

A Natural Language Processing System for Extracting Evidence of Drug Repurposing from Scientific Publications

Shivashankar Subramanian,^{1,2} Ioana Baldini,¹ Sushma Ravichandran,¹
Dmitriy A. Katz-Rogozhnikov,¹ Karthikeyan Natesan Ramamurthy,¹ Prasanna Sattigeri,¹
Kush R. Varshney,¹ Annmarie Wang,^{3,4} Pradeep Mangalath,^{3,5} Laura B. Kleiman³

¹IBM Research, Yorktown Heights, NY, USA, ²University of Melbourne, VIC, Australia,

³Cures Within Reach for Cancer, Cambridge, MA, USA,

⁴Massachusetts Institute of Technology, Cambridge, MA, USA, ⁵Harvard Medical School, Boston, MA, USA

Abstract

More than 200 generic drugs approved by the U.S. Food and Drug Administration for non-cancer indications have shown promise for treating cancer. Due to their long history of safe patient use, low cost, and widespread availability, repurposing of these drugs represents a major opportunity to rapidly improve outcomes for cancer patients and reduce healthcare costs. In many cases, there is already evidence of efficacy for cancer, but trying to manually extract such evidence from the scientific literature is intractable. In this emerging applications paper, we introduce a system to automate non-cancer generic drug evidence extraction from PubMed abstracts. Our primary contribution is to define the natural language processing pipeline required to obtain such evidence, comprising the following modules: querying, filtering, cancer type entity extraction, therapeutic association classification, and study type classification. Using the subject matter expertise on our team, we create our own datasets for these specialized domain-specific tasks. We obtain promising performance in each of the modules by utilizing modern language processing techniques and plan to treat them as baseline approaches for future improvement of individual components.

1 Introduction

Sir James Black, the 1998 Nobel Prize laureate in physiology and medicine, famously said, “The most fruitful basis of the discovery of a new drug is to start with an old drug.” *Drug repurposing* is the process of identifying and validating a new therapeutic indication for an existing drug. A major benefit of drug repurposing is that it can lead to faster drug development and approval as well as safer and less costly treatments for patients (Hernandez et al. 2017).

This work focuses on repurposing non-cancer generic drugs for the treatment of cancer. Scientific publications, such as pre-clinical studies (in vitro and in vivo studies) and early-clinical studies (case reports, observational studies, and small clinical trials), contain evidence on these drugs being tested for cancer use. For example, the Repurposing Drugs in Oncology (ReDO) project, through manually searching research articles indexed by PubMed, found anti-cancer evidence for more than 200 non-cancer generic drugs

(Pantziarka et al. 2017; Bouche, Pantziarka, and Meheus 2017; Verbaanderd et al. 2017). However, manual review to identify and analyze potential evidence is time-consuming and intractable to scale, as PubMed indexes millions of articles and the collection is continuously updated. For example, the PubMed query *Neoplasms [mh]* yields more than 3.2 million articles.¹ Hence there is a need for a (semi-) automatic approach to identify relevant scientific articles and provide an easily consumable summary of the evidence.

Cures Within Reach for Cancer is developing the Oncology Repurposing Engine™, which synthesizes the plethora of scientific and real-world data on non-cancer generic drugs to identify the most promising therapies to repurpose for cancer. Drug repurposing is a complex process involving synthesizing and prioritizing evidence, conducting clinical trials and, upon success, influencing the standard of care. The focus of this paper is on the key evidence synthesis step.

Synthesizing scientific evidence for drug repurposing can be divided into three major categories: network-based, natural language processing (NLP) approaches, and semantic techniques (Xue et al. 2018). Network-based approaches aim to infer relationships between biological entities (drug–disease or drug–target relationships), inspired by the fact that biologic entities (disease, drug, protein, etc.) in the same module of biological networks share similar characteristics (Martínez et al. 2015). NLP approaches aim to both identify biological entities and mine new knowledge from scientific literature (Li, Zhu, and Chen 2009). Semantic approaches require a semantic network to be built first, which can be used with various approaches to mine relationships between entities (Palma, Vidal, and Raschid 2014).

We focus on NLP approaches in this paper. Our primary contributions are as follows:

1. Formulating the set of tasks required to identify appropriate evidence of non-cancer generic drug repurposing as a pipeline of individual NLP tasks

¹Neoplasm is a technical term for abnormal and excessive growth of tissue; cancers are a type of neoplasm. The [mh] in PubMed’s query syntax instructs the search to expand the query using Medical Subject Headings (MeSH), which are a taxonomic organization of medical terms.

2. Precisely specifying those tasks in terms of input and output (not an easy endeavor)
3. Creating domain-specific datasets that support the task definition
4. Designing models using modern techniques for each of the domain-specific tasks, along with their performance evaluation

In the remainder of the paper, we precisely formulate a set of NLP tasks on scientific papers for the purpose of evidence extraction and overview a solution architecture. We describe the dataset we have assembled using the subject matter expertise of our team. We provide details on our initial best attempts at techniques for the different NLP problems and present empirical results.

2 Problem Description

PubMed, provided by the National Center for Biotechnology Information (NCBI), is a comprehensive source of biomedical studies, comprising more than 30 million biomedical abstracts and citations from various sources such as MEDLINE, life science journals, and online books.

Given a list of generic drugs, the goal of our work is to automatically select from the large PubMed collection of abstracts, the articles that describe non-cancer generic drugs being tested for the treatment of cancer and measure cancer-relevant phenotypic outcomes.² Once relevant documents are identified, the next step is to correctly identify the type of drug and cancer studied, classify the therapeutic association between the two (e.g., positive, negative, inconclusive) and also the type of study that was performed (e.g., pre-clinical or clinical). We refer to this collection of information (i.e., drug, cancer, therapeutic association, study type) as the *evidence* discussed in the PubMed abstract.

Identifying such evidence from scientific abstracts is non-trivial for the following reasons:

Abstracts use domain-specific language: The articles that discuss cancer interventions use domain-specific jargon which makes the text hard to comprehend by both humans with non-expert background and machines that are not trained with domain-specific data. Hence a strong collaboration between domain-experts and data scientists is required to define machine learning tasks, collect and annotate the appropriate information and design and evaluate machine learning models that address the designed tasks. This is the main focus and contribution of our work.

Drug or cancer may not be the focus of the study: Even when an abstract contains given drug and cancer terms, it may turn out to be irrelevant because either the drug or the cancer are not the focus of the study. This can happen when

the drug and/or cancer are mentioned as background information, and not employed in the article's methods, objective or results. A subtle case is when the drug addresses side-effects of cancer treatment, e.g. to reduce the pain caused by chemotherapy, but does not actually treat cancer.

Studies may not evaluate phenotypic outcomes: Some articles describe interventions of non-cancer generic drugs on cancer, without discussing phenotypic outcomes. For example, the mechanism of action of the drug for cancer is analyzed, without any mention of the therapeutic association between the drug and the disease. Such articles are not of interest to our study and filtering them out from the rest of the articles is a challenging NLP task. In addition, the complexity of the task is exacerbated by the fact that there are several possible phenotypic outcomes (e.g., cell growth, apoptosis, angiogenesis, tumor progression, overall survival) that can be described in various ways.

Incomplete/unreliable metadata: While PubMed articles are tagged with MeSH terms and publication type, as these tags are manually created, they are error-prone (Marshall et al. 2018) or incomplete. In addition, current publication types do not include information on pre-clinical studies.

Once the relevant abstracts are identified, the therapeutic association is identified for a pair of drug and cancer type, similar to drug-disease relation extraction (Peng, Wei, and Lu 2016). This step involves the following challenges:

Need for identifying cancer types: Studies of a given drug may address various types of cancer (especially true in pre-clinical studies), and hence it is important to identify the cancer types that represent the focus of a particular study, e.g., *non-small cell lung cancer*.

Presence of multiple drugs and cancer types: An abstract can discuss more than one drug and/or cancer type and hence it is necessary to consider all the pairs of drug-cancer combinations for analysis (Peng, Wei, and Lu 2016).

Last but not least, for those relevant abstracts with drug-cancer therapeutic associations classified, it is important to predict the study types, which can be used to prioritize the drug-cancer pairs for further analysis. For example, results from clinical studies with patients carry more weight than pre-clinical laboratory studies.

2.1 Solution Overview

We propose an evidence discovery pipeline, shown in Figure 1, to achieve the goals described above. First we query PubMed using the provided search engine to narrow the collection of articles we analyze. Note that querying PubMed, even with a sophisticated search string as explained shortly, may not yield only *relevant* articles. Hence we have a (shallow) filtering stage to reject the easy irrelevant cases. Using the resulting abstracts, cancer types are identified using a named entity recognition (NER) model. With the abstract and pairs of drug-cancer types, therapeutic association is classified and also the type of study is categorized.

As PubMed collection is comprehensive, different projects, such as the Cochrane Crowd project, established querying strategies to narrow down the collection to a subset of interest. For example, the Cochrane Crowd project developed a Cochrane highly sensitive search (CHSS) strategy

²Phenotype is the observable physical properties of an organism; these include the organism's appearance, development, and behavior. We focus on phenotypic outcomes (such as proliferation/death of cells grown in culture or tumor progression/overall survival rates for clinical trials) since they are a more direct measure of outcomes that matter to cancer patients and represent stronger therapeutic evidence (as opposed to, for example, the effects of drugs on protein levels).

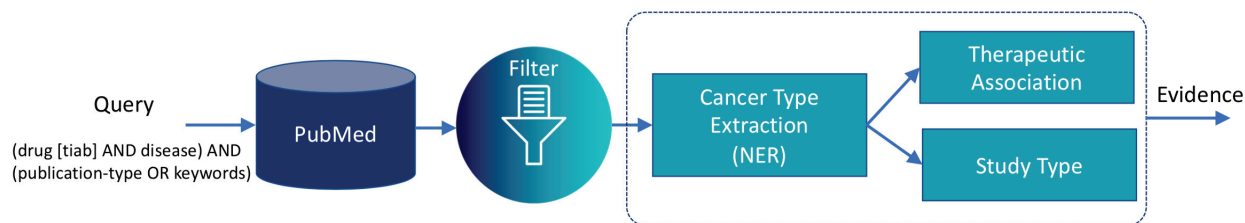


Figure 1: Solution overview: The proposed evidence discovery pipeline that takes a generic drug as input and provides pre- and early-clinical evidence of repurposing this drug for cancer treatment.

(Dickersin, Scherer, and Lefebvre 1994) to identify RCTs . In this work, we adapt the search strategy to our context. The PubMed query is detailed in Table 1. The additional filtering we perform is explained in Section 3. NER, therapeutic association and study type classification methods are discussed in Section 4. As not all the irrelevant articles are filtered with the shallow filter, we incorporate the irrelevant categories in the therapeutic associations class schema, as given below:

- Irrelevant:
 - Drug has no relation to cancer (cases where either the drug or the cancer is not the focus of the study).
 - Abstract does not discuss a *phenotypic outcome*.
- Relevant:
 - Effective: the drug was shown to be effective for treating the cancer.
 - Detrimental: the drug has a detrimental effect on the cancer.
 - No effect: the drug has no effect on the cancer.
 - Inconclusive: the results of the study are inconclusive.

The study types we consider are defined as follows:

- pre-clinical studies (in vitro, in vivo)
- observational studies (incl. case reports)
- clinical trials.

An example of a *relevant* abstract with cancer type, therapeutic association and study type annotation is given in Figure 2. The next section describes how we created an initial annotated dataset that enabled us to build machine learning models for the cancer type, therapeutic association and study type identification.

3 Dataset

Our team of machine learning scientists and biomedical scientists worked closely together to fine-tune the querying/filtering strategy and to annotate cancer types, along with the therapeutic associations and study types. We started by considering a subset of FDA-approved non-cancer generic drugs. We followed the sequence of steps described below to create our dataset. In the beginning, the process was highly iterative. As we learned and discovered more details on the articles of interest, the process became streamlined.

Retinoids can block cell proliferation and induce apoptosis in tumor cells. The antitumoral effect of synthetic retinoids like **Adapalene (ADA)** on **hepatoma** cells (HepG2, Hep1B) was investigated. **ADA** at 10(-4)M efficiently induced apoptosis, reaching 61.7% in HepG2 and 79.1% in Hep1B after 72 h incubation. This was accompanied by up-regulation of pro-apoptotic bax and caspase 3, while bcl-2 was down-regulated, shifting the bax/bcl-2 ratio to >2.3 in hepatoma cells. *ADA inhibits hepatoma cell growth in vitro and is a powerful inducer of hepatoma cell apoptosis.*

Figure 2: Sample *relevant* abstract annotation. PubMed # 15105045, *Adapalene* is the non-cancer generic drug, used to treat *hepatoma cancer*. It is a pre-clinical evidence (in vitro study), and has an *effective* association. Evidence for association with phenotypic outcome measured is italicized.

Querying for high recall: We curated a list of about 150 positive examples (largely extracted from the citations from the ReDO project) which represent abstracts that contain pre-/early-clinical evidence of non-cancer generic drugs used in cancer treatment. We developed a query, which is a variant of the CHSS strategy for fetching potentially relevant abstracts in the context of the non-cancer drug and cancer diseases specified in the query. Note in Table 1 the drug is specified while the cancer is not. We obtained a recall of over 97% on the list of positive examples.

Entity filtering for improving precision: We sampled abstracts obtained from the query and manually analyzed the contents of the abstract to evaluate precision, which was around 13%. From this initial set of examples, we observed that cancer may not be present in the abstract itself; we speculate that the search on MeSH terms such as *Neoplasms [mh]* include such articles. To improve precision, we filtered articles which do not contain *cancer* or *cell line*³ entities in the abstract, and we refer to this process as *entity filtering*. This had no impact on the recall of positive examples. Choosing 100 random abstracts from the output of the query, after applying entity filtering and manually annotating the text, we obtain a precision of over 40% in identifying relevant abstracts (See class schema in Section 2.1 for the definition of relevant abstracts). A majority of the abstracts were irrelevant because drugs or cancer types were mentioned as background information, without being the fo-

³Since the article is indexed with *Neoplasms*, we expect either cancer or cell lines to be discussed in the abstract.

Component	Definition
DISEASE	Neoplasms [mh] OR cancer [tiab] OR tumor [tiab] OR leukemia [tiab] OR carcinoma [tiab] OR blastoma [tiab] OR sarcoma [tiab] OR myeloma [tiab] OR lymphoma [tiab]
PUBLICATION-TYPE	clinical trial [pt] OR clinical observational study [pt] OR case reports [pt] OR controlled clinical trial [pt] OR randomized controlled trial [pt]
KEYWORDS	preclinical [tiab] OR in vitro [tiab] OR in vivo [tiab] OR animals [mh] OR humans [mh] OR mice* [tiab] OR mouse [tiab] OR placebo [tiab] OR clinical [tiab] OR observ* [tiab] OR Drug Evaluation, Preclinical [mh:noexp]
QUERY	(drug [tiab] AND disease) AND (publication-type OR keywords)

Table 1: Querying approach modified from CHSS strategy. **QUERY** enforces that output abstracts should contain the given drug, and should be related to cancer disease (DISEASE), and also be a pre-/early-clinical evidence (PUBLICATION-TYPE, KEYWORDS).

cus of the study.

Shallow filtering with proximity features: To improve precision further, we developed a *shallow filtering* step, which is based on drug and cancer type proximity features in the given abstract (Peng, Wei, and Lu 2016). By manually inspecting relevant abstracts, we devised the following set of features:

- Occurrence count for given drug in the abstract. We speculate that the higher the count, the more likely the drug is the focus of the study.
- Binary indicator of whether the drug and cancer entities are co-located at sentence level, and the position of such a sentence in the abstract. The collocation at the sentence level increases the chance of a relationship between the two, while the sentence position within the abstract may suggest importance (e.g., conclusive results are usually discussed towards the end)
- Binary-valued co-location features (as explained above) computed with and without expanding abbreviations.
- Binary indicator of whether the drug and cancer entities are co-located in the title.
- Binary indicator of whether the drug and cancer entities are co-located in the last sentence.

Dataset creation: We considered the drugs and corresponding abstracts from the positive examples list as positive instances, and the instances from random samples where *drug had no relation to cancer* as negative cases. We trained a naïve Bayes model (due to small sample-size) using the proximity features from 275 instances (repeated for multiple drugs in an abstract). The classification threshold was tuned using 5-fold cross validation, and we used this trained model to filter instances further. We obtained a sample of over 200 drug and abstract combinations from the query after passing the output through both the entity and shallow filters, and analyzed it manually. The ratio of cases where drug or cancer entities not relevant to the objective of study decreased where 81% of those cases were filtered (compared to 71% without the shallow filter), and the overall precision for *relevant* cases was 53%. We used this model to sample more abstracts for several iterations. All abstracts were manually annotated by one domain-expert annotator, and expanding the effort with more annotators deserves attention in future work. The dataset annotated so far comprises 1085 abstracts, out of which 522 are relevant and 563 irrelevant.

Coarse-level	Association	Count
Irrelevant	Drug has no relation to cancer	553
	No phenotypic outcome measured	155
Relevant	Effective	555
	Detrimental	50
	Inconclusive	216
	No effect	68

Table 2: Therapeutic association distribution in the dataset collected. Coarse-level: broad category of associations

Study type	Count
Pre-clinical	318
Clinical observational study	107
Clinical trial	69
Other	28

Table 3: Stage of evidence (study type) distribution in the relevant abstracts.

The therapeutic association distribution considering pairs of drug-cancer types is given in Table 2. Since there can be more than one (*drug, cancer*) pair in an abstract, the sum of all counts in the table are higher than the number of unique abstracts. The distribution of study types for the 522 relevant abstracts is given in Table 3. Note that for the relevant cases, the articles belong to a single study type, hence the study type classification corresponds to the entire abstract, and not to a drug-cancer pair, as is the case for the therapeutic association.

4 Solution Components

We discuss the approaches for entity extraction, therapeutic association classification, and study type classification, which form the key components of the proposed evidence discovery pipeline.

Named Entity Recognition: For cancer-related entity identification, we use various NER methods. The entities of interest include: cancer, cell lines and organ of the body. Cancer entities indicate the type of cancer discussed in the abstract, cell lines are used to filter abstracts (as explained in previous section), and organ of the body is used for post-processing. We train sequential token-level IOB (inside, outside, beginning) tag prediction model using the

Tag	...	O	O	O	O	O	O	O	B-Cancer	I-Cancer
Token	...	to	evaluate	a	potential	therapeutic	agent	for	breast	cancer

Table 4: NER example with cancer and other tokens tagged using IOB format.

Model	Recall	OS
CRF	54.2	66.4
spaCy	67.7	77.6

Table 5: NER Performance

BioNLP13CG dataset (Pyysalo et al. 2015). Tokens that are not of interest are treated as ‘O’. Table 4 shows an example. We use the well-known conditional random field (CRF) (Song et al. 2018) and convolutional neural network (CNN) based spaCy models (Honnibal and Montani 2017) for entity extraction as discussed below.

CRF: The following word features are used with a linear chain CRF-based probabilistic model: token, word shape (capitalization, punctuation, digits), Penn POS tags based on spaCy, ClearNLP dependency labels, relative position in the sentence, and features for the adjacent words.

spaCy: spaCy NER is a transition-based system based on the chunking model from (Lample et al. 2016). It uses a CNN to score actions, where tokens are represented as hashed, embedded representations of the prefix, suffix, shape and lemmatized features of individual words. This model has been shown to provide state-of-the-art results on many datasets (Neumann et al. 2019).

Once we obtain the chunks of entities (sequence of BI tokens), we perform post-processing as follows: (i) merge tokens of the form ‘Cancer (entity) of the Organ (entity)’ as ‘Organ Cancer’, for e.g., ‘tumor of the lung’ as ‘lung tumor’. (ii) ‘lung/breast cancer’ as ‘lung cancer’ and ‘breast cancer’. (iii) abbreviations and expansions are considered as entities.

Therapeutic Association Classification: Given a drug-cancer pair and the corresponding abstract text, we use the following approaches for association classification.

Logistic Regression: A standard logistic regression model trained with feature vectors that are a concatenation of term frequency bag-of-words representations of abstracts, drug and cancer type (Lehman et al. 2019). The tokens are stemmed using the Porter stemmer.

Deep Averaging Networks (DAN): This model is similar to logistic regression, except that the tokens of abstract, drug and cancer type are initialized with word vectors trained using skip-gram objective over a large set of PubMed abstracts (Pyysalo et al. 2013). The text of a given abstract is passed through deep averaging networks (Iyyer et al. 2015) where the word vectors are re-trained (with the training data and classification objective), and the representations of abstract, drug and cancer type are concatenated, and passed through a final logistic layer. During experimentation, we observed that the DAN model performs better than a simple feed-forward network where the word vectors are static.

SciBERT: BERT (Devlin et al. 2019) (deep bidirectional

Class	Log. Reg	DAN	SciBERT
Irrelevant	0.83	0.81	0.81
Relevant	0.80	0.79	0.85

Table 6: Binary therapeutic association classification

Class	Log. Reg	DAN	SciBERT
No relation	0.74	0.71	0.80
No phenotypic outcome	0.30	0	0.34
Effective	0.72	0.67	0.73
Detrimental	0.12	0	0.33
No effect	0.18	0	0.18
Inconclusive	0.30	0.03	0.25

Table 7: Fine-grained therapeutic association classification.

transformers that learn representations by predicting masked tokens and next sentence) trained on scientific literature text (Beltagy, Cohan, and Lo 2019). The drug and cancer type entities are encapsulated with special characters and concatenated with the input abstract text. The task is framed as a multi-class classification problem. The sequence representation is obtained using SciBERT’s encoding of the [CLS] token⁴ (from the last hidden layer). This encoding captures the entire sequence representation and is used for the multi-class classification with a logistic layer.

Study Type Classification: We train logistic regression models with different choices of features (Marshall et al. 2018): bag-of-words (BoW), publication type (PT), MeSH terms and combining all.

5 Experimental Results

NER: Since annotations are obtained only for the target cancer type in each abstract, we evaluate the performance (on the 1085 abstracts) using recall⁵ with exact match and a token-level overlap score (OS) (Moreau, Yvon, and Cappé 2008), where the predicted entity with highest overlap is used to compute the score. Performance comparing the CRF and spaCy model is given in Table 5: spaCy performs better than CRF.

Therapeutic Association: We perform a 5-fold cross-validation split at the document level, and evaluate performance for drug-cancer type pairs, given the abstract text. Note that we use gold-standard cancer type annotations for this analysis. We evaluate two different settings to understand the complexity of the task: (1) irrelevant vs. relevant

⁴[CLS] is inserted as a special beginning token for every input sequence.

⁵Ratio between *count of unique cancer entities predicted correctly by the model.* and *number of unique cancer entities*

Class	PT	BoW	MeSH	All
Pre-clinical	0.86	0.96	0.95	0.96
Clinical observational study	0.25	0.81	0.66	0.84
Clinical trial	0.78	0.78	0.60	0.80
Other	0.38	0.34	0.37	0.40

Table 8: Study type classification using logistic regression.

binary classification (Table 6) and (2) all six classes given in Section 2.1 (Table 7). Performance is measured using F-score. SciBERT performs better the best in most cases.

Study Type: Results using logistic regression with different choices of features are given in Table 8. Performance is measured using F-score. Using all the features together provides the best performance.

6 Conclusion and Future Work

We proposed an end-to-end evidence discovery pipeline that fetches potential candidate abstracts from PubMed for further evaluation with the goal of identifying non-cancer generic drug activity against different cancer types. We discuss the components in the pipeline, and use NLP approaches along with a number of well-thought-out heuristics to provide solutions for each component. From the experimental results, the task of distinguishing between *irrelevant* and *relevant* abstracts has promising performance. However, classifying the associations in a fine-grained manner is more challenging due to class imbalance and a relatively small dataset. This task calls for future research, including the development of semi-supervised approaches. Study type classification observes good performance. Future research involves generating a database of evidence that can be used by different modules in the Oncology Repurposing Engine to prioritize the most promising drug-cancer combinations.

7 Acknowledgments

This work was conducted under the auspices of the IBM Science for Social Good initiative.

References

Beltagy, I.; Cohan, A.; and Lo, K. 2019. SciBERT: Pretrained contextualized embeddings for scientific text. arXiv:1903.10676.

Bouche, G.; Pantziarka, P.; and Meheus, L. 2017. Beyond aspirin and metformin: The untapped potential of drug repurposing in oncology. *Eur. J. Cancer* 172:S121–S122.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

Dickersin, K.; Scherer, R.; and Lefebvre, C. 1994. Identifying relevant studies for systematic reviews. *BMJ* 309:1286.

Hernandez, J. J.; Prysizlak, M.; Smith, L.; Yanchus, C.; Kurji, N.; Shahani, V. M.; and Molinski, S. V. 2017. Giving drugs a second chance: Overcoming regulatory and financial hurdles in repurposing approved drugs as cancer therapeutics. *Front. Oncol.* 7:273.

Honnibal, M., and Montani, I. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Iyyer, M.; Manjunatha, V.; Boyd-Graber, J.; and Daumé III, H. 2015. Deep unordered composition rivals syntactic methods for text classification. In *ACL-IJCNLP*.

Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; and Dyer, C. 2016. Neural architectures for named entity recognition. In *NAACL-HLT*.

Lehman, E.; DeYoung, J.; Barzilay, R.; and Wallace, B. C. 2019. Inferring which medical treatments work from reports of clinical trials. In *NAACL-HLT*.

Li, J.; Zhu, X.; and Chen, J. Y. 2009. Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLoS Comput. Biol.* 5(7):e1000450.

Marshall, I. J.; Noel-Storr, A.; Kuiper, J.; Thomas, J.; and Wallace, B. C. 2018. Machine learning for identifying randomized controlled trials: An evaluation and practitioner’s guide. *Res. Synth. Methods* 9:602–614.

Martínez, V.; Navarro, C.; Cano, C.; Fajardo, W.; and Blanco, A. 2015. DrugNet: Network-based drug–disease prioritization by integrating heterogeneous data. *Artif. Intell. Med.* 63(1):41–49.

Moreau, E.; Yvon, F.; and Cappé, O. 2008. Robust similarity measures for named entities matching. In *COLING*.

Neumann, M.; King, D.; Beltagy, I.; and Ammar, W. 2019. ScispaCy: Fast and robust models for biomedical natural language processing. arXiv:1902.07669.

Palma, G.; Vidal, M.-E.; and Raschid, L. 2014. Drug-target interaction prediction using semantic similarity and edge partitioning. In *ISWC*.

Pantziarka, P.; Sukhatme, V.; Meheus, L.; Sukhatme, V.; and Bouche, G. 2017. Repurposing non-cancer drugs in oncology — how many drugs are out there? bioRxiv:197434.

Peng, Y.; Wei, C.-H.; and Lu, Z. 2016. Improving chemical disease relation extraction with rich features and weakly labeled data. *J. Cheminformatics* 8:53.

Pyysalo, S.; Ginter, F.; Moen, H.; Salakoski, T.; and Ananiadou, S. 2013. Distributional semantics resources for biomedical text processing. In *LBM*.

Pyysalo, S.; Ohta, T.; Rak, R.; Rowley, A.; Chun, H.-W.; Jung, S.-J.; Choi, S.-P.; Tsujii, J.; and Ananiadou, S. 2015. Overview of the cancer genetics and pathway curation tasks of BioNLP shared task 2013. *BMC Bioinformatics* 16(Suppl. 10):S2.

Song, H.-J.; Jo, B.-C.; Park, C.-Y.; Kim, J.-D.; and Kim, Y.-S. 2018. Comparison of named entity recognition methodologies in biomedical documents. *BioMed. Eng. OnLine* 17(Suppl. 2):158.

Verbaanderd, C.; Meheus, L.; Huys, I.; and Pantziarka, P. 2017. Repurposing drugs in oncology: Next steps. *Trends Cancer* 3(8):543–546.

Xue, H.; Li, J.; Xie, H.; and Wang, Y. 2018. Review of drug repositioning approaches and resources. *Int. J. Biol. Sci.* 14(10):1232–1244.