Features • PERSPECTIVE

feature

# Molecular clinical safety intelligence: a system for bridging clinically focused safety knowledge to early-stage drug discovery – the GSK experience

Dana E. Vanderwall[1,7], Nancy Yuen[1,*], Mohammad Al-Ansari[2,7], James Bailey[3], David Fram[2,8], Darren V.S. Green[4], Stephen Pickett[4], Giovanni Vitulli[5], Juan I. Luengo[6] and June S. Almenoff[1,9]

**Drug toxicity is a major cause of late-stage product attrition. During lead identification and optimization phases little information is typically available about which molecules might have safety concerns. A system was built linking chemistry, preclinical and human safety information, enabling scientists to lever safety knowledge across multiple disciplines. The system consists of a data warehouse with chemical structures and chemical and biological properties for ~80 000 compounds and tools to access and analyze clinical data, toxicology, in vitro pharmacology and drug metabolism data. Tapping into this safety knowledge enables rapid clinically focused risk assessments of drug candidates. Use of this strategy adds value to the drug discovery process at GSK via efficient triage of compounds based on their potential for toxicity.**

## Introduction

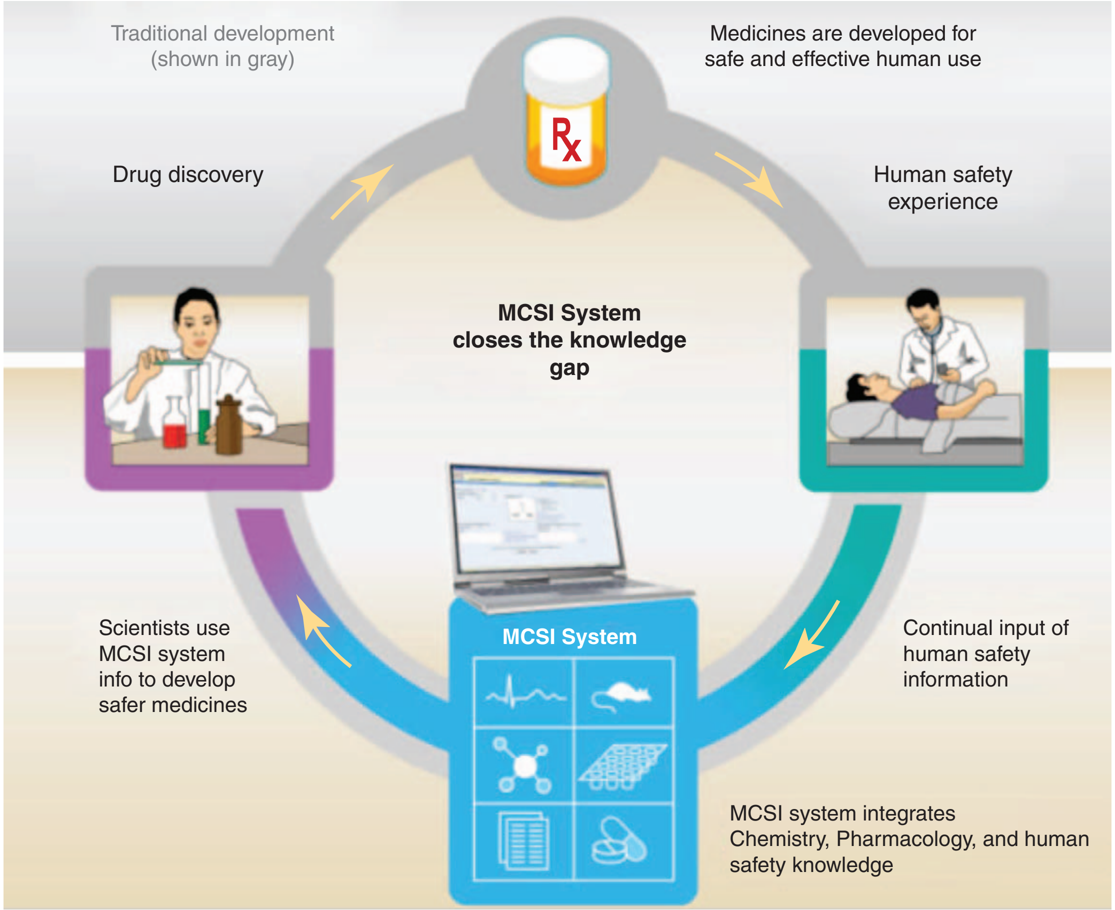
The pharmaceutical industry has made great strides in the discovery of effective medicines, but making early predictions about the safety profile of new medicines remains challenging. Between 1991 and 2000 only one in nine compounds that made it through development was approved by European or US regulatory authorities [1]. Unforeseen adverse reactions can have serious health consequences for patients and can be responsible for serious negative impacts on the productivity of R&D efforts. Although several *in silico* systems have been developed to predict preclinical toxicology and drug metabolism [2–4], there is a significant unmet need for tools that can help predict the human safety profile of novel drug candidates. A data mining approach linking chemical substructures to adverse drug reactions has been previously described [5]. The molecular clinical safety intelligence (MCSI) system described here is a first-in-class prototype system that harnesses a much broader range of knowledge, encompassing the body of existing preclinical, chemistry, toxicology, metabolism, pharmacology and clinical safety information, to improve the detection and management of clinical safety risk in early drug discovery.

## Closing the knowledge gap that exists in traditional drug discovery

New compounds in the drug discovery pipeline are first carefully tested in the laboratory and in animals, and if no obvious safety problems are found using these models the compounds are progressed to human clinical studies. Drugs found to be safe and effective in human clinical studies are then introduced to the market where they can be widely prescribed. As shown in Fig. 1, the traditional flow of knowledge in drug discovery has been one-way, without systematic feedback of the large body of human safety data to early phases of drug discovery.

As illustrated by the complete cycle shown in Fig. 1, MCSI was developed to close this knowledge gap between early drug discovery and human safety experience by capturing human safety data for thousands of drugs and adverse events and integrating it with chemistry, toxicology and pharmacology data. This enables scientists working in early drug discovery to have

1359-6446/06/$ - see front matter © 2011 Elsevier Ltd. All rights reserved. doi:10.1016/j.drudis.2011.05.001


Traditional development (shown in gray)
Medicines are developed for safe and effective human use
Drug discovery
Human safety experience
MCSI System closes the knowledge gap
MCSI System
Scientists use MCSI system info to develop safer medicines
Continual input of human safety information
MCSI system integrates Chemistry, Pharmacology, and human safety knowledge

Drug Discovery Today

**FIGURE 1**

The MCSI system. Currently, the cumulative knowledge associated with a drug candidate flows through the pipeline in one direction, from the early to late phases, with little systematic feedback to early drug discovery. With scientists in different disciplines separated by as many as 10 years along the progression path, this leaves a gap between the clinical knowledge and the next generation of medicines. MCSI creates a feedback loop that closes the knowledge gap between early drug discovery and human safety experience.

access to practical tools to draw directly from this knowledge base and shape the human safety profile of the next generation of drugs.

## Overview of the approach: broad data integration with a focused delivery to the target audience

The MCSI software was developed through a collaborative effort between GlaxoSmithKline (GSK) and the Phase Forward Lincoln Safety Group. The objective was to create a software system that helps minimize safety-related attrition of clinical candidates by linking, and making available for quantitative analysis and modeling, a broad array of data about tens of thousands of drugs and drug-like compounds. A schematic overview of the system architecture is shown in Fig. 2.

The core of the system is a specially engineered data warehouse integrating public and internal data on: chemical structures, chemical descriptors and physicochemical properties; drug metabolism, pharmacology and toxicology; and quantitative and qualitative human safety data (examples of the data are illustrated below). Surrounding this core is a suite of query, visualization and modeling tools that can be applied to the data in the warehouse to identify potential safety liabilities for new compounds and to inform the understanding of observed toxicities during preclinical and clinical testing.

The data warehouse contains safety-relevant information on ~80 000 drug and drug-like compounds from three primary, and initially highly disparate, streams:

- Proprietary data consisting of:
  - i. *in vitro* pharmacology profiling of marketed drugs generated using the GSK internal compound screening platform
  - ii. physicochemical properties and ADME models [6,7] calculated using GSK-internal algorithms
  - iii. a GSK-internal library of chemical structural alerts
- Pharmacological, metabolic and toxicological data abstracted from >50 000 peer-reviewed articles and reference sources
- Quantitative human safety data derived from databases of spontaneously reported adverse drug reactions

To populate the first stream of proprietary data, GSK mounted a systematic effort to screen 1800 marketed drugs through pharmacological target assays available across GSK discovery research.[a] In addition, several *in silico* models

[a] Sets of marketed drugs were screened in all available GSK macro-molecular target assays, including targets in the drug discovery portfolio, as well as targets associated with a known safety liability. Every compound was tested in every assay at a single concentration, followed by a determination of full dose–response curves for any compounds active above a threshold determined to be appropriate for each assay. In total, dose-response curves were determined for 54 000 compound-target combinations.

Features • PERSPECTIVE

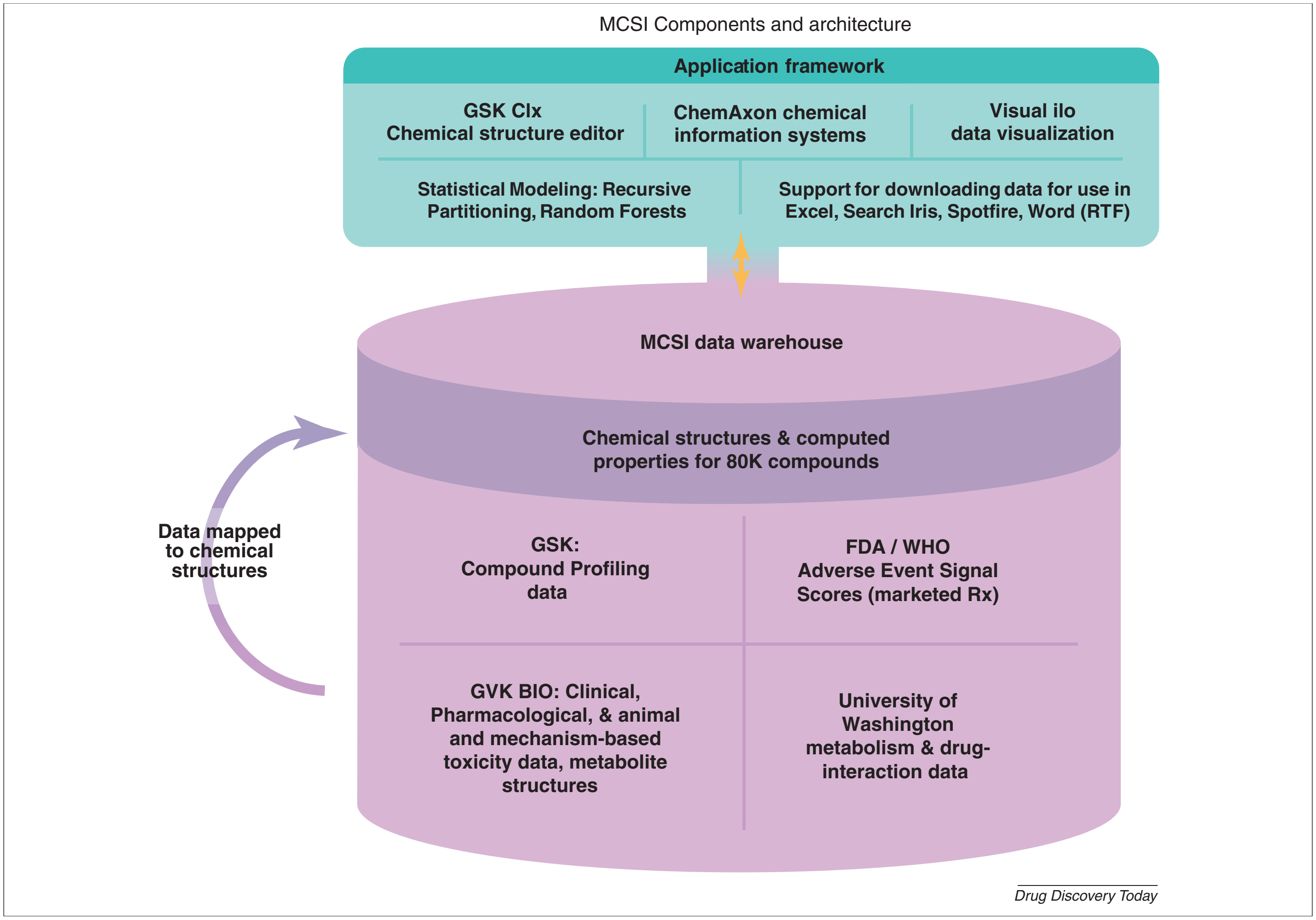

MCSI Components and architecture
Application framework
GSK Clx Chemical structure editor
ChemAxon chemical information systems
Visual ilo data visualization
Statistical Modeling: Recursive Partitioning, Random Forests
Support for downloading data for use in Excel, Search Iris, Spotfire, Word (RTF)
MCSI data warehouse
Chemical structures & computed properties for 80K compounds
Data mapped to chemical structures
GSK: Compound Profiling data
FDA / WHO Adverse Event Signal Scores (marketed Rx)
GVK BIO: Clinical, Pharmacological, & animal and mechanism-based toxicity data, metabolite structures
University of Washington metabolism & drug-interaction data
Drug Discovery Today

**FIGURE 2**

MCSI system architecture.

were used to calculate physicochemical and pharmacokinetic parameters important for improving drug 'developability' [8]. Finally, the compounds in MCSI were indexed against a library of chemical fragments associated with toxicity liability that GSK staff had previously compiled from the literature.

To populate the second stream (i.e. pharmacological, metabolic and toxicological data), GSK contracted with GVK BIO, a discovery-stage contract research organization (CRO) that assembles and curates such data on marketed compounds, clinical candidate compounds and other drug-like compounds from the published literature (i.e. journal articles, patents, and drug labels) http://www.gvkbio.com/informatics.html (Accessed April 2011).

To populate the third stream (i.e. quantitative human safety data), GSK made use of the disproportionality analysis (DPA) scores generated by the Empirica$^{TM}$ signal safety data mining software. This safety data mining software operates over large-scale spontaneous reporting system databases, including the FDA Adverse Event Reporting System (AERS) database and the World Health Organization (WHO) Vigibase database [9–12]. The DPA scores provide a quantitative estimate, based on relative reporting rate, of the strength of association between each given drug and a given adverse event term; their inclusion is fundamental to MCSI, enabling identification and modeling of relationships between empirically observed human safety outcomes and the biological or chemical properties of drug compounds.

A major engineering effort was required to integrate the information from these disparate sources into a 'coherent whole'. The central element tying all of the information together is unique chemical structure: from a given chemical structure one can navigate to all of the information associated with the corresponding drug compound, whether it originated from internal pharmacological screening, abstracted published literature or DPA scores from adverse event databases. Because compounds were represented by various flavors of SMILES or Molefiles in their respective databases or, in the case of AERS and Vigibase, simply as generic drug names, significant challenges were faced in standardizing representations, matching drugs and compounds, and ultimately representing their chemical structures in a format that is compatible with the front- and back-end chemically aware software components used by the system. Manual curation was required to resolve discrepancies in chemical structure for the same chemical, as well as possible synonyms for drug names, species, targets or toxicology terms, and to ascertain that the same term in different data sources had the same meaning.

To fully realize the potential of this large, integrated body of data a suite of query, visualization and modeling tools was built to support specifically targeted queries and analyses, as well as the flexible exploration of the data. Because an innovative drug implies a novel chemical structure, one of the primary means of querying the system is to draw the chemical structure of a compound (or of an idea for a compound) and look for any compounds in the database that are similar in structure (whole or partial) – the most common query a medicinal chemist poses in all systems or tools. The data associated with any of the 'neighbor' compounds subsequently retrieved by this type of query provide insight, by association, into the potential properties or toxicities of the novel compound used as a query probe. Thus, the system can be interrogated using compounds (or ideas for compounds prior to synthesis) from any phase of development.

Query and analysis results are presented as interactive, scrollable tables and graphs that offer point-and-click drilldown to full compound details. After a list of compounds related to the novel compound has been generated the following questions can be evaluated:

- Are there compounds with similar structures that have associated safety concerns?
- Which particular structural features appear most associated with a particular adverse event and/or toxicity?
- Is the biological target for a novel compound associated with adverse events?
- Is there evidence that the novel compound could be metabolized to toxic intermediates? Are there options to potentially block this metabolism?
- Is the novel compound likely to have unanticipated biological effects (i.e. off-target activity)?

It is also possible to construct queries, build datasets and examine the properties for all drugs that:

- Exhibit a particular adverse event
- Work via a particular mechanism of action or have affinity for a particular pharmacological target
- Have specific physicochemical or pharmacokinetic properties.

## Applications in drug discovery

### *Structural similarity to compounds with safety concerns*

The most common usage scenario for MCSI is to start out with a compound of interest and then to search for structurally related compounds for which there exist known safety or toxicity data. The structural relationship can be based on overall structural similarity or by inclusion of a specified fragment of the compound of interest (i.e. a 'substructure search'). The basic premise is that structurally similar compounds can have similar chemical or biological properties. This 'similar property principle' is the basis of the science and methods behind the field of computational chemistry [13]. If a user of MCSI identifies compounds with documented toxicity that have a clear relationship to the user's novel compound then the user has learned about important potential chemical or biological properties of the novel compound, which might not have been previously anticipated or tested for. This is an exploratory approach intended to generate hypotheses, not a prediction based on a statistical model with an estimated measure of confidence. All information identified in the analysis of query results has to be carefully interpreted in the context of all additional knowledge, and with the relevant content and therapeutic area experts engaged. This can be valuable at the lead identification stage to assist in the prioritization of potential chemical series for further testing and elaboration, when there is typically a dearth of data available with which to distinguish among them. In a lead optimization program fragments or chemical motifs with a history of safety or stability issues can be avoided in the design of new molecules.

The example in Fig. 3a illustrates a query designed to investigate the core structure containing a thio-urea within the 6-membered ring (see TQZ-001). This type of query can be drawn with more or less restriction on the definition of the probe chemical structure, to look more broadly at related compounds or to enable a more focused investigation. In the example shown the list of identified compounds contains a marketed drug, propylthiouracil, which is, in fact, a very close match for the original structure of interest (Fig. 3b). The data associated with propylthiouracil within the MCSI system reveal metabolic transformation of the core structure of propylthiouracil to a reactive metabolite, and subsequent toxicity and adverse events associated with that metabolite. MCSI contains a pointer to a study that demonstrates formation of oxidative metabolites of the sulfur [14]. It also points to a study that demonstrates that these metabolites can react in a non-specific way with cysteine residues on the surface of proteins (by way of a nucleophilic attack by the cysteine thiol on the labile oxidized sulfur of the drug metabolite) [15]. Further, MCSI points to another informative study in which protein adducts of the metabolite have been detected and linked to subsequent immune-sensitization and toxicities [16]. Finally, MCSI contains the information that reveals manifestations of these toxicities were clinically observed in adverse events reported in post-marketing databases.

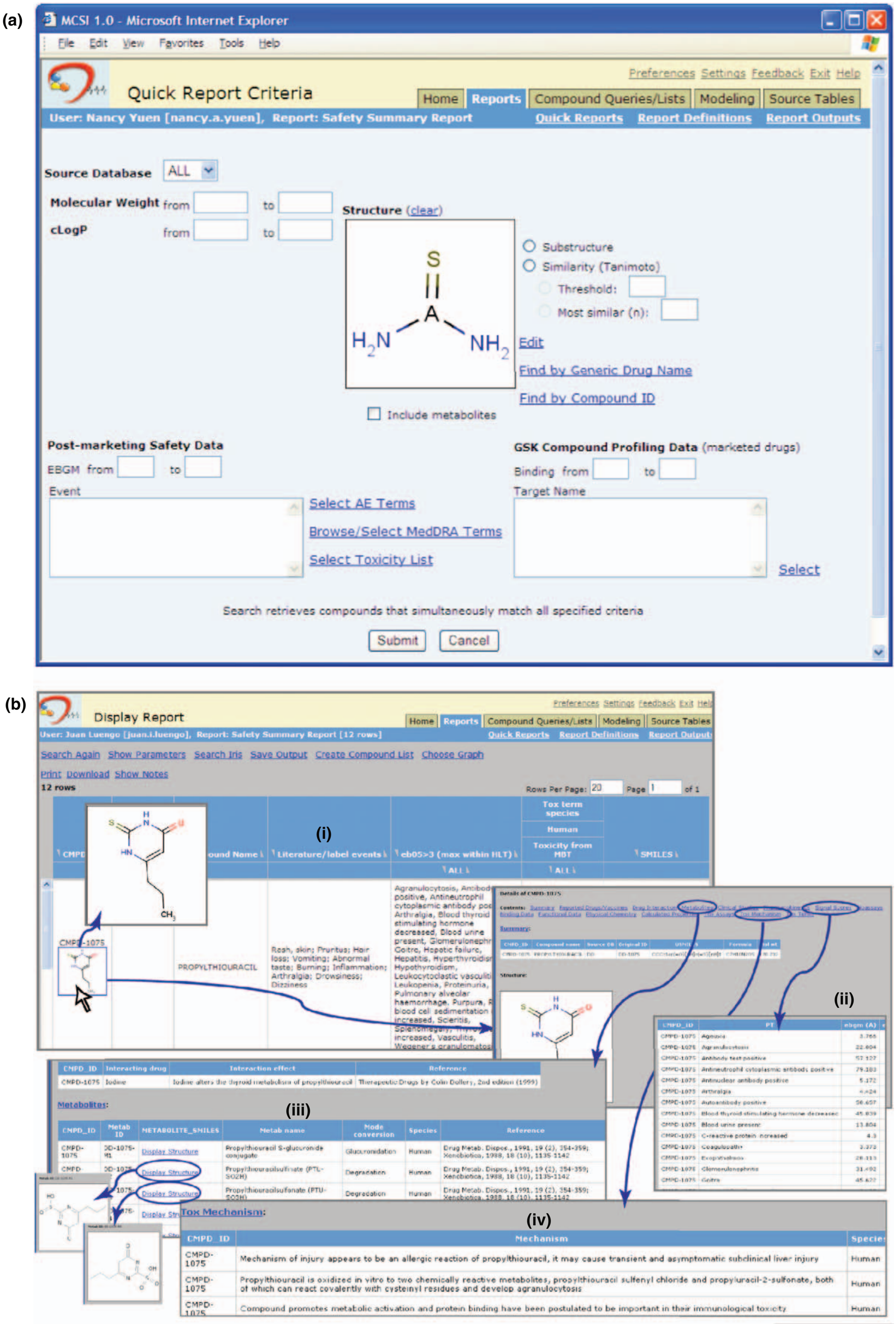
The information relating to propylthiouracil retrieved by the MCSI query, after consideration by the program team chemists along with scientists in preclinical drug metabolism, pharmacokinetics and safety assessment, could assist in the design of experimental follow-up to confirm the potential metabolism or toxicity of the novel compound(s) in development, or suggest specific kinds of monitoring of future data in the program for emerging evidence of similar issues. In the historical case in question at GSK the program team working on the chemical series was at a very early stage in the discovery process, so the information uncovered by MCSI led to the decision to discontinue further work on this chemical series and to shift resources toward alternative series. Although the marketed drug propylthiouracil and its associated safety issues were familiar to personnel in clinical safety, the medicinal chemists in early drug discovery were not familiar with either the drug or the studies characterizing the metabolic-induced liability. This further highlights the value of a system that encompasses thousands of drugs and clinical candidates and spans data and expertise across multiple and diverse scientific and medical domains.

### Identification of off-target activity

Although access to a broad cross-screening capability is often available either internally, or as an outsourced capability, few of the two million compounds that comprise the GSK HTS library will have been profiled across the full range of targets for which assays exist. Consequently, the ability to gain insight into possible secondary target activities for compounds identified as active (i.e. primary) in a HTS exercise would contribute to objective prioritization of chemical series worthy of resource commitment. The data integrated in MCSI include biochemical assay data from two types of sources: (i) assay data extracted from peer-reviewed literature and (ii) assay data generated at GSK in profiling a set of marketed drugs [17].

If the result of a query reveals compounds with similar chemical structure for which biological assay data are present in the MCSI database, these data can be interpreted as suggesting possible biological activities for our probe compound. There are limits to the interpretation of these data, given the lack of enough data to establish a true structure–activity relationship. Nevertheless, if the similarity between the molecules is judged reasonable by an expert in medicinal chemistry, and the documented alternate activity could impact the development of that chemical series, then that activity could be validated experimentally.

An example below illustrates the result of using compounds identified by HTS for a


(a)
MCSI 1.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Preferences Settings Feedback Exit Help
Home Reports Compound Queries/Lists Modeling Source Tables
Quick Reports Report Definitions Report Outputs
S
A
$H_2N$
$NH_2$
Edit
Find by Generic Drug Name
Find by Compound ID
Select AE Terms
Browse/Select MedDRA Terms
Select Toxicity List
Select
Submit
Cancel
(b)
Preferences Settings Feedback Exit Help
Home Reports Compound Queries/Lists Modeling Source Tables
Quick Reports Report Definitions Report Outputs
Search Again Show Parameters Search Iris Save Output Create Compound List Choose Graph
Print Download Show Notes
(i)
(ii)
(iii)
(iv)
S
H
N
O
HN
$CH_3$
HO
OH
Display Structure

Quick Report Criteria

User: Nancy Yuen [nancy.a.yuen], Report: Safety Summary Report

Source Database ALL

Molecular Weight from ___ to ___

cLogP from ___ to ___

Structure (clear)

○ Substructure
○ Similarity (Tanimoto)
○ Threshold: ___
○ Most similar (n): ___

□ Include metabolites

Post-marketing Safety Data

EBGM from ___ to ___

Event

GSK Compound Profiling Data (marketed drugs)

Binding from ___ to ___

Target Name

Search retrieves compounds that simultaneously match all specified criteria

Display Report

User: Juan Luengo [juan.i.luengo], Report: Safety Summary Report [12 rows]

12 rows

Rows Per Page: 20 Page 1 of 1

| CMPD | | ound Name | Literature/label events | eb05>3 (max within HLT) | Tox term species<br>Human<br>Toxicity from MBT | SMILES |
|---|---|---|---|---|---|---|
| | | | | ALL | ALL | |
| CMPD-1075 | | PROPYLTHIOURACIL | Rash, skin; Pruritus; Hair loss; Vomiting; Abnormal taste; Burning; Inflammation; Arthralgia; Drowsiness; Dizziness | Agranulocytosis, Antibod positive, Antineutrophil cytoplasmic antibody pos Arthralgia, Blood thyroid stimulating hormone decreased, Blood urine present, Glomerulonephr Goitre, Hepatic failure, Hepatitis, Hyperthyroidisr Hypothyroidism, Leukocytoclastic vasculiti Leukopenia, Proteinuria, Pulmonary alveolar haemorrhage, Purpura, R blood cell sedimentation increased, Scleritis, Splenomegaly, Thyro increased, Vasculitis, Wegener's granulomatos | | |

Details of CMPD-1075

Contents: Summary Reported Drugs/Vaccines Drug Interaction Metabolites Clinical Studies Pharmacokinetics Signal Scores Assays Binding Data Functional Data Physical Chemistry Calculated Properties Tox Assays Tox Mechanism Tox Terms

Summary:

| CMPD_ID | Compound name | Source DB | Original ID | USMILES | Formula | Mol wt |
|---|---|---|---|---|---|---|
| CMPD-1075 | PROPYLTHIOURACIL | DD | DD-1075 | CCCc1cc(=O)[nH]c(=S)[nH]1 | C7H10N2OS | 170.230 |

Structure:

| CMPD_ID | PT | ebgm (A) |
|---|---|---|
| CMPD-1075 | Ageusia | 3.766 |
| CMPD-1075 | Agranulocytosis | 22.004 |
| CMPD-1075 | Antibody test positive | 57.127 |
| CMPD-1075 | Antineutrophil cytoplasmic antibody positive | 79.183 |
| CMPD-1075 | Antinuclear antibody positive | 5.172 |
| CMPD-1075 | Arthralgia | 4.424 |
| CMPD-1075 | Autoantibody positive | 56.657 |
| CMPD-1075 | Blood thyroid stimulating hormone decreased | 45.839 |
| CMPD-1075 | Blood urine present | 13.804 |
| CMPD-1075 | C-reactive protein increased | 4.3 |
| CMPD-1075 | Coagulopathy | 3.373 |
| CMPD-1075 | Exophthalmos | 28.113 |
| CMPD-1075 | Glomerulonephritis | 31.492 |
| CMPD-1075 | Goitre | 45.672 |

| CMPD_ID | Interacting drug | Interaction effect | Reference |
|---|---|---|---|
| CMPD-1075 | Iodine | Iodine alters the thyroid metabolism of propylthiouracil | Therapeutic Drugs by Colin Dollery, 2nd edition (1999) |

Metabolites:

| CMPD_ID | Metab ID | METABOLITE_SMILES | Metab name | Mode conversion | Species | Reference |
|---|---|---|---|---|---|---|
| CMPD-1075 | DD-1075-M1 | Display Structure | Propylthiouracil S-glucuronide conjugate | Glucuronidation | Human | Drug Metab. Dispos., 1991, 19 (2), 354-359; Xenobiotica, 1988, 18 (10), 1135-1142 |
| CMPD- | DD-1075 | Display Structure | Propylthiouracilsulfinate (PTU-SO2H) | Degradation | Human | Drug Metab. Dispos., 1991, 19 (2), 354-359; Xenobiotica, 1988, 18 (10), 1135-1142 |
| | -1075- | Display Structure | Propylthiouracilsulfonate (PTU-SO3H) | Degradation | Human | Drug Metab. Dispos., 1991, 19 (2), 354-359; Xenobiotica, 1988, 18 (10), 1135-1142 |
| | 75- | Display Str | | | | |

Tox Mechanism:

| CMPD_ID | Mechanism | Specie |
|---|---|---|
| CMPD-1075 | Mechanism of injury appears to be an allergic reaction of propylthiouracil, it may cause transient and asymptomatic subclinical liver injury | Human |
| CMPD-1075 | Propylthiouracil is oxidized in vitro to two chemically reactive metabolites, propylthiouracil sulfenyl chloride and propyluracil-2-sulfonate, both of which can react covalently with cysteinyl residues and develop agranulocytosis | Human |
| CMPD-1075 | Compound promotes metabolic activation and protein binding have been postulated to be important in their immunological toxicity | Human |

*Drug Discovery Today*

**FIGURE 3**

**(a)** MCSI's 'quick report' interface. **(b)** Query results displays top-line safety information for retrieved compounds (i). Drilldown on a compound gives further detail such as quantitative clinical safety scores (ii), metabolites (iii) and toxicity mechanisms (iv).

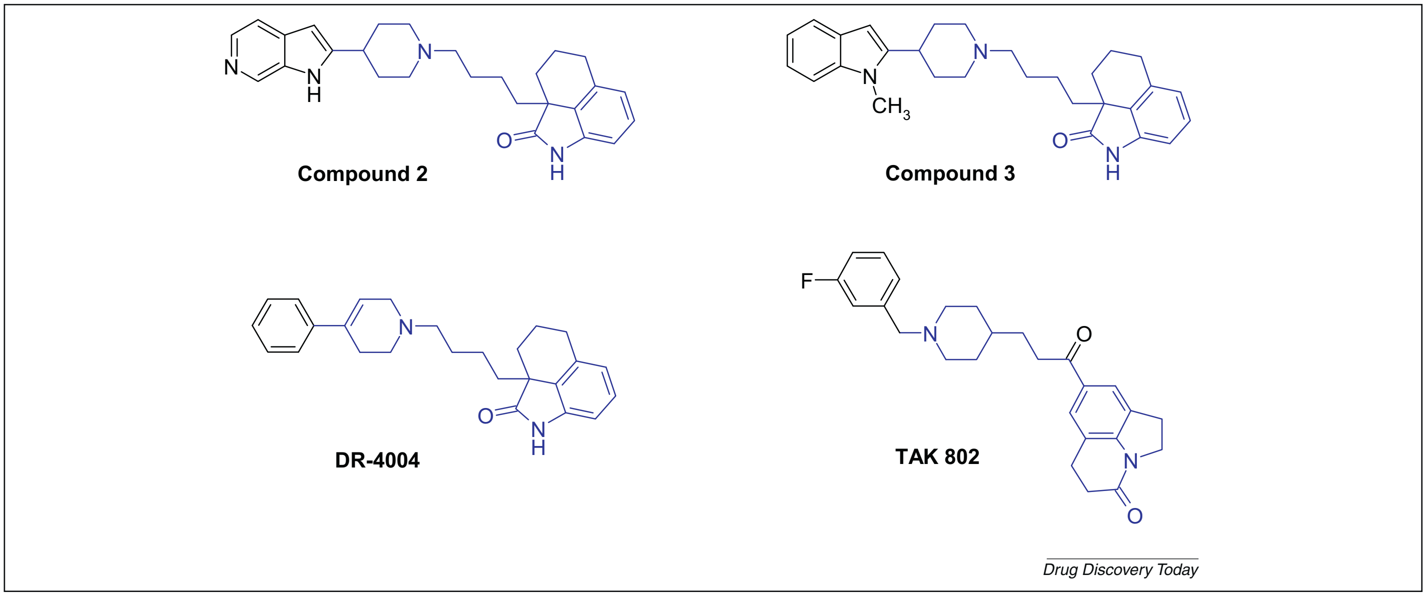


**FIGURE 4**

Compound structures. Use of chemical similarity searches for compounds 2 and 3 'reveal' DR4004 and TAK802, which have off-target activity.

G-protein-coupled receptor (compounds 2 and 3; Fig. 4) as query probes. A query using chemical similarity highlighted the compounds named DR4004 and TAK802 within the MCSI database. Information stored within MCSI indicates that DR4004 is an antagonist of the 5HT-7 and dopamine-2 receptors; it was subsequently shown that similar activities were measured for compounds 2 and 3. MCSI contains the information that TAK802 inhibits human erythrocyte-derived acetylcholinesterase with a $K_i$ of 2 nM. The program team might not otherwise have been able to anticipate this as a potential liability of compounds 2 and 3, because no assay was available in-house for acetylcholinesterase, an activity with important biological implications. However, such an assay could easily be developed and the candidate compounds optimized to eliminate the effect of that off-target activity.

## Identification of potentially toxic metabolites

As a program optimizing early leads eventually progresses to preclinical safety studies the potential for a compound to undergo metabolic transformation is studied. A system like MCSI can be used to study the identified or suspected metabolites to: (i) identify other parent drug structures shown to form similar metabolites and the routes by which they are formed; and (ii) to assess, given other compounds or drugs that form similar metabolites, whether a link between those metabolites and toxicity has been observed. A substructure search using the quinone-imine fragment (shown in Fig. 5) identifies 280 compounds for which that fragment is present as a substructure within a metabolite, as well as 20 compounds where that fragment is part of the parent structure. Nine of the 280 compounds were metabolites of a marketed drug; of all the types of compounds included in MCSI the marketed drugs (with the associated human safety data) are the most valuable subset, and take precedence in the interpretation of a set of results. The remainder of the compounds were preclinical candidates, many of which contained limited data, for example only toxicology data. Of the nine marketed compounds, eight of them displayed a reactive metabolite with associated clinical cytotoxicities, including hepatotoxicity; and for one compound the quinone-imine was part of a larger ring structure. Among the nine compounds were acetaminophen and minocycline, for which liver toxicity is associated with the formation of the quinone-imine-containing metabolite via oxidation by cytochrome P450 subtypes E and D [18,19]. The ability of the system to include the metabolite structures in the chemical search for drugs (and associated data) obviates the need for an individual medicinal chemist to possess an encyclopedic knowledge of the metabolism of all the chemical structures and anticipate the routes by which the same metabolite could be formed from different parent structures.

## Exploring the relationship between adverse events and target activity

The exploration of possible relationships between clinical adverse events and the bioassay-measured target activity of drugs starts with a query of MCSI for a particular adverse event. The user can retrieve all drugs with reported DPA scores for an adverse event as well as the values from measured experimental data from the GSK

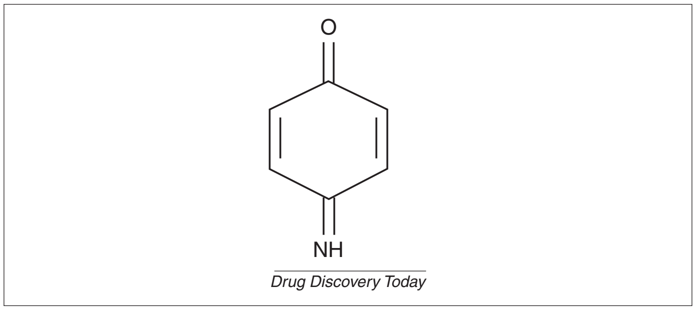
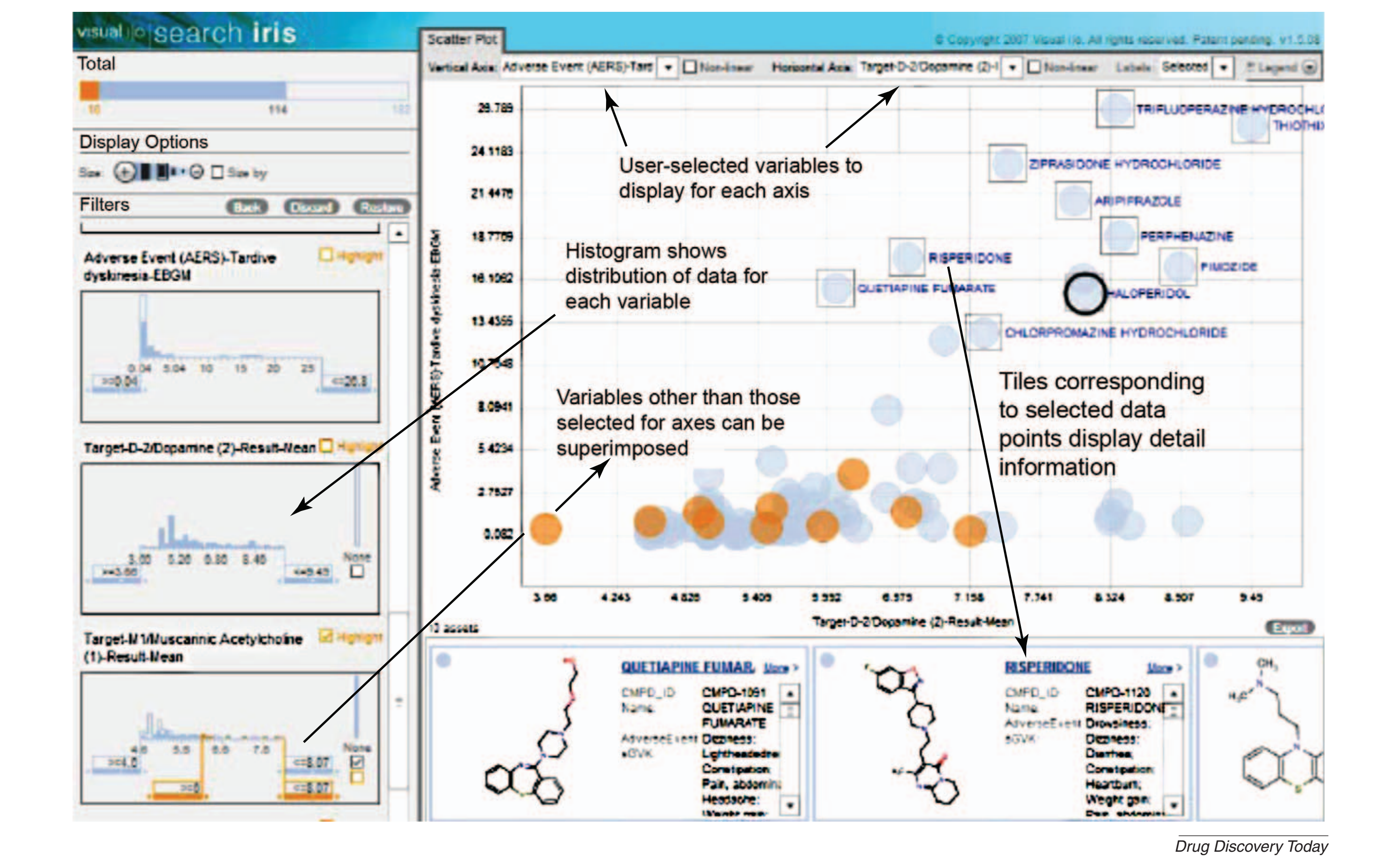


**FIGURE 5**

Quinone imine substructure searched within MCSI.


visual i|o search iris
Total
Display Options
Size
Size by
Filters
Back
Discard
Restore
Adverse Event (AERS)-Tardive dyskinesia-EBGM
Highlight
Target-D-2/Dopamine (2)-Result-Mean
Target-M 1/Muscarinic Acetylcholine (1)-Result-Mean
None
Scatter Plot
© Copyright 2007 Visual i|o. All rights reserved. Patent pending. v1.5.08
Vertical Axis: Adverse Event (AERS)-Tard
Non-linear
Horizontal Axis: Target-D-2/Dopamine (2)-
Labels: Selected
Legend
Adverse Event (AERS)-Tardive dyskinesia-EBGM
28.789
24.1183
21.4476
18.7769
16.1062
13.4355
10.7648
8.0941
5.4234
2.7527
0.082
3.66
4.243
4.826
5.409
5.992
6.575
7.158
7.741
8.324
8.907
9.49
Target-D-2/Dopamine (2)-Result-Mean
TRIFLUOPERAZINE HYDROCHL
THIOTHI
ZIPRASIDONE HYDROCHLORIDE
ARIPIPRAZOLE
PERPHENAZINE
RISPERIDONE
PIMOZIDE
QUETIAPINE FUMARATE
HALOPERIDOL
CHLORPROMAZINE HYDROCHLORIDE
13 assets
Export
QUETIAPINE FUMAR
More >
CMPD_ID CMPD-1091
Name QUETIAPINE FUMARATE
AdverseEvent Dizziness; Lightheadedne; Constipation; Pain, abdomin; Headache
RISPERIDONE
CMPD_ID CMPD-1120
Name RISPERIDONE
AdverseEvent Drowsiness; Dizziness; Diarrhea; Constipation; Heartburn; Weight gain
CH₃

User-selected variables to display for each axis

Histogram shows distribution of data for each variable

Variables other than those selected for axes can be superimposed

Tiles corresponding to selected data points display detail information

Drug Discovery Today

**FIGURE 6**

On-demand visualization of relationships between a quantitative clinical adverse event and receptor binding with simple drill-down access to molecular and other relevant information.

internal biochemical target assays. An on-demand visualization tool in MCSI using SearchIRIS from Visual i|o[b] enables one to explore such potential relationships. An illustration of the relationship between dopamine receptor D2 target activity and the adverse event tardive dyskinesia is shown in Fig. 6.

The integration of *in vitro* target data with clinical adverse event data establishes a tangible link between human safety knowledge and the GSK experimental platforms. This could be an area of potential applicability of MCSI to biotherapeutics. If there is a relationship between a pharmacological activity and an adverse effect the drugs that target this particular pharmacological activity *in vitro* (through any modality) and the associated clinical adverse event can serve as benchmarks against which novel compounds can be screened for toxicity potential.

## Concluding remarks

MCSI breaks with conventional processes and information flow in two important ways: (i) by providing tools that enable the analysis of potential safety issues of drug molecules in the early phases of the drug discovery process; and (ii) by providing a way to link human safety information to the experimental platforms used in early drug discovery. In conventional drug discovery and development across the industry, hundreds of variants of new drug molecules are evaluated before a small number of candidates are selected for subsequent laboratory safety testing in cells and animals. Much of the early evaluation relies on scientists' prior experience and domain knowledge of the therapeutic area. MCSI offers medicinal chemists and biologists tools to connect the structure of new drug candidates with vast and distributed sources of data relevant for understanding their potential human safety impact, and it rapidly informs GSK scientists of potential safety implications of new drug molecules – often before they are even synthesized, and well before they reach the stage of laboratory safety testing.

The examples described show early promise of an approach to safety-driven drug design at GSK based on providing previously unavailable knowledge and analytical tools to early-stage drug development. Possible future enhancements include access to more complete datasets through potential collaboration with other organizations, and integration with other tools in use in the research environment.

A further benefit was the fostering of collaborative problem solving across disciplines that have not generally interacted to date. Traditional drug development involves the hand-off of drug molecules from medicinal chemists into preclinical testing, followed by a transition to clinical study. The expertise of clinically oriented medicine or pharmacology is generally not tapped during very early development by medicinal chemists. Interaction between medicinal chemists, drug metabolism experts and clinicians was increased; each area brings acuity to the interpretation of data from the MCSI tool, for the

[b] SearchIRIS is a trademark of Visual i|o. Product image © Copyright 2002–2008 Visual i|o.

overall goal of reducing the attrition of potential drugs in development.

## Acknowledgments

The authors thank and acknowledge members of the GSK cross functional MCSI user team for their valuable feedback, which has helped the authors optimize the design of the system. MCSI is a research tool developed as a prototype system to manage safety risk in drug development; it is not a commercial or for-profit software system.

## References

1 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev.* 3, 711–715

2 Greene, N. *et al.* (1999) Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR, and METEOR. SAR & QSAR. *Environ. Res.* 10, 299–313

3 Matthews, E.J. *et al.* (2009) Identification of structure–activity relationships for adverse effects of pharmaceuticals in humans: part C: use of QSAR and an expert system for the estimation of the mechanism of action of drug-induced hepatobiliary and urinary tract toxicities. *Regul. Tox. Pharmacol.* 54, 43–65

4 Gomba, V.K. *et al.* (1995) Assessment of developmental toxicity potential of chemicals by quantitative structure–toxicity relationship models. *Chemosphere* 31, 2499–2510

5 Schreiber, J. *et al.* (2009) Mapping adverse drug reactions in chemical space. *J. Med. Chem.* 52, 3103–3107

6 Norinder, U. and Bergstrom, C.A.S. (2006) Prediction of ADMET properties. *Chem. Med. Chem.* 1, 920–937

7 Gleeson, M.P. (2008) Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* 51, 817–834

8 Leeson, P.D. and Springthorpe, B. (2007) The influence of drug-like concepts on decision making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890

9 DuMouchel, W. and Pregibon, D. (2001) Empirical bayes screening for multi-item associations. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp. 67–76

10 Fram, D.M. *et al.* (2003) Empirical Bayesian data mining for discovering patterns in post-marketing drug safety. In *Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp. 359–368

11 Almenoff, J.S. *et al.* (2007) Novel statistical tools for monitoring the safety of marketed drugs. *Clin. Pharmacol. Ther.* 82, 157–166

12 Almenoff, J.S. *et al.* (2007) Online signal management: a systems-based approach that delivers new analytical capabilities and operational efficiency to the practice of pharmacovigilance. *Drug Inf. J.* 41, 779–789

13 Johnson, M. et al. (1989) *Qsar: Quantitative Structure–Activity Relationships in Drug Design* (Fauchere, J.I., ed.), pp. 167–171, Alan R. Liss, New York

14 Lee, E. *et al.* (1988) Oxidative metabolism of propylthiouracil by peroxidases from rat bone marrow. *Xenobiotica* 18, 1135–1142

15 Park, B.K. *et al.* (1998) Role of drug disposition in drug hypersensitivity: a chemical, molecular, and clinical perspective. *Chem. Res. Toxicol.* 11, 969–988

16 Waldhauser, L. and Uetrecht, J. (1991) Oxidation of propylthiouracil to reactive metabolites by activated neutrophils: implications for agranulocytosis. *Drug Metab. Dispos.* 19, 354–359

17 Macarron, R. *et al.* (2011) Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195

18 Losanoff, J.E. *et al.* (2007) Minocycline toxicity requiring liver transplant. *Dig. Dis. Sci.* 52, 3242–3244

19 Harvison, P.J. *et al.* (1998) Cytochrome P-450 isozyme selectivity in the oxidation of acetaminophen. *Chem. Res. Toxicol.* 47, 47–52

*Dana E. Vanderwall[1,7]*
*Nancy Yuen[1,*]*
*Mohammad Al-Ansari[2,7]*
*James Bailey[3]*
*David Fram[2,8]*
*Darren V.S. Green[4]*
*Stephen Pickett[4]*
*Giovanni Vitulli[5]*
*Juan I. Luengo[6]*
*June S. Almenoff[1,9]*

*[1]GlaxoSmithKline, 5 Moore Drive, Mailstop 17.1315K.1C, Research Triangle Park, NC 27709, United States*
*[2]Lincoln Safety Group, Phase Forward, United States*
*[3]EpiNova DPU, ImmunoInflammation CEDD, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY UK*
*[4]Computational and Structural Chemistry, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY UK*
*[5]Respiratory DMPK, GlaxoSmithKline Medicines Research Centre Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY UK*
*[6]Chemistry Cancer Metabolism Oncology, R&D GlaxoSmithKline, 1250 South Collegeville Road UP1110, P.O. Box 5089 Collegeville, PA 19426-0989 United States*

**Corresponding author*
*E-mail: nancy.a.yuen@gsk.com*

*Current affiliations:*
*[7]Oracle Corporation, United States.*
*[8]Commonwealth Informatics, Inc., United States.*
*[9]Furiex Pharmaceuticals, United States.*