Expert Opinion

Ashley Publications
www.ashley-pub.com

Theragenomic knowledge management for individualised safety of drugs, chemicals, pollutants and dietary ingredients

Joseph Gut† & Dario Bagatto
†*TheraSTrat AG, Gewerbestrasse 25, 4123 Allschwil, Switzerland*

Severe adverse drug reactions (ADRs) are a major problem in drug development and clinical practice and the most common cause of market withdrawals of drugs. Individualised drug safety aims at the prospective identification of single patients who carry genetic predispositions for the development of serious or fatal ADRs under drug treatment. For a comprehensive individualised drug safety evaluation a clearly structured organisation, linkage and representation of diverse and heterogeneous, yet related, knowledge is imperative. To efficiently support experts in this process a platform, coined OKAPI, was specified that combines multiple concepts in knowledge management to perform ontological knowledge acquisition, processing and integration. SafeBase™ (TheraSTrat) is an ontology driven implementation of the OKAPI specification and an innovative, user-friendly and interactive platform for storage, management and visualisation of knowledge across multiple scientific disciplines pertinent to current and future theragenomics-based drug discovery and development and to strategies of individualised drug safety.

Keywords: adverse drug reaction, individualised drug safety, knowledge management, personalised medicine, pharmacogenetic, predisposition, toxicogenetic

Expert Opin. Drug Metab. Toxicol. (2005) 1(3):537-554

1. Introduction

Severe adverse drug reactions (ADRs), including, but not restricted to, idiosyncratic drug reactions, which occur unpredictably and at low frequency on exposure of human individuals to drugs (and many chemicals beyond drugs) are a major problem in drug development, clinical practice and the assessment and regulation of the safety of drugs and chemicals. Meta-analyses in the past have suggested that in the US in 1994, ADRs were responsible for > 100,000 deaths, making them between the fourth and sixth leading cause of death [1]. Although these numbers were criticised [2], they nevertheless emphasise the importance of ADRs. Indeed, there is good evidence that ADRs account for up to 5% of all hospital admissions [3] and increase the length of stay in the hospital at the expense of increased cost of ~ $2500/patient [4,5].

ADRs are one of the most common causes of drug withdrawals and bans of drug use in postmarketing phases (Table 1), which lead to substantial registration delays when at the end of the clinical development (Phase III) not all safety issues are resolved (for example, omalizumab, tegaresod, lumiracoxib and ximelagatran), and often lead to the discontinuation or temporary suspension of further clinical development (as, for example, in the cases of etomixir, raglitazar, repaglinide and tysabri). Each of these scenarios has enormous financial implications for affected pharmaceutical companies.

The situation of frequent discontinuation in clinical development of promising new drug candidates and of market withdrawals of drugs that have been both effective and safe in the vast majority of patients is increasingly creating an unmet medical need, namely to develop integrated therapeutic strategies in order to keep

Table 1. Recent examples of drugs withdrawn from the market because of subpopulation (idiosyncratic) forms of severe adverse drug reactions. Withdrawals may have been worldwide or restricted to selected markets (not specified here).

Drug	Adverse drug reaction	Company	Withdrawal
Valdecoxib	Cardiovascular disease	Pfizer	April 2005
Tysabri*	Progressive multifocal leucoencephalopathy	Biogen Idec/Elan	February 2005
Rofecoxib	Cardiovascular disease	Merck	September 2004
Serzone	Liver failure	Bristol-Meyer Squibb	June 2004
Epdedra	Haemorrhagic stroke	Several‡	April 2004
KavaKava	Liver failure	Several§	January 2003
Cerivastatin	Rhabdomyolysis	Bayer	August 2001
Rapacuronium	Bronchospasm	Organon	March 2001
Alosetron	Ischaemic colitis	GlaxoSmithKline	November 2000
Cisapride	Cardiac arrhythmia	Jansen Pharmaceuticals	June 2000
Troglitazone	Liver toxicity	Warner-Lambert	May 2000
Grepafloxacin	Prolonged QT intervals	Glaxo	October 1999
Pemoline	Liver toxicity	Abbott	September 1999
Astemizole	Cardiac arrhythmia	Jansen Pharmaceuticals	June 1999
Bromfenac	Liver toxicity	Wyeth	June 1998
Mibefradil	Drug–drug interaction	Roche	June 1998
Terfenadine	Cardiac arrhythmia	Hoechst-Marion Roussel	December 1997
Dexfenfluramine	Heart valve disease	Interneuron	September 1997
Fenfluramine	Heart valve disease	Wyeth	September 1997

*Market suspension; ‡Mostly suppliers of dietary supplements; §Mostly suppliers of herbal foods and herbal medicinal products.

successful drugs on the market, even though there exist a few individuals within the targeted patient population who, based on their genetic predisposition, cannot tolerate a given drug.

2. Adverse drug reactions

ADRs can be classified in terms of their clinical, pharmacological and chemical characteristics [6,7]. Thus, from a clinical point of view, ADRs comprise type A and B reactions. Type A, or augmented, reactions can be predicted from the known pharmacology, and often represent an exaggeration of the pharmacological effects of the drug. These reactions are usually dose-dependent and may be reversed by dose reduction. Type B, or bizarre (idiosyncratic), reactions cannot be predicted from a knowledge of the basic pharmacology of the drug and have no simple dose–response relationship, that is, there is a lack of correlation between dose and risk of toxicity. Host-dependent factors seem to be important in predisposition to these reactions, and they are thought to have both metabolic and immunological components, which may determine individual susceptibility. These reactions tend to be serious and account for many drug-induced deaths. From a chemical (i.e., structural) point of view, ADRs may comprise type C and D reactions. Type C, or chemical, reactions are those reactions whose biological characteristics can be either rationalised or even predicted based on the chemical structure of the parent drug, or of reactive intermediates and metabolites thereof. Type D or delayed reactions may occur many years after treatment such as, for example, second tumours years after treatment with chemotherapeutic agents. Also included in this category of ADRs are teratogenic effects seen in children after drug intake by the mother during pregnancy.

Serious, but rare, ADRs (including idiosyncratic reactions) are usually detected only once a given drug has been used widely in a large patient population in the postmarketing phase of the drug's life cycle. Serious ADRs do occur, despite extensive preclinical evaluation of a given new chemical entity (i.e., drug candidate) in laboratory animals and large numbers of patients enrolled in clinical trials in order to evaluate the efficacy and safety of this drug candidate in humans. Several reasons for this phenomenon may exist, including the fact that large clinical trials, exposing perhaps up to 10,000 patients to the new drug prior to registration, may not suffice to reliably detect rare ADRs. Complicating the understanding of rare ADRs is the occurrence of confounding factors, for example, pre-existing disease, diet, occupational and/or social exposure to a wide variety of compounds, and so on, all of which may contribute to the predisposition of a single individual for the development

Table 2. Selected examples of clinically important pharmacogenetic polymorphisms in genes that influence drug metabolism and safety.

Gene	Frequency of major polymorphism	Drug(s)	Adverse drug reaction
CYP2C9	14 – 28% (heterozygotes) 0.2 – 1% (homozygotes)	Warfarin Tolbutamide Phenytoin Losartan	Haemorrhage Hypoglycaemia Phenytoin toxicity Decreased antihypertensive effect
CYP2D6	5 – 10% (poor metabolisers) 0.1 – 10% (ultrarapid metabolisers)	Antiarrhythmics Antidepressants Antipsychotics Opoids	Proarrhythmic and other toxic affects Inefficacy in ultrarapid metabolisers; toxicity in poor metabolisers Tardive dyskinaesia Inefficacy of codeine as analgesic; narcotic side effects; dependency
CYP2C19	3 – 6% (Caucasians) 10 – 20% (Asians)	Omeprazole Diazepam	Higher cure rates when given with clarithromycin Prolonged sedation
DPD	0.1%	5-Fluorouracil	Myelotoxicity Neurotoxicity
TPMT	0.3%	Mercaptopurine Thioguanine Azathioprine	Myelotoxicity
NAT2	40 – 70% (Caucasians) 10 – 20% (Asians)	Sulfonamides Amonafide	Hypersensitivity Myelotoxicity (in rapid acetylators)

CYP: Cytochrome P450; DPD: Dihydropyrimidine dehydrogenase; NAT: *N*-acetyltransferase; TPMT: Thiopurine methyltransferase.

of a serious ADR. The exposure of an organism to a drug or other chemical (xenobiotic and/or endobiotic) induces complex interactions between the molecular structure of the compound and genes, receptors, transporters, metabolising enzymes, and so on, with various contributions to a beneficial or adverse response towards the compound, and most of these interactions may depend on the genetic outfit of single individuals; some of the deleterious functional consequences may depend on genetic constellations (i.e., alleles, haplotypes) that occur very rarely in the population.

3. Genetic and structural predispositions for adverse drug reactions

Several new concepts of pharmacogenetics/-genomics (PGx) and toxicogenetics/-genomics (TGx) are emerging, with the hope to alleviate the dilemma posed by the rare, seemingly unpredictable, severe ADRs in patient populations. For the sake of simplicity, in the present report, the collective term 'theragenomics' is used, in order to refer to the combination of PGx/TGx, which aims at providing safer and better therapies to patients, including those with predisposition for ADRs.

Among the best-studied causes of individual differences in human response to drugs and other chemicals are differences in the activities of xenobiotic-metabolising enzymes, caused either by genetic or environmental factors [8,9]. Genetic polymorphisms of these enzymes have been shown to cause frequent interindividual variation in the ability to metabolise drugs and chemicals to either inactive (detoxication) or reactive metabolites (toxication). Prototypic and very well-characterised pharmacogenetic polymorphisms are those associated with the cytochrome P450 (CYP) family of drug-metabolising enzymes. In fact, one of the most extensively studied enzymes is CYP2D6, which plays a role in the metabolism of ~ 25 – 30% of all prescribed drugs. The rate of metabolism of a given drug whose metabolism is catalysed in a rate-limiting fashion and to a considerable (> 30%) proportion through CYP2D6 may vary by a factor of 100 between 'extensive' and 'poor' metabolisers. It is interesting to note that Roche Diagnostics has developed the first CYP450 microarray test, coined AmpliChip CYP450 Test, which searches for 29 polymorphisms and mutations for the CYP2D6 gene and two polymorphisms for the CYP2C19 gene, and runs on a chip developed by Affymetrix. The test is already widely used in clinical trials to help select the appropriate patients or healthy volunteers, and has very recently received FDA approval [101] as an *in vitro* diagnostic tool.

Such tests may prove useful in clinical practice also. An example could be atomoxetine, which is a selective noradrenaline reuptake inhibitor and was approved by the FDA for attention deficit hyperactivity disorder in 2002. Atomoxetine is primarily metabolised through the CYP2D6 enzymatic pathway [10]. CYP2D6 poor metabolisers have a 10-fold higher area under the curve (AUC) and a 5-fold higher peak concentration to a given dose of atomoxetine compared with extensive metabolisers. The blood levels in poor metabolisers are similar to those attained by taking strong inhibitors of CYP2D6. The higher blood levels in poor metabolisers may lead to a higher rate of some adverse effects of atomoxetine,

Table 3. Selected examples of clinically important toxicogenetic traits.

Gene	Compound(s)	Adverse drug effect
ALDH2	Cyclophosphamide Vinyl chloride	SCE-frequency in lymphocytes
DIA4	Ubiquinones Menadione Mitomycin C	Menadione-associated urolithiasis
KCNH2	Quinidine Cisapride	Drug-induced long QT syndrome Drug-induced *Torsade de Pointes*
KCNQ1	Terfenadine disopyramide meflaquine	Drug-induced long QT syndrome
hKCNE2	Clarithromycin	Drug-induced arrhythmia
SCN5A	Mexiletine	Efficacy for long QT syndrome secondary to SCN5A mutations, but not to hERG mutations
RYR1	Halothane succinylcholine	Drug-induced malignant hyperthermia
UGTA1	Irinotecan	Diarrhoea, myelosuppression
UGTA1	Tranilast	Hyperbilirubinea
*HLA-B*5701*, *HLA-DR7*, *HLA-DQ3*	Abacavir	Hypersensitivity reactions*

*Example of the presence of a haplotype, which shows very high association of drug exposure with adverse clinical effect [24].
ALDH: Aldehyde dehydrogenase; DIA4: NAD(P)H dehydrogenase (quinone); hKCNE2: Potassium voltage-gated channel subfamily E member 2; HLA: Human leukocyte antigen; KCNH2; Potassium voltage-gated channel, subfamily H (eag-related), member 2; KCNQ1: Potassium voltage-gated channel, KQT-like subfamily, member 1; RYR1: Ryanodine receptor 1 (skeletal muscle-type ryanodine receptor); SCE: Sister chromatid exchange; SCN5A: Sodium channel, voltage-gated, type V, α (long QT syndrome 3); UGT: Uridine 5′-diphosphate glucuronosyltransferase.

some of which may be serious forms of liver toxicity [102]. Although in its approval the FDA did not request a restriction of the drug to certain phenotypes, because clinical safety signals were not definitive enough to warrant genotype-guided dosing, the prescription information accompanying atomoxetine warns expressly and repeatedly of the polymorphisms' impact on the metabolism of the drug, and mentions that laboratory tests are available to identify CYP2D6 poor metabolisers and propose dosage modifications.

More recently, genetic polymorphisms of drug receptors and transporters have been associated with ADRs. These polymorphisms usually were discovered by the occurrence of ADRs in some patients, followed by family studies, identification of the target enzyme and its gene, and demonstration that the mutations in the gene cause the adverse reaction phenotype. As discussed above, from a clinical perspective, ADRs may be classified as augmented reactions (type A), which are predictable from the pharmacodynamics (PD) and/or pharmacokinetics (PK) of the drug or chemical in questions. Pharmacogenetic (Table 2) and to some extent toxicogenetic (Table 3) traits have been identified that form the basis for type A adverse reactions, mostly because they affect rate-limiting processes, which are dominated by one or very few genes, and which play a predominant role in the clearance of drugs with a narrow therapeutic window.

In contrast, many of the type B (idiosyncratic), C (chemical) and D (delayed) ADRs, which are unpredictable from the knowledge of the basic PD/PK parameters and show marked individual susceptibility and no simple dose-dependency, are mediated by toxic metabolites and interactions of these metabolites with cellular targets. Consequently, complex processes beyond drug metabolism are expected to contribute to individual susceptibility [11-13]. This means that a single trait that is associated with an ADR might constitute a risk factor, but might neither be necessary nor sufficient to produce the adverse reaction by itself. In fact, very complex patterns of chemical (i.e., reactive intermediates, reactive metabolites, etc.) risk factors combined with pharmaco- and toxicogenetic traits are expected to form the molecular basis for the type B (idiosyncratic), C (chemical) and D (delayed) ADRs. Some recently identified examples of such traits may include the TNF-α promoter region G308A polymorphism associated with serious hypersensitivity reactions to carbamazepine [14], and the R576A polymorphism of the α-subunit of the IL-4 receptor, which is strongly associated with atopy [15]. Moreover, animal models for acetaminophen hepatotoxicity such as the Nrf2 knockout mouse [16,17] or mice transgenic for a dominant-negative human K18 mutant [18] may help to identify genetic traits that in addition to traits in drug or xenobiotic metabolism contribute significantly to the predisposition for ADRs.

In clinical practice, besides polymorphisms of the CYP2D6 gene, the identification of polymorphisms of at least two additional genes has become increasingly important in determining the predisposition of carriers of selected alleles for the occurrence of untoward drug effects. Thus, genotyping for the rare variant ABCB1 1236C > T, as well as for other rare genetic variants of the uridine 5′-diphosphate glucuronosyltransferase (UGT)1A1 gene [19], may be factors

assisting with dose optimisation of irinotecan chemotherapy in cancer patients and avoidance of severe toxicity (i.e., neutropoenia, myelotoxicity) in carriers of these alleles. Additional investigation is required to confirm these findings in a larger population and to assess relationships between irinotecan disposition and the rare variant genotypes, especially in other ethnic groups. Moreover, recently, a DNA repeat polymorphism of the UGT1A1 gene, which has previously been identified as an aetiological factor for Gilbert's disease [20], was identified as the susceptibility locus for hyperbilirubinaemia in patients exposed to the antirestenosis drug tranilast [21].

Similarly, thiopurine methyltransferase (TPMT) catalyses the *S*-methylation of thiopurines. TPMT exhibits genetic polymorphism(s), and TPMT-deficient patients treated with standard doses of thiopurines accumulate excessive thioguanine nucleotides in haematopoietic tissues, leading to severe, potentially fatal haematological toxicity. TPMT-genotyping is increasingly used to identify the carriers of the offending allelic variants in order to adjust doses accordingly or to exclude altogether the carriers of said alleles from therapies with thiopurines [22]. To that end, Prometheus Laboratories proposes its PRO-PredictRxTPMT genetic test to help the physician to tailor drug therapy and lower doses for TPMT-deficient or intermediate patients [103].

Genetic factors influencing drug metabolism and the immune response, including the human leukocyte antigen (HLA) genotype, might be involved in drug hypersensitivity reactions, including carbamazepine-induced Stevens–Johnson syndrome (CBZ-SJS). Interestingly, genotyping 157 CYP single-nucleotide polymorphisms (SNPs) revealed no significant association between any of the CYP SNPs and occurrence of CBZ-SJS [23]. However, genotyping the same patients for all *HLA-B*, *-C*, *-A* and *-DRB1* alleles indicated that the alleles *B*1502*, *Cw*0801*, *A*1101* and *DRB1*1202* within the *HLA* region occurred at increased frequency in CBZ-SJS patients relative to the control patients. In particular, *HLA-B*1502* was present in 100% (44/44) of CBZ-SJS patients, but in only 3% (3/101) of CBZ-tolerant patients and in 8.6% (8/93) of the general population [23]. When the CBZ-tolerant group was used as the control, the presence of *B*1502* achieved a 93.6% positive-prediction value for CBZ-SJS, whereas its absence has a negative-prediction value of 100%. In a test for CBZ-SJS, the *HLA-B*1502* allele should, therefore, have 100% sensitivity and 97% specificity. Further analysis revealed an ancestral haplotype, that is, *B*1502*, *Cw*0801*, *A*1101* and *DRB1*1202*, which was present in 66% of the CBZ-SJS patients and in only 3% of the normal subjects, but was absent in CBZ-tolerant patients [23].

Abacavir is a nucleoside analogue reverse transcriptase inhibitor taken in combination with other anti-HIV medications. A potentially fatal hypersensitivity, or allergic reaction, has been associated with the use of abacavir in at least 5% of patients; abacavir hypersensitivity was identified in 18 patients and drug tolerance was confirmed in 167 patients after 6 weeks of therapy in a recent study in a predominantly white, male study population [24]. Patients with the allele *HLA-B*5701* were > 100-times more likely to be hypersensitive to abacavir; this allele was present in 14 (78%) of the 18 patients with abacavir hypersensitivity, and in 4 (2%) of the 167 abacavir-tolerant patients. The genetic region was marked by the combination of *HLA-B*5701* and *-DR7*, and *HLA-DQ3* (the 57.1 ancestral haplotype) was present in 13 (72%) hypersensitive patients and none of the tolerant patients [24]. All patients with the full 57.1 ancestral haplotype experienced abacavir hypersensitivity. Because four patients in the study developed abacavir hypersensitivity without presenting the susceptibility haplotype, there seem to exist additional susceptibility factors, which are possibly related to the disposition of abacavir in the nontolerant patient population.

Overall, it appears that theragenomics-based patient selection for targeted therapy approaches with safety aspects at centre stage is increasingly being used in the clinical setting. One should be aware, however, that at present a narrow line of distinction between the causation trough and the correlation with a genetic predisposition for ADRs might exist. Although the aberrant handling of atomoxetine, irinotecan and the thiopurines through allelic variants of CYP2D6, UGT1A1 and TPMT, respectively, might reflect the cause of the associated ADRs, a mere albeit significant correlation of the carbamazepine- and abacavir-induced hypersensitivity reactions with HLA allelic variants seems to exist [23,24]. Only a more profound analysis of all involved components in the pathways that form the etiological basis for the observed drug-induced hypersensitivity reactions will reveal the causative genetic variants as well and eventually yield positive predictive values sufficient to identify all patients at risk for hypersensitivity reactions to carbamazepine or abacavir [25] across diverse patient populations of varying ethnic origin.

4. Theragenomics-based knowledge base and knowledge management platform

As is easily evident from the few examples cited above, handling of the knowledge and data space surrounding theragenomics is a daunting task. Individualised drug safety aims at the prospective identification of individual patients who carry genetic predispositions for the development of serious or fatal ADRs under drug treatment. The analysis and predictive understanding at the molecular level of the combination of chemical and genetic risk factors that eventually predispose a patient for the development of a severe ADR rely on various underlying data such as chemical structures and metabolic pathways, gene and protein sequences, theragenomics-derived data including genotype and haplotype determinations and clinical data on a single patient basis, which all need to be integrated for comprehensive drug safety evaluation on the basis of a single individual. In fact, in order to perform thorough analyses and reach insightful decisions of high impact in a field as complex as theragenomics-based individualised drug safety, a holistic view on relevant knowledge, information and

data are necessary. Such knowledge/data are heterogeneous and originate from many different scientific disciplines, and its collection is arduous and convoluted. Often, in the public domain and within organisations, the information is dispersed in a variety of databases and/or systems in noncompatible formats. To efficiently address these challenges comprehensive knowledge bases and platforms with a well-defined and clearly structured organisation, linkage and representation of the dispersed yet related knowledge are imperatively needed and have been missing so far.

4.1 Filling the gap

The traditional approaches of creating relational data stores for storage of this sort of information have been considered problematic, and with the increasing desire to integrate multiple data sources new approaches are needed. An ideal platform should be able to consistently support the acquisition, integration and processing of information from several sources. It should provide a concise representation of actual knowledge available at a given point in time, and last, but not least, it should encourage by design the sharing and augmentation of knowledge, and to use it in a forward-looking way.

In order to enable and support scientific experts in efficient and high-impact assessment of theragenomics-based drug safety, a functional platform has been designed that combines various concepts in the area of knowledge management to perform ontological knowledge acquisition, processing and integration. This platform has been coined OKAPI; together with the adverse drug reactions information scheme (ADRIS; see below), it comprises the framework for the implementation of SafeBase™ (TheraSTrat): the knowledge base and platform for theragenomics-based individualised drug safety.

The value of scientific knowledge management relates directly to the effectiveness with which the managed knowledge enables a single expert (e.g., molecular toxicologist) or a group of experts (e.g., cross-disciplinary international development team) to make a high impact decision based on a given situation (e.g., unresolved serious safety issue). Without on-demand access to managed knowledge, every situation is addressed based on what the single expert or a group of experts can contribute to the resolution of the safety issue in question. In contrast, with on-demand access to managed knowledge, every issue can be addressed with the sum total of knowledge and experience anyone in an organisation has ever gathered about an issue of a similar nature. Learning and acting from experience, using known cases serving as patterns for solutions within similar new situations is highly relevant in corporate knowledge management, as the value of being able to retain and learn from experience should be evident. This methodology is the basis for case-based reasoning (CBR) [26]. To approach a problem using CBR one has to retrieve the most similar case(s), then based on the information and knowledge gathered using these cases one could propose or revise a solution for the problem. Much of importance is to retain the parts of such a CBR approach in a knowledge management system, as it is likely to be useful for future problems in similar areas.

Commonly used linear storage and representation of knowledge, whether in reports, tables, slide presentations or other form, fail to capture the networked nature of knowledge. All information and data are interconnected, and only in its interconnected representation does one see the context in which a particular piece of information is placed. Seeing information and data within its context allows the extraction of knowledge from this information and data. The OKAPI platform stores and represents its data in a semantic network that organises logically, relates and integrates the diverse data needed for a CBR-based safety assessment. A semantic network is a graphical notation for representing knowledge in patterns consisting of nodes and relations. The nodes represent objects, concepts or situations within a specific knowledge domain. The relations represent and define relationships between the nodes. Semantic networks are often used to represent the knowledge of human experts. Computer implementations of semantic networks were first developed for artificial intelligence and machine translation, but earlier versions have long been used in philosophy, psychology and linguistics.

With different information resources coming together in one system designed for the use of experts from various scientific disciplines, a common understanding of the structure of this information is crucial. To address this, ontologies are emerging as a key aspect of information management in many areas, from the interchange of data to corporate knowledge management. Simply put, an ontology is a way of defining a shared common understanding about the kinds of objects and relationships that are being discussed within a knowledge domain, so that dependable communication can happen between people using this information. Of course, ontologies are dependent on the subject matter (i.e., knowledge space) they represent. To that end the ADRIS has been designed to address conceptionally the ontological organisation as well as the logical and semantic relation of the factors that are relevant in the knowledge space of individualised drug safety. Thus, ADRIS provides the holistic yet structured view on molecular factors that form the basis for the aetiology of ADRs at the level of individual patients. ADRIS served as the ontological prerequisite for the creation of the theragenomics-based knowledge base and knowledge management platform, termed SafeBase. Recently [27], in a first publication, the possibilities of ontology based knowledge management in drug safety were discussed and presented, using troglitazone-induced hepatotoxicity as a model case [28]. Such analyses help to identify alleles in the population that associate with huge variations in single mechanistic steps of drug metabolism and toxicity. In the case of troglitazone metabolism, it became immediately evident that there exist huge gaps in the knowledge on the ST1A3 and SULT1A1 genes and their handling in the population of a major metabolic step in troglitazone disposition *in vivo* [28].

Driven by and based on the ADRIS ontology, the OKAPI platform provides predefined terms for physical and logical

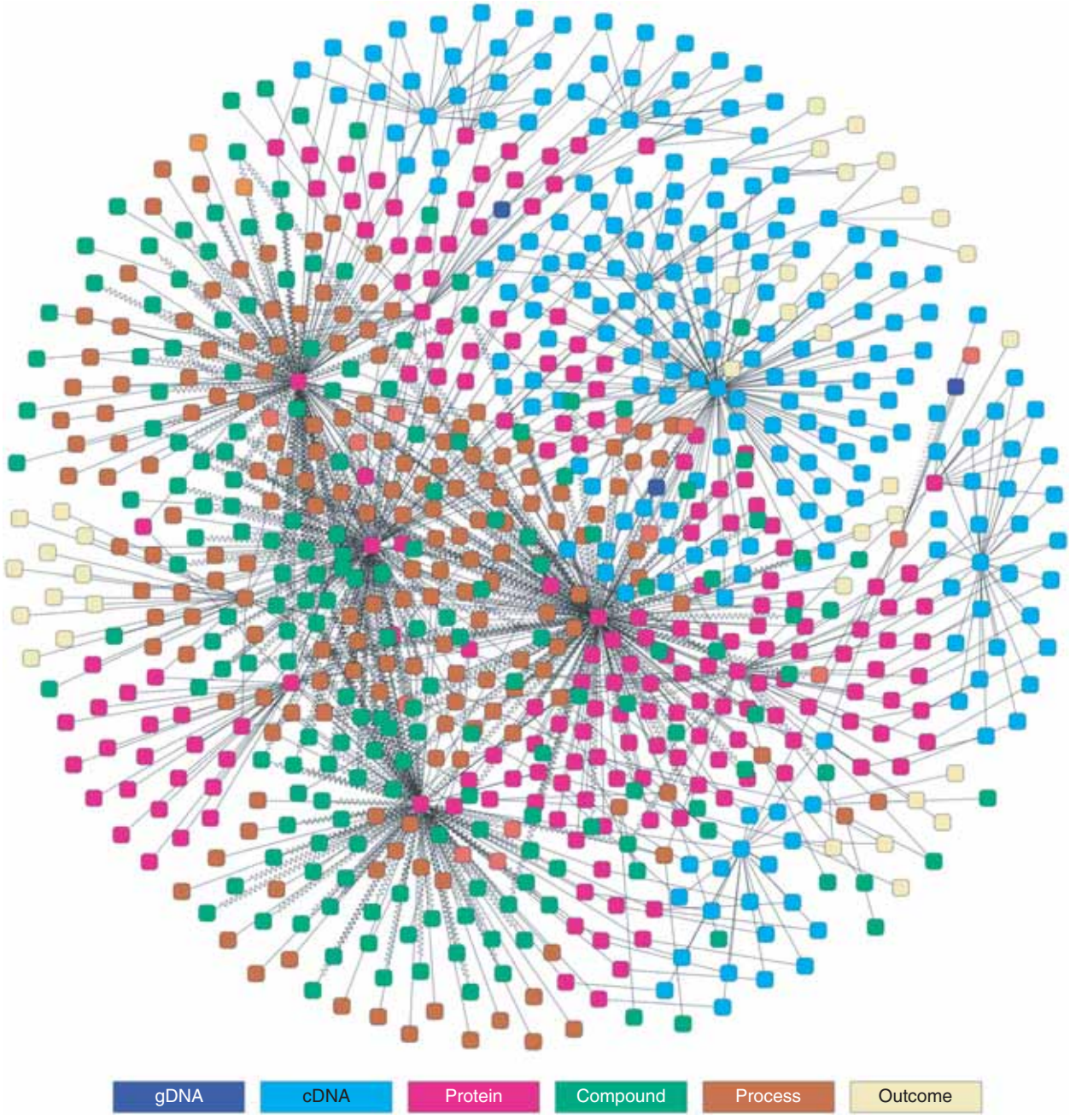

Figure 1. Simplified interaction map providing a top level overview of the knowledge (knowledgelets and their relations) centered around a selection of human cytochrome P450 enzymes (i.e., CYP1A1, -2C9, -2C19, -2D6 and -3A4), including their coding genes and allelic variants, a limited number of their substrates, inhibitors, inducers, and primary steps of drug metabolic processes they catalyse, as well as a selection of outcomes of these processes. Note that for clarity reasons, the representation here is in simple view; in the life SafeBase™ platform, views for medium and expert levels are available, which reveal medium and high levels of detail, respectively, for each single represented knowledgelet and relation. Concept and graphical rendering are from within SafeBase.

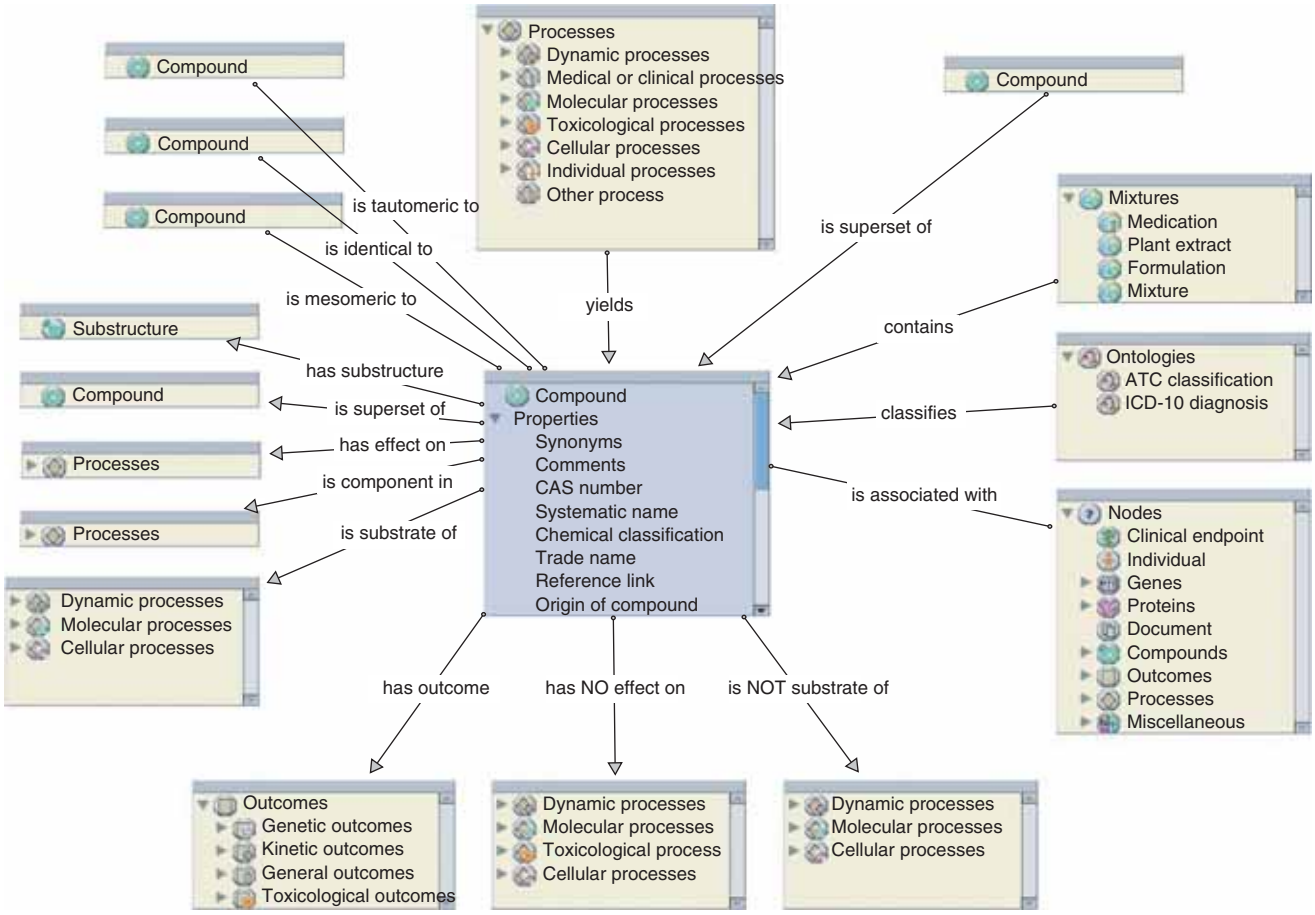

Figure 2. Knowledgelets have knowledge about their nearest neighbours through directed (i.e., relation-in or relation-out) and nondirected relations. The relations are logical, semantic and may carry qualitative and/or quantitative information themselves and as nodes may carry properties as well. Here, the example of a chemical compound and its possible relations with its nearest neighbouring nodes (i.e., processes, other compounds, proteins, genes, etc.) is given. Concept and graphical rendering are from within SafeBase™.
ATC: Anatomical Therapeutic Chemical Classification; CAS: Chemical Abstracts Service; ICD-10: International Classification of Diseases, version 10.

objects, their properties and the relationships and interactions between these objects. By implementing these conceptual pieces in one system the OKAPI platform facilitates the travelling of the knowledge space on individualised drug safety. When it comes to an unresolved drug safety issue, the mapping of this space allows researchers and regulators alike to ask seemingly simple yet possibly high-impact questions such as i) do you know what you know; ii) do you not know what you know; iii) do you know what you don't know; and iv) do you not know what you don't know? The facile representation of some of the answers to such questions lets researchers and physicians in drug development, clinical practice and regulatory authorities quickly identify strengths, weaknesses, threats and opportunities in theragenomics-guided drug development and safety evaluations, make theragenomics-based knowledge broadly available and allow such knowledge to be used both in retrospective and prospective ways. Figure 1 illustrates an example of a top level overview of the interactions in the knowledge space centred around a selection of human CYP enzymes (i.e., CYP1A1, -2C9, -2C19, -2D6 and -3A4), including their coding genes and allelic variants, a limited number of their substrates, inhibitors and inducers, and primary steps of metabolic processes they catalyse.

4.2 Innovative use of technology and methodology to achieve results

SafeBase is an actual implementation of the specification of the OKAPI platform, and as such offers an innovative and powerful solution to integrate knowledge and data across multiple scientific disciplines pertinent to both drug development and drug safety. SafeBase integrates information from chemistry, pharmacology, toxicology, kinetics, pharmacogenetics, toxicogenetics, preclinical and clinical research, as well as pharmacovigilance, including data for individual patients. SafeBase allows the selection of any discrete entity of knowledge, observation and investigation of its relationships and details, and

determination of the source of any attribute information. Users can easily identify the entity's ontological relatives and navigate through their relationships to investigate their characteristics in more detail.

SafeBase is a multitier, multiuser system. It consists of a rich Java 2, Swing [104] client application, that is, the 'Intelligent Knowledge Browser', which provides the functionality to store, process, query and visualise the information in the central data libraries. The client application connects to an internet- or intranet-based server architecture using the standard secure HTTPS protocol. The server architecture consists of web servers (in this case based on Apache 1.3.x [1051]), which distribute requests to a Java-based application server layer called 'SafeBase Toxilico'. The application servers access relational databases containing the core data libraries, using an object-relational mapping layer. The application servers can also access Web pages and/or Web services and integrate their content and visualise it within the system. This results in a very scalable platform architecture, which can easily be adapted and/or expanded to the needs of individual users. In addition, the use of standard protocols, Java and standard SQL92 relational databases supported by a large variety of platforms allows running SafeBase client as well as the SafeBase server part on multiple platforms and with various databases.

4.3 Innovations for the field

SafeBase is primarily designed to handle theragenomic knowledge and information, but not raw data, as information is much richer than simple data sets. Whereas data are a syntactic correct container of characters, information contains concrete relations to and representations of the real world, thus making information significantly more valuable than raw data. The need to integrate this information was realised and, consequently, SafeBase is designed to contain, besides the information itself, semantic relationships and annotations that add a logical meaning to the information. The system supports the identification of knowledge among pieces of information stored in SafeBase, and because this information covers a broad range of know-ledge disciplines, SafeBase also allows easy crossing of boundaries between multiple scientific disciplines.

ADR-relevant information is stored, maintained, constantly updated and curated by scientific experts in the data libraries of SafeBase. Within the semantic network maintained by the system, this information is abstracted in an object-oriented way and consists of discrete predefined entities termed 'knowledgelets'. These knowledgelets are ontologically organised and represent individual elements such as compounds. Figure 2 shows the ontological relatives of the knowledgelet 'compound'. Typically, a knowledgelet networks with its nearest neighbours through directed relations (node-in and -out) as well as through nondirected relations. This allows each single knowledgelet to be part of several concepts depending on the context in which the knowledgelet plays a role; each knowledgelet 'knows' about its relative position within a selected knowledge space.

There are many types of elements that can faithfully represent theragenomics knowledge. The top-level categories of knowledgelets include:

- individuals (including single patients, groups of patients, subpopulations, populations)
- end points (toxicological and clinical)
- genes (e.g., coda, gDNA, mRNA, SNPs, alleles)
- proteins (e.g., adducted proteins, antibodies, allelic variant proteins, etc.)
- compounds (e.g., drugs and other chemicals, substructures, mixtures, formulations, combination medications, biomolecules, etc.)
- processes (cellular, dynamic, molecular, mechanistic, toxicological, etc.)
- outcomes (genetic, kinetic, dynamic, mechanistic and numeric, including genotype, haplotype, expression level, pharmacokinetic, toxicokinetic, drug–drug interaction, pharmaco- and toxicodynamic, covalent binding, adduct formation, molecular mimicry, antigenicity, K_m, V_{max}, K_i, IC_{50}, half-life, C_{max}, etc.)
- relations (logical, semantic, distant, others)

Knowledgelets are annotated with entity specific predefined properties that add relevant information with specific semantic characteristics. These properties can be entered in various formats such as numeric, textual, pick-list and byte data, and can cover attributes such as chemical structures, gene sequences, covalent modifications, synonyms, Chemical Abstracts Service numbers, and so on.

Knowledgelets relate to each other, and the relations themselves can be annotated with relation-specific properties. SafeBase integrates actual and metadata (i.e., descriptive information) in one system in order to efficiently organise and query ADR-related information. Unlike other data sources that mainly contain links to external data repositories, SafeBase organises and stores the relevant data locally. Links pointing to external data sources are included where applicable. As a result, the immediate availability and the integration of this vast amount of heterogeneous data allows to efficiently mine and visualise this information in one single system, thus enabling the user to gain new knowledge related to ADRs.

New ADR data are created constantly. For example, SNPs, splice variants and gene expression analyses are revealing new insights in drug action and disease mechanisms. SafeBase is designed to accommodate this emerging information in the form of knowledgelets. Their properties can easily be defined and added to the system in a consistent way, thus allowing the system to grow with current and future needs.

Within the client front-end of SafeBase, knowledgelets, including their properties and relationships to other entities, are represented as semantic networks. The client provides an efficient and easy way to navigate through large amounts of information while maintaining the overview. For example, starting from an individual's genotype, one can navigate via the corresponding gene to the protein and

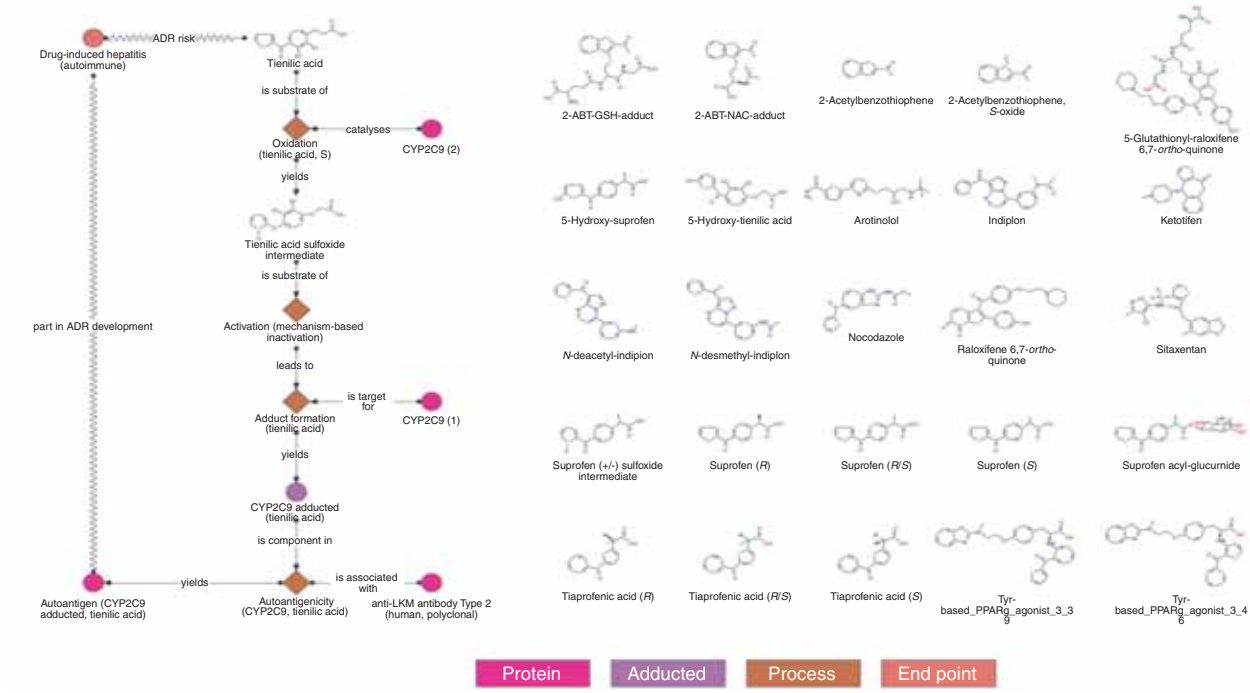

Figure 3. Metabolism of tienilic acid leads to mechanism-based inactivation and adduction of CYP2C9 in humans *in vivo*. Adducted CYP2C9 is an autoantigen for LKM-2 autoantibodies and is associated with the induction of tienilic acid-induced hepatitis (pathway represented on the left side of the concept). Using the thiophene substructure of tienilic acid, chemical similarity search from within SafeBase™ reveals numerous parent molecules and metabolites that carry the thiophene substructure and may exhibit behaviour similar to that of tienilic acid in patient populations (right side of the concept). Concept and graphical rendering are from within SafeBase.

ABT: Acetylbenzothiophene; ADR: Adverse drug reaction; CYP: Cytochrome P450; GSH: Glutathione; LKM: Liver kidney microsomal; NAC: *N*-acetyl cysteine; PPAR: Peroxisome proliferator-activated receptor.

its variants, to metabolic pathways and to clinical end points all within one system.

SafeBase provides knowledge management, visualisation and querying tools such as nearest neighbours identification, elucidation of mechanistic paths between knowledgelets, automated layouts, and so on. Critical functionalities to manipulate information such as molecular structure drawing and similarity search tools, along with DNA- and protein-sequence management modules are seamlessly integrated. Besides drawing and inspecting chemical compounds, automatic generation and parsing of canonical SMILES and chemical hashed fingerprinting necessary for the search of similar molecular structures are provided as well. Various DNA- and protein-sequence-specific tasks (e.g., evolution of allele-specific variations in gene sequences and introduction of drug-specific covalent modification sites in protein sequences) are also integrated.

4.4 Other systems in this field

Several databases and platforms, both in the public domain and from commercial outfits, try to address the demands of the scientist when dealing with theragenomics to varying degrees. Initially, one might to refer to nucleic acid and protein sequence information, which is mainly obtained and edited from public database collections such as those available at the National Center for Biotechnology Information [106], the European Molecular Biology Laboratory Nucleotide Sequence Database [107] and SWISS-PROT [108]. Allelic variants and information on their functions are accessible from databases such as the Human Gene Mutation Database [109] and, with particular emphasis on their familial inheritance, from the Online Mendelian Inheritance in Man™ (OMIM™; Johns Hopkins University) database [110]. The Biomolecular Interaction Network Database (BIND) [111] is a publicly available Internet tool containing a growing collection of records on molecular interactions including information gathered from scientific literature. Interactions are the basic units of BIND and can be linked together to form molecular complexes and pathways. BIND supports researchers in understanding publicly available experimental data in a global context. PharmGKB [112] is a publicly available internet-based research tool. It is a central repository for genetic and clinical information about individuals who have participated in research studies at various medical centres. In addition, genomic data, molecular and cellular phenotype data, and clinical phenotype data are accepted from the scientific community at large. PharmGKB's aim is to aid researchers in understanding how genetic variation among individuals

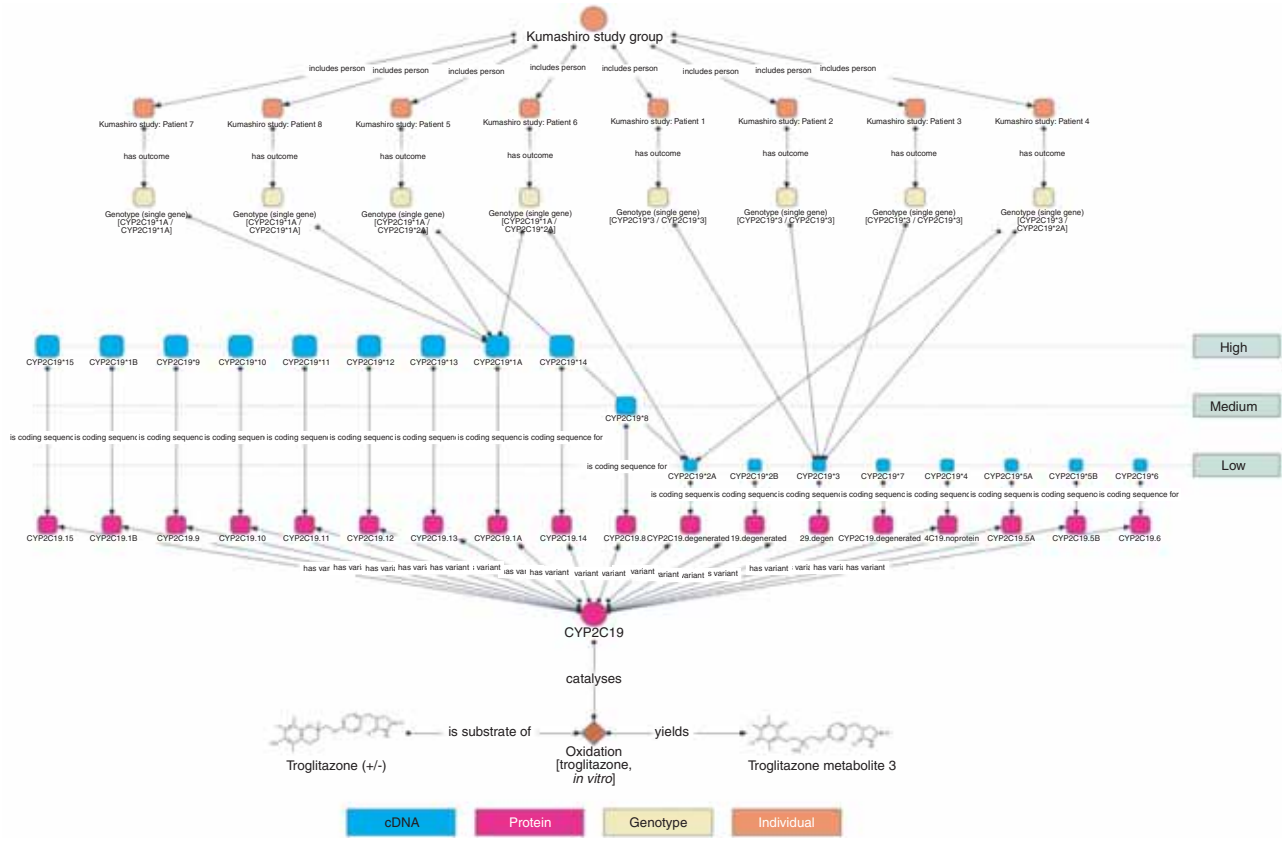

Figure 4. Presentation of the allelic variability in a patient cohort (Kumashiro study group [35]**) of a single metabolic step in the disposition of troglitazone in humans, catalysed by CYP2C19.** Note that the alleles of CYP2C19 (nodes in green) are arranged according to the functional activity of the coded for protein variants (high, medium, low). Note that the representation here is in medium view; in the life SafeBase™ platform, expert and simple levels of views are available that reveal high and low levels of detail, respectively, for each single represented knowledgelet and relation. Concept and graphical rendering are from within SafeBase.
CYP: Cytochrome P450.

contributes to differences in reactions to drugs. PharmGKB is the nearest to the SafeBase knowledge base with respect to content, and might hold more pharmacogenetic data in that the population of its knowledge base occurs through an established network of scientific experts in pharmacogenetics. PharmGKB offers its users PharmGKB Pathways, which are drug-centred, gene-based interactive pathways that aim to highlight candidate genes and gene groups and associated genotype and phenotype data. These overviews are similar to concepts available in SafeBase (see Figures 3 – 5), which extend, however, far beyond pharmacogenetic data and include numerous molecular toxicological and mechanistic concepts. In both systems, these pathways and concepts may be updated as new evidence is published. Functionally, SafeBase goes much further than PharmGKB in that, for example, chemical similarity searches are implemented, recognition of stereoselectivity of parent compounds and metabolites is automated (inclusive the generation of stereo SMILES), concepts can be travelled for crossing and intersecting pathways, and graphical representation of pathways is fully under control of the individual user by the system's inherent rendering capabilities (see Figures 3 – 6).

Several commercial systems may resemble by design the SafeBase approach, although their focus might vary. For example, GeneGo [113] supports with their products the assembly of experimentally determined enzymatic reactions and signalling protein interactions into interconnected networks. By analysis of databases on a fundamental understanding of human biology they assist scientists in drug target selection, validation and identification of molecular biomarkers for disease states. A recent publication by scientists of the company [29] presented a novel method for visualising nuclear hormone receptor networks relevant to drug metabolism in their MetaDrug™ database product. Ingenuity Systems [114] delivers systems biology expertise through pathways analysis databases and platforms. The Ingenuity Pathways Knowledge Base provides value throughout the drug discovery process, from early-stage target identification and validation, to later-stage activities such as toxicology, pharmacogenomics, biomarker identification and alternative indications of approved drugs. Moreover, their

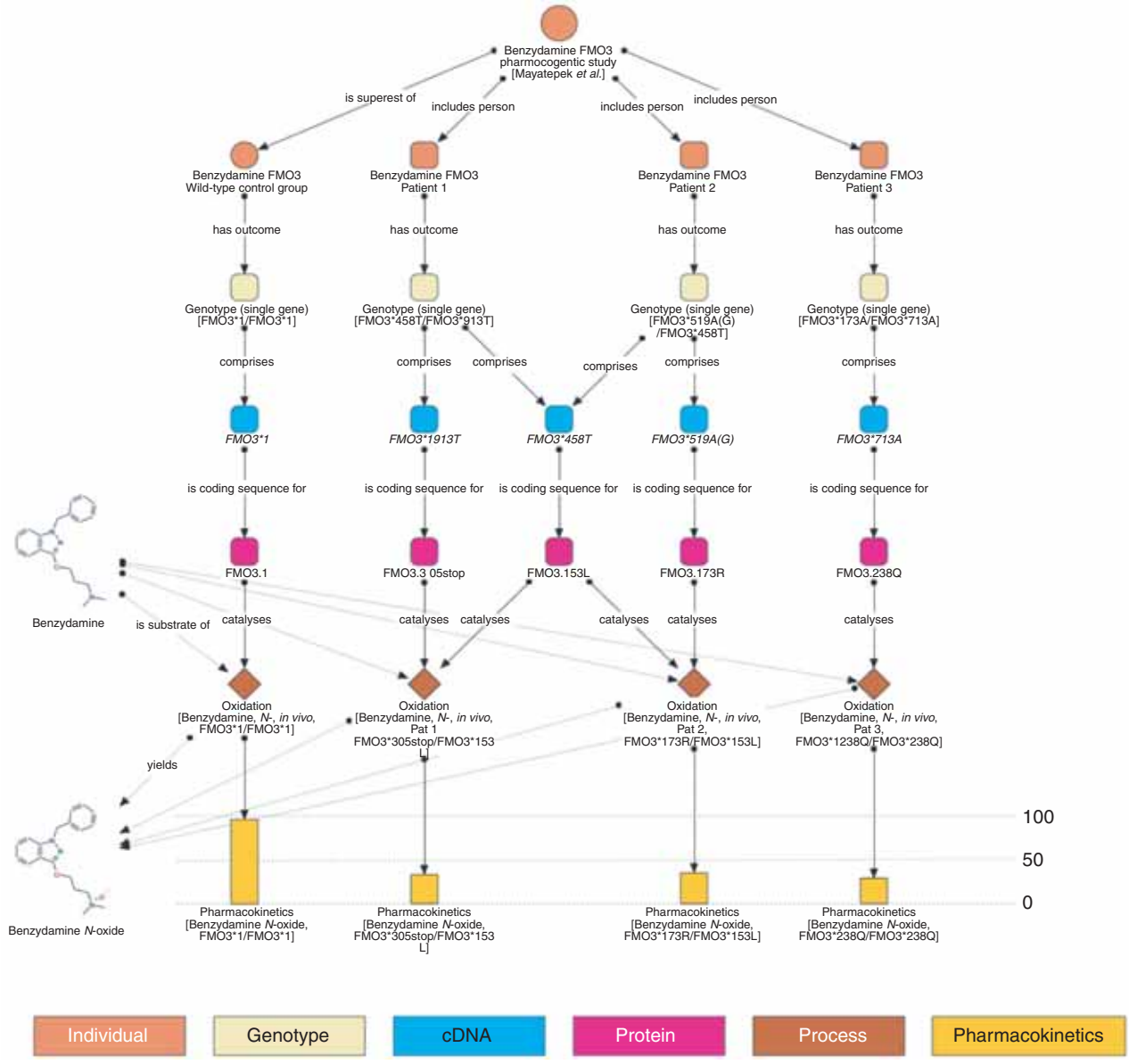

Figure 5. Visualisation of the results of a pharmacogenetic study *in vivo* on the genotype-dependent metabolism of benzydamine by human individuals (Mayatepek *et al.* [36]**).** The benzydamine *N*-oxygenation capacity of a group of individuals homozygous for the *FMO3*1* allele (wild-type control group) and of single heterozygote (*FMO*458T*/*FMO3*913T* or *FMO3*519A*/*FMO3*458T*) and homozygote (*FMO3*713A*/*FMO3*713A*) patients was analysed. The pharmacokinetic output parameter was the metabolic ratio benzydamine *N*-oxide/benzydamine in the urine collected for 24 h [36]. Note that the representation here is in medium view; in the life SafeBase™ platform, expert and simple levels of views are available that reveal high and low levels of detail, respectively, for each single represented knowledgelet and relation. Concept and graphical rendering are from within SafeBase.
FMO: Flavin-containing monooxygenase.

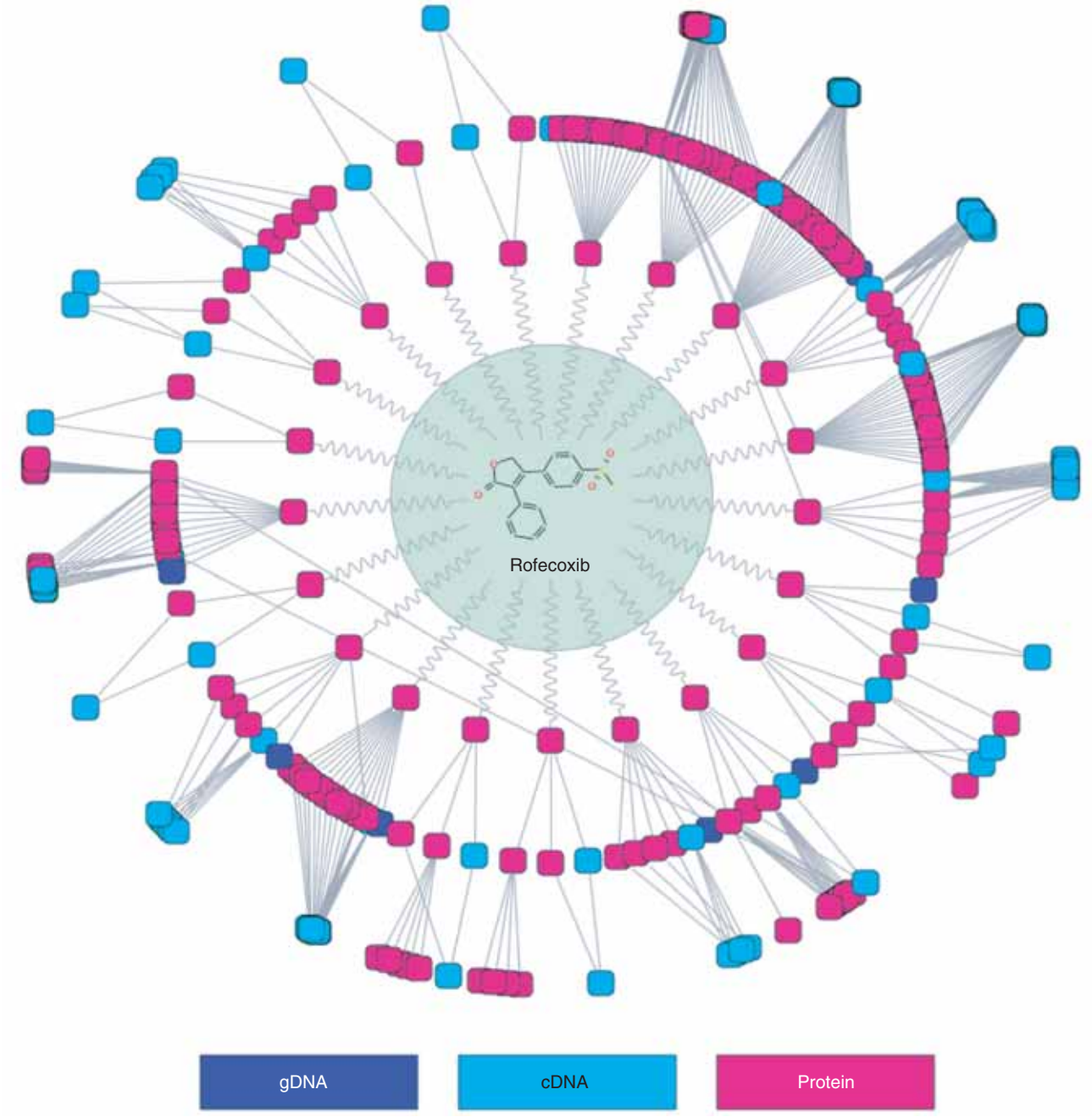

Figure 6. Simplified interaction map giving a top-level graphical overview on candidate genes (and allelic variants thereof) responsible for rofecoxib-induced myocardial infarction. Note that, with the exception of the central node on rofecoxib, which is in medium level view, for clarity reasons, the representation here is in simple view; in the life SafeBase™ platform, views for expert and low levels are available that reveal higher and lower levels of detail, respectivley, for each single represented knowledgelet and relation. Concept and graphical rendering are from within SafeBase.

approach enables companies to create and maintain a common, enterprise-wide knowledge asset that captures, structures and shares biological knowledge. Iconix Pharmaceuticals [115] is active in the field of chemogenomics: the integration of chemistry and genomics to profile drug candidates. By investigating the broad spectrum of effects of chemical compounds on a living system the company provides genomic patterns that are predictive of specific forms of toxicity and mechanisms of action, all of which are accessible and stored using the DrugMatrix® database.

One should also be aware of concepts that have the potential to serve as the technological basis for the design of platforms and knowledge systems similar to the ones describe here. For example, Protégé [116] is a tool that allows users to construct domain ontologies, customise data entry forms, and enter data. It is a platform that can be extended to include graphical components such as graphs and tables. Protégé is designed to allow experts to reuse domain ontologies and problem-solving methods, thereby shortening the time needed for the maintenance of knowledge-based systems [30]. Moreover, Cytoscape [117] is an open source software project for integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework. Cytoscape is most powerful when used in conjunction with large databases of protein–protein, protein–DNA, and genetic interactions that are increasingly available for humans and model organisms. Cytoscape provides basic functionality to layout and query the network, in order to visually integrate the network with expression profiles, phenotypes, and other molecular states, and to link the network to databases of functional annotations [31].

5. Use of retrospective knowledge on adverse drug reactions in a forward-looking way

Theragenomics-based knowledge, which is available and augmented through knowledge management platforms, is used at several stages of the drug development process.

5.1 Identification of candidate genes of predisposition

In one approach, theragenomics-based knowledge is being used to identify candidate genes, and allelic variants thereof, that may render single patients susceptible to the development of serious ADRs when exposed to a given drug. Such candidate gene profiles will then be used in cohort studies to verify the positive and negative predictive values of single alleles, combinations thereof, or haplotypes to identify susceptible individuals. Eventually, such profiles will end up in diagnostic companions for drugs that carry a certain risk of provoking serious adverse reactions in susceptible individuals. Theragenomics-based knowledge bases and platforms are facilitating the initial steps in this process, namely the candidate gene selection.

Except for situations where a single gene dominates in a rate-limiting manner a single step in the disposition of a drug with a narrow therapeutic window, the aetiology of ADRs expectedly is based on the interaction of multiple genes. Extensive analyses in the SafeBase knowledge base, in order to identify candidate genes for selected ADRs that are frequently the reason for discontinuations of clinical development of drug candidates or for the withdrawal of drugs from the market, are in line with this expectation (Table 4). Figure 6 shows the graphical output of such an analysis for the identification of candidate genes (and their allelic variants) from within SafeBase, which may render single patients susceptible to myocardial infarction, as induced by the COX-2 inhibitor rofecoxib. Candidate gene selection for a selected end point can be achieved by examining and analysing any combination of data repositories, holding the necessary data (for example, the OMIM database [110] as a starting point). However, this might be a cumbersome and extremely time consuming process and offers few clues to interactions of selected candidate genes with other clinical end points. In contrast, within the fully expanded knowledge space of candidate genes for rofecoxib-induced myocardial infarction, for the expert user, it becomes immediately clear that many of the candidate genes identified so far, including, but not limited to, ALOX5AP, APP, BRI, CST3, NOTCH3, CYP3A4, CYP8A1, PLA2G7 and TBXAS1, are likely to play a role in rofecoxib-induced myocardial infraction, in pre-existing susceptibility for the same condition, in ischaemic and/or haemorrhagic stroke and in other ADRs that affect the cardiovascular system (Gut *et al.*, manuscript in preparation). Moreover, immediate overviews on how other '-coxibs', for example, celecoxib, valdecoxib, lumiracoxib, etoricoxib, cimicoxib, deracoxib, firocoxib, parecoxib and tilmacoxib, may be associated with a candidate gene profile similar or identical to that obtained with rofecoxib are obtainable. The ontology based networking of knowledge also leads to the identification of 'hidden' or 'confounding' risk factors such comedications, ingredients of dietary supplements (e.g., ephedrine), phytohormones and designer drugs with genes that are constituents of the candidate gene set for rofecoxib-induced myocardial infarction (Gut *et al.*, manuscript in preparation).

5.2 Guidance of the drug development process

In a second approach, such knowledge becomes usable for guidance in the drug discovery process. Modern lead selection and optimisation processes may be guided by questions such as:

- What are the most important overlooked factors in the chemical structures of a lead compound that may, later in preclinical and clinical development, lead to metabolism- and disposition-related problems of the molecule?
- Are there (possibly rare) allelic variants of genes present in the future target patient population that may handle the lead compound dramatically different from the corresponding wild-type allele?
- Have chemical substructures present in the lead compound led to severe toxic effects in patient populations or subpopulations?

Table 4. Overview on the number of candidate genes and the resulting number of candidate allelic variants that have been identified from within the SafeBase™ knowledge base that are associated with the aetiology of a selection of adverse drug reactions frequently occurring in drug development or leading to drug withdrawals.

Adverse drug reaction	Candidate genes	Candidate alleles
Myocardial infarction (rofecoxib)	48	247
Rhabdomyolysis	15	2746
Liver toxicity	68	3683
Cholestasis	37	510
Agranulocytosis	9	248
Ischaemic colitis	9	198
Myopathy	167	389
Haemolytic anaemia	84	2202
Thrombocytopoenia	87	2256
Neutropoenia	55	1449
Thrombosis	54	2124
Heparin related	73	890
Lactic acidosis	41	838
Genes of the immune system*	741	4891
Genes of metabolism and disposition*	74	1252

*The total number of candidate genes identified so far that may a play role in any aetiology of adverse drug reactions with any given drug or chemical beyond drugs. For a drug-induced disease caused by a selected single drug, the number of candidate genes from this group will be considerably lower and largely depend on the pathways of disposition of the selected drug.

Taken together, theragenomics-based evidence may lead to optimisation of molecules at this stage of development in order to increase the likelihood of success in preclinical and clinical development, thereby achieving an early attrition as opposed to late attrition of drug candidates. In fact, collaborations are emerging that take advantage of guidance in the lead optimisation process based on knowledge from theragenomics-based knowledge bases according to the concept shown schematically in Figure 7. The goal is to select lead molecules for the further development process that do not carry structural risk factors, or produce them in the course of their disposition as reactive intermediates or metabolites thereof, of which there exists evidence in the patient population for serious ADRs. As an example, the case of tienilic acid is presented, which produces, by a fairly well-understood mechanism, tienilic acid-induced autoimmune-type hepatitis in some patients [32]. The process involves the mechanism-based inactivation of, and adduct formation to, CYP2C9 [33,34] with the ensuing recognition of the adducted CYP2C9 by autoimmune-type liver kidney microsome-2 autoantibodies and the development of tienilic acid-induced hepatitis in some patients (see left side of concept in Figure 3) [32-34]. Mechanism-based inhibition and adduction to CYP2C9 (and possibly other protein targets) is a general, unfavourable property of thiophene substructures [34] and a quick chemical substructure search from within the SafeBase knowledge base reveals within seconds a multitude of chemical compounds (parent molecules and metabolites alike), all of which carry thiophene-type substructures with the potential of mechanism-based inactivation of and adduction to CYP2C9 (see the right panel of the concept in Figure 3).

Theragenomic-based knowledge may come into play at these very early stages of the drug development process because allelic variants of a given gene may handle compounds very different from wild-type alleles, and each single step in the disposition of a molecule may be subjected to this variability. This is illustrated in Figure 4 for one single metabolic step of troglitazone biotransformation, catalysed by allelic variants of CYP2C19 [35]. This single biotransformation step will encounter huge variability, depending on whether individual patients are homozygotes or heterozygotes for the wild-type alleles or functional aberrant allelic variants of CYP2C19 (Figure 4). Particularly the frequency by which such alleles occur in the targeted patient population may influence the decision process on a lead selection. Such early knowledge on critical alleles could prove very useful in the case of safety concerns during drug development in clinical Phases I and II, in which the early and rapid identification on safety relevant alleles (SNPs, haplotypes, genotypes) could provide confidence for a commitment to full drug development. It may guide the development towards certain subpopulations of patients not predisposed for the development of undesired adverse events. A likely strategy has been presented for the development of tranilast [20,21].

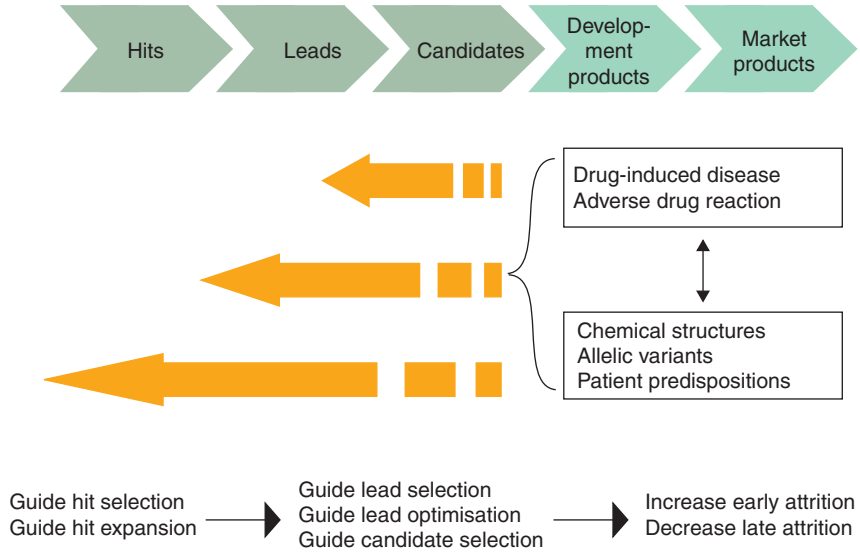

Figure 7. Schematic representation on how theragenomics-based knowledge on compounds in late development or on the market, and obtained from public domain and company databases, may be (re-)used in the very early stages of the drug development process (i.e., hit finding, lead optimisation, candidate selection).

6. Conclusion

The question of whether theragenomics-based tests for drug safety help to make drugs safer in the population may still be unanswered; however, the fact that in the clinical setting, genotyping is beginning to be used for dose adjustment for drugs or for exclusion of certain patients from therapies, as discussed above, indicate that the use of theragenomics-based knowledge will soon play an important role at all stages of the drug development process and eventually increase the safety of drugs considerably. What remains is the formidable task for pharmaceutical companies and regulators to transfer the huge amount of data that is currently generated in the field of theragenomics into contextual knowledge for the benefit of the patient, and to make this knowledge accessible in a productive manner.

This task is where platforms such as SafeBase and others, as discussed above, come into play. The advantages and strengths of ontology driven platforms are: i) the handling of diverse pieces of knowledge; ii) the integration of holistic knowledge spaces amenable for easy evaluation from very different points of view (i.e., experts from different disciplines can access the same knowledge); and iii) the visualisation of relations and interactions of such knowledge at any level of detail. Accordingly, knowledge bases and platforms need to be able to integrate the necessary knowledge(lets). As an outlook, in Figure 5, a view is presented on how knowledge bases and platforms in theragenomics might be able to make information accessible to scientists in early and late stages of drug development and to physicians in clinical practice; for example, how the genetic background of individual patients forms the basis of handling single steps in the disposition of a drug *in vivo*, how this might affect the safety (and of course the efficacy) of this drug and for which patients adjustments in therapeutic strategies (dose adjustments, initiation of alternative therapies because of intolerance, etc.) need to be considered. The example chosen here is the visualisation of the pharmacokinetic outcome of the exposure of patients of defined flavin-containing monooxygenase 3 genotypes to benzydamine, an NSAID [36]. This type of data/knowledge is emerging rapidly from the theragenomics field and will take centre stage in decision making in the drug development process.

7. Expert opinion

Theragenomic knowledge bases and platforms will not only have a role in support of future theragenomics-based risk analysis and safety assessments for drugs, it is becoming increasingly clear that the disposition of chemicals, pollutants and dietary ingredients, for example, is under pharmacogenetic control, as is the susceptibility of humans for adverse reactions towards these agents. Thus, paraoxonase 1 genotype-dependent disposition of the organophosphorus insecticide diazinon [37], seemingly elevated risks for carriers of certain combinations of alleles of CYP2D6 and *N*-acetyltransferase 2 for susceptibility to multiple chemical sensitivity [38], the emerging association of certain alleles of CYP2E1, myeloperoxidase, NAD(P)H dehydrogenase quinone 1, glutathione-*S*-transferase (GST)M1 and GSTT1 with benzene exposure related haematotoxicity [39,40], and the vast array of pathways by which dietary ingredients may interact with the disposition of drugs and chemicals (giving way to a new research field coined 'nutritional genomics' [41]) are but a few examples for this development.

These knowledge spaces need to be integrated into meaningful theragenomics-based risk assessment, as they may contribute hidden factors to the predisposition of a single patient for the development of seemingly unpredictable ADRs.

Few databases or knowledge management systems today address this problem satisfactorily. SafeBase is one of these tools and provides the user with a range of possibilities to easily visualise and communicate very complex problems on the level of both an expert and a layperson. It is the firm belief of the authors of this review that these knowledge integrative and communicative capabilities of knowledge management systems will be a key success factor in applying theragenomics concepts successfully in individualised drug safety and personalised medicine.

Acknowledgements

The authors thank A Wess for reading the manuscript and R Dannecker, R Schindler, H Hug, M Scarsi, C Rücker, R Boeuf and S Gonukula for helpful discussions.

Bibliography

1. LAZAROU J, POMERANZ BH, COREY PN: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* (1998) 279:1200-1205.
2. KVASZ M, ALLEN IE, GORDON MJ *et al.*: Adverse drug reactions in hospitalized patients: a critique of a meta-analysis. *Med. Gen. Med.* (2000) 6:E3.
3. EINARSON TK: Drug-related hospital admissions. *Ann. Pharmacother.* (1993) 27:832-840.
4. BATES DW, CULLEN DJ, LAIRD N *et al.*: Incidence of adverse drug events and potential adverse drug events: implications for prevention. ADE Prevention Study Group. *JAMA* (1995) 274:29-34.
5. BATES DW, SPELL N, CULLEN DJ *et al.*: The cost of adverse drug events in hospitalized patients. *J. Am. Med. Assoc.* (1997) 277:307-311.
6. RAWLINS MD, THOMSON JW. Mechanisms of adverse drug reactions. In: *Textbook of Adverse Drug Reactions.* DM Davies (Ed.), Oxford University Press, Oxford, UK (1991):18-45.
7. PARK BK, PIRMOHAMED M, KITTERINGHAM NR: Role of drug disposition in drug hypersensitivity: a chemical, molecular, and clinical perspective. *Chem. Res. Toxicol.* (1998) 11:969-988.
8. MEYER UA: Pharmacogenetics and adverse drug reactions. *Lancet* (2000) 356:1667-1671.
9. MEYER UA, GUT J: Genomics and the prediction of xenobiotic toxicity. *Toxicology* (2002) 182:463-466.
10. SAUER JM, PONSLER GD, MATTUIZ EL *et al.*: Disposition and metabolic fate of atomoxetine hydrochloride: the role of CYP2D6 in human disposition and metabolism. *Drug Metab. Dispos.* (2003) 31:98-107.
11. GUT J, CHRISTEN U, FREY N *et al.*: Molecular mimicry in halothane hepatitis: biochemical and structural characterization of lipoylated autoantigens. *Toxicology* (1995) 97:199-224.
12. PARK BK, PIRMOHAMED M: Toxicogenetics in drug development. *Toxicol. Lett.* (2001) 120:281-291.
13. PIRMOHAMED M, PARK BK: Genetic susceptibility to adverse drug reactions. *TIPS* (2001) 22:298-305.
14. PIRMOHAMED M, LIN K, CHADWICK D *et al.*: TNFalpha promoter region gene polymorphisms in carbamazepine-hypersensitive patients. *Neurology* (2001) 56:890-896.
15. HERSHEY GK, FRIEDRICH MF, ESSWEIN LA *et al.*: The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. *N. Engl. J. Med.* (1997) 337:1720-1725.
16. CHAN K, HAN XD, KAN YW: An important function of Nrf2 in combating oxidative stress: detoxification of acetaminophen. *Proc. Natl. Acad. Sci. USA* (2001) 98:4611-4616.
17. ENOMOTO A, ITOH K, NAGAYOSHI E *et al.*: High sensitivity of Nrf2 knockout mice to acetaminophen hepatotoxicity associated with decreased expression of ARE-regulated drug metabolizing enzymes and antioxidant genes. *Toxicol Sci.* (2001) 59:169-177.
18. KU NO, LIAO J, CHOU CF *et al.*: Susceptibility to hepatotoxicity in transgenic mice that express a dominant-negative human keratin 18 mutant. *J. Clin. Invest.* (1996) 98:1034-1046.
19. MATHIJSSEN RH, MARSH S, KARLSSON MO *et al.*: Irinotecan pathway genotype analysis to predict pharmacokinetics. *Clin. Cancer Res.* (2003) 9:3246-3253.
20. DANOFF TM, CAMPBELL DA MCCARTHY LC *et al.*: Gilbert's syndrome UGT1A1 variant confers susceptibility to tranilast-induced hyperbilirubinanemia. *Pharmacogenomics J.* (2004) 4:49-53.
21. ROSES A: Pharmacogenetics and drug development: the path to safer and more effective drugs. *Nat. Rev. Genet.* (2004) 5:645- 656.
22. MCLEOD HL, KRYNETSKI EY, RELLING MV *et al.*: Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastoic leukemia. *Leukemia* (2000) 14:567-572.
23. CHUNG WH, HUNG SI, HONG HS *et al.*: Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* (2004) 428:486.
24. MALLAL S, NOLAN D, WITT C *et al.*: Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* (2002) 359:727-732.
25. HUGHES AR, MOSTELLER M, BANSAL AT *et al.*: Association of genetic variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. *Pharmagenomics* (2004) 5:203-211.
26. AAMODT A, PLAZA E: Case-base reasoning: foundational issues, methodological variations, and system approaches. In: *AICom-Artificial Intelligence Communications.* A Aamodt, E Plaza (Eds), IOS Press, Amsterdam, The Netherlands (1994):39-59.
27. HUG H, BAGATTO D, DANNECKER R *et al.*: ADRIS – The adverse drug reactions

information scheme. *Pharmacogenetics (*2003) 13:767-722.

28. HUG H, DANNECKER R, SCHINDLER R *et al.*: Ontology-based knowledge management of troglitazone-induced hepatotoxicity. *Drug Discov. Today* (2004) 9:948-954.
29. EKINS S, KIRILLOV E, RAKHMATULIN EA *et al.*: A novel method for visualizing nuclear hormone receptor networks relevant to drug metabolism. *Drug Metab. Dispos.* (2005) 33:474-481.
30. NOY NF, CRUBEZY M, FERGERSON RW *et al.*: Protege-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Ann. Symp Proc.* (2003) 953.
31. SHANNON P, MARKIEL A, OZIER O *et al.*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 132:498-504.
32. LECOEUR S, ANDRE C, BEAUNE PH: Tienilic acid-induced autoimmune hepatitis: anti-liver and anti kidney microtonal type 2 autoantibodies recognize a three-site conformational epitope on cytochrome P4502C9. *Mol. Pharmacol.* (1996) 50:326-333.
33. OBERMAYER-STRAUB P, STRASSBURG CP, MANNS MP: Target proteins inhuman autoimmunity: cytochromes P450 and UDP-glucuronosyltransferases. *Can. J. Gastroenterol.* (2000) 14:429-439.
34. LOPEZ-GARCIA MP, DANSETTE PM, MANSUY D: Thiophene derivatives as new mechanism-based inhibitors of cytochrome P450: inactivation of yeast-expressed human liver cytochrome P450 2C9 by tienilic acid. *Biochemistry* (1994) 33:166-175.
35. KUMASHIRO R, KUBOTA T, KOGAY *et al.*: Association of troglitazone-induced liver injury with mutation of the cytochrome P450 gene. *Hepatol. Res.* (2003) 26:337-342.
36. MAYATEPEK E, FLOCK B, ZSCHOKE J: Benzydamine metabolism *in vivo* is impaired in patients with deficiency of flavin-containing monooxygenase 3. *Pharmacogenetics* (2004) 11:775-777.
37. O'LEARY KA, EDWARDS RJ, TOWN MM *et al.*: Genetic and other sources of variation in the activity of serum paraoxonase/diazonase in humans: consequences for risk from exposure to diazinon. *Pharmacogenet. Genomics* (2005) 15:51-60.
38. McKEOWN-EYSSEN G, BAINES C, COLE DEC *et al.*: Case-control study of genotypes in multiple chemical sensitivity: CYPD26, NAT1, NAT2, PON1, PON2, and MTHFR. *Int. J. Epidemiol.* (2004) 33:971-978.
39. WAN J, SHI J, HUI L *et al.*: Association of genetic polymorphisms in CYP2E1, MPO, NQO1, GSTM1, and GSTT1 genes with benzene poisoning. *Environ. Health Persp.* (2002) 110:1213-1218.
40. LAN Q, ZHANG L, LI G *et al.*: Hematotoxicity in workers exposed to low levels of benzene. *Science* (2004) 306:1774-1776.
41. KAPUT J, RODRIGUEZ RL: Nutritional genomics: the next frontier in the postgenomic era. *Physiol. Genomics* (2004) 16:166-177.

Websites

101. http://www.roche-diagnostics.com/products_services/amplichip_cyP450.html AmpliChip CYP450 Test. Accessed May 2005.
102. http://www.fda.gov/medwatch/SAFETY/2004/safety04.htmStrattera Strattera. Accessed May 2005.
103. http://www.prometheuslabs.com/212a.asp?nav=products PRO-PredictRxTPMT. Accessed May 2005.
104. http://java.sun.com Sun Developer Network. Accessed May 2005.
105. http://httpd.apache.org Apache HTTP Server Project. Accessed May 2005.
106. http://www.ncbi.nlm.nih.gov/ National Center for Biotechnology Information. Accessed May 2005.
107. http://www.ebi.ac.uk/embl/ EMBL Nucleotide Sequence Database. Accessed May 2005.
108. http://www.expasy.org/sprot/ SwissProt protein knowledge. Accessed May 2005.
109. http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff. Accessed May 2005.
110. http://www.ncbi.nlm.nih.gov/omim/ Online Mendelian Inheritance in Man. Accessed May 2005.
111. http://bind.ca/ Biomolecular Interaction Network Database. Accessed May 2005.
112. http://www.pharmgkb.org/ The Pharmacogenetics and Pharmcogenomics Knowledge Database. Accessed May 2005.
113. http://www.genego.com/ Systems Biology for Drug Discovery. Accessed May 2005.
114. http://www.ingenuity.com/ Ingenuity. Accessed May 2005.
115. http://www.iconixpharm.com/ Iconix Pharmaceuticals: Bridging Genomics and Chemistry. Accessed May 2005.
116. http://protege.stanford.edu/ Protégé. Accessed May 2005.
117. http://www.cytoscape.org/ Cytospace. Accessed May 2005.

Affiliation

Joseph Gut[†1], Dr, Prof & Dario Bagatto[2]
[†]Author for correspondence
[1]TheraSTrat AG, Gewerbestrasse 25, 4123 Allschwil, Switzerland
Tel: +41 614 855 010; Fax: +41 614 855 029;
E-mail: joseph.gut@jgut.com
[2]TheraSTrat AG, Gewerbestrasse 25, 4123 Allschwil, Switzerland
E-mail: dario.bagatto@therastrat.com