Semantic approaches to enable drug discovery in
biomedical big data

by

Tanya Hiebert

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology with Specialization in Bioinformatics

Carleton University
Ottawa, Ontario

© 2014, Tanya Hiebert

Abstract

Drugs such as penicillin and insulin have treated human disease for years, leading to both an improved quality of life, as well as an increase in overall life expectancy. Despite large amounts of biomedical data being available, the data is not being harnessed due to integration issues between datasets and heterogeneity between ontologies. Here, the use of semantic technologies in drug repurposing and drug safety is explored as it would greatly help the pharmaceutical industry answer difficult questions. Hypotheses studied include the use of mappings between model phenotypes and drug effects to identify human drug targets, and using pre-existing data as evidence to profile drug safety. Mappings done manually proved to be of better quality in comparison to automatic mappings. Evaluating drug related cardiotoxicity based on existing data was demonstrated to be successful for drug safety profiling. The results of this work demonstrate the usefulness of using pre-existing data to discover new knowledge.

Acknowledgements

I would like to thank my supervisor Dr. Michel Dumontier for taking me in as a graduate student in his lab and giving me the opportunity to learn informatics skills even though I did not have an informatics background. I would like to thank him for his time, patience and guidance throughout my thesis.

I would also like to thank my committee members, Dr. Myron Smith and Dr. Vance Trudeau for their constructive suggestions and guidance during my thesis.

I would like to acknowledge my lab mates Dr. Jose Toledo Cruz and Dr. Alison Callahan who were great role models, were always willing to help me, and who often shared valuable advice with me.

I would also like to thank my family and friends who supported me throughout my theses, who were always there to listen to me, and who often provided me with moral support when I needed it most.

Contributions

Within the PhenomeDrug approach, my supervisor Michel Dumontier and our collaborators Robert Hoehndorf, Nigel W. Hardy, Paul N. Schofield and Georgios V. Gkoutos have all worked hard and contributed to the success of the approach.

The HyQue-Cardiotoxicity approach made use of HyQue, a semantic web tool created by Dr. Alison Callahan.

Beatriz E. Lujan-Toro, an undergraduate student, participated in the HyQue-Cardiotoxicity project by creating some of the rules which are used to evaluate cardiotoxicity.

Table of Contents

Contents

Abstract.....	ii
Acknowledgements	iii
Contributions.....	iii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
List of Appendices.....	viii
Acronyms	ix
1 Chapter: Introduction.....	1
1.1 Drug development	2
1.2 Data integration	7
1.3 Semantic mappings.....	9
1.4 Hypothesis testing	13
1.5 Thesis and structure	14
1.6 Broader Impact.....	18
2 Chapter: Mapping drug effects to model phenotypes in the PhenomeDrug approach	20
2.1 Abstract	20
2.2 Introduction	20
2.3 Materials and methods.....	28
2.4 Results	32
2.5 Discussion	35
2.6 Conclusion.....	38

3	Chapter: HyQue-Cardiotoxicity	40
3.1	Abstract	40
3.2	Introduction	40
3.3	Materials and methods.....	46
3.4	Results	55
3.5	Discussion	67
3.6	Conclusion.....	75
4	Chapter: Conclusion	76
4.1	Overall summary	76
	Appendices.....	87
	References.....	89

List of Tables

Table 2-1: Examples of mappings made between SIDER UMLS terms and ontologies .	33
Table 2-2: Number of mappings made between ontologies to SIDER UMLS terms.....	33
Table 2-3: Result of the comparison made between the manual and automatic mappings between HP and UMLS SIDER terms.....	34
Table 3-1: Datasets used for evidence in the scope of this project.....	62
Table 3-2: Categorization of evidence scores for drug cardiotoxicity.....	63
Table 3-3: Result of evaluating the evidence for TKI induced cardiotoxicity	63
Table 3-4: Database rules which seem to be responsible for imatinib's higher than expected score.....	64
Table 3-5: Result of evaluating the evidence for non-TKI drugs known to be cardiotoxic inducing cardiotoxicity	64
Table 3-6: Result of evaluating the evidence for non-TKI drugs not identified to be cardiotoxic inducing cardiotoxicity	64
Table 3-7: Comparison of evidence scores for cardiotoxic and non-cardiotoxic drugs ...	65
Table 3-8: Results of evaluations made by HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for cardiotoxic drugs	65
Table 3-9: Results of evaluations made by HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for non-cardiotoxic drugs	66
Table 3-10: Summary of performance of HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for non-cardiotoxic drugs	67

List of Figures

Figure 1-1: Overview of the scope of this thesis	15
Figure 2-1: Role of mappings within the PhenomeDrug approach	27
Figure 2-2: Overview of the rational for the PhenomeDrug approach	28
Figure 3-1: Example of the structure of the cardiotoxic side-effect rule	50
Figure 3-2: Overview of the HyQue-Cardiotoxicity approach	51

List of Appendices

Appendix A - Manual mapping of UMLS SIDER concepts to ontologies.....	87
Appendix B - Comparison of manual vs automatic mappings	87
Appendix C - Evidence rules and associated datasets	87
Appendix D - List of terms used within rule set related to cardiotoxicity.....	87
Appendix E - List of cardiotoxic, non-TKI drugs, and score for HyQue evaluation	87
Appendix F - List of non-cardiotoxic, non-TKI drugs, and score for HyQue evaluation	87
Appendix G - List of cardiotoxic drugs created using DailyMed for method comparison	88
Appendix H - List of non-cardiotoxic drugs created using DailyMed for method comparison.....	88
Appendix I - Reasons for comparison of drugs not working.....	88
Appendix J - Breakdown of scores for HyQue-Cardiotoxicity rules.....	88

Acronyms

ADME (absorption, distribution, metabolism, and excretion)
hERG (human ether-a-go-go-related gene)
HP (human phenotype)
ID (identifier)
LOD (Linked Open Data)
LVEF (left ventricular ejection fraction)
MGI (Mouse Genome Informatics)
MP (mouse phenotype)
OWL (Web Ontology Language)
QSAR (qualitative structure–activity relationships)
RDF (resource description framework)
RDFS (RDF Schema)
SIDER (Side Effect Resource)
SKOS (Simple Knowledge Organization System)
SPARQL (SPARQL Protocol and RDF Query Language)
SMILES (Simplified molecular-input line-entry system)
SPIN (SPARQL inferencing notation)
TKI (tyrosine kinase inhibitor)
UMLS (Unified Medical Language System)
URI (Uniform Resource Identifier)

1 Chapter: Introduction

Recent advances in experimental research such as high throughput technologies has led to large amounts of data being produced (Zhu *et al.*, 2010). However, this biomedically relevant big data requires substantial time and effort to adequately prepare due to issues with identifying, retrieving and integrating data from various data sources (Masseroli *et al.*, 2014). Finding data which may be hidden or distributed in various datasets is currently a challenge such as differences in the data type, the quality of the data, as well as data access issues due to copyright (Williams *et al.*, 2012). Currently, there is a lack of computational approaches available to effectively find and retrieve this dispersed data (Samwald *et al.*, 2011). For example, there are Linked Data browsers which help navigate from one dataset to another which tend to be too generic for specific uses, and issues such as the reduced ability to query different datasets at once due to the heterogeneity between the structure of the data (Samwald *et al.*, 2011). Another issue faced when working with data from different sources is the issue that the data must first be normalized, for example ensuring that the data share the same format prior to computational analysis (Sansone *et al.*, 2012). In response, different attempts to data integration have been made, for example using data warehouses, integrating services together, and the use of federated queries (Zhang *et al.*, 2011). These attempts at integration have been performed within different approaches such as knowledge-representation, databases, and the semantic web (Meilicke & Stuckenschmidt, 2009). Another means of increasing integration is through the use of ontology mapping, which is to find and map equivalent terms between ontologies (Jung, 2010). The activity of generating mappings between terms allows for correspondences to be created between

equivalent terms found in different ontologies which can then be exploited to bridge disparate datasets.

Utilizing data from various datasets would allow for the discovery of new knowledge in the life sciences, such as finding new relationships between pharmaceuticals and genes (Zhu *et al.*, 2010). In particular, drug discovery research could greatly benefit from knowledge obtained through the analysis of large datasets to help answer difficult pharmaceutical related questions (Williams *et al.*, 2012). Solving such data integration problems and exploiting the power of big data could lend itself to both drug discovery and drug safety profiling. This thesis is concerned with aspects of integrating multiple resources for the purpose of drug discovery and drug safety.

1.1 Drug development

Currently the pharmaceutical industry is faced with a high drug attrition rate as only 8% of drugs will ever make it to market. This represents a low success rate for drugs for which the typical investments can cost up to 1.8 billion dollars to develop, and can take 10-15 years before ever reaching the market (Paul *et al.*, 2010). Furthermore, the recent patent expirations of big blockbuster drugs will lead to reduced revenue for Big Pharma to generics, which are estimated to represent roughly 209 billion dollars in sales (Khanna, 2012). The lack of new chemical entities reaching the market has increased the pressure to find leads which will materialize into the next blockbuster drug as attrition is costly not only to the struggling pharmaceutical industry, but for patients awaiting better treatments as well (Holland, 2013).

Drug development is an important process in the pharmaceutical industry. It is the process by which new drugs are discovered or synthesized to be marketed for the use of treating various ailments (Knoop & Worden, 1998). Traditionally, drug development relied on serendipity or the modification of known molecules by chemists to give them desired traits (Ratti & Trist, 2001). Today, the pharmaceutical industry uses a pipeline in order to develop a drug which begins with target identification, followed by target validation to identify leads, clinical studies to assess the lead, and finally marketing the drug for use (Ratti & Trist, 2001).

Drug target identification is the initial process of drug development. Currently, it is done by target based and phenotype based methods (Patel, 2013), as well as in silico methods (Tertappen & Reggiani, 2001). Target based screening involves identifying the target gene likely involved in a disease, and then screening against the identified target with candidate molecules. Target based screening makes use of genome sequencing studies and disease observations to identify genes of interest, as well as high throughput screening to identify lead compounds for the target (Sams-Dodd, 2005). In comparison, phenotypic screening involves the administration of drugs in animal models in order to find a desirable effect on a disease phenotype, thus identifying a possible treatment for disease. This method has the advantage of possibly identifying the mechanism of action behind a drugs action (Patel, 2013).

The preclinical stage of drug discovery makes use of in vitro, in vivo and in silico methods in order to assess drug tolerability, organ-specific toxicity, and dose-response relationships (Roberts *et al.*, 2014). The preclinical stage is important as any indication of toxicity would end the development of the studied drug lead, saving on development

costs which would have been incurred if the lead would have moved onto the very costly clinical stage (Roberts *et al.*, 2014). The clinical stage begins with an application for IND (Investigational New Drug), followed by phase I, phase II and phase III clinical trials, and finally a NDA (new drug application) is filed if the phase III trial is successful (Sataloff, 2008). Phase I clinical trials are performed using a small group of healthy individuals in order to assess safety, as well as to gather information on the pharmacokinetics and pharmacodynamics of a drug candidate (Tonkins, 2005). Phase II consists of testing the drug in a group of patients who have the illness, and allows for the observation of dose-response relationships, and how efficacious a drug is (Tonkins, 2005). Finally, phase III trials consist of testing the drug in a larger group of patients afflicted with the illness under study for studies which tend to last for a longer duration than the two first phases. Phase III allows for a continued study of drug efficacy, as well as weighing the risks and benefits of a drug to determine if a NDA should be filed in order to have the FDA review the information for approval to market (Tonkins, 2005). Drug companies may also opt to continue into Phase IV, which allows them to monitor the usage population for the drug and compare its usage against that of competing brands, monitor reports of adverse events, and follow drug efficacy (Sataloff, 2008).

The majority of drug development failures occur largely due to safety concerns in the earlier stages of development encompassing the preclinical stage of development as well as phase I trials, whereas efficacy is the largest contributor to failures in the later phases of development (Cook *et al.*, 2014). This is supported by an additional study which found that drug attrition occurs mostly in phase I and phase II trials, and the attrition is largely attributed to efficacy (56%) or safety (28%) issues (Arrowsmith &

Miller, 2013). As such, there is a need to improve both the safety and efficacy of new drugs. Several papers have been written on the subject of reducing attrition rates, the majority of these papers offering various solutions to reduce efficacy and safety related attrition. For example, suggested solutions include improving target identification to improve efficacy (Paul *et al.*, 2010), introducing toxicology assays earlier into the discovery phase to identify safety concerns sooner (Hornberg *et al.*, 2014), and reducing the costs of clinical trials by conducting them remotely, such as from the patients home via video conferencing (Holland, 2013).

In silico methods have been developed to fulfill one of the suggestions to lower drug attrition, which was to improve target identification. In silico methods seeking to identify drug targets do so by using different approaches such as drug side-effect similarity (Campillo *et al.*, 2007), correlations between gene and disease expression profiles (Lamb *et al.*, 2006), and the similarity between drugs and genes found within drug-gene associations (Perlman *et al.*, 2011). Databases have also been conceived that contain drug-target and/or drug-chemical information such as the Therapeutic Target DataBase (Zhu *et al.* 2009), STITCH (Kuhn *et al.*, 2012) and DrugBank (Law *et al.*, 2014), which can serve as sources of data for predictive methods.

A second major issue in drug development is drug toxicity, which is responsible for many drug withdrawals as well as adverse events reported in patients. Toxicity is the poisonous effect that a substance can have on the body, having the ability to negatively affect an organism at different levels, such as from the cellular level to the organism level (Bai & Abernethy, 2013).

Of particular concern when describing drug related toxicology is cardiotoxicity. Within this work cardiotoxicity is defined by an abnormality of the heart which that can be identified by abnormal function or anatomy. The importance of identifying drug related cardiotoxicity is that it can lead to possibly life threatening outcomes such as arrhythmias, myocardial infarction and heart failure (Schlitt *et al.*, 2014). Cardiotoxicity is the leading cause of drug recalls due to safety concerns (McNaughton *et al.*, 2014), as well as the leading cause of preclinical failures observed by a leading pharmaceutical company (Cook *et al.*, 2014). During the pre-clinical phase of drug development, pharmaceutical developers are recommended to demonstrate safety across three different cardiac safety assessments prior to entering human trials which include studying the inhibition of the hERG channel and action potential duration in vitro, and studying QT prolongation in in vivo models (Wiśniowska *et al.*, 2014). Methods used to identify drug-related cardiotoxicity include both experimental and in silico methods.

Manual methods to assess cardiotoxicity in drug development include the use of animal models such as the zebrafish (Cheng *et al.*, 2011), testing on human stem cell derived cardiomyocytes (Mandenious *et al.*, 2011), in vitro assays to measure hERG channel blockage, and QT studies in large animals (Heijman *et al.*, 2014).

In silico methods in toxicology seek to identify serious safety concerns for drugs in development or drugs already on the market. Computational methods used to predict toxicity issues can be classified as expert systems or data-driven approaches (Simon-Hettich *et al.*, 2006). Expert systems make use of rules created to evaluate data in the hopes of mimicking decisions which would be expected by a domain expert. In comparison, data-driven approaches make use of calculations such as finding correlations

in order to assign a value to a prediction based on the input data (Simon-Hettich *et al.*, 2006). Data-driven methods usually make use of QSAR (quantitative structure–activity relationship) and try to predict toxicity based on what is known of a similar chemical structure (Muster *et al.*, 2007). In drug discovery, data-driven approaches are usually used in the early stages of drug development whereas expert systems are usually used at the later stages of drug development (Segall *et al.*, 2014).

1.2 Data integration

Data integration has the ability to greatly aid scientific research as it can provide a more complete assessment by using more comprehensive data. Data integration is a step where data collected from different data sources are processed and are integrated together. Integrating data will help solve two major problems, which includes the necessity of accessing data sources separately and retrieving the desired information and allows for computational approaches to uncover new knowledge (Zhang *et al.*, 2011).

The data integration problem has been acknowledged within pharmaceutical research. To tackle the data integration issue with pharmaceutical research, several methods such as WENDI (Zhu *et al.*, 2010), BioGateway (Antezana *et al.*, 2009) and research initiatives such as Open PHACTS (Williams *et al.*, 2012) have been developed. These different initiatives and methods try to approach the data integration problem in different ways. For example the WENDI (Web Engine for Non-obvious Drug Information) approach is a web service which makes calls to different data sources, and returns information which is then pooled together and presented to the user. The input to the method is a chemical, which is used to query the various databases for similar content to the input query, and the output is the search result made available directly on the web

page (Zhu *et al.*, 2010). BioGateway is a semantic tool based on performing analyses on data which is expressed in RDF, which is a model for data within the semantic web. It works by querying and integrating data from different datasets and can create hypotheses related to systems biology (Antezana *et al.*, 2009). Data structured in RDF for the life sciences includes datasets made available by Bio2RDF (Callahan *et al.*, 2013a), or more chemically related datasets made accessible by Chem2BIO2RDF (Chen *et al.*, 2010). These sources aid integration issues by providing data as RDF in a common format that allows for traversing across the datasets. Open PHACTS (Open Pharmacological Concept Triple Store), is a collaboration initiative between pharmaceutical, academic and enterprises which seeks to make drug-related data openly available to a set of tools which they plan to develop in order to find new knowledge to improve drug discovery. The underlying premise of their approach is to offer a platform that utilizes the semantic web to gather datasets and allow for interoperability (Williams *et al.*, 2012). Linking Open Drug Data (LODD) is another initiative based on using the Semantic Web. One of their goals is to demonstrate the usefulness of using Linked Data for the integration of various data. Linked Data is data that contains direct references to other datasets and is represented in a web-friendly format (Erkimbaev *et al.*, 2013). Members within this group have made several datasets related to drug discovery available as Linked Data. Even though making use of Linked Data is promising, this group is still aware of the limitations of using Linked Data such as the heterogeneity between datasets (Samwald *et al.*, 2011). A way to reduce the heterogeneity of datasets is through semantic mappings, which formalize the relationship between terms used in different datasets. The integration of datasets can thus be done by the use of semantic mappings, but this requires that data

be annotated with ontology terms, for example using the SIO ontology (Dumontier *et al.*, 2014) and the translational medicine ontology (Luciano *et al.*, 2011). The use of SIO to integrate biological data was demonstrated when Bio2RDF datasets were mapped to relations in SIO to allow the connection of datasets which were not previously connected, as well as to use queries which also consider the meaning of terms used (Callahan *et al.*, 2013b). The SIO ontology (Semanticscience Integrated Ontology) is an ontology designed to improve the integration of biomedical datasets via ontology terms. SIO increases semantic interoperability by making types and relations available to describe data such as a biological process (Dumontier *et al.*, 2014). The translational medicine ontology allows the integration of life sciences datasets with medical datasets by the use of this single ontology. Bridging such varied data such as electronic health records and genomics data could aid computational approaches such as those seeking to provide a means of personalised medicine for a given patient (Luciano *et al.*, 2011).

1.3 Semantic mappings

A current topic of research is the improvement of interoperability between different ontologies. This is important as doing so would lead to the expansion of current knowledge that can be used to create, answer or study new hypotheses (Hoehndorf *et al.*, 2007). A way to improve this interoperability is by finding equivalent terms between ontologies. Uncovering new knowledge from different data sources is difficult, and as such, research initiatives such as the Monarch Initiative (<http://monarchinitiative.org/>) have been assembled to develop tools which will allow for the exploitation of biomedical data such as model phenotypes and genomics, and apply it to the advancement of

translational research. Increasing interoperability between ontologies would allow for the exploitation of the large amount of data found as Linked Open Data (LOD) in the life sciences (Gross *et al.*, 2011).

Linked Open Data is a web of open data which is linked, and expressed in RDF (Erkimbaev *et al.*, 2013). RDF is a model which allows for data to be expressed as statements to describe a resource. The statements are modeled as triples consisting of a subject, predicate and object. An example triple is given below. This triple states that imatinib (DB00619) has a target ABL1 (target:17).

`<http://bio2rdf.org/drugbank:DB00619> <http://bio2rdf.org/drugbank_vocabulary:target> <http://bio2rdf.org/drugbank.target:17>`

Data expressed in RDF can be queried using a SPARQL (SPARQL Protocol and RDF Query Language) query. The advantage of RDF is that it is machine readable, and allows for datasets to connect to and from other datasets based on a common URI (Bizer *et al.*, 2009). URI's are unified resource identifiers which are unique identifiers used to identify resources (Heath & Bizer, 2011). Use of RDF in the life sciences has the advantage of querying across several datasets such as those containing information on proteins, pathways and chemical structures all at once, and has led to important hosts of biological databases such as EBI (European Bioinformatics Institute) to format the content of the data found in their databases into RDF (Jupp *et al.*, 2014).

In consequence of the lack of interoperability between ontologies, several automatic approaches to mapping have been developed (Shvaiko & Euzenat, 2013). A few examples of automatic approaches include AgreementMaker (Cruz *et al.*, 2009), BLOOMS (Jain *et al.*, 2010), ALIGN (Hayden *et al.*, 2012), and ASMOV (Jean-Mary *et*

al., 2009). AgreementMaker is a matching system capable of matching ontologies by making use of three layers. The first uses TF-IDF (term frequency-inverse document frequency), which is a measure of how many times a term shows up in demonstration of its importance, in order to identify similarities between concepts by the use of concept features (such as labels). The second layer matches based on ontology structure, and finally the third layer creates mappings based on the first two layers. BLOOMS is a tool used to map the schema of two ontologies for LOD that makes use of a bootstrapping technique which validates the alignments made by the method by comparing the frequency of the branches found within various trees. The bootstrapping technique is used as a way to align the schema of ontologies, or the relation between terms, based on the large amount of data currently found in Wikipedia (Jain *et al.*, 2010). BLOOMS first makes a web service call to Wikipedia using the classes from ontologies to return strings to create trees. These trees form a forest, and the forests obtained from each ontology are compared to one another. If trees are similar, they are verified for the confidence of the alignment. Alignments with confidence values of at least 0.95 are kept, and are put through a reasoner to create the mappings. The types of relations which can be found in the mappings include `rdfs:subClassOf` (in both directions) or `owl:equivalentClass`, which describe how one term relates to another in the structure between ontology terms. ALIGN is a hybrid method which matches concepts based on schema and instance similarity. The schema matching step looks for lexical matches between concepts, whereas the instance similarity between classes step is calculated using a Jaccard coefficient. The Jaccard coefficient is a measure of how similar two sets of data are to one another by dividing the intersection of the terms over their union. Finally, ASMOV (automated Semantic

Matching of Ontologies with Verification), matches ontologies by using similarity features between concepts of the input ontologies, and then verifies the mappings using semantics. Similarity features used within this method include lexical matching (ex. strings of terms), similarity between entities (ex. structural), as well as a few other similarity measures.

Several challenges exist when trying to map from one ontology to another such as ontologies not using the same syntax, not using the same terms to distinguish the same entity, differences in coverage and granularity between ontologies, as well as the way that concepts are organised (Euzenat & Shvaiko, 2013). Furthermore, matching ontologies can lead to mapping errors such as redundant mappings, inconsistent mappings, and imprecise mappings which are usually the result of homonymy (Wang & Xu, 2007). Oftentimes, automated mappings need to be reviewed manually by a domain expert in order to identify erroneous mappings (Meilicke *et al.*, 2009).

The alternative to automatic mapping is manual mapping. Manual mappings are disadvantageous as they tend to be impractical as they require a lot of time and effort, and can be prone to human error (Köhler *et al.*, 2011). The advantage is that manual mappings can find mappings which may not be found by automatic methods (Bodenreider *et al.*, 2005). Tools are available to aid in the process of manual mapping such as methods to help identify ontology alignment errors which can be identified using algorithms to alert of possible errors to the expert who is manually verifying mappings (Meilicke *et al.*, 2009).

1.4 Hypothesis testing

Traditionally, hypothesis generation and testing in the life sciences is performed by validation experiments. The generation of hypotheses is often based on background knowledge. As large amounts of scientific data are currently being produced due to technological advances, there is a potential to use this existing data to generate (Biesecker, 2013) and evaluate hypotheses. GenePath is an example of a system able to create new hypotheses in the form of gene-process relationships. This system does so by creating inference patterns based on both experimental data and background knowledge that is used towards creating a genetic network (Zupan *et al.*, 2003). Although this approach is useful for identifying genetic networks, there is no measure of the correctness of the hypotheses. It however does make use of the large amounts of data available to make its inferences. In contrast, HyQue is a semantic web tool created for the evaluation of hypotheses related to galactose metabolism in yeast (Callahan *et al.*, 2011). HyQue conducts its evaluations by first structuring hypotheses as events. Once a hypothesis is inputted to HyQue, the tool is able to query a knowledge base represented in RDF in a

SPARQL endpoint, retrieve the data to serve as evidence, and finally apply a set of domain-specific rules and scoring functions before providing a final overall score based on the evidence gathered (Callahan *et al.*, 2011). HyQue offers several advantages such as making use of Linked Data as evidence, being easily expandable, and easily modified to different use cases such as drug safety profiling.

1.5 Thesis and structure

Given the remarkable growth and diversity in structured biological data, and the outstanding need to effectively integrate these data towards developing safe and effective drugs, I hypothesize that the use of Semantic Web technologies will not overcome limitations in existing approaches and enable new applications in drug discovery and drug safety (see Figure 1-1). In this work, I address two key problems: developing semantic mappings to enable the integration of drug effects with model organism phenotypes to identify new drug targets, and second, developing rules to enable the efficient retrieval and evaluation of biopharmaceutical data pertinent to drug safety. These methods were used within approaches developed to aid drug target identification (PhenomeDrug) and improving drug safety through safety profiling using evidence (HyQue-Cardiotoxicity).

Figure 1-1: Overview of the scope of this thesis

The overall aim of this work is to develop semantic technologies to overcome existing limitations in drug development. This was done by two methods, creating mappings between ontologies, and querying and integrating Linked Open Data. These methods were applied towards drug target identification and drug safety within two approaches. The first approach, PhenomeDrug, was designed to identify human drug targets by associating mouse model phenotypes to drug effects. The second approach, HyQue-Cardiotoxicity, is an approach developed to evaluate the hypothesis that a drug is cardiotoxic based on evidence available as Linked Open Data (LOD).

1.5.1 Chapter 1

In Chapter 1, mappings between drug effects and model phenotype ontologies were developed to identify human drug targets in the context of the PhenomeDrug project. Described is the method by which manual mappings were created, a comparison of manual and automatically-derived mappings, and issues faced when mapping ontologies are discussed.

1.5.1.1 Chapter 1: Hypothesis

I hypothesize that mappings between model phenotypes and drug effects will enable the identification of drug targets.

1.5.1.2 Chapter 1: Aims

1. Provide high quality mappings between drug effects and model phenotypes for terms that are otherwise unmapped after the use of lexical mapping and existing cross-references
2. Evaluate mapping quality by comparing manual mappings to automatically generated mappings

1.5.2 Chapter 2

In Chapter 2, the HyQue-Cardiotoxicity approach will be described. HyQue-Cardiotoxicity is an approach designed to evaluate the hypothesis that a drug is cardiotoxic based on the evaluation of evidence found as Linked Open Data using a semantic web tool known as HyQue (Callahan *et al.*, 2011). The key idea was to design a set of rules to evaluate data that would mimic decisions that approximate those of a domain expert (Segall & Barber, 2014), while leveraging the large amount of relevant data available on the semantic web (Callahan *et al.*, 2013a). In this chapter, the results of the evaluation of HyQue's ability to differentially score cardiotoxic and non-cardiotoxic drugs, as well as a comparison against two toxicological predictors will be presented. It is important to mention that the semantic web tool used within the HyQue-Cardiotoxicity project was developed by a colleague, and was modified for its use within this work.

1.5.2.1 Chapter 2: Hypothesis

I hypothesize that biomedical evidence found as Linked Open Data related to various levels (ex. cellular, organ, and observable phenotype) will enable drug safety profiling of drug-related cardiotoxicity.

1.5.2.2 Chapter 2: Aims

1. Develop HyQue compatible data retrieval and data evaluation rules to gather evidence of drug cardiotoxicity.
2. Demonstrate HyQue-Cardiotoxicity's ability to differentially score cardiotoxic and non-cardiotoxic drugs

3. Evaluate HyQue-Cardiotoxicity's performance in comparison to two freely-available predictive methods of toxicity

1.6 Broader Impact

This work is significant because it not only examines the feasibility of using semantic web technologies to identify novel drug targets and evaluate the safety of marketed drugs, but also addresses key limitations arising from difficulties in integration of large and heterogeneous data. The work on developing semantic mappings in Chapter 1 is significant because it enables the use of animal model data to inform drug development efforts. The observation that the on-target effects of inhibitors are well predicted by knockout animal models (Zambrowicz & Sands, 2003) is converted into a systematic analysis in which the matching of drug effects with knockout phenotypes identifies drug targets. Moreover, the semantic mappings could be used for additional investigations into profiling drugs for their safety and their application, thereby creating new avenues by which the pharmaceutical industry can leverage both public and private data to reduce drug failure in human trials.

The work in Chapter 2 is significant because it presents an approach to profile drug safety using knowledge that is otherwise difficult to harness and use in evaluation. Drug related safety issues are a major issue and are responsible for putting patients at serious risk such as for the drug Vioxx which led to the drugs withdrawal (Hileman, 2005). As toxicity can affect a patient from the cellular level to the organ level, it is necessary to consider all of these potential safety signals when profiling drug safety (Bai & Abernethy, 2013). A method such as HyQue-Cardiotoxicity is able to gather different

types of evidence ranging from cellular assays to predicted drug side-effects, evaluate the evidence against a set of rules, and provide a final score to represent the level of evidence for a drugs cardiotoxicity.

2 Chapter: Mapping drug effects to model phenotypes in the PhenomeDrug approach

2.1 Abstract

PhenomeDrug is a semantic approach that compares model organism phenotypes with known drug effects to identify human drug targets (Hoehndorf *et al.*, 2014). A key aspect of the approach requires that drug effects and model phenotypes be compared to each other. In this chapter, the development and evaluation of the semantic correspondences in the form of mappings that made the PhenomeDrug approach possible is described. The method by which both automatic and manual mappings were created and subsequently evaluated for their quality is also described. Results indicate that the automatic method led to degenerate mappings which included terms that are imprecise or overly broad. As manual mappings are time consuming and still prone to error, it is suggested that a hybrid mapping method would result in quality mappings in less time. This work is significant because it enables drug effects to be incorporated into a larger network of cross-species phenotypes, essential for a wide variety of biomedical and clinical applications.

2.2 Introduction

2.2.1 Motivation

As scientific research advances and new discoveries are made, more data has become available such as phenotype information for model organisms of disease. Making use of research data and integrating it with other datasets would allow for the discovery

of new knowledge (Hoehndorf *et al.*, 2007). An issue with wanting to make use of data from various sources is that they are not easily integrated, for example due to differences in format. One way of integrating datasets is by the use of mappings between ontologies which are currently used within datasets (Gkoutos *et al.*, 2012). Ontologies are a set of predefined terms that are used to describe concepts or annotate data within a domain (Brown *et al.*, 2009). Mapping terms from one ontology to another is critical as it allows for finding equivalent terms which can be used within projects which require datasets to be integrated together, relate datasets to one another in order to find new links, or to analyze data (Gross *et al.*, 2011). In order to map ontologies and try to integrate datasets, researchers from different domains have developed automatic and semi-automatic approaches to map ontologies (Meilicke *et al.*, 2009).

In the context of this work, mouse phenotypes will be associated with drug effects in order to identify new drug targets. The basis for associating animal model phenotypes with drug side-effects is that the effects of an inhibitory drug can approximate that of a knockout or knock-down model (Zambrowicz & Sands, 2003). If the drug effects mimic the phenotype of a knockout model, then the observation of the inhibited gene could be very informative on how a drug works. For example, it could indicate that the drug acts directly on the “knockout gene”. Other possible explanations include the possibility of the drug acting on a different target, or that the drug acts on a protein upstream from the gene, or that the drug acts on the translation of the gene, or even the possibility of the drug acting on the pathway leading to the gene product.

In order to associate the mouse model phenotypes to drug effects, mappings need to be made as there are currently no existing mappings available between model

phenotypes and drug effects. Mapping model phenotypes to drug effects is a challenge as different vocabularies are used and ontologies do not always have the same coverage, as described in the following example. Attempting to map the drug effect for the UMLS term “decreased interest” (UMLS: C0424091) is difficult to map to phenotype ontologies as the phenotype ontologies do not have an existing term for decreased interest as it is a challenging phenotype to observe. Thus, mapping drug effects to model phenotypes is a necessary requirement to allow the PhenomeDrug approach to identify human drug targets as these mappings do not currently exist. Not only are mappings important to the PhenomeDrug approach, but they are important to other computational approaches seeking to uncover new knowledge.

2.2.2 Issues with mapping ontologies

One of the biggest issues of integrating datasets via ontologies is that it is difficult to produce quality mappings. Oftentimes, this is due to ontologies differing from one another in structure and content, for example two distinct ontologies describing abnormal phenotypes and drug effects differing from one another as they are species-specific (Mungall *et al.*, 2010). In order for approaches which make use of various datasets to be successful, it is necessary that mappings produced are quality mappings. Poor quality mappings are often due to errors made during the mapping process. The types of errors found within ontology mapping include incorrect mappings, inexact mappings, inconsistent mappings and redundant mappings (Wang & Xu, 2008). Both manual and computational methods are available to assess the quality of mappings made between ontologies.

Automatic methods for measuring the quality of mappings include comparing the mappings made to those made by a pre-existing alignment and providing a measure of the quality of the mappings (for example, using a precision recall curve). When a reference alignment is not available, a relative reference set of mappings can be created by sampling from the created mappings and having a domain expert identify the correct mappings from this sample. Once the reference set has been created, the mappings can then be evaluated using a relative precision recall curve (Euzenat & Shvaiko, 2013). Alternatively, the evaluation of the mappings can be done manually by using a domain expert to validate mappings made between ontologies. The use of a domain expert to create a gold standard of mappings to which automated mappings can be compared is one of the best ways to evaluate mappings (Cruz *et al.*, 2009).

2.2.3 Mapping strategies

Current computational approaches for mapping ontologies make use of machine learning methods, lexical matching, and links found within the ontologies to map to a common intermediate ontology (Noy, 2009). Examples of computational methods for creating mappings between ontologies include AgreementMaker (Cruz *et al.*, 2009) and BLOOMS (Jain *et al.*, 2010). Although computational approaches are often employed, they are not perfect, and their competency often depends on the ontologies to which they are applied (Ghazvinian *et al.*, 2009).

The alternative to computational mapping between ontologies is manual mapping. Manual mappings have the advantage of finding mappings which may not be obvious to computational methods (ex. lexical matching), and also make use of the domain

knowledge of the expert creating the manual mappings. Inversely, manual mappings are impractical to map large ontologies, are prone to human errors, and are costly when compared to computational methods (Bodenreider *et al.*, 2005).

2.2.4 Inter-annotator agreement

When two approaches are used to map terms, they can also be viewed as annotators. The mappings made between both approaches can then be compared for their level of agreement. Kappa's statistic is a popular measure used to determine the agreement between two annotators based on whether the two separate annotators will assign the same value for a variable when two annotators are being compared (McHugh, 2012). The Kappa statistic has been used in different contexts, for example in medical research to compare patient diagnoses made by medical professionals (Chmura Kraemer *et al.*, 2002), to measure co-occurrence in gene annotation studies (Glass *et al.*, 2012), and to measure inter-annotator agreement between domain experts (Boeker *et al.*, 2011). The final Kappa statistic is a value between 0 and 1, 0 indicating no agreement and a value of 1 being a perfect agreement between the two annotators (McHugh, 2012). Kappa's statistic was used within this work in order to compare the agreement between the manual and automated mappings which were generated.

2.2.5 PhenomeDrug

PhenomeDrug is an approach that could lend itself to drug repurposing, and was developed to identify human drug targets by associating knockout mouse phenotypes and drug effects. The rationale for wanting to associate knockout model phenotypes with drug effects is based on the observation that phenotypes seen in a gene knockout model would

closely approximate drug side-effects (Zambrowicz & Sands, 2003), which could be indicative of a common target. For example, the similarity between drug and knockout model profiles could indicate that the drug acts directly on the gene in question, that the drug acts on a protein upstream from the gene, or even that the drug is acting on the pathway leading to the gene product (see Figure 2-1). The similarity between phenotypes and drug effects was calculated using semantic similarity. Semantic similarity is a measure which is used to compare two groups of terms and calculate the similarity between the two (Pesquita *et al.*, 2009). Semantic similarity is based on the Jaccard metric, and is calculated by dividing the intersection of terms over the union. To make use of mouse phenotypes to find human drug targets, PhenomeNet (Hoehndorf *et al.*, 2011) was used to integrate mouse and human phenotypes. The PhenomeNet system is a cross-species ontology which combines phenotypes from different animal model ontologies using semantic similarity. PhenomeNet was altered for the scope of this project to integrate mouse model phenotypes and human drug effects, and made use of a modified SimGIC (Pesquita *et al.*, 2008) measure of semantic similarity. SimGIC is a measure based on the Jaccard index weighted by the information content of a term. This measure was altered to take into account the fact that a drug may act on several different targets (Hoehndorf *et al.*, 2014), also known as drug promiscuity (Hopkins, 2009). Prior to using PhenomeNet to integrate phenotypes and calculate the semantic similarity, mappings had to be made between drug effects and model phenotypes. This was done both automatically and manually. It is the manual mappings which are covered in this chapter.

The performance of the PhenomeDrug approach was tested by comparing the generated drug target associations with those described in the DrugBank (Law *et al.*, 2014) and STITCH human and mouse databases (Kuhn *et al.*, 2012). PhenomeDrug was realized with collaborators, and had a demonstrated ability to identify human drug targets. The details related to the work done by collaborators can be found within the publication for the approach which was recently published in the Journal of Bioinformatics (Hoehndorf *et al.*, 2014).

2.2.6 Manual mappings of drug effects to model phenotypes

The focus of this chapter is on the manual mappings of model phenotypes to human drug effects. This is done within the PhenomeDrug approach (Hoehndorf *et al.*, 2014) in order to identify human drug targets (see Figure 2-2). The hypothesis of this work is that creating mappings between model phenotypes and human drug effects can allow for the identification of human drug targets. The hypothesis will be evaluated by reviewing and comparing the mappings done with those done automatically. Furthermore, high quality mappings should lead to good performance by the PhenomeDrug approach.

Figure 2-1: Role of mappings within the PhenomeDrug approach

In order to allow for the similarity between mouse phenotypes and drug effects to be calculated to identify drug targets, drug effects needed to be mapped to phenotype terms found in ontologies such as the HP, MP, MPATH, NBO and DO ontologies. The mapping step was first performed automatically, followed by manual mapping for unmapped terms.

Figure 2-2 Overview of the rationale for the PhenomeDrug approach

Associating knockout mouse model phenotypes with effects of an inhibitory drug and calculating the semantic similarity between them allows for the identification of drug targets in humans.

2.3 Materials and methods

Within the PhenomeDrug project, manual mappings were made between SIDER UMLS drug effect terms and HP and MP phenotype terms for 2132 unmapped terms. Prior to the manual mapping, 4717 UMLS SIDER terms had already been mapped automatically (Hoehndorf *et al.*, 2014). SIDER is a resource created by text mining drug package inserts from which drug indications and side-effects were identified, along with their frequencies (Kuhn *et al.*, 2010). Side-effects in SIDER are annotated using UMLS (United Medical Language System Thesaurus) which is a thesaurus that integrates vocabularies across the biomedical domain (Bodenreider, 2004). The goal of the

mappings was to find mouse phenotype (MP) and human phenotype (HP) terms from ontologies which would be equivalent to the UMLS drug effect terms which were unmapped. The importance of this step is that the method relies on the associations made between mouse model phenotypes and drug effects, thus it is important that the mappings be made and that they be accurate.

Furthermore, an evaluation of mappings made between UMLS and HP was performed to assess the manual mappings, and compare the quality of the manual mappings to those made automatically. The 344 manual mappings made from UMLS to HP were compared with the HP to UMLS mappings generated by a computational method for the same 344 UMLS terms.

2.3.1 Storing terms to map and mappings to ontologies

A list of 2132 unmapped UMLS SIDER drug effect terms were stored in a spreadsheet. Within this spreadsheet, there were columns designated for storing found ontology terms and their respective identifier, as well as a description of the type of mapping which was made. Three different types of mappings were possible; an exact match, an inexact match and no match, which are discussed further in this chapter (see section 2.3.3).

2.3.2 Searching ontologies for terms

BioPortal (Noy *et al.*, 2009) was used to search ontologies for matching terms to the unmapped UMLS drug effect terms. BioPortal is a repository of over 300 biomedical ontologies. BioPortal enables users to search across all ontologies for terms of interest.

The advanced search, which was the search used within the scope of this project, can be used to restrict the search to ontologies of interest. Unmapped UMLS terms were searched against the following ontologies; mouse phenotype ontology (Smith *et al.*, 2004), human phenotype ontology (Robinson *et al.*, 2008), mouse pathology (Schofield *et al.*, 2010), disease ontology (Schriml *et al.*, 2011), and neurobehavioral ontology (Gkoutos *et al.*, 2012) as they are ontologies which contain terms that best approximate the UMLS terms. The returned result of an advanced search made in BioPortal is a list of matching terms found within the specified ontologies. Matching ontology terms returned from a search can be matched by preferred term, but also by synonym terms which are associated with the preferred term.

2.3.3 Selection of ontology terms.

After performing the search query for matching terms within specified ontologies using BioPortal, the best fit terms were selected based on the similarity of meaning between terms and kept in a spreadsheet along with their identifier and the type of match (see Table 2-1). Precedence was given to mouse phenotype ontology (MP) and human phenotype ontology (HPO) terms as they are the phenotypes most likely to correspond to human drug effects based on species similarity. The type of possible matches for each term included an exact match, an inexact match or no match. In the event of an exact match, an ontology term was an exact match to the query term, for example the UMLS SIDER term “rhinitis” matching the MP ontology term “rhinitis” (MP:0001867). The event of an inexact match occurred when an ontology term was found, but is not quite the same as the query term. For example, the drug effect term “hypercholesterolemia” (UMLS:C0020443) was most closely matched to “increased circulating cholesterol level”

(MP:0005178). When no match was possible, it meant that the search did not find a matching ontology term. In this case, the ontologies were reviewed for similar terms, and new terms were created using the existing similar terms. These newly suggested terms were stored in the spreadsheet and presented to a project collaborator (Georgios V. Gkoutos) so that the new terms could be created in existing ontologies such as HP and included in the final mappings.

2.3.4 Retrieving UMLS drug effect to model phenotype ontologies mappings for the PhenomeDrug Approach

3858 pre-existing cross references between SIDER UMLS terms and HPO were obtained from the HPO website. The file was released on May 16th 2012. 859 terms were mapped by lexical matching from UMLS to either HPO or MP. 2132 drug effect terms remained unmapped.

2.3.5 Automatic HPO-UMLS mappings obtained for the mapping comparison

The HPO to UMLS mappings file was downloaded from the HPO website (<http://compbio.charite.de/svn/hpo/trunk/src/mappings/term2umls.out>) which contained the computer generated mappings between UMLS SIDER terms and the HP ontology. The file was retrieved on June 17th 2014. This file contained 33670 rows of mappings from HPO terms to UMLS terms.

2.3.6 Comparison of manually curated and computer generated UMLS to HP mappings

A spreadsheet was created with the 344 manual mappings made to HP for the unmapped UMLS terms. Several columns were created to categorize the manual mapping in comparison with the computer generated mapping. To compare the mappings, a common UMLS identifier was used in order to ensure that a common UMLS term was being used to compare the manual and computational mappings. The computational mappings used for the comparison are those described in section 2.3.5. The spreadsheet was used to tally the amount of mappings in common, and if the automatically generated mappings were correct or incorrect. An online calculator, ReCal2 (Reliability for 2 Coders, available: <http://dfreelon.org/utis/recalfront/recal2/>), was used in order to assess the inter-rater agreement between manual and automatic mappings.

2.4 Results

2.4.1 Mapping Drug Effects to Organismal Phenotypes

Manual mappings were made between drug effect terms and terms found within phenotype ontologies. Prior to the manual mapping, automated mapping was performed to map the most drug effect terms as possible to phenotype ontologies. After automated mapping, 2132 drug effect terms remained unmapped. Manual mapping was used to map 950 unmapped drug effect terms to ontologies which included HP (human phenotype), MP (mouse phenotype), NBO (neurobehavioral), MPATH (mouse pathology) and DO (disease ontology), with preference given to MP and HP terms (see Appendix A). 1182 drug effect terms remained unmapped even after a second pass of manual mappings. Terms which were left unmapped were mostly terms which involved medical events or

medical equipment, for example IUD expulsion (C0021899), which are not described in phenotype ontologies. Of the 950 manually mapped terms, 293 mappings were made to MP, 344 mappings were made to HP, 232 mappings were made to DO, 55 mappings to MPATH and finally 26 mappings were made to NBO (see Table 2-2).

Table 2-1: Examples of mappings made between SIDER UMLS terms and ontologies

SIDER UMLS term	UMLS cui	Frequency	Primary mammalian phenotype term ID	Primary mammalian phenotype term label	Inexact match	No match
infection	C0021311	5122	MP:0002406	increased sensitivity to infection	x	
pain	C0030193	2241	MP:0005407	hyperalgesia	x	
diabetes mellitus	C0011849	1954	HP:0000819	diabetes mellitus		
cancer	C0006826	1557	MP:0002018	increased malignant tumor incidence	x	
ulcers	C0041582	1417				x

Table 2-2: Number of mappings made between ontologies to SIDER UMLS terms

Ontology	Number of terms mapped to SIDER (#)
Human Phenotype (HP)	344
Mouse Phenotype (MP)	293
Disease Ontology (DO)	232
Mouse Pathology (MPATH)	55
Neurobehavioral (NBO)	26
TOTAL	950

Table 2-3: Result of the comparison made between the manual and automatic mappings between HP and UMLS SIDER terms

Comparison classification	Number of cases (#)
Number of terms for which the automated method had mappings	300/344
Mappings in common	93/505*
Correct automatic mappings	94/505*
Incorrect automatic mappings	411/505*

* 505 mappings were made for 300 of the 344 terms mapped by the automatic mappings

2.4.2 Comparing manual with automatic mappings

344 UMLS terms were manually mapped to the HPO. In comparison, only 300 of the 344 terms were mapped by the automated method, some terms mapped multiple times leading to 505 total mappings. Of the 505 mappings, only 103 mappings were common to both the manual and the automatic method. Importantly, the automatic method mapped the same side effect term to several different phenotype terms (See Appendix B). Table 2-3 shows the different kinds of mappings and their frequency. Using the Recal2 calculator, it was determined that only 19% of the mappings were actually in agreement. Kappa's statistic was used to establish the degree of inter-annotator agreement between the manual and automated approach. The Kappa statistic had a value of 0.05, indicating that the agreement was negligible between the two approaches. On closer examination, I found that 200 of the 505 mappings made by the automated method were incorrect based on manual review.

2.5 Discussion

2.5.1 Unmapped concepts within manual mapping

When manually mapping UMLS drug effect terms to HP phenotypes for the PhenomeDrug approach, 1182 drug effect terms could not be mapped between UMLS and the selected biomedical ontologies (HP, MP, MPATH, NBO and DO). One reason that drug effect terms could not be mapped manually to ontologies is that some drug effect terms are used to describe medical procedures or equipment that are not expressed in the phenotype ontologies considered. For example the drug effect term “TUD expulsion” (C0021899) describes both a medical device and an event, which are not described in phenotype ontologies which describe phenotypic abnormalities (Smith *et al.*, 2005). At times, no matches could be made due to no equivalent term being found in the biomedical ontologies. For example, the term “ulcers” is not found in any of the selected biomedical ontologies. This example demonstrates a case where the term is too broad and could be found anywhere on an organism, which is not possible to describe in ontologies such as HP and MP where phenotypes are organized by anatomy. The difference in terminology is likely due to differences in the domain of the ontologies, the phenotype ontologies being specific to a given species (Mungall *et al.*, 2010).

2.5.2 Inexact matches within manual mapping

Along with no matches there existed the case of inexact matches. Inexact matches are matches in which the terms approach a similar meaning, but are not completely equivalent. Within the manual mappings, a high volume of mappings are inexact matches. This is due to ontologies such as the Mouse Phenotype Ontology, expressing their terms differently. For example, the UMLS drug effect “Infection” (C0021311) is an

inexact match to the Mouse Phenotype Ontology's "Increased sensitivity to infection" (MP:0002406) term. The terms found in the Mouse Phenotype ontology were created to describe phenotypic abnormalities (Smith *et al.*, 2005) hence the mention of "increased sensitivity" in the term, whereas UMLS terms group biomedical terms (Bodenreider, 2004). Inexact matches are also due to the modifiers used in MPO such as "abnormal". Within manual mappings, these modifiers led to mappings being classified as inexact as the modifier changes the meaning in the term.

2.5.3 Comparison of manual and automatic mappings between UMLS concepts and HPO terms

Mappings made both manually and computationally for 344 terms between UMLS and HPO were compared. The automated method made several mappings from HPO to UMLS terms. When the automated mappings were reviewed, it was shown that two thirds of them were incorrect upon review. This was due to the terms being either too general, too specific or in some cases even unrelated (see Table 2-4). The high level of disagreement with automatically-generated mappings indicates the need for expert-driven review (Meilicke *et al.*, 2009).

A recent second automatic mapping study sought to create mappings between UMLS concepts and HPO (Winnenburg & Bodenreider, 2014). Several findings from this study's analysis are in common with the small study of mappings between UMLS and HPO done within this thesis. The manual mappings within this thesis led to 45% coverage, whereas this second automatic mapping resulted in coverage of 54% with lexical matching alone. The authors of the second automatic mapping approach state a few reasons which could have contributed to the reduced coverage. First, they state that

the difference in representation of phenotypes in different ontologies such as the difference in the vocabulary used between ontologies to describe the same term, would have contributed to the inability to map HPO to UMLS. Secondly, the lexical matching was not very successful due to variations in the ontology terms between equivalent concepts. The results of this study which sought to map a much larger number of terms could thus help provide better insight into possible explanations as to why automatic mappings between HPO to UMLS drug effect terms was not as successful as the manual ones.

Within the manual mappings made within the scope of this thesis, it was found that a lot of the inexact terms were due to modifiers within the terms found in phenotype ontologies, such as the modifier “abnormal” for terms found in MPO, such as “abnormal clotting”. When mapping automatically, such modifiers can lead to terms being unmapped through lexical matching, and require that the terms undergo normalization to remove such modifiers to increase the odds of mappings (Burgun *et al.*, 2009). Thus, not treating the modifiers in lexical mapping could lead to an increase in unmapped terms in comparison to manual terms. This is supported by the observation that manual mappings can often find mappings which may be missed using automatic methods based on lexical matching (Bodenreider *et al.*, 2005). The Burgun *et al* study demonstrated that 20% of unmapped terms could be mapped once such modifiers were removed from MPO terms.

As the sample size of terms used in this thesis used to compare the manual mappings and the computational mappings is small in comparison to the size of whole ontologies, it is not possible to make any general statements between the merits of the two mapping methods on this study alone. However it would seem that through this

analysis that the manual and automatic mappings were in agreement most of the time if the automatic methods multiple mappings are not considered as the automatic method produced both multiple and inaccurate mappings. When the multiple mappings are included, as made obvious by the low Kappa coefficient, the manually and automatically derived mappings are not in agreement. These conclusions as well as other conclusions are also described in an anterior study which sought to map 2500 terms between mouse and human anatomical terms using both manual and automatic means (Bodenreider *et al.*, 2005). Within this study, the manual and automated methods generated 715 and 781 mappings respectively. The mappings made by the two methods were evaluated by a domain expert. The conclusions from the comparison of the two mapping methods highlighted that although manual methods may uncover mappings not found by automated methods, they have the disadvantage of being costly and time consuming, and could still be prone to error (Bodenreider *et al.*, 2005).

A caveat of the manual and automatic mapping evaluation described in this thesis is that the manual mappings were made by the same “expert” who compared the manual and automated mappings. This could have contributed to a bias when examining automatic mappings based on the evaluation of the mappings made is dependent on the knowledge of the “expert”. Nonetheless, the analysis allowed for a review of the manual mappings to verify their completeness, and supports the findings of larger studies which mapped UMLS to phenotype ontologies.

2.6 Conclusion

In conclusion, manual mappings were made between drug effects and model phenotypes were used within the PhenomeDrug approach to identify drug targets. A

review of the mappings demonstrated that the manual HP mappings subset provided mappings which were more accurate than those done automatically. As manual mappings between large ontologies would be unrealistic, automated methods followed by manual curation could provide quality mappings without the disadvantage of the cost of manual mappings. The review of alternate studies demonstrated that common issues such as mismatches between ontologies caused by differences in vocabulary and domain specific terms are still problematic during the mapping process, regardless of the mapping technique used. Consequently, it is necessary to continue research efforts into mapping methods and improving ontologies to increase the interoperability between ontologies in order to increase the ability to gain new knowledge by the integration of various datasets.

3 Chapter: HyQue-Cardiotoxicity

3.1 Abstract

Cardiotoxicity is the leading cause of safety related recalls of pharmaceuticals. Recently, tyrosine kinase inhibitors (TKIs) have been unexpectedly associated with cardiotoxicity. Cases such as TKIs demonstrate that the current methods of identifying drug-related cardiotoxicity are not adequate. Presented in this work is HyQue-Cardiotoxicity, a semantic web tool which makes use of recent data in form of Linked Open Data as evidence and scores it based on rules related to cardiotoxicity. Upon evaluation, this approach performs better than other approaches. The advantages of HyQue-Cardiotoxicity include its ability to reason over existing data found within different datasets, its ease of expandability, and the ability to retrace the specific evidence used within the evaluation to understand the final overall score.

3.2 Introduction

3.2.1 Motivation

Drug safety is of utmost importance to patients, physicians, public health agencies and the pharmaceutical companies that develop, manufacture, and distribute drugs. A major issue in drug development is drug toxicity, which is the cause of many reported adverse events in patients, and is responsible for many drug withdrawals. Toxicity can be defined as the poisonous effect that a substance can have on the body, which can affect different levels of an organism from the cellular level to the organs (Bai & Abernethy, 2013). An example of drug induced-toxicity is cardiotoxicity, which whose toxic effects manifest themselves in ways such as lowered left ventricular ejection fraction, a

manifestation associated with serious cardiac events. Cardiotoxicity events may also only appear at higher doses of a drug, for example the cardiotoxicity seen after starting treatment with doxorubicin (Shi *et al.*, 2011). Although drug toxicity is usually observed during preclinical testing (Segall & Barber, 2014), additional adverse effects may be uncovered during clinical trials or during the post-marketing surveillance period (Dearden, 2003). Once a drug has reached the market, the evaluation of a drug's safety profile continues through epidemiology studies, electronic medical records, post-marketing trials, and clinical registries (Platt *et al.*, 2008). The combination of preclinical testing and post-marketing surveillance can lead to the attrition of a drug when toxicities are identified (Guengerich, 2011). Also, approximately one third of all drugs on the market are discontinued due to observations that a drug may be unsafe or toxic (Ferri *et al.*, 2013). As such, it would be of interest to detect these drug toxicities sooner, and ideally during development of a drug to prevent adverse events caused by unforeseen toxicities.

3.2.2 Cardiotoxicity and tyrosine kinase inhibitors

Cardiotoxicity is the leading type of drug toxicity responsible for drug recalls over safety concerns (McNaughton *et al.*, 2014). Not surprisingly, cardiotoxicity is currently a limitation in the treatment of cancers as it can negatively impact the overall health and survival of the patient by introducing life threatening complications such as congestive heart failure (Berardi *et al.*, 2013). Recently, the issue of cardiotoxicity has become a concern with the use of tyrosine kinase inhibitors (Raschi *et al.*, 2010). Tyrosine kinase inhibitors (TKIs) are monoclonal antibodies or small molecules which are designed to inhibit kinases which may be overexpressed, and which may play a role in cancer (Force

& Kolaja, 2011). Tyrosine kinases have been linked to several serious adverse cardiac events such as heart failure, left ventricular dysfunction and QT prolongation (Force *et al.*, 2007). As such, there is a need to identify cardiotoxic events sooner for drugs such as TKIs to improve the safety of patients as well as to protect pharmaceutical developers.

3.2.3 Current methods for the evaluation of cardiotoxicity

In an effort to predict or study drug-induced cardiotoxicity to improve drug safety, various types of methods and approaches have been developed ranging from the use of animal models such as zebrafish (Cheng *et al.*, 2011), human cardiomyocytes and bioanalytics (Mandenious *et al.*, 2011), and QSAR (Frid & Matthews, 2010).

One strategy to detect safety concerns related to drug toxicity earlier is the use of *in silico* methods which are developed to predict or identify toxicity related events. The advantage of using computational methods is that they allow the use of evidence to support drug safety decisions based on pre-existing data (Valerio, 2011). The majority of methods seeking to predict cardiotoxicity rely mostly on the identification of the hERG channel. When blocked, this channel has been linked to the development of Torsades de pointes and long QT syndrome (Rampe & Brown, 2013). The method most used for predicting the inhibition of hERG channels is QSAR (qualitative structure–activity relationships) methods such as the use of pharmacophores, support vector machines and genetic programming (Ekins, 2014). hERG is an important measure when evaluating the potential for cardiotoxicity, as supported by the fact that the FDA now requires that hERG assays be done prior to testing in human subjects (Rampe & Brown, 2013). The hERG channel is a potassium channel implicated in the electrical functioning of the heart.

When this channel is blocked it can cause QT prolongation, which can increase the risk of arrhythmia (Sanguinetti & Tristani-Firouzi, 2006). Unfortunately the results of hERG assays are not always reliable as they are associated with a 30% chance of being either a false positive or false negative error (Rampe & Brown, 2013). As such, a method which could include other sources of evidence beyond the hERG channel would be of interest.

Other types of methods have been conceived to predict drug toxicity or cardiotoxicity based on a chemical-structure and chemical-property relationships. For example, Cramers rules, which can be found in the OpenTox framework, makes use of a set of rules used to evaluate the structure of a compound through a decision tree. The rules are based on identifying chemical structure elements which are associated with toxicity, such as a structure containing a functional group that has been found to be associated with toxicity (Lapenna & Worth, 2011). The result of the rules within Cramers rules are used to classify compounds as being either class I, class II or class III. Class I compounds have known information which indicates a low risk for toxicity, whereas class III compounds have no previous known information related to their substructures and as such, could potentially be highly toxic. Class II compounds are considered to be of intermediate risk of being toxic (Patlewicz, 2008).

Commercial predictors can also be used to evaluate the potential of health effects on an organ such as the heart. For example the Percepta platform created by ACD Labs can provide toxicity predictions for target organs. This platform takes in a chemical structure as the input and compares the input structure with structures found in databases. Within the databases, compounds are associated to adverse events which are then

analysed by various in-house algorithms to predict the probability of a compound having toxic health effects.

Admittedly, there are also some challenges when using *in silico* methods, such as determining which factors can contribute (or not) to toxicity such as dosing, ADME, drug-drug interactions, as well as having access to relevant data (Egan *et al.*, 2004).

3.2.4 Linked Open Data

In recent years, some important biological and chemical databases such as CTD, PharmGKB, SIDER and DrugBank have made their data publicly available (Bai & Abernethy, 2013). These databases contain pertinent information about toxicity-related endpoints of interest and biological knowledge which can be used to improve drug safety.

The semantic web is an extension of the web that is machine readable and allows for the use and exchange of data from different sources through Linked Data. Linked Data is a means of connecting data in order to make use of as much information which is available on a given subject by the use of common unique resource identifiers (URIs) and RDF. URIs are identifiers used to identify things. RDF, or the Resource Description Framework, is a data model which allows the structuring of data to describe resources (Heath & Bizer, 2011). RDF is structured in triples composed of a subject, predicate and an object. For example, in Bio2RDF's DrugBank endpoint, the triple:

`<http://bio2rdf.org/drugbank:DB00619> <http://bio2rdf.org/drugbank_vocabulary:target> <http://bio2rdf.org/drugbank.target:17>`

describes that imatinib (DB00619) has a target ABL1 (target:17). An object can be a literal (ex. "imatinib"), a boolean (ex. "true") or a URI. RDF can provide links between datasets, known as Linked Data. Linked Data allows for the integration of information from various sources to provide a more complete picture. RDF can be stored in SPARQL

(SPARQL Protocol and RDF Query Language) endpoints, and the RDF data of interest can be retrieved using the SPARQL querying language in a SPARQL query. Bio2RDF hosts many endpoints which contain biological data found in databases such as DrugBank and SIDER. The Linked Data found in Bio2RDF includes genes, proteins, drug targets, drug effects, pathways and database identifiers (Callahan *et al.*, 2013a). SPIN, which is the SPARQL inferencing notation, is a RDF vocabulary which allows for the creation of rules and constraints using SPARQL that can be applied onto data. The use of constraints and rules is useful when working with such large amounts of data. For example, when retrieving side-effects for a drug, it is necessary to verify for the presence of side-effects which are cardiotoxic within the list of side-effects which were retrieved. This is done by the use of constraints and rules. Use of the semantic web is advantageous in scientific research as it allows for the combination of data from different research branches in order to make new predictions and assumptions which may not have been discovered if only one source of data would have been used (Pasquier, 2008).

3.2.5 HyQue-Cardiotoxicity

In this work, the HyQue-Cardiotoxicity approach was designed to assess a drug's potential for being cardiotoxic using Linked Open Data (LOD). The success of this approach will be measured based on the ability of the method to differentially score cardiotoxic and non-cardiotoxic drugs, as well as through a comparison with two predictors of toxicity. The method makes use of publicly available data accessed through various SPARQL endpoints. Unlike other methods which were developed to predict cardiotoxicity using specially developed metrics from only a single source of data, this approach gathers different types of evidence and evaluates the evidence against the

hypothesis that a drug is cardiotoxic. Furthermore, unlike other methods which do not allow a measure of how correct their predictions are or which part of the data contributed to the final score, this method is evidence-based, and produces a final score representative of the amount of evidence to support the hypothesis, and allows the score to be traced back to the data which was used for the evaluation. This is done by expanding HyQue, a semantic web tool for evaluating evidence, to the use of evaluating hypotheses related to drug related cardiotoxicity.

3.3 Materials and methods

3.3.1 Datasets

The data used is made available by Bio2RDF (Callahan *et al.*, 2013a), ChEMBL (Gaulton *et al.* 2011), and data created by manual curation of scientific literature. Bio2RDF includes an RDF representation and SPARQL endpoints for several databases (<http://download.bio2rdf.org/release/3/release.html>). The Bio2RDF endpoints used for this project include endpoints which store data from CTD (Davies *et al.*, 2013), DrugBank (Law *et al.*, 2014), PharmGKB (Whirl-Carrillo *et al.*, 2012), PubMed (Acland *et al.*, 2013), SIDER (Kuhn *et al.*, 2010), OFFSIDES & TWOSIDES (Tatonetti *et al.*, 2012), MGI (<http://www.informatics.jax.org>), and ClinicalTrials.gov (www.clinicaltrials.gov).

CTD was used as a source for papers linked to a disease, pathways predicted to be altered by a drug, genes which are targeted by a drug, the interaction type between a drug and a given gene, and the identification of diseases which are related to genes targeted by a drug. DrugBank was used as a source for the identification of known drug targets, drug-

drug interactions, and the interaction type between a drug and a target. PharmGKB includes drug-drug interactions (TWO SIDES), drug pathways, and predicted side-effects (OFF SIDES). TWO SIDES provides predicted drug-drug interactions and OFF SIDES provides predicted side-effects for drugs. PubMed was used to find articles and abstracts about drug-drug interactions. The SIDER endpoint was used to identify drug effects obtained from drug package inserts. The MGI endpoint was used to identify the phenotypes associated with mouse models based on an altered allele matching a drug target. ChEMBL hosts its own SPARQL endpoint (<http://www.ebi.ac.uk/rdf/services/chembl/sparql>). The ChEMBL endpoint was used to query for IC₅₀ assays (assays used to measure the concentration at which a drug acts as an inhibitor (Sebaugh, 2011)), assays related to cardiotoxicity, TUNEL assays, and QTc drug assays, which measure interval when there is repolarization and depolarization of the heart ventricles (Pollard *et al.*, 2010). Finally, local endpoints for clinicaltrials.gov as well as a manually curated database were hosted locally. To find clinical trials related to a drug, the clinicaltrials.gov endpoint was used. The manually curated database contained extracted facts from reading papers related to drug off targets, drug pathways, and clinical observations where positive evidence of a drug being cardiotoxic were described. The source of the datasets used can be found in Table 3-1.

3.3.2 HyQue

HyQue is a semantic web tool which allows for the evaluation of a hypothesis based on evidence which was originally developed for hypotheses in yeast (Callahan *et al.*, 2011). HyQue makes use of its own ontology in order to define the input hypothesis, describing the proposition, event type, and input (ex. DrugBank drug URI). The ontology

is used to link the SPIN functions to the event. For this approach, HyQue was extended for the domain specific case of drug related cardiotoxicity. To make use of HyQue, the HyQue ontology (<http://semanticscience.org/ontology/hyque.owl>) and HyQue Spin rules (<https://github.com/alisoncallahan/hyque-spin-rules/blob/master/hyque-system-rules.spin.rdf>) were imported into TopBraid Composer Free Edition, v.4.2.0 (<http://www.topquadrant.com/>), which is an integrated development environment for building semantic web applications. Once imported into TBC, HyQue is ready to use as per the documentation.

3.3.3 Rule creation

Within a rule, there are 3 types of functions. The first function type is a data retrieval function that can specify a SPARQL query to obtain results from one or multiple SPARQL endpoints. For example, a data retrieval function can be used in order to retrieve side-effects for a drug such as sunitinib from the SIDER endpoint. The second type of function is an evaluation function which assesses the extent to which the data retrieved meets a given set of conditions. For example, an ASK statement can be used to ask if one of the side-effects for sunitinib in the SIDER endpoint is found within the list of manually curated side-effects linked to cardiotoxicity. The result of this function will be a boolean. If a retrieved side-effect from this endpoint is present, “true” will be returned. If none of the retrieved side-effects match those on the list, then “false” will be returned. The third type of function is a scoring function. The scoring function will award a score based on the result of the evaluation function, awarding a positive score when the evidence supports the hypothesis (thus when the data evaluation function returns “true”) and a negative score when the evidence negates a hypothesis (when the data evaluation

function returns “false”). The scoring function can be grouped with other scoring functions in a final aggregation function for the event which combines the respective scores into one final overall score. The overall maximum score for a given hypothesis is a score of 1, which would represent positive evidence of drugs toxicity for each of the 20 functions. Thus, each function is worth 0.05 of the final score. Of these functions, 3 relied on predictions made by other methods such as for predicted side-effects (OFFSIDES), predicted drug-cardiotoxic disease pathways (CTD) and predicted DDIs (TWOSIDES). For these functions, p-values were used to restrict associations to those with a p-value less than 0.0001 or smaller (a p-value of less than 1E-11 for drug-disease pathway). See Figure 3-1 for an example of a HyQue-Cardiotoxigenicity rule. This method is composed of 20 rules (see Appendix C) which are used to evaluate evidence of drug related cardiotoxicity.

Evaluating side effect related evidence for cardiotoxicity in TKI's

Figure 3-1: Example of the structure of the cardiotoxic side-effect rule

A) Data retrieval. Initially, a drug is specified in the form of a Bio2RDF DrugBank URI. The DrugBank endpoint is queried with the input DrugBank drug URI, for example the URI for sunitinib, and a cross-reference to the SIDER drug URI for sunitinib is returned. This SIDER drug URI is inputted in the SIDER endpoint, and a query to retrieve side-effects for sunitinib retrieves the associated side-effects. B) Data evaluation. The retrieved side-effects for sunitinib are then compared to those found within the list of manually curated side-effects linked to cardiotoxicity. C) Scoring. If the data evaluation resulted in a match between the retrieved effects and the curated cardiotoxic effects list (like is shown in this figure), a score of 1 is assigned due to the presence of supporting evidence. A lack of evidence results in a score of 0 for this rule.

Figure 3-2: Overview of the HyQue-Cardiotoxicity approach

HyQue takes in a hypothesis, evaluates it, and then provides a final score based on the evaluation of evidence. HyQue does the evaluation by making use of rules (see Figure 3-1) specified using SPIN. These rules are used to evaluate the data retrieved from several SPARQL endpoints. SPARQL endpoints used in this project include those hosted by BIO2RDF (SIDER, PharmGKB, CTD, PubMed, ClinicalTrials.gov and MGI) as well as an endpoint hosted by a third party (ChEMBL), and a local endpoint (manually curated database).

3.3.4 List of terms used to evaluate evidence related to cardiotoxicity

Manually curated lists of cardiotoxic terms related to mouse phenotypes, drug effects, MeSH disease and phenotypes, genes, and pathways were created. The lists are based on the examination of all terms related to the heart in the mammalian phenotype ontology (MP) for mouse phenotypes, SIDER UMLS terms used to describe drug-effects, and in CTD-associated MeSH terms. Overall in MP, there are 51 MP terms that were considered phenotypes that were linked to cardiotoxicity. In SIDER, 56 terms were considered to be cardiotoxic. In CTD, 23 MeSH terms were identified which are related

to cardiotoxicity. Gene cardiotoxicity lists were created based on literature, where a human gene was identified as playing a role in cardiotoxicity. The gene list consisted of 12 genes. Lists of pathways linked to cardiotoxicity were created based on pathways mentioned in literature as possibly or known to be involved in cardiotoxicity. Overall, 6 KEGG pathways, 4 Reactome pathways and 6 PharmGKB pathways were considered to be cardiotoxic pathways. Drug targets involved in cardiotoxicity were also obtained through literature, which resulted in a list of 9 drug targets. All lists used within this project can be found in the supplementary data (Appendix D).

3.3.5 Manually created database

A database was created with the purpose of including further data which may not yet have been curated from literature, and thus not yet found in existing BIO2RDF endpoints. This task was performed by manually curating eleven papers related to TKI and cardiotoxicity from which facts related to drug off targets, affected pathways, and clinical observations were extracted. The papers used to create the database include the work published in Chen *et al.* 2008, Choueiri *et al.*, 2010, Chu *et al.*, 2007, Force *et al.*, 2007, Force & Kolaja, 2011, Kerkela *et al.*, 2006, Korennykh *et al.*, 2009, Levitzki, 2013, Mellor *et al.*, 2011, Orphanos *et al.*, 2009, Spector *et al.*, 2007. This information was stored in a spreadsheet with the related publication ID until the contents of the spreadsheet were structured into RDF using OpenRefine v.2.5 (<http://openrefine.org/>) and the RDF Refine plugin (<http://refine.deri.ie/>). The produced RDF was then stored in a SPARQL endpoint.

3.3.6 Initial evaluation of HyQue-Cardiotoxicity's ability to score evidence for cardiotoxicity

HyQue's ability to assess the evidence supporting the cardiotoxicity of TKIs (Table 2-3) was compared against the rate of cardiotoxicity as reported in (Chen *et al.*, 2008). HyQue was run over 76 non-TKI cardiotoxic drugs (see Appendix E), and 39 non-TKI non-cardiotoxic drugs and nutraceuticals (see Appendix F) from which no toxicity was expected. The non-TKI drugs known to be cardiotoxic were chosen for the evaluation based on a paper which listed drugs known to be cardiotoxic (Fermini & Fossa, 2003). The non-TKI drugs and nutraceuticals that are not cardiotoxic were chosen after consultation with a pharmacist to identify drug classes with the least risk of being cardiotoxic. The non-TKI cardiotoxic drugs tested included doxorubicin, epirubicin, daunorubicin, idarubicin, fluorouracil, mitoxantrone, cyclophosphamide, phentermine, ergotamine, and paclitaxel. The non-TKIs drugs and nutraceuticals which were not expected to be cardiotoxic included drugs such as dexrazoxane, metformin, acetylsalicylic acid, acetaminophen, vitamin C, and vitamin A.

3.3.7 Creating lists of drugs for the comparison of HyQue-Cardiotoxicity with other methods

Lists consisting of 24 cardiotoxic (see Appendix G) and 30 non-cardiotoxic (see Appendix H) non-TKI drugs were created in order to allow the comparison of other toxicity methods to HyQue-Cardiotoxicity. To create the lists, DailyMed was searched using the advanced search function in order to create a list of cardiotoxic and non-cardiotoxic drugs. DailyMed is a resource run by the NIH which provides the most up to date drug information as well as access to drug package inserts

(<http://dailymed.nlm.nih.gov/dailymed/>). DailyMed's advanced search was used to look for the occurrence of the words "cardiovascular" or "card" within the adverse reactions, warnings and precautions, and boxed warning sections of package inserts to have the search return a list of drugs which would likely be cardiotoxic. The same was done using the "not in" modifier to create a list of drugs likely to not be cardiotoxic. The entries returned from the search result were then manually verified to make sure that the drug was cardiotoxic (or non-cardiotoxic) based on the information found within the package insert. Within the cardiotoxic set, 7 drugs could not be run in all three methods and 5 drugs in the non-cardiotoxic set could not be run for reasons such as some structures not having SMILES available, or the HyQue method not being able to complete its analysis for reasons still unknown (see Appendix I for an example). A spreadsheet was created which listed the drug name of the drugs found in DailyMed along with the DrugBank identifier and SMILES for the drug. The DrugBank identifier was used to create 30 hypothesis files for non-cardiotoxic drugs and 24 hypothesis files for the cardiotoxic drugs to act as the input for the HyQue-Cardiotoxicity method using a PHP script. The SMILES associated with the drugs were used as the input for the other two predictive methods; OpenTox and ACD/Labs's Percepta Predictor.

3.3.8 Other predictive methods used to compare with HyQue-Cardiotoxicity

OpenTox is a framework which hosts different applications to predict chemical toxicities (Hardy *et al.*, 2010). Within this framework, the Extended Cramer rules found under the Human Health Effects section of the tool were used to help evaluate the method's ability to discern if a drug is toxic based on chemical structure. Cramers Extended Rules. The second predictive method used was the ACD Lab's Percepta

Predictor v5.0.0.184 (<http://www.acdlabs.com/products/percepta/predictors.php>), a commercial predictor of ADME and toxicity endpoints, from which a compounds cardiovascular toxicity was assessed using the Health Effects module. The Health Effects module bases its predictions on a collection of more than 100,000 compounds that are associated with adverse events and toxicities found in the databases (http://www.acdlabs.com/download/docs/datasheets/datasheet_health.pdf). Both the OpenTox predictor and the Percepta platform were run using the SMILES associated with the drugs found within the cardiotoxic and non-cardiotoxic drug lists created using DailyMed, using the preset parameters of these approaches.

3.4 Results

In total, there were 20 rules used to evaluate the evidence in support of drug and TKI cardiotoxicity. These rules are divided into cardiotoxic events (15), and events found in the database (5). They are reported together in the final overall score. Each rule contributes 0.05 to the final overall score.

Functions found within the cardiotoxicity event evaluation have a weight of 5% within the event (or 1/20). The manually curated database contained only information on TKIs. As such, a maximum score of 15/20 (or 0.75) is possible for non-TKIs.

3.4.1 HyQue-Cardiotoxicity's evaluation of TKIs and other drugs

A few drugs were selected to help identify the most discriminating functions within the method, which included 8 TKIs, 9 non-TKI drugs known to be cardiotoxic, and 10 non-TKI drugs not known to be cardiotoxic (see Appendix J). After reviewing the

scores, the most discriminating functions were found to be the inhibited target, drug side-effect and predicted side-effect functions.

The inhibited target function, a function that verifies if it is known that a drug inhibits a known target involved in cardiotoxicity, was shown to be quite discriminating when identifying TKI related cardiotoxicity. Of the drugs reviewed, only TKIs had positive evidence for this function, and in particular, the TKIs known to be cardiotoxic (dasatinib, imatinib, sorafenib and sunitinib). For both the side-effect and predicted side-effect functions, all TKIs received a score of 1, with the exception of Lapatinib for the side-effect function. Non-TKIs known to be cardiotoxic all got a score of 1 for the known side-effects being cardiotoxic. Of the non-TKI drugs known to be cardiotoxic, only 3 of these drugs (fluorouracil, mitoxantrone and phentermine) also had predicted side-effects which were cardiotoxic. For non-TKI drugs known to not be cardiotoxic, only two drugs had known side-effects of cardiotoxicity (dexrazoxane and vitamin A), and 4 drugs (acetylsalicylic acid, vitamin A, econazole, and acetazolamide) had predicted side-effects associated with them. The cardiotoxic disease function which verifies if a drug is associated with a cardiotoxic MeSH term in CTD. It was found that all TKIs and non-TKI cardiotoxic drugs in the sample all had positive evidence for this function, whereas the non-TKIs not known to be cardiotoxic only had half of the drugs being associated with a cardiotoxic MeSH term (acetylsalicylic acid, ketoconazole, penicillin V and vitamin A). One of the least distinguishing functions was `gene_disease_association`, which verifies if genes targeted by a drug are also involved with a MeSH term linked to cardiotoxicity in the PharmGKB database. This is evidenced by the fact that all drugs in

the sample except for two non-cardiotoxic drugs (acetazolamide and insulin (regular)), were found to get a positive score for this function.

Although the HyQue-Cardiotoxicity method is meant to evaluate the evidence of cardiotoxicity, it could be interesting to compare the scores based on the evidence with the results described in literature. Results found for the run with TKIs are in accordance with the current literature (Chen *et al.*, 2008) describing the cardiotoxicity of TKIs, except in the case of imatinib (see Table 3-3). Within the manually curated database, imatinib is the most represented drug, and gets a positive score for evidence supporting its cardiotoxicity for each of the 5 functions within the curated database rule (see Table 3-4). The presence of positive evidence in the database for all 5 database functions would thus contribute 0.25 to the overall score. Evidence found in the database are from various journal publications, and the observations made within these publications may not yet been evaluated or accepted as drug related cardiotoxic events for this drug. Also, the article which is being used to compare HyQue's classification was published in 2008 (Chen *et al.*, 2008), and other evidence. Thus a re-evaluation of a drugs toxicity may one day reveal a higher level of toxicity.

Apart from imatinib, sunitinib and sorafenib were found by this approach to have the highest score. Sunitinib is a multitarget TKI known to be associated with cardiac adverse events such as changes in left ventricular ejection fraction (LVEF), chronic heart failure and hypertension. The cardiotoxicity associated with Sunitinib was found to be due to off-target action of sunitinib, where it would inhibit the 5' adenosine monophosphate-activated protein kinase (AMPK) (Kerkela *et al.*, 2009). Lapatinib was the TKI with the lowest score, meaning that it had the least amount of evidence

supporting its cardiotoxicity. The low score for lapatinib is in accordance with an analysis of 44 clinical trials performed for this drug (Perez *et al.*, 2008). The study concluded that lapatinib was estimated to have a low level of cardiotoxicity as the observed toxicity involved only a relatively low decrease in LVEF in a few patients, which had also been reversible and asymptomatic. For erlotinib and gefitinib, TKIs which target the epidermal growth factor receptor (EGFR), this method did not find there was much evidence supporting their cardiotoxicity, which is also in accordance with what is known for these drugs (Orphanos *et al.*, 2009).

3.4.2 HyQue-Cardiotoxicity's evaluation of non-TKI cardiotoxic drugs

When the approach is run on the dataset of drugs which are not TKIs, but are known to be cardiotoxic, mitoxantrone received the highest score based on evidence with a score of 0.32 (see Table 3-5), followed by doxorubicin. Mitoxantrone is an anthraquinone derivative which has been associated with a high risk of cardiovascular events such as chronic heart failure (Gharib & Burnett, 2002). Doxorubicin is an antineoplastic agent known to be cardiotoxic, and responsible for causing cardiomyopathy. It was recently suggested that the cause of this cardiotoxic effect is due to iron accumulating within the mitochondria (Ichikawa *et al.*, 2014). The lowest score within this dataset was for idarubicin, with a score of 0.12, followed by ergotamine with a score of 0.15. Ergotamine is an alkaloid used to treat migraines. Ergotamine is safe for use with most patients, however has been associated with myocardial infarction where abuse has been identified (Roithinger *et al.*, 1993). The low score for this drug can thus be explained by the fact that cardiovascular toxicity is only really reported when using

ergotamine at higher doses, and thus the likeliness of finding evidence supporting cardiotoxicity is low.

3.4.3 HyQue-Cardiotoxicity's evaluation of non-TKI non-cardiotoxic drugs

Within the non-TKI dataset of drugs which are not identified as being cardiotoxic, there were some surprising results (see Table 3-6). For example acetaminophen had evidence of cardiotoxicity. A study which monitored the occurrence of major adverse cardiac events related to the use of acetaminophen in females demonstrated that the drug was associated with a dose-related increase in risk when 15 tablets or more were consumed per week (Chan *et al.*, 2006). Upon further review of the data used to evaluate the cardiotoxicity associated with acetylsalicylic acid and acetaminophen, it would seem that the data obtained from CTD, SIDER, and PHARMGKB (also contains OFFSIDES results) would be the sources of the positive evidence of cardiotoxicity. For example in the case of acetaminophen, the drug was found to affect a few pathways associated with heart toxicity such as the MAPK and the protein processing in endoplasmic reticulum pathways. The drug was also found to have cardiotoxic side-effects in SIDER such as tachycardia and hypertension. Predicted side-effects for acetaminophen identified by OFFSIDES were also found such as bradycardia, and the PharmGKB contained a gene-disease association for this drug due to the fact that it acts on PTGS1 (prostaglandin H2 synthase 1), a gene found to be associated with myocardial infarction in this knowledge base.

3.4.4 HyQue-Cardiotoxicity's ability to differentially score cardiotoxic and non-cardiotoxic drugs

In order to determine if HyQue-Cardiotoxicity could differentially score cardiotoxic and non-cardiotoxic drugs, the distribution of the scores of cardiotoxic and non-cardiotoxic drugs was determined. As the distributions were both normal distributions, the t-test was used in order to see if the two distributions were significantly different (see Table 3-7). The t-test resulted in a p-value of 0.0009. In order to further confirm HyQue's ability to evaluate evidence of cardiotoxicity, HyQue-Cardiotoxicity was compared to two toxicity predictors, ACD labs Percepta platform and Cramers Rules found in the OpenTox predictor.

3.4.5 HyQue-Cardiotoxicity in comparison with other methods

A toxicity evaluation was performed for 17 cardiotoxic drugs and 25 non-cardiotoxic drugs using HyQue-Cardiotoxicity, Cramers Rules (found in OpenTox) and the Percepta platform to see if the predictive methods could discern between the lists of cardiotoxic and non-cardiotoxic drugs.

Within the cardiotoxic set of drugs, the Percepta platform developed by ACD Labs considered 14 drugs to have a toxic effect on the cardiovascular system with a probability of 0.5 or higher (see Table 3-8). Three drugs were predicted to not have a probability to be toxic to the cardiovascular system with three drugs (meloxicam, cyclophosphamide, ifosfamide) having a predicted probability being less than 0.1. Of the 14 drugs identified by the Percepta Platform with probabilities of affecting the cardiovascular system, five were found to have a full probability (1.0) of being toxic to

the cardiovascular system, which included haloperidol, Nuvaring, epirubicin hydrochloride, doxorubicin, and daunorubicin. The OpenTox method which made use of Cramers Extended rules had classified all drugs as being Class III. HyQue-Cardiotoxicity's evaluation returned no evidence for cardiotoxicity for 6 drugs (score <0.05), and weak evidence for the remaining 13 drugs (score <0.4). Within the evaluation using HyQue-Cardiotoxicity, the highest score was 0.33 for haloperidol, which had received a probability of 0.93 for affecting the cardiovascular system by the Percepta platform. The HyQue-Cardiotoxicity method evaluated four drugs to have no evidence of cardiotoxicity, everolimus, saphris, diclofenac and amphetamine.

The non-cardiotoxic set of drugs (see Table 3-9) was found to be cardiotoxic for 13 drugs (probability of 0.5 and over) when evaluated by the Percepta platform. Only two non-cardiotoxic drugs were found to have a probability of less than 0.1. The OpenTox evaluation led to all drugs being classified as Class III (high toxicity) as a result of Cramer's Rules, with the exception of tretinoin, which was classified as Class I (low toxicity). The HyQue-Cardiotoxicity found 13 non-cardiotoxic drugs to have absolutely no evidence of cardiotoxicity and 10 drugs to have weak evidence of cardiotoxicity, the highest score being 0.167 for miconazole.

Table 3-1: Datasets used for evidence in the scope of this project

Dataset	Endpoint url	Original database url
DrugBank (Law <i>et al.</i> , 2014)	http://cu.DrugBank.bio2rdf.org/sparql	http://www.DrugBank.ca/
ChEMBL (Gaulton <i>et al.</i> , 2011)	http://www.ebi.ac.uk/rd/services/chembl/sparql	https://www.ebi.ac.uk/chembl/
Clinicaltrials.gov	Local endpoint	http://clinicaltrials.gov/
CTD (Davis <i>et al.</i> , 2013)	http://cu.ctd.bio2rdf.org/sparql	http://ctdbase.org/
Manually curated database	Local endpoint	N/A
MGI	Local endpoint	http://www.informatics.jax.org/
OFFSIDES (Tatonetti <i>et al.</i> , 2012)	http://cu.pharmgkb.bio2rdf.org/sparql	http://www.pharmgkb.org/downloads/
PharmGKB (Whirl-Carrillo <i>et al.</i> , 2012)	http://cu.pharmgkb.bio2rdf.org/sparql	https://www.pharmgkb.org/
PubMed	http://cu.pubmed.bio2rdf.org/sparql	http://www.ncbi.nlm.nih.gov/pubmed
SIDER (Kuhn <i>et al.</i> , 2010)	Local endpoint	http://sideeffects.embl.de/
TWOSIDES (Tatonetti <i>et al.</i> , 2012)	http://cu.pharmgkb.bio2rdf.org/sparql	http://www.pharmgkb.org/downloads/

Table 3-2: Categorization of evidence scores for drug cardiotoxicity

Score	Strength of Evidence	Definition
<0.05	None	When the evidence does not (or barely) support the hypothesis that a drug is cardiotoxic.
0.05-0.4	Weak	When there is little evidence to support the hypothesis that a drug is cardiotoxic.
0.4-0.6	Intermediate	When the evidence somewhat supports the hypothesis that a drug is cardiotoxic.
0.6-0.8	Strong	There is a sufficient amount positive support to suggest that a drug is cardiotoxic.
>0.80	Strongest	There is a large body of positive evidence to support that a drug is cardiotoxic.

Table 3-3: Result of evaluating the evidence for TKI induced cardiotoxicity

TKI	Score	HyQue-Cardiotoxicity's classification of cardiotoxicity based on the score (see Table 2)	Known cardiotoxicity based on Chen <i>et al.</i>	Rate of cardiotoxicity based on Chen <i>et al.</i>
dasatinib	0.50	Intermediate	Yes	low-moderate
erlotinib	0.22	Weak	No	N/A
gefitinib	0.22	Weak	No	N/A
imatinib	0.63	Strong	Yes	low
lapatinib	0.12	Weak	No	N/A
nilotinib	0.33	Intermediate	Yes	low
sorafenib	0.52	Intermediate	Yes	low
sunitinib	0.48	Intermediate	Yes	moderate

Table 3-4: Database rules which seem to be responsible for imatinib's higher than expected score

Database evidence rule	Score
hasCardiotoxicEffectScore	1
hasCardiotoxicTargetScore	1
hasClinicalObservationScore	1
hasCardiotoxicOffTargetScore	1
hasCardioProtectiveOffTargetScore	0

Table 3-5: Result of evaluating the evidence for non-TKI drugs known to be cardiotoxic inducing cardiotoxicity

Drug	Score	Strength of cardiotoxic evidence
daunorubicin	0.22	Weak
doxorubicin	0.27	Weak
epirubicin	0.17	Weak
ergotamine	0.15	Weak
fluorouracil	0.22	Weak
idarubicin	0.12	Weak
mitoxantrone	0.32	Weak
paclitaxel	0.15	Weak
phentermine	0.17	Weak

Table 3-6: Result of evaluating the evidence for non-TKI drugs not identified to be cardiotoxic inducing cardiotoxicity

Drug	Score	Strength of cardiotoxic evidence
acetylsalicylic acid	0.18	Weak
acetaminophen	0.32	Weak
acetazolamide	0.12	Weak
dexrazoxane	0.05	Weak
fluticasone	0.10	Weak
insulin (regular)	0.00	None
ketoconazole	0.12	Weak
metformin	0.32	Weak
rosuvastatin	0.10	Weak
vitamin A	0.27	Weak

Table 3-7: Comparison of evidence scores for cardiotoxic and non-cardiotoxic drugs

	Cardiotoxic	Non-Cardiotoxic
Number of drugs	76	39
Mean Score	0.20 ± 0.07	0.15 ± 0.06
Normality	0.98	0.96
T-Test Result	P = 0.0009	

Table 3-8: Results of evaluations made by HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for cardiotoxic drugs

DrugBank ID	Drugname	Percepta platform	Cramer's Extended Rules (OpenTox)	HyQue - Cardiotoxicity
DB01590	everolimus	1	High (class III)	0.00
DB00502	haloperidol	0.93	High (class III)	0.33
DB00814	meloxicam	0.03	High (class III)	0.05
DB00294	Nuvaring	1	High (class III)	0.05
DB06216	Saphris *	0.94	High (class III)	0.00
DB00991	oxaprozin	0.38	High (class III)	0.07
DB00445	epirubicin hydrochloride	1	High (class III)	0.17
DB00586	diclofenac	0.88	High (class III)	0.00
DB00997	doxorubicin	1	High (class III)	0.27
DB01204	mitoxantrone	0.95	High (class III)	0.22
DB00182	amphetamine	0.66	High (class III)	0.00
DB01577	methamphetamine	0.61	High (class III)	0.10
DB00694	daunorubicin	1	High (class III)	0.17
DB01181	ifosfamide	0.06	High (class III)	0.07
DB01101	capecitabine	0.72	High (class III)	0.07
DB01099	flucytosine	0.78	High (class III)	0.05
DB00252	phenytoin	0.56	High (class III)	0.05

* A similar structure used in OpenTox as the same structure not available. Structure resemblance was of 0.9 or higher.

Table 3-9: Results of evaluations made by HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for non-cardiotoxic drugs

DrugBank ID	Drug	Percepta platform	Cramer's Extended Rules (OpenTox)	HyQue - Cardiotoxicity
DB00223	diflorasone	1	High (class III)	0.00
DB00254	doxycycline	0.36	High (class III)	0.15
DB00595	oxytetracycline	0.95	High (class III)	0.05
DB01045	rifampicin	0.98	High (class III)	0.05
DB01110	miconazole	0.94	High (class III)	0.17
DB00836	loperamide	0.98	High (class III)	0.00
DB01075	diphenhydramine	0.98	High (class III)	0.12
DB00684	tobramycin	1	High (class III)	0.00
DB00588	fluticasone propionate	1	High (class III)	0.10
DB01106	levocabastine	0.86	High (class III)	0.00
DB00755	tretinoin	0.48	Low (Class I)	0.05
DB00532	Mesantoin	0.18	High (class III)	0.00
DB00253	medrysone	0.96	High (class III)	0.00
DB01157	trimetrexate	0.97	High (class III)	0.02
DB00545	pyridostigmine	0.6	High (class III)	0.12
DB00764	mometasone	1	High (class III)	0.00
DB06817	Isentress	0.05	High (class III)	0.00
DB00600	monobenzone	0.47	High (class III)	0.05
DB00258	calcium acetate	0.04	High (class III)	0.00
DB00446	chloramphenicol	0.7	High (class III)	0.00
DB01294	bismuth subsalicylate	0.33	High (class III)	0.00
DB00763	methimazole	0.25	High (Class III)	0.02
DB06699	degarelix	0.54	High (Class III)	0.02
DB00682	warfarin	0.41	High (Class III)	0.07
DB01032	probenecid	0.34	High (Class III)	0.02

Table 3-10: Summary of performance of HyQue-Cardiotoxicity, Cramers Extended Rules and the Percepta Platform for non-cardiotoxic drugs

Method	Methods Measure	Result Range for evaluation	Mean
Cramers Rules	Classification (I, II, III)	Class I or Class III (high)	N/A
HyQue-Cardiotoxicity	Evidence score (0.00-1.00)	0.00 - 0.17	0.04 ± 0.05
Percepta Platform	Probability (0.00-1.00)	0.04 - 1.00	0.65 ± 0.34

3.5 Discussion

3.5.1 Scoring of evidence

Within this approach, scoring was based on the presence of evidence. Each of the functions were worth 1 part of 20 (or 0.05), added together to reach a potential overall final score of 1. A maximum score (score of 1), would require that there exist positive evidence spanning all 20 rules. This would require that there is evidence that a drug is cardiotoxic in each of the various datasets, as well as in the curated database. This case is unlikely as it would require that a drug be well known and understood, and that a drug's cardiotoxicity be established and documented. The reason why this would not be the case is due to the fact that understanding a drug's action within the body can be very complex. A drug's action can affect many different aspects of the human body from the genetic level to the phenotypic level (Bai & Abernethy, 2012), thus making it necessary to consider all of the possible affectations that a drug can directly, indirectly or

unknowingly cause. For example drug effects can be the result of off-targets and interactions, or a misunderstood mechanism of action. As such, some of the rules conceived to verify whether a certain aspect of known cardiotoxicity is associated with a drug, such as inhibiting the hERG ion channel, may result in a score of 0 for that function as the root of a drugs' cardiotoxicity may not be due to the inhibition of this channel.

Within the scoring approach used in this method, there is the possibility that there is dependence between variables that could be confounding the final scores of this approach. For example, if the same evidence is present in two resources due to the duplication in a database from information found at the second database, this would lead to evidence being indirectly evaluated twice which would affect the overall evaluation. The presence of dependence between variables was not evaluated in this work; however it could be studied using Principle Component Analysis (PCA). Studying the rules using Principle Component Analysis could identify the features which are the most informative to evaluating cardiotoxicity, as well as ensure that the features are not correlated or dependent on one another (Jolliffe, 2002).

Scoring could also be further improved by weighing rules differently based on the degree to which a feature is associated to cardiotoxicity. This would allow for a type of evidence which is more associated with an increased risk of cardiotoxicity to be weighted higher than a type of evidence which does not contribute as much to the risk of developing cardiotoxicity.

A second aspect of scoring which would need to be addressed would include a measure of the statistical significance for the evaluation of the cardiotoxic and non-cardiotoxic datasets. This is because a measure of significance would have provided a

measure for how likely (or unlikely) that a drug would have had substantial evidence (or not) of cardiotoxicity by chance. For example the DrugBank database has information on thousands of drugs, yet only 76 cardiotoxic and 39 non-cardiotoxic drugs were evaluated during this evaluation. A way to take into consideration the size of the database could include developing a measure which is similar to BLAST's E-Value (Madden, 2003), which is a measure of the significance of an alignment that also considers the size of the database. Assigning a level of significance of assessments made by the HyQue-Cardiotoxicity approach which would also consider the size of the database would demonstrate whether or not the evaluations made by the approach are significant.

3.5.2 Null functions in ruleset

A few of the functions obtained scores of 0 for all drugs within the sample. These functions included cardiotoxic pathway in PharmGKB, and cardiotoxic mouse phenotypes. For the case of the cardiotoxic mouse phenotype rule, this is suspected to be due to the use of presence tests to evaluate phenotypes, which can lead to the lack of reporting of negative results within databases. Unfortunately after applying restrictions such as the mouse model having to be a knockout and be associated with a phenotype which appears in the curated cardiotoxic phenotype list, no mouse models currently fulfill these evaluation criteria. For the cardiotoxic pathway in PharmGKB rule, the lack of hits could be caused by the list of pathways not being extensive enough to cover all pathways related to cardiotoxicity. The null functions will be kept in this approach in the eventuality that more information about a drug becomes available in the future which could serve as positive evidence of a drug being cardiotoxic.

3.5.3 Comparing methods for the list of cardiotoxic drugs

When comparing HyQue-Cardiotoxicity to the ACD lab's Percepta platform and the Cramers Extended Rules found in the OpenTox framework, it can be seen that assessments vary between platforms (see Table 3-8). The ACD Percepta platform seems to predict a high amount of drugs from the cardiotoxic dataset as being cardiotoxic, whereas HyQue-Cardiotoxicity does not find any evidence of cardiotoxicity for 4 drugs within this dataset, and next to no evidence for another 4 drugs (score 0.05 or less). The Extended Cramers Rules in OpenTox deemed all drugs within the cardiotoxic dataset as being cardiotoxic, but also does so for non-cardiotoxic drugs.

3.5.4 Comparing methods for the list of non-cardiotoxic drugs

The HyQue-Cardiotoxicity approach seems best able to identify non-cardiotoxic drugs (Table 3.9). The HyQue-Cardiotoxicity method was able to evaluate the majority of non-cardiotoxic drugs as not having any (or much) evidence of cardiotoxicity. In fact, 11 of the 25 drugs had no evidence of cardiotoxicity (score of 0.00). The Percepta platform oddly evaluated a few drugs in the non-cardiotoxic dataset with a full probability (probability of 1) for affecting the cardiovascular system which included diflorasone, tobramycin, fluticasone propionate, mometasone, and also had a high mean score and high score deviation for non-cardiotoxic drugs (0.65 ± 0.34), see Table 3-10. The Extended Cramers Rules deemed all drugs within the non-cardiotoxic dataset as being cardiotoxic except for tretinoin, which was classified as Class I. As such, the Extended Cramers Rules did not perform well on this dataset.

3.5.5 Overall method comparison result between methods

Overall, HyQue-Cardiotoxicity is the approach best able to differentiate between cardiotoxic and non-cardiotoxic drugs (see Table 3-10). The HyQue-Cardiotoxicity method seems to better evaluate evidence for non-cardiotoxic drugs in comparison to cardiotoxic drugs. A possible explanation for this would be that the rules designed within the HyQue-cardiotoxicity project may need to be extended to include a more extensive list of side-effects, drug targets, genes and pathways linked to cardiotoxicity to improve the ability to score evidence of cardiotoxicity when appropriate. As the rules are designed to allow for the evaluation of evidence which would be comparable to that of a domain expert, it would be important to ensure that the manually curated lists, for example the list of cardiotoxic drug effects, cover all effects related to cardiotoxicity. As such, a review of the current lists of effects, targets, genes and pathways by a cardiotoxicity expert such as a physician could possibly add more terms to the curated lists which could help improve the HyQue-Cardiotoxicity approach's ability to identify evidence of drug-related cardiotoxicity. Another possibility would be to extend the rules to include more datasets and sources of evidence which could be evaluated by the HyQue-Cardiotoxicity method.

Upon review, it would seem that the Percepta platform and Cramers Extended Rules are unlikely to be able to differentiate between cardiotoxic and non-cardiotoxic drugs. Cramers Extended Rules evaluated all drugs as being cardiotoxic except for tretinoin, demonstrating its inability to differentially score cardiotoxic and non-cardiotoxic drugs. The Percepta platform also seems to have a bias towards classifying drugs as being cardiotoxic as several non-cardiotoxic drugs had a high probabilities of

being cardiotoxic. As both of these approaches rely only on the structure of the drugs, it is possible that the evaluation of the drug structure prior to being metabolized in the body could be the cause for the both approaches having a decreased ability to discern between cardiotoxic and non-cardiotoxic drugs. Patient-related factors such as age, gender, race, and pre-existing medical conditions, as well as other factors such as drug response linked to the dosage administered (Alomar, 2014) can contribute to the added complexity of predicting drug toxicities which cannot be accounted for when using a method based solely on structure-based methods to predict toxicity. It should also be mentioned that a few drugs found within the non-cardiotoxic dataset were topical treatments such as diflorasone, which are not considered in Cramers Extended rules as the rules were created for the assessment of drugs taken orally (Lapenna & Worth, 2011), such as doxycycline and loperamide. As drugs taken orally found within the non-cardiotoxic evaluation set such as loperamide were still considered to be of Class III and had obtained a high probability of negatively affecting the cardiovascular system (score of 0.98 given by the Percepta platform, see table 3-9), it would seem that the drug administration is unlikely the cause for the inaccurate predictions.

3.5.6 Caveats of HyQue-Cardiotoxicity

Although the HyQue-Cardiotoxicity approach was able to evaluate the evidence of drug related cardiotoxicity of TKIs approaching that of what is known in literature, it does have a few caveats. One caveat is that the manually created database only contains information on TKIs. Thus, any drug that is not in this database can only get a maximum score of 0.75 as the database is worth 0.25 of the final overall score. Ways of reducing this potential problem could be to extend the amount of curated knowledge to include

evidence of cardiotoxicity from other drugs which are not TKIs, or to not run the rules based on the curated database when a set of drugs does not contain TKIs.

A second issue which was found with drugs which were recently marketed as drug related information had not yet been gathered and added into online databases such as CTD. An example of such a case is ruxolitinib, a TKI, for which the entry found in CTD was still bare at the time of evaluation. This type of event also affects cases where rules depend on cross-references from one database for another to retrieve further information, such as when using a DrugBank URI as an input to retrieve drug information in CTD. For example the drug vandetanib did not have a cross-reference from DrugBank to CTD at the time the method was run.

Another issue with using information found in many databases is that databases do not tend to report negative results. For example many databases related to animal models only report phenotypes which were observed during the evaluation, and not those which were absent. As such, it is unknown whether the result of a test which was done for a specific drug was negative or if the test was simply not done.

3.5.7 Advantages of HyQue-Cardiotoxicity

Although the use of public data is not perfect, making use of Linked Open Data is still advantageous. One of the advantages of the HyQue-Cardiotoxicity approach and its use of Linked Open Data is that the evaluations will most likely improve as the information found in public databases grows. A second advantage of this approach is its ease of expandability which allows for further functions to be added including new databases or rules as more knowledge becomes available. Finally, another advantage of

this approach is its ability to incorporate data from various sources to help identify drug safety issues earlier than what could have been identified if only one type of evidence had been considered.

This method is by no means attempting to predict or classify drugs based on their cardiotoxicity as this would best be performed by computational methods based on machine learning methods, or logistic regression. This method is a means of providing a baseline of evidence of a drugs cardiotoxicity, and could serve itself during the feature selection stage when developing prediction models of cardiotoxicity. A current problem with computational predictive methods is the task of feature selection when domain knowledge is not available. With the large amount of data which is available in the biological sciences as well as the sometimes small sample sizes, it can be difficult to identify the most relevant variables to use when making predictions, for example when making predictions using data from microarray experiments (Hua *et al.*, 2009). The importance of feature selection within predictive methods is that it is correlated with a model's predictive performance, but can be difficult to do without domain specific knowledge (Guyon & Elisseeff, 2003). Currently, feature selection to make a model more efficient at making predictions can be done by using wrappers and filters, but use of these methods may lead to the removal of variables which are dependent of other variables, or increase the risk of overfitting the data (Saeys *et al.*, 2007). As such, it is hypothesized that HyQue-Cardiotoxicity could be used in order to assist in the identification of features for predictive models of cardiotoxicity.

3.6 Conclusion

In conclusion, the HyQue-Cardiotoxicity approach is able to differentially score cardiotoxic and non-cardiotoxic drugs by evaluating Linked Open Data. As the knowledge related to drugs continues to grow, the evaluation of this approach should improve as there will be more evidence available to support or refute the hypothesis that a drug is cardiotoxic. This approach has the benefit of being easily expandable to include further sources of evidence as they become available in the future. Future endeavors would include consulting with a pharmaceutical expert or cardiologist to help identify more sources of evidence to evaluate cardiotoxicity and expanding the manually curated lists in the event that they do not currently identify all known effects, pathways and genes involved in cardiotoxicity. An analysis of the rules to evaluate the evidence of drug-related cardiotoxicity would ensure that the current evidence being evaluated is uncorrelated.

4 Chapter: Conclusion

4.1 Overall summary

Within this work, I hypothesized that the use of Semantic Web technologies will overcome the limitations in existing approaches and enable new applications in drug discovery and drug safety. This thesis demonstrates that semantic technologies can be applied to overcome the limitations of using biomedical big data, such as the issue of heterogeneity in data formats. For example, biomedical data made available by Bio2RDF, for which the representation of the data was normalized and the cross-references between the databases, allowed for querying across various datasets for related data, such as from a drug in DrugBank to its targeted gene in PharmGKB. Overcoming the limitations of using big data is an important task as it is a current bottleneck to gaining new knowledge or investigating new hypotheses is the heterogeneity between data sources.

As a use case, this thesis applied semantic technologies to drug target discovery and drug safety profiling. The integration of various sources of data was done by the development of mappings between ontologies within the PhenomeDrug project and by the development of the HyQue-Cardiotoxicity approach which queried various datasets for existing knowledge prior to evaluation.

The importance of this integration in pharmaceutical research is that this domain requires that knowledge from both the chemical and biological points of view be integrated in order to gain a better understanding of underlying mechanisms. These approaches to integrating datasets are not only restricted to applications in pharmacology, but can be applied to many domains such as building management for studying building energy consumption (Curry *et al.*, 2013), in public health when analyzing government

survey data to associate regional childhood obesity to the proximity of fast food outlet types (Shadbolt & O'Hara, 2013), and in disaster management to identify early disaster warnings (Silva *et al.*, 2013).

Further work is still required in order to enable semantic technologies to make use of big data. Such issues include increasing the amount of mappings between datasets as well as improving the quality of existing and future mappings, increasing the amount of data available as RDF, as well as encouraging researchers to publish their data openly.

In summary, creating quality semantic mappings between ontology terms enabled datasets to be connected that enabled the association of model phenotypes to drug effects to identify drug targets. The use of Linked Open Data as evidence enabled the evaluation of drug-related toxicity within the HyQue-Cardiotoxicity approach.

4.1.1 Summary of PhenomeDrug Chapter

Existing approaches for the identification of drug targets are often based on drug-drug or disease-disease similarity, whereas the PhenomeDrug approach uses the similarity between drug effects and phenotypes observed in model organisms. Large amounts of data are available from animal model organisms, for example mouse models found in MGI. This data is useful as animal models have been used in the past to identify new drugs (as well as for safety profiling).

The PhenomeDrug method was able to identify human drug targets by associating mouse model phenotypes to drug effects. This required that mappings be made between UMLS side effect concepts and ontologies. Mapping these terms allowed for creating the interoperability needed between drug effects and mouse phenotypes to allow for the

similarity between the two to be calculated leading to the identification of human drug targets which was the goal of this approach. Lexical mapping and existing cross references between UMLS and HPO described in HPO were first leveraged to map terms, however 2132 UMLS terms still remained unmapped. The unmapped terms were mapped manually in order to increase the number of terms which could be used within this approach. The majority of unmapped terms were due to UMLS terms containing medical language not used to describe phenotypes such as medical events related to the use of medical devices such as IUD expulsion. A common reason for inexact matches was due to the vocabulary between the two ontologies differing, for example how the HP and MP ontologies are used to describe abnormalities such as the term “abnormal blood vessel morphology” described in MP.

To investigate the quality of the manual mappings to HP, the manually mapped HP terms were compared with those done automatically. Overall, the manual mappings allowed for better mappings than those produced by the computational method for the terms verified. This assessment is based on manual revision of the automatic mappings. This was largely due to the automated mapping method creating several mappings for a single UMLS term which tended to be either too general or too specific. For example the automated mapping for the HP term “Lactic Acidosis, Chronic” is too specific for the UMLS term “Lactic Acidosis”. Clearly incorrect mappings were also made by the automatic method, for example between the UMLS concept “celiac disease” and the HP term “Arteriovenous Fistulas Of Celiac And Mesenteric Vessels”. The incorrect mappings made by the automatic method demonstrate the need for a review of mappings by an expert to ensure quality mappings. Even though the manual mappings seemed more

accurate, mapping ontologies manually is not an easy feat and does have disadvantages. Manual mappings are time consuming thus it would not be efficient to map two large ontologies to one another manually, and manual mappings are prone to human error. A combination of automated mappings followed by manual review by an expert is likely to produce better quality mappings in a shorter amount of time.

4.1.2 Summary of HyQue-Cardiotoxicity

The HyQue-Cardiotoxicity approach is able to differentially score cardiotoxic and non-cardiotoxic drugs by evaluating Linked Open Data. This approach is based on the retrieval, evaluation and scoring of evidence related to drug-related cardiotoxicity. By leveraging HyQue, the hypothesis of a drug being cardiotoxic can be supported or negated based on evidence found as Linked Open Data. By using data from various data sources spanning different levels affected by cardiotoxicity from the cellular to organism level, a more comprehensive evaluation based on evidence can be made to assess a drug's cardiotoxicity. What distinguishes this approach from others is that not only can it make use of data found in public databases and integrate them into an evaluation, but this approach allows for the functions which make up the evaluation to be broken down so that specific pieces of evidence can be reviewed. As such, this method demonstrates the usefulness of using the semantic web and various datasets in drug discovery, in particular within the domain of drug safety.

As the knowledge related to drugs continues to grow, the evaluation of this approach should improve as there will be more evidence available to support or refute the hypothesis that a drug is cardiotoxic. This approach has the benefit of being easily

expandable to include further sources of evidence as they become available in the future. Future endeavors would include consulting with a pharmaceutical expert or cardiologist to help identify more sources of evidence to evaluate cardiotoxicity, and possibly expanding the current manually curated lists if they do not currently identify all known effects, pathways and genes which are involved in cardiotoxicity. Another aspect which could greatly improve this method would be to study each function separately, and research evidence types to see if certain functions are better types of evidence of cardiotoxicity over others. In these circumstances, the value of each function would be altered to be weighted heavier than other functions.

4.1.3 Future work and insight

In order for future work involving semantic approaches such as PhenomeDrug and HyQue-Cardiotoxicity to be successful, certain issues will need to be resolved. For example the issue of interoperability between species-specific ontologies will need to be improved. This issue became obvious when mapping terms between species-specific ontologies as several difficulties and inconsistencies were present. One issue encountered was the meaning of terms. For example the HPO term “Proptosis”, which is also found in UMLS. Albeit the two ontologies both have terms with this same label, they have definitions which distinctively vary. The definition for the HP term is only for one eye “An eye that is protruding anterior to the plane of the face to a greater extent than is typical.”, whereas the UMLS term described two eyes “Abnormal protrusion of both eyes”. A second example of different meanings is found in the case of mapping using synonyms. For example, the HPO term “intellectual disability” (HP:0001249) can be

mapped to the UMLS term “Mental deficiency” (UMLS: C0025362) through the HP synonym “Mental deficiency”. Notice that the meaning between intellectual and mental are not one of the same, and neither is a disability and a deficiency as these words have slight nuances with distinguish them in their meaning. A way of overcoming such issues would be specify the relation type between terms mapped by defining them using relations described by SKOS (Alistair *et al*, 2005), OWL or RDFS (Allemang & Hendler, 2011). Defining the relation type between mapped terms which are not directly equivalent allows for more accurate mappings. For example non-equivalent terms can be identified as being a subclass of the other term, or as a related term. This would be necessary while mapping ontologies as equivalent mappings are not always possible based on the current terms available in ontologies. A way to help integrate terms between species-specific ontologies would possibly be to create an overall ontology which could contain terms which could be used to describe the majority of animal models. Such an ontology is the UBERON ontology which integrates anatomy terms across species, which could be used within approaches which would make use of cross-species data for inferencing (Mungall *et al*, 2012). Similarly, to integrate phenotypes across species, a cross-species phenotype ontology called UberPheno was developed to provide terms applicable to model organisms such as mice, zebrafish and humans (Köhler *et al*, 2013).

In a collaborative mapping effort, several activities would need to take place in order to create the best quality mappings. One such activity would be the ability of mappings to be rated by experts involved in the project, which would involve the collaboration of ontologists, curators as well as domain specific experts. This step would allow the input of experts across different fields with different knowledge to share their

input and concerns which may address issues such as the meaning of terms, the level of specificity, or considering the hierarchy of mapped terms between ontologies. The rating of terms would allow for terms to be mapped based on agreement, and mappings associated with disagreements would not be included in the final mappings, and could be tagged as problematic, and open to review. The idea of allowing for the open review of data to improve data selection and evaluations is a small part of the overall vision of Force11, an interdisciplinary group which seeks to address issues with scholarly publishing (Bourne *et al.*, 2011). Beyond the ability to rate the mappings would be ability to suggest an improved mapping or new terms which could then be submitted to the involved ontologies. This would allow for the utmost best mapping to be used in order to bridge terms between ontologies to improve interoperability and leverage the ability to increase knowledge and understanding.

Another avenue for future work would include increasing the amount of data currently available as Linked Open Data. This is an important step to improving new approaches which rely on LOD as it allows for a larger amount of data to be used for predictions, to identify new correlations or to support or refute hypotheses. The need for more data and the usefulness of Linked Data is slowly becoming more obvious as large databases such as ChEMBL (Gaulton *et al.*, 2011) and EBI (Jupp *et al.*, 2014) have made the data found in their databases available as RDF. There are also groups such as Bio2RDF (Callahan *et al.*, 2013a) and Chem2Bio2RDF (Chen *et al.*, 2010) who are making the data found in important databases, which are often curated, available in RDF.

Although the data found in larger databases is being represented in RDF, data from individual researchers, which may span several years of research, are not available

as Linked Open Data. Possible hindrances to sharing and representing data in RDF by the scientific researchers in the life sciences can include issues such as not knowing about the semantic web, not knowing how to represent data in RDF, and not wanting to share data openly with others due to the level of competition when publishing research. The issue of expressing data in RDF to the scientific researchers without a background in computer science can be quite problematic. In order to allow researchers to publish their data, it would be of interest to create an application which could convert research data into RDF. This application would need to be able to take in different file formats such as spreadsheets and have a user friendly interface which would allow a user with little informatics experience to easily select settings and options for the data conversion. Another feature could include a preview screen so that the user can visualize the output prior to it being generated. This application would also need to have the ability to load and store the data with little involvement from the user. For example to load the data formatted into RDF into a triple store so that it can later be queried in the future. The suggested application would still require the user to have some understanding of Linked Data and RDF representation. However, it would allow more researchers to publish their data due to the reduced requirement for understanding and implementing programming. Such an application could consist of a workflow, which take an input and follows a series of predefined steps in order to generate an output. An example is the virtual research environment myExperiment (De Roure *et al*, 2009) which allows users to complete tasks using workflows, which offers some of the characteristics mentioned earlier such as a simple interface (De Roure *et al*, 2009).

A second issue foreseen by the scientific researchers with Linked Open Data would be having to share data openly. This can induce fear within a domain which prides itself on publications and making new discoveries. Only when findings are published and credited are they usually shared (Mons *et al*, 2007). In order to encourage sharing, it would be of interest to either impose or encourage the publication of research results as open Linked Data after a certain time frame after publication, for example within a year of publication. This would allow the researcher to first publish their paper without the risk of losing recognition for their research. A suggestion for associating data or findings to its author is by the use of nanopublications, which is a statement made between concepts which has its own unique identifier as well as metadata containing author and provenance information (Mons *et al*, 2011). The initiative to publish data as Linked Open Data could be imposed by funding agencies, research groups, and journals. Authors could also pressure journals to publish their data as well (Mons & Velterop, 2009), as using nanopublications may lead to an increase in citations to the authors work (Mons *et al*, 2007). For example, a research group that receives its funding from parents of children with orphaned diseases could be asked to publish their data as Linked Open Data post-publication in order to receive funding. This condition for funding would allow for the data to be usable by novel approaches (Mons & Velterop, 2009), which in line with the previous example, could possibly identify cures for this orphaned disease now that the information is publicly available. Some research funders have started to require that data be published openly in order to receive funding (Bourne *et al*, 2011), as such, a grant received by a research group promising to publish their data as LOD might be more likely to receive funding. This idea could also be imposed by government funding

agencies and journal publishers in order to increase the amount of data available for approaches seeking to gain new knowledge. Another issue faced by Linked Open Data is the quality of the information.

Previously, heuristics have been used to assess the quality of data, such as by comparing data content against the content of the rest of the data retrieved, by assessing the context of the data using the metadata, or by making use of ratings which may already be associated with the data (Heath & Bizer, 2011). The quality of data used as the input to generate predictions or perform evaluations by various approaches making use of Linked Open Data is important. The importance of this is due to the fact that bad data will lead to inaccurate predictions or conclusions. One way of doing this would be to make use of provenance to include from where data comes from, and to make it a requirement to include this information when publishing Linked Open Data. Other important information would be to create standards which require that metadata contain other relevant information. This information would include the date, any modifications made to the data, versioning information (Zhao *et al*, 2009), information of how analyses were derived, etc. An example use of provenance in Linked Data includes a study using RDF graphs to store provenance information of fly-related genomic data in order to ensure trust in this data, for example that the links to other databases are correct when entries in the current database are altered (Zhao *et al*, 2009). The use of data provenance would allow users of the data to decide whether or not they trust the source, and if the data produced in one dataset is comparable to that produced in another dataset, which would be useful for comparing assay results such as microarrays.

Future endeavors should seek to improve interoperability of datasets through the mitigation of heterogeneity between datasets, for example by improving mappings between ontologies. The use of semantic technologies and Linked Open Data can reduce interoperability issues, and aid in the discovery of new knowledge. In order to increase the amount of available data, quality data such as data found in various curated databases needs to be made available as Linked Open Data. As more Linked Data becomes available and the heterogeneity between datasets becomes better resolved, approaches such as PhenomeDrug and HyQue-Cardiotoxicity can be leveraged to their full potential to help identify new information from disparate datasets.

Appendices

To access appendices, permission is required via GoogleDocs.

Appendix A - Manual mapping of UMLS SIDER concepts to ontologies

<http://goo.gl/WLCoq3>

Appendix B - Comparison of manual vs automatic mappings

<http://goo.gl/A7WkBQ>

Appendix C - Evidence rules and associated datasets

<http://goo.gl/nSjxcD>

Appendix D - List of terms used within rule set related to cardiotoxicity

<http://goo.gl/3PM4fj>

Appendix E - List of cardiotoxic, non-TKI drugs, and score for HyQue evaluation

<http://goo.gl/g4HfoM>

Appendix F - List of non-cardiotoxic, non-TKI drugs, and score for HyQue evaluation

<http://goo.gl/iD8zeB>

Appendix G - List of cardiotoxic drugs created using DailyMed for method comparison

<http://goo.gl/qdtKIX>

Appendix H - List of non-cardiotoxic drugs created using DailyMed for method comparison

<http://goo.gl/Obfg7a>

Appendix I - Reasons for comparison of drugs not working

<http://goo.gl/4MY1QE>

Appendix J - Breakdown of scores for HyQue-Cardiotoxicity rules

<http://goo.gl/5zKdby>

References

- Acland, A., R. Agarwala, T. Barrett, J. Beck, DA Benson, C. Bollin, E. Bolton, et al. 2013; 2012. Database resources of the national center for biotechnology information. *Nucleic Acids Research* 41 (Database issue): D8-D20.
- Alistair, M, Brian Matthews, Michael Wilson, and Dan Brickley. 2005. **SKOS core: simple knowledge organisation for the web**. In *International Conference on Dublin Core and Metadata Applications* 3.
- Allemang, Dean, and James Hendler. 2011. **The RDF Schema Language**. In *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. 2nd ed., Elsevier.
- Alomar, Muaed Jamal. 2014. Factors affecting the development of adverse drug reactions (review article). *Saudi Pharmaceutical Journal : SPJ : The Official Publication of the Saudi Pharmaceutical Society* 22 (2): 83-94.
- Antezana, Erick, Ward Blondé, Mikel Egaña, Alistair Rutherford, Robert Stevens, Bernard De Baets, Vladimir Mironov, and Martin Kuiper. 2009. BioGateway: A semantic systems biology tool for the life sciences. *BMC Bioinformatics* 10 (Suppl 10): S11.
- Arrowsmith, J., and P. Miller. 2013. Trial watch: Phase II and phase III attrition rates 2011-2012. *Nature Reviews.Drug Discovery* 12 (8): 569.
- Bai, Jane P. F., and Darrell R. Abernethy. 2013. Systems pharmacology to predict drug toxicity: Integration across levels of biological organization. *Annual Review of Pharmacology and Toxicology* 53 : 451-73.
- Berardi, Rossana, Miriam Caramanti, Agnese Savini, Silvia Chiorrini, Chiara Pierantoni, Azzurra Onofri, Zelmira Ballatore, Mariagrazia De Lisa, Paola Mazzanti, and Stefano Cascinu. 2013. State of the art for cardiotoxicity due to chemotherapy and to targeted therapies: A literature review. *Critical Reviews in oncology/hematology* 88 (1): 75-86.
- Biesecker, Leslie G. 2013. Hypothesis-generating research and predictive medicine. *Genome Research* 23 (7): 1051-3.
- Bizer, Christian, Heath Tom, and Berners-Lee Tim. 2009. Linked Data - the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (3): 1-22.

- Bodenreider, O., T. F. Hayamizu, M. Ringwald, S. De Coronado, and S. Zhang. 2005. Of mice and men: Aligning mouse and human anatomies. *AMIA ...Annual Symposium Proceedings / AMIA Symposium*. *AMIA Symposium*: 61-5.
- Bodenreider, Olivier. 2004. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Research* 32 (Database issue): D267-70.
- Boeker, Martin, Ilinca Tudose, Janna Hastings, Daniel Schober, and Stefan Schulz. 2011. Unintended consequences of existential quantifications in biomedical ontologies. *BMC Bioinformatics* 12 (1): 456.
- Bourne PE, Clark TW, Dale R, de Waard A, Herman I, et al. 2012. Improving the future of research communication and e-scholarship. *Dagstuhl Manifestos* 1: 41-60.
- Brown, Steve D. M., Wolfgang Wurst, Ralf Kühn, and John M. Hancock. 2009. The functional annotation of mammalian genomes: The challenge of phenotyping. *Annual Review of Genetics* 43 (1): 305-33.
- Bult, C. J., J. T. Eppig, J. A. Blake, J. A. Kadin, J. E. Richardson, and Mouse Genome Database Group. 2013. The mouse genome database: Genotypes, phenotypes, and models of human disease. *Nucleic Acids Research* 41 (Database issue): D885-91.
- Burgun, A., F. Mougin, and O. Bodenreider. 2009. Two approaches to integrating phenotype and clinical information. *AMIA Annu Symp Proc. 2009*; 2009: 75–79.
- Callahan, A., J. Cruz-Toledo, P. Ansell, and M. Dumontier. 2013a. Bio2RDF release 2: Improved coverage, interoperability and provenance of life science Linked Data. *Eswc 2013*.: 200-12.
- Callahan, Alison, José Cruz-Toledo, and Michel Dumontier. 2013b. Ontology-based querying with Bio2RDF's linked open data. *Journal of Biomedical Semantics* 4 (Suppl 1):S1.
- Callahan, Alison, Michel Dumontier, and Nigam H. Shah. 2011. HyQue: Evaluating hypotheses using semantic web technologies. *Journal of Biomedical Semantics* 2 (Suppl 2): S3.
- Campillos, M., M. Kuhn, A. C. Gavin, L. J. Jensen, and P. Bork. 2008. Drug target identification using side-effect similarity. *Science* 321 (5886): 263-6.
- Chan, AT, JE Manson, CM Albert, CU Chae, KM Rexrode, GC Curhan, EB Rimm, WC Willett, and CS Fuchs. 2006. Nonsteroidal antiinflammatory drugs, acetaminophen, and the risk of cardiovascular events. *Circulation* 113 (12): 1578-87.
- Chen, Bin, Xiao Dong, Dazhi Jiao, Huijun Wang, Qian Zhu, Ying Ding, and David J. Wild. 2009; 2010. Chem2Bio2RDF: A semantic framework for linking and data

- mining chemogenomic and systems chemical biology data. *BMC Bioinformatics* 11 (1): 255.
- Chen, MH, R. Kerkelae, and T. Force. 2008. Mechanisms of cardiac dysfunction associated with tyrosine kinase inhibitor cancer therapeutics. *Circulation* 118 (1): 84-95.
- Cheng, H., G. Kari, AP Dicker, U. Rodeck, WJ Koch, and T. Force. 2011. A novel preclinical strategy for identifying cardiotoxic kinase inhibitors and mechanisms of cardiotoxicity. *Circulation Research* 109 (12): 1401-1409.
- Chmura Kraemer, Helena, Vyjeyanthi S. Periyakoil, and Art Noda. 2002. Kappa coefficients in medical research. *Statistics in Medicine* 21 (14): 2109-29.
- Choueiri, TK, FAB Schutz, YJ Je, JE Rosenberg, and J. Bellmunt. 2010. Risk of arterial thromboembolic events with sunitinib and sorafenib: A systematic review and meta-analysis of clinical trials. *Journal of Clinical Oncology* 28 (13): 2280-5.
- Chu, Tammy F., Maria A. Rupnick, Risto Kerkela, Susan M. Dallabrida, David Zurakowski, Lisa Nguyen, Kathleen Woulfe, et al. 2007. Cardiotoxicity associated with tyrosine kinase inhibitor sunitinib. *Lancet* 370 (9604): 2011-9.
- Cook, D., D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, and M. N. Pangalos. 2014. Lessons learned from the fate of AstraZeneca's drug pipeline: A five-dimensional framework. *Nature Reviews. Drug Discovery* 13 (6): 419-31.
- Cruz, I. F., F. P. Antonelli, and C. Stroe. 2009. AgreementMaker: Efficient matching for large real-world schemas and ontologies. *Proceedings of the VLDB Endowment* 2 (2): 1586-9.
- Curry, E., J. O'Donnell, E. Corry, S. Hasan, M. Keane, and S. O'Riain. 2013. Linking building data in the cloud: Integrating cross-domain building data using Linked Data. *Advanced Engineering Informatics* 27 (2): 206-19.
- Davis, Allan Peter, Cynthia Grondin Murphy, Robin Johnson, Jean M. Lay, Kelley Lennon-Hopkins, Cynthia Saraceni-Richards, Daniela Sciaky, et al. 2013. The comparative toxicogenomics database: Update 2013. *Nucleic Acids Research* 41 (Database issue): D1104-14.
- Dearden, John C. 2003. In silico prediction of drug toxicity. *Journal of Computer-Aided Molecular Design* 17 (2): 119-27.
- De Roure, D., C. Goble, and R. Stevens. 2009. The design and realisation of the (my)experiment virtual research environment for social sharing of workflows. *Future Generation Computer Systems-the International Journal of Grid Computing-Theory Methods and Applications* 25 (5): 561-7.

- Dumontier, Michel, Christopher Jo Baker, Joachim Baran, Alison Callahan, Leonid Chepelev, José Cruz-Toledo, Nicholas R. Del Rio, et al. 2014. The semanticscience integrated ontology (SIO) for biomedical research and knowledge discovery. *Journal of Biomedical Semantics* 5 (1): 14.
- Egan, William J., Gregor Zlokarnik, and Peter D. J. Grootenhuis. 2004. In silico prediction of drug safety: Despite progress there is abundant room for improvement. *Drug Discovery Today: Technologies* 1 (4): 381-7.
- Ekins, Sean. 2013; 2014. Progress in computational toxicology. *Journal of Pharmacological and Toxicological Methods* 69 (2): 115-40.
- Erkimbaev, A. O., V. Yu Zitserman, G. A. Kobzev, V. A. Serebrjakov, and K. B. Teymurazov. 2013. Publishing scientific data as linked open data. *Scientific and Technical Information Processing* 40 (4): 253-63.
- Euzenat, Jérôme, and Pavel Shvaiko. 2013. **Evaluation of matching systems.** In *Ontology matching*. 2ed., 285-317. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Euzenat, Jérôme, and Pavel Shvaiko. 2013. **The Matching Problem.** In *Ontology matching*. 2ed., 25-54. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Fermini, Bernard, and Anthony A. Fossa. 2003. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nature Reviews. Drug Discovery* 2 (6): 439-47.
- Ferri, Nicola, Peter Siegl, Alberto Corsini, Joerg Herrmann, Amir Lerman, and Renee Benghozi. 2013. Drug attrition during pre-clinical and clinical development: Understanding and managing drug-induced cardiotoxicity. *Pharmacology & Therapeutics* 138 (3): 470-84.
- Force, Thomas, Daniela S. Krause, and Richard A. Van Etten. 2006; 2007. Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nature Reviews Cancer* 7 (5): 332-44.
- Frid, Anna A., and Edwin J. Matthews. 2010. Prediction of drug-related cardiac adverse effects in humans-B: Use of QSAR programs for early detection of drug-induced cardiac toxicities. *Regulatory Toxicology and Pharmacology : RTP* 56 (3): 276-89.
- Gaulton, Anna, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, et al. 2012. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40 (D1): D1100-7.
- Gharib, M. I., and A. K. Burnett. 2002. Chemotherapy-induced cardiotoxicity: Current practice and prospects of prophylaxis. *European Journal of Heart Failure* 4 (3): 235-42.

- Ghazvinian, A., N. F. Noy, and M. A. Musen. 2009. Creating mappings for ontologies in biomedicine: Simple methods work. *AMIA ...Annual Symposium Proceedings / AMIA Symposium*. *AMIA Symposium* 2009: 198-202.
- Gkoutos, G. V., P. N. Schofield, and R. Hoehndorf. 2012. The neurobehavior ontology: An ontology for annotation and integration of behavior and behavioral phenotypes. *International Review of Neurobiology* 103: 69-87.
- Glass, Kimberly, Edward Ott, Wolfgang Losert, and Michelle Girvan. 2012. Implications of functional similarity for gene regulatory interactions. *Journal of the Royal Society, Interface / the Royal Society* 9 (72): 1625-36.
- Gross, Anika, Michael Hartung, Kirsten Toralf, and Erhard Rahm. 2011. Mapping composition for matching large life science ontologies. Boston, NY, http://icbo.buffalo.edu/ICBO-2011_Proceedings.pdf.
- Guengerich, F. Peter. 2011. Mechanisms of drug toxicity and relevance to pharmaceutical development. *Drug Metabolism and Pharmacokinetics* 26 (1): 3-14.
- Guyon, I., and A. Elisseeff. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* (3): 1157-82.
- Hardy, B., N. Douglas, C. Helma, M. Rautenberg, N. Jeliaskova, V. Jeliaskov, I. Nikolova, et al. 2010. Collaborative development of predictive toxicology applications. *Journal of Cheminformatics* 2:7.
- Hayden, Wimmer, Yoon Victoria, and Rada Roy. 2012. Applying semantic web technologies to ontology alignment. *International Journal of Intelligent Information Technologies (IJIT)* 8 (1): 1-9.
- Heath, Tom, Christian Bizer. 2011. *Linked Data: Evolving the web into a global data space*. Vol. 1. San Rafael, Calif. (1537 Fourth Street, San Rafael, CA 94901 USA): Morgan & Claypool Publishers.
- Heijman, Jordi, Niels Voigt, Leif G. Carlsson, and Dobromir Dobrev. 2014. Cardiac safety assays. *Current Opinion in Pharmacology* 15: 16-21.
- Hileman, Bette. 2005. Vioxx. *Chemical & Engineering News [H.W.Wilson - AST]* 83 (25): 134.
- Hoehndorf, R., Hiebert, N. W. Hardy, P. N. Schofield, G. V. Gkoutos, and M. Dumontier. 2014. Mouse model phenotypes provide information about human drug targets. *Bioinformatics* 30 (5): 719-25.

- Hoehndorf, R., F. Loebe, J. Kelso, and H. Herre. 2007. Representing default knowledge in biomedical ontologies: Application to the integration of anatomy and phenotype ontologies. *BMC Bioinformatics* 8: 377.
- Hoehndorf, R., P. N. Schofield, and G. V. Gkoutos. 2011. PhenomeNET: A whole-phenome approach to disease gene discovery. *Nucleic Acids Research* 39 (18): e119.
- Holland, John. 2013. Fixing a broken drug development process. *Journal of Commercial Biotechnology* 19 (1): 5.
- Hopkins, Andrew L. 2009. Drug discovery predicting promiscuity. *Nature* 462 (7270): 167-8.
- Hornberg, Jorrit J., Morten Laursen, Nina Brenden, Mikael Persson, Annemette V. Thougard, Dorthe B. Toft, and Tomas Mow. 2013. Exploratory toxicology as an integrated part of drug discovery—Part I: Why and how. *Drug Discovery Today*.
- Hua, Jianping, Waibhav D. Tembe, and Edward R. Dougherty. 2009. Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition* 42 (3): 409-24.
- Ichikawa, Yoshihiko, Mohsen Ghanefar, Marina Bayeva, Rongxue Wu, Arineh Khechaduri, Sathyamangla V. Naga Prasad, R. Kannan Mutharasan, Tejaswitha Jairaj Naik, and Hossein Ardehali. 2014. Cardiotoxicity of doxorubicin is mediated through mitochondrial iron accumulation. *The Journal of Clinical Investigation* 124 (2): 617-30.
- Jain, P., P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh. 2013. **Ontology alignment for linked open data**. In *The semantic web – ISWC 2010.*, eds. P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, Z. F. Pan, I. Horrocks and B. Glimm, 402-417, Springer Berlin Heidelberg.
- Jean-Mary, Yves R., E. Patrick Shironoshita, and Mansur R. Kabuka. 2009. Ontology matching with semantic verification. *Web Semantics: Science, Services and Agents on the World Wide Web* 7 (3): 235-51.
- Jolliffe, Ian T. 2002. **Introduction**. In *Principal Component Analysis*. 1-9.
- Jung, Jason J. 2010. Reusing ontology mappings for query routing in semantic peer-to-peer environment. *Information Sciences* 180 (17): 3248-57.
- Jupp, Simon, James Malone, Jerven Bolleman, Marco Brandizi, Mark Davies, Leyla Garcia, Anna Gaulton, et al. 2014. The EBI RDF platform: Linked open data for the life sciences. *Bioinformatics* 30 (9): 1338-9.

- Kerkelä, Risto, Luanda Grazette, Rinat Yacobi, Cezar Iliescu, Richard Patten, Cara Beahm, Brian Walters, et al. 2006. Cardiotoxicity of the cancer therapeutic agent imatinib mesylate. *Nature Medicine* 12 (8): 908-16.
- Kerkela, Risto, Kathleen C. Woulfe, Jean-Bernard Durand, Ronald Vagnozzi, David Kramer, Tammy F. Chu, Cara Beahm, Ming Hui Chen, and Thomas Force. 2008; 2009. Sunitinib-induced cardiotoxicity is mediated by off-target inhibition of AMP-activated protein kinase. *Clinical and Translational Science* 2 (1): 15-25.
- Khanna, Ish. 2012. Drug discovery in pharmaceutical industry: Productivity challenges and trends. *Drug Discovery Today* 17 (19-20): 1088-102.
- Knoop, S. J., and D. E. Worden. 1988. The pharmaceutical drug development process: An overview. *Drug Information Journal* 22 (2): 259-68.
- Köhler, Sebastian, Sebastian Bauer, Chris J. Mungall, Gabriele Carletti, Cynthia L. Smith, Paul Schofield, Georgios V. Gkoutos, and Peter N. Robinson. 2011. Improving ontologies by automatic reasoning and evaluation of logical definitions. *BMC Bioinformatics* 12 (1): 418.
- Köhler, Sebastian, Sandra C. Doelken, Barbara J. Ruef, Sebastian Bauer, Nicole Washington, Monte Westerfield, George Gkoutos, et al. 2013. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. *F1000research*. 2.
- Kolaja, Kyle L., and Thomas Force. 2011. Cardiotoxicity of kinase inhibitors: The prediction and translation of preclinical models to clinical outcomes. *Nature Reviews Drug Discovery* 10 (2): 111-26.
- Korennykh, A. V., P. F. Egea, A. A. Korostelev, J. Finer-Moore, C. Zhang, K. M. Shokat, R. M. Stroud, and P. Walter. 2009. The unfolded protein response signals through high-order assembly of Ire1. *Nature* 457 (7230): 687-93.
- Kuhn, M., M. Campillos, I. Letunic, L. J. Jensen, and P. Bork. 2010. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology* 6: 343.
- Kuhn, Michael, Damian Szklarczyk, Andrea Franceschini, Christian von Mering, Lars Juhl Jensen, and Peer Bork. 2012. STITCH 3: Zooming in on protein-chemical interactions. *Nucleic Acids Research* 40 (D1): D876-80.
- Lamb, Justin, Emily D. Crawford, David Peck, Joshua W. Modell, Irene C. Blat, Matthew J. Wrobel, Jim Lerner, et al. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313 (5795): 1929-35.

- Lapenna, Silvia, and Andrew Worth. 2011. *Analysis of the cramer classification scheme for oral systemic toxicity - implications for its implementation in toxtree*. Luxembourg: Publications Office of the European Union, EUR 24898 EN.
- Law, V., C. Knox, Y. Djoumbou, T. Jewison, AC Guo, YF Liu, A. Maciejewski, et al. 2013; 2014. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research* 42 (D1): D1091-7.
- Levitcki, Alexander. 2013. Tyrosine kinase inhibitors: Views of selectivity, sensitivity, and clinical performance. *Annual Review of Pharmacology and Toxicology* 53: 161-85.
- Luciano, Joanne S., Bosse Andersson, Colin Batchelor, Olivier Bodenreider, Tim Clark, Christine K. Denney, Christopher Domarew, et al. 2011. The translational medicine ontology and knowledge base: Driving personalized medicine by bridging the gap between bench and bedside. *Journal of Biomedical Semantics*. 2(S2): S1.
- Madden, T. 2002. The BLAST Sequence Analysis Tool. In *The NCBI Handbook [Internet]*. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. <http://www.ncbi.nlm.nih.gov/books/NBK21097/>
- Mandenius, Carl-Fredrik, Daniella Steel, Fozia Noor, Thomas Meyer, Elmar Heinzle, Julia Asp, Sarina Arain, et al. 2011. Cardiotoxicity testing using pluripotent stem cell-derived human cardiomyocytes and state-of-the-art bioanalytics: A review. *Journal of Applied Toxicology* 31 (3): 191-205.
- Masseroli, Marco, Barend Mons, Erik Bongcam-Rudloff, Stefano Ceri, Alexander Kel, François Rechenmann, Frederique Lisacek, and Paolo Romano. 2014. Integrated bio-search: Challenges and trends for the integration, search and comprehensive processing of biological information. *BMC Bioinformatics* 15 (Suppl 1): S2.
- McHugh, Mary L. 2012. Interrater reliability: The kappa statistic. *Biochemia Medica* 22 (3): 276-82.
- McNaughton, Rhian, Gwenaël Huet, and Saad Shakir. 2014. An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. *BMJ Open* 4 (1): e004221.
- Meilicke, Christian, Heiner Stuckenschmidt, and Andrei Taminlin. 2008; 2009. Reasoning support for mapping revision. *Journal of Logic and Computation* 19 (5): 807-29.
- Mellor, Howard R., Alex R. Bell, Jean-Pierre Valentin, and Ruth R. A. Roberts. 2011. Cardiotoxicity associated with targeting kinase pathways in cancer. *Toxicological Sciences: An Official Journal of the Society of Toxicology* 120 (1): 14-32.

- Mons, Barend, and Jan Velterop. 2009. Nano-publication in the e-science era. *8th International Semantic Web Conference (ISWC 2009)*.
- Mons, Barend, Herman van Haagen, Christine Chichester, Peter-Bram 't Hoen, Johan T. den Dunnen, Gertjan van Ommen, Erik van Mulligen, et al. 2011. The value of data. *Nature Genetics* 43 (4): 281-3.
- Mungall, CJ, GV Gkoutos, CL Smith, MA Haendel, SE Lewis, and M. Ashburner. 2010. Integrating phenotype ontologies across multiple species. *Genome Biology* 11 (1): R2.
- Mungall, CJ, C. Torniai, GV Gkoutos, SE Lewis, and MA Haendel. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome Biology* 13 (1): R5.
- Muster, Wolfgang, Alexander Breidenbach, Holger Fischer, Stephan Kirchner, Lutz Müller, and Axel Pähler. 2008. Computational toxicology in drug development. *Drug Discovery Today* 13 (7): 303-10.
- Noy, Natalya F. 2009. **Mapping ontologies**. In *Handbook on ontologies*., ed. Rudi Studer Steffen Staab (eds.). 2nd ed., 573-590. Berlin; Springer, c2009.
- Noy, Natalya F., Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, et al. 2009. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* 37 (Suppl 2): W170-3.
- Orphanos, GS, GN Ioannidis, and AG Ardavanis. 2009. Cardiotoxicity induced by tyrosine kinase inhibitors. *Acta Oncologica* 48 (7): 964-70.
- Pasquier, C. 2008. Biological data integration using semantic web technologies. *Biochimie* 90 (4): 584-94.
- Patel, Anand C. 2013. Clinical relevance of target identity and biology: Implications for drug discovery and development. *Journal of Biomolecular Screening* 18 (10): 1164-85.
- Patlewicz, G., N. Jeliaskova, R. J. Safford, A. P. Worth, and B. Aleksiev. 2008. An evaluation of the implementation of the cramer classification scheme in the toxtree software. *SAR and QSAR in Environmental Research* 19 (5-6): 495-524.
- Paul, Steven M., Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, and Aaron L. Schacht. 2010. How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews. Drug Discovery* 9 (3): 203-14.

- Perez, Edith A., Maria Koehler, Julie Byrne, Alaknanda J. Preston, Erica Rappold, and Michael S. Ewer. 2008. Cardiac safety of lapatinib: Pooled analysis of 3689 patients enrolled in clinical trials. *Mayo Clinic Proceedings* 83 (6): 679-86.
- Perlman, Liat, Assaf Gottlieb, Nir Atias, Eytan Ruppin, and Roded Sharan. 2011. Combining drug and gene similarity measures for drug-target elucidation. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 18 (2): 133-45.
- Pesquita, C., D. Faria, H. Bastos, A. E. Ferreira, A. O. Falcao, and F. M. Couto. 2008. Metrics for GO based protein semantic similarity: A systematic evaluation. *BMC Bioinformatics* 9 (Suppl 5): S4.
- Pesquita, Catia, Daniel Faria, André O. Falcão, Phillip Lord, and Francisco M. Couto. 2009. Semantic similarity in biomedical ontologies. *PLoS Computational Biology* 5 (7): e1000443.
- Platt, Richard, Leanne Madre, Robert Reynolds, and Hugh Tilson. 2008. Active drug safety surveillance: A tool to improve public health. *Pharmacoepidemiology and Drug Safety* 17 (12): 1175-82.
- Pollard, CE, N. Abi Gerges, MH Bridgland-Taylor, A. Easter, TG Hammond, and J-P Valentin. 2010. An introduction to QT interval prolongation and non-clinical approaches to assessing and reducing risk. *British Journal of Pharmacology* 159 (1): 12-21.
- Rampe, David, and Arthur M. Brown. 2013. A history of the role of the hERG channel in cardiac risk assessment. *Journal of Pharmacological and Toxicological Methods* 68 (1): 13-22.
- Raschi, Emanuel, Valentina Vasina, Maria Grazia Ursino, Giuseppe Boriani, Andrea Martoni, and Fabrizio De Ponti. 2010. Anticancer drugs and cardiotoxicity: Insights and perspectives in the era of targeted therapy. *Pharmacology and Therapeutics* 125 (2): 196-218.
- Ratti, E., and D. Trist. 2001. The continuing evolution of the drug discovery process in the pharmaceutical industry. *Farmaco (Societa Chimica Italiana:1989)* 56 (1-2): 13-9.
- Roberts, R. A., S. L. Kavanagh, H. R. Mellor, C. E. Pollard, S. Robinson, and S. J. Platz. 2014. Reducing attrition in drug development: Smart loading preclinical safety assessment. *Drug Discovery Today* 19 (3): 341-7.
- Robinson, P. N., S. Kohler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. 2008. The human phenotype ontology: A tool for annotating and analyzing human hereditary disease. *American Journal of Human Genetics* 83 (5): 610-5.

- Roithinger, F. X., C. Punzengruber, F. Gremmel, M. Hinterreiter, F. Holzner, and O. Pachinger. 1993. Myocardial infarction after chronic ergotamine abuse. *European Heart Journal* 14 (11): 1579-81.
- Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23 (19): 2507-17.
- Sams-Dodd, Frank. 2005. *Target-based drug discovery: Is something wrong?*. Vol. 10. England: ELSEVIER SCI LTD.
- Samwald, Matthias, Anja Jentzsch, Christopher Bouton, Claus Stie Kallesøe, Egon Willighagen, Janos Hajagos, M. Scott Marshall, et al. 2011; 2010. Linked open drug data for pharmaceutical research and development. *Journal of Cheminformatics* 3 (1): 1-6.
- Sanguinetti & Tristani-Firouzi, 2006. hERG potassium channels and cardiac arrhythmia. *Nature* 440: 463-469.
- Sansone, Susanna-Assunta, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, et al. 2012. Toward interoperable bioscience data. *Nature Genetics* 44 (2): 121-6.
- Sataloff, Robert T. 2008. Drug development research: The process. *Ear, Nose, & Throat Journal* 87 (8): 420.
- Schlitt, Axel, Karin Jordan, Dirk Vordermark, Jürgen Schwamborn, Thorsten Langer, and Christoph Thomssen. 2014. Cardiotoxicity and oncological treatments. *Deutsches Ärzteblatt International* 111 (10): 161.
- Schofield, P. N., M. Gruenberger, and J. P. Sundberg. 2010. Pathbase and the MPATH ontology community resources for mouse histopathology. *Veterinary Pathology* 47 (6): 1016-20.
- Schriml, L. M., C. Arze, S. Nadendla, Y. W. Chang, M. Mazaitis, V. Felix, G. Feng, and W. A. Kibbe. 2012. Disease ontology: A backbone for disease semantic integration. *Nucleic Acids Research* 40 (Database issue): D940-6.
- Sebaugh, J. L. 2011. Guidelines for accurate EC50/IC50 estimation. *Pharmaceutical Statistics* 10 (2): 128-34.
- Segall, MD, and C. Barber. 2014. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discovery Today* 19 (5): 688-93.

- Shadbolt, Nigel, and Kieron O'Hara. 2013. Linked Data in government. *IEEE Internet Computing* 17 (4): 72-7.
- Shi, Y., M. Moon, S. Dawood, B. McManus, and P. P. Liu. 2011. Mechanisms and management of doxorubicin cardiotoxicity. *Herz* 36 (4): 296-305.
- Shvaiko, P., and J. Euzenat. 2013. Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering* 25 (1): 158-76.
- Silva, Thushari, Vilas Wuwongse, and Hitesh Nidhi Sharma. 2013; 2012. Disaster mitigation and preparedness using linked open data. *Journal of Ambient Intelligence and Humanized Computing* 4 (5): 591-602.
- Simon-Hettich, Brigitte, Andreas Rothfuss, and Thomas Steger-Hartmann. 2006. Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicology* 224 (1-2): 156-62.
- Smith, C. L., C. A. Goldsmith, and J. T. Eppig. 2005. The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology* 6 (1): R7.
- Spector, Neil L., Yosef Yarden, Bradley Smith, Ljuba Lyass, Patricia Trusk, Karen Pry, Jason E. Hill, Wenle Xia, Rony Seger, and Sarah S. Bacus. 2007. Activation of AMP-activated protein kinase by human EGF receptor 2/EGF receptor tyrosine kinase inhibitor protects cardiac cells. *Proceedings of the National Academy of Sciences of the United States of America* 104 (25): 10607-12.
- Tatonetti, N. P., P. P. Ye, R. Daneshjou, and R. B. Altman. 2012. Data-driven prediction of drug effects and interactions. *Science Translational Medicine* 4 (125):125ra31.
- Terstappen, G. C., and A. Reggiani. 2001. In silico research in drug discovery. *Trends in Pharmacological Sciences* 22 (1): 23-6.
- Tonkens, Ross. 2005. An overview of the drug development process. *Physician Executive* 31 (3): 48.
- Valerio, Jr, Luis G. 2011. In silico toxicology models and databases as FDA critical path initiative toolkits. *Human Genomics* 5 (3): 200-7.
- Wang, P., and B. Xu. 2008. Debugging ontology mappings: A static approach. *Computing and Informatics* 27 (1): 21-36.
- Whirl-Carrillo, M., EM McDonagh, JM Hebert, L. Gong, K. Sangkuhl, CF Thorn, RB Altman, and TE Klein. 2012. Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics* 92 (4): 414-7.

- Williams, Antony J., Lee Harland, Paul Groth, Stephen Pettifer, Christine Chichester, Egon L. Willighagen, Chris T. Evelo, et al. 2012. Open PHACTS: Semantic interoperability for drug discovery. *Drug Discovery Today* 17 (21-22): 1188-98.
- Winnenburg, R., and O. Bodenreider. 2014. Coverage of phenotypes in standard terminologies. *Proceedings of the ISMB 2014 SIG Meeting "BioLINK"*(in press).
- Wiśniowska, Barbara, Aleksander Mendyk, Kamil Fijorek, and Sebastian Polak. 2014. Computer-based prediction of the drug proarrhythmic effect: Problems, issues, known and suspected challenges. *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology : Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 16 (5): 724-35.
- Zambrowicz, Brian P., and Arthur T. Sands. 2003. Knockouts model the 100 best-selling drugs-will they model the next 100? *Nature Reviews Drug Discovery* 2 (1): 38-51.
- Zhao, Jun, Alistair Miles, Graham Klyne, and David Shotton. 2009. Linked Data and provenance in biological data webs. *Briefings in Bioinformatics* 10 (2): 139-52.
- Zhang, Z., V. B. Bajic, J. Yu, K-H Cheung, and J. P. Townsend. 2011. **Data integration in bioinformatics: Current efforts and challenges.** In *Bioinformatics - trends and methodologies.*, ed. M. A. Mahdavi, 41-56. Rijeka, Croatia: InTech - Open Access Publisher.
- Zhu, Feng, BuCong Han, Pankaj Kumar, XiangHui Liu, XiaoHua Ma, Xiaona Wei, Lu Huang, et al. 2009; 2010. Update of TTD: Therapeutic target database. *Nucleic Acids Research* 38 (Database issue): D787-91.
- Zhu, Qian, Michael S. Lajiness, Ying Ding, and David J. Wild. 2010. WENDI: A tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly publications. *Journal of Cheminformatics* 2 (1): 1-9.
- Zupan, Blaz, Ivan Bratko, Janez Demsar, Peter Juhan, Tomaz Curk, Urban Borstnik, J. Robert Beck, John Halter, Adam Kuspa, and Gad Shaulsky. 2003. GenePath: A system for inference of genetic networks and proposal of genetic experiments. *Artificial Intelligence in Medicine* 29 (1-2): 107-30.