Journal of Biomedical Informatics xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Disambiguation of PharmGKB drug–disease relations with NDF-RT and SPL

Qian Zhu *, Robert R. Freimuth, Jyotishman Pathak, Matthew J. Durski, Christopher G. Chute

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

ARTICLE INFO

Article history:
Received 26 June 2012
Accepted 18 May 2013
Available online xxxx

Keywords:
Pharmacogenomics
Clinical associations
PharmGKB
NDF-RT
SPL

ABSTRACT

PharmGKB is a leading resource of high quality pharmacogenomics data that provides information about how genetic variations modulate an individual's response to drugs. PharmGKB contains information about genetic variations, pharmacokinetic and pharmacodynamic pathways, and the effect of variations on drug-related phenotypes. These relationships are represented using very general terms, however, and the precise semantic relationships among drugs, and diseases are not often captured. In this paper we develop a protocol to detect and disambiguate general clinical associations between drugs and diseases using more precise annotation terms from other data sources. PharmGKB provides very detailed clinical associations between genetic variants and drug response, including genotype-specific drug dosing guidelines, and this procedure will be adding information about drug–disease relationships not found in PharmGKB. The availability of more detailed data will help investigators to conduct more precise queries, such as finding particular diseases caused or treated by a specific drug.

We first mapped drugs extracted from PharmGKB drug–disease relationships to those in the National Drug File Reference Terminology (NDF-RT) and to Structured Product Labels (SPLs). Specifically, we retrieved drug and disease role relationships describing and defining concepts according to their relationships with other concepts from NDF-RT. We also used the NCBO (National Center for Biomedical Ontology) annotator to annotate disease terms from the free text extracted from five SPL sections (indication, contraindication, ADE, precaution, and warning). Finally, we used the detailed drug and disease relationship information from NDF-RT and the SPLs to annotate and disambiguate the more general PharmGKB drug and disease associations.

© 2013 Published by Elsevier Inc.

1. Introduction

The Pharmacogenomics Knowledge Base (PharmGKB [1]) is a publicly available internet resource for pharmacogenomics data and knowledge that provides information about genes involved in modulating the response to drugs. PharmGKB includes data about genetic variations, pharmacokinetic and pharmacodynamic pathways, and the effects of genetic variations on drug-related phenotypes. PharmGKB also provides integrated knowledge including relationships among genes, drugs, and diseases. The importance of this resource is widely recognized, and it has been used in several investigations. For example, Pathak et al. [2] investigated pharmacodynamics (PDs) and pharmacokinetics (PKs) relationships between drugs and diseases in PharmGKB, and compared them with the drug–disease relationships in NDF-RT. Theobald et al. [3] extracted the relationships between drugs, disease and genes from PubMed guided by the relationships from PharmGKB.

While PharmGKB collects, encodes, and disseminates knowledge about the impact of human genetic variations on drug response [4], the detailed semantic relationships between drugs and diseases are not often captured during the curation process. For example, PharmGKB contains a generic relationship between the drug donepezil (PA449394) and the disease dementia (PA443853), but no detailed information about that relationship is provided. The availability of more precisely annotated drug and disease associations would be invaluable to the research community, especially for drug repositioning studies.

Detailed information about drug–disease relationships would also be valuable to studies that focus on Adverse Drug Events (ADEs). Vilar et al. [5] combined data from the US Food and Drug Administration's (FDA) Adverse Event Reporting System (AERS) with similarity measures of molecular structure, a process that achieved significant improvement in precision in computationally detecting ADEs. Benton et al. [6] conducted a study to identify potential ADEs by using medical message boards and online resources. Jiang et al. [7] developed ADEpedia by using ADE information from SPLs to create a RDF triple store. Each of these studies improve our ability to detect ADEs, which is a crucial step for

* Corresponding author. Tel.: +1 507 538 0187.
E-mail address: zhu.qian@mayo.edu (Q. Zhu).

1532-0464/$ - see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jbi.2013.05.005

Please cite this article in press as: Zhu Q et al. Disambiguation of PharmGKB drug–disease relations with NDF-RT and SPL. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.05.005

UNCORRECTED PROOF

identifying diseases associated with particular drugs. The detailed annotation of drug and disease associations from PharmGKB that is described in this paper adds significant value to existing resources, thereby facilitating studies like those mentioned above.

The overall goal of this study is to enrich the drug and disease associations in PharmGKB with more detailed semantic relationships. We present a method to disambiguate the clinical associations between drugs and diseases in PharmGKB. We utilized the Veterans Affairs National Drug File Reference Terminology (NDF-RT [8]) and DailyMed Structured Product Labeling (SPL [9]) as the sources of detailed information about drug–disease associations, and demonstrated how existing data sources can be used to semantically enrich each other.

2. Background

We annotated drug and disease relationships from PharmGKB using publicly available data sources. Each source is described below, along with rationale for its use in this study.

2.1. PharmGKB

PharmGKB contains genomic, phenotype and clinical information collected from pharmacogenomics studies. More specifically, PharmGKB provides information regarding variant annotations, drug-centered pathway, pharmacogene summaries, clinical annotations, pharmacogenomics based drug-dosing guidelines, drug labels with pharmacogenomics information. PharmGKB also provides to registered research network members tools to browse, query, download, submit, edit and process the information [4].

2.2. Veterans affairs National Drug File Reference Terminology (NDF-RT)

NDF-RT is produced and maintained by the U.S. Department of Veterans Affairs (VA). It is used for modeling drug characteristics including ingredients, chemical structure, dose form, physiologic effect, mechanism of action, pharmacokinetics, and related diseases [10]. NDF-RT concepts are grouped into several general categories, including drug and disease. The drug category includes VA classifications of medications, generic ingredient preparations used in medications, and orderable (clinical) VA drug products. The disease category includes pathophysiologic data as well as certain non-disease physiologic states that are treated, prevented, or diagnosed by an ingredient or drug product. In addition, the disease category may also describe contraindications.

Although PharmGKB provides mappings with ATC (Anatomical Therapeutic Chemical) and DrugBank, we used NDF-RT in this study for several reasons: (1) ATC is a proprietary terminology; (2) neither ATC nor DrugBank provide information about drug and disease associations; (3) NDF-RT has been integrated into RxNorm, which is specified by U.S. Meaningful Use regulations; (4) NDF-RT is updated weekly.

2.3. Structured Product Labeling (SPL)

The SPL standard has been adopted by the FDA for exchanging information about drugs and drug ingredients, including dosage, strength, usage, and known ADEs [10]. SPL documents contain structured sections that in turn contain unstructured content that comprise the product label (including all text, tables and figures). SPL defines the content of human prescription drug labeling in an XML format that are organized by section headings; 76 section headings have been defined and coded by LOINC [11].

SPLs were used in this study as a data source that compliments NDF-RT. Other studies have used this resource in a similar manner, which allows us to extend our current work with other resources and applications.

2.4. National Center for Biomedical Ontology (NCBO) annotator

The NCBO annotator is an ontology-based web service for annotating textual biomedical data with concepts from more than 200 ontologies, which are part of two important repositories of biomedical ontologies and terminologies: the UMLS Metathesaurus and NCBO BioPortal [12]. The NCBO annotator provides the capability for detecting and annotating disease terms within the SPL free text used in this study.

3. Methods

Data sets were retrieved from the PharmGKB July 16th, 2011 snapshot. Of the 24,227 relations in the dataset, 2698 drug–disease relationships were extracted, including 363 relationships between a drug class and a disease. The drug class-disease relationships were excluded from the analysis since NDF-RT only provides relationships between individual drugs and diseases. Therefore, 2335 drug–disease relationships from PharmGKB were annotated in this study. These relationships included 579 distinct drugs (called "PharmGKB drug" in this report), and 444 distinct diseases. The annotation process is illustrated in Fig. 1.

We mapped PharmGKB drugs to records in the NDF-RT and SPL data sets, then used the disease information obtained from these two resources to more precisely annotate the PharmGKB drug–disease relationships. The annotation workflow is shown in Fig. 2. The steps involved are described in the following sections.

3.1. NDF-RT extraction

NDF-RT provides role relationships that describe and define concepts according to their relationships with other concepts. Each role has a domain (the kind of concept whose definition may use the role) and a range (the kind of concept to which the role can refer) [10]. In this study, we focused on the role relationships between NDF-RT concepts that are "Generic Ingredients or Combinations" or "Disease", as shown in Table 1.

The NDF-RT API [13] was used to retrieve information for the 579 PharmGKB drugs, using the drug name as an input parameter (Fig. 1). The query results included the NDF-RT drug identifier (NUI) and its corresponding role relationships (Table 1).

To ensure the accuracy of the mappings between PharmGKB and NDF-RT, we manually reviewed all mappings in which the NDF-RT concept names did not exactly match the PharmGKB drug names. The review process confirmed that all non-exact mappings were due only to differences in spelling, use of synonymous terms, and differences in drug representation format, and therefore they were retained for further analysis.

3.2. SPL extraction

SPLs provide high quality information for marketed drugs, which includes generic names, ingredients, dosage forms, routes of administration, and usage of the drug. While the SPLs are only semi-structured, they are a valuable resource that can be used to enrich the annotation for PharmGKB drug and disease pairs, and therefore we utilized them in this study.

The SPL data set was downloaded from DailyMed on 11/4/2011, and stored in a local database. We extracted the disease information from the five SPL sections listed in Table 2 as detailed below,

UNCORRECTED PROOF

Please cite this article in press as: Zhu Q et al. Disambiguation of PharmGKB drug–disease relations with NDF-RT and SPL. J Biomed Inform (2013), http://dx.doi.org/10.1016/j.jbi.2013.05.005

Download English Version:

https://daneshyari.com/en/article/10355488

Download Persian Version:

https://daneshyari.com/article/10355488

Daneshyari.com