Foundations for a Realism-based Drug Repurposing Ontology

James Schuler, William Mangione, Ram Samudrala*, Werner Ceusters*

Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences at the University at Buffalo, Buffalo, NY, USA

**senior and corresponding authors*

Abstract

Several ontologies represent entities pertinent to the domain of medicinal drugs. An analysis of these ontologies and the related literature shows that they primarily do so from the perspective of treatment and that the definitions for many of the core entities fall short when applied to drug discovery in general and drug repurposing in particular. We therefore redefined or created new elucidations and definitions for terms which are most important to understanding what is meant by 'drug repurposing' using guidelines of ontological realism, thereby making judicious use of the Basic Formal Ontology, the Ontology for Biomedical Investigations, the Ontology for General Medical Science, and the Drug Ontology. We tested the appropriateness of these modifications for the description of a use case on what is involved, and inferred when using the Computational Analysis of Novel Drug Opportunities (CANDO) drug repurposing platform. We found that the definitions proposed remove some of the shortcomings of other ontologies but that still more work is needed to address all issues.

Keywords:

drug repurposing, ontological realism, Basic Formal Ontology

Introduction

It is critical for all involved in any aspect of biomedicine to stay on top of advances in the state of the art of the interplay between drugs and the human body. This is true at all levels of granularity: from the level at which basic science researchers study how drug molecules interact with cellular and subcellular structures, all the way up to the level at which clinicians are aiming to provide optimal direct patient care by prescribing the best suited medicinal products for the diseases from which their patients are suffering.

The amount of information generated is enormous and sifting through it a tedious task unless it could be supported by accurate and reliable automatic methods. This requires, for instance, that such automatic methods would come with some form of understanding what it means for something to be a drug, and to understand what it means for something to be a treatment. It would require also that researchers present their findings in a way that minimizes the risk for automatic methods to misunderstand what is being conveyed. This requires formalization and standardization at all levels of representation ranging from data to information over knowledge, using methods that avoid ambiguities, redundancies, and information loss. One such method is realism-based ontology.

Aspects of biomedicine that have yet to be described ontologically are drug discovery and drug repurposing.

Any drug discovery pipeline involves scientists from numerous disciplines working at different levels of granularity. This leads to numerous, perhaps conflicting, understandings of terms such as 'drug' and 'drug discovery'.

A typical process of drug discovery begins when a biomedical researcher identifies a protein involved in some disease. A computational researcher then uses digital models of the protein and some drug, together with some protocol to use molecular docking to measure the energy of binding (how strong the chemical interaction is) and find the binding pose (the spatial relationship between all atoms in the compound-protein system) of the drug to the protein. Based on these results, the next experiment undertaken may be measuring cell growth in a petri dish, when those cells containing the protein are treated with the drug, i.e., subjected to the presence of some preparation containing the small molecule, e.g., in a liquid preparation. This is an *in vitro* experiment. In some *in vivo* work which follows, some pill or injectable solution containing the drug may be given to some animal model, e.g., an animal such as a mouse which has a disease that is assessed to be similar to a disease which occurs in humans. If these preclinical studies are successful, then clinical trials can be undertaken, going through different phases (I to IV in the United States), with different formulations of the drug and different patient populations. The Food and Drug Administration (FDA) or relevant government authority may then approve the drug for sale and distribution for the studied disease.

Some compounds hypothesized to have useful medicinal properties do not have known 'targets', so a pharmaceutical company or research group may perform a 'high throughput screening' experiment (1). In these experiments, the action of many different compounds against many different proteins are measured in a large well-plate, with promising compounds ('hits') moving on to more careful and specific investigations,

ideally culminating in clinical trials and safe, efficacious human use.

Under some circumstances clinicians may prescribe some drug for another type of disease if they believe it is medically sound. In common language, 'drug repurposing' and its synonym 'drug repositioning' mean finding a new use for an old or previously approved drug. A classic example of drug repurposing is sildenafil (Viagra) (2). Originally developed to treat high blood pressure and chest pain, the male participants in the early clinical trials noticed peculiar side effects pop up. Sildenafil was then studied and sold for treating erectile dysfunction; it was successfully 'repurposed' from one indication to another. Sildenafil has in fact been repurposed for a second time, in this case, to treat pulmonary hypertension (3).

In the above example, drug repurposing was driven by coincidental observations. A better approach would be to turn it into an active search process. That is the goal of the Computational Analysis of Novel Drug Opportunities (CANDO) platform for shotgun drug repurposing (4–10). The platform uses large-scale molecular modeling and docking simulations to calculate drug-target interactions to infer similarity of drug behavior on a proteomic scale. CANDO is composed of several key components such as drug/compound and protein structural data and drug-indication associations (data on whether a particular drug is used in the treatment of a given indication). Although CANDO has already demonstrated success (4), our hypothesis is that a better ontological understanding of drug repurposing experiments and of the relationship between drugs/compounds and diseases will increase the benchmarking performance of the platform and the fidelity of our models to reality. Furthermore, we believe that the integration of realism-based ontologies in CANDO will ensure our work to be directly comparable with other drug discovery, development, and repurposing approaches.

The data sources we have used thus far in CANDO versions include non-ontologic understandings of compounds and disease. For example, in version 1 of CANDO (v1) we used a compound-indication association mapping from the Comparative Toxicogenomics Database (CTD) where the indications are labeled with a Medical Subject Headings (MeSH) identification (11). MeSH is not an ontology, and there are known issues (12). Additionally, our drug and protein structure data sets have never been curated with any ontologies. Therefore, we hypothesize by integrating Open Biomedical Ontologies (OBO) Foundry ontologies which follow ontological realism into CANDO, we will obtain more accurate results with an increased fidelity to reality from our models enabling us to bring repurposed drugs to the market quicker and in a more cost efficient manner.

This paper aims to lay the foundations for this effort.

Methods

We followed a three-step approach: 1) extensive search for relevant literature in drug repurposing, 2) identification of useful existing resources, and 3) ontological analysis, definition and elucidation of key entities to jumpstart a future ontology for drug repurposing.

Literature review

We used the general Google search engine, Google Scholar, and PubMed to look for research articles using combinations of the following terms anywhere in the document or all in the title: 'drug repurposing' ('drug repositioning'), 'ontology', and 'BFO'. The search parameters and counts were established on April 5, and the searches themselves conducted on April 9. The number of articles found is listed in *Table 1*, but relevant articles are scarce.

Table 1: Search for Relevant Publications

search term	search method	Google search	Google Scholar	PubMed
drug repurposing, ontology	anywhere in document	25,300	1,530	40
	in title	64	4	3
drug repositioning, ontology	anywhere in document	24,000	1,610	29
	in title	4	1	0
drug repurposing, BFO	anywhere in document	252	9	0
	in title	0	0	0
drug repositioning, BFO	anywhere in document	355	7	0
	in title	0	0	0
drug repurposing, ontology, BFO	anywhere in document	135	8	0
	in title	0	0	0
drug repositioning, ontology, BFO	anywhere in document	254	7	0
	in title	0	0	0

Most within the scope is 'An Ontology for Description of Drug Discovery Investigations' which follows OBO Foundry principles, uses BFO as its upper-level ontology, and makes judicious use of definitions from Ontology for Biomedical Investigations (OBI) and the Information Artifact Ontology (IAO). This research focuses on the use case of a robot for screening compounds and individual results as opposed to answering, 'What is Drug Discovery/Repurposing?' (13).

Also pertinent is An Ontology for Pharmaceutical Ligands and Its Application for in Silico Screening and Library Design' (14). The researchers sought to fill what they saw as a void in annotation schemes for pharmaceutical ligands, as at the time annotation efforts were focused on genomic sequences. The theme of this work was on development of databases. Additionally, they claimed – oddly – a function of a 'drug' to be at the level of an individual molecular entity.

Finally, we can mention Gómez-Pérez *et al.*, who reviewed several important ontologies used in medicinal chemistry (15). They write short characterizations of ontologies without delving into much detail or describing strengths and weaknesses of a particular tool. The ontologies they enumerate

are grouped into the following categories: ontologies about the classification of chemical compounds, ontologies about the classification of drugs, and ontologies about drug discovery, design, and development.

We thus did not identify any attempt towards formal constructions of a drug repurposing ontology, but only work which uses ontology as part of a drug repurposing experiment.

To prepare for the second step, we took a broad view in analyzing these works, thereby critically analyzing key aspects of drugs, treatment, drug discovery, and drug repurposing as documented in the literature and identifying shortcomings in these attempts.

Relevant existing ontologies

In our attempt to define drug repurposing and build a Drug Repurposing Ontology of related and important terms, we have made judicious use of established ontologies, especially those espousing ontological realism and adhering to the principles of the Open Biomedical Ontologies (OBO) Foundry (16,17). Most of the OBO Foundry ontologies have been built using Basic Formal Ontology (BFO) as a top level ontology, and we retain this for its use (18).

The BioAssay Ontology (BAO) was originally developed to support standardization of data generation, collection, and searching from high-throughput screening (HTS) experiments (19). It was then extensively further developed, expanding its scope to assays and screening results beyond HTS. This included many entities relevant to drug discovery and drug repurposing (20–22).

Recently, efforts have been made to work with other ontologies, such as the Ontology for Biomedical Investigations (OBI) (23,24). The GPCR Ontology is an effort to describe one specific type of 'drug targets', G-protein coupled receptors (GPCRs), and was intended to integrate with the BAO (25). The Drug Target Ontology hopes to describe the sorts of entities with which the molecular entity of 'drug' may interact and cause some effect (26).

The most relevant previous work is the Drug Ontology (DrOn) (27–30), developed by practitioners of ontological realism and aligned with OBO Foundry ontologies. It turned out to be an adequate tool as a starting point for our work.

Ontological analysis

Through careful reading of the biomedical ontology literature and through analysis of definitions and elucidations found using Ontobee (31), we attempted to describe a drug repurposing experiment using available terms, but we found these terms and their definitions, insofar available, inadequate. With this in mind we delved into redefining or creating new definitions for terms which are most important to understanding what is meant by 'drug repurposing' using guidelines of ontological realism, thereby making judicious use of BFO, OBI, the Ontology for General Medical Science (OGMS) (32), and with a focus on the Drug Ontology.

Finally, we applied our new understanding of the entities involved in drug repurposing to describe a use case example, namely, to ontologically describe what is involved, and what is inferred when using the Computational Analysis of Novel Drug Opportunities (CANDO) drug repurposing platform (4).

Results

Definitions

Our definitions or elucidations for all terms we have created or changed are listed in *Table 2*.

Ontological description of a CANDO use case

A key aspect of CANDO is modeling the interaction of compounds with proteins. We have many instances of models of ChEBI:molecules, including ChEBI:protein and ChEBI:compound. Using an instance of some molecular docking software (which is some subtype of OBI:software), e.g., 'CANDOCK' (33), we predict the pose of an interacting compound and protein structure, as well as the corresponding interaction score/energy. After combining individual OBI:datum together, we can complete a process of OBI:drawing a conclusion based on data and then participate in a OBI:prediction about what scattered molecular aggregate whose parts are individual molecular compounds from the earlier computational experiment can be used in some DRO:treatment of a given OGMS:disease after ingestion using an appropriate DrOn:drug product.

The entire process of using CANDO is an occurrent part of some DRO:drug repurposing. Other researchers may use hypotheses generated by us to inform them of which further occurrent parts of the drug repurposing process need to occur, for example, a preclinical study using a mouse model, or a clinical trial with human participants.

Discussion

What counts as 'drug'?

The creators of DrOn recognize different levels of granularity when discussing drugs. First and foremost are the individual molecular entities, namely the single instances of compounds. Next are collections of instances of molecular entities, i.e., the 'portion of pure substances', and the subtypes 'portion of compound' and 'portion of element', or 'portion of mixture'. Finally there is the 'drug product', e.g., a tablet with a specific amount of some 'scattered molecular aggregate' which has an 'active ingredient role' and another scattered molecular aggregate, with an 'excipient role'. Additionally, parts of DrOn include realizable entities that inhere in molecular entities, such as the disposition of an individual molecule to bind to a protein. The DrOn also reveals issues of drug-related entities of other terminologies and ontologies, including those present in the: NDF-RT (National Drug File - Reference Terminology) (34), SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms) (35), ChEBI (Chemical Entities of Biological Interest) (36), OBI, and ATC (Anatomical Therapeutic Chemical Classification System) (37).

Table 2: Foundational definitions for drug repurposing applications. Proposed terms are in bold, re-used terms from existing ontologies are in italics.

Discovery: *process* that creates *information content entities* about aspects of a *portion of reality* which were not documented in some existing body of *information content entities* generally available to some community.
Drug discovery: **discovery** documenting the *disposition* of a **scattered molecular aggregate** to regain or maintain homeostasis.
Drug repurposing: drug discovery documenting the *disposition* of a **scattered molecular aggregate** to **treat** some *disease*, when another such *disposition* is already documented.
Treatment / to treat: *process* that influences the *realization* of a *disease* toward homeostasis.
Scattered molecular aggregate: *object aggregate* that consists of all molecules that are located in some bounded region.
Scattered molecular aggregate delivery: *function* of a *drug product* to enable some **scattered molecular aggregate** to be located in the appropriate *spatiotemporal region* such that the **scattered molecular aggregate** can *participate* in **treatment**
Prodrug: *role* inhering in a **scattered molecular aggregate** x_i composed out of molecules which have the *disposition* to undergo a chemical transformation to *molecules* of another type resulting in x_i becoming the bearer of a *disposition* to participate in a **treatment**.

While we found the Drug Ontology to be the best and most relevant ontology for our work in describing drug repurposing, we do not commit to the existence and definition of certain entities committed to in DrOn. This precludes us from accurately describing our drug repurposing research in their terms.

Firstly, we believe there is an inconsistency with two critical terms used by DrOn. OBI defines a 'scattered molecular aggregate' (SMA) to be 'a material entity that consists of all the molecules of a specific type that are located in some bounded region and which is part of a more massive material entity that has parts that are other such aggregates'[1]. DrOn uses SMA in related definitions. A 'drug product' is defined as 'a material entity (1) containing at least one scattered molecular aggregate as part that is the bearer of an active ingredient role and (2) that is itself the bearer of a clinical drug role'[2]. The definition as written implies that if a scattered molecular aggregate exists, then it exists necessarily as part of a larger entity with other scattered molecular aggregate parts. Nonetheless, the definition for drug product uses the phrase 'at least one scattered molecular aggregate as part', which implies a drug product could exist with a single scattered molecular aggregate as a part. This seems to be inconsistent.

One way to solve this inconsistency, and to better represent the reality of drugs and drug repurposing, is to use a term to signify an object aggregate consisting of molecular entities. There are related terms in DrOn, chiefly, 'portion of pure substance', 'portion of mixture' and 'scattered molecular aggregate'. We believe changing the definition of SMA to, 'an object aggregate that consists of all molecules that are located in some bounded region', provides nice solutions, namely, removing the inconsistency, and giving us the ability to talk about both portions of pure substances and portions of mixtures.

A drug product is not generally without use, however. Indeed, a function which inheres in a given drug product may be an instance of an entity we call 'scattered molecular aggregate delivery', which we define as, 'a function of a drug product to enable some molecular aggregate to be located in the appropriate spatiotemporal region such that the molecular aggregate can participate in treatment'. It is critical a scattered molecular aggregate is at the appropriate location at the correct time to realize its disposition.

Drug Discovery and Drug Repurposing as a process

Drug repurposing is a subtype of drug discovery, which is a subtype of discovery, which is a subtype of process. We do not claim to have proposed a general definition of 'discovery' as we recognize that the very notion crosses many boundaries of sciences and that the term is also used in non-scientific contexts. We do not, for instance, include uses of the word 'discovery' as when a child 'discovers' an Easter egg under some plant while hunting for Easter eggs.

Treatment

We found the term for 'treatment' from OGMS to be problematic, both in general usage and for our current needs. Based on version 1.0 of BFO, the OGMS definition is 'a processual entity whose completion is hypothesized (by a healthcare provider) to alleviate the signs and symptoms associated with a disorder'[3]. Although present in the OWL-version of OGMS, this term was not defined in the foundational paper which is at the basis of the OGMS (32). Entities on the side of the patient should insofar possible never be defined on the basis of what is known or hypothesized about them. In this case, the definition allows for a physician to say 'I hypothesize some homeopathic regimen will decrease the size of your tumor'. As any homeopathic treatment would never be the causative agent in shrinking the size of the tumor, the hypothesis is false (38), but by the current definition, the homeopathic regimen would be a treatment.

We define 'treatment' as a 'process that influences the realization of a disease toward homeostasis'. The consequence is that a 'treatment' that doesn't work is not a treatment under this definition. In other words: what one in general language would call 'an unsuccessful treatment' is under our definition

[1] http://purl.obolibrary.org/obo/OBI_0000576

[2] http://purl.obolibrary.org/obo/DRON_00000005

[3] http://purl.obolibrary.org/obo/OGMS_0000090

no treatment at all. Note that when such a process about which we hypothesize it will benefit the patient is started, we will only know whether the process is an instance of treatment after observing the desired results of the process. This is similar with the side effect involved in the common definition of chronic pain as 'a pain that is present for at least 3 months': it means that when presented with a patient exhibiting pain since one day, that pain might already be a chronic pain but we have to wait 3 months before we are able to identify that pain as such. Note also that it does not matter what kind of process is done or on what something is done as long as the disease realization is changed towards homeostasis.

A scattered molecular aggregate may have the disposition to influence the homeostasis of an organism. If this disposition is to regain or maintain homeostasis, and the scattered molecular aggregate exists in a sufficient amount, and the disposition is realized, a treatment has occurred. If this disposition of a scattered molecular aggregate was specifically evolved or designed for, then it is a function.

Besides 'homeostasis', we are using also the OGMS terms and definitions of disorder, disease, and disease course by Scheuermann et al. (32) to justify our definition for treatment. With a disorder being the physical basis of some disposition to undergo pathological processes (disease), and a disease course the totality of all processes through which a disease is realized. Eliminating the disorder gets rid of the corresponding disease and any potential disease course thereof (although, of course, further disorders for which the former diseases was a pre-disposition might continue to exist). For example, if there is a mutation in one's DNA which causes a protein to misfold and perform some actions which, if left 'untreated' would cause problems in the heart leading to death, and if the totality of misfolded proteins is successfully inhibited using some 'drug', then a disorder is still present, in the form of instances of misfolded proteins. The formation of a misfolded protein is itself a pathological process, and so the disease is still being realized. However, the temporal parts of the disease course that are realized after the drug is doing its job, are of different types than the parts before: the disease has been influenced toward homeostasis so that the person will not, for example, experience heart problems or death; there will just be the production of misfolded proteins.

For every instance of a scattered molecular aggregate composed of particular molecules, the disposition to treat a particular disease inheres in all such instances. This is not to say any instance of an SMA has some disposition to treat a disease: only those whose parts consist of particular molecules, i.e. those that have the disposition to interact with bodily components such as proteins that participate in the realization of some disease. The disposition exists whether it is known to science or not.

A function to treat a disease only inheres in some portion of compound if the molecular entity parts have evolved or been designed to participate in the treatment process.

If a company manufactures some portion of aspirin with only the specific intent to treat headaches, this portion has the function to treat headaches, but has no other function. The disposition, but not the function, to prevent or minimize the consequences of a heart attack inheres in that particular portion of aspirin, but if it has not been manufactured for that purpose, it is not its function. This is consistent with the Drug Ontology to some degree, but we disagree about in what entity the function inheres. According to DrOn, it inheres in the drug product (e.g., pill). We believe this to be false, and claim that any realizable entity related to treatment inheres in some scattered molecular aggregate (a term for which we are suggesting an updated definition).

Consider a person consuming a drug product for which it is claimed that there inheres some function to treat renal cell carcinoma. If the drug product is a tablet which is meant to be chewed, and if the person chews the tablet, then the tablet is no longer in existence, but no function to treat the cancer has been realized. However, a portion of compound which was previously a part of the tablet is appropriately distributed throughout the body. The molecular entities which make up the SMA realize their disposition to bind to and inhibit certain disordered proteins, i.e., the disorder. In the ultimate case, the renal cell carcinoma tumor is destroyed and the treatment process is complete. In this situation, there is indeed some entity which was pivotal in the treatment, but it cannot have been the tablet, as it was not in existence during the entire temporal region during which the treatment, i.e., the elimination of the tumor, occurred. As we agree with the creators of the Drug Ontology such a realizable entity does not inhere in individual molecules, we therefore say it must have been some scattered molecular aggregate.

One question might be: which one precisely? There are indeed widely variable amounts of portions of compound in which these treatment functions may inhere. For example, a function to treat a bacterial infectious disease may inhere in the scattered molecular aggregate which is contained in 20 tablets of some antibiotic pill. In the case of a chronic illness such as essential hypertension, a function inheres in the portion of compound contained in all the tablets a person with essential hypertension ingests over the course of some treatment.

Consider another example where a portion of compound has some function to treat a disease, i.e., scientists have discovered it has such a disposition, and the portions of compound are manufactured specifically for this purpose. If we have a powder of this portion of compound which can be absorbed into the body through the buccal mucosa, enter the bloodstream, and end up in the correct location where it will be able to realize its function, then by simply placing the powder underneath the tongue, one is enabling the portion of compound to begin the process of realizing its function. In this case, no drug product is ever present as a Drug Ontology drug product contains, by definition, at least several scattered molecular aggregates as parts. The entity which participates in the treatment which results in the beneficial amelioration of some disorder, disease, or disease course is the molecular aggregate of compound. Similarly, chewing tree bark which contains a portion of aspirin to relieve headache involves no drug product (39).

Prodrugs

The view of some treatment disposition or function to inhere in a scattered molecular aggregate and not in a drug product

also lends itself to better understand 'prodrugs' and combination therapies. A prodrug is generally described as 'a drug for which the dosed ingredient is an inactive or only mildly efficacious entity, but once in the body it is converted to the active ingredient by either a spontaneous or an enzyme-catalysed reaction' (40). Sofosbuvir, a drug used in the treatment of hepatitis C, is an example of a prodrug (41). The scattered molecular aggregate which is in a drug product may not have the disposition or function to engage in some treatment for a given disease. Each individual molecular entity does have the disposition to be modified in some way to a molecular entity of a different type, and the resulting molecular aggregate, composed of different molecular entities, is where any realizable entity related to treatment inheres.

Our new understanding of prodrugs can be highlighted with several cases. A particular disease treatment may consist of taking more than one drug product at a time. In one scenario, one or both of the molecular aggregates in the drug products may have the disposition to treat the disease by themselves. In another, none of the molecular aggregates have any disposition to treat the disease by themselves, but rather only when both are in the body at the same time does some therapeutic effect occur. This type of interaction has been discovered through analysis of electronic health record (EHR) data by Tatonetti et al (42). In all of these scenarios, none of the combinations of molecular aggregates may exist in any individual drug product, and yet some disposition or function to treat the disease certainly exists in the combination of molecular aggregates.

Limitations and Future Work

While we have suggested changing the definition of scattered molecular aggregate to better fit our understanding, we recognize this may be too dramatic, and perhaps we could simply create a new term, and keep SMA as a term to refer to some 'molecular aggregates' in a drug product, specifically. However, we wish to define some entity which subsumes both 'portion of compound' and 'portion of mixture', as in the Drug Ontology they are both currently subtypes of BFO:object. We believe some new supertype, if we keep the original definition for SMA, would be a subtype of BFO:object aggregate.

There remains difficulty in creating an ontology so general it can accurately describe every aspect of pharmaceuticals, both from the clinical and research perspective. The entire drug discovery or drug repurposing process is complex and sometimes one claim may not be applicable to another instance of how it is believed some other drug 'works'.

Armed with our improved understanding of the drug repurposing process, we aim to incorporate a more rigorous ontological understanding in future computational experiments with CANDO to better describe the compounds, proteins, diseases, and related associations.

Conclusions

We have found what we believe are errors in the understanding and definitions of core entities in drug discovery, drug repurposing and drug treatment. Chief among them are 'treatment' and several entities in the Drug Ontology describing basic tenants of 'drugs', which made it difficult to accurately describe the reality of drug discovery and drug repurposing. The definitions proposed here remove some of the shortcomings of other ontologies. More work is however needed for 'scattered molecular aggregate': the revision proposed here eliminates inconsistencies but leaves further questions open.

Acknowledgements

This work was supported in part by a 2010 National Institute of Health Director's Pioneer Award [1DP1OD006779], a National Institute of Health Clinical and Translational Sciences Award [UL1TR001412], a National Library of Medicine T15 Award [T15LM012495], a NCI/VA BD-STEP Fellowship in Big Data Sciences, and startup funds from the Department of Biomedical Informatics at the University at Buffalo.

We wish to thank all members of the Fall 2018 Biomedical Ontology course from the Departments of Philosophy and Biomedical Informatics, at the University at Buffalo, who offered feedback on an early iteration of this work.

Address for correspondence

The corresponding authors are Ram Samudrala (ram@compbio.org) and Werner Ceusters (wceusters@gmail.com); Gateway Building, 77 Goodell Street, Suite 540, Buffalo, NY 14203.

References

1. Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell. 2009 Aug 21;138(4):645–59.
2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004 Aug;3(8):673–83.
3. Nelson SJ, Oprea TI, Ursu O, Bologa CG, Zaveri A, Holmes J, et al. Formalizing drug indications on the road to therapeutic intent. J Am Med Inform Assoc. 2017 Nov 1;24(6):1169–72.
4. Minie M, Chopra G, Sethi G, Horst J, White G, Roy A, et al. CANDO and the infinite drug discovery frontier. Drug Discov Today. 2014 Sep;19(9):1353–63.
5. Sethi G, Chopra G, Samudrala R. Multiscale modelling of relationships between protein classes and drug behavior across all diseases using the CANDO platform. Mini Rev Med Chem. 2015;15(8):705–17.
6. Chopra G, Samudrala R. Exploring Polypharmacology in Drug Discovery and Repurposing Using the CANDO Platform. Curr Pharm Des. 2016;22(21):3109–23.
7. Chopra G, Kaushik S, Elkin PL, Samudrala R. Combating Ebola with Repurposed Therapeutics Using the CANDO Platform. Molecules [Internet]. 2016 Nov 25;21(12). Available from: http://dx.doi.org/10.3390/molecules21121537
8. Mangione W, Samudrala R. Identifying Protein Features Responsible for Improved Drug Repurposing Accuracies

Using the CANDO Platform: Implications for Drug Design. Molecules [Internet]. 2019 Jan 4;24(1). Available from: http://dx.doi.org/10.3390/molecules24010167
9. Schuler J, Samudrala R. Fingerprinting CANDO: Increased Accuracy with Structure and Ligand Based Shotgun Drug Repurposing [Internet]. bioRxiv. 2019 [cited 2019 Apr 11]. p. 591123. Available from: https://www.biorxiv.org/content/10.1101/591123v1.abstract
10. Falls Z, Mangione W, Schuler J, Samudrala R. Exploration of interaction scoring criteria in the CANDO platform [Internet]. bioRxiv. 2019 [cited 2019 Apr 11]. p. 591578. Available from: https://www.biorxiv.org/content/10.1101/591578v1.abstract
11. Lipscomb CE. Medical Subject Headings (MeSH). - PubMed - NCBI [Internet]. [cited 2019 Apr 10]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10928714
12. Cowell LG, Smith B. Infectious Disease Ontology [Internet]. Infectious Disease Informatics. 2010. p. 373–95. Available from: http://dx.doi.org/10.1007/978-1-4419-1327-2_19
13. Qi D, King RD, Hopkins AL, Bickerton GRJ, Soldatova LN. An ontology for description of drug discovery investigations. J Integr Bioinform [Internet]. 2010 Mar 25;7(3). Available from: http://dx.doi.org/10.2390/biecoll-jib-2010-126
14. Schuffenhauer A, Zimmermann J, Stoop R, van der Vyver J-J, Lecchini S, Jacoby E. An ontology for pharmaceutical ligands and its application for in silico screening and library design. J Chem Inf Comput Sci. 2002 Jul;42(4):947–55.
15. Gómez-Pérez A, Martínez-Romero M, Rodríguez-González A, Vázquez G, Vázquez-Naya JM. Ontologies in medicinal chemistry: current status and future challenges. Curr Top Med Chem. 2013;13(5):576–90.
16. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251–5.
17. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Appl Ontol. 2010 Jan 1;5(3-4):139–88.
18. Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology [Internet]. 2015. Available from: http://dx.doi.org/10.7551/mitpress/9780262527811.001.0001
19. Vempati U, Visser U, Abeyruwan S, Sakurai K, Przydzial M, Chung C, Smith RP, Koleti A, Mader C, Lemmon VP, Schürer SC. Bioassay Ontology to Describe High-Throughput Screening Assays and their Results. In University of Miami; 2011. p. 209–16.
20. Schürer SC, Vempati U, Smith R, Southern M, Lemmon V. BioAssay ontology annotations facilitate cross-analysis of diverse high-throughput screening data sets. J Biomol Screen. 2011 Apr;16(4):415–26.
21. Visser U, Abeyruwan S, Vempati U, Smith RP, Lemmon V, Schürer SC. BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. BMC Bioinformatics. 2011 Jun 24;12(1):257.
22. Vempati UD, Przydzial MJ, Chung C, Abeyruwan S, Mir A, Sakurai K, et al. Formalization, Annotation and Analysis of Diverse Drug and Probe Screening Assay Datasets Using the BioAssay Ontology (BAO). PLoS One. 2012 Nov 14;7(11):e49198.
23. Abeyruwan S, Vempati UD, Küçük-McGinty H, Visser U, Koleti A, Mir A, et al. Evolving BioAssay Ontology (BAO): modularization, integration and applications. J Biomed Semantics. 2014 Jun 3;5(1):S5.
24. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLoS One. 2016 Apr 29;11(4):e0154556.
25. Przydzial MJ, Bhhatarai B, Koleti A, Vempati U, Schürer SC. GPCR ontology: development and application of a G protein-coupled receptor pharmacology knowledge framework. Bioinformatics. 2013 Dec 15;29(24):3211–9.
26. Lin Y, Mehta S, Küçük-McGinty H, Turner JP, Vidovic D, Forlin M, et al. Drug target ontology to classify and integrate drug discovery data. J Biomed Semantics. 2017 Nov 9;8(1):50.
27. Hogan WR, Hanna J, Joseph E, Brochhausen M. Towards a Consistent and Scientifically Accurate Drug Ontology. CEUR Workshop Proc. 2013;1060:68–73.
28. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. J Biomed Semantics. 2013 Dec 18;4(1):44.
29. Hanna J, Bian J, Hogan WR. An accurate and precise representation of drug ingredients. J Biomed Semantics. 2016 Apr 19;7:7.
30. Hogan WR, Hanna J, Hicks A, Amirova S, Bramblett B, Diller M, et al. Therapeutic indications and other use-case-driven updates in the drug ontology: anti-malarials, anti-hypertensives, opioid analgesics, and a large term request. J Biomed Semantics. 2017 Mar 3;8(1):10.
31. Ong E, Xiang Z, Zhao B, Liu Y, Lin Y, Zheng J, et al. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. Nucleic Acids Res. 2017 Jan 4;45(D1):D347–52.
32. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit Transl Bioinform. 2009 Mar 1;2009:116–20.
33. Fine JA, Konc J, Samudrala R, Chopra G. CANDOCK: Chemical atomic network based hierarchical flexible docking algorithm using generalized statistical potentials [Internet]. Available from: http://dx.doi.org/10.1101/442897
34. Brown SH, Elkin PL, Rosenbloom ST, Husser C, Bauer BA, Lincoln MJ, et al. VA National Drug File Reference Terminology: a cross-institutional content coverage study. Stud Health Technol Inform. 2004;107(Pt 1):477–81.

35. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform. 2006;121:279–90.
36. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 2016 Jan 4;44(D1):D1214–9.
37. Anatomical Therapeutic Chemical Classification System (WHO) [Internet]. The SAGE Encyclopedia of Pharmacology and Society. Available from: http://dx.doi.org/10.4135/9781483349985.n37
38. Angell M, Kassirer JP. Alternative medicine--the risks of untested and unregulated remedies. N Engl J Med. 1998 Sep 17;339(12):839–41.
39. Weissmann G. Aspirin. Sci Am. 1991 Jan;264(1):84–90.
40. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017 Jan;16(1):19–34.
41. Sofia MJ, Bao D, Chang W, Du J, Nagarathnam D, Rachakonda S, et al. Discovery of a β-d-2'-deoxy-2'-α-fluoro-2'-β-C-methyluridine nucleotide prodrug (PSI-7977) for the treatment of hepatitis C virus. J Med Chem. 2010 Oct 14;53(19):7202–18.
42. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012 Mar 14;4(125):125ra31.