Schafer P, Kosek J, et al. 2013. Absorption, metabolism and excretion of [^{14}C]pomalidomide in humans following oral administration. *Cancer Chemother. Pharmacol.* 71:489–501
55. Chowdhury G, Shibata N, Yamazaki H, Guengerich FP. 2014. Human cytochrome P450 oxidation of 5-hydroxythalidomide and pomalidomide, an amino analogue of thalidomide. *Chem. Res. Toxicol.* 27:147–56
56. Cometriq® [package insert]. South San Francisco, CA: Exelixis Inc.; 2012
57. Votrient® [package insert]. Research Triangle Park, NC: GlaxoSmithKline; 2013
58. Iclusig® [package insert]. Cambridge, MA: Ariad Pharmaceuticals Inc.; 2012
59. Kenny JR, Mukadam S, Zhang C, Tay S, Collins C, et al. 2012. Drug-drug interaction potential of marketed oncology drugs: in vitro assessment of time-dependent cytochrome P450 inhibition, reactive metabolite formation and drug-drug interaction. *Pharm. Res.* 29:1960–76
60. Narasimhan NI, Dorer DJ, Niland K, Haluska F, Sonnichsen D. 2013. Effects of ketoconazole on the pharmacokinetics of ponatinib in healthy subjects. *J. Clin. Pharmacol.* 53:974–81
61. Martin P, Oliver S, Kennedy SJ, Partridge E, Hutchison M, et al. 2012. Pharmacokinetics of vandetanib: three phase I studies in healthy subjects. *Clin. Ther.* 34:221–37
62. Zhou HJ, Aujay MA, Bennett MK, Dajee M, Demo SD, et al. 2009. Design and synthesis of an orally bioavailable and selective peptide epoxyketone proteasome inhibitor (PR-047). *J. Med. Chem.* 52:3028–38

63. Wang Z, Yang J, Kirk C, Fang Y, Alsina M, et al. 2013. Clinical pharmacokinetics, metabolism, and drug-drug interaction of carfilzomib. *Drug Metab. Dispos.* 41:230–37
64. Herman SE, Gordon AL, Hertlein E, Ramanunni A, Zhang X, et al. 2011. Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood* 117:6287–96
65. Dalvie D, Kang P, Zientek M, Xiang C, Zhou S, Obach RS. 2008. Effect of intestinal glucuronidation in limiting hepatic exposure and bioactivation of raloxifene in humans and rats. *Chem. Res. Toxicol.* 21:2260–71
66. Obach RS, Kalgutkar AS, Soglia JR, Zhao SX. 2008. Can in vitro metabolism-dependent covalent binding data in liver microsomes distinguish hepatotoxic from nonhepatotoxic drugs? An analysis of 18 drugs with consideration of intrinsic clearance and daily dose. *Chem. Res. Toxicol.* 21:1814–22
67. Dalvie D, Obach RS, Kang P, Prakash C, Loi CM, et al. 2009. Assessment of three human in vitro systems in the generation of major human excretory and circulating metabolites. *Chem. Res. Toxicol.* 22:357–68
68. Genève J, Degott C, Letteron P, Tinel M, Descatoire V, et al. 1987. Metabolic activation of the tricyclic antidepressant amineptine: II. Protective role of glutathione against in vitro and in vivo covalent binding. *Biochem. Pharmacol.* 36:331–37
69. Fromenty B, Pessayre D. 1995. Inhibition of mitochondrial beta-oxidation as a mechanism of hepatotoxicity. *Pharmacol. Ther.* 67:101–54
70. Farid NA, Kurihara A, Wrighton SA. 2010. Metabolism and disposition of the thienopyridine antiplatelet drugs ticlopidine, clopidogrel, and prasugrel in humans. *J. Clin. Pharmacol.* 50:126–42
71. Plavix® [package insert]. Bridgewater, NJ: Sanofi-Aventis; 2013
72. Mekinist™ [package insert]. Research Triangle Park, NC: GlaxoSmithKline; 2014
73. Bauman JS, Kelly JM, Tripathy S, Zhao SX, Lam WM, et al. 2009. Can in vitro metabolism-dependent covalent binding data distinguish hepatotoxic from nonhepatotoxic drugs? An analysis using human hepatocytes and liver S-9 fraction. *Chem. Res. Toxicol.* 22:332–40
74. Thompson RA, Isin EM, Li Y, Weidolf L, Page K, et al. 2012. In vitro approach to assess the potential for risk of idiosyncratic adverse reactions caused by candidate drugs. *Chem. Res. Toxicol.* 25:1616–32
75. Teng WC, Oh JW, New LS, Wahlin MD, Nelson SD, et al. 2010. Mechanism-based inactivation of cytochrome P450 3A4 by lapatinib. *Mol. Pharmacol.* 78:693–703
76. Didiuk MT, Griffith DA, Benbow JW, Liu KK, Walker DP, et al. 2009. Short-acting 5-(trifluoromethyl)pyrido[4,3-*d*]pyrimidin-4(3*H*)-one derivatives as orally-active calcium-sensing receptor antagonists. *Bioorg. Med. Chem. Lett.* 19:4555–59
77. Kalgutkar AS, Griffith DA, Ryder T, Sun H, Miao Z, et al. 2010. Discovery tactics to mitigate toxicity risks due to reactive metabolite formation with 2-(2-hydroxyaryl)-5-(trifluoromethyl)pyrido[4,3-*d*]pyrimidin-4(3*H*)-one derivatives, potent calcium-sensing receptor antagonists and clinical candidate(s) for the treatment of osteoporosis. *Chem. Res. Toxicol.* 23:1115–26

Insights from Genome-Wide Association Studies of Drug Response

Kaixin Zhou and Ewan R. Pearson

Medical Research Institute, University of Dundee, Scotland, United Kingdom DD1 9SY; email: k.zhou@dundee.ac.uk, e.z.pearson@dundee.ac.uk

Annu. Rev. Pharmacol. Toxicol. 2013. 53:299–310

First published online as a Review in Advance on October 16, 2012

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-011112-140237

Keywords

GWAS, pharmacogenomics, drug response, genetics

Abstract

Early genome-wide association studies (GWAS) using relatively small samples have identified both rare and common genetic variants with large impact on severe adverse drug reactions, dosing, and efficacy. Here we outline the challenges and recent successes of the GWAS approach in disease genetics and the ways in which these can be applied to pharmacogenomics for biological discovery, determination of heritability, and personalized treatment. We highlight that the genetic architecture of drug efficacy reflects a complex trait yet that of adverse drug reactions more closely mirrors the architecture of Mendelian diseases and how this difference affects future study design. Given that multiple layers of biological data are increasingly available on large samples from biorepositories linked to electronic medical records, GWAS will remain a key component of the systems biology approach to uncovering small to moderate genetic determinants of drug response; these discoveries should move us closer to a personalized approach to health care.

INTRODUCTION

GWAS: genome-wide association study/studies

Considerable variability exists in patients' response to drug treatment, as manifested by poor treatment outcomes such as low efficacy, adverse events, and toxicity. The one-size-fits-all ideology that drives drug development to maximize its target population has contributed to this observed spectrum of variation in response. It is generally believed that understanding the biological mechanisms and genetic determinants of drug response will help to guide therapeutic strategies toward a better safety and efficacy profile (1, 2) and, ultimately, to a more stratified approach to therapy.

Knowledge-based candidate gene studies have been useful for identifying genetic variants that alter the pharmacokinetic properties of drug transporter or metabolizing enzymes that lead to treatment outcome difference among patients (3). However, the biological mechanisms through which many commonly used drugs affect treatment outcomes are poorly understood, and this has limited the success of genetic studies of pharmacodynamics.

Genome-wide association studies (GWAS) offer a hypothesis-free approach that systematically tests hundreds of thousands or more variants in the genome without prior knowledge of the location of the causal variants (4). Over the past few years, thousands of variants affecting the genetic basis of many common diseases and complex traits have been identified through GWAS (see, e.g., NHGRI GWAS catalog) (5). Although clinical translation of this burst in gene discovery has not been straightforward, novel insights have been revealed, which should allow biologists to uncover the complex mechanisms that lead to the various phenotypes (6, 7).

Early drug response GWAS have mostly focused on critical clinical problems such as individualized dosing or severe adverse drug reactions (ADRs). These studies have successfully identified a few dozen large effect genetic variants, for which clinical translation has been eagerly pursued (8–11). The fundamentals behind these successful studies and where they lead have been reviewed previously (1, 12, 13). Drug efficacy GWAS are slowly beginning to emerge, offering first glimpses at the complex genetic architecture underlying the variable treatment outcomes for some commonly prescribed drugs.

In this review, we focus on issues related to drug efficacy GWAS. Like complex traits such as height, drug efficacy is often observed as a continuous spectrum in the target population. It is, therefore, reasonable to anticipate that the genetic architecture for drug efficacy and other complex traits share many common features. Here, we first review the basics of GWAS design and the consensus view on genetic architecture from studies of many phenotypes and then summarize existing drug response GWAS results from both ADR and efficacy studies. Finally, we propose a few future directions in which drug efficacy GWAS may progress in the context of experience from GWAS of other complex traits.

GWAS OF COMMON DISEASES

The GWAS design became realistic only after the International HapMap Project comprehensively catalogued more than 10 million common polymorphisms across all the major ethnic groups (14, 15). The primary idea of GWAS is to systematically test the association between a phenotype and all the common polymorphisms in thousands of individuals. Given the strong linkage disequilibrium (LD) between common polymorphisms, a selection of a few hundred thousand independent single-nucleotide polymorphisms (SNPs), coupled with imputation, was found to be enough to tag the vast majority of common polymorphisms in the whole genome (16). Indeed, most commercially available GWAS genotyping arrays similarly focus on a subset of carefully selected SNPs to achieve more than 90% coverage of the common SNPs in the genome. These robust high-throughput genotyping platforms have finally made it possible to genotype thousands of individuals in an affordable and timely fashion (4).

Over the past few years, researchers across the world have joined forces to assemble large cohorts with many thousands of subjects for powerful GWAS. These have led to the discovery of well over 2,000 robustly replicated loci associated with common diseases and complex traits in the human population (17). It is worth noting here that by design, the associated SNPs identified through GWAS are likely simply tagging the causal variants in LD, and further investigations are often required to identify the true functional variants (18).

Genetic Architecture of Common Diseases

GWAS are designed and powered to detect associations through LD between common SNPs and causal variants. Therefore, it is not surprising that most of the variants identified through GWAS are common. With few exceptions, the majority of these variants have small to moderate genetic effects. For example, of the 40 published type 2 diabetes risk variants, only *TCF7L2* has an allelic odds ratio (OR) of 1.35; the ORs of all the other signals, which collectively explain around 10% of the European population's familial risk, range from 1.05 to 1.25 (17). The numbers of independent loci identified for different phenotypes vary greatly due to the differences in sample size between studies. One of the most fruitful complex traits under intensive investigation through GWAS is height, for which more than 100 loci have been identified. Collectively, these 100 loci account only for around 20% of the total variance in the population, which is considerably lower than the 80% heritability—that is, the overall genetic contribution to population variance—estimated by traditional twin studies (19). Simulation and theoretical modeling have demonstrated that incomplete LD between the causal variants and common SNP markers may explain a small part of the heritability underestimation (20, 21). However, almost without exception, the GWAS-identified loci can explain only a small proportion of the genetic variance for common diseases and complex traits, revealing the so-called missing heritability problem (22).

Alternative hypotheses of common disease genetic architecture have been proposed to address the missing heritability issue. The focus of debate is whether a large number of small effect variants across the whole allele frequency spectrum or a large number of rare variants with large effects are missed by current GWAS (23, 24). Empirical evidence points to a combination of the two models. For example, in type 2 diabetes, multiple independent common variants at the *TP53INP2* locus have independent contributions to disease risk (25), and the *HNF1A* locus harbors both GWAS-identified common risk variants and other highly penetrant variants causing monogenic forms of diabetes. Furthermore, a targeted sequencing study of the GWAS-identified *MTNR1B* locus found that multiple rare variants with significant functional consequences also contribute to type 2 diabetes susceptibility independent of the GWAS signal (26). Thus, multiple variants contributing to disease risk are segregating in the population at a wide allele frequency spectrum, and a combination of different genetic experiments, such as GWAS and deep sequencing, may be required to get a more complete picture of the genetic architecture of complex traits.

Revealing Novel Biological Mechanism

GWAS have delivered a breakthrough in our understanding of common disease susceptibility and variation in complex human phenotypes (23). The vast majority of GWAS signals were unknown to the field, as most of them fall outside of the protein coding regions and therefore were difficult to link to known biological pathways. For the large number of signals in intergenic regions or gene deserts, follow-up genetic fine mapping and functional work is usually required to identify the causal variants and the exact molecular mechanism through which the genetic effects are exerted. However, candidate genes can be nominated in many cases based upon prior functional knowledge of the sequences in and around the physical proximity of the GWAS signals. Anchored by

these candidate genes, documented protein-protein interactions and functional pathways can be integrated with GWAS results to reveal novel biological mechanisms for diseases and to identify novel drug targets beyond the GWAS-identified genes. A good example is the GWAS-identified association of variants in *IL12B* and *IL23R* with ankylosing spondylitis (27, 28). This observation pointed to the involvement of the *IL-23R* pathway, which had not previously been considered as biologically important for this disease. As a result, new therapies that block IL-17, another member of the IL-23 cytokine processing pathway, were assessed in the treatment of ankylosing spondylitis and have shown promising results in early small clinical trials (29). Due to the complex genetic architecture of disease and the length of time it often takes to develop and implement an effective treatment strategy, the translation of GWAS results into such successful clinical interventions has been limited to date. In the medium- to long-term, it is likely that the greatly enhanced understanding of the genetic determinants of common diseases and the resulting better understanding of their pathogenic pathways will accelerate progress toward a personalized medicine approach to treatment.

GWAS OF DRUG RESPONSE

The past few years have witnessed exciting pharmacogenomic discoveries from GWAS. As compared with those of common diseases or other complex traits, these studies typically use smaller samples and identify genetic markers that are strong predictors of severe ADR, dosing, and efficacy.

Rare Adverse Drug Reactions

Severe ADRs are often rare but have profound implications on patients, health care providers, regulatory agencies, and the pharmaceutical industry (30). Early GWAS of ADRs have successfully identified a few highly penetrant variants (31–34). It is worth noting that only one variant with large genetic effect was identified for each type of severe ADR, indicating that the genetic architecture of these phenotypes may be similar to that of traditional Mendelian diseases. However, none of the variants identified thus far fully account for all the episodes of a specific ADR. As GWAS platforms were designed to capture common genetic variants, most rare variants or low frequency variants are probably not well tagged by the current platforms. It is likely that there are more rare variants to be identified for these severe ADRs. With the rapidly falling cost of sequencing and the successful discovery of additional causal variants for Mendelian diseases through exome sequencing (35), further discoveries of rare variants underlying severe ADRs are expected in the near future (36).

Drug Efficacy Studies

Another area of active pharmacogenetic investigation through GWAS has been drugs with a narrow therapeutic window, such as warfarin and acenocoumarol. Candidate gene studies focusing on genes acting in the known pharmacokinetic and pharmacodynamic pathways found that variants in *CYP2C9* and *VKORC1* affect the treatment dose in a patient population (37, 38). In the initial GWAS of warfarin dosing with 181 patients, only *VKORC1* was found to have genome-wide significance. Further replication of the top hits confirmed *CYP2C9* as another main contributor, and the two variants can explain more than 40% variance in the warfarin drug dose required to maintain appropriate levels of anticoagulation (39). In a subsequent GWAS using 1,053 Swedish patients, a further variant in *CYP4F2* was found to explain 1.5% of the variation after adjusting for the major genetic effects from *CYP2C9* and *VKORC1* (40). This example clearly shows that, in addition to the variants with major impact, common variants with small effects still contribute to

variation in drug response, suggesting response to these drugs may have similar genetic architecture to common diseases and other complex traits in the human population.

An example of a common variant having a profound impact on drug efficacy is the recent discovery of a polymorphism near *IL28B* associated with interferon-α treatment success in hepatitis C virus infection. Three groups independently reported GWAS evidence pointing to common variants at the locus affecting drug efficacy across major ethnic groups (41–43). The C allele of rs12979860 is associated with better sustained virological response, and its frequency varies dramatically between major ethnic groups, ranging from 90% in East Asians to around 75% in Europeans and less than 50% in African Americans. This allele frequency difference can explain approximately half of the difference in response rates between African American patients and those of European ancestry (41). Moreover, in a subsequent GWAS, the same allele was shown to be strongly associated with spontaneous hepatitis C viral clearance among individuals of both European and African ancestry (44). This example demonstrates that the same common variant affects not only the complex host immune response to the virus but also the efficacy of drug intervention. Therefore, it is not unreasonable to expect that additional variants associated with the risk of common diseases will affect the response to commonly prescribed drugs given that the pathways involving the risk genes are often the targets of intervention. A further example is that carriers of the type 2 diabetes risk allele of variant rs7903146 in *TCF7L2* have weaker glycemic response to oral antidiabetic agent sulfonylureas (45).

Drug response GWAS have provided insight into the mechanism by which some commonly used drugs work, especially those agents that have been widely used but for which there remains uncertainty about their action mechanism (46, 47). For example, the first-line antidiabetic agent metformin has been used in clinic for more than half a century with good therapeutic outcomes and few dramatic adverse responses. However, little is known about its exact action mechanism apart from evidence that it activates AMP-activated protein kinase (AMPK) by inhibition of the mitochondrial respiratory chain (48), although its ability to lower blood glucose does not require AMPK (49). A recent GWAS using 3,920 type 2 diabetes patients found robust statistical evidence that polymorphisms at a locus that encompasses the *ATM* gene are associated with glycemic response to metformin (47). Although the observed effect size of the leading variant at this locus is small (allelic risk of treatment failure is 1.25, and it accounts for 2.5% of treatment outcome variation) and the causal variant and mechanism for the association have yet to be fully elucidated, these results indicate that variants in the well-known cancer gene *ATM* may potentially alter the glycemic response to metformin, revealing novel mechanisms for this old drug.

Genetic Architecture of Drug Response

To summarize the features of successful pharmacogenetic GWAS, we carried out a survey of the studies that reported at least one genome-wide significant ($p < 10^{-7}$) signal, as documented by the NHGRI GWAS catalog Web site (**http://www.genome.gov/gwastudies/**) by the end of February 2012. The fundamental difference between pharmacogenomics studies and studies of common diseases and other complex phenotypes lies in the sample size. The typical sample size for drug response GWAS is in the hundreds with only a few studies managing just over a thousand patients. In sharp contrast, GWAS of common diseases and other complex traits typically use a few thousand subjects in their initial screening, with recent GWAS meta-analyses including tens of thousands of subjects. The sample size difference has resulted in some distinct features of the signals found to date by pharmacogenomics GWAS.

The first important feature of the drug response GWAS findings lies in their allele frequency distribution. **Figure 1** plots the minor allele frequency distribution of GWAS-identified variants

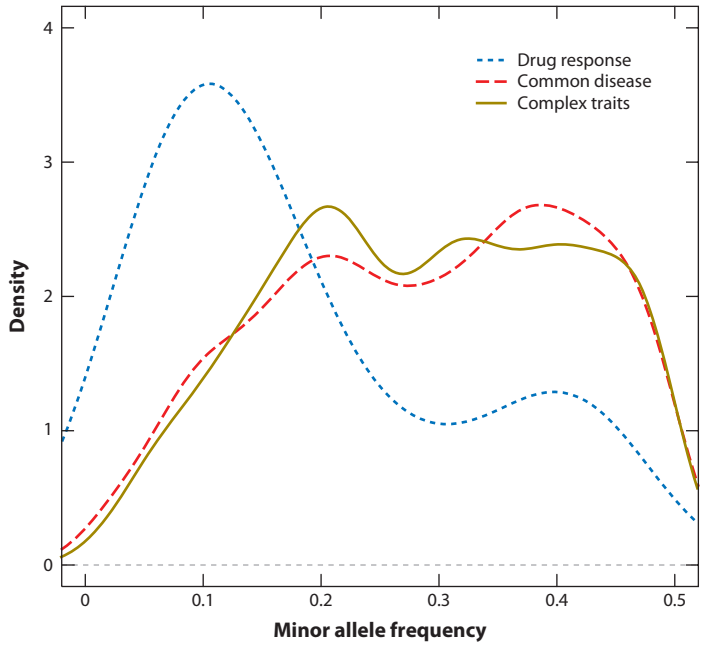


Figure 1

Density plot of minor allele frequency of the genome-wide significant ($p < 10^{-7}$) GWAS signals from the NHGRI GWAS catalog Web site (5).

for pharmacogenomics ($n = 31$), common diseases ($n = 1018$), and other complex traits ($n = 1285$). Although there is little difference between common diseases and other complex traits, an obvious excess of low allele frequency variants exists in drug response signals. This is likely due to the investigation of severe ADRs that found low frequency variants associated with the rare events. However, the main finding is that a significant number of drug response variants reached high allele frequency in the population.

A large proportion of the current pharmacogenetic GWAS findings have reported large genetic effects. When drug response is measured as a dichotomous trait, allelic OR values are often well over 3, as compared with the typical small to moderate effects seen in GWAS findings for common diseases (OR < 1.5). In quantitative phenotypes, it is not unusual that a single variant accounts for more than 10% of the variation in drug dosing or efficacy, whereas in other complex traits, one variant typically explains no more than a few percent of the variance. Clearly, common variants with large genetic effects on drug responses do exist. But whether this is the norm or the exception is still unclear given the limited number of pharmacogenomic GWAS findings published to date.

In summary, GWAS have provided a first glimpse at the genetic architecture of drug response. Rare severe ADRs may be similar to monogenic diseases that are driven by a few highly penetrant risk alleles, and using the next generation sequencing for variant discovery would be a better study design than GWAS to reveal more rare variants with high impact. The genetic architectures of drug efficacy and dosing are likely to be similar to those of common diseases and other complex traits that are determined by a combination of multiple common and rare variants (2). Therefore conducting well-powered GWAS has the potential to reveal additional variants, either of large effect with direct clinical implication or of novel biological mechanism, affecting the variation in drug response. Although array genotyping is becoming more affordable, assembling large well-phenotyped samples remains a hurdle for future drug response GWAS (1).

FUTURE DIRECTIONS

Electronic Medical Records

To meet the increasing demand of large samples for complex trait GWAS, the application of electronic medical records (EMRs) to high-throughput genomic research has been eagerly pursued in the past few years (50). Structural data such as patients' diagnoses, clinical biochemistry test results, and dispensed medications can be retrieved from EMRs to build a detailed picture of health-care history on each individual and identify subjects for genetic studies of certain phenotypes. Such EMR-based genetic research is better positioned to investigate the genetic determinants of drug response compared with data derived from clinical trials (51). In addition to the obvious advantage of cost effectiveness in sample recruitment, EMR-based studies can produce results that are more relevant to clinical practice. All the complexity of polypharmacy and environmental exposures that are present in clinical practice are not necessarily captured by clinical trials but can be identified through EMRs. Another essential advantage of EMR-based genetic research is the ability to assemble sufficient samples in a timely fashion. This is particularly important for investigations into rare ADRs that are limited in clinical trials but can be screened from large population–based EMRs. Further advantages of EMR-based genetic research are the flexibility offered in phenotypes available for study and the ability to standardize phenotypes across multiple EMR data sets.

Many large medical centers are currently constructing large biobanks alongside their EMR database to facilitate high-throughput genomic investigations and other omic research (52, 53). Despite the many hurdles that exist in adopting EMRs for genetic research, including sparse data collection and the low sensitivity in case identification, several proof-of-principle studies have been published using EMR and biobanking data to replicate the GWAS-identified variants of common diseases and other complex traits (54–56). For drug response, the recent metformin pharmacogenomics GWAS relied on EMRs for both the GWAS discovery and the initial replication step (47). Similarly, in a recent study, an EMR-linked biobank was used to replicate the association between *CYP2C19*2* and clopidogrel resistance (57). Given the sample size and statistical power advantages offered by EMR-based GWAS, novel insights into the genetic determinants of drug response with more ready application to clinical practice are expected from these ongoing endeavors.

Extreme Phenotypes

Drug response and efficacy often follow a continuous distribution in patients due to collective contributions from multiple genetic and environmental factors. Although both extreme responders and nonresponders exist, partial responders make up the bulk of the distribution. Evaluations of treatment efficacy in partial responders are often guided by a simple threshold in clinical guidelines without taking into account all other clinical features. Failure to account for such features could result in misclassification of patients with borderline outcomes and hence reduce the statistical power in subsequent genetic analysis (30). One way of maintaining reasonable statistical power while reducing the cost of data collection is through recruitment of extreme responders and nonresponders for genetic investigation. **Figure 2** plots the statistical power of different sample recruitment strategies to detect a common variant with small to moderate effect on treatment outcome. Assuming the minor allele frequency of a variant is 25% and it accounts for 0.5% or 1% of drug response variation in the whole population, selecting the top and bottom quartile of extreme responders and nonresponders for a case control analysis will have similar statistical power to studying the whole spectrum of responders but with only half of the data collection cost (58). Similarly, for rare variants with larger effects, selecting samples with extreme phenotypes for sequencing-based association tests could reduce the required sample size as much as fourfold compared with recruiting from the general population (59).

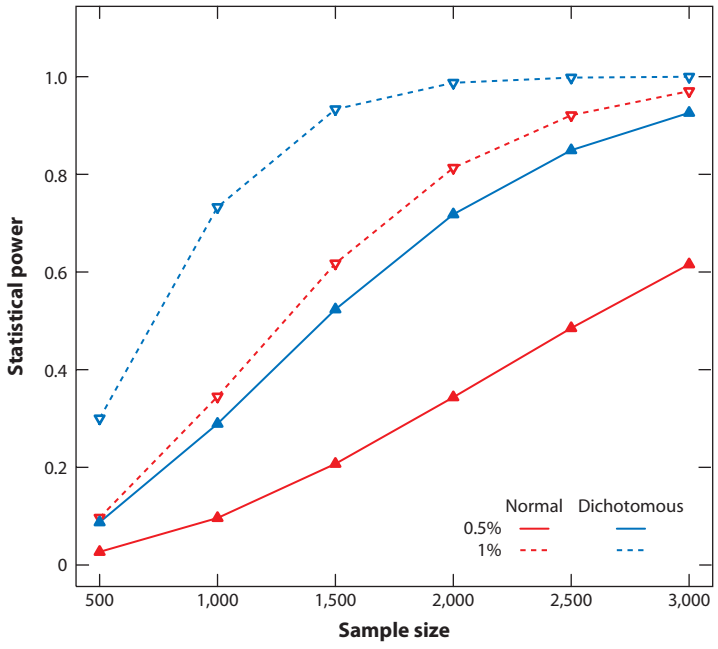


Figure 2

Statistical power calculations ($\alpha = 0.001$) for association tests of a quantitative locus accounting for 0.5% (*solid lines*) or 1% (*dotted lines*) variance with allele frequency of 25%. Studies using a normal population sample are plotted in red; studies using a dichotomous trait by selecting the top and bottom quarter of the population are plotted in blue.

Systems Biology Approach

Drug response GWAS have identified genes with large effects, but it is likely that additional variants with moderate effects may have failed the stringent threshold for genome-wide statistical significance in small to moderate studies. Given the limitation in sample sizes, one way of recovering this information is through pathway analysis that in principle tests the consistent moderate association signals in a group of genes involved in certain pathways (60). Pathways based on prior biological knowledge can be tested with summary statistics from signals in and around the genes involved and then adjusted for multiple testing at the pathway level. Such pathway-based analysis is particularly useful when genes that are most suitable for targeted intervention have only moderate statistical signals compared with other genes in the pathway, as demonstrated by the example described above of the association of the *IL-23* pathway with ankylosing spondylitis and the early promise of drugs targeting *IL-17* (27–29).

In addition to genetics, there is increasing interest in other omics, and in the next few years we are likely to see integration of EMR and other phenotype data with multiple layers of other omics, such as transcriptomics, proteomics, and metabolomics, on top of a GWAS or sequencing scaffold. Such data would benefit the interpretation of GWAS results in many ways. Over the past few years, several GWAS have been published documenting the loci that affect gene expression levels (eQTLs) in different tissues, and these eQTLs have been widely used to prioritize candidate genes at GWAS-identified loci (61–64). More importantly, eQTL data has been used to shed light on the biological networks through which genes affect phenotypes. For example, a *cis* eQTL has been identified for transcription factor *KLF14* in the adipose tissue, and this eQTL in turn significantly affects the expression levels of at least ten other genes in *trans* (i.e., distal to the

variant on the same chromosome or on a different chromosome) (65). Interestingly, large GWAS have shown that this eQTL and the variants in the ten *trans*-regulated genes are associated with a spectrum of metabolic traits, such as type 2 diabetes and HDL level, indicating a key role of *KLF14* in metabolic syndrome. A large proportion of pharmacogenomic GWAS-identified loci are also eQTLs, suggesting that genetic variants may also contribute to drug response diversity by affecting gene expression levels (66). Biological data at the other levels, such as proteomics and metabolomics, are expected to leverage the power of integrative genomics to provide additional insights into the biological mechanism of drug response.

Drug Response Heritability Estimation with GWAS Data

Most of the pharmacogenetic studies performed to date are based on empirical evidence that a patient's genome plays a role in his or her response to drugs. Very little is known about the heritability of drug response (i.e., the proportion of variation in drug response that can be explained by genetic factors). This is largely due to the fact that twin and family studies, which have been used successfully to estimate heritability for common complex traits, such as height, and common diseases, such as type 2 diabetes, are impractical for investigating drug response because family members are rarely exposed to the same drug in real life.

Given that most of the GWAS genotyping platforms can capture the majority of common genetic variation in the genome, it is now possible to estimate complex trait heritability through GWAS, and this is particularly helpful for traits such as drug response that have been difficult to investigate in traditional family or twin studies.

A few methods have been developed to use GWAS data to estimate total variance explained by SNPs in the GWAS panel (21, 67). Most prominently, Yang et al. (21, 68) have taken advantage of the relatedness between seemingly unrelated individuals in a GWAS to fit a linear model. This method can accurately estimate complex traits heritability close to values derived from traditional family and twin studies (68, 69). We have recently applied this method to the GWAS data on metformin response in 1,117 patients (47). Early results indicate that 36% (standard error = 32%) of the variation in metformin response can be explained by the variants covered on the GWAS panel, suggesting metformin glycemic response is likely to have a strong genetic determinant (K. Zhou, C. Palmer, and E. Pearson, unpublished data). It is worth noting that the heritability estimation in this preliminary study is high but not statistically significant, largely due to the fact that genetic variations contributed by causal variants are relatively small compared with those from the majority neutral variation captured by the whole GWAS panel. The successful application of this method to drug response heritability estimation therefore also relies on large sample sizes to achieve robust measurements (21).

CONCLUSION

Over the past few years, GWAS of drug response have identified a few dozen genetic determinants of severe ADR, dosing, and efficacy. Due to sample size limitation, most of the variants identified confer large genetic effects on drug response. As with GWAS of common diseases and other complex traits, these significant pharmacogenomic discoveries have added to our understanding of the mechanism of drug actions, leading to possible indirect benefits in health care but little immediate translation into clinical practice. Emerging drug efficacy GWAS suggest that the genetic architecture of drug response may be similar to that of common disease. Additional variants with small to moderate effects on a full allele frequency spectrum remain to be discovered. Using EMRs and biobanking to break the sample-size barrier and incorporating improved study design, including extreme phenotyping and systems biology analysis of multiple layers of biological data,

future GWAS can expand identification of genetic determinants of drug response. GWAS data can also provide a global estimation of total genetic contribution to variation in drug response that was not possible prior to the GWAS era.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Daly AK. 2010. Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.* 11:241–46
2. Roden DM, Wilke RA, Kroemer HK, Stein CM. 2011. Pharmacogenomics: the genetics of variable drug responses. *Circulation* 123:1661–70
3. Goldstein DB, Tate SK, Sisodiya SM. 2003. Pharmacogenetics goes genomic. *Nat. Rev. Genet.* 4:937–47
4. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9:356–69
5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106:9362–67
6. McCarthy MI, Zeggini E. 2007. Genome-wide association scans for type 2 diabetes: new insights into biology and therapy. *Trends Pharmacol. Sci.* 28:598–601
7. Perry JR, Frayling TM. 2008. New gene variants alter type 2 diabetes risk predominantly through reduced beta-cell function. *Curr. Opin. Clin. Nutr. Metab. Care* 11:371–77
8. Altman RB, Kroemer HK, McCarty CA, Ratain MJ, Roden D. 2011. Pharmacogenomics: Will the promise be fulfilled? *Nat. Rev. Genet.* 12:69–73
9. Guttmacher AE, McGuire AL, Ponder B, Stefansson K. 2010. Personalized genomic information: preparing for the future of genetic medicine. *Nat. Rev. Genet.* 11:161–65
10. Roden DM, Tyndale RF. 2011. Pharmacogenomics at the tipping point: challenges and opportunities. *Clin. Pharmacol. Ther.* 89:323–27
11. Rogowski WH, Grosse SD, Khoury MJ. 2009. Challenges of translating genetic tests into clinical and public health practice. *Nat. Rev. Genet.* 10:489–95
12. Wang L, McLeod HL, Weinshilboum RM. 2011. Genomics and drug response. *N. Engl. J. Med.* 364:1144–53
13. Daly AK. 2012. Using genome-wide association studies to identify genes important in serious adverse drug reactions. *Annu. Rev. Pharmacol. Toxicol.* 52:21–35
14. International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
15. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, et al. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58
16. Barrett JC, Cardon LR. 2006. Evaluating coverage of genome-wide association studies. *Nat. Genet.* 38:659–62
17. Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90:7–24
18. Singleton AB, Hardy J, Traynor BJ, Houlden H. 2010. Towards a complete resolution of the genetic architecture of disease. *Trends Genet.* 26:438–42
19. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832–38
20. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, et al. 2010. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42:570–75
21. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69

22. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
23. Hirschhorn JN. 2009. Genomewide association studies—illuminating biologic pathways. *N. Engl. J. Med.* 360:1699–701
24. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. 2010. Rare variants create synthetic genome-wide associations. *PLoS Biol.* 8:e1000294
25. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* 42:579–89
26. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, et al. 2012. Rare *MTNR1B* variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* 44:297–301
27. Danoy P, Pryce K, Hadler J, Bradbury LA, Farrar C, et al. 2010. Association of variants at *1q32* and *STAT3* with ankylosing spondylitis suggests genetic overlap with Crohn's disease. *PLoS Genet.* 6:e1001195
28. Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, et al. 2010. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.* 42:123–27
29. Song IH, Poddubnyy D. 2011. New treatment targets in ankylosing spondylitis and other spondyloarthritides. *Curr. Opin. Rheumatol.* 23:346–51
30. Crowley JJ, Sullivan PF, McLeod HL. 2009. Pharmacogenomic genome-wide association studies: lessons learned thus far. *Pharmacogenomics* 10:161–63
31. Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, et al. 2009. *HLA-B*5701* genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* 41:816–19
32. Link E, Parish S, Armitage J, Bowman L, Heath S, et al. 2008. *SLCO1B1* variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* 359:789–99
33. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, et al. 2011. *HLA-A*3101* and carbamazepine-induced hypersensitivity reactions in Europeans. *N. Engl. J. Med.* 364:1134–43
34. Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, et al. 2010. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat. Genet.* 42:711–14
35. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, et al. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12:745–55
36. di Iulio J, Rotger M. 2011. Pharmacogenomics: What is next? *Front. Pharmacol.* 2:86
37. Lindh JD, Lundgren S, Holm L, Alfredsson L, Rane A. 2005. Several-fold increase in risk of overanticoagulation by *CYP2C9* mutations. *Clin. Pharmacol. Ther.* 78:540–50
38. Wadelius M, Chen LY, Downes K, Ghori J, Hunt S, et al. 2005. Common *VKORC1* and *GGCX* polymorphisms associated with warfarin dose. *Pharmacogenomics J.* 5:262–70
39. Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, et al. 2008. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* 112:1022–27
40. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, et al. 2009. A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet.* 5:e1000433
41. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, et al. 2009. Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* 461:399–401
42. Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, et al. 2009. *IL28B* is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. *Nat. Genet.* 41:1100–4
43. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. 2009. Genome-wide association of *IL28B* with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* 41:1105–9
44. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, et al. 2009. Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* 461:798–801
45. Pearson ER, Donnelly LA, Kimber C, Whitley A, Doney AS, et al. 2007. Variation in *TCF7L2* influences therapeutic response to sulfonylureas: a GoDARTs study. *Diabetes* 56:2178–82
46. Tantisira KG, Lasky-Su J, Harada M, Murphy A, Litonjua AA, et al. 2011. Genomewide association between *GLCCI1* and response to glucocorticoid therapy in asthma. *N. Engl. J. Med.* 365:1173–83
47. Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, et al. 2011. Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. *Nat. Genet.* 43:117–20

48. Zhou G, Myers R, Li Y, Chen Y, Shen X, et al. 2001. Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Investig.* 108:1167–74
49. Foretz M, Hebrard S, Leclerc J, Zarrinpashneh E, Soty M, et al. 2010. Metformin inhibits hepatic gluconeogenesis in mice independently of the LKB1/AMPK pathway via a decrease in hepatic energy state. *J. Clin. Investig.* 120:2355–69
50. Kohane IS. 2011. Using electronic health records to drive discovery in disease genomics. *Nat. Rev. Genet.* 12:417–28
51. Wilke RA, Xu H, Denny JC, Roden DM, Krauss RM, et al. 2011. The emerging role of electronic medical records in pharmacogenomics. *Clin. Pharmacol. Ther.* 89:379–86
52. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, et al. 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* 84:362–69
53. McCarty CA, Chapman-Stone D, Derfus T, Giampietro PF, Fost N. 2008. Community consultation and communication for a population-based DNA biobank: the Marshfield clinic personalized medicine research project. *Am. J. Med. Genet. A* 146A:3026–33
54. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, et al. 2010. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am. J. Hum. Genet.* 86:560–72
55. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, et al. 2011. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am. J. Hum. Genet.* 88:57–69
56. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. 2010. A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS One* 5:e13011
57. Delaney JT, Ramirez AH, Bowton E, Pulley JM, Basford MA, et al. 2012. Predicting clopidogrel response using DNA samples linked to an electronic health record. *Clin. Pharmacol. Ther.* 91:257–63
58. Purcell S, Cherny SS, Sham PC. 2003. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–50
59. Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, et al. 2011. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet. Epidemiol.* 35:236–46
60. Wang K, Li M, Hakonarson H. 2010. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11:843–54
61. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. 2007. Population genomics of human gene expression. *Nat. Genet.* 39:1217–24
62. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. 2008. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 6:e107
63. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, et al. 2007. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* 39:1208–16
64. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. 2009. Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* 325:1246–50
65. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, et al. 2011. Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes. *Nat. Genet.* 43:561–64
66. Kasarskis A, Yang X, Schadt E. 2011. Integrative genomics strategies to elucidate the complexity of drug response. *Pharmacogenomics* 12:1695–715
67. So HC, Li M, Sham PC. 2011. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.* 35:447–56
68. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, et al. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43:519–25
69. Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88:76–82

Annual Review of Pharmacology and Toxicology

Volume 53, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml

Systems Pharmacology to Predict Drug Toxicity: Integration Across Levels of Biological Organization*

Jane P.F. Bai and Darrell R. Abernethy

Office of Clinical Pharmacology, Office of Translational Science, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland 20993; email: darrell.abernethy@fda.hhs.gov

Annu. Rev. Pharmacol. Toxicol. 2013. 53:451–73

First published online as a Review in Advance on November 5, 2012

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-011112-140248

Keywords

adverse drug effects, bioinformatics, systems analysis

Abstract

To achieve sensitive and specific mechanism-based prediction of drug toxicity, the tools of systems pharmacology will be integrated using structured ontological approaches, analytics, mathematics, and statistics. Success of this effort is based on the assumption that a systems network that consists of drug-induced perturbations of physiological functions can be characterized. This network spans the hierarchy of biological organization, from gene to mRNA to protein to intracellular organelle to cell to organ to organism. It is populated with data from each of these levels of biological organization. These data, from disparate sources, include the published literature, drug development archives of all approved drugs and drug candidates that did not complete development, and various toxicity databases and adverse event reporting systems. The network contains interrelated genomics, transcriptomics, and metabolomics data, as well as organ and physiological functional data that are derived from the universe of information that describes and analyzes drug toxicity. Here we describe advances in bioinformatics, computer sciences, next-generation sequencing, and systems biology that create the opportunity for integrated systems pharmacology–based prediction of drug safety.

INTRODUCTION

Mechanism-based prediction and evaluation of drug toxicity constitute an evolving science whose development is critical to drug discovery, development, and regulatory evaluation and whose goal is to make advances in human therapeutics available while protecting public health. Attaining this objective requires new approaches in integrative pharmacology that have only recently become available. Rapidly evolving data resources in genomics, transcriptomics, metabolomics, proteomics, integrated cellular and organ function, and their relationship to human physiology are used to populate the predictive systems. System structure and the computational framework and capacity to support a predictive capability are currently being developed. Efforts to utilize the same approaches for drug discovery have begun to shift the paradigm of traditional drug discovery, aiming at the goal of driving discovery through computerized information networking among human diseases, biology, and chemistry (1). This effort is prediction driven, whereby calculated statistical probabilities project a framework for defining the work plan that will proceed in development. The stated goal is to allow for better prediction of clinical efficacy and toxicity for a drug using mathematics, statistics, biology, chemistry, and computational science.

The drugs approved by the US Food and Drug Administration (FDA), as well as those that failed before or after approval, and their treatment-related effects including adverse reactions constitute a large data and knowledge resource of chemistry, clinical and animal phenotypes, and pharmacological and toxicological measurements. These data resources reside in fragmented forms and in many places, including the FDA's Drugs@FDA web page (2), PubMed, PubChem, the FDA's Adverse Event Reporting System (AERS) (3), primary medical records, new drug applications and investigational new drug submissions, similar data from international regulatory agencies, the pharmaceutical industry's archival data, and the published medical literature. For a clinical event that is suspected to be an adverse drug reaction (ADR), at the present time one typically has to explore many or all of these data sources to evaluate the potential linkage between the pharmacological mechanisms of a drug and the clinical phenotype of the individuals who have experienced the event. Extension of such an approach to sensitively and specifically predict such an ADR before it has been clinically observed is an important next step toward advancing systems pharmacology. Approaches to predictive toxicity are already being used in early drug discovery efforts to select compounds to move forward in development, reduce drug attrition, and reduce pharmaceutical research and development costs (4). These early discovery activities, although a starting place, are designed to have high sensitivity but are less concerned about specificity. In other words, they identify potential significant toxicity but have limited ability to determine whether the prediction is correct. Efforts that leverage human physiology, cellular biology and biochemistry, genomics, and drug chemistry are resulting in the creation of publicly accessible data sources that can provide the information support for systems pharmacology–based drug safety prediction (**Table 1**). Each of these data sources has its own unique focus and constitutes an essential part of a network that characterizes drug toxicity. This review highlights recent advances in the prediction and assessment of drug safety that will chart future directions for developing safer drugs and for identifying, in a prospective manner, patient subgroups that are either susceptible to, or protected from, a drug toxicity.

THE GLOBAL TOXICOLOGICAL NETWORK

Framed in a systems analysis context, physiological homeostasis is maintained by a hierarchy of functional domains of genetic sequence, gene transcription, transcriptional regulation, protein function and interaction, organelles, cells, and organs that are interconnected at each level of functional organization and across levels. Drug exposure may simply be viewed as a perturbation that alters the system. An important factor in understanding drug exposure is the characterization

Table 1 Key knowledge bases and databases

Database or knowledge base	URL
SIDER (computer-readable side effect resource)	**http://sideeffects.embl.de**
DrugBank	**http://www.drugbank.ca**
Chemical Effects in Biological Systems (CEBS)	**http://cebs.niehs.nih.gov/**
NCBI Database of Genotypes and Phenotypes (dbGaP)	**http://www.ncbi.nlm.nih.gov/gap/**
Comparative Toxicogenomics Database	**http://ctd.mdibl.org/**
Genetic Association Database (archive of human genetic association studies of complex diseases and disorders)	**http://geneticassociationdb.nih.gov**
Kyoto Encyclopedia of Genes and Genomes (KEGG) (bioinformatics resource for linking genomics to life)	**http://www.genome.jp/kegg**
The Pharmacogenomics Knowledgebase (PharmGKB) (resource describing how variation in human genetics leads to variation in response to drugs)	**http://www.pharmgkb.org**
Gene Expression Omnibus (GEO) (database repository of high-throughput gene expression data and hybridization arrays, chips, and microarrays)	**http://www.ncbi.nlm.nih.gov/geo**
Connectivity Map (detailed map that links gene patterns associated with disease to corresponding patterns produced by drug candidates and a variety of genetic manipulations)	**http://www.broadinstitute.org/genome_bio/connectivitymap.html**
The Gene Ontology (GO) (standardized representation of gene and gene product attributes across species and databases)	**http://www.geneontology.org**
Tox21 (Computational Toxicology Research program)	**http://epa.gov/ncct/Tox21**
International HapMap Project (database of genes associated with human disease and response to pharmaceuticals)	**http://hapmap.ncbi.nlm.nih.gov**
Human Interactome Database (database of human binary protein-protein interaction networks)	**http://interactome.dfci.harvard.edu/H_sapiens**
European Bioinformatics Institute (EBI) ArrayExpress Archive	**http://www.ebi.ac.uk/microarray-as/ae/**
NCI-60 DTP Human Tumor Cell Line Screen	**http://dtp.nci.nih.gov/branches/btb/ivclsp.html**
Library of Integrated Network-Based Cellular Signatures (LINCS)	**http://commonfund.nih.gov/lincs/**
Reactome	**http://www.reactome.org/ReactomeGWT/entrypoint.html**
Online Mendelian Inheritance in Man®	**http://www.ncbi.nlm.nih.gov/omim**

of organ, cellular, and subcellular drug distribution, in addition to systemic drug exposure that is derived from absorption, distribution, metabolism, and elimination (ADME) studies (**Figure 1*a***). The molecules of a drug interact with proteins and also interact directly or indirectly with RNAs, DNAs, and perhaps other cellular structural elements. These loci of interaction, termed connectivity nodes, interact with other nodes across the entire system. Some drug-induced stimuli have widespread system effects, whereas others cause more localized effects. For a drug-induced stimulus to propagate within the hierarchically structured network from one functional domain to the next, a critical threshold of functional disturbance presumably has to be reached. Perturbation of homeostasis propagates from the initial interaction(s) with cellular molecular target(s) to

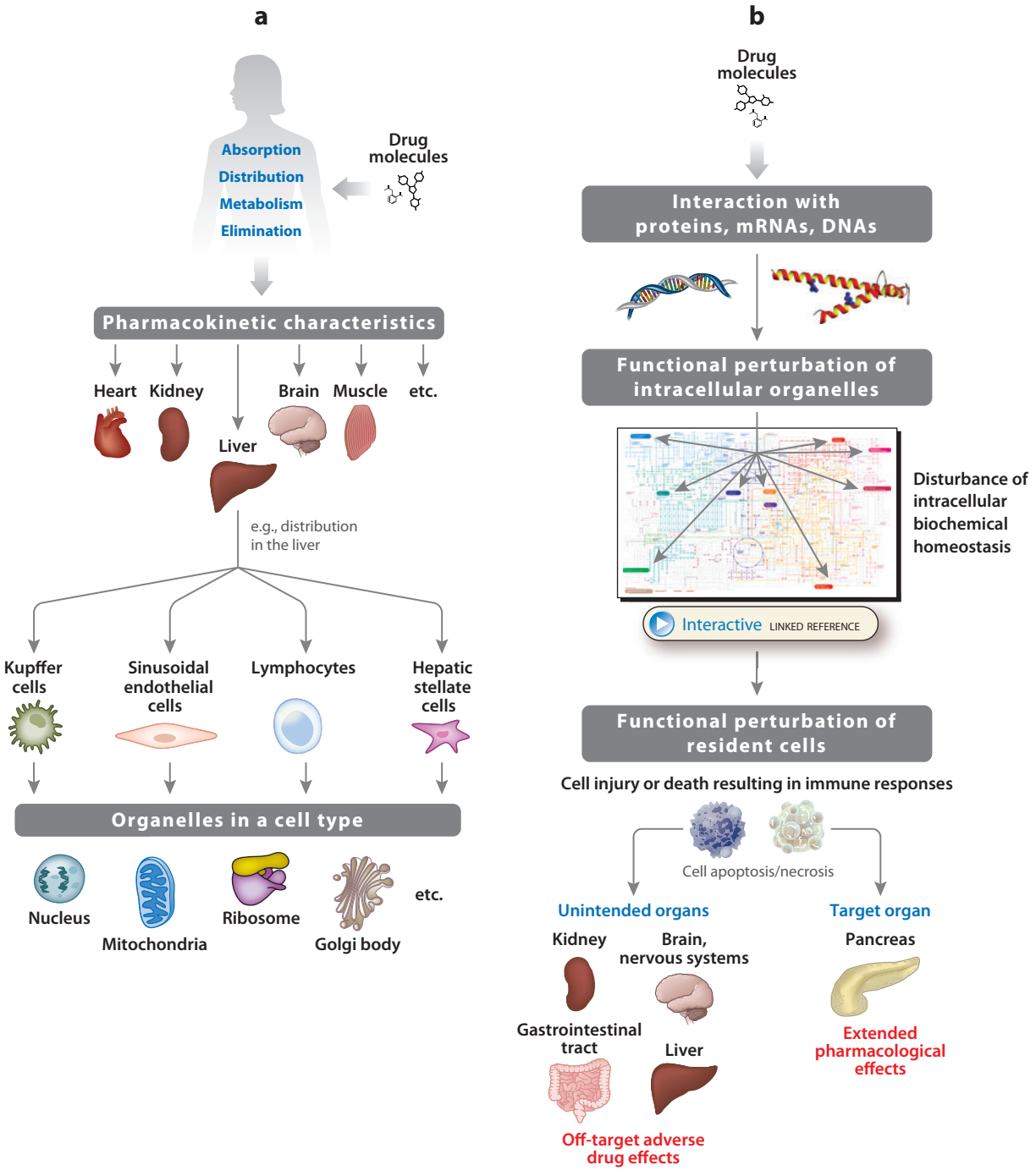


Figure 1

(*a*) Distribution hierarchy of drug molecules in the human body. (*b*) Propagation schematic of drug actions in the human body. The colored biochemical pathways map is taken from Reference 15 (**http://www.genome.jp/kegg**).

organelles [such as mitochondria (5, 6)], to cells (7), and to organs (8), with the final result being detectable or measurable drug action(s). One action may be the desired therapeutic effect, but other actions may result in parallel propagations that cause other (off-target) effects and systemic adverse reactions (8) (**Figure 1*b***). Both drug chemistry and individual patient phenotype form the basis for drug response, either therapeutic or toxic. Although individual drugs possess unique and diverse chemical structures, they are linked to one another by structural fingerprints and similarity (9), by overlapping transcriptomic patterns (10), by adverse reactions (11), by shared mechanisms of biotransformation such as cytochrome 450 enzyme-mediated metabolism (12), and by Anatomical Therapeutic Chemical (ATC) classification of pharmacological target (13) and pathway (14).

Human genomics and environmental factors are the backbone of the biochemical and physiological diversity of the human population, forming the blueprint for toxicological networks that can be described for each drug. These individual networks are then expanded into a global toxicological network based on the universe of drugs, drug candidates that failed, and perhaps other chemicals with known toxic signatures. The blueprint based on genetic makeup is modified by mutations (16) and environmental exposures (17). Included within the hierarchical structure of the entire toxicological network are detectable and evaluable phenotypes of clinical ADRs that are characterized as completely as possible using a structured vocabulary. As shown in **Figure 2**, the features of an ADR include severity and frequency at the population level, clinical signs and symptoms, and measurable metabolomic (18, 19) and functional biomarkers (20). Generally, little or no information is available about clinical phenotype measured at the molecular, cellular, and organ levels; however, as imaging methods advance, more information in this sphere will emerge. These layers of phenotype constitute the hierarchy of the drug-ADR network in an individual and in the patient population that is exposed to a drug. In addition to the genetic signature of the clinical phenotype, accumulated and new environmental influences (epigenetic factors) also impact this multidimensional toxicological network (21, 22). As a result, the global toxicological network will be a complex system, and using it to do predictive analysis will be computationally intensive. As experience with this approach accumulates, subsets of the global network likely will be used for specific drugs and toxicities.

Drug chemical structure can be linked to intracellular molecular pathways via protein-ligand binding profiles and protein-protein interactions (7, 23), to cellular biochemical biomarkers and phenotypes via metabolic pathways (7, 24, 25), to organ phenotypes (26), and eventually to clinical phenotypes and systemic biomarkers (27) (**Figure 3**). Although investigation of the safety profile of a drug candidate is required during its development, present clinical trial methodology is not developed to statistically investigate its ADR profile—instead, it is characterized in a purely descriptive manner. Consequently, observed ADRs are simply enumerated, and a systematic analysis of biological mechanism is often not performed. Reverse-engineering approaches have been conducted on a limited number of drug-ADR pairs, with the aim of understanding the translational linkage of the ADR across individual levels of the biological hierarchy (19). Now, with advances in computational sciences and the richness in data from various levels, FDA-approved labels (2), AERS (3), SIDER (computer-readable side effect resource; see Reference 28 and the website **http://sideeffects.embl.de/**), and electronic medical records are able to provide the necessary population-level phenotypic data to link clinical drug safety information to the global toxicological network (29, 30).

Each individual drug can be characterized by an ADR profile and toxicity network that consist of single or multiple sites of action that either may be completely independent of one another or may share different levels of connectivity. Probably in no case are these sites and connectivity links completely characterized. These sites can be thought of as biological hubs that consist of gene and gene product (including mRNA and protein), and mathematically they are characterized as nodes

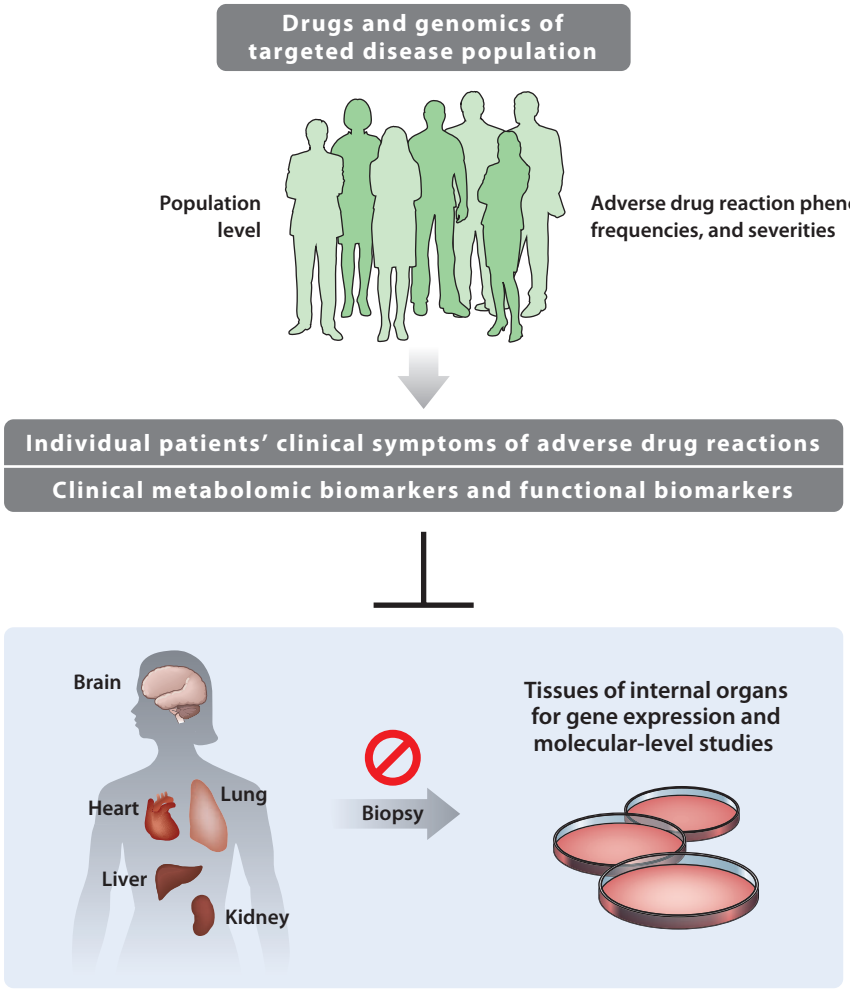


Figure 2

Current paradigm of drug safety needs translational bridging between clinical phenotype and molecular phenotype.

interconnected with one another via edges. Each edge is statistically weighted to represent the interaction intensity between each pair of nodes. The size of each network depends on the protein hub (see the Human Interactome Database at **http://interactome.dfci.harvard.edu/H_sapiens/**) or on a hub that represents another type of drug target (mRNA or DNA) with which the drug interacts. If the protein interacting with a drug molecule is upstream of a signaling pathway, then the cellular network that it perturbs may be larger than that perturbed by a drug molecule interacting with a protein downstream of a signaling pathway (15). The central hubs have been termed master regulators, and interaction with one of them may have a larger impact than interaction with downstream proteins. Techniques to establish the components and size of a network for a given drug or drug target are in development. The intensity of a drug-induced perturbation depends on the interaction (affinity or dissociation constant, also termed IC_{50}) between the drug and protein and on subsequent protein-protein interactions in the pathways downstream of the

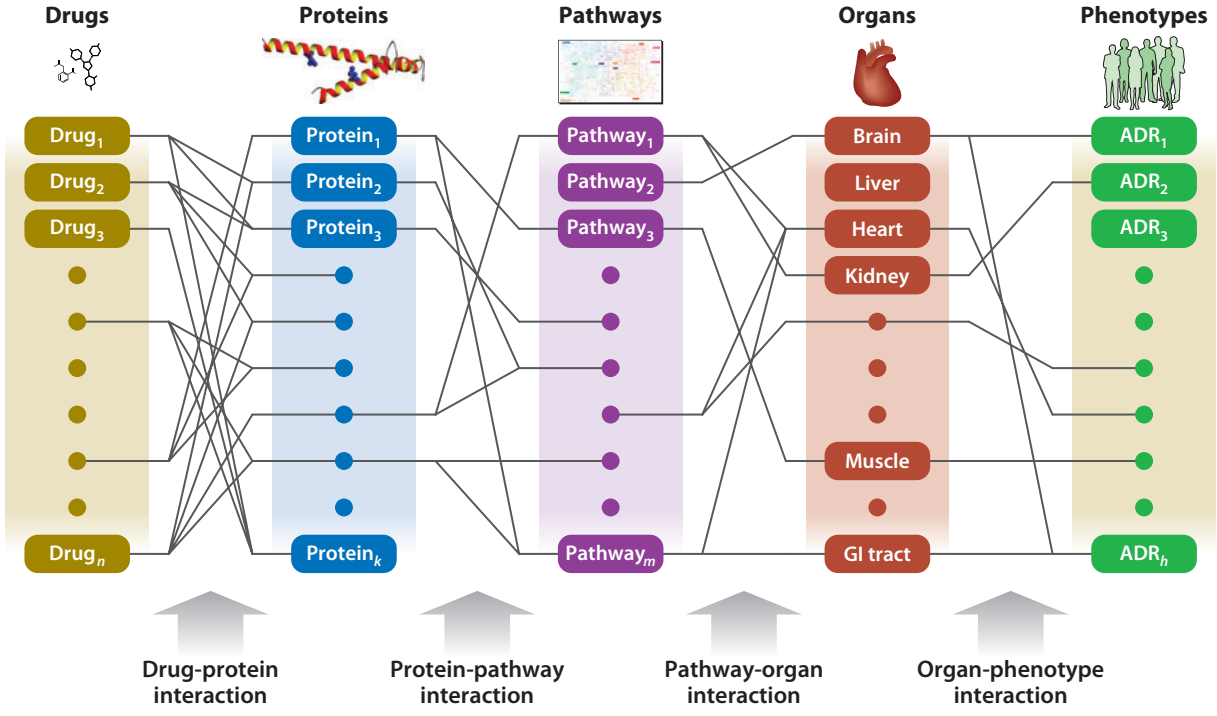


Figure 3

Interaction map showing *n* number of drugs, *k* number of proteins, *m* number of pathways, and *h* number of adverse drug reactions (ADRs). Abbreviation: GI, gastrointestinal.

targeted protein. In general, a gene network underlies an mRNA network, and a corresponding mRNA network underlies a protein network (10). Some drugs can interact with mRNAs or DNAs (31). A drug that interacts with a protein as a target can interact with multiple protein molecules (the therapeutic target and off targets) with different affinities (32); consequently, it can activate different signaling pathways or interact with multiple functional pathways. The signaling or functional pathways that are perturbed may have overlapping toxicological networks, resulting in synergistic or canceling system consequences. Similar toxicity profiles (33) of the same class of drugs are considered class effects (34) that are related to the on-target effects in off-target organs, and this similarity is expected to lead to toxicological networks that originate from hubs of the same or a similar nature and that share similar toxicological topologies. By extension of this concept, different drugs can cause similar ADRs that result from sharing similar toxicological pathways or networks (11, 35, 36). Housekeeping genes are expressed across organs and tissues, whereas tissue-specific genes are expressed differentially (37). A good understanding of tissue specificity is required to adequately define relevant genes and pathways in a specific organ and to identify the key nodes underlying the organ-specific safety profile of a particular drug.

INTEGRATION OF CONNECTIVITY WITHIN THE GLOBAL TOXICOLOGICAL NETWORK

Within the hierarchy of the global toxicological network (**Figure 1**), connectivity linkages can be made within each hierarchical level, between adjacent levels, or between two levels far apart in an interaction map. Approaches derived from applications of computer sciences, mathematics, and

statistics have been used to establish the connectivity within each level or among a few levels to derive a statistical prediction of ADRs. New methods are needed to understand and determine the degree of statistical uncertainty for a prediction within or across hierarchical levels.

Drug Chemical Characteristics as Predictors of Connectivity

The use of log $P_{(octanol/water)}$ and other molecular properties based on the chemical structure and substructural fragments of a drug (38) has led to quantitative structure-activity relationship (QSAR) methods that associate the structural characteristics of approved drug molecules with clinical ADRs described in the FDA-approved labels or in the databases that report adverse events (14, 39–42). Efforts to predict the toxicity potential of new drug candidates using mathematical network correlations of the ATC code, ADRs, and chemical properties have achieved approximately 50% sensitivity (43). Bender et al. (44) bridged chemical structures and ADRs using selected molecular targets that met a prespecified IC_{50} cutoff and thus incorporated drug-target interaction potency into safety prediction. Low et al. (45) applied QSAR methodology to rat transcriptomics data from a set of hepatotoxic and nonhepatotoxic compound pairs and concluded that QSAR-transcriptomic hybrid prediction did not perform better than transcriptomic prediction alone. QSAR approaches used alone may be more successful in predicting a specific drug-induced cellular toxicity phenotype, such as phospholipidosis (46), than in predicting a wide range of complex ADRs simultaneously. These approaches have been applied only to organic small-molecule drugs but not to peptide and protein drugs.

Genetic Signature as a Predictor of Connectivity

Single-nucleotide polymorphisms and haplotypes have been extensively used to explore as well as to describe the genetic basis of many human diseases and their response to treatment [see the Pharmacogenomics Knowledgebase (PharmGKB) at **http://www.pharmgkb.org/]**. Development of statistical connectivity between a genetic signature and a patient phenotype (47, 48) has been attempted for several diseases; for example, one study (49) links phenome to genome to describe and genetically define neuropsychiatric disorders.

Genome-wide association studies and candidate gene studies have been successfully used to identify causal genetic mutations associated with idiosyncratic ADRs. A recent review that highlighted research and accomplishments in this area (16) noted some successes, including association of HLA-B*5701 with abacavir hypersensitivity. Many strains of genetically manipulated (transgenic, knockout, and knockin) mice have been developed to investigate the relationship between a specific gene and animal phenotype (50, 51). Perhaps owing to the uncertainty about the interaction networks of a gene and its gene product, in some cases the observed phenotype is unlike that predicted from either the function of the protein encoded or the clinical phenotype in individuals carrying that genetic mutation. Protein expression assays may be conducted in transgenic animals to better understand the phenotype actually observed (52). One example: 36 inbred mouse strains were used to explore the high variability in acetaminophen drug-induced liver injury (DILI) in mice, and the genetic polymorphisms of these mice were selected to extensively cover human genetic variability (53). Humanized mouse models are increasingly used in drug toxicity research (54, 55), with most efforts centered on hepatotoxicity (55, 56). Most of the useful extrapolations of results from animals to humans have come from studies utilizing genes that are highly preserved across the species.

Transcriptomic Signature as a Predictor of Connectivity

Advances in microarray and next-generation sequencing technologies have allowed affordable gene expression profiling and enabled integration of transcriptomic profiling into drug discovery

and development (57). Tox21 is a National Institutes of Health (NIH) program that focuses on high-throughput toxicity testing (58), and the Gene Expression Omnibus (GEO) is a functional genomics data repository sponsored by the National Center for Bioinformatics for data mining (**http://www.ncbi.nlm.nih.gov/geo/**). The Japanese Toxicogenomics Project involving 15 companies and 150 compounds offers a large animal gene expression database for bioinformatics computational purposes (59). The NIH's Library of Integrated Network-Based Cellular Signatures program is creating a database for public use that will consist of molecular and cellular phenotypic signatures resulting from exposure of cell lines or primary cells or stem cells to drugs (see the Library of Integrated Network-Based Cellular Signatures at **http://commonfund.nih.gov/lincs**). Emerging methods for single-cell transcriptomics may have great use for the discovery of new targets, both therapeutic and toxic, that would otherwise be obscured by the heterogeneity of cells in transcriptomic profiling of cell populations (60). These approaches likely will allow investigators to explore the mechanisms of rare ADRs. Transcriptomics profiles can be obtained at multiple time points, which afford the opportunity to explore temporal interactions between drugs and biological systems.

Both whole-genome microarray (61) and differential expression of candidate genes (19) have been used to identify genes that are up- or downregulated, to allow generation of hypotheses and deduction of the biological pathways and networks that are perturbed (62) in relation to specific ADRs. Pinpointing the causative biological pathways perturbed by a drug that leads to an ADR, however, requires confirmation by changes at the translational level of corresponding proteins and requires availability of detection methods such as monoclonal antibodies (19). Furthermore, mass spectrometry methods are being developed for low-level protein identification and quantification (63). Clinical transcriptomics analysis to delineate the mechanisms of drug toxicity in humans is limited by inaccessibility of the target organ's tissue (**Figure 2**). Human peripheral blood mononuclear cells are therefore often used as a surrogate for inaccessible organs (64, 65). The validity of such an approach must be ascertained for each specific drug toxicity. A large data repository of transcriptomic signatures of human diseases is publicly accessible at the GEO website (**http://www.ncbi.nlm.nih.gov/geo/**). The processes of comparing and contrasting in vitro drug signatures from cell lines with human disease signatures have been helpful in repurposing approved drugs for new indications (66, 67) and may also be useful in discovering and understanding ADRs.

Epigenetic mechanisms, including DNA methylation at the cytosine residue and histone modifications, may be controlled in part by microRNAs (miRNAs) (68, 69), whose expression may be subject to epigenetic influences (70). Experimental findings show that changes in miRNA expression induced by chemicals or drugs can cause perturbation of cellular homeostasis (17, 71, 72). More than one thousand human miRNAs have been identified (see the miRBase database at **http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=hsa**). In addition to profiling mRNA expression, profiling miRNA expression following exposure to a drug may eventually be useful in establishing the molecular mechanism of ADRs (72, 73, 74). The study of fetal alcohol syndrome in a mouse embryo animal model showed that alcohol exposure during neurulation caused widespread alcohol-induced DNA methylation (22). Prolonged systemic administration of phenobarbital caused hypomethylation of *Cyp2b10* and its transcriptional activation in mouse liver (31). RNA editing by adenosine deaminases acting on RNA expands the functional complexity of the transcriptome and is far from completely described or understood. This editing may result in a large expansion beyond what would be predicted from variability in the genome, as suggested by identification of thousands of possible RNA editing sites (75). RNA editing commonly occurs in immature single-stranded mRNAs termed precursor mRNAs (pre-mRNAs); as a result, a given receptor has numerous isoforms. For example, more than 30 transcripts of the serotonin-2C receptor (5-HT_{2C}R) have been described (76). Haloperidol and risperidone altered RNA editing

of the 5-HT_{2C}R at different sites, resulting in differential translation of receptor isoforms, but neither clozapine nor chlorpromazine caused any significant editing (21). Therefore, drugs not only may alter gene transcription and epigenetic control of transcription but also may impact posttranscriptional editing, possibly resulting in a larger repertoire of potential adverse effects on cellular and organ homeostasis than is commonly appreciated.

Proteomic Signature as a Predictor of Connectivity

Signatures of the effects of drugs on cellular homeostatic networks can be readily detected in in vitro cell lines, allowing identification of the genes affected (10). A single gene can be transcribed and translated into several protein products through numerous complex pre-mRNA splicing and posttranslational modifications. In addition to causing changes at the transcript level, a drug can cause perturbation directly at the protein level by altering phosphorylation of a protein (77), which then impacts downstream processes. Publicly accessible protein interactome databases contain data that have been curated and supported with experimental findings (15; see also the Human Interactome Database at **http://interactome.dfci.harvard.edu/H_sapiens/** and the Reactome at **http://www.reactome.org/ReactomeGWT/entrypoint.html**). Protein-protein interactions can also be deduced through large-scale computational mapping (78) or text mining of study reports (79). Publicly accessible databases are useful for constructing subsets of protein-protein interactions to investigate toxicological pathways and networks that can be used in conjunction with the pathway networks that are also available in these databases. These databases can be used to construct a network hierarchy for any single FDA-approved drug (80, 81). The ADR profile of each drug can be overlaid on a more general toxicology network hierarchy to define the drug-specific toxicology network (81–83).

Correlation of drug-protein interactions with ADRs on the basis of drug-protein docking scores has been explored to determine the utility of this approach for the prediction of a new chemical entity's safety profile (84). Investigators can use drug-protein docking simulation to explore the result of genetic mutation by manipulating the amino acid sequence of a protein and consequently its 3D binding pocket, and this result may be useful in understanding the role of genetic mutations in ADRs. One study showed that MHC1 B*5701 rather than MHC1 B*5703 was the risk allele for abacavir hypersensitivity (85). In the case of clozapine-induced agranulocytosis, gene expression data are in agreement with the prediction of an off-target protein based on drug-protein docking (86). Such prediction based on drug-protein interactions was further expanded to include the putative biological pathways associated with individual protein targets, and, in at least one case, it identified a specific ADR-pathway association (87).

Correlations of the ADR profiles of drugs and their individual drug-protein binding assays or IC_{50} values across a large number of drug-protein pairs provide an important part of the description of the pharmacological basis for safety signals (35, 44). Yang et al. (88) took advantage of the availability of a large number of in vitro binding constants (Kd) between kinases and inhibitors to develop a computational predictive correlation between the Kd values for individual drug-kinase pairs and the frequencies of clinical diarrhea, rash, and fatigue. A large-scale bioinformatics computation by Yamanishi et al. (36) attempted to integrate the pharmacological similarity derived from the approved labels of drug products, the structural similarity between drugs and Tanimoto coefficients, and the protein sequence similarity derived from the Kyoto Encyclopedia of Genes and Genomes (KEGG) website (15). Four types of drug target were studied: enzymes, ion channels, G protein–coupled receptors, and nuclear receptors. This approach was more effective in linking pharmacological similarity to drug target than in linking chemical similarity to pharmacological similarity; nonetheless, it will likely stimulate further work in this area. Computational linkage

between disease neighborhoods and human protein-protein interactions has also been attempted, in an effort to identify the putative biological network associated with a physiologically defined drug-induced arrhythmia. Knowledge of the electrical mechanism that defines the physiology of the arrhythmia was utilized (89). Mass spectrometry–based proteomics can add further details about proteins and pathways involved in drug toxicity (90) for integration into computational models.

Organelle-Based Approaches as Predictors of Connectivity

Drug-induced mitochondrial toxicity is responsible, at the organelle level, for several clinical ADRs. The key role of mitochondria in the production of cellular energy via ATP formation due to aerobic respiration is consistent with the adverse cellular consequences of drug-induced mitochondrial dysfunction. Toxicity can be measured by the change in mitochondria permeability and the production of reactive oxygen species (oxidative stress). Such changes are the hallmarks of mitochondrial dysfunction in acetaminophen hepatotoxicity, presumably caused by its reactive metabolite NAPQI (*N*-acetyl-*p*-benzoquinone imine) (5, 6). In addition, inhibition of mitochondrial fatty acid oxidation has also been implicated in DILI (91, 92). Amiodarone and tetracycline differ greatly from each other, but both cause treatment-related hepatic steatosis (25). Exposure to either drug can result in the formation of intracytoplasmic lipid droplets and significant elevations of intracellular triglycerides. These histological features result from altered expression profiles of the genes involved in fatty acid transport and lipid metabolism (25). The mitochondrial basis of drug toxicity prediction is usually assessed using a composite score, such as the one that includes the mitochondrial DNA (mtDNA) expression profile, mitochondrial membrane potential, ATP level, electron transport chain, cell viability, and cellular reactive oxygen species level (93). The human mtDNA has been completely sequenced (94). Future understanding of mtDNA mutations and mtDNA transcriptomics in relation to ADRs may help shed light on the incidence of rare ADRs. The data supporting drug-induced mitochondrial dysfunction will constitute the information base that will support organelle-level connectivity in a systems pharmacology predictive network. The database for drug-induced toxicity for other cellular organelles is more limited and will require further development.

Organ-Based Approaches as Predictors of Connectivity

Within an organ, a single cell is surrounded by the same or different types of resident cells; consequently, its cellular phenotype, manifested either biochemically or functionally, is influenced by its neighboring cells. Coculture of a hepatoma cell line and an inflammatory cell line, compared with hepatoma cells cultured alone, reportedly increased the sensitivity of detecting the cytotoxic effect of troglitazone (95). Environmental milieu can impact the phenotypes of the biological networks perturbed by a drug even in different cell culture conditions (24). Coculture studies between cardiomyocytes and fibroblasts and between endothelial cells and vascular smooth muscle cells also demonstrated that cell-cell interactions caused changes in gene expression profile and in cellular phenotype (96, 97), indicating the importance of considering intercellular interactions in drug toxicity. In the widely used two-dimensional (2D) cell culture environment, the cells lose the 3D structural interactions with their neighboring cells that are otherwise present in vivo in the organ where they reside. Clearly, 2D cell culture does not completely capture the local organ milieu, including the close contact with local microvasculature and blood circulation to which a cell is exposed. A 3D microfluidic biochip of HepG2/C3A cells (98) and a microscale human-on-a-chip tissue of hepatocyte/fibroblast coculture (99) that mimics the 3D organ microenvironment are

being developed to explore metabolomics-on-a-chip (100). These organs-on-a-chip are designed to simulate the local in vivo structural, biochemical, and physical environment in order to replicate relevant spatiotemporal dynamics that cannot be replicated and captured by static cell cultures (101). Huh et al. (101) reported that an in vitro lung-on-a-chip provided an expression readout of pulmonary proinflammatory responses to silica nanoparticles in the presence of breathing motion that would not have been observed in in vitro resting cell cultures. The organ-on-a-chip system, if successfully constructed for an organ of interest, is hypothesized to be better than the static cell culture system at mimicking the in vivo milieu to which a cell is exposed. Expression array data from these systems may provide more relevant information to populate the support structure for a systems pharmacology predictive network.

Regional structural characteristics within an organ can dictate regional phenotypic differences in a specific drug-induced toxicity. For example, the liver has differences in the hepatic zonal distribution of xenobiotic induction of cytochrome 450 enzymes (102) that cannot be characterized by in vitro cell culture studies. In addition to traditional histological studies, gene expression profiling (103) and binding of radiolabeled drug molecules (104) have been used to describe organ region-specific drug toxicity. Transcriptomic signatures of organs following exposure to a drug may be quite similar to one another, with some organ-specific differences (103, 105). These various types and multiple levels of data characterize the toxicity phenotype of a drug-exposed organ, and they highlight the fact that limited experimental data are available to populate a systems pharmacology predictive toxicity model across organs and drugs. Efforts are under way to identify clinically measurable biomarkers that predict and/or reflect organ-specific toxicity, such as the efforts of the Predictive Safety Testing Consortium (**http://c-path.org/pstc.cfm**).

Detection of specific clinical organ toxicities usually relies on serum biomarkers and functional testing. Routinely used in clinical settings, serum ALT, AST, and bilirubin are associated with liver toxicity (20), whereas serum cardiac troponin-T (106), N-terminal pro-B-type natriuretic peptide, and midregional proatrial natriuretic peptide (107) are associated with cardiac toxicity or cardiac functional impairment. These biomarkers are insufficiently organ specific and in many cases not predictive of clinical drug toxicity; thus, developing better organ-specific and predictive biomarkers is an urgent need. Different toxicity phenotypes can occur in a single organ; for example, QT prolongation and non-QT cardiac toxicity, such as congestive heart failure, have different physiological and pathological origins and are clinically diagnosed using electrocardiography and echocardiography or cardiac magnetic resonance imaging, respectively. Clearly, linking biological pathways that lead to a specific organ ADR requires bridging with specific clinical translational biomarkers.

Interorgan Relationships as Predictors of Connectivity

Clinically, drug-induced injury to one organ often is associated with dysfunction in other organs, and clinical drug toxicity is defined by the entire set of organ dysfunctions. Acute kidney injury, for instance, causes adverse functional impact on distant organs including the brain, gastrointestinal tract, liver, lung, heart, and bone marrow (108). Interrelationships among organ dysfunctions can sometimes be linked by protein biomarkers. Cystatin C and β_2-microglobulin, for example, are markers of renal function but also are predictors of cardiovascular diseases (109). Another example is adenosine 5′-monophosphate-activated protein kinase, a metabolic master switch (110) that, in conjunction with adiponectin, also plays a key role in hepatic steatosis (111, 112) and renal disease (112, 113). These proteins are considered to be nodes that connect one disease to another. Interorgan interactions can also be defined through a connectivity network that is created by drug actions, both therapeutic and toxic. Cisplatin is known to be nephrotoxic, to be ototoxic, and to have some myelosuppressive effects. The constellation of these toxicities would then be

the connectivity network that characterizes its toxicity. Furthermore, drug-induced toxicity in one organ can result in secondary toxicity in another; for example, DILI results in encephalopathy that is secondary to the liver injury (114). Accounting for organ-organ interactions in the algorithms of drug safety prediction and assessment will be necessary to create a comprehensive systems pharmacology predictive network.

Medical Dictionary for Regulatory Activities (MedDRA): a standardized medical terminology developed by the International Conference on Harmonisation (ICH) to classify adverse event information associated with the use of biopharmaceuticals and other medical products

Clinical Description of Drug Toxicity via Medical Dictionary for Regulatory Activities Terms

Injury of an organ following a drug treatment is described clinically using the Medical Dictionary for Regulatory Activities (MedDRA), a standardized vocabulary of multiple diverse clinical phenotypes (115). For example, MedDRA defines at least 13 phenotypes of DILI, including acute hepatic failure, acute fatty liver with lactic acidosis, cirrhosis, acute hepatic necrosis, bland cholestasis, acute viral hepatitis–like liver injury, autoimmune-like hepatitis, cholestatic hepatitis, nodular regeneration, immunoallergic hepatitis, nonalcoholic fatty liver, vanishing bile duct syndrome, and sinusoidal obstruction syndrome (114). The molecular basis for most of these DILI phenotypes is different (25, 116, 117). As the MedDRA terms constitute an international standard, it is necessary to group terms under headings that can be effectively mapped to molecular mechanisms of drug toxicity. This effort is currently under way. To properly understand clinical ADRs following exposure to a drug, a complete patient phenotype including laboratory, imaging, and other diagnostic data in addition to signs and symptoms will be critical. As the development of a standardized electronic health record evolves, this information will become more available (115). ExPub, an integrated toxicity database of more than 400,000 compounds that uses MedDRA terms (**http://www.ebscohost.com/academic/expub**), is another data resource for enriching the global toxicological network.

FRAMEWORK FOR CONSTRUCTING THE GLOBAL TOXICOLOGICAL NETWORK

In the effort to predict and assess the safety profiles of future drug candidates, the immediate goal is to construct a global toxicological network that is expandable and that allows the continually growing data and knowledge resources to be effectively integrated and utilized. We envision that this network will be composed of local toxicological subnetworks of individual drugs and that it can be built by utilizing the ADR phenotypes of FDA-approved drugs and failed drugs (ones that are either withdrawn or aborted prematurely). This goal will be accomplished by linking the drugs' ADR phenotypes to molecular targets of toxicity, then mapping these targets up through increasingly complex hierarchical levels of human physiology.

Translation of Molecular Perturbations to Organ Dysfunction at the System Level

The hierarchy of drug actions beginning with molecular interactions and intracellular pathways and eventually leading to manifestations of ADRs, as laid out in **Figures 1–3**, lacks definitive pathways that translate disturbances at the cellular and organ levels of the functional connectivity network into clinical ADR phenotypes. For example, acetaminophen, troglitazone, amiodarone, tetracycline, and flucloxacillin cause different clinical phenotypes of DILI (8, 25, 27, 118, 119), 13 of which are defined by the MedDRA, as noted above (114). These different phenotypes present the challenge of differentially associating the underlying mechanism of drug action (therapeutic and/or toxic) with its unique phenotype of organ toxicities. Achieving that goal requires a comprehensive mapping of the perturbed pathways/networks to the corresponding biochemical

functional (metabolic) consequences across the hierarchy from molecular pathways to organelles to cells to organs. Metabolomic profiles are linked to cellular metabolic pathways, which form the connectivity blueprint of functional homeostasis and are expected to provide much of the data to support this part of the predictive network. Metabolomic profiles can often be assessed by sampling serum, urine, feces, body fluid, and tissue samples from animals and/or humans. Metabolic pathways of carbohydrates, energy, lipids, nucleotides, amino acids, cofactors, vitamins, terpenoids, polyketides, and xenobiotics as well as biosynthesis pathways have been mapped, and details are continuously being added (15). Metabolic pathways will be integrated into a global metabolic network (82) to allow the association of the metabolic pathways involved in cellular function homeostasis with diseases (15, 19, 120) and with drug toxicity (19, 121). As the data to populate a global metabolic network are developed, it can be expanded to enable identification of specific metabolomic and other biomarkers as the translational bridge between ADR phenotypes and the corresponding cellular functional and metabolic pathways that are perturbed. This underway effort is informed by the creation of human disease metabolic networks (120). Successful translational mapping can be facilitated by ontological inference and statistical inference to bridge the gaps (122).

Ontologies as the Integrating Axis

For the purposes of this review, an ontology is defined as a combination of complex relational databases that are based on semantic web technology and on artificial intelligence (the ability to reason internally). Many databases are becoming available for data mining, and advances in the science of bioinformatics enable compilation and information association across and within different levels of biological organization. These advances include the creation of interacting relationships between diseases and genes, between diseases and metabolic pathways, among proteins, among genes, between genes and proteins, between drugs and proteins, and between genes and RNAs. However, these data are produced from different experimental study designs using different organs/tissues or animal species. It is challenging to annotate and index these data to establish the framework of a defined biological or clinical domain with a common language for the purpose of computationally translating the data into a meaningful application for the systems pharmacology predictive toxicity network. For example, a cell line ontology is needed when integrating the "omics" data from different cell lines. Isolated primary cells (123) or immortalized cell lines (25) are convenient tools for investigating drug toxicity mechanistically and phenotypically. However, there are differences between any two transformed hepatocyte cell lines or between different preparations of primary human hepatocytes (PHH). PHH and transformed hepatocytes (HepG2) behave differently in their unique immune and inflammatory responses; the latter seems to exhibit downregulated immune responses (24). In response to treatment with troglitazone (PPARγ agonist) or muraglitazar (PPARα/γ agonist), HepRG is more sensitive than PHH in showing a concentration-dependent reduction in cellular ATP content (123). Extensive cell line array databases are available for mining and are being expanded continuously (see, e.g., **http://www.broadinstitute.org/genome_bio/connectivitymap.html**, **http://dtp.nci.nih.gov/branches/btb/ivclsp.html**, **http://commonfund.nih.gov/lincs/**, **http://epa.gov/ncct/Tox21/**) (**Table 1**). Databases for human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs) are in an earlier stage of development; both cell types appear to be functionally similar and to have similar signaling pathways involved in controlling differentiation (124), and they can potentially differentiate into a diverse range of cell types for drug toxicity screening. Information about both hESCs and iPSCs plus their differentiated cell types will be included in the production of in vitro omics-scale data. Integrating the data generated from various cell lines

requires a cell line ontology. Such an ontology will facilitate categorical inference and linkage of data from diverse sets of cell lines, as well as allow identification of the toxicological pathways perturbed by a drug across aggregated cell line data sets.

In addition to the challenges posed by the aggregation and interpretation of data from diverse cell lines, there are often differences in how biological and clinical studies are designed and carried out. Different ways of expressing and analyzing the study results and of describing clinical phenotypes have created further challenges for data integration and for correlating clinical phenotypes with genomics (48, 125). Integrated linkage of transcriptomics, proteomics, metabolomics, and genomics data across the hierarchical structure of the global toxicological network (**Figure 1**) requires a systematic utilization of ontologies for logical between-level inference. Ontologies of all hierarchical levels are therefore needed to facilitate large-scale integration and mining of data and to enable linkage and integrated mapping of data from various hierarchical levels. Existing ontologies that provide an excellent platform to do this work include the Gene Ontology (GO), Cell Line Ontology, Systematized Nomenclature of Medicine Clinical Terms, and National Center for Biomedical Ontology (NCBO) (see **http://www.bioontology.org/**). We are also seeing the development of ontologies and ontological tools as well as languages to facilitate data integration and mining using integrated analytics approaches (122, 126–128; see also the NCBO BioPortal at **http://bioportal.bioontology.org**). GO is an extensively used cross-species comparative toxicogenomics resource (129). Orthologous genes (genes in different species that evolved directly from a single ancestral gene) constitute the axis for integrating data from different tissues, in vitro cell lines, iPSCs, hESCs, and animal species. Ontologies and ontological tools will form the axis for integrating data from various sources to construct the global toxicological network.

CONCLUSION

There are several areas in which few data are likely to be available to construct the systems pharmacology predictive toxicity network. For example, little or no in vivo mechanistic data will be available at the patient level. As a result, probabilistic and translational linkages among many of the nodes in the network will be required. The linkages will be established mathematically and statistically by utilizing biomedical and pharmaceutical databases and biomedical as well as omics databases that contain large amounts of curated data of adequate quality. Connecting individual toxicological networks of approved drugs and failed drug candidates by integrating the network components of the hierarchy will help define and construct this global network (**Table 1**). Metabolic pathways are continuously being curated (15) and compiled into metabolic networks, and approaches will be adapted from those already in use. For example, estimation of the likelihood ratio based on the panel of test results has been proposed to facilitate diagnosis of a patient's disease (130). Integrating uncertainty in the interaction intensity (IC_{50} values) between a drug molecule and a target protein into pharmacokinetic and pharmacodynamic modeling of a drug along with its C_{max} and dose/exposure enables a probabilistic assessment of clinical efficacy of drug candidates (131). Similar approaches can be developed for probabilistic safety assessment by integrating the molecular-level data into population-level modeling and also integrating pharmacogenomics and epigenetics data to include individual variability. The global network, when constructed, will consist of edges that allow mathematical and/or statistical modeling of individual connectivity strengths and will fully incorporate in vitro, animal, and clinical data.

With ontologies as the unifying axis, integrated approaches to analyses that combine computing sciences (such as quantum or cloud computing) (132, 133), mathematics, and statistical approaches such as Bayesian networks (134) will be required to evaluate individual connectivity intensities among the large volume of genomics, transcriptomics, and phenomics data and information

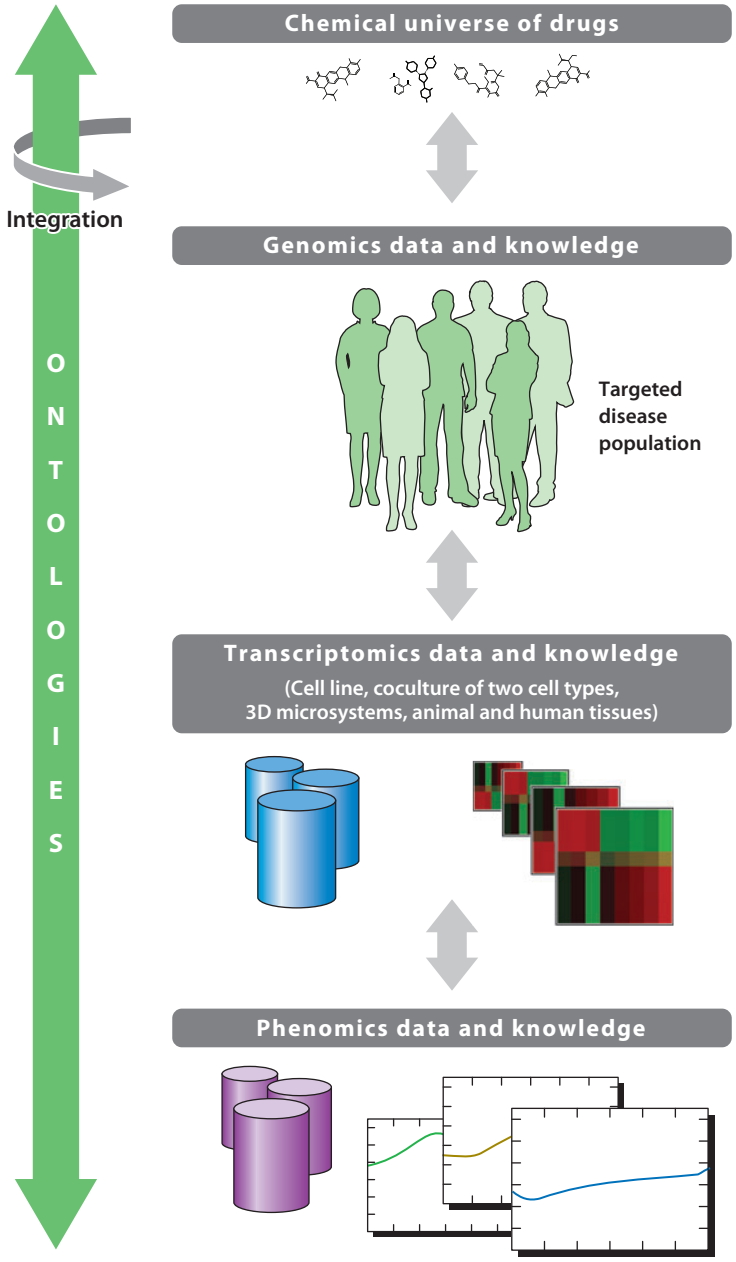


Figure 4

Schematic framework of analytic integration of genomics, transcriptomics, and phenomics.

from humans and animal species (**Figure 4**). Continuing efforts and successes with the mouse phenome database (135), with cross-species (mouse-to-human) phenotype mapping in relation to phenotype-genome association (136), with neuropsychiatric phenomics (49, 125), and with analytical methods to describe connectivity associations among phenome, genome, and transcriptome (48, 137–141) are integral to the development of the systems pharmacology predictive toxicity

network. At the population level, physiologically based human exposure-response modeling that includes genomics data may provide a means to predict rare ADRs, although that hypothesis will require careful testing. At this time, specific genetic mutations in relation to human genetic disorders (see Online Mendelian Inheritance in Man® at **http://www.ncbi.nlm.nih.gov/omim**) and drug ADME (see PharmGKB at **http://www.pharmgkb.org**) have been most reproducible and useful in establishing the connectivity between genomics and clinical phenotype. The PharmGKB database, a knowledge source containing information about drugs and related genetic mutations that affect their ADME, is useful for constructing the network between drugs and their exposure variability in individual organs in relation to drug toxicity and ADRs. Biomarkers associated with organ dysfunction are widely utilized in clinical diagnosis of diseases and determination of patients' responses to treatment and will be useful to bridge transcriptomics and phenotype.

A substantial resource for clinical response data following drug treatment of human diseases resides within individual pharmaceutical companies; there have been calls for precompetitive collaborations to enhance the availability and access to these data (142). These data, if combined, will facilitate construction of the global toxicological network. Merging toxic and therapeutic chemical spaces that have already been explored or are being explored and integrating transcriptomics, genomics, and phenomics data associated with those chemical spaces will be required to move this program forward in order to improve the sensitivity and, of equal importance, the specificity of toxicity prediction for future drug candidates (59, 143, 144). This major collaborative effort will require the active participation of the pharmaceutical industry, academic investigators and centers, research institutes such as the NIH, and the FDA and other regulatory bodies. We believe that such systems pharmacology approaches represent the next steps to advance drug safety science.

DISCLOSURE STATEMENT

The views expressed are those of the authors and do not necessarily represent the position of, or imply endorsement from, the US Food and Drug Administration or the US government.

LITERATURE CITED

1. Schadt EE, Friend SH, Shaywitz DA. 2009. A network view of disease and compound screening. *Nat. Rev. Drug Discov.* 8:286–95
2. US Food Drug Admin. 2012. *Drugs@FDA: FDA Approved Drug Products*. Silver Spring, MD: US Food Drug Admin. **http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm**
3. US Food Drug Admin. 2012. *Adverse Event Reporting System* (*AERS*). Silver Spring, MD: US Food Drug Admin. **http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm**
4. Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3:711–15
5. Haouzi D, Cohen I, Vieira HL, Poncet D, Boya P, et al. 2002. Mitochondrial permeability transition as a novel principle of hepatorenal toxicity in vivo. *Apoptosis* 7:395–405
6. Reid AB, Kurten RC, McCullough SS, Brock RW, Hinson JA. 2005. Mechanisms of acetaminophen-induced hepatotoxicity: role of oxidative stress and mitochondrial permeability transition in freshly isolated mouse hepatocytes. *J. Pharmacol. Exp. Ther.* 312:509–16
7. Bae MA, Pie JE, Song BJ. 2001. Acetaminophen induces apoptosis of C6 glioma cells by activating the c-Jun NH_2-terminal protein kinase-related cell death pathway. *Mol. Pharmacol.* 60:847–56
8. Bechmann LP, Marquitan G, Jochum C, Saner F, Gerken G, Canbay A. 2008. Apoptosis versus necrosis rate as a predictor in acute liver failure following acetaminophen intoxication compared with acute-on-chronic liver failure. *Liver Int.* 28:713–16

9. Bender A, Jenkins JL, Scheiber J, Sukuru SC, Glick M, Davies JW. 2009. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J. Chem. Inf. Model.* 49:108–19
10. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–35
11. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. 2008. Drug target identification using side-effect similarity. *Science* 321:263–66
12. Takarabe M, Shigemizu D, Kotera M, Goto S, Kanehisa M. 2010. Characterization and classification of adverse drug interactions. *Genome Inform.* 22:167–75
13. Keiser MJ, Irwin JJ, Shoichet BK. 2010. The chemical basis of pharmacology. *Biochemistry* 49:10267–76
14. Scheiber J, Chen B, Milik M, Sukuru SC, Bender A, et al. 2009. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J. Chem. Inf. Model.* 49:308–17
15. Kanehisa Lab. 2012. *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Kyoto: Kyoto Univ. **http://www.genome.jp/kegg/**
16. Daly AK. 2012. Using genome-wide association studies to identify genes important in serious adverse drug reactions. *Annu. Rev. Pharmacol. Toxicol.* 52:21–35
17. Rager JE, Smeester L, Jaspers I, Sexton KG, Fry RC. 2011. Epigenetic changes induced by air toxics: Formaldehyde exposure alters miRNA expression profiles in human lung cells. *Environ. Health Perspect.* 119:494–500
18. Shah AD, Wood DM, Dargan PI. 2011. Understanding lactic acidosis in paracetamol (acetaminophen) poisoning. *Br. J. Clin. Pharmacol.* 71:20–28
19. Fannin RD, Russo M, O'Connell TM, Gerrish K, Winnike JH, et al. 2010. Acetaminophen dosing of humans results in blood transcriptome and metabolome changes consistent with impaired oxidative phosphorylation. *Hepatology* 51:227–36
20. US Dep. Health Hum. Serv., US Food Drug Admin., Cent. Drug Eval. Res., Cent. Biol. Eval. Res. 2009. *Guidance for industry: drug-induced liver injury: premarketing clinical evaluation.* Silver Spring, MD: US Food Drug Admin. **http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf**
21. Sodhi MS, Airey DC, Lambert W, Burnet PW, Harrison PJ, Sanders-Bush E. 2005. A rapid new assay to detect RNA editing reveals antipsychotic-induced changes in serotonin-2C transcripts. *Mol. Pharmacol.* 68:711–19
22. Liu Y, Balaraman Y, Wang G, Nephew KP, Zhou FC. 2009. Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. *Epigenetics* 4:500–11
23. Xie L, Li J, Bourne PE. 2009. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* 5:e1000387
24. Alexopoulos LG, Saez-Rodriguez J, Cosgrove BD, Lauffenburger DA, Sorger PK. 2010. Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol. Cell. Proteomics* 9:1849–65
25. Antherieu S, Rogue A, Fromenty B, Guillouzo A, Robin MA. 2011. Induction of vesicular steatosis by amiodarone and tetracycline is associated with up-regulation of lipogenic genes in HepaRG cells. *Hepatology* 53:1895–905
26. Chen C, Krausz KW, Shah YM, Idle JR, Gonzalez FJ. 2009. Serum metabolomics reveals irreversible inhibition of fatty acid β-oxidation through the suppression of PPARα activation as a contributing mechanism of acetaminophen-induced hepatotoxicity. *Chem. Res. Toxicol.* 22:699–707
27. Julie NL, Julie IM, Kende AI, Wilson GL. 2008. Mitochondrial dysfunction and delayed hepatotoxicity: another lesson from troglitazone. *Diabetologia* 51:2108–16
28. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. 2010. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* 6:343
29. Hauben M, Aronson JK. 2009. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. *Drug Saf.* 32:99–110

30. Szarfman A, Machado SG, O'Neill RT. 2002. 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.* 25:381–92
31. Lempiainen H, Muller A, Brasa S, Teo SS, Roloff TC, et al. 2011. Phenobarbital mediates an epigenetic switch at the constitutive androstane receptor (CAR) target gene *Cyp2b10* in the liver of B6C3F1 mice. *PLoS ONE* 6:e18216
32. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. 2011. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* 29:1039–45
33. Cheng H, Kari G, Dicker AP, Rodeck U, Koch WJ, Force T. 2011. A novel preclinical strategy for identifying cardiotoxic kinase inhibitors and mechanisms of cardiotoxicity. *Circ. Res.* 109:1401–9
34. Padala S, Thompson PD. 2012. Statins as a possible cause of inflammatory and necrotizing myopathies. *Atherosclerosis* 222:15–21
35. Fliri AF, Loging WT, Thadeio PF, Volkmann RA. 2005. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat. Chem. Biol.* 1:389–97
36. Yamanishi Y, Kotera M, Kanehisa M, Goto S. 2010. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26:i246–54
37. She X, Rohl CA, Castle JC, Kulkarni AV, Johnson JM, Chen R. 2009. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* 10:269
38. Hansch C, Leo A. 1979. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. New York: Wiley
39. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, et al. 2009. Predicting new molecular targets for known drugs. *Nature* 462:175–81
40. Clark M, Wiseman JS. 2009. Fragment-based prediction of the clinical occurrence of long QT syndrome and torsade de pointes. *J. Chem. Inf. Model.* 49:2617–26
41. Hammann F, Gutmann H, Vogt N, Helma C, Drewe J. 2010. Prediction of adverse drug reactions using decision tree modeling. *Clin. Pharmacol. Ther.* 88:52–59
42. Rodgers AD, Zhu H, Fourches D, Rusyn I, Tropsha A. 2010. Modeling liver-related adverse effects of drugs using *k* nearest neighbor quantitative structure-activity relationship method. *Chem. Res. Toxicol.* 23:724–32
43. Cami A, Arnold A, Manzi S, Reis B. 2011. Predicting adverse drug events using pharmacological network models. *Sci. Transl. Med.* 3:114ra27
44. Bender A, Scheiber J, Glick M, Davies JW, Azzaoui K, et al. 2007. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* 2:861–73
45. Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, et al. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem. Res. Toxicol.* 24:1251–62
46. Kruhlak NL, Choi SS, Contrera JF, Weaver JL, Willard JM, et al. 2008. Development of a phospholipidosis database and predictive quantitative structure-activity relationship (QSAR) models. *Toxicol. Mech. Methods* 18:217–27
47. Freimer N, Sabatti C. 2003. The human phenome project. *Nat. Genet.* 34:15–21
48. Butte AJ, Kohane IS. 2006. Creation and implications of a phenome-genome network. *Nat. Biotechnol.* 24:55–62
49. Bilder RM. 2008. Phenomics: building scaffolds for biological hypotheses in the post-genomic era. *Biol. Psychiatry* 63:439–40
50. Scott NJ, Cameron VA, Raudsepp S, Lewis LK, Simpson ER, et al. 2012. Generation and characterization of a mouse model of the metabolic syndrome: apolipoprotein E and aromatase double knockout mice. *Am. J. Physiol. Endocrinol. Metab.* 302:E576–84
51. Harrill AH, Ross PK, Gatti DM, Threadgill DW, Rusyn I. 2009. Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. *Toxicol. Sci.* 110:235–43
52. Nakagawa H, Maeda S, Hikiba Y, Ohmae T, Shibata W, et al. 2008. Deletion of apoptosis signal-regulating kinase 1 attenuates acetaminophen-induced liver injury by inhibiting c-Jun N-terminal kinase activation. *Gastroenterology* 135:1311–21

53. Harrill AH, Watkins PB, Su S, Ross PK, Harbourt DE, et al. 2009. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res.* 19:1507–15
54. Cai S, Wang H, Bailey B, Ernstberger A, Juliar BE, et al. 2011. Humanized bone marrow mouse model as a preclinical tool to assess therapy-mediated hematotoxicity. *Clin. Cancer Res.* 17:2195–206
55. Chen AA, Thomas DK, Ong LL, Schwartz RE, Golub TR, Bhatia SN. 2011. Humanized mice with ectopic artificial liver tissues. *Proc. Natl. Acad. Sci. USA* 108:11842–47
56. Strom SC, Davila J, Grompe M. 2010. Chimeric mice with humanized liver: tools for the study of drug metabolism, excretion, and toxicity. *Methods Mol. Biol.* 640:491–509
57. DiBlasio-Smith EA, Arai M, Quinet EM, Evans MJ, Kornaga T, et al. 2008. Discovery and implementation of transcriptional biomarkers of synthetic LXR agonists in peripheral blood cells. *J. Transl. Med.* 6:59
58. Shukla SJ, Huang R, Austin CP, Xia M. 2010. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discov. Today* 15:997–1007
59. Uehara T, Ono A, Maruyama T, Kato I, Yamada H, et al. 2010. The Japanese Toxicogenomics Project: application of toxicogenomics. *Mol. Nutr. Food Res.* 54:218–27
60. Bartfai T, Buckley PT, Eberwine J. 2012. Drug targets: single-cell transcriptomics hastens unbiased discovery. *Trends Pharmacol. Sci.* 33:9–16
61. Yilmaz S, Boffito M, Collot-Teixeira S, De Lorenzo F, Waters L, et al. 2010. Investigation of low-dose ritonavir on human peripheral blood mononuclear cells using gene expression whole genome microarrays. *Genomics* 96:57–65
62. Erler JT, Linding R. 2010. Network-based drugs and biomarkers. *J. Pathol.* 220:290–96
63. Zhao Y, Widen SG, Jamaluddin M, Tian B, Wood TG, et al. 2011. Quantification of activated NF-κB/RelA complexes using ssDNA aptamer affinity-stable isotope dilution-selected reaction monitoring-mass spectrometry. *Mol. Cell. Proteomics* 10:M111.008771
64. Mrozek E, Kolesar J, Young D, Allen J, Villalona-Calero M, Shapiro CL. 2008. Phase II study of sequentially administered low-dose mitomycin-C (MMC) and irinotecan (CPT-11) in women with metastatic breast cancer (MBC). *Ann. Oncol.* 19:1417–22
65. Liu G, Kolesar J, McNeel DG, Leith C, Schell K, et al. 2008. A phase I pharmacokinetic and pharmacodynamic correlative study of the antisense Bcl-2 oligonucleotide g3139, in combination with carboplatin and paclitaxel, in patients with advanced solid tumors. *Clin. Cancer Res.* 14:2732–39
66. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, et al. 2011. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3:96ra77
67. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, et al. 2011. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3:96ra76
68. Bao N, Lye KW, Barton MK. 2004. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev. Cell* 7:653–62
69. Tuddenham L, Wheeler G, Ntounia-Fousara S, Waters J, Hajihosseini MK, et al. 2006. The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells. *FEBS Lett.* 580:4214–17
70. Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, et al. 2006. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9:435–43
71. Jetten MJ, Gaj S, Ruiz-Aracama A, de Kok TM, van Delft JH, et al. 2012. 'Omics analysis of low dose acetaminophen intake demonstrates novel response pathways in humans. *Toxicol. Appl. Pharmacol.* 259(3):320–28
72. Fukushima T, Hamada Y, Yamada H, Horii I. 2007. Changes of micro-RNA expression in rat liver treated by acetaminophen or carbon tetrachloride–regulating role of micro-RNA for RNA expression. *J. Toxicol. Sci.* 32:401–9
73. Lizarraga D, Gaj S, Brauers KJ, Timmermans L, Kleinjans JC, van Delft JH. 2012. Benzo[a]pyrene-induced changes in microRNA-mRNA networks. *Chem. Res. Toxicol.* 25:838–49
74. Wilson KD, Hu S, Venkatasubrahmanyam S, Fu JD, Sun N, et al. 2010. Dynamic microRNA expression programs during cardiac differentiation of human embryonic stem cells: role for miR-499. *Circ. Cardiovasc. Genet.* 3:426–35

75. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* 22:1001–5
76. Niswender CM, Herrick-Davis K, Dilley GE, Meltzer HY, Overholser JC, et al. 2001. RNA editing of the human serotonin 5-HT_{2C} receptor: alterations in suicide and implications for serotonergic pharmacotherapy. *Neuropsychopharmacology:* 24:478–91
77. Bae MA, Song BJ. 2003. Critical role of c-Jun N-terminal protein kinase activation in troglitazone-induced apoptosis of human HepG2 hepatoma cells. *Mol. Pharmacol.* 63:401–8
78. Tyagi M, Hashimoto K, Shoemaker BA, Wuchty S, Panchenko AR. 2012. Large-scale mapping of human protein interactome using structural complexes. *EMBO Rep.* 13:266–71
79. Fliri AF, Loging WT, Volkmann RA. 2009. Drug effects viewed from a signal transduction network perspective. *J. Med. Chem.* 52:8038–46
80. Sarkar IN, Butte AJ, Lussier YA, Tarczy-Hornoch P, Ohno-Machado L. 2011. Translational bioinformatics: linking knowledge across biological and clinical realms. *J. Am. Med. Inform. Assoc.* 18:354–57
81. Dudley JT, Schadt E, Sirota M, Butte AJ, Ashley E. 2010. Drug discovery in a multidimensional world: systems, patterns, and networks. *J. Cardiovasc. Transl. Res.* 3:438–47
82. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et al. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* 104:1777–82
83. Cosgrove BD, Alexopoulos LG, Hang TC, Hendriks BS, Sorger PK, et al. 2010. Cytokine-associated drug toxicity in human hepatocytes is associated with signaling network dysregulation. *Mol. BioSyst.* 6:1195–206
84. Luo H, Chen J, Shi L, Mikailov M, Zhu H, et al. 2011. DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. *Nucleic Acids Res.* 39:W492–98
85. Yang L, Chen J, He L. 2009. Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. *PLoS Comput. Biol.* 5:e1000441
86. Yang L, Wang K, Chen J, Jegga AG, Luo H, et al. 2011. Exploring off-targets and off-systems for adverse drug reactions via chemical-protein interactome–clozapine-induced agranulocytosis as a case study. *PLoS Comput. Biol.* 7:e1002016
87. Wallach I, Jaitly N, Lilien R. 2010. A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways. *PLoS ONE* 5:e12063
88. Yang X, Huang Y, Crowson M, Li J, Maitland ML, Lussier YA. 2010. Kinase inhibition-related adverse events predicted from in vitro kinome and clinical trial data. *J. Biomed. Inform.* 43:376–84
89. Berger SI, Ma'ayan A, Iyengar R. 2010. Systems pharmacology of arrhythmias. *Sci. Signal.* 3:ra30
90. Moro S, Chipman JK, Antczak P, Turan N, Dekant W, et al. 2012. Identification and pathway mapping of furan target proteins reveal mitochondrial energy production and redox regulation as critical targets of furan toxicity. *Toxicol. Sci.* 126:336–52
91. Fulgencio JP, Kohl C, Girard J, Pegorier JP. 1996. Troglitazone inhibits fatty acid oxidation and esterification, and gluconeogenesis in isolated hepatocytes from starved rats. *Diabetes* 45:1556–62
92. Inoue I, Takahashi K, Katayama S, Harada Y, Negishi K, et al. 1995. Effect of troglitazone (CS-045) and bezafibrate on glucose tolerance, liver glycogen synthase activity, and β-oxidation in fructose-fed rats. *Metabolism* 44:1626–30
93. Wagner BK, Kitami T, Gilbert TJ, Peck D, Ramanathan A, et al. 2008. Large-scale chemical dissection of mitochondrial function. *Nat. Biotechnol.* 26:343–51
94. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–65
95. Edling Y, Sivertsson LK, Butura A, Ingelman-Sundberg M, Ek M. 2009. Increased sensitivity for troglitazone-induced cytotoxicity using a human in vitro co-culture model. *Toxicol. In Vitro* 23:1387–95
96. Heydarkhan-Hagvall S, Helenius G, Johansson BR, Li JY, Mattsson E, Risberg B. 2003. Co-culture of endothelial cells and smooth muscle cells affects gene expression of angiogenic factors. *J. Cell. Biochem.* 89:1250–59
97. LaFramboise WA, Scalise D, Stoodley P, Graner SR, Guthrie RD, et al. 2007. Cardiac fibroblasts influence cardiomyocyte phenotype in vitro. *Am. J. Physiol. Cell Physiol.* 292:C1799–808

98. Prot JM, Briffaut AS, Letourneur F, Chafey P, Merlier F, et al. 2011. Integrated proteomic and transcriptomic investigation of the acetaminophen toxicity in liver microfluidic biochip. *PLoS ONE* 6:e21268
99. Khetani SR, Bhatia SN. 2008. Microscale culture of human liver cells for drug development. *Nat. Biotechnol.* 26:120–26
100. Shintu L, Baudoin R, Navratil V, Prot JM, Pontoizeau C, et al. 2012. Metabolomics-on-a-chip and predictive systems toxicology in microfluidic bioartificial organs. *Anal. Chem.* 84:1840–48
101. Huh D, Matthews BD, Mammoto A, Montoya-Zavala M, Hsin HY, Ingber DE. 2010. Reconstituting organ-level lung functions on a chip. *Science* 328:1662–68
102. Santostefano MJ, Richardson VM, Walker NJ, Blanton J, Lindros KO, et al. 1999. Dose-dependent localization of TCDD in isolated centrilobular and periportal hepatocytes. *Toxicol. Sci.* 52:9–19
103. Zordoky BN, Anwar-Mohamed A, Aboutabl ME, El-Kadi AO. 2010. Acute doxorubicin cardiotoxicity alters cardiac cytochrome P450 expression and arachidonic acid metabolism in rats. *Toxicol. Appl. Pharmacol.* 242:38–46
104. Vickers AE. 2009. Tissue slices for the evaluation of metabolism-based toxicity with the example of diclofenac. *Chem. Biol. Interact.* 179:9–16
105. Zordoky BN, Anwar-Mohamed A, Aboutabl ME, El-Kadi AO. 2011. Acute doxorubicin toxicity differentially alters cytochrome P450 expression and arachidonic acid metabolism in rat kidney and liver. *Drug Metab. Dispos.* 39:1440–50
106. Herman EH, Zhang J, Lipshultz SE, Rifai N, Chadwick D, et al. 1999. Correlation between serum levels of cardiac troponin-T and the severity of the chronic cardiomyopathy induced by doxorubicin. *J. Clin. Oncol.* 17:2237–43
107. von Haehling S, Jankowska EA, Morgenthaler NG, Vassanelli C, Zanolla L, et al. 2007. Comparison of midregional pro-atrial natriuretic peptide with N-terminal pro-B-type natriuretic peptide in predicting survival in patients with chronic heart failure. *J. Am. Coll. Cardiol.* 50:1973–80
108. Grams ME, Rabb H. 2012. The distant organ effects of acute kidney injury. *Kidney Int.* 81:942–48
109. Astor BC, Shafi T, Hoogeveen RC, Matsushita K, Ballantyne CM, et al. 2012. Novel markers of kidney function as predictors of ESRD, cardiovascular disease, and mortality in the general population. *Am. J. Kidney Dis.* 59:653–62
110. Merrill GF, Kurth EJ, Hardie DG, Winder WW. 1997. AICA riboside increases AMP-activated protein kinase, fatty acid oxidation, and glucose uptake in rat muscle. *Am. J. Physiol. Endocrinol. Metab.* 273:E1107–12
111. Song Z, Deaciuc I, Zhou Z, Song M, Chen T, et al. 2007. Involvement of AMP-activated protein kinase in beneficial effects of betaine on high-sucrose diet-induced hepatic steatosis. *Am. J. Physiol. Gastrointest. Liver Physiol.* 293:G894–902
112. Decleves AE, Mathew AV, Cunard R, Sharma K. 2011. AMPK mediates the initiation of kidney disease induced by a high-fat diet. *J. Am. Soc. Nephrol.* 22:1846–55
113. Zoccali C, Mallamaci F, Panuccio V, Tripepi G, Cutrupi S, et al. 2003. Adiponectin is markedly increased in patients with nephrotic syndrome and is related to metabolic risk factors. *Kidney Int. Suppl.* 84:S98–102
114. Tujios S, Fontana RJ. 2011. Mechanisms of drug-induced liver injury: from bedside to bench. *Nat. Rev. Gastroenterol. Hepatol.* 8:202–11
115. Int. Conf. Harmon. 2005. *E2B(R) clinical safety data management: data elements for transmission of individual case safety reports, revision 2*. Silver Spring, MD: US Food Drug Admin. **http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm129399.pdf**
116. Funk C, Ponelle C, Scheuermann G, Pantze M. 2001. Cholestatic potential of troglitazone as a possible factor contributing to troglitazone-induced hepatotoxicity: in vivo and in vitro interaction at the canalicular bile salt export pump (Bsep) in the rat. *Mol. Pharmacol.* 59:627–35
117. Gunawan BK, Liu ZX, Han D, Hanawa N, Gaarde WA, Kaplowitz N. 2006. c-Jun N-terminal kinase plays a major role in murine acetaminophen hepatotoxicity. *Gastroenterology* 131:165–78
118. Mahadevan SB, McKiernan PJ, Davies P, Kelly DA. 2006. Paracetamol induced hepatotoxicity. *Arch. Dis. Child.* 91:598–603
119. Russmann S, Kaye JA, Jick SS, Jick H. 2005. Risk of cholestatic liver disease associated with flucloxacillin and flucloxacillin prescribing habits in the UK: cohort study using data from the UK General Practice Research Database. *Br. J. Clin. Pharmacol.* 60:76–82

120. Li X, Li C, Shang D, Li J, Han J, et al. 2011. The implications of relationships between human diseases and metabolic subpathways. *PLoS ONE* 6:e21131
121. Chang RL, Xie L, Bourne PE, Palsson BO. 2010. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* 6:e1000938
122. Jupp S, Klein J, Schanstra J, Stevens R. 2011. Developing a kidney and urinary pathway knowledge base. *J. Biomed. Semantics* 2(Suppl. 2):S7
123. Rogue A, Lambert C, Josse R, Antherieu S, Spire C, et al. 2011. Comparative gene expression profiles induced by PPARγ and PPARα/γ agonists in human hepatocytes. *PLoS ONE* 6:e18816
124. Vallier L, Touboul T, Brown S, Cho C, Bilican B, et al. 2009. Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem Cells* 27:2655–66
125. Bilder RM, Sabb FW, Cannon TD, London ED, Jentsch JD, et al. 2009. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience* 164:30–42
126. Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R, Musen MA. 2009. Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinform.* 10(Suppl. 2):S1
127. Shah NH, Bhatia N, Jonquet C, Rubin D, Chiang AP, Musen MA. 2009. Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinform.* 10(Suppl. 9):S14
128. Hur J, Xiang Z, Feldman EL, He Y. 2011. Ontology-based *Brucella* vaccine literature indexing and systematic analysis of gene-vaccine association network. *BMC Immunol.* 12:49
129. Mattingly CJ, Rosenstein MC, Davis AP, Colby GT, Forrest JN Jr, Boyer JL. 2006. The comparative toxicogenomics database: a cross-species resource for building chemical-gene interaction networks. *Toxicol. Sci.* 92:587–95
130. Morgan AA, Chen R, Butte AJ. 2010. Likelihood ratios for genome medicine. *Genome Med.* 2:30
131. Mudd PN Jr, Groenendaal H, Bush MA, Schmith VD. 2010. Probabilistic risk analysis: improving early drug development decision making. *Clin. Pharmacol. Ther.* 88:871–75
132. Barz S, Kashefi E, Broadbent A, Fitzsimons JF, Zeilinger A, Walther P. 2012. Demonstration of blind quantum computing. *Science* 335:303–8
133. Chouvarine P, Cooksey AM, McCarthy FM, Ray DA, Baldwin BS, et al. 2012. Transcriptome-based differentiation of closely-related *Miscanthus* lines. *PLoS ONE* 7:e29850
134. Gopalakrishnan V, Lustgarten JL, Visweswaran S, Cooper GF. 2010. Bayesian rule learning for biomedical data mining. *Bioinformatics* 26:668–75
135. Maddatu TP, Grubb SC, Bult CJ, Bogue MA. 2012. Mouse Phenome Database (MPD). *Nucleic Acids Res.* 40:D887–94
136. Sardana D, Vasa S, Vepachedu N, Chen J, Gudivada RC, et al. 2010. PhenoHM: human-mouse comparative phenome-genome server. *Nucleic Acids Res.* 38:W165–74
137. Curtis RE, Yin J, Kinnaird P, Xing EP. 2012. Finding genome-transcriptome-phenome association with structured association mapping and visualization in GenAMap. *Pac. Symp. Biocomput.* 2012:327–38
138. Gormanns P, Mueller NS, Ditzen C, Wolf S, Holsboer F, Turck CW. 2011. Phenome-transcriptome correlation unravels anxiety and depression related pathways. *J. Psychiatr. Res.* 45:973–79
139. Hoehndorf R, Schofield PN, Gkoutos GV. 2011. PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Res.* 39:e119
140. Avery CL, He Q, North KE, Ambite JL, Boerwinkle E, et al. 2011. A phenomics-based strategy identifies loci on *APOC1*, *BRAP*, and *PLCG1* associated with metabolic syndrome phenotype domains. *PLoS Genet.* 7:e1002322
141. Hoehndorf R, Oellrich A, Rebholz-Schuhmann D, Schofield PN, Gkoutos GV. 2012. Linking PharmGKB to phenotype studies and animal models of disease for drug repurposing. *Pac. Symp. Biocomput.* 2012:388–99
142. Friend SH. 2010. The need for precompetitive integrative bionetwork disease model building. *Clin. Pharmacol. Ther.* 87:536–39
143. McBurney RN, Hines WM, Von Tungeln LS, Schnackenberg LK, Beger RD, et al. 2009. The liver toxicity biomarker study: phase I design and preliminary results. *Toxicol. Pathol.* 37:52–64
144. Suter L, Schroeder S, Meyer K, Gautier JC, Amberg A, et al. 2011. EU framework 6 project: Predictive Toxicology (PredTox)—overview and outcome. *Toxicol. Appl. Pharmacol.* 252:73–84

Annual Review of Pharmacology and Toxicology

Volume 53, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml