Systematic drug repositioning through mining adverse event data in ClinicalTrials.gov (#14793)

First submission

Please read the **Important notes** below, the **Review guidance** on page 2 and our **Standout reviewing tips** on page 3. When ready **submit online**. The manuscript starts on page 4.

Important notes

Editor and deadline

Walter de Azevedo Jr. / 24 Dec 2016

Files	1 Table file(s) 1 Other file(s) Please visit the overview page to **download and review** the files not included in this review PDF.
Declarations	No notable declarations are present

Review guidelines

Please read in full before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standards**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

7 Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that your international audience can clearly understand your text. I suggest that you have a native English speaking colleague review your manuscript. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Give specific suggestions on how to improve the manuscript	*Line 56: Note that experimental data on sprawling animals needs to be updated. Line 66: Please consider exchanging "modern" with "cursorial".*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Systematic drug repositioning through mining adverse event data in ClinicalTrials.gov

Eric Wen Su Corresp., [1] , **Todd M Sanger** [1]

[1] Advanced Analytics Hub, Eli Lilly and Company, Indianapolis, Indiana, United States of America

Corresponding Author: Eric Wen Su
Email address: ewsu@lilly.com

Drug repositioning (i.e. drug repurposing) is the process of discovering new uses for marketed drugs. Historically, such discoveries were serendipitous. However, the rapid growth in electronic clinical data and text mining tools makes it feasible to systematically identify drugs with the potential to be repurposed. Described here is a novel method of drug repositioning by mining ClinicalTrials.gov. The text mining tools I2E (Linguamatics) and PolyAnalyst (Megaputer) were utilized. An I2E query extracts “Serious Adverse Events” (SAE) data from randomized trials in ClinicalTrials.gov. Through a statistical algorithm, a PolyAnalyst workflow ranks the drugs where the treatment arm has fewer predefined SAEs than the control arm, indicating that potentially the drug is reducing the level of SAE. Hypotheses could then be generated for the new use of these drugs based on the predefined SAE that is indicative of disease (for example, cancer).

Systematic Drug Repositioning through Mining Adverse Event Data in ClinicalTrials.gov

Eric Wen Su[1], Todd M. Sanger[1]

[1]Advanced Analytics Hub, Eli Lilly and Company, Indianapolis, IN, USA.

Corresponding Author:
Eric Su[1]

Email address: ewsu@lilly.com

Abstract

Drug repositioning (i.e. drug repurposing) is the process of discovering new uses for marketed drugs. Historically, such discoveries were serendipitous. However, the rapid growth in electronic clinical data and text mining tools makes it feasible to systematically identify drugs with the potential to be repurposed. Described here is a novel method of drug repositioning by mining ClinicalTrials.gov. The text mining tools I2E (Linguamatics) and PolyAnalyst (Megaputer) were utilized. An I2E query extracts "Serious Adverse Events" (SAE) data from randomized trials in ClinicalTrials.gov. Through a statistical algorithm, a PolyAnalyst workflow ranks the drugs where the treatment arm has fewer predefined SAEs than the control arm, indicating that potentially the drug is reducing the level of SAE. Hypotheses could then be generated for the new use of these drugs based on the predefined SAE that is indicative of disease (for example, cancer).

Introduction

Drug repositioning (i.e. drug repurposing) involves the identification and development of new uses for existing drugs (Ashburn & Thor 2004). The best known example of drug repositioning is the serendipitous discovery of the additional use of thalidomide for the treatment of painful sores associated with leprosy. In 1964, Dr. Jacob Sheskin used thalidomide to help a patient sleep, unexpectedly, the thalidomide also healed the patient's sores and eliminated his pain (Ashburn & Thor 2004; Sheskin 1965). This discovery shows that clinical data could be the most direct and reliable source of drug repositioning.

However, systematic drug repositioning efforts since 1964 have not been based on clinical data. Typical approaches include high-throughput screening of marketed drugs (Qosa et al. 2016), targeted testing of a class of drugs for a new disease area (Wu et al. 2016a), and *in silico* methods (Hodos et al. 2016; Mullen et al. 2016), usually based on drug-target interactions (Zheng et al. 2015).

Described here is a novel approach to drug reposition using data from randomized clinical trials. Text mining tools have been used to extract serious adverse event (SAE) data, identify drugs with fewer events related to diseases or associated symptoms in the drug arm than in the control arm, and rank the drugs based on the z-score of log odds ratio.

Materials & Methods

A text mining query was developed to extract SAE data from clinical trial data posted at ClinicalTrials.gov. ClinicalTrials.gov (https://clinicaltrials.gov/) is a registry of federally and privately funded clinical trials conducted in the United States and around the world, and contains rich biomedical data from over 220,000 studies in 191 countries. The query was built using Linguamatics' text mining tool I2E (Cormack et al. 2015).

The query (shown in Figure 1) has 4 main elements:

- To extract Serious Adverse Events classified as cancerous, the combined cancer terms and synonyms from MeSH (https://www.nlm.nih.gov/mesh/) and NCI (http://www.cancer.gov/research/resources/terminology) were loaded into the query region "Serious Event Subtitle" of ClinicalTrials.gov (the "Neoplasms" class).

- The same "Neoplasms" class was negated in the "Condition" region to exclude cancer trials.
- To link the SAE counts to the relevant study arm (i.e. drug or placebo etc.), the group (study arm) IDs and description ("Title") were extracted from the Reporting Groups region
- The wildcard "random*" was required in Study Design or Official Title region to ensure only randomized trials are reported

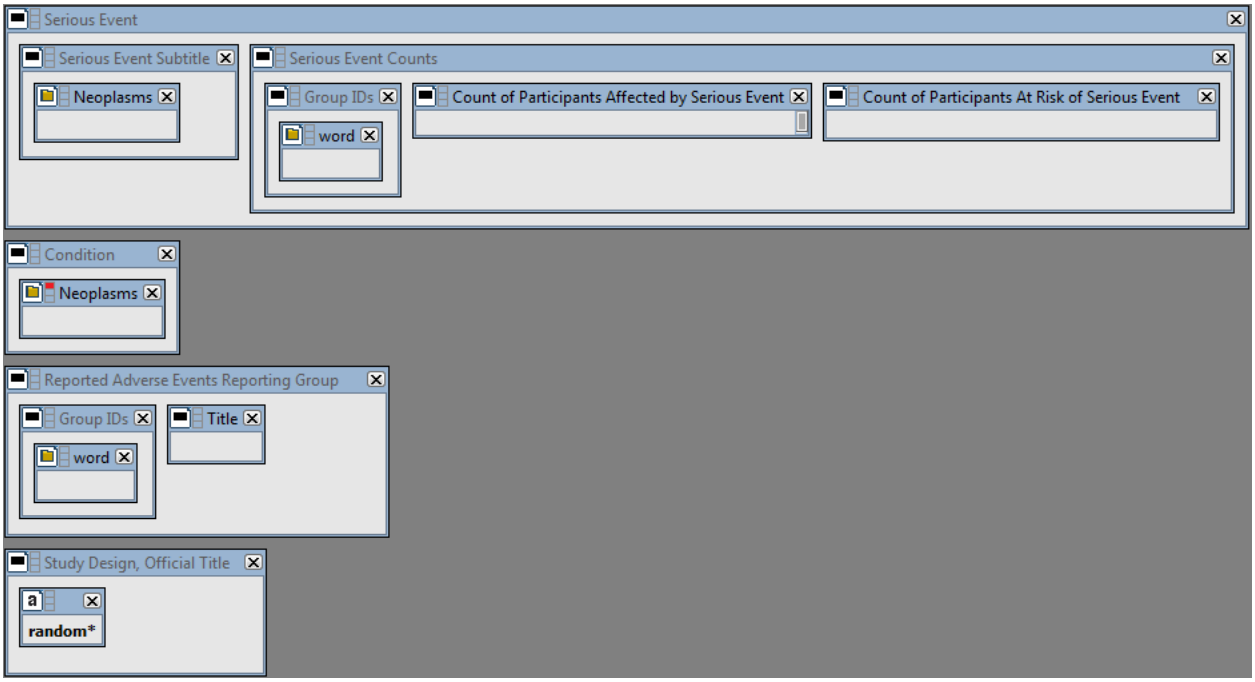

Figure 1. The I2E query. See the "Supplementary information" to reproduce the query by copying and pasting the YAML script into the I2E Pro interface.

The Excel output from the I2E query in Figure 1 was loaded into PolyAnalyst (Megaputer) for reformatting and calculating odds ratios (OR) and z-score. The final table was sorted by z-score.

The formula for calculating odds ratio (OR), standard error (SE), 95% confidence interval lower and upper limits (*LowerLimit* and *UpperLimit*), and z-score are as follows:
Let *Cs* = Number of patients with **S**AE in **C**ontrol arm; *Cn* = **N**umber of patients in **C**ontrol arm and *Ds* = Number of patients with **S**AE in **D**rug arm; *Dn* = **N**umber of patients in **D**rug arm

$$OR = \frac{Ds/(Dn - Ds)}{Cs/(Cn - Cs)}$$

The distribution of log(OR) is approximately normal with:

$$SE = \sqrt{\frac{1}{Cs} + \frac{1}{Cn - Cs} + \frac{1}{Ds} + \frac{1}{Dn - Ds}}$$

$$LowerLimit = \exp(\log(OR) - 1.96SE)$$
$$UpperLimit = \exp(\log(OR) + 1.96SE)$$

The null hypothesis is that there is no difference between drug and control arm (expected mean OR =1). Therefore,

$$z = \frac{\log(OR) - \log(1)}{SE} \text{ or } z = \log(OR) / SE$$

Since the *Cs* and *Ds* are usually small, SE, lower and upper limits, and z-score may not be meaningful for hypothesis testing. However, z-scores are still useful to rank drugs for hypothesis generation on drug repurposing.

Also because of the multiple comparison nature of the algorithm, the results should be only used for hypothesis generation, not for making any conclusion.

For drugs with z-scores ≤ -1.96, we reviewed the biomedical literature on the drugs, the drug targets, and the disease pathways to see if the hypothesis is consistent with the current scientific knowledge. Any drug repositioning hypothesis would have to be tested by a new randomized clinical trial with the hypothesis predefined in its protocol.

Results

The I2E query in Figure 1 was run on the ClinicalTrails.gov index updated on August 14, 2016. The report contains 105,399 SAE events classified as cancer, from 2861 randomized trials. An example of the extracted data is shown in Table 1.

ClinicalTrials.gov ID	Serious Adverse Event	Study Arm	Number of Patients with SAE	Number of Patients
NCT00048165	Basal cell carcinoma	Daclizumab	4	216
NCT00048165	Basal cell carcinoma	Placebo	3	207
NCT00048581	BASAL CELL CARCINOMA	Abatacept (ABA)	1	258
NCT00048581	BASAL CELL CARCINOMA	Placebo (PLA)	0	133
NCT00089661	Benign breast neoplasm	Denosumab 60 mg Q6M	0	129
NCT00089661	Benign breast neoplasm	Placebo	1	120
NCT00089661	Benign ovarian tumour	Denosumab 60 mg Q6M	1	129
NCT00089661	Benign ovarian tumour	Placebo	0	120

Table 1. A sample of the data extracted from ClinicalTrials.gov by the I2E query described above.

The I2E output table was reformatted as illustrated in Table 2 to have one row per trial per SAE (type of cancer).

ClinicalTrials.gov ID	Serious Adverse Event	Number of patients with SAE in control arm	Number of patients in control arm	Control Arm	Number of patients with SAE in drug arm	Number of patients in drug arm	Drug Arm
NCT00089791	Bladder cancer	3	3876	Placebo	4	3886	Denosumab 60 mg Q6M
NCT00089791	Breast cancer	25	3876	Placebo	34	3886	Denosumab 60 mg Q6M
NCT00089791	Colon cancer	8	3876	Placebo	11	3886	Denosumab 60 mg Q6M
NCT00120289	Lung neoplasm malignant	14	1696	Placebo + Simvastatin	8	1718	ERN + Simvastatin
NCT00120289	Malignant melanoma	4	1696	Placebo + Simvastatin	1	1718	ERN + Simvastatin
NCT00120289	Non-small cell lung cancer	4	1696	Placebo + Simvastatin	0.3	1718	ERN + Simvastatin
NCT00143507	Colon cancer	7	5430	Placebo	5	5477	Ivabradine
NCT00143507	Rectal cancer	6	5430	Placebo	3	5477	Ivabradine

Table 2. A sample of the reformatted table.

If a row has less than 3 patients with SAE in the control arm, it is deleted. This is because the goal is to find drugs that have fewer cancer SAEs in the drug arm than in the control arm. After the deletions, the table has only 601 rows left.

If a row has 0 patients with SAE in the drug arm, the 0 value is replaced with 0.3. These replacements enable the ranking of the drugs that have no cancer SAE in the drug arm. Without the replacements, all such rows will have zero for OR and minus infinity for z-score.

The final table with calculated columns is shown in Table 3. The drugs were ranked by sorting the z-score from the lowest value to the highest.

Drug	Serious Adverse Event	Ds	Dn	Cs	Cn	Control	SE	OR	Lower Limit	Upper Limit	z	ClinicalTrials.gov ID
V501	Cervical dysplasia	20	480	46	468	Placebo	0.28	0.40	0.23	0.69	-3.33	NCT00378560
Clopidogrel / Telmisartan	Colon cancer	4	5000	14	5023	Clopidogrel / Placebo	0.57	0.29	0.09	0.87	-2.20	NCT00153062
Vorapaxar	RECTAL CANCER	4	13186	13	13166	Placebo	0.57	0.31	0.10	0.94	-2.06	NCT00526474
Phylloquinone	Cancer	3	217	11	223	Placebo	0.66	0.27	0.07	0.98	-1.99	NCT00150969
Clopidogrel + ASA	Pancreatic carcinoma	1	3772	8	3782	Placebo + ASA	1.06	0.13	0.02	1.00	-1.96	NCT00249873
Core-phase: Aliskiren	Gastric cancer	1	4272	8	4285	Core-phase: Placebo	1.06	0.13	0.02	1.00	-1.96	NCT00549757

Table 3. The final table with calculated columns. The rows are sorted by z-score. Only the top 6 rows are shown (see Supplementary information for all 162 rows with z < -1). *Ds* = Number of patients with **S**AE in **D**rug arm; *Dn* = **N**umber of patients in **D**rug arm; *Cs* = Number of patients with **S**AE in **C**ontrol arm; *Cn* = **N**umber of patients in **C**ontrol arm

The results in Table 3 could range from false positive findings to possible signals for drug repositioning hypotheses. Therefore, we evaluated the drugs for cancer by other research from the current biomedical literature.

The V501 vaccine (Table 3, Row 1) arm had less cervical dysplasia events than control in a clinical trial on the prevention of papillomavirus infection. Papillomavirus is already known to be associated with cervical dysplasia (Firnhaber et al. 2009), a precursor lesion of cancer of the cervix (Kesic et al. 1990). We consider this top hit as a positive control that supports the credibility of our approach, since the prevention of the viral infection would naturally lead to the prevention of cervical dysplasia.

The data in Table 3, Row 2 suggest that the hypertension drug Telmisartan might be useful to prevent colon cancer (note that Clopidogrel is in both Drug and Control arm, so we did not investigate Clopidogrel further). Recent cell-based studies reported that Telmisartan exerts anti-tumor effects by activating peroxisome proliferator-activated receptor-γ (Li et al. 2014; Pu et al. 2016; Wu et al. 2016b). The algorithm presented here provides the first evidence from a randomized clinical trial, indicating that Telmisartan may be viable as a repurposed prevention for colon cancer.

Phylloquinone (Table 3, Row 4) is a vitamin (vitamin K1) supplement rather than a prescription drug. K vitamins + sorafenib induce apoptosis in human pancreatic cancer cell lines (Wei et al. 2010). A prospective cohort analysis found that individuals who increased their intake of dietary phylloquinone might have a lower risk of cancer than those who did not (Juanola-Falgarona et al. 2014). The data from the randomized trial in Table 3 suggest that vitamin K1 might actually help prevent cancer (OR=0.27, 95% CI=0.07-0.98). The potential cancer prevention by vitamin K1 is especially intriguing because one can get more than 1000% daily value of vitamin K1 by

simply eating one cup of cooked kale or spinach (https://www.healthaliciousness.com/articles/food-sources-of-vitamin-k.php).

The clinical trial in Table 3, row 6 tested Aliskiren for cardiovascular and renal disease in patients with type 2 diabetes. The SAE data from this study show that only 1 out of 4272 patients in the Aliskiren arm reported gastric cancer versus 8 out of 4285 patients in the placebo arm. A recent paper described that Aliskiren inhibits renal carcinoma cell lines proliferation *in vitro* (Hu et al. 2015). The data from this randomized clinical trial suggest the possible repurposing of Aliskiren for cancer.

Lastly, our literature search found no direct link between Vorapaxar (Table 3, Row 3) or Clopidogrel (Table 3, Row 5) and cancer prevention or treatment. Thus, these data in Table 3 could be the first sign that Vorapaxar or Clopidogrel might be useful for cancer or could be interpreted as false positive findings since we have made no attempt to adjust the multiplicity (multiple comparisons) in this exploratory analysis.

Above are only six outputs from our repositioning algorithm for one type of disease. The method described here could be used to identify other candidates for repositioning on any diseases that are reported as serious adverse events in ClinicalTrials.gov.

Discussion

Presented here is a novel drug repositioning method that reveals potential new uses of existing drugs directly from clinical trial data. This article provides only a rudimentary way to conduct drug repositioning using text mining tools on ClinicalTrials.gov. However, it could serve to stimulate other investigational initiatives to use clinical data to repurpose drugs, supplements, or even food to help prevent or treat diseases.

Serious adverse event data from randomized trials in the ClinicalTrials.gov were used because randomized trials are controlled experiments. However, ClinicalTrials.gov is only a tiny part of clinical data that could lead to the discovery of new use of existing drugs. Electronic medical record databases have much more clinical data than ClinicalTrials.gov. Other large sources of clinical data include Federal Adverse Event Reporting System and social media (Nugent et al. 2016). These data could provide new information not only on marketed drugs, but also on supplements and food.

Compared to traditional drug development, repositioned drugs have the advantage of decreased development time and costs given that significant toxicology and safety data will have already been accumulated, drastically reducing the risk of attrition during the drug discovery and development process.

The method we described here could merely help identify possible new uses of existing drugs to be investigated further. Prospective clinical trials would be required to provide the necessary evidence to have such new uses approved by regulatory agencies.

Conclusions

The rapidly growing clinical data could be extracted and analyzed for drug repositioning utilizing text mining tools. Repositioning non-cancer drugs with low toxicity or even vitamin supplements for cancer might provide tangible benefits for patients.

Acknowledgements

The authors are grateful for the statistical review by Dr. Margaret Gamalo, the expert editing by Drs. Marjo Gazak and Jane Reed, and insightful advice and review by Dr. Stephen Ruberg

References

Ashburn TT, and Thor KB. 2004. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3:673-683.

Cormack J, Nath C, Milward D, Raja K, and Jonnalagadda SR. 2015. Agile text mining for the 2014 i2b2/UTHealth Cardiac risk factors challenge. *J Biomed Inform* 58 Suppl:S120-127. 10.1016/j.jbi.2015.06.030

Firnhaber C, Zungu K, Levin S, Michelow P, Montaner LJ, McPhail P, Williamson AL, Allan BR, Van der Horst C, Rinas A, and Sanne I. 2009. Diverse and high prevalence of human papillomavirus associated with a significant high rate of cervical dysplasia in human immunodeficiency virus-infected women in Johannesburg, South Africa. *Acta Cytol* 53:10-17.

Hodos RA, Kidd BA, Shameer K, Readhead BP, and Dudley JT. 2016. In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip Rev Syst Biol Med* 8:186-210. 10.1002/wsbm.1337

Hu J, Zhang LC, Song X, Lu JR, and Jin Z. 2015. KRT6 interacting with notch1 contributes to progression of renal cell carcinoma, and aliskiren inhibits renal carcinoma cell lines proliferation in vitro. *Int J Clin Exp Pathol* 8:9182-9188.

Juanola-Falgarona M, Salas-Salvado J, Martinez-Gonzalez MA, Corella D, Estruch R, Ros E, Fito M, Aros F, Gomez-Gracia E, Fiol M, Lapetra J, Basora J, Lamuela-Raventos RM, Serra-Majem L, Pinto X, Munoz MA, Ruiz-Gutierrez V, Fernandez-Ballart J, and Bullo M. 2014. Dietary intake of vitamin K is inversely associated with mortality risk. *J Nutr* 144:743-750. 10.3945/jn.113.187740

Kesic V, Petkovic S, and Milacic D. 1990. [Smoking and nonmalignant changes in the uterine cervix]. *Srp Arh Celok Lek* 118:201-204.

Li J, Chen L, Yu P, Liu B, Zhu J, and Yang Y. 2014. Telmisartan exerts anti-tumor effects by activating peroxisome proliferator-activated receptor-gamma in human lung adenocarcinoma A549 cells. *Molecules* 19:2862-2876. 10.3390/molecules19032862

Mullen J, Cockell SJ, Tipney H, Woollard PM, and Wipat A. 2016. Mining integrated semantic networks for drug repositioning opportunities. *PeerJ* 4:e1558. 10.7717/peerj.1558

Nugent T, Plachouras V, and Leidner JL. 2016. Computational drug repositioning based on side-effects mined from social media. *PeerJ Computer Science* 2:e46. 10.7717/peerj-cs.46

Pu Z, Zhu M, and Kong F. 2016. Telmisartan prevents proliferation and promotes apoptosis of human ovarian cancer cells through upregulating PPARgamma and downregulating MMP9 expression. *Mol Med Rep* 13:555-559. 10.3892/mmr.2015.4512

Qosa H, Mohamed LA, Al Rihani SB, Batarseh YS, Duong QV, Keller JN, and Kaddoumi A. 2016. High-Throughput Screening for Identification of Blood-Brain Barrier Integrity Enhancers: A Drug Repurposing Opportunity to Rectify Vascular Amyloid Toxicity. *J Alzheimers Dis*. 10.3233/jad-151179

Sheskin J. 1965. Thalidomide in the Treatment of Lepra Reactions. *Clin Pharmacol Ther* 6:303-306.

Wei G, Wang M, and Carr BI. 2010. Sorafenib combined vitamin K induces apoptosis in human pancreatic cancer cell lines through RAF/MEK/ERK and c-Jun NH2-terminal kinase pathways. *J Cell Physiol* 224:112-119. 10.1002/jcp.22099

Wu CH, Bai LY, Tsai MH, Chu PC, Chiu CF, Chen MY, Chiu SJ, Chiang JH, and Weng JR. 2016a. Pharmacological exploitation of the phenothiazine antipsychotics to develop novel antitumor agents-A drug repurposing strategy. *Sci Rep* 6:27540. 10.1038/srep27540

Wu TT, Niu HS, Chen LJ, Cheng JT, and Tong YC. 2016b. Increase of human prostate cancer cell (DU145) apoptosis by telmisartan through PPAR-delta pathway. *Eur J Pharmacol* 775:35-42. 10.1016/j.ejphar.2016.02.017

Zheng C, Guo Z, Huang C, Wu Z, Li Y, Chen X, Fu Y, Ru J, Ali Shar P, Wang Y, and Wang Y. 2015. Large-scale Direct Targeting for Drug Repositioning and Discovery. *Sci Rep* 5:11970. 10.1038/srep11970