Using the Literature-Based Discovery Paradigm to Investigate Drug Mechanisms

Caroline B. Ahlers, MD[1], Dimitar Hristovski, PhD,[2] Halil Kilicoglu,[1] MS, Thomas C. Rindflesch, PhD[1]
[1]National Library of Medicine, Bethesda, MD
[2]Institute of Biomedical Informatics, Medical Faculty, University of Ljubljana, Slovenia

Abstract

Drug therapies are often used effectively without their underlying mechanism being completely understood. We exploit the literature-based discovery paradigm to investigate these mechanisms and propose a discovery pattern that draws on semantic predications extracted from MEDLINE citations. The use of semantic predications and the discovery pattern provides a way to uncover previously unnoticed associations between pharmacologic and bioactive substances on the one hand and bioactive substances and disorders on the other. In this paper, we concentrate on research investigating the use of antipsychotic agents used for treatment of cancer. Our method resulted in five biomolecules that may provide a link between the antipsychotic agents and cancer: brain-derived neurotrophic factor, CYP2D6, glucocorticoid receptor, PRL, and TNF.

Introduction

There has been a longstanding informal observation that schizophrenics have lower incidence of cancer than the general population [1-3]. Assuming this correlation is valid, Mortensen [4] discusses the role of neuroleptic medication. Carrillo and Benítez [5] suggest a mechanism involving the inhibition of some of the cytochrome P450 microsomal enzymes (specifically, CYP1A2 and CYP2D6) by antipsychotic drugs. Additional research has further investigated the potential of antipsychotic agents to treat cancer (for example [6-7]).

Drug therapies are often used effectively, even though the exact cause of action may be either poorly understood or unknown. In this paper we exploit the literature based discovery paradigm [8] as the basis for a methodology investigating the underlying mechanisms of drug therapies, concentrating on the use of antipsychotic agents to treat cancer.

Background

Literature-Based Discovery

Literature based discovery (LBD) is a method for uncovering relationships not overtly asserted in the research literature. Swanson [8] defined the original paradigm, in which an association between two concepts A and C not directly asserted in the research literature may be uncovered via a third concept (B). Swanson stipulated that A and C be in literature domains that do not overlap. The possible relationship between A and C is considered to be a discovery and a hypothesis for future research. For example, after noting an association between fish oil and blood viscosity (A-B) and another association between blood viscosity and Raynaud's disease (B-C), Swanson [8] proposed fish oil (A) as a new treatment for Raynaud's disease (C).

Swanson's system, as well as many that followed [9-14] were based on finding co-occurrence of (typically) words or phrases. Srinivasan and Libbus [15] use MeSH terms assigned to MEDLINE citations. Hristovski, et al. [16] extended Swanson's paradigm. Analogous to Swanson's A, B, and C literature domains, they defined concepts X, Y, and Z. They also augmented co-occurrences with semantic predications giving specific information about the nature of the association. They argue that the more specific information provided by semantic predications benefits the discovery process by being more understandable, lowering the number of relations that have to be assessed by humans (at an acceptable cost of some missed relations), and providing explanation capabilities.

Hristovski et al. [16] further defined the notion of a *discovery pattern*, which contains a set of conditions to be satisfied for the discovery of new relations between concepts. Using two such patterns, based on changes of a substance, body function, or body measurement associated with a disease, they suggested insulin as a novel treatment for Huntington's disease.

In the preceding, LBD is used for open discovery, in which X-Y and Y-Z relations are used to discover an X-Z relation. Another way to exploit LBD is through closed discovery. In this method, X-Z is known (or assumed). X-Y and Y-Z relations are then scrutinized to determine what Y concepts they have in common,

as a way of explicating the relationship between X and Z. Examples of closed discovery are those given in [17] (Y concepts to explain the relation between migraine and magnesium) and [18], which proposes an explanation for the epidemiologic evidence that estrogen protects against Alzheimer's disease.

Natural-Language Processing

Semantic predications represent relations asserted between two entities in text. In this study we rely on SemRep [19] to extract semantic predications from MEDLINE citations. Medical domain knowledge is provided by the Unified Medical Language System (UMLS). The UMLS Metathesaurus is accessed using MetaMap [20], and permissible semantic relations are defined by the UMLS Semantic Network. Examples of predications extracted from (1) are given in (2).

(1) IL-4 production was inhibited by haloperidol and chlorpromazine, but not by clozapine.
(2) Haloperidol INHIBITS Interleukin-4
    Chlorpromazine INHIBITS Interleukin-4

Methods

Based on the methodology introduced in [17] we suggest a discovery pattern, *May_Disrupt* (3), for explicating the mechanisms underlying drug therapies that are currently used but poorly understood.

 (3)   Substance X <inhibits> Substance Y
        Substance Y <causes> Pathology Z
        Substance X <may_disrupt> Pathology Z

The *May_Disrupt* pattern concentrates on pharmacogenomics (relationship among drugs, genes, and diseases). The lines in the pattern match SemRep predications in this domain. The first line matches predications with predicate INHIBITS, representing the inhibitory action of one bioactive substance on another (X-Y relations). The second line matches a SemRep predication with predicate CAUSES, PREDISPOSES, or ASSOCIATED_WITH, representing etiological relations between a bioactive substance and a pathological process (Y-Z relations). The third line matches predications with predicate TREATS or PREVENTS (X-Z relation).

When used for open discovery, *May_Disrupt* states that if substance X inhibits substance Y and if substance Y causes disease Z, then substance X may disrupt (prevent or treat) disease Z. When used for closed discovery, the *May_Disrupt* pattern states that for a drug X that treats disease Z, if drug X inhibits Y and Y causes Z, then Y is (part) of the mechanism of action in X treating Z. Cole and Bruza [21] discuss an alternative mechanism for both open and closed discovery.

In this paper we exploit *May_Disrupt* for closed discovery. Rather than suggesting a new drug therapy for a disease, we attempt to explicate the mechanism underlying drug therapies already in use. We followed the following procedure in exploiting SemRep predications and the discovery pattern *May_Disrupt* for this purpose.

We first obtained two sets of MEDLINE citations by using an X term (substance) and a Z term (pathology) as PubMed queries. We then processed these citations with SemRep, producing two sets of semantic predications. The first includes X-Y relations for the known X term and various unknown Y terms. The second includes Y-Z relations for the known Z term and various unknown Y terms.

In order to locate useful X-Y and Y-Z relations, the two sets of predications were subjected to further processing. First, predications containing arguments that occur near the root of a hierarchy in the UMLS Metathesaurus (such as "Pharmacologic Substance," "Disease," or "Gene") were eliminated as being too general to be useful. Second, arguments in each set were filtered for the relevant X or Y term. In the X set, only those predications were kept that had the X term as subject. In the Z set, those with the Z term as object were kept.

The remaining predications were matched to lines one and two of the *May_Disrupt* discovery pattern (3). In the set of predications generated from the X term citations, only those with predicate INHIBITS were kept. These match line one and constitute X-Y relations. To locate Y-Z relations (line two), in the set of predications generated from the Z term citations, only those with predicate CAUSES, PREDISPOSES, and ASSOCIATED_WITH were kept. A list of Y arguments shared by the X-Y and Y-Z relations was then generated. These serve as potential explanatory links between the two relations.

Finally, we conducted a novelty check to determine to what extent the Y terms the system proposed have already been discussed in the research literature.

Results

In applying our methodology to investigate antipsychotic agents (X) used to treat cancer (Z), we first produced a set of MEDLINE citations for both terms. For the antipsychotic agents, we issued a PubMed query containing "(antipsychotic agents[mh] OR psychoses/drug therapy[mh] OR antipsychotic agents[pa])" and several specific names of antipsychotic drugs; this query returned 113,243 citations. For cancer, we issued the PubMed query

"neoplasms[mh]" and retained the most recent 100,000 citations. The retrieved citations constitute X and Z sets and were processed with SemRep, resulting in 721,257 and 903,808 predications, respectively.

The predications in each set were then further processed to locate useful X-Y and Y-Z relations. In both sets, predications containing non-specific arguments were eliminated. We then defined the X and Z terms in each set. In the X set, this was "Antipsychotic Agents," and all predications not containing this concept as subject were eliminated, leaving 16,704 predications. In the Z set, the Z term was defined as concepts having the UMLS semantic type 'Neoplastic Process'. After eliminating predications not having an object with this semantic type, 37,535 predications were left.

The X and Z predications were then filtered through lines one and two of the *May_Disrupt* discovery pattern. In the X set, only predications with predicate INHIBITS were kept (line one); this produced 568 X-Y relations representing an antipsychotic agent inhibiting a bioactive substance. In the Z set, only predications with predicates CAUSES, PREDISPOSES, and ASSOCIATED_WITH were kept (line two), resulting in 16,943 Y-Z relations representing a bioactive substance playing a role in the etiology of cancer. From the remaining X-Y and Y-Z relations, we then isolated the list of Y arguments they shared.

Before further analysis, we eliminated some concepts from the Y list which were unlikely to be useful. We erased all drugs (concepts with UMLS semantic type 'Pharmacologic Substance') since our current goal was not to assess X concepts interacting with drugs. Further, concepts referring to classes (such as "Tumor Suppressor Genes"), rather than specific substances, were also eliminated. Fifteen Y terms remained, as listed in (4).

(4) Y Terms

APOD gene
APOE gene
*Brain-Derived Neurotrophic Factor
Calmodulin-Dependent Phosphodiesterase
CASP4
Concanavalin A
CRH gene
*CYP2D6 gene
Dopamine D2 Receptor
EPO gene
GAG gene
*Glucocorticoid Receptor
Heat shock proteins
*PRL gene
Receptors, Purinergic P1
*TNF gene

These are substances both inhibited by antipsychotic agents and involved in the etiology of cancer; they can potentially contribute to our understanding of the mechanisms underlying antipsychotic agents treating cancer.

The accuracy of the predications in which the fifteen Y terms occurred was assessed manually by the first author. The ten unstarred terms in (4) were determined to have come from predications generated due to SemRep errors. Almost all of these were due to ambiguous words or acronyms. For example, the text token *Ca2+* was wrongly mapped to the gene concept "CASP4," which then became an argument in an incorrect SemRep predication. The five concepts marked with an asterisk are those remaining after validation and may serve as (partial) explanatory links between the antipsychotic agents and cancer.

These five terms were subjected to an assessment for novelty (manually by the first author) to determine to what extent they had already been discussed in the literature as involved in antipsychotic agents and cancer treatment. MEDLINE was searched for citations that discussed antipsychotic agents and cancer along with one of the starred concepts in (4). Five PubMed searches were conducted, one for each of the starred concepts in (4). Each search consisted of three components consisting of a combination of MeSH terms and text words to match: a) any antipsychotic agent, b) any cancer, and c) one of the starred terms in (4).

Fifteen total citations were returned by these five searches. Each citation was examined and it was determined that only one (PMID 10492064, see [6] above) discussed one of the relevant terms (CYP2D6) as explaining an antipsychotic agent treating cancer. In nine of these fifteen citations, the three relevant terms (an antipsychotic, a cancer, and one of the concepts from (4)) did not in fact appear. For example, PMID 15056479 refers to perazine (a phenothiazine antipsychotic agent) and CYP2D6; however, cancer is not mentioned. The research is about the details of the metabolism of this drug. The citation was returned because it was indexed with the MeSH term "Neoplasm Metastasis." In five citations, the three relevant terms occurred, but the Y term from (4) was not discussed as an explanation for an antipsychotic used for cancer. For example, PMID 11071396 discusses the well-known **stimulation** of

the PRL gene by the antipsychotic agents reserpine and haloperidol, as well as the cancer-predisposing actions of PRL. However, our system extracted predications on the lesser-known **inhibition** of PRL by antipsychotic agents at high doses.

Discussion

In order to assess the viability of our method in explicating the disruptive link between antipsychotic drugs and cancer, we examined some of the citations from which SemRep had extracted the terms in (4). We note several citations which state either that the antipsychotic agents inhibit one of the relevant substances or one of these causes cancer. A few citations discuss these facts as an explanation of the effect of the antipsychotic agents on cancer. We note where the results of this project contribute to and extend those ideas.

Brain-Derived Neurotrophic Factor (BDNF)
Statements supporting antipsychotic agents as inhibiting BDNF include "In recent studies, the BDNF expression was reduced by typical neuroleptics," (PMID 15526143). BDNF was asserted to be associated with primary carcinoma of the liver cells and neoplasm progression (PMID 17089044) and multiple myeloma (PMID 16875931).

PRL gene
Though antipsychotic agents are well known as stimulators of PRL gene expression, our system identified assertions that antipsychotic agents inhibit the PRL gene at certain doses and under certain conditions. (PMID 10530797, PMID 436760). The role of PRL gene as an etiological agent for many forms of cancer (breast, prostate, rectum, hematopoietic system, etc.) is well documented. Assertions include "Genetic variation in the PRL and PRLR genes was shown to influence breast cancer risk" (PMID 16434456) and "Prolactin promotes growth of a spontaneous T cell lymphoma: role of tumor and host derived cytokines" (PMID 16982465).

CYP2D6 gene:
As noted earlier, the CYP2D6 gene has been discussed as providing a link between the antipsychotic agents and cancer [6]. Our results support and expand that notion. Statements obtained by our system which support the inhibition of this gene include, "One-day exposure of rats to the classic neuroleptics decreased the activity of CYP2D in rat liver microsomes" (PMID 15572279). Citations referring to the etiological association between CYP2D6 gene and various types of cancer include those discussing carcinogenic agents and other bioactive molecules in organ tissues including the prostate (PMID 16716118), the pituitary gland (PMID 16611538), and the hematopoietic system (PMID 16493615).

TNF gene:
Tumor necrosis factor alpha has an etiological effect on cancer. Our system returned predications on its role as an angiogenic switch (PMID 16935777, PMID 16263219, PMID 16114015); TNF mutations involved in cancer predisposition (PMID 16476505, PMID 16643431, PMID 16839795); and the role of TNF in cell growth stimulation (PMID 16643431). Previous research has discussed phenothiazines treating cancer by inhibiting TNF (PMID 17017885). We also found a statement about the inhibitory relationship between antipsychotic agents and TNF: "Antipsychotic drugs and PCP significantly reduced the levels of TNF in the prefrontal cortex compared to vehicle-treated animals, whilst other cytokines remained unchanged." (PMID 16478754)

Glucocorticoid Receptor:
Our system identified etiological associations between glucocorticoid receptor and gastric carcinoma (PMID 16713543) and breast carcinoma (PMID 16639692). It also identified an assertion of antipsychotic agents inhibiting the glucocorticoid receptor "Previously, we have found that some antipsychotic drugs are able to inhibit glucocorticoid receptor (GR)-mediated gene transcription." (PMID 14730115).

The current implementation of this approach is limited in several ways. Effectiveness is dependent on SemRep accuracy. As SemRep improves, particularly regarding resolution of word sense ambiguity, we expect the number of false positives (the unstarred terms in (4)) to decrease. More generally, the discovery pattern, which underpins our method, was limited in this study to relationships that can be represented as two predications. In principle, more complex relationships can be accommodated by incorporating chains of predication schemas into discovery patterns; however, we have so far not investigated this possibility. Finally, we processed the most recent 100,000 MEDLINE citations on cancer, rather than the total retrieval of 1,800,000.

Conclusion

Working in the literature-based discovery paradigm, we investigated the mechanisms underlying drug therapies, concentrating on research discussing the antipsychotic agents for cancer. We define a discovery pattern that guides the discovery of these

mechanisms, focusing on drug-bioactive substance relations as well as associations between bioactive substances and disorders. The discovery pattern draws on semantic predications extracted from MEDLINE citations using SemRep. Our method resulted in five bioactive substances that may provide a link between the antipsychotic agents and cancer: brain-derived neurotrophic factor, CYP2D6, glucocorticoid receptor, PRL, and TNF.

References

1. Commissioners in Lunacy for England and Wales (1909) Annual Report. London: HMSO.
2. Mortensen PB. The occurrence of cancer in first admitted schizophrenic patients. Schizophr Res. 1994 Jun;12(3):185-94.
3. Dalton SO, Mellemkjaer L, Thomassen L, Mortensen PB, Johansen C. Risk for cancer in a cohort of patients hospitalized for schizophrenia in Denmark, 1969-1993. Schizophr Res. 2005 Jun 15;75(2-3):315-24. Epub 2004 Dec 13. Review.
4. Mortensen PB. Neuroleptic medication and reduced risk of prostate cancer in schizophrenic patients.
Acta Psychiatr Scand. 1992 May;85(5):390-3.
5. Carrillo JA, Benitez J. Are antipsychotic drugs potentially chemopreventive agents for cancer? Eur J Clin Pharmacol. 1999 Aug;55(6):487-8.
6. Schleuning M, Brumme V, Wilmanns W. Growth inhibition of human leukemic cell lines by the phenothiazine derivative fluphenazine. Anticancer Res. 1993 May-Jun;13(3):599-602.
7. Frussa-Filho R, Monteiro Mdo C, Soares CG, Decio RC. Effects of haloperidol, bromocriptine and amphetamine on the development of Ehrlich ascites carcinoma in mice. Pharmacology. 1992;45(1):58-60.
8. Swanson, D.R.: Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspect Biol Med. 1986 Autumn;30(1):7-18.
9. Swanson, D.R., Smalheiser, N.R.: An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artif. Intell. 91 (1997) 183-203.
10. Hristovski D, Stare J, Peterlin B, Dzeroski S. Supporting discovery in medicine by association rule mining in Medline and UMLS. MEDINFO 2001.
11. Hristovski, D., Peterlin, B., Mitchell, J. A. and Humphrey, S. M. (2005), 'Using literature-based discovery to identify disease candidate genes', Int. J. Med. Inform., Vol. 74(2–4), pp. 289–298.
12. Weeber M, Klein H, Aronson AR, Mork JG, Jong-Van Den Berg L, Vos R. Text-based discovery in biomedicine: the architecture of the DAD-system. Proc AMIA Symp. 2000;(20 Suppl):903-7.
13. Gordon MD, Lindsay RK. Toward discovery support systems: A replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. J Am Soc Inf Sci 1996; 47(2):116-128.
14. Fuller, S. S., Revere, D., Bugni, P. F. and Martin, G. M. (2004), 'A knowledgebase system to enhance scientific discovery: Telemakus', Biomed. Digit Libr., Vol. 1(1), p. 2.
15. Srinivasan, P. and Libbus, B. (2004), 'Mining MEDLINE for implicit links between dietary substances and diseases', Bioinformatics, Vol. 20, Suppl 1, pp. I290–I296.
16. Hristovski D, Friedman C, Rindflesch TC, Peterlin B. Exploiting semantic relations for literature-based discovery. AMIA Annu Symp Proc. 2006:349-53.
17. Swanson DR. Migraine and magnesium: eleven neglected connections. Perspect Biol Med. 1988 Summer;31(4):526-57.
18. Smalheiser NR, Swanson DR. Linking estrogen to Alzheimer's disease: an informatics approach. Neurology. 1996 Sep;47(3):809-10.
19. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J of Biomed Inf. 2003 Dec;36(6):462-77.
20. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp. 2001;17-21.
21. Cole RJ, Bruza PD. A Bare Bones Approach to Literature-Based Discovery: An Analysis of the Raynaud's/Fish-Oil and Migraine-Magnesium Discoveries in Semantic Space. Discovery Science 2005:84-98.