A peer-reviewed version of this preprint was published in PeerJ on 13 February 2017.

McCusker JP, Dumontier M, Yan R, He S, Dordick JS, McGuinness DL. 2017. Finding melanoma drugs through a probabilistic knowledge graph. PeerJ Computer Science 3:e106
https://doi.org/10.7717/peerj-cs.106

A peer-reviewed version of this preprint was published in PeerJ on 13 February 2017.

McCusker JP, Dumontier M, Yan R, He S, Dordick JS, McGuinness DL. 2017. Finding melanoma drugs through a probabilistic knowledge graph. PeerJ Computer Science 3:e106
https://doi.org/10.7717/peerj-cs.106

Finding Melanoma Drugs Through a Probabilistic Knowledge Graph

Jamie P. McCusker[1*], Michel Dumontier[4], Rui Yan[1], Sylvia He[1] , Jonathan S. Dordick[2,3], and Deborah L. McGuinness[1,3]

[1]**Department of Computer Science,**
[2]**Department of Chemical & Biological Engineering,**
[3]**Center for Biotechnology & Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, US**
[4]**Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, US**

Corresponding Author:
James P. McCusker
110 8th Street, Troy, NY 12180, USA
Email address: mccusj@cs.rpi.edu

ABSTRACT

Metastatic cutaneous melanoma is an aggressive skin cancer with some progression-slowing treatments but no known cure. The omics data explosion has created many possible drug candidates, however filtering criteria remain challenging, and systems biology approaches have become fragmented with many disconnected databases. Using drug, protein, and disease interactions, we built an evidence-weighted knowledge graph of integrated interactions. Our knowledge graph-based system, ReDrugS, can be used via an API or web interface, and has generated 25 high quality melanoma drug candidates. We show that probabilistic analysis of systems biology graphs increases drug candidate quality compared to non-probabilistic methods. Four of the 25 candidates are novel therapies, three of which have been tested with other cancers. All other candidates have current or completed clinical trials, or have been studied in *in vivo* or *in vitro*. This approach can be used to identify candidate therapies for use in research or personalized medicine.

Keywords: melanoma, drug repositioning, knowledge graphs, uncertainty reasoning

1 INTRODUCTION

Metastatic cutaneous melanoma is an aggressive cancer of the skin with low prevalence but very high mortality rate, with an estimated 5 year survival rate of 6 percent (Barth et al., 1995) There are currently no known therapies that can consistently cure metastatic melanoma. Vemurafenib is effective against BRAF mutant melanomas (Chapman et al., 2011) but resistant cells often result in recurrence of metastases (Le et al., 2013) Melanoma itself may be best approached based on the individual genetics of the tumor, as it has been shown to involve mutations in many different genes to produce the same disease (Krauthammer et al., 2015). Because of this, an individualized approach may be necessary to find effective treatments.

Drug repurposing, or the discovery of new uses for existing approved drugs, can often lead to effective new treatments for diseases. A wide range of computational methods have been developed in support of drug repositioning. Computational approaches (Sanseau and Koehler, 2011) include topic modeling, (Bisgin et al., 2012, 2014) side effect similarity,(Yang and Agarwal, 2011; Ye et al., 2014) drug and/or disease similarity (Chiang and Butte, 2009; Gottlieb et al., 2011), genome-wide association studies (Kingsmore et al., 2008; Grover et al., 2014), and gene expression (Lamb et al., 2006; Sirota et al., 2011) Systems biology has also provided a number of network analysis approaches (Yang and Agarwal, 2011; Wu et al., 2013b; Cheng et al., 2012; Emig et al., 2013; Harrold et al., 2013; Wu et al., 2013a; Vogt et al., 2014) but the field has been limited by a fragmentation of databases. Most systems biology databases

are not aligned with each other, and typically leave out crucial information about how other biological entities, like drugs and diseases, interact with the systems biology graph. Further, while some interaction databases provide human curation and validation of pathway interactions, and others provide experimental evidence for the recorded interactions, there has not yet been, to our knowledge, a resource that combines the two approaches and quantifies the reliability of the evidence used to assert the interactions.

A knowledge graph is a compilation of facts and figures that can be used to provide contextual meaning to searches. Google is using knowledge graphs to improve its search and to analyze the information graph of the web; Facebook is using them to analyze the social graph. We built our knowledge graph with the goal of unifying large parts of biomedical domain knowledge for both mining and interactive exploration related to drugs, diseases, and proteins. Our knowledge graph is enhanced by the provenance of each fragment of knowledge captured, which is used to compute the confidence probabilities for each of those fragments. Further, we use open standards from the World Wide Web Consortium (W3C), including the Resource Description Framework (RDF) (Klyne and Carroll, 2005), Web Ontology Language (OWL) (Group et al., 2009), and SPARQL (Harris et al., 2013). The representation of the knowledge in our knowledge graph is aligned with best practice vocabularies and ontologies from the W3C and the biomedical community, including the PROV Ontology (Lebo et al., 2013), the HUPO Proteomics Standards Initiative Molecular Interactions (PSI-MI) Ontology (Hermjakob et al., 2004), and the Semanticscience Integrated Ontology (SIO) (Dumontier et al., 2013). Use of these standards, vocabularies, and ontologies make it simple for ReDrugS to integrate with other similar efforts in the future with minimal effort.

We proposed and built a novel computational drug repositioning platform, that we refer to as ReDrugS, that applies probabilistic filtering over individually-supported assertions drawn from multiple databases pertaining to systems biology, pharmacology, disease association, and gene expression data. We use our platform to identify novel and known drugs for melanoma.

2 RESULTS

We used ReDrugS to examine the drug-target-disease network and identify known, novel, and well supported melanoma drugs. The ReDrugS knowledge base contained 6,180 drugs, 3,820 diseases, 69,279 proteins, and 899,198 interactions.

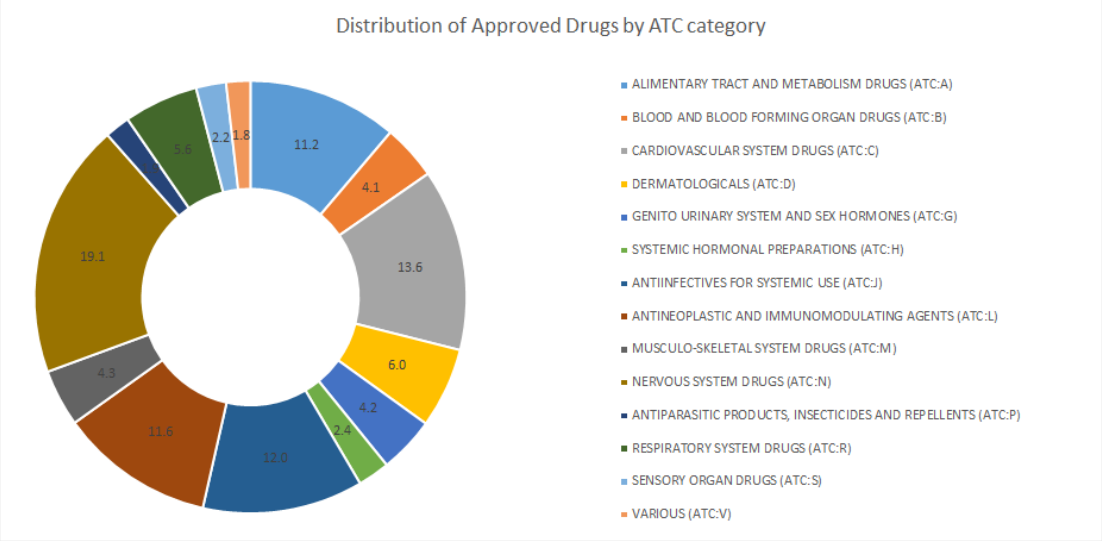

Figure 1. Percentage approved drugs in each of the categories of the Anatomic Therapeutic Classification (ATC) system.

We examined drug and gene connections that were 3 or less interaction steps from melanoma, and additionally filtered interactions with a joint probability greater or equal to 0.93. We identified 25 drugs in the resulting drug-gene-disease network surrounding melanoma as illustrated in Figure 2 .

We then validated the set of 25 drugs by determining their position in the drug discovery pipeline for

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

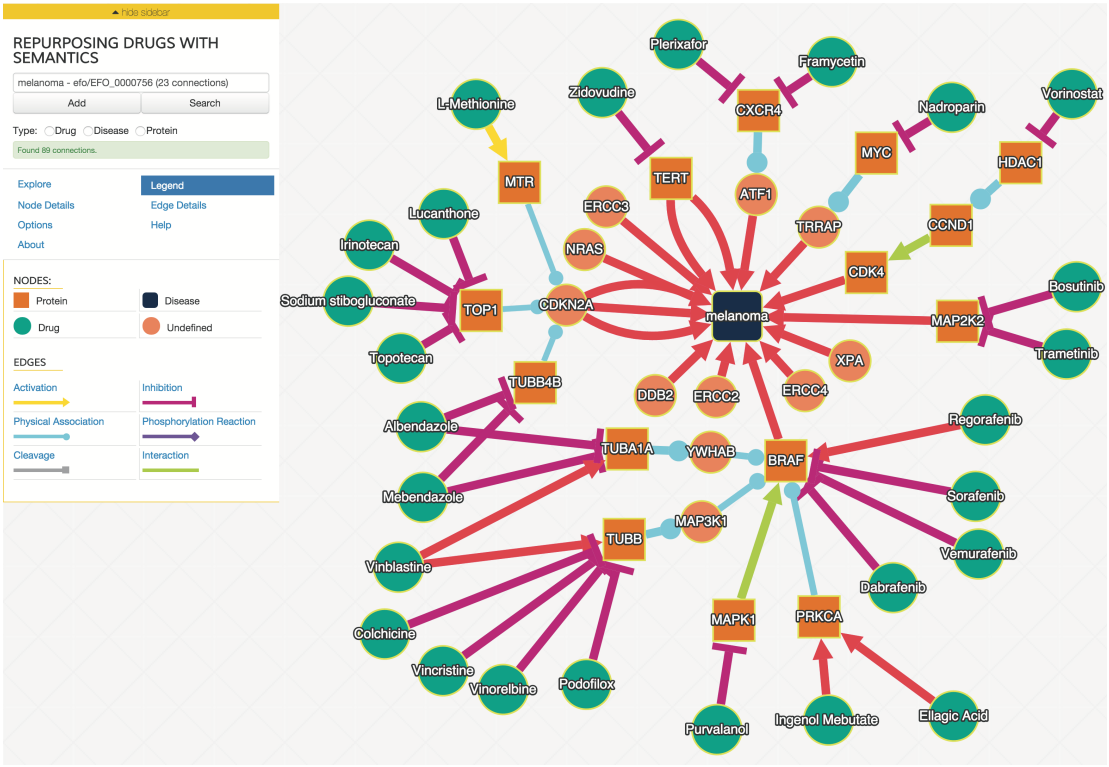

Figure 2. The interaction graph of predicted melanoma drugs with a probability of 0.93 or higher and have three or fewer intervening interactions between drug and disease. The "Explore" tab contains the controls to expand the network in various ways, including the filtering parameters. Node and edge detail tabs provide additional information about the selected node or edge, including the probabilities of the edges selected. Users can control the layout algorithm and related options using the "Options" tab.

melanoma. Table 1 shows that nearly all drugs uncovered by ReDrugS were previously been identified as potential melanoma therapies either in clinical trials or *in vivo* or *in vitro*. Of the 25 drugs, 12 have been in Phase I, II, or III clinical trials, 5 have been studied *in vitro*, 4 *in vivo*, 1 was investigated as a case study, and 3 are novel.

To further evaluate our system, we examined the impact of decreasing the joint probability or increasing the number of interaction steps. Figures 3 A and B show precision, recall, and f-measure curves while varying each parameter. Using these information retrieval performance curves we found that using a joint probability of 0.93 or greater with 3 or less interaction steps maximizes the precision and recall as shown in Figure 3.

By performing a sampled literature search on hypothesis candidates with a joint probability of 0.5 or higher and 6 or fewer interaction steps, we were able to generate precision, recall, and f-measure curves for both cutoffs to find our cutoff of 0.93 with 3 or fewer interaction steps. The precision, recall, and f-measure curves are shown for varying joint probability thresholds in Figure 3 A and for varying interaction step counts in Figure 3 B.

3 DISCUSSION

We designed ReDrugS to quickly and automatically integrate and filter a heterogeneous biomedical knowledge graph to generate high-confidence drug repositioning candidates. Our results indicate that ReDrugs generates clinically plausible drug candidates, in which half are in various stages of clinical trials, while others are novel or are being investigated in pre-clinical studies. By helping to consolidate the three main datatypes - drug targets, protein interactions, and disease genes, ReDrugs can amplify the ability of researchers to filter the vast amount of information into those that are relevant for drug discovery.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Table 1. Drug discovery status for 25 drug candidates identified using ReDrugS. "Pathway" refers to the target or pathway that the drug acts on. "Steps" is distance in number of interactions between the drug and the disease, and "Joint p" is the joint probability that all of those interactions occur.

Status	Drug	Pathway	Steps	Joint p
Approved	Vemurafenib (Chapman et al., 2011)	BRAF	2	0.98
Phase III	Dabrafenib (Hauschild et al., 2012)	BRAF	2	0.98
	Sorafenib (National Cancer Institute, 2005)	BRAF	2	0.98
	Vinblastine (Luikart et al., 1984)	MAP kinase	3	0.93
Phase II	Zidovudine (Humer et al., 2008)	TERT	2	0.98
	Trametinib (Kim et al., 2012)	MAP kinase	2	0.98
	Regorafenib (Istituto Clinico Humanitas, 2015)	BRAF	2	0.98
	Nadroparin (Nagy et al., 2009)	MYC	3	0.97
	Vinorelbine (Whitehead et al., 2004)	MAP kinase	3	0.93
	Irinotecan (Fiorentini et al., 2009)	CDKN2A	3	0.93
	Topotecan (Kraut et al., 1997)	CDKN2A	3	0.93
Phase I	Sodium stibogluconate (Naing, 2011)	CDKN2A	3	0.93
Case Study	Ingenol Mebutate (Mansuy et al., 2014)	PRKCA/BRAF	3	0.95
In Vitro	Bosutinib (Homsi et al., 2009)	MAP kinase	2	0.98
	Purvalanol (Smalley et al., 2007)	MAP kinase/TP53	3	0.97
	Ellagic Acid (Kim et al., 2008)	PRKCA/BRAF	3	0.95
	Albendazole (Patel et al., 2011)	CDKN2A	3	0.93
	Colchicine (Lemontt et al., 1988)	MAP kinase	3	0.93
In Vivo	Plerixafor (D'Alterio et al., 2012)	CXCR4	3	0.97
	Vincristine (Sawada et al., 2004)	MAP kinase	3	0.93
	L-Methionine (Clavo and Wahl, 1996)	CDKN2A	3	0.93
	Mebendazole (Doudican et al., 2008)	CDKN2A	3	0.93
Novel	Framycetin	CXCR4	3	0.97
	Lucanthone	CDKN2A	3	0.93
	Podofilox	MAP kinase	3	0.93

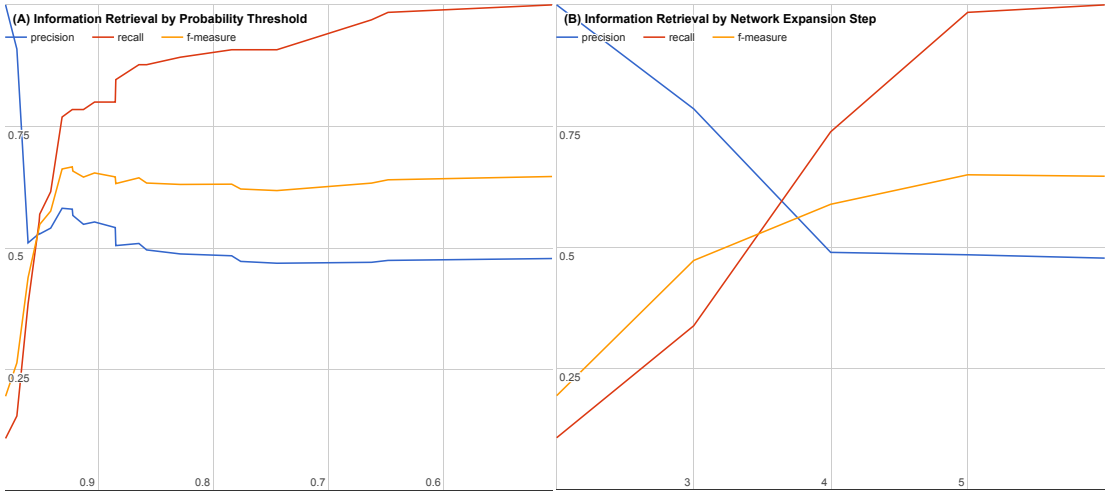

Figure 3. Precision, recall, and f-measure by (A) varying thresholds for joint probability and (B) varying number of interaction steps. Precision is the percentage of returned candidates that have been validated experimentally or have been in a clinical trial (a "hit") versus all candidates returned. Recall is the percentage of all known validated "hits". F-measure is the geometric mean of precision and recall that provides a balanced evaluation of the quality and completeness of the results.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

3.1 Candidate Significance

Three drugs were identified that have not previously been studied for melanoma treatment. Framycetin, a CXCR4 inhibitor, has not previously been considered for melanoma treatment. While it is nephrotoxic when administered orally (Greenberg, 1965), is used topically as an antibacterial treatment. While it may not be of use for metastasis, it might serve as a simple, inexpensive prophylactic treatment after excision of primary tumors. Additionally, Lucanthone and Podofilox were identified as having potential effects on melanoma through CDKN2A and MAP kinase, respectively.

One drug we identified, Vemurafenib, is approved for treatment of late stage melanoma has been shown to inhibit the BRAF protein in BRAF-V600 mutant melanomas (Chapman et al., 2011). However, cells can become resistant to Vemurafenib, thereby leading to metastasis (Le et al., 2013).

A number of the drugs we identified are in clinical trials for treatment of melanoma. We identified BRAF-oriented drugs, Dabrafenib (Hauschild et al., 2012), Sorafenib (National Cancer Institute, 2005), and Regorafenib (Istituto Clinico Humanitas, 2015), that have been evaluated in clinical trials, but have not yet been approved. Zidovudine, or Azidothymidine (AZT) is a TERT inhibitor that has shown significant melanoma tumor reductions in mouse models (Humer et al., 2008). Three MAP kinase-related compounds, Vinblastine (Luikart et al., 1984), Trametinib (Kim et al., 2012), Vinorelbine (Whitehead et al., 2004) were identified that are in clinical trials for melanoma treatment. CDKN2A was another popular target, as Irinotecan (Fiorentini et al., 2009) Topotecan (Kraut et al., 1997) Sodium stibogluconate (Naing, 2011) are all drugs in clinical trial that we identified as potential therapies.

Many other drugs were identified that are being studied in the lab. Additional drugs were identified that target the MAP kinase pathway, including Bosutinib (Homsi et al., 2009), Purvalanol (Smalley et al., 2007), Colchicine (Lemontt et al., 1988) Vincristine (Sawada et al., 2004). Podofilox has not yet been investigated in melanoma treatments, but preliminary investigations have focused on treating Chronic Lymphocytic Leukemia (CLL) (Shen et al., 2013) and Non-Small Cell Lung Cancer (NSCLC) (Peng et al., 2014). Since these drugs attack MAPK2 and related proteins rather than BRAF or NRAS, they can potentially synergize with other treatments (Homsi et al., 2009). Bosutinib in particular has been investigated as a synergistic treatment for melanoma (Held et al., 2012). Another possible treatment pathway is CXCR4 inhibition. Mouse models suggest that CXCR4 inhibitors like Plerixafor can reduce tumor metastasis and primary tumor growth (D'Alterio et al., 2012). We identify both Plerixafor and Framycetin (Neomycin B) as useful CXCR4 inhibitors. Two PKRCA activators, Ingenol Mebutate and Ellagic Acid, were also identified. PKRCA binds with BRAF (Pardo et al., 2006), but it is mechanistically unclear how PKRCA activation would result in treatment of melanoma. A number of other therapies are also notable. Purvalenol can inhibit GSK3β, which in turn activates TP53. Some, but not all, melanomas have TP53 deactivation (Smalley et al., 2007). Nadroparin, a MYC inhibitor, may inhibit tumor progression (Nagy et al., 2009). More broadly, heparins can potentially inhibit the metastatic process in melanoma and other cancers (Maraveyas et al., 2010).

The approach that we present here offers a novel, mechanism-focused exploration to identify and examine drugs and targets related to cancer. This approach filters our noisy or poorly supported parts of the knowledge graph to identify more confident mechanisms between drugs, targets and diseases. Thus, our approach can be used to explore high confidence associations that are produced as a result of large scale computational screens that use network connectivity (Yang and Agarwal, 2011; Wu et al., 2013b; Cheng et al., 2012; Emig et al., 2013; Harrold et al., 2013; Wu et al., 2013a; Vogt et al., 2014), the complementarity in drug-disease gene expression, and the similarity of chemical fingerprints, side-effects, targets, or indications (Yang and Agarwal, 2011; Ye et al., 2014; Chiang and Butte, 2009; Gottlieb et al., 2011; Lamb et al., 2006; Sirota et al., 2011). Importantly, since we focus on protein networks that are strongly linked with diseases, we believe that our mechanism focused approach will also aid in the identification of disease-modifying drug candidates, rather than solely those that would be useful for the treatment of symptomatic phenotypes or related co-morbid conditions.

3.2 Architecture

ReDrugS uses a fairly straightforward web architecture, as shown in Figure 4. It uses the Blazegraph RDF database backend. The database layer is interchangeable except that the full text search service needs to use Blazegraph-only properties to perform text searches as text indexing is not yet standardized in the SPARQL query language. All other aspects are standardized and should work with other RDF databases without modification. ReDrugs currently uses the Python-based TurboGears web application

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

framework hosted using the Web Services Gateway Interface (WSGI) standard via an Apache HTTP server. TurboGears in turn hosts the SADI web services that drive the application and access the database. It also serves up the static HTML and supporting files.

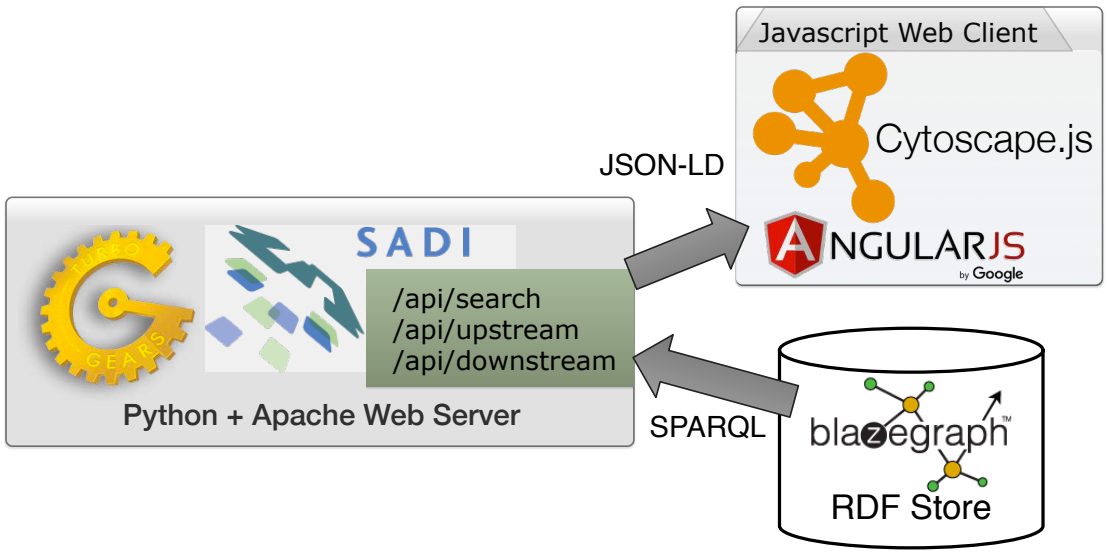

Figure 4. The ReDrugS software architecture. Using web standards and a three layer architecture (RDF store, web server, and rich web client), we were able to build a complete knowledge graph analysis platform.

The user interface is implemented with AngularJS and Cytoscape.js, which submits queries to the SADI web services using JSON-LD and aggregates results into the networked view. The software relies exclusively on standardized protocols (HTTP, SADI, SPARQL, RDF, and others) to make it simple to replace technologies as needed. The data itself is processed using conversion scripts as shown in Figure 6.

We have also adapted and featured ReDrugS in an immersive visualization laboratory called the Collaborative-Research Augmented Immersive Virtual Environment (CRAIVE) Lab at RPI, as shown in Figure 5. The goal of the demonstration was to explore new ways to visualize, sonify, and interact with big data in large-scale virtual reality systems. We also leveraged a gesture controller (Microsoft kinect) to interact with the visualization. With the 360-degree projection, multiple people can explore the visualization concurrently, which accelerates the exploration and discovery speed.

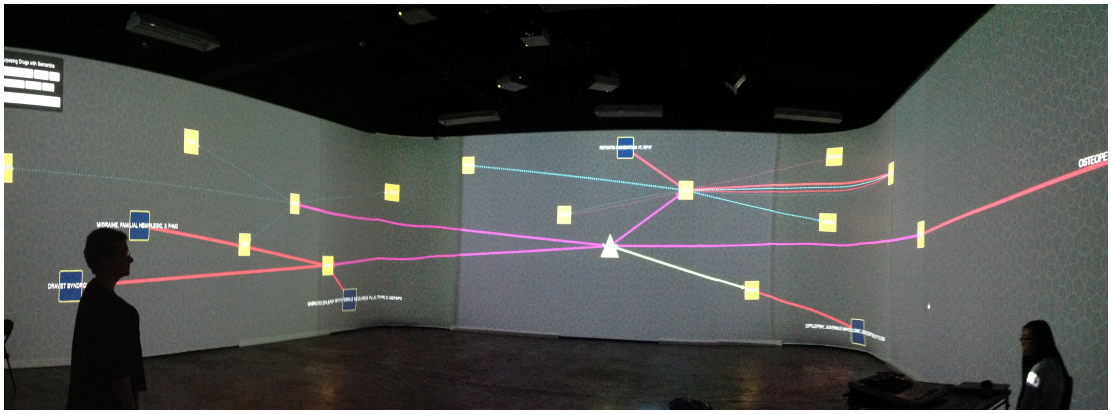

Figure 5. The authors demonstrate the ReDrugS user interface in the Collaborative-Research Augmented Immersive Virtual Environment (CRAIVE) Lab at RPI.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

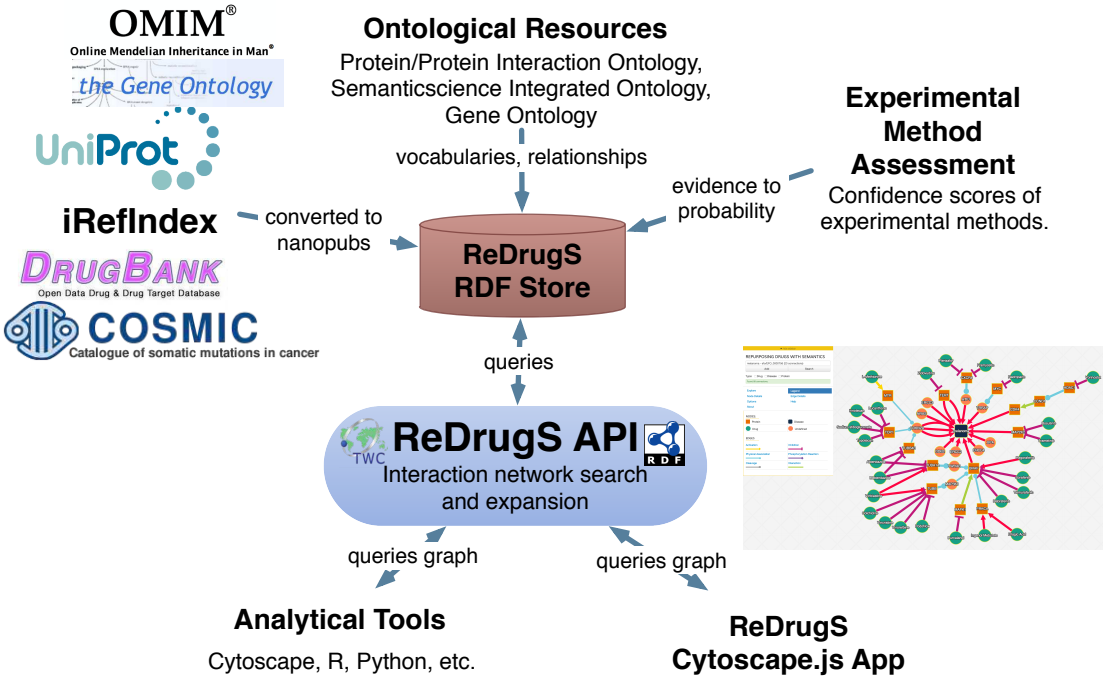

Figure 6. The ReDrugS data flow. Data is selected from external databases and converted using scripts into nanopublication graphs, which are loaded into the ReDrugS data store. This is combined with experimental method assessments, expressed in OWL, and public ontologies into the RDF store. The web service layer queries the store and produces aggregate analyses of those nanopublications, which is consumed and displayed by the rich web client. The same APIs can be used by other tools for further analysis.

3.3 Limitations and Future Work

Our study has a some limitations. First, our study is limited by the sources of data used. We used 3 databases (DrugBank, iRefIndex, and OMIM) to construct the initial knowledge graph. These databases are continuously changing and necessarily incomplete with respect to the total number of drugs, targets, protein interactions, diseases, and disease genes. For instance, as of 8/15/2016 there are over 2000 additional FDA approved drugs in DrugBank than in the version that was initially used. Second, the focus of our work is on the potential repositioning of FDA approved drugs, which means that tens of thousands of chemical compounds with protein binding activity cannot be considered as candidates in the current study. Third, our path expansion is currently limited to pairwise protein-protein interactions, which excludes interactions as a result of protein complexes or regulatory pathways. Having a more sophisticated understanding of non direct interactions will help identify candidate drugs that can regulate entire pathways in a more rational manner. Additionally, we aim to incorporate knowledge of the complementarity of drug and disease gene expression patterns as evidenced by the Connectivity Map (Lamb et al., 2006), which could suggest therapeutic and adverse interactions. Finally, as we develop new hypotheses about potential new drug effects, we plan to test them using a new three-dimensional cellular microarray to perform high throughput drug screening (Lee et al., 2008) with reference samples. The integration of computational predictions and high throughput screening platform will enable the systematic evaluation of any drug or mechanism of action against any disease or adverse event.

4 MATERIALS AND METHODS

This research project did not involve human subjects. The ReDrugS platform consists of a graphical web application, an application programming interface (API), and a knowledge base. The graphical web application enables users to initiate a search using drug, gene, and disease names and synonyms. Users can then interact with the application to expand the network at an arbitrary number of interactions away from

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

the entity of interest, and to filter the network based on a joint probability between the source and target entities. Drug-protein, protein-protein, and gene-disease interactions were obtained from several datasets and integrated into ontology-annotated and provenance and evidence bearing representations called nanopublications. The web application obtains information from the knowledge base using semantic web services. Finally, we evaluated our approach by examining the mechanistic plausibility of the drug in having melanoma-specific disease modifying ability. We evaluated a large number of possible drug/disease associations with varying joint probabilities and interaction steps to determine the thresholds with the highest F-Measure, resulting in our thresholds of three or less interactions and a joint probability of 0.93 or higher.

Using the ReDrugS application page[1], we initiate our search for "melanoma", and select the first suggestion obtained from the Experimental Factor Ontology (EFO).[2] The application then provides immediate neighborhood of drugs and genes that are associated with melanoma. We expanded the network by first selecting the melanoma node and expanding the link distance to $|I| \leq 3$ and the changing the minimum joint probability to $p \geq 0.93$ in the search options. Importantly, we also limit the node type to "Drug". Finally, we click on the "find incoming links" button (two left-facing arrows). When finished the network will show all drugs interacting with melanoma that meet the above criteria, as well as any intervening entities and their interactions. The resulting network can be downloaded as an image, or a summary CSV file. We used the CSV file to validate the links by searching Google Scholar and ClinicalTrials.gov for each proposed drug/disease combination. We consider a "hit" to be a pairing with a published positive experiment *in vivo* or *in vitro* or any pairing that has been tested in a clinical trial. While this level of validation does not guarantee efficacy, it does determine if the resulting connection is a plausible hypothesis that might be tested.

4.1 Data Fusion

We developed a structured knowledge base containing data pertaining to drugs, targets, interactions, and diseases. We used five data sources: iRefIndex (Razick et al., 2008) DrugBank (Wishart et al., 2006), UniProt Gene Ontology Annotations (GOA) (Camon et al., 2004), the Online Mendelian Inheritance in Man (OMIM) (Hamosh et al., 2005), and the COSMIC Gene Census (Futreal et al., 2004).

iRefIndex contains protein-protein interactions and protein complexes and is an amalgam of the Biomolecular Interaction Network Database (BIND) (Bader et al., 2003), BioGRID (Stark et al., 2006), the Comprehensive Resource of Mammalian protein complexes (CORUM) (Ruepp et al., 2008), Database of Interacting Proteins (DIP), (Xenarios et al., 2002), Human Protein Reference Database (HPRD), (Prasad et al., 2009), InnateDB (Lynn et al., 2008), IntAct (Kerrien et al., 2012), MatrixDB (Chautard et al., 2011), Molecular INTeraction database (MINT) (Chatr-aryamontri et al., 2008), MPact (Güldener et al., 2006), microbial protein interaction database (MPIDB) (Goll et al., 2008), MIPS mammalian protein-protein interaction database (MPPI) (Pagel et al., 2005), and Online Predicted Human Interaction Database (OPHID) (Brown and Jurisica, 2005). DrugBank provides information about experimental/approved drugs and their targets, and UniProt GOA describes proteins in terms of their biological processes, cellular locations, and molecular functions. OMIM provides associations between genes and inherited or genetically-driven diseases. The COSMIC Gene Census is a curated list of genes that have causal associations with one or more cancer types.

Each association (e.g. drug-target, protein-protein, disease-gene) was captured using the nanopublication (Groth et al., 2010) scheme. A nanopublication is a digital artifact that consists of an assertion, its provenance, and information about the digital publication. Our nanopublications are represented as Linked Data: Each data item is identified using an dereferenceable HTTP Uniform Resource Identifier (URI) and statements are represented using the Resource Description Framework (RDF). Each nanopublication corresponds to a single interaction assertion from one of the databases. We used a number of automated scripts to produce the nanopublications and load them into the SPARQL endpoint. An example nanopublication is shown in Figure 7. We used the Semanticscience Integrated Ontology (SIO) (Dumontier et al., 2013) as a global schema to describe the nature and components of the associations, and coupled this with the PSI-MI Ontology (Hermjakob et al., 2004) to denote the types of interactions. We used the World Wide Web Consortium's Provenance Ontology (PROV-O) (Lebo et al., 2013) to capture provenance of the assertion (which data source it originated from). We loaded our nanopublications into

[1] http://redrugs.tw.rpi.edu
[2] http://www.ebi.ac.uk/efo/EFO_0000756

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Blazegraph, an RDF nanopublication compatible database. The data is accessed using its native SPARQL endpoint by the web application.

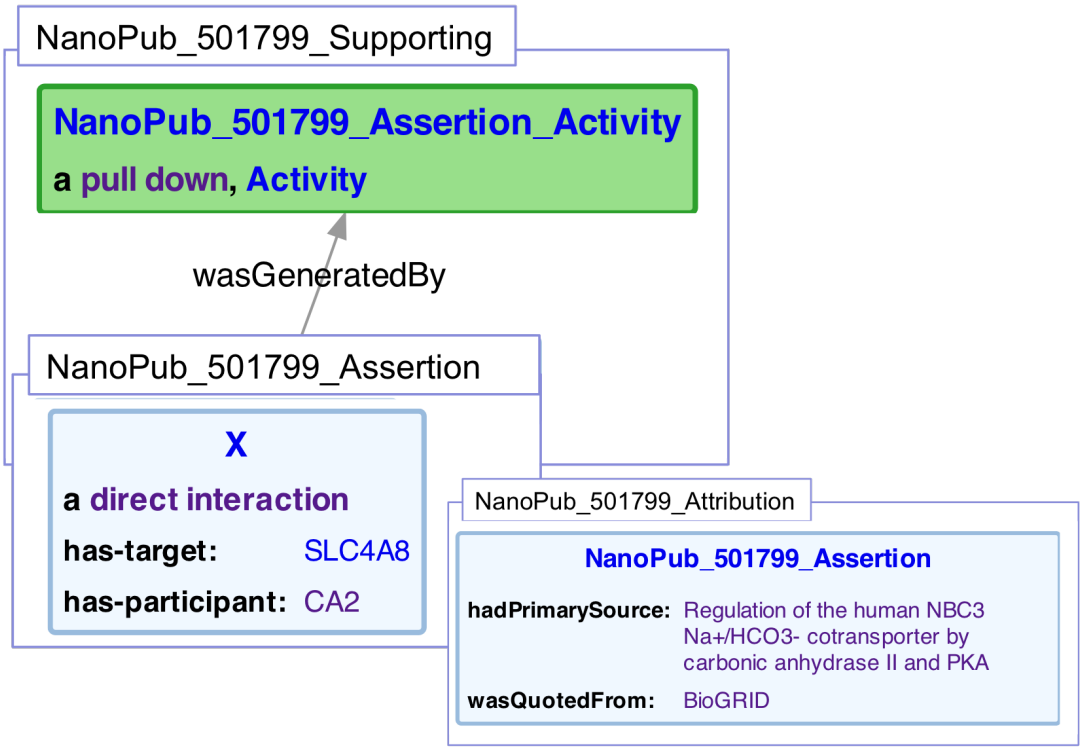

Figure 7. Representation of a protein/protein interaction within a nanopublication. Three graphs are represented. The assertion graph (NanoPub_501799_Assertion), states that an interaction (X) is of type *sio:DirectInteraction*, and has the target of SLC4A8, and a participant of CA2. The supporting graph (NanoPub_501799_Supporting), states that the assertion graph was generated by a pull down experiment (one of many encoded experiment types used in , a subclass of *prov:Activity*. The attribution graph (NanoPub_501799_Attribution), in turn, states that the assertion had a primary source of (Loiselle et al., 2004) and that the interaction was quoted from BioGrid.

4.2 Assertion Probability

Each knowledge graph fragment, enclosed in a nanopublication, is assigned a probability based on the quality of the methods used to create the assertions in the fragment. We compute probabilities based on two different methods. Manually curated assertions, from DrugBank, OMIM, and COSMIC Gene Census, are directly given a probability $p = 0.999$. Assertions that have been derived from a specific experimental method are given probabilities appropriate for that method. These probabilities are derived from a expert-driven measure of the reliability of the experimental method used to derive the association. Factors involved in the assessment of confidence include the degree of indirection in the assay, the sensitivity and specificity of the approach, and reproducibility of results under different conditions based on the comparative analyses of techniques (Obenauer and Yaffe, 2004; Sprinzak et al., 2003). Two expert bioinformaticians rated the reliability of each method and assigned a score of 1-3, where 1 corresponds to low confidence and 3 high confidence. After their initial assessment, they conferred on their reasoning for each score to resolve differences where possible. The experts considered level 1 to correspond to weak evidence that needs independent verification. Level 2 methods are generally reliable, but should have additional biological evidence. Level 3 methods are high quality method that produces few false positives. We calculated inter-annotator agreement between the two annotators over the three categories using Scott's Pi. Scott's Pi is similar to Cohen's kappa in that it improves on simple observed agreement by factoring in the extent of agreement that might be expected by chance. We determined the agreement to

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Service Name	Description	URL	Input	Output
Resource text tearch	Look up resources using free text search against their RDFS labels. This service is optimized for typeahead user interfaces.	search	*pml:Query*	*pml:AnsweredQuery*
Find interactions in a biological process	Find interactions whose participants or targets also participate in the input process.	process	*sio:Process*	*sio:Process*
Find upstream participants	Find interactions that the input entity is a target of in and have explicit participants.	upstream	*sio:MaterialEntity*	*sio:Target*
Find downstream targets	Find interactions that the input entity participates in and have explicit targets.	downstream	*sio:MaterialEntity*	*sio:Agent*

Table 2. The API endpoint prefix is `http://redrugs.tw.rpi.edu/api/`.

be 0.56 (Scott's Pi value of 0.26) across 104 experimental methods comprising of 99.9999% of interaction annotations (SCOTT, 1955).

The scores of 1, 2, and 3 were then assigned provisional probabilities of $p = 0.8$, $p = 0.95$, and $p = 0.99$ respectively. We chose these probabilities as approximations of the conceptual levels of probability for each rating by the experts, and feel that those probabilities correspond to how often an experiment at that confidence level can be expected to be accurate. We plan to provide a more rigorous assessment of the accuracy of each method against gold standards in future work. These confidence values were encoded into an OWL ontology along with the evidence codes. The full inferences were extracted using Pellet [3] and loaded into the SPARQL endpoint, where they were used to apply the probabilities to each assertion in the knowledge graph that had experimental evidence.

4.3 Semantic Web Services

We developed four Semantic Automated Discovery and Integration (SADI) web services (Wilkinson et al., 2009) in Python[4] to support easy access to the nanopubications (see Table 2) in ReDrugS. The four services are enumerated in Table 2.

The first service is a simple free text lookup, that takes an *pml:Query*[5] McGuinness et al. (2007) with a *prov:value* as a query and produces a set of entities whose labels contain the substring. This is used for interactive typeahead completion of search terms so users can look up URIs and entities without needing to know the details.

The other three SADI services look up interactions that contain a named entity. Two of them look at the entity to find upstream and downstream connections, and the third service assumes that the entity is a biological process and finds all interactions that related to that process. The services return only one interaction for each triple (source, interaction type, target). There are often multiple probabilities per interaction, and more than one interaction per interaction type. This is because the interaction may have been recorded in multiple databases, based on different experimental methods. To provide a single probability score for each interaction of a source and target, the interactions are combined. A single probability is generated per identified interaction by taking the geometric mean of the probabilities for that interaction. However, this method is undesirable when combining multiple interaction records of the same type. We instead combine the interaction records using a form of probabilistic voting using composite

[3] `https://github.com/complexible/pellet`

[4] For further information on developing web services in Python using SADI, see this tutorial: https://github.com/markwilkinson/SADI-Semantic-Web-Services-Core/wiki/Building-Services-in-Python

[5] PML 3, in development: https://github.com/timrdf/pml. This includes PML 2 constructs that are not covered in PROV-O.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Z-Scores. This is done to model that multiple experiments that produce the same results reinforce each other, and should therefore give a higher overall probability than would be indicated by taking their mean or even by Bayes Theorem. We do this by converting each probability into a Z Score (aka Standard Score) using the Quantile Function ($Q()$), summing the values, and applying the Cumulative Distribution Function ($CDF()$) to compute the corresponding probability:

$$P(x_{1...n}) = CDF\left(\sum_{i=1}^{n} Q(P(x_i))\right)$$

These composite Z Scores, which we transform back into probabilities, are frequently used to combine multiple indicators of the same underlying phenomena, as in (Moller et al., 1998). It has a drawback, however. One concern is that the strategy does not account for multiple databases recording the same non-independent experiment. This can possibly inflating the probabilities of interactions described by experiments that are published in more than one database.

4.4 Graph Expansion Using Joint Probability

In order to compute the probability that a given entity affects another, we compute the joint probability that each of the intervening interactions are true. Joint probability is the probability that every assertion in the set is true. This is computed by taking the product of probabilities of each interaction:

$$P(x_1 \wedge ... \wedge x_n) = \prod_{i=1}^{n} P(x_i)$$

This joint probability is used as a threshold that users can set to stop graph expansion. We also provide expansion limits using the number of interaction steps that are needed to connect the two entities.

4.5 User Interface

The user interface was developed using the above SADI web services and uses Cytoscape.js,[6], angular.js,[7] and Bootstrap 3.[8] An example network is shown in Figure 2 Users can search for biological entities and processes, which can then be autocompleted to specific entities that are in the ReDrugS graph. Users can then add those entities and processes to the displayed graph and retrieve upstream and downstream connections and link out to more details for every entity. Cytoscape.js is used as the main rendering and network visualization tool, and provides node and edge rendering, layout, and network analysis capabilities, and has been integrated into a customized rich web client.

In order to evaluate this knowledge graph, we developed a demonstration web interface[9] based on the Cytoscape.js[10] JavaScript library. The interface lets users enter biological entity names. As the user types, the text is resolved to a list of entities. The user finishes by selecting from the list, and submitting the search. The search returns interactions and nodes associated with the entity selected, which are added to the Cytoscape.js graph. Users are also able to select nodes and populate upstream or downstream connections. Figure 2 is an example output of this process.

5 ACKNOWLEDGEMENTS

A special thanks to Pascale Gaudet, who, with Michel Dumontier, evaluated the experimental methods and evidence codes listed in the Protein/Protein Interaction Ontology and Gene Ontology. Thank you also to Kusum Solanki and John Erickson for evaluation, feedback, and planning in the initial stages of this project.

[6] http://cytoscape.github.io/cytoscape.js
[7] https://angularjs.org
[8] http://getbootstrap.com
[9] `http://redrugs.tw.rpi.edu`
[10] `http://cytoscape.github.io/cytoscape.js`

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

REFERENCES

Bader, G. D., Betel, D., and Hogue, C. W. (2003). BIND: the biomolecular interaction network database. *Nucleic acids research*, 31(1):248–250.

Barth, A., Wanek, L., and Morton, D. (1995). Prognostic factors in 1,521 melanoma patients with distant metastases. *J Am Coll Surg*, 181:193–201.

Bisgin, H., Liu, Z., Fang, H., Kelly, R., Xu, X., and Tong, W. (2014). A phenome-guided drug repositioning through a latent variable model. *BMC bioinformatics*, 15:267.

Bisgin, H., Liu, Z., Kelly, R., Fang, H., Xu, X., and Tong, W. (2012). Investigating drug repositioning opportunities in FDA drug labels through topic modeling. *BMC bioinformatics*, 13 Suppl 1:S6.

Brown, K. R. and Jurisica, I. (2005). Online predicted human interaction database. *Bioinformatics*, 21(9):2076–2082.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., and Apweiler, R. (2004). The Gene Ontology annotation (GOA) database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic acids research*, 32(suppl 1):D262–D266.

Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., Hogg, D., Lorigan, P., Lebbe, C., Jouary, T., Schadendorf, D., Ribas, A., O'Day, S. J., Sosman, J. A., Kirkwood, J. M., Eggermont, A. M., Dreno, B., Nolop, K., Li, J., Nelson, B., Hou, J., Lee, R. J., Flaherty, K. T., and McArthur, G. A. (2011). Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine*, 364(26):2507–2516.

Chatr-aryamontri, A., Zanzoni, A., Ceol, A., and Cesareni, G. (2008). Searching the protein interaction space through the MINT database. In *Functional Proteomics*, pages 305–317. Springer.

Chautard, E., Fatoux-Ardore, M., Ballut, L., Thierry-Mieg, N., and Ricard-Blum, S. (2011). MatrixDB, the extracellular matrix interaction database. *Nucleic acids research*, 39(suppl 1):D235–D240.

Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., Zhou, W., Huang, J., and Tang, Y. (2012). Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS computational biology*, 8(5):e1002503.

Chiang, A. P. and Butte, A. J. (2009). Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. *Clinical pharmacology and therapeutics*, 86:507–510.

Clavo, A. and Wahl, R. (1996). Effects of hypoxia on the uptake of tritiated thymidine, L-leucine, L-methionine and FDG in cultured cancer cells. *J Nucl Med*, 37:502–6.

D'Alterio, C., Barbieri, A., Portella, L., Palma, G., Polimeno, M., Riccio, A., Ieranò, C., Franco, R., Scognamiglio, G., Bryce, J., Luciano, A., Rea, D., Arra, C., and Scala, S. (2012). Inhibition of stromal CXCR4 impairs development of lung metastases. *Cancer Immunology Immunotherapy*, 61(10):1713–1720.

Doudican, N., Rodriguez, A., Osman, I., and Orlow, S. J. (2008). Mebendazole Induces Apoptosis via Bcl-2 Inactivation in Chemoresistant Melanoma Cells. *Molecular Cancer Research*, 6(8):1308–1315.

Dumontier, M., Baker, C. J., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., Del Rio, N. R., Duck, G., Furlong, L. I., Keath, N., Klassen, D., McCusker, J. P., Queralt-Rosinach, N., Samwald, M., Villanueva-Rosales, N., Wilkinson, M. D., and Hoehndorf, R. (2013). The Semanticscience Integrated Ontology (SIO) for Biomedical Research and Knowledge Discovery. *Draft available at http://code.google. com/p/semanticscience/wiki/SIO*.

Emig, D., Ivliev, A., Pustovalova, O., Lancashire, L., Bureeva, S., Nikolsky, Y., and Bessarabova, M. (2013). Drug target prediction and repositioning using an integrated network-based approach. *PloS one*, 8(4):e60618.

Fiorentini, G., Aliberti, C., Del, C. A., Tilli, M., Rossi, S., Ballardini, P., Turrisi, G., and Benea, G. (2009). Intra-arterial hepatic chemoembolization (TACE) of liver metastases from ocular melanoma with slow-release irinotecan-eluting beads. Early results of a phase II clinical study. *In Vivo*, 23:131–7.

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M. R. (2004). A census of human cancer genes. *Nature Reviews Cancer*, 4(3):177–183.

Goll, J., Rajagopala, S. V., Shiau, S. C., Wu, H., Lamb, B. T., and Uetz, P. (2008). MPIDB: the microbial protein interaction database. *Bioinformatics*, 24(15):1743–1744.

Gottlieb, A., Stein, G., Ruppin, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol*, 7:496.

Greenberg, L. H. (1965). Audiotoxicity and Nephrotoxicity Due to Orally Administered Neomycin. *JAMA*, 194(7):827.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Groth, P., Gibson, A., and Velterop, J. (2010). The anatomy of a nanopublication. *Information Services and Use*, 30(1):51–56.

Group, W. O. W. et al. (2009). OWL 2 Web Ontology Language Document Overview.

Grover, M. P., Ballouz, S., Mohanasundaram, K. A., George, R. A., Sherman, C. D. H., Crowley, T. M., and Wouters, M. A. (2014). Identification of novel therapeutics for complex diseases from genome-wide association data. *BMC medical genomics*, 7 Suppl 1:S8.

Güldener, U., Münsterkötter, M., Oesterheld, M., Pagel, P., Ruepp, A., Mewes, H.-W., and Stümpflen, V. (2006). MPact: the MIPS protein interaction resource on yeast. *Nucleic acids research*, 34(suppl 1):D436–D441.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., and McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, 33(suppl 1):D514–D517.

Harris, S., Seaborne, A., and Prud'hommeaux, E. (2013). SPARQL 1.1 query language. *W3C Recommendation*, 21.

Harrold, J. M., Ramanathan, M., and Mager, D. E. (2013). Network-based approaches in drug discovery and early development. *Clinical pharmacology and therapeutics*, 94(6):651–8.

Hauschild, A., Grob, J.-J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., Rutkowski, P., Blank, C. U., Miller, W. H., Kaempgen, E., Martín-Algarra, S., Karaszewska, B., Mauch, C., Chiarion-Sileni, V., Martin, A.-M., Swann, S., Haney, P., Mirakhur, B., Guckert, M. E., Goodman, V., and Chapman, P. B. (2012). Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre open-label, phase 3 randomised controlled trial. *The Lancet*, 380(9839):358–365.

Held, M. A., Langdon, C. G., Platt, J. T., Graham-Steed, T., Liu, Z., Chakraborty, A., Bacchiocchi, A., Koo, A., Haskins, J. W., Bosenberg, M. W., and Stern, D. F. (2012). Genotype-Selective Combination Therapies for Melanoma Identified by High-Throughput Drug Screening. *Cancer Discovery*, 3(1):52–67.

Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C., et al. (2004). The hupo psi's molecular interaction format—a community standard for the representation of protein interaction data. *Nature biotechnology*, 22(2):177–183.

Homsi, J., Cubitt, C. L., Zhang, S., Munster, P. N., Yu, H., Sullivan, D. M., Jove, R., Messina, J. L., and Daud, A. I. (2009). Src activation in melanoma and Src inhibitors as therapeutic agents in melanoma. *Melanoma Research*, 19(3):167–175.

Humer, J., Ferko, B., Waltenberger, A., Rapberger, R., Pehamberger, H., and Muster, T. (2008). Azidothymidine inhibits melanoma cell growth in vitro and in vivo. *Melanoma Research*, 18(5):314–321.

Istituto Clinico Humanitas (2015). Regorafenib in Patients With Metastatic Solid Tumors Who Have Progressed After Standard Therapy (RESOUND). Accessed 2016-1-10.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction database in 2012. *Nucleic acids research*, 40(D1):D841–D846.

Kim, K. B., Kefford, R., Pavlick, A. C., Infante, J. R., Ribas, A., Sosman, J. A., Fecher, L. A., Millward, M., McArthur, G. A., Hwu, P., Gonzalez, R., Ott, P. A., Long, G. V., Gardner, O. S., Ouellet, D., Xu, Y., DeMarini, D. J., Le, N. T., Patel, K., and Lewis, K. D. (2012). Phase II Study of the MEK1/MEK2 Inhibitor Trametinib in Patients With Metastatic BRAF-Mutant Cutaneous Melanoma Previously Treated With or Without a BRAF Inhibitor. *Journal of Clinical Oncology*, 31(4):482–489.

Kim, S., Liu, Y., Gaber, M. W., Bumgardner, J. D., Haggard, W. O., and Yang, Y. (2008). Development of chitosan-ellagic acid films as a local drug delivery system to induce apoptotic death of human melanoma cells. *Journal of Biomedical Materials Research Part B: Applied Biomaterials*, 90B(1):145–155.

Kingsmore, S. F., Lindquist, I. E., Mudge, J., Gessler, D. D., and Beavis, W. D. (2008). Genome-wide association studies: progress and potential for drug discovery and development. *Nature reviews. Drug discovery*, 7(3):221–30.

Klyne, G. and Carroll, J. J. (2005). Resource Description Framework (RDF): Concepts and Abstract Syntax. *W3C Recommendation*.

Kraut, E. H., Walker, M. J., Staubus, A., Gochnour, D., and Balcerzak, S. P. (1997). Phase II Trial of Topotecan in Malignant Melanoma. *Cancer Investigation*, 15(4):318–320.

Krauthammer, M., Kong, Y., Bacchiocchi, A., Evans, P., Pornputtapong, N., Wu, C., McCusker, J., Ma,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

S., Cheng, E., Straub, R., Serin, M., Bosenberg, M., Ariyan, S., Narayan, D., Sznol, M., Kluger, H., Mane, S., Schlessinger, J., Lifton, R., and Halaban, R. (2015). Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nat Genet*, 47:996–1002.

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *science*, 313(5795):1929–1935.

Le, K., Blomain, E. S., Rodeck, U., and Aplin, A. E. (2013). Selective RAF inhibitor impairs ERK1/2 phosphorylation and growth in mutant NRAS vemurafenib-resistant melanoma cells. *Pigment Cell Melanoma Res*, 26(4):509–517.

Lebo, T., Sahoo, S., and McGuinness, D. (2013). PROV-O: The PROV Ontology. `http://www.w3.org/TR/prov-o/`.

Lee, M.-Y., Kumar, R. A., Sukumaran, S. M., Hogg, M. G., Clark, D. S., and Dordick, J. S. (2008). Three-dimensional cellular microarray for high-throughput toxicology assays. *Proceedings of the National Academy of Sciences*, 105(1):59–63.

Lemontt, J., Azzaria, M., and Gros, P. (1988). Increased mdr gene expression and decreased drug accumulation in multidrug-resistant human melanoma cells. *Cancer Res*, 48:6348–53.

Loiselle, F. B., Morgan, P. E., Alvarez, B. V., and Casey, J. R. (2004). Regulation of the human nbc3 na+/hco3- cotransporter by carbonic anhydrase ii and pka. *American Journal of Physiology-Cell Physiology*, 286(6):C1423–C1433.

Luikart, S., Kennealey, G., and Kirkwood, J. (1984). Randomized phase III trial of vinblastine, bleomycin, and cis-dichlorodiammine-platinum versus dacarbazine in malignant melanoma. *J Clin Oncol*, 2:164–8.

Lynn, D. J., Winsor, G. L., Chan, C., Richard, N., Laird, M. R., Barsky, A., Gardy, J. L., Roche, F. M., Chan, T. H., Shah, N., et al. (2008). InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Molecular systems biology*, 4(1).

Mansuy, M., Nikkels-Tassoudji, N., Arrese, J. E., Rorive, A., and Nikkels, A. F. (2014). Recurrent In Situ Melanoma Successfully Treated with Ingenol Mebutate. *Dermatol Ther (Heidelb)*, 4(1):131–135.

Maraveyas, A., Johnson, M. J., Xiao, Y. P., and Noble, S. (2010). Malignant melanoma as a target malignancy for the study of the anti-metastatic properties of the heparins. *Cancer Metastasis Rev*, 29(4):777–784.

McGuinness, D. L., Ding, L., Silva, P. P. D., and Chang, C. (2007). PML 2: A Modular Explanation Interlingua. In *Proceedings of the AAAI 2007 Workshop on Explanation-Aware Computing*, pages 22 – 23.

Moller, J., Cluitmans, P., Rasmussen, L., Houx, P., Rasmussen, H., Canet, J., Rabbitt, P., Jolles, J., Larsen, K., Hanning, C., Langeron, O., Johnson, T., Lauven, P., Kristensen, P., Biedler, A., van Beem, H., Fraidakis, O., Silverstein, J., Beneken, J., and JS, G. (1998). Long-term postoperative cognitive dysfunction in the elderly: ISPOCD1 study. *The Lancet*, 351(9106):857–861.

Nagy, Z., Turcsik, V., and Blaskó, G. (2009). The Effect of LMWH (Nadroparin) on Tumor Progression. *Pathology & Oncology Research*, 15(4):689–692.

Naing, A. (2011). Phase I Dose Escalation Study of Sodium Stibogluconate (SSG) a Protein Tyrosine Phosphatase Inhibitor, Combined with Interferon Alpha for Patients with Solid Tumors. *J. Cancer*, page 81.

National Cancer Institute (2005). Carboplatin and Paclitaxel With or Without Sorafenib Tosylate in Treating Patients With Stage III or Stage IV Melanoma That Cannot Be Removed by Surgery. Accessed 2016-1-10.

Obenauer, J. C. and Yaffe, M. B. (2004). Computational prediction of protein-protein interactions. In *Protein-Protein Interactions*, pages 445–467. Springer.

Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Mark, P., Stümpflen, V., Mewes, H.-W., et al. (2005). The MIPS mammalian protein–protein interaction database. *Bioinformatics*, 21(6):832–834.

Pardo, O., Wellbrock, C., Khanzada, U., Aubert, M., Arozarena, I., Davidson, S., Bowen, F., Parker, P., Filonenko, V., Gout, I., Sebire, N., Marais, R., Downward, J., and Seckl, M. (2006). FGF-2 protects small cell lung cancer cells from apoptosis through a complex involving PKCepsilon, B-Raf and S6K2. *EMBO J*, 25:3078–88.

Patel, K., Doudican, N. A., Schiff, P. B., and Orlow, S. J. (2011). Albendazole sensitizes cancer cells to ionizing radiation. *Radiation Oncology*, 6(1):160.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ:

Peng, X., Wang, F., Li, L., Bum-Erdene, K., Xu, D., Wang, B., Sinn, A. A., Pollok, K. E., Sandusky, G. E., Li, L., Turchi, J. J., Jalal, S. I., and Meroueh, S. O. (2014). Exploring a structural protein–drug interactome for new therapeutics in lung cancer. *Mol. BioSyst.*, 10(3):581.

Prasad, T. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human protein reference database—2009 update. *Nucleic acids research*, 37(suppl 1):D767–D772.

Razick, S., Magklaras, G., and Donaldson, I. M. (2008). iRefIndex: a consolidated protein interaction database with provenance. *BMC bioinformatics*, 9(1):405.

Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegele, B., Schmidt, T., Doudieu, O. N., Stümpflen, V., et al. (2008). CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic acids research*, 36(suppl 1):D646–D650.

Sanseau, P. and Koehler, J. (2011). Editorial: computational methods for drug repurposing. *Briefings in bioinformatics*, 12(4):301–2.

Sawada, N., Kataoka, K., Kondo, K., Arimochi, H., Fujino, H., Takahashi, Y., Miyoshi, T., Kuwahara, T., Monden, Y., and Ohnishi, Y. (2004). Betulinic acid augments the inhibitory effects of vincristine on growth and lung metastasis of B16F10 melanoma cells in mice. *Br J Cancer*, 90(8):1672–1678.

SCOTT, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19(3):321–325.

Shen, M., Zhang, Y., Saba, N., Austin, C. P., Wiestner, A., and Auld, D. S. (2013). Identification of Therapeutic Candidates for Chronic Lymphocytic Leukemia from a Library of Approved Drugs. *PLoS ONE*, 8(9):e75252.

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., Sage, J., and Butte, A. J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine*, 3(96):96ra77.

Smalley, K. S., Contractor, R., Haass, N. K., Kulp, A. N., Atilla-Gokcumen, G. E., Williams, D. S., Bregman, H., Flaherty, K. T., Soengas, M. S., Meggers, E., and Herlyn, M. (2007). An Organometallic Protein Kinase Inhibitor Pharmacologically Activates p53 and Induces Apoptosis in Human Melanoma Cells. *Cancer Research*, 67(1):209–217.

Sprinzak, E., Sattath, S., and Margalit, H. (2003). How reliable are experimental protein–protein interaction data? *Journal of molecular biology*, 327(5):919–923.

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic acids research*, 34(suppl 1):D535–D539.

Vogt, I., Prinz, J., and Campillos, M. (2014). Molecularly and clinically related drugs and diseases are enriched in phenotypically similar drug-disease pairs. *Genome Medicine*, 6:52.

Whitehead, R. P., Moon, J., McCachren, S. S., Hersh, E. M., Samlowski, W. E., Beck, J. T., Tchekmedyian, N. S., and Sondak, V. K. (2004). A Phase II trial of vinorelbine tartrate in patients with disseminated malignant melanoma and one prior systemic therapy. *Cancer*, 100(8):1699–1704.

Wilkinson, M., Vandervalk, B., and McCarthy, L. (2009). SADI Semantic Web Services - 'cause you can't always GET what you want! In *2009 IEEE Asia-Pacific Services Computing Conference (APSCC)*. Institute of Electrical & Electronics Engineers (IEEE).

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*, 34(suppl 1):D668–D672.

Wu, C., Gudivada, R. C., Aronow, B. J., and Jegga, A. G. (2013a). Computational drug repositioning through heterogeneous network clustering. *BMC systems biology*, 7 Suppl 5:S6.

Wu, Z., Wang, Y., and Chen, L. (2013b). Network-based drug repositioning. *Molecular bioSystems*, 9(6):1268–81.

Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S.-M., and Eisenberg, D. (2002). DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic acids research*, 30(1):303–305.

Yang, L. and Agarwal, P. (2011). Systematic drug repositioning based on clinical side-effects. *PloS one*, 6(12):e28025.

Ye, H., Liu, Q., and Wei, J. (2014). Construction of drug network based on side effects and its application for drug repositioning. *PloS one*, 9(2):e87864.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2007v2 | CC BY 4.0 Open Access | rec: 8 Dec 2016, publ: