

**EXPLORING THE PHARMACOGENOMICS KNOWLEDGE BASE (PHARMGKB) FOR  
REPOSITIONING BREAST CANCER DRUGS BY LEVERAGING WEB ONTOLOGY LANGUAGE (OWL)  
AND CHEMINFORMATICS APPROACHES\***

QIAN ZHU

*Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA  
Email: zhu.qian@mayo.edu*

CUI TAO

*School of Biomedical Informatics, University of Texas Health Science Center at Houston, TX 77030, USA  
Email: cui.tao@uth.tmc.edu*

FEICHEN SHEN

*School of Computing and Engineering, University of Missouri-Kansas City, Kansas City, MO 64110, USA  
Email: fsm89@mail.umkc.edu*

CHRISTOPHER G. CHUTE

*Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA  
Email: chute@mayo.edu*

Computational drug repositioning leverages computational technology and high volume of biomedical data to identify new indications for existing drugs. Since it does not require costly experiments that have a high risk of failure, it has attracted increasing interest from diverse fields such as biomedical, pharmaceutical, and informatics areas. In this study, we used pharmacogenomics data generated from pharmacogenomics studies, applied informatics and Semantic Web technologies to address the drug repositioning problem. Specifically, we explored PharmGKB to identify pharmacogenomics related associations as pharmacogenomics profiles for US Food and Drug Administration (FDA) approved breast cancer drugs. We then converted and represented these profiles in Semantic Web notations, which support automated semantic inference. We successfully evaluated the performance and efficacy of the breast cancer drug pharmacogenomics profiles by case studies. Our results demonstrate that combination of pharmacogenomics data and Semantic Web technology/Cheminformatics approaches yields better performance of new indication and possible adverse effects prediction for breast cancer drugs.

---

\* This work was supported by the Pharmacogenomic Research Network (NIH/NIGMS-U19 GM61388) and the Cancer Prevention & Research Institute of Texas (CPRIT R1307).

## 1. Introduction

Traditional drug development is costly and labor-intensive, and scientists are devoted to finding an alternative way to facilitate the drug discovery process. Drug repositioning, finding new therapeutic uses for existing drugs, is one of the most efficient and efficacious approaches to speed drug discovery. With the advance in computational technology, computational drug repositioning has shown its advantage as many studies been published recently. Ye et al. [1] explored a disease-oriented strategy for evaluating the relationship between drugs and disease on the basis of their pathway profile; Napolitano et al. [2] investigated machine-learning algorithms to predict drug repositioning; Li and Lu[3] presented an approach for identifying potential new indications of an existing drug through its relation to similar drugs. Butte's lab has reported their efforts on computational drug repurposing by exploring gene expression data [4, 5]. These studies drew on different technologies to address the problem of computational drug repositioning. However, none of them attempted to leverage data from emerging pharmacogenomics (PGx) studies in an integrated and transformable manner and explore Semantic Web technology as core implementation tool to address drug repositioning, which is our proposed aim for this study. PGx study investigates how genetic variations affect drug responses for the individual patient, consequently high volume of PGx information including relations among drugs, genes, single nucleotide polymorphisms (SNPs), etc. has been accumulated. The overarching goal of this study was to provide PGx profiles for FDA approved breast cancer drugs (BCDs) by leveraging informatics approaches and Semantic Web technologies, and ultimately to facilitate oncology-relevant biomedical and clinical studies and to support breast cancer drug repositioning.

Currently in the PGx world, different formats are being used for different data resources, which is the main obstacle to integration of PGx data to support development of relevant applications. Different formats might be preferred to represent scientific data, based on the nature of the source, the way the data are to be queried or visualized, or the type of analyses to be performed. Traditionally, investigators have relied heavily on tools such as Excel spreadsheets and relational databases to store and represent their research findings. However, these tools lack interoperability and capability to make inferences. In contrast, Semantic Web technology can manage scientific data in a more integrative and intelligent way. It is "a rigorous mechanism for defining and linking data using Web protocols in such a way that the data can be used by machines not just for display, but also for automation, integration, and reuse across various applications"[6]. Web Ontology Language (OWL), as a Semantic Web standard, can formally represent domain knowledge; it "organizes concepts or entities within classification (specialization or "is-a") hierarchies that provide for inheritance of attributes"[7]. Reusing existing resources in an integrative manner is essential, but exploring new associations is much more challenging. A Semantic Web reasoner enables identification of new BCD PGx associations, with an ultimate goal of repositioning BCDs. Dumontier [10] has demonstrated some advantages by expressing PGX data, PharmGKB in OWL for personalized medicine purpose.

Additionally, novel PGx information may be detected from a chemical perspective. Drugs with chemical structure similar to that of cancer drugs or genes associated with drugs with similar chemical structure can be identified using cheminformatics approaches[8]. Cheminformatics, a suite of computational technologies to solve a range of chemical problems, can be used to identify and evaluate new PGx associations. More precisely, we implemented a similar-structure searching algorithm to identify drugs similar to BCDs and find potential new uses for these drugs.

The paper is organized into the following sections. First, we introduce materials being used in this study; second, in the Methods section, we introduce details about PGx OWL profiles generation for BCDs and case study; third, we illustrate our results generated from each step in the Results section, which is followed by Discussion and Conclusion.

## 2. Materials

### 2.1. PharmGKB

The PharmGKB contains genomic, phenotype and clinical information collected from PGx studies. PharmGKB provides information regarding variant annotations, drug-centered pathway, pharmacogene summaries, clinical annotations, PGx-based drug-dosing guidelines, and drug labels with PGx information[9].

In this study, we used PGx information extracted from a relationship file received from PharmGKB by May 8, 2013, to generate the PGx profile for FDA-approved BCDs. Figure 1 shows some concrete PGx related association examples from the PharmGKB relationship file. Particularly, we extracted “Entity id”, “Entity name”, and “Entity type” for this study. Other fields, such as PubMed IDs (PMIDs), will be explored and integrated in a future study to support selection of the best PGx associations with publications as evidence.

A	B	C	D	E	F	G	H	I	J	K
Entity1_id	Entity1_name	Entity1_type	Entity2_id	Entity2_name	Entity2_type	Evidence	Association	PK	PD	PMIDs
PA150481189	taxanes	Drug	PA162387925	FAM82A1	Gene	ClinicalAnnotation	associated		PD	17224914;
PA150481189	taxanes	Drug	PA267	ABCB1	Gene	ClinicalAnnotation,VariantAnnotation	associated	PK	PD	12684679;
PA150481189	taxanes	Drug	PA27094	CYP1B1	Gene	ClinicalAnnotation	associated		PD	17224914;
PA150481189	taxanes	Drug	PA29028	GSTP1	Gene	VariantAnnotation	associated		PD	19203783
PA150481189	taxanes	Drug	rs1045642	rs1045642	VariantLocation	VIP,VariantAnnotation	associated		PD	12684679
PA150481189	taxanes	Drug	rs1056836	rs1056836	VariantLocation	ClinicalAnnotation	associated		PD	17224914;
PA150481189	taxanes	Drug	rs1695	rs1695	VariantLocation	VariantAnnotation	associated		PD	19203783
PA150481189	taxanes	Drug	rs2032582	rs2032582	VariantLocation	ClinicalAnnotation,VariantAnnotation	associated	PK	PD	16467099;
PA449412	doxorubicin	Drug	CYP2C8 *1A, CYP2C8 *3	CYP2C8 *1A, CYP2C8 *3	Haplotype	VariantAnnotation	not associated		PD	22527101
PA449412	doxorubicin	Drug	PA116	ABCC2	Gene	ClinicalAnnotation,Pathway,VariantAnnotation	associated	PK	PD	16330681,2
PA449412	doxorubicin	Drug	PA123	CYP2B6	Gene	ClinicalAnnotation,VariantAnnotation	associated		PD	20179710
PA449412	doxorubicin	Drug	PA124	CYP2C19	Gene	ClinicalAnnotation,VariantAnnotation	associated		PD	20179710
PA449412	doxorubicin	Drug	PA125	CYP2C8	Gene	VariantAnnotation	not associated		PD	22527101
PA449412	doxorubicin	Drug	PA130	CYP3A4	Gene	VariantAnnotation	associated		PD	20459744
PA449412	doxorubicin	Drug	PA134911502	SLC22A16	Gene	ClinicalAnnotation,Pathway,VariantAnnotation	associated	PK	PD	17559346,2
PA449412	doxorubicin	Drug	PA134956204	C18orf56	Gene	VariantAnnotation	not associated		PD	19159907
PA449412	doxorubicin	Drug	PA142671588	KLC3	Gene	VariantAnnotation	associated		PD	21826087
PA449412	doxorubicin	Drug	PA164724093	NO52	Gene	Pathway	associated		PD	21048526

Fig. 1. Examples of PGx relations available in PharmGKB

In addition to the PGx information from the PharmGKB relationship file shown in Figure 1, PharmGKB also provides pathway information, which includes associations between pathway and

drug, pathway and gene, and pathway and disease. Overall ten associations among drugs, genes, diseases, pathways, SNPs are available from PharmGKB. Table 1 shows these associations from two PharmGKB data files. Haplotype related associations are beyond the scope of this study.

Table 1. PGx related associations available from PharmGKB

Associations Resources	Drug- Drug	Drug- Gene	Drug- Pathway	Drug- SNP	Gene- Pathway	Gene- Disease	Disease- Pathway	Disease- SNP	Gene- Disease	Gene- Gene
PharmGKB Relationship file	√	√		√		√		√	√	√
PharmGKB Pathway data			√		√		√			

## 2.2. FDA approved BCDs

The National Cancer Institute (NCI) maintains cancer drugs approved by the FDA for breast cancer[11]. In this study, we did not consider drug combinations that are not approved by the FDA, even though the individual drugs are approved. Of 23 BCDs from NCI, a total of 18 BCDs have been manually mapped to the PharmGKB relationship file. The PGx profiles have been generated for these 18 BCDs, as described in the following sections. Table 2 shows the 23 BCDs from NCI vs 18 BCDs mapped to PharmGKB.

Table 2. BCDs from NCI and PharmGKB<sup>a</sup>

<i>BCDs available from NCI</i>	<i>BCDs identified in PharmGKB relationship file</i>
<b>ado-trastuzumab emtansine</b>	anastrozole
anastrozole	capecitabine
capecitabine	cyclophosphamide
cyclophosphamide	docetaxel
docetaxel	doxorubicin
doxorubicin hydrochloride	epirubicin
epirubicin hydrochloride	everolimus
everolimus	exemestane
exemestane	fluorouracil
fluorouracil	fulvestrant
fulvestrant	gemcitabine
gemcitabine hydrochloride	lapatinib
<b>ixabepilone</b>	letrozole
lapatinib ditosylate	methotrexate
letrozole	paclitaxel
<b>megestrol acetate</b>	pertuzumab
methotrexate	tamoxifen
paclitaxel	trastuzumab
<b>paclitaxel albumin-stabilized nanoparticle formulation</b>	
pertuzumab	
tamoxifen citrate	
trastuzumab	
<b>toremifene</b>	

<sup>a</sup>Drugs that failed to map to PharmGKB are shown in bold.

### 2.3. Semantic Web Technologies

Emerging Semantic Web technologies provide a formal mechanism to represent domain knowledge and data and to perform semantic reasoning on top of this knowledge. Semantic Web technology supports flexible, extensible, and evolvable knowledge transfer and reuse. It has been widely used in biomedical domains to formalize and model medical and biological systems. The Resource Description Framework (RDF)[12] is a World Wide Web Consortium (W3C) standard that specifies a graph-based data model for representing Semantic Web data. Each piece of information is represented in three parts (a triple): subject, predicate, and object. The RDF representations allow efficient querying and visualization of relationships between important biomedical entities. OWL [13] is a standard ontology language for the Semantic Web. A distinguishing characteristic of RDF and ontologies compared with the conventional relational database is “their degree of connectedness, their ability to model coherent, linked relationships”[14]. Representing the associations using OWL will enable powerful data integration among heterogeneous data sets, which is a well-known challenge in the translational science study community.

### 3. Methods

In this study, we focused on FDA approved BCDs and generated PGx OWL profiles by leveraging PharmGKB data and semantic web technologies. The OWL profiles explicitly capture BCD concepts and relationships and enable the semantic inference for novel drug associations. The overall architecture of the proposed project is shown in Figure 2. The details about each step are described in the following sections.

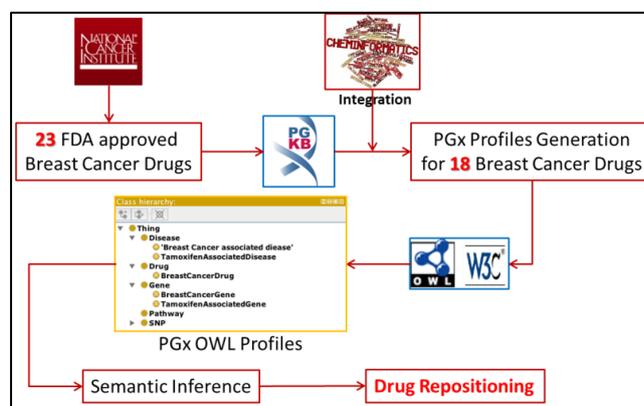


Fig. 2. Building blocks for the overall architecture

### 3.1. Generation of Integrative Breast Cancer PGx Profiles

#### 3.1.1. BCD PGx related association extraction

The PGx related associations shown in Table 1 were explored in this study for generation of PGx profiles. We programmatically extracted the PGx related associations from the relationship file that is tab delimited. In addition, we manually identified associations among pathways, drugs, genes and diseases for 18 BCDs from the PharmGKB pathway file that is a plain text file. Additional associations were inferred by invoking a rule-based OWL reasoner described in section 3.2.

### 3.1.2. *Chemical structure based similarity calculation*

To identify inferred associations for BCDs from a chemical perspective, two steps were involved: retrieval of chemical representations (by the simplified molecular-input line-entry system [SMILES] [15] or the IUPAC International Chemical Identifier InChI [16]) and structural similarity calculation. Except for the drugs with SMILES annotated by PharmGKB, we first converted active ingredient names to chemical representations through publically accessible services, such as the PubChem Entrez web service [17] and the NCI Chemical Identifier Resolver [18]. We then translated such chemical representations to chemical fingerprints and compared chemical structure similarity between BCDs and drugs from the PharmGKB by calculating the Tanimoto coefficient [19]. A cheminformatics toolkit, the Chemical Development Kit [20], has been explored to automate these two steps. Finally, PharmGKB drugs with similarity scores higher than 0.7 compared with BCDs were marked as structurally similar BCDs. Thus, more PGx related associations were transformable to BCDs via similar PharmGKB drugs. Appropriate properties for describing the similar structural relationships have been defined and used for inference in PGx OWL profiles for BCDs.

## 3.2. *BCD PGx OWL profile construction and semantic inference*

We captured and integrated PGx related associations for BCDs as PGx profiles. These integrated PGx profiles can then serve as a knowledge base to further infer new drug targets or associations. We established an OWL ontology-based approach for this purpose. More specifically, we developed an OWL ontology that captures 1) comprehensive BCDs' PGx profiles and 2) rules to infer drug targets or other associations based on the profiles. We used the Protégé system[21] for OWL ontology development.

### 3.2.1. *Meta-ontology model definition*

We first defined a meta-ontology model to describe base classes and relationships for the BCD profiles. Base classes include "Drug," "Gene," "Disease," "SNP," and "Pathway." Specific subclasses of these base classes, such as "Breast Cancer Drug" or "Breast Cancer Drug Associated SNP," can also be defined. Relationships between these classes, such as "associatedwithDrug," "associatedwithDisease," "associatedwithSNP," and "associatedwithPathway," have also been defined as object properties with appropriate domains and ranges.

### 3.2.2. *PGx profile representation*

Specific BCDs, SNPs, genes, and pathways were represented as OWL individuals with appropriate types. For example, line 1 in Figure 3 defines Tamoxifen as an instance of the Drug class. Lines 2-5 further represent additional PGx profile information about the Drug Tamoxifen. Similarly, information about particular genes, SNPs, diseases, and pathways can also be stored using RDF

triples. For example, lines 8-10 and 13-14 represent a partial profile of SNP rs2234693 and the drug clomifene, respectively.

### 3.2.3. Identifying new indications for BCDs via semantic inference

New indication candidates identification for BCDs is built on the

basis of PGx related associations and predefined axioms. We used Description Logic (DL)[22] to define axioms shown in Figure 4. For instance, we defined that a disease *di* may associate with a drug *dr* if *di* is either directly associated with *dr* or associated with any gene, pathway, or SNP that is associated with *dr*. For example, we can find tamoxifen-associated diseases using the first axiom listed in Figure 4. Similarly, we can define a tamoxifen-associated SNP, gene, and pathway using OWL DL. Another way to find tamoxifen-associated disease is to search on the basis of its chemical structure. Our method is based on the fact that drugs with the similar structure (isStructuralSimilarto) are very likely to share the same biological properties, which would likely lead to the same disease profile. The second axiom in Figure 4 defines this feature.

1. Tamoxifen rdf:type Drug
2. Tamoxifen associatedwithGene ESR1
3. Tamoxifen associatedwithGene BRCA1
4. Tamoxifen associatedwithDisease Breast\_Cancer
5. Tamoxifen associatedwithPathway Aromatase\_Inhibitor\_Pathway\_(Breast\_Cell)\_Pharmacodynamics
6. Tamoxifen isStructuralSimilarto Clomifene
7. ....
8. rs2234693 rdf:type SNP
9. rs2234693 associatedwithGene ESR1
10. rs2234693 associatedwithDisease Rheumatoid\_Arthritis
11. ESR1 associatedwithDisease Ovarian\_Neoplasms
12. ....
13. Clomifene rdf:type Drug
14. Clomifene associatedwithDisease Polycystic\_Ovary\_Syndrome

Fig. 3. RDF representation for PGx profiles

#### tamoxifen-associated disease:

1. Disease and (associatedwithDrug value Tamoxifen or associatedwithSNP some TamoxifenSNP or associatedwithGene some TamoxifenGene or associatedwithPathway some TamoxifenPathway)
2. Disease and associatedwithDrug some (Drug and isStructuralSimilarto value Tamoxifen)

Fig. 4. Rule representation for PGx OWL profiles.

## 4. Case Study

Using the above semantic definitions, we can infer more information about a particular BCD. We chose tamoxifen, as a use case testbed. “Tamoxifen treats advanced breast cancer in men and women, and early breast cancer in women. And it may prevent breast cancer in women who are at a high risk because of age, family history, or other factors”[23]. We did not invite domain experts to evaluate our inference results for this study, hence, we attempted to validate the performance and usability of PGx OWL profiles by detecting existing hints from the literature as evidence.

Tamoxifen is associated with the BRCA1 gene (a TamoxifenGene, in Figure 3) and BRCA1 is associated with the disease “Ovarian Neoplasms”.

The reasoner can infer ovarian cancer might be associated with tamoxifen via the first axiom listed in

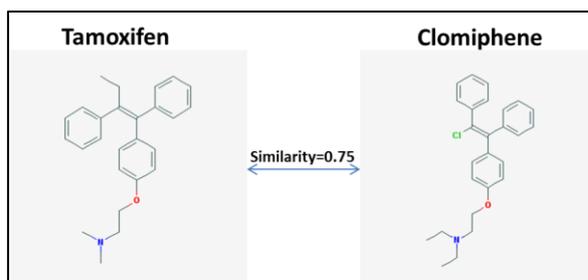


Fig. 5 Structural comparison between tamoxifen and clomifene.

Figure 4. That is to say, tamoxifen can not only treat breast cancer, but also may be used to treat ovarian cancer. Several publications and clinical trials have reported this use of tamoxifen.[24, 25]

“Clomifene treats ovulation problems in women who want to become pregnant”[26]. There are no explicit hints to tie together an ovulation drug and a BCD. However, PGx OWL profiles identified a possible linkage between these two agents. As shown in Figure 5, clomifene and tamoxifen are structurally similar with a similarity score 0.75, which is higher than the threshold 0.7 that we setup. Then the reasoner can infer that tamoxifen may be associated with diseases associated with clomifene (eg, Polycystic\_Ovary\_Syndrome) via the second axiom shown in Figure 4. In 2011, Dhaliwal et al [27] reported that tamoxifen can be prescribed as an alternative to clomifene in women with polycystic ovary syndrome.

In addition to repositioning tamoxifen with other therapeutic uses, we also can identify potential adverse effects by running our PGx OWL profiles based reasoner. From our OWL profiles, as shown in Figure 3, we identified that tamoxifen is associated with the *ESR1* gene as a “TamoxifenGene.” Since the SNP rs2234693 is associated with *ESR1* (a “TamoxifenGene”), rs2234693 is classified as a “TamoxifenSNP” by the reasoner. Furthermore, since rs2234693 is “associatedwithDisease” Rheumatoid Arthritis, then rheumatoid arthritis is identified as a disease that might be associated with tamoxifen by the reasoner. In the real world, as of June 24, 2013, a total of 7,947 people have been reported to have adverse effects when taking tamoxifen citrate. Among them, 35 people (0.44%) have rheumatoid arthritis. [28]

## 5. Results

We generated and presented PGx profiles for 18 breast cancer drugs from NCI by exploring PGx information from PharmGKB. To enable semantic reasoning and to identify more novel PGx associations for BCDs, we created OWL ontology to capture and represent the concepts and relations from PGx profiles.

### 5.1. BCD PGx profile generation

We identified 955 associations for 18 BCDs from the PharmGKB relationship file, which include associations among drugs, genes, and SNPs. We manually identified 287 associations for 18 BCDs from the PharmGKB pathway file, which include associations among pathways, drugs, genes, and diseases.

### 5.2. Chemical structural similarity calculation

To integrate structural similarity, we calculated drug pairs between BCDs and drugs from the PharmGKB. Of 679 unique PharmGKB drugs (including drug classes) extracted from the PharmGKB relationship file, 339 are without SMILES. We invoked NCI chemical resolver to generate SMILES for

these 339 drugs by given drug names, 193 have retrieved SMILES. For the rest of 146 drugs and drug classes without SMILES, we ran PubChem entrez web service to generate SMILES and 37 drugs assigned with SMILES. In total 78 drug classes and 31 drugs were excluded from similarity calculation because no SMILES were generated. For pathway file, we have identified another 71 unique drugs. Among these drugs, there are 65 drugs assigned SMILES via PubChem Entrez web service. Total 5 drugs and 26 drug classes without SMILES were excluded for similarity calculation.

### 5.3. PGx OWL profile generation

BCDs relevant PGx profiles were converted to OWL representation, the drugs, genes, diseases, SNPs from the PharmGKB relationship file and pathway file were also imported into the OWL ontology for inference purpose. A snapshot of the PGx OWL ontology is shown in Figure 6. This ontology includes 294 diseases, 750 drugs including 18 breast cancer drugs, 4277 genes including 215 breast cancer associated genes, 1,426 pathways including 15 breast cancer drugs involved pathways, and 1744 SNPs including 346 breast cancer associated SNPs. It also includes the similarity scores of 10,159 pairs of drugs.

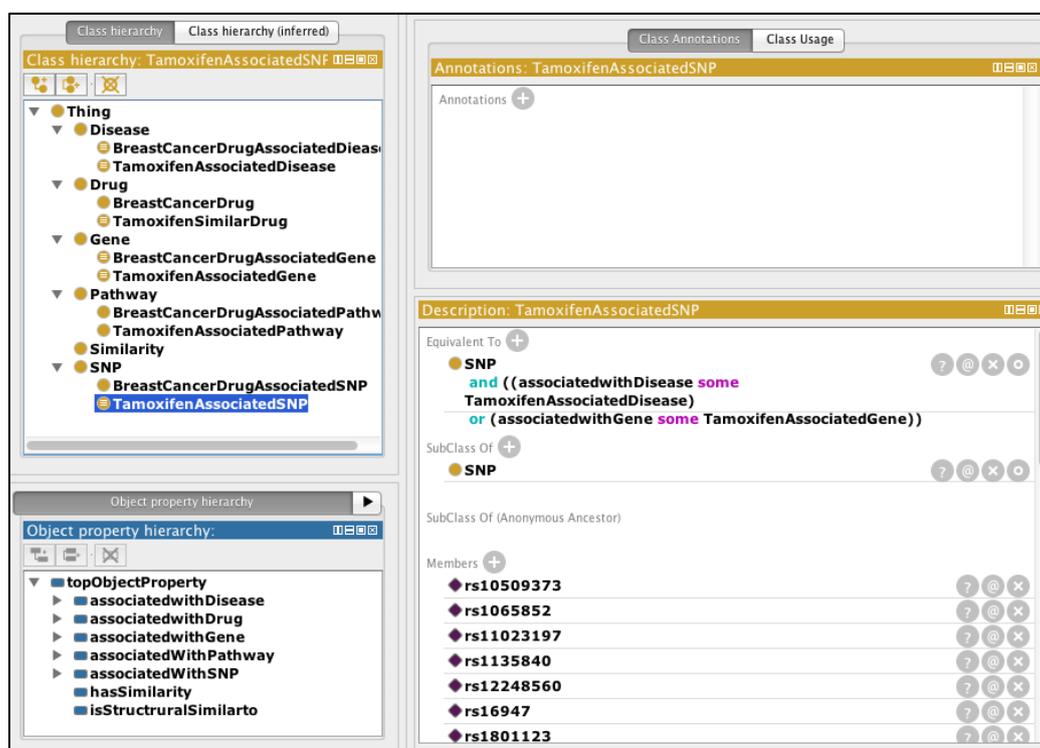


Fig. 6. PGx OWL ontology snapshot

## 6. Discussion and Conclusion

This report presents our preliminary work focusing on computational drug repositioning application development leveraging PGx information integration and Semantic Web technology exploration for

FDA approved BCDs. We have successfully demonstrated the utility of this application to reposition existing BCDs with new uses, and detect potential adverse effects. Our work illustrates that PGx data provides sufficient information to support drug repositioning and, furthermore, that Semantic Web technology provides technical support for formal representation and semantic inference of data.

This is our first attempt to use a PGx resource and Semantic Web technology to address drug repositioning in a computational way. With the promising results of this study, we will expand this investigation in several directions: 1) In the current study, we explored only PharmGKB as a PGx resource, which is not enough to identify more novel associations for BCDs. We will integrate additional PGx-related resources, such as an FDA biomarkers table, the DrugBank database, the Comparative Toxicogenomics Database, and the Kyoto Encyclopedia of Genes and Genomes. 2) Once more PGx resources are integrated, one drug might be inferred to multiple PGx associations. Then we will propose to define some “gold standards” for prioritizing the relevance of these associations to particular drugs. The standards might be built on the number of co-occurrences of the PGx associations, as supported by publications, etc. 3) We worked only on BCDs in this study. In future studies, we will extend our effort to other cancer drug categories or other categories of drugs, such as antidepressants, using the same strategy that we applied in this study.

## 7. Acknowledgments

This work was supported by the Pharmacogenomic Research Network (NIH/NIGMS-U19 GM61388) and the Cancer Prevention & Research Institute of Texas (CPRIT R1307).

## References

1. Ye, H., et al., *A pathway profile-based method for drug repositioning*. Chinese Science Bulletin, 2012. **57**(17): p. 2106-2112.
2. Napolitano, F., et al., *Drug Repositioning: A Machine-Learning Approach through Data Integration*. Journal of Cheminformatics, 2013. **5**(1): p. 30.
3. Li, J. and Z. Lu. *A new method for computational drug repositioning using drug pairwise similarity*. in *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*. 2012: IEEE.
4. Dudley, J.T., et al., *Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease*. Science Translational Medicine, 2011. **3**(96): p. 96ra76.
5. Sirota, M., et al., *Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data*. Science Translational Medicine, 2011. **3**(96): p. 96ra77.
6. Pathak, J., R.C. Kiefer, and C.G. Chute, *Using Semantic Web Technologies for Cohort Identification from Electronic Health Records for Clinical Research*. AMIA Summits on Translational Science Proceedings, 2012. **2012**: p. 10.
7. Searls, D.B., *Data integration: challenges for drug discovery*. Nature Reviews Drug Discovery, 2005. **4**(1): p. 45-58.

8. Olsson, T. and T.I. Oprea, *Cheminformatics: a tool for decision-makers in drug discovery*. Current opinion in drug discovery & development, 2001. **4**(3): p. 308.
9. Hewett, M., et al., *PharmGKB: the pharmacogenetics knowledge base*. Nucleic acids research, 2002. **30**(1): p. 163-165.
10. Dumontier, M. and N. Villanueva-Rosales, *Towards pharmacogenomics knowledge discovery with the semantic web*. Briefings in Bioinformatics, 2009.
11. *Drugs Approved for Breast Cancer*. [cited 2013 July 3rd]; Available from: <http://www.cancer.gov/cancertopics/druginfo/breastcancer>.
12. *Resource Description Framework (RDF)*. [cited 2013 July 17th]; Available from: <http://www.w3.org/RDF/>.
13. *OWL Web Ontology Language*. [cited 2013 July 17th]; Available from: <http://www.w3.org/TR/owl-features/>.
14. *An Executive Intro to Ontologies*. 2009; Available from: <http://www.mkbergman.com/900/an-executive-intro-to-ontologies/>.
15. *Simplified molecular-input line-entry system (SMILES)*. [cited 2013 July 3rd]; Available from: [http://en.wikipedia.org/wiki/Simplified\\_molecular-input\\_line-entry\\_system](http://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system).
16. *The IUPAC International Chemical Identifier (InChI)*. [cited 2013 July 3rd]; Available from: <http://www.iupac.org/home/publications/e-resources/inchi.html>.
17. *PubChem Entrez Programming Utilities*. [cited 2013 July 3rd]; Available from: <http://www.ncbi.nlm.nih.gov/books/NBK25500/>.
18. *NCI Chemical Identifier Resolver*. [cited 2013 July 3rd]; Available from: <http://cactus.nci.nih.gov/chemical/structure>.
19. Butina, D., *Unsupervised data base clustering based on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets*. Journal of Chemical Information and Computer Sciences, 1999. **39**(4): p. 747-750.
20. Steinbeck, C., et al., *The Chemistry Development Kit (CDK): An open-source Java library for chemo-and bioinformatics*. Journal of Chemical Information and Computer Sciences, 2003. **43**(2): p. 493-500.
21. *Protégé*. [cited 2013 July 3rd]; Available from: <http://protege.stanford.edu/>.
22. Baader, F. 2003: Cambridge university press.
23. *Tamoxifen*. [cited 2013 July 3rd]; Available from: <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0012290/?report=details>.
24. Lee, J.-Y., et al., *Effect of combined treatment with progesterone and tamoxifen on the growth and apoptosis of human ovarian cancer cells*. Oncology reports, 2012. **27**(1): p. 87-93.
25. *Clinical Trial for Tamoxifen Compared With Thalidomide in Treating Women With Ovarian Epithelial Cancer, Fallopian Tube Cancer, or Primary Peritoneal Cancer*. [cited 2013 July 3rd]; Available from: <http://clinicaltrials.gov/ct2/show/record/NCT00041080>.
26. *Clomiphene* [cited 2013 July 3rd]; Available from: <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0009673/?report=details>.
27. Dhaliwal, L.K., et al., *Tamoxifen: An alternative to clomiphene in women with polycystic ovary syndrome*. Journal of human reproductive sciences, 2011. **4**(2): p. 76.
28. *Tamoxifen citrate and Rheumatoid arthritis*. [cited 2013 July 3rd]; Available from: <http://www.ehealthme.com/ds/tamoxifen+citrate/rheumatoid+arthritis>.