Characterizing drug-related adverse events by joint analysis of biomedical and genomic data: A case study of drug-induced pulmonary fibrosis

Alex Jiang[1,2], Anil G. Jegga[1,3,4]
[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [2]Cornell University, Ithaca, New York, USA; [3]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; [4]Department of Computer Science, University of Cincinnati College of Engineering, Cincinnati, Ohio, USA

Abstract
Spontaneous reporting systems such as the FDA's adverse event reporting system (FAERS) present a great resource to mine for and analyze real-world medication usage. Our study is based on a central premise that FAERS captures unsuspected drug-related adverse events (AEs). Since drug-related AEs result for several reasons, no single approach will be able to predict the entire gamut of AEs. A fundamental premise of systems biology is that a full understanding of a biological process or phenotype (e.g., drug-related AE) requires that all the individual elements be studied in conjunction with one another. We therefore hypothesize that integrative analysis of FAERS-based drug-related AEs with the transcriptional signatures from disease models and drug treatments can lead to the generation of unbiased hypotheses for drug-induced AE-modulating mechanisms of action as well as drug combinations that may target those mechanisms. We test this hypothesis using drug-induced pulmonary fibrosis (DIPF) as a proof-of-concept study.

Introduction

In the United States, more than 2 million cases of prescription drug-related adverse events (AE) occur annually, including 100,000 deaths. Spontaneous reporting systems such as the FDA's adverse event reporting system (FAERS) present a great resource to mine for and analyze real-world medication usage (1-5). FAERS represents real-world iatrogenic observations which can be clinically validated (2, 6) and indeed has been instrumental in the identification of serious drug-related AEs leading to either modification of the product labeling information or market withdrawal of an approved drug (7). FAERS data mining has also been shown to be successful in reproducing well-established clinical associations (8). Our hypothesis therefore is that systematic deep-mining of the FAERS data can identify novel drug pairs that either aggravate or reduce the risk of drug-induced disease (DID). Any clinical phenotype, including the drug-related AEs, are a result of perturbations of complex biological interactions. As a result, unimodal approaches may not be effective in predicting or in understanding the molecular basis of the entire gamut of drug-related or drug-induced AEs. Therefore, systems biology-based approaches that allow integration of multiple heterogeneous types of data sets can provide an overarching framework to explore the different types of drug-related AEs. Further, studying drugs in the context of cellular networks can provide insights into AEs caused by off-targets of those drugs. We therefore hypothesize that integrative analysis of drug-related AEs from FAERS with the differential transcriptome data sets from small molecules and disease phenotype (human patients and model systems) can lead to generation of unbiased hypotheses for DID-mitigating drug combinations and mechanisms of action.

Methods

AERS*Mine* - FAERS datamining

We used AERSMine (4) to mine FAERS and identify drugs that are significantly associated with a side-effect. Relative risk was calculated using a ratio of the rate of AE occurrence in patients given a certain drug or drug combination over the rate of the same AE occurrence in patients not given those drugs. Therefore, a higher relative risk suggests a more likely association between a given drug and AE, with 0 being the lowest possible. We also used safety signal, a Bayesian probability-based measure to analyze the relationship between a given drug and AE. Positive safety signals represent a positive correlation, negative scores denote a negative correlation, and scores close to zero signify independence (9). Safety signal was slightly preferred over relative risk, as a low relative risk score is ambiguous, potentially signifying either no relationship or an inverse relationship between drug or drug combination and AE. All AERSMine queries were performed using a standard Benjamini and Hochberg False Discovery Rate (FDR) correction (adjusted p-value < 0.05).

Differentially expressed gene signatures for phenotypes and drugs

For drug and disease transcription profiles we used curated data sets, including those from NCBI's GEO database as available from Illumina's BaseSpace Correlation Engine (BSCE; http://www.nextbio.com/b/nextbio.nb; Illumina, Cupertino, CA, USA).

Gene annotations and functional enrichment analysis
We used ToppGene Suite (TGS) Knowledgebase (10) for compiling the fibrosis gene sets and for performing the functional enrichment analysis.

Results
Case Study: Drug-induced pulmonary disease (DIPD)
Drug-induced pulmonary disease (DIPD) is a serious but relatively understated risk for millions of patients in the US and globally. Previous research shows that more than 600 FDA-approved drugs may cause DIPD. By one estimate, as many as 10% of chemotherapy patients develop DIPD (11). Other classes of DIPD-implicated drugs include cardiovascular medications, anti-microbial, and anti-inflammatory drugs. Currently, no alternatives exist for patients subjected to chemotherapeutics-induced pulmonary toxicity because the cause of DIPD is the primary medication. On the other hand, treatment of complex multifactorial disorders may lead to inadvertent polypharmacy predisposing to drug-interaction induced pulmonary disease. Thus, understanding the molecular mechanisms underlying DIPD and finding safer and effective therapeutic regimens is of paramount importance. As outlined in the methods, we used the drug-related AE data from the FAERS and differential transcriptome data sets for pulmonary diseases and drugs from the BSCE compendia as the basis for computational models that integrate network analyses with systems biology approaches to find and characterize drug combinations that can reduce the occurrence of DIPD, specifically drug-induced pulmonary fibrosis (DIPF).

DIPF candidate causal drugs
To identify DIPF candidate causal drugs we queried FAERS using AERSMine to get a list of drugs with the greatest safety signal scores (≥2.0) for AE pulmonary fibrosis. We also searched AERSMine with known DIPF-causal drugs and found several of them to have either relative risk >2.0 or safety signal >0. For example, bleomycin, amiodarone, and gefitinib – all well-known DIPF-causal drugs had safety signal >2.0 (Table 1). To identify potentially novel DIPF-

Table 1: Causal Drugs

Class	Drug	Relative Risk	Safety Signal	Pneumotox	Drug Label Warnings
Known Causal	Bleomycin	35.421	4.977	Yes	Yes, Boxed
	Methotrexate	4.675	1.933	Yes	Yes, Boxed
	Amiodarone	16.823	3.820	Yes	Yes, Boxed
	Gemcitabine	3.280	1.549	Yes	Yes
	Gefitinib	5.068	2.196	Yes	Yes
	Docetaxel	3.761	1.743	Yes	Yes**
	Paclitaxel	**2.224**	**0.988**	**Yes**	**Yes****
	Oxaliplatin	4.989	2.149	Yes	Yes
	Leflunomide	4.873	2.112	Yes*	Yes
	Cyclophosphamide	4.008	1.806	Yes	Yes
	Lansoprazole	2.250	0.992	ILD	Yes
	Nitroglycerin	2.224	0.993	No	No
Chemo Agent	**Rituximab**	**4.292**	**1.905**	**Yes**	**No**
	Doxorubicin	3.899	1.780	Yes	No
	Vinblastine	3.316	1.575	ILD	No
	Vincristine	7.059	2.174	ILD	No
	Vinorelbine	3.937	1.823	ILD	Yes
	Tamoxifen	3.301	1.566	Yes	No
	Cytarabine	1.260	0.183	N/A	Yes
Arthritis	Sulfasalazine	4.581	2.022	Yes	Yes
Immune System	**Rituximab**	**4.292**	**1.905**	**Yes**	**No**
	Thalidomide	2.161	0.950	ILD	No
	Tocilizumab	3.216	1.529	Yes*	Yes+
Other	**Nitrofurantoin**	**9.878**	**3.140**	**Yes**	**Yes**
	Dronedarone	4.952	2.156	ILD	Yes
	Cimetidine	3.406	1.614	?	No
	Acetylcysteine	7.527	2.762	Yes*	No
	Propoxyphene=	3.679	1.712	ILD	No

All label warnings for pulmonary fibrosis; *- linked to worsening of existing fibrosis; **- warning found in "Adverse Reactions" and not "Warnings and Precautions"; +- when taken with causal drug methotrexate; =- withdrawn from market for heart problems; N/A- compound not in database; ?- marked as "questionable signal"; "ILD"- linked to interstitial lung diseases but not specifically fibrosis

causal drugs, we repeated the AERSMine query excluding patients who were on known DIPF-causal drugs. The relative risk and safety signal scores from these queries are shown in the table 1. An important step in our approach was to exclude patients with any history of pre-existing respiratory disorders. By doing so, our analysis was not confounded by lung-related AEs resulting from exacerbation of the underlying clinical condition. We also used the Pneumotox database (http://www.pneumotox.com), which tracks and stores drug-induced and iatrogenic respiratory disease. Additional confirmation of the causal candidates was accomplished by comparing our findings with the FDA drug label warnings (DailyMed) and with data from the Canada Vigilance Adverse Reaction Online Database (CVAROD - http://webprod3.hc-sc.gc.ca/arquery-rechercheei/index-eng.jsp). Table 1 shows the top DIPF candidate causal drugs, including both novel suspects and those previously reported in FDA label warnings or clinical reports in the literature.

DIPF candidate therapeutics

Following a similar approach as described in the previous section, we compiled a list of pulmonary fibrosis-mitigating DIPF candidate therapeutics from AERSMine that had safety signal scores <0. This resulted in several candidate drugs. Interestingly, among these were drugs that have been previously reported to have anti-fibrotic effects. We checked Pneumotox database and FDA Label warnings to ascertain that the DIPF candidate therapeutics we have identified are not reported to cause pulmonary fibrosis. Among the DIPF candidate therapeutics were antidiabetic (sitagliptin, linagliptin, liraglutide, and canagliflozin) and antipsychotic medications (ziprasidone, risperidone, and paliperidone). Linagliptin is reported to attenuate pulmonary (12) and kidney fibrosis (13). Among the drugs under antipsychotics class, ziprasidone, risperidone, and paliperidone show significantly reduced risk for DIPF (Table 2).

Table 2: Therapeutic Drugs

Class	Drug	Relative Risk	Relative Risk w/ Causal*	Safety Signal	Safety Signal w/ Causal*	Cases/Patient w/ Causal*
Neural Drugs	Ziprasidone	0.095	N/A	-3.393	N/A	0/609
	Risperidone	0.614	0.669	-0.697	-5.051	2/2698
	Paliperidone	0.027	N/A	-5.221	N/A	0/129
Diabetes	Linagliptin	0.114	N/A	-3.130	N/A	0/597
	Liraglutide	0.065	1.259	-3.937	-5.201	1/717
	Canagliflozin	0.126	N/A	-2.983	N/A	0/297
Osteoporosis	Denosumab	0.630	1.631	-0.664	-2.784	6/3321
HIV	Efavirenz	0.051	N/A	-4.294	N/A	0/405
	Didanosine	0.124	N/A	-3.006	N/A	0/139
Other	Levonorgestrel	0.019	N/A	-5.668	N/A	0/724
	Fampridine	0.075	2.015	-3.737	-4.626	1/4485
	DMF	0.044	N/A	-4.483	N/A	0/379
	Vardenafil	0.100	N/A	-3.327	N/A	0/302
	Orlistat	0.077	1.971	-3.698	-4.652	1/458
	Lamivudine	0.379	0.651	-1.394	-6.064	1/1386
	Inter. beta 1a	0.445	1.216	-1.152	-4.229	2/1485
	Clonidine	0.848	2.471	-0.239	-1.757	13/4752

*- patients taking candidate w/ at least 1 known causal drug, ex. 609 such patients for ziprasidone; N/A- no cases in cohort

DIPF- mitigating or aggravating drugs – Mechanistic hypotheses

We next tested the hypothesis that integrative analysis of drug-related AEs with drug-induced transcriptomics data can lead to generation of mechanistic hypotheses about plausible causes, and intervention, for DIPFs. To delineate this, we leverage public gene annotation data sets normalized and integrated as part of our ToppGene Suite (TGS) Knowledgebase (10) and BSCE (see Methods). To do this using TGS Knowledgebase, we first compiled a list of 782 pulmonary fibrosis-related genes based on human and mouse phenotype gene associations as well as known pulmonary fibrosis genes from DisGeNet and genetic markers of pulmonary fibrosis from BSCE. We then used BSCE to compare these 782 genes against the BSCE compendia of differential transcriptomes. We filtered the results for differential transcriptomes from bleomycin mouse models of pulmonary fibrosis. Among the 124 intersecting genes, 85 genes were significantly overexpressed in the lungs of bleomycin mouse model of fibrosis (Fig. 1).

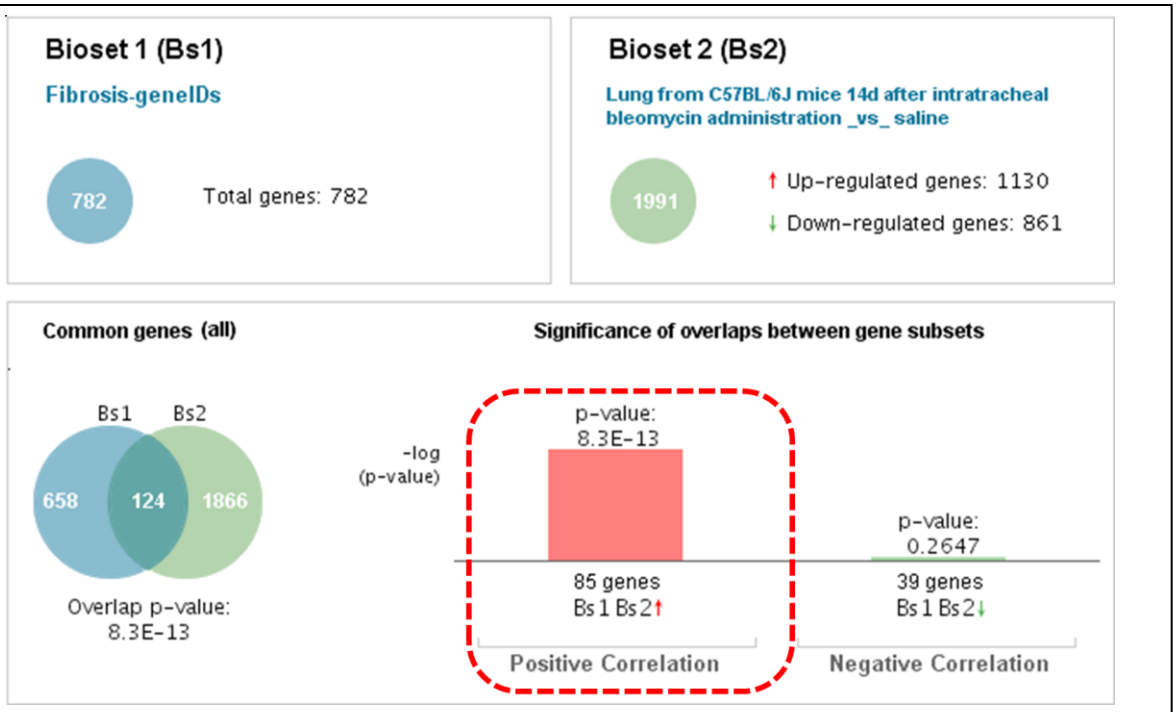

Fig. 1. BSCE was used to compare the 782 fibrosis gene set with the differential transcriptome of bleomycin mouse models of pulmonary fibrosis (BSCE curated data sets including those from NCBI's GEO database). Of the 124 intersecting genes, 85 genes (red box) were significantly upregulated in mouse models of pulmonary fibrosis. The numbers do not add up accurately because of some genes occurring in both up- and down- regulated gene sets owing to differential expression of different probe sets mapped to a same gene.

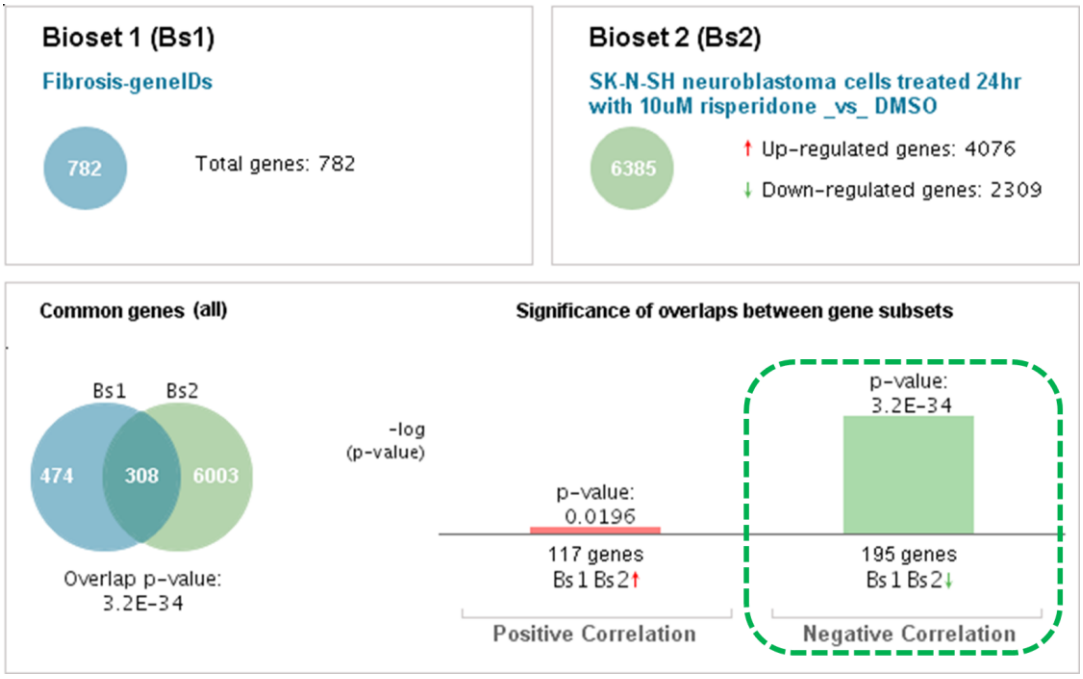

Fig. 2. Comparison of the 782 fibrosis gene set with the differential transcriptome of risperidone treatment. Of the 308 intersecting genes, 195 genes (green box) were significantly downregulated in risperidone study. The numbers do not add up accurately because of some genes occurring in both up- and down- regulated gene sets owing to differential expression of different probe sets mapped to a same gene.

Next, we compared the 782 fibrosis genes with compendia of differential transcriptomes (using BSCE),

specifically with risperidone (DIPF candidate therapeutic) related data sets. Interestingly, we found a significant negative correlation. Of the 308 intersecting genes, 195 genes were significantly downregulated following the risperidone treatment (Fig. 2).

Finally, using BSCE's meta-analysis, we directly compared the differentially expressed pulmonary fibrosis mouse model data set with that of the risperidone differential transcriptome. This resulted in an overlap of 789 genes, of which 263 and 202 genes were reciprocally related. In other words, 263 genes upregulated in pulmonary fibrosis mouse model were downregulated in the risperidone data set while 202 genes downregulated in pulmonary fibrosis mouse model were upregulated in the risperidone differential transcriptome data set (Fig. 3). Functional analysis of

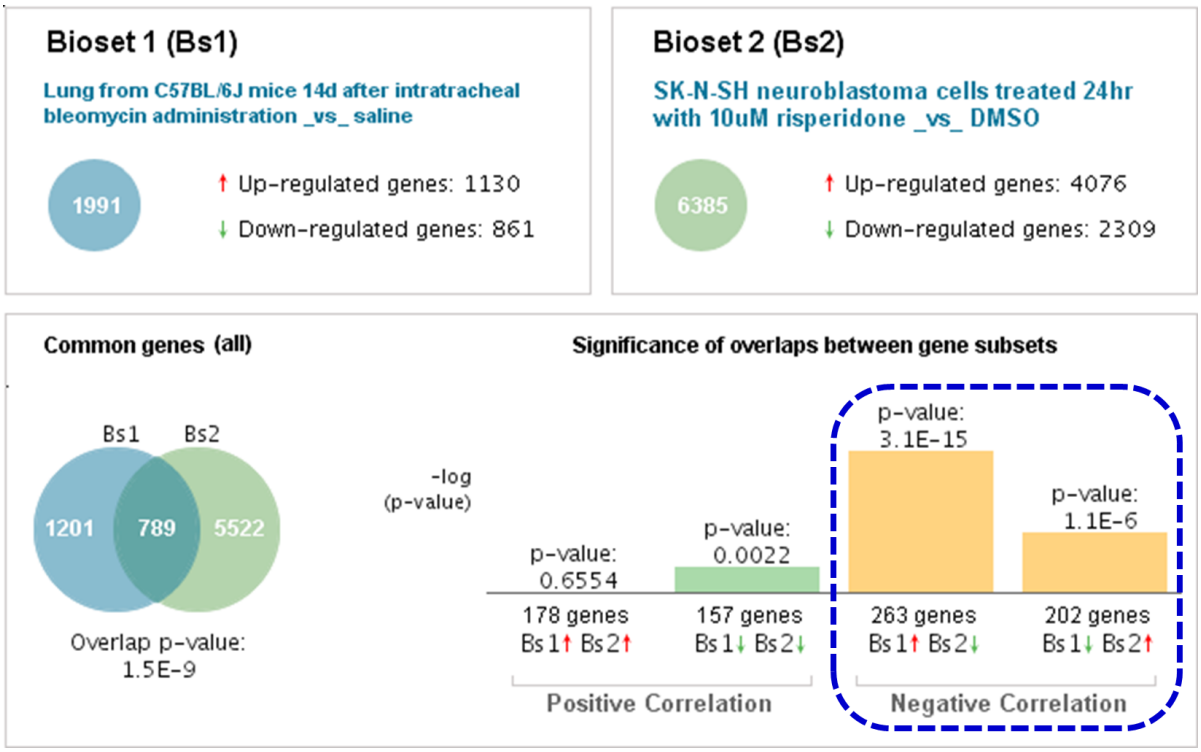

Fig. 3. Correlation of the differential transcriptomes of mouse model for pulmonary fibrosis and risperidone treatment. Of the 789 intersecting genes, 465 (263 + 202) genes (blue box) are reciprocally regulated. The numbers do not add up accurately because of some genes occurring in both up- and down- regulated gene sets owing to differential expression of different probe sets mapped to a same gene.

the 263 genes (upregulated in bleomycin mouse model of pulmonary fibrosis but downregulated following risperidone treatment) using ToppFun application of TGS showed enrichment (FDR p-value 0.05) for pathways and biological processes such as ECM, fibroblast proliferation and migration, collagen binding, abnormal alveolar morphology, etc. – all hallmarks of pulmonary fibrosis.

Discussion

Spontaneous reporting systems such as FAERS present a great resource to mine for and analyze real-world concomitant medication usage. FAERS data mining has been shown to be successful in reproducing well-established clinical associations (e.g., statins and muscular events, oxaliplatin and peripheral sensory neuropathy, proton pump inhibitors and hypomagnesaemia, etc. (14)). Further, even though polypharmacy – the concurrent use of multiple medications – has been shown to have both positive and negative outcomes, there have been no studies to find and investigate the role of polypharmacy in DIPD systematically. Polypharmacy is ubiquitous – an estimated 29% of the elderly patients use at least 5 prescription medications while 46% of prescription users also used over-the-counter medications (15).

Similar to any phenotype, drug-related AEs or phenotypes also result because of several, and largely unknown and yet to be discovered, genetic and environmental components, and their interactions. Therefore, any single approach will fail to predict the entire range of drug-induced AEs. Systems biology-based integrative

approaches that permit joint analysis of individual heterogeneous elements and their interactions can enable a relatively complete understanding of the underlying molecular basis of drug-induced AESs. Further, studying the drugs in the context of cellular networks can provide insights into AEs caused by off-targets of drugs (16, 17).

Our study has certain limitations. Since FAERS is a spontaneous reporting system, apart from the data quality-related issues, there can be potential reporting biases. Confounding (18-21), for instance, impacts understanding true correlations and presents a significant challenge for drug-AE hypothesis generation. To limit the effects of confounders, *a priori* clinical knowledge can be applied to exclude known confounders. However, confounders are not always known beforehand. Hence, automatic confounder control methods (22) based on propensity scores, direct adjustment, similarity matching and ensemble resampling can assist in mitigating the effects of unknown confounders. In addition, clinical data from EHRs can serve as a "gold standard" while also complementing and strengthening drug-related AE hypotheses. Last but not least, FAERS data lacks the denominators (i.e., total number of patients using the drug globally but did not experience an AE) to estimate the true attributable risks. In the current study, we relied on BSCE for drug and disease phenotype transcription profiles. However, there is a possibility that some of the drugs we discover to be potentially increasing or decreasing DIPF risk may not be represented in BSCE.

Conclusion

Delineating and characterizing the AE-mitigating effects of drugs and drug combinations through these systematic pharmacological approaches has a direct impact on precision medicine when combined with genomic sequencing and electronic medical records in clinical settings. Although we have focused on DIPF in this current study, the same methodology can be extended for elucidating other drug-induced system-wide AEs. As part of ongoing and future studies, we are focusing on identifying drug combinations that are potential DIPF risk modifiers. Using the DIPF candidate causal (DIPF+ve) and therapeutic (DIPF-ve) drugs identified, we will generate pairwise combinations (DIPF modifier matrices) and re-query FAERS data using AERSMine as described previously for the incidence of DIPF in patients on these combinations. In other words, if we have *m* DIPF+ve drugs and *n* DIPF-ve drugs, we will generate *m* X *n* combinations to facilitate an unprecedented deep-dive into the realm of DIPF therapeutics.

Acknowledgement: This research was supported in part by the NIH NHLBI's 1R21 HL135368. We thank Dr. Mayur Sarangdhar, Cincinnati Children's Hospital Medical Center, for assistance with AERSMine usage and for useful discussions during the analysis and manuscript preparation.

References

1. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data. 2016;3:160026.
2. Tatonetti NP, Denny JC, Murphy SN, Fernald GH, Krishnan G, Castro V, et al. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. Clin Pharmacol Ther. 2011;90(1):133-42.
3. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012;4(125):125ra31.
4. Sarangdhar M, Tabar S, Schmidt C, Kushwaha A, Shah K, Dahlquist JE, et al. Data mining differential clinical outcomes associated with drug regimens using adverse event reporting data. Nature Biotechnology. 2016;34(7):697-700.
5. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-8.
6. Sarangdhar M, Gruenzel AR, Yacyshyn MB, Aronow BJ, Yacyshyn B. Lower Expression of PPAR. in AVM Tissue Provides a Novel Therapeutic Target for Use of Glitazones in Treatment of GI Bleeding: Validation of Our Large-Scale Clinical Effects Data Analysis. GASTROENTEROLOGY. 2015;148(4):S642-S.
7. Wysowski DK, Swartz L. Adverse drug event surveillance and drug withdrawals in the United States, 1969-2002: the importance of reporting suspected reactions. Archives of internal medicine. 2005;165(12):1363-9.
8. Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. International journal of medical sciences. 2013;10(7):796-803.
9. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol. 1998;54(4):315-21.

10. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research. 2009;37(Web Server issue):W305-11.
11. Coultas DB, Zumwalt RE, Black WC, Sobonya RE. The epidemiology of interstitial lung diseases. Am J Respir Crit Care Med. 1994;150(4):967-72.
12. Kabel AM, Abd Elmaaboud MA, Atef A, Baali MH. Ameliorative potential of linagliptin and/or calcipotriol on bleomycin-induced lung fibrosis: In vivo and in vitro study. Environ Toxicol Pharmacol. 2017;50:216-26.
13. Kanasaki K, Shi S, Kanasaki M, He J, Nagai T, Nakamura Y, et al. Linagliptin-mediated DPP-4 inhibition ameliorates kidney fibrosis in streptozotocin-induced diabetic mice by inhibiting endothelial-to-mesenchymal transition in a therapeutic regimen. Diabetes. 2014;63(6):2120-31.
14. Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. Int J Med Sci. 2013;10(7):796-803.
15. Qato DM, Alexander GC, Conti RM, Johnson M, Schumm P, Lindau ST. Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States. JAMA. 2008;300(24):2867-78.
16. Berger SI, Iyengar R. Role of systems pharmacology in understanding drug adverse events. Wiley Interdiscip Rev Syst Biol Med. 2011;3(2):129-35.
17. Berger SI, Iyengar R. Network analyses in systems pharmacology. Bioinformatics. 2009;25(19):2466-72.
18. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clinical pharmacology and therapeutics. 2013;93(6):539-46.
19. Hauben M, Hung EY. Pneumothorax as an adverse drug event: an exploratory aggregate analysis of the US FDA AERS database including a confounding by indication analysis inspired by Cornfield's condition. International journal of medical sciences. 2013;10(8):965-73.
20. Napoli AA, Wood JJ, Coumbis JJ, Soitkar AM, Seekins DW, Tilson HH. No evident association between efavirenz use and suicidality was identified from a disproportionality analysis using the FAERS database. Journal of the International AIDS Society. 2014;17:19214-.
21. Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, et al. Systems pharmacology of adverse event mitigation by drug combinations. Science translational medicine. 2013;5(206):206ra140-206ra140.
22. Low YS, Gallego B, Shah NH. Comparing high-dimensional confounder control methods for rapid cohort studies from electronic health records. Journal of comparative effectiveness research. 2015.