Systematic Evaluation of Adverse Drug Reactions and Drug Repurposing Candidates

By

Patrick Wu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Informatics

May 14, 2021

Nashville, Tennessee

Approved:

Cynthia Gadd, Ph.D., M.B.A.

Wei-Qi Wei, M.D., Ph.D.

QiPing Feng, Ph.D.

Scott D. Nelson, Pharm.D., M.S.

Qingxia (Cindy) Chen, Ph.D.

Joshua C. Denny, M.D., M.S.

For my parents and brother

ACKNOWLEDGEMENTS

This work would not have been possible without the generous support of the Vanderbilt University Medical Scientist Training Program, School of Medicine, and Department of Biomedical Informatics. This work was also supported by multiple grants from the National Institutes of Health (T32 GM007347, T15 LM007450, P50 GM115305).

I am grateful for all of the people whom I have worked with throughout my research career over the past ten years. Members of the Huffaker Lab at Cornell University were instrumental in helping me develop a passion for research and continued interest in basic cell biology. Members of the McGavern Lab at the National Institutes of Health helped me appreciate the dynamic nature of the immune system. I am also grateful for the opportunity to have learned from my professors, many mentors, and classmates in DBMI. Thank you for the thought-provoking discussions. It has been a pleasure working with the members of the Center for Precision Medicine and the Wei Lab.

For the work presented in this dissertation, I would like to thank my committee members, Drs. Cindy Gadd, QiPing Feng, Scott Nelson, Cindy Chen, and Joshua Denny for providing extensive professional guidance and teaching me valuable lessons about scientific research. I would like to especially thank my dissertation mentor, Dr. Wei-Qi Wei, for his generosity, patience, and guidance in research direction.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 3

Chapter 4

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

INTRODUCTION

Definitions of “Biomedical Informatics” include (1) “the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving and decision making, motivated by efforts to improve human health”[1] and (2) “[the translation of] data to knowledge and discovery”.[2] Over the past decade, the amount of clinical data in electronic health records (EHRs) and human genetic data has grown exponentially.

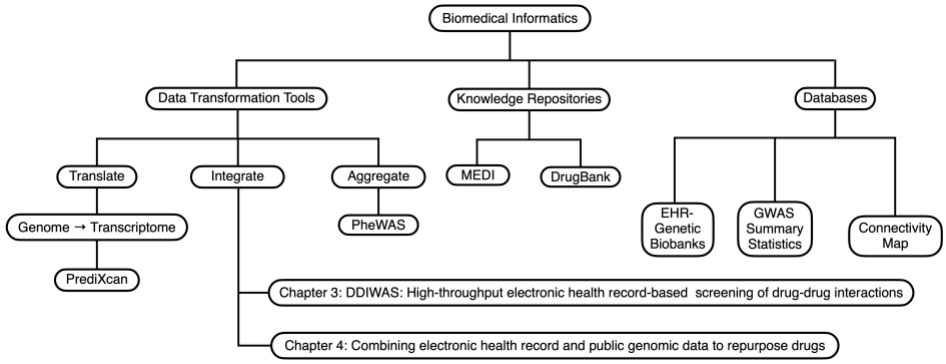


Figure 1. Overview of dissertation projects in the context of Biomedical Informatics.
Shows the data transformation tools, knowledge repositories, and databases used in the development of new methods for using data, to answer questions in biomedicine. EHR: electronic health record; GWAS: genome-wide association study; PheWAS: phenome-wide association study; MEDI: MEDication Indication Resource.

In the US, the Affordable Care Act increased the adoption of EHRs by hospitals across the country. Increased use of EHRs generated large amounts of clinical data, but researchers did not have tools to repurpose the aggregated clinical data for biomedical research. In the late 2000s/early 2010s, researchers published the first studies showing that genomics research could be performed in EHR-linked DNA biobanks.[3] These biobanks connected genetic data to phenotype data contained in de-identified EHRs, allowing researchers to perform high-

throughput genotype-phenotype association studies in EHRs.[4,5] These studies motivated the creation of the Electronic Medical Records and Genomics (eMERGE) Network,[6] a trans-institutional consortium focused on developing, applying, and sharing approaches to combine clinical data in EHRs and biobanks for genomics research. The eMERGE network provided the opportunity for researchers to develop portable informatics methods, like phenotyping algorithms[7] to perform genetic association tests. In the 2010s, large national databases containing EHR data linked to genetic data were created, like the UK Biobank[8] and the National Institutes of Health (NIH) *All of Us* Research Program.[9] These publicly available databases contain unique data sources that not only allow genomics research using EHR data, but also allow biomedical research focused on answering questions outside the field of human genetics.

Schemas that have facilitated the secondary use of EHR data for biomedical research are Common Data Models (CDMs). An early challenge in developing portable methods for reusing EHR data for research, was that institutions used idiosyncratic terminologies to represent clinical data and used local models to store data. To address the problem, initiatives like Observational Health Data Sciences and Informatics (OHDSI) created and have continuously updated the Observational Medical Outcomes Partnership (OMOP) CDM, which allows data coming from disparate sources to be harmonized in a consistent and standardized manner.[10,11] The OHDSI/OMOP CDM defines the data types collected and how data is organized, so researchers can develop and use methods on data sources that have adopted the format.

In previous studies, researchers developed automated approaches to identify drug-drug interaction (DDI) signals[12] and to clinically validate drug repurposing signals[13] in EHRs. However, these approaches were often not portable, as the automated approaches relied upon bespoke natural language processing (NLP) tools tuned specifically for their database. Further, when those studies were conducted, EHR databases were only beginning to adopt CDMs. Since then, large de-identified EHR databases, like the Vanderbilt University Medical Center (VUMC) Synthetic Derivative (SD) and national databases, like the NIH *All of Us* Research Program, have adopted the OHDSI/OMOP CDM. This change provided the opportunity to develop portable methods to reuse EHR data for biomedical research.

In this dissertation, I describe the development and application of portable methods to repurpose EHR data for identifying DDI signals and validating drug repurposing candidates (Figure 1). In

Chapter 1, I discuss the motivation for my research. In Chapter 2, I provide the background for the dissertation. In Chapter 3, I describe the development and evaluation of a novel approach called drug-drug interaction wide association study (DDIWAS). DDIWAS uses signals from the EHR allergy list to identify clinically-significant DDIs. Using allergy list information makes DDIWAS portable, because like International Classification of Diseases (ICD) billing codes, the allergy list section is widely used, as the module was required for US health systems to meet Meaningful Use criteria. In Chapter 4, I describe the development of a portable approach to integrate human transcriptomic data, drug perturbation data, and clinical data in EHRs to identify and clinically validate drug repurposing candidates. In chapter 5, I summarize the work with a discussion of the limitations and future directions.

CHAPTER 2

BACKGROUND

Electronic Health Record Data

In the last two decades, a large amount of human biomedical data has been generated and stored. Such data have provided the opportunity for researchers to make new discoveries about human biology, discoveries leading to new treatments for human diseases. To make those discoveries, researchers need analytic tools to make those discoveries.

Since the year 2000, there has been a large growth in clinical data from EHRs and human genetics. These two domains once separate, have come together due to the development of EHRs linked to DNA biobanks.[3] Integration of EHR and human genetic data has allowed researchers to conduct studies once very difficult due the high costs associated with collecting clinical and genetic data.

When learning from data, standardized models for representing structured data and amount of data are important. The work in this dissertation, integrating clinical and genetic data, was only possible due to the standard terminologies developed prior to the start of this dissertation. These terminologies are the results of decades of public and private monetary investment and studies by researchers in the field of biomedical informatics. In this chapter, I provide background on the standardized vocabularies used in this dissertation and how data has been generated. First, I will focus on clinical data stored in EHRs, and then transition to genetic data. I will then describe the specific work that was done for my dissertation, providing the proper context in the areas of drug-drug interactions and drug repurposing in informatics.

Standardized terminologies in EHR data

When a patient visits their doctor or is admitted to the hospital, various data types are generated to document their care. For each patient, their visit generates information to document their conditions, drug exposures, and measurements performed to aid in diagnosing the patient or for preventive care. These data types are collected in the EHR and represented as machine-readable information made possible by the development of standard vocabularies. This work

primarily used the following standard vocabularies: RxNorm (drugs), International Classification of Diseases (ICD) codes (conditions), and LOINC codes (lab measurements) (Figure 1).

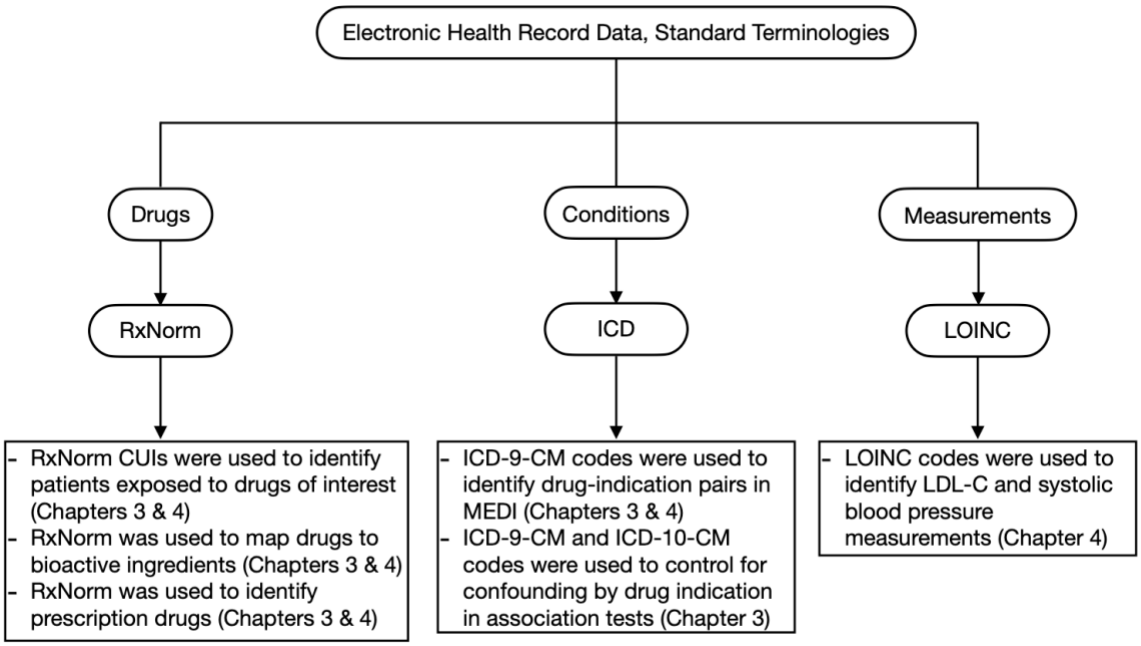


Figure 1. Overview of EHR data standard terminology used in studies.
CUI: Concept Unique Identifier; ICD: International Classification of Diseases; LOINC: Logical Observation Identifiers Names and Codes; LDL-C: low-density lipoprotein cholesterol.

Created in 2002, RxNorm is a terminology allowing semantic mapping of drugs.[14] With RxNorm, the same drug represented differently among sources are consolidated and mapped to RxNorm concepts represented by CUIs (RxCUIs). Thus, simvastatin, a common lipid-lowering drug different sources represented in the Anatomical Therapeutic Chemical (ATC) classification system[15] as C10AA01 and in DrugBank[16] as DB00641 both map to RXCUI 36567. In my chapters 3 and 4, I used RxNorm to map drugs to their active drug ingredients and to identify prescription drugs.

To define patient cohorts in EHRs, researchers have leveraged ICD codes. In the US, the two most recent iterations of the coding system are ICD-9-CM and ICD-10-CM.[17] In 1979, ICD-9 was developed by the World Health Organization (WHO) to track mortality and morbidity.

Subsequently, ICD-9 was modified by the US National Center for Health and Statistics (NCHS) to create ICD-9-CM, improving its application to clinical billing. In 1990, the WHO developed ICD-10,[18] which the NCHS used to replace ICD-9-CM, with ICD-10-CM. In EHR-based studies, ICD codes are commonly used in rule-based approaches to create disease cohorts, a process known as phenotyping.[19] Advantages of phenotyping algorithms based on ICD codes include portability, as ICD codes are used commonly across hospital systems, so data to identify patients with phenotypes of interest are widely available in EHR databases. Disadvantages of phenotyping based on ICD codes include its primary use for billing, as the specificity can differ depending on the phenotype of interest. However, combining ICD billing codes with other structured data like drugs have been shown to significantly improve the performance of phenotyping algorithms.[20] To meet Meaningful Use requirements, EHR systems were required to use Systematized Nomenclature of Medicine – Clinical Terminology (SNOMED-CT) to represent diagnoses in problem lists. SNOMED-CT is also the terminology chosen to represent diagnoses in the OHDSI/OMOP CDM because it is much more expressive and has an extensive hierarchical structure.[21] Diagnoses represented in SNOMED-CT can be mapped to ICD codes using existing publicly available mappings.[22] In Chapter 3, I use ICD codes to control for DDIs potentially confounded by indication. In both chapters, ICD-9-CM codes were used to identify drug-indication pairs, using the MEDI knowledge base.[23]

In 1994, Logical Observation Identifiers Names and Codes (LOINC) codes were created as a standard terminology for laboratory tests.[24] LOINC codes were necessary, as institutions used local, idiosyncratic systems to map their laboratory results, making it difficult to integrate laboratory data from different sites.[25] In chapter 4, I used LOINC codes to obtain low-density lipoprotein cholesterol (LDL-C) and systolic blood pressure measurements.

ICD codes, RxNorm CUIs, and LOINC codes are all integrated by the Unified Medical Language System (UMLS).[26] Started in 1990, the UMLS Metathesaurus by the US National Library of Medicine (NLM) allows semantic translation of biomedical concepts, thereby allowing researchers to integrate biomedical data from different resources, be it from the EHR or from the biomedical literature. Biomedical concepts are mapped to UMLS concepts, represented by CUIs. The most recent (2020AA) version of the UMLS contains 4.28 million concepts from 214 vocabularies.[27]

The OMOP/OHDSI CDM,[11,28] builds on the standard terminologies by providing a template to organize data in relational databases, thereby allowing query code written at one institution to be used, with minimal modifications, in external databases organized also using OMOP/OHDSI CDM. CDMs reduce potential biases in findings related to variability in data extraction and processing. The OMOP/OHDSI CDM allowed researchers to integrate data from eleven hospitals across four countries, to answer questions about how treatment pathways vary for type 2 diabetes, hypertension, and depression, using data from 250 million patients.[29] Integration of large data also allowed researchers to examine the effectiveness of second-line treatment options for type 2 diabetes, across a diverse cohort of 246 million patients.[30] Though there are other CDMs,[31] large-scale initiatives like the NIH *All of Us* Research Program have chosen to organize their data using the OMOP/OHDSI CDM.[9] In both chapters, I provide publicly available code templates to extract data from EHR databases organized using the OMOP/OHDSI CDM, thereby increasing the ability of other researchers to see whether findings described in these studies are replicable in their database. Publicly available code also allows them to build upon the tools developed in this study.

Large EHR data

The recent large growth in the amount of EHR data in the US has especially influenced the work in this dissertation. In 2008, only 7.6% of US hospitals had a basic EHR system.[32] In 2009, the US passed the Health Information Technology for Economic and Clinical Health (HITECH) Act. HITECH incentivized US hospitals to adopt EHR systems;[33,34] as a result, EHR use increased from 72% in 2011 to 96% in 2017 (Figure 2).[35] In the UK, a national EHR system was also implemented in the past decade.[36] This growth in EHR data has coincided with the development of the NIH *All of Us* Research Program, an initiative to create a large publicly available source of clinical data from a diversity cohort, linked to biobanks.[9]

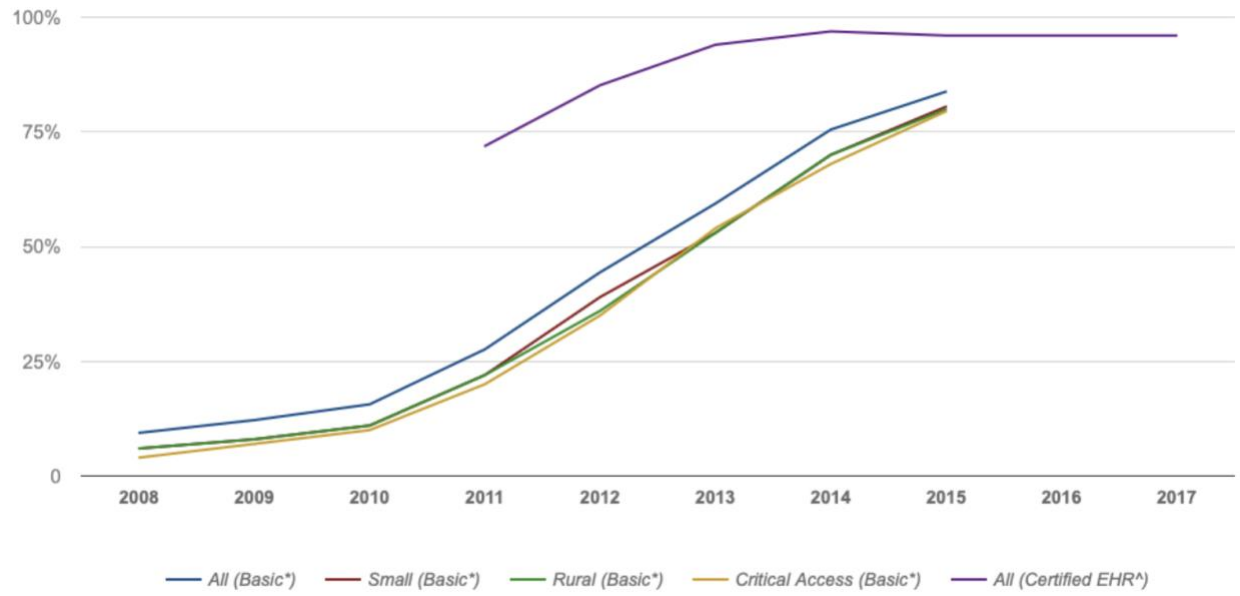


Figure 2. Between 2008-2017, EHR adoption in the US increased.
This figure was adapted from `https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php` . [37] EHR: electronic health record.

The exponential increase of EHR data allowed researchers to cost-effectively obtain phenotypic data from many patients. Hospitals use EHRs to document patient care, which facilitates healthcare provider communication and medical billing. Routinely collected healthcare data has been increasingly used with basic research to better understand disease pathophysiology,[38] and used as a source of Real-World Evidence (RWE) to support findings in clinical trial studies.[39] The amount of global EHR data available for biomedical research has grown at an exponential pace and will continue to accumulate in the future.[19,29]

Human Transcriptomic Data

In Chapter 4, I identified drug repurposing candidates using transcriptomic signature matching. Obtaining useful human transcriptomic data is difficult because of technological and ethical barriers. A major technological barrier to obtaining transcriptomic data at scale, is that technologies to measure gene expression, using methods like genome-wide gene expression microarrays and RNA-sequencing (RNA-seq) remains expensive.[40,41] Further, since gene expression differs from one tissue to another, even if a researcher had transcriptomic data from the blood of their patients of interest, knowing also how a gene is expressed in just the blood is

not enough, as tissue-specific expression is important to understanding disease biology.[42–44] Ethical and practical concerns preclude taking tissue biopsies from humans, like from the kidney or brain. In the past, researchers often used data from pre-clinical animal disease models. But, there are limitations to using animal disease models, since animal disease models often poorly simulate human diseases.[45,46]

Fortunately, researchers recognized these two limitations and have developed methods to estimate transcriptomic signatures for phenotypes of interest. In Chapter 4, I used one of these methods, called S-PrediXcan.[47,48] S-PrediXcan models were trained using the Genotype-Tissue Expression (GTEx) reference set.[49] S-PrediXcan estimates genetically-regulated transcriptomic signature for a phenotype of interest using just GWAS summary statistics, a widely available resource.

High-quality GWAS summary statistics used by S-PrediXcan to estimate phenotype transcriptomic signatures are only available due to single-nucleotide polymorphism (SNP) imputation. SNP imputation is possible due to the substantial investment and work in the creation of tools and resources to annotate the human genome. In most GWAS, genotypes are obtained using microarrays,[50–53] where only 300,000-500,000 SNPs are directly genotyped and used to infer missing genotypes through imputation, expanding the number of SNPs to >2.2 million.[54] SNP imputation is possible due to reference data sets from initiatives like the International HapMap Project,[55] 1000 Genomes Project,[56] Haplotype Reference Consortium[57], and gnomAD.[58]

The GTEx project was formed to better understand how genetic variants affect gene expression across tissues and how variation-mediated gene expression changes impact human disease. The project has allowed the research community to understand some of the associations between genetic variants and human disease.[59] The most recent version of GTEx has RNA-seq data linked to patient genotypes, across 49 different tissues.[60] For S-PrediXcan models, the SNP weights are calculated using an elastic net model, with the SNPs as independent variables, and gene expression as dependent variables.[47]

Drug Perturbation Data

In Chapter 4, I searched for drugs that reversed the S-PrediXcan estimated phenotype transcriptomic signatures using the Integrated Library of Integrated Network-Based Cellular

Signatures (iLINCS).[61] iLINCS is a data repository comprised of drug databases like ConnectivityMap (CMap),[62] L1000,[63] and DrugMatrix.[64] The data in these drug databases are drug-induced gene expression signatures. For CMap and L1000, gene expression signatures are from cancer cell lines treated *in vitro* with a library of drugs at varying concentrations and exposure times. For DrugMatrix, the gene expression signatures were obtained from primary non-human animal tissues, like rat liver induced by drug exposures.

Drug-Drug Interactions

Adverse DDIs cause patient harm and are responsible for a significant amount of withdrawn drugs. DDIs occur between an object drug (affected by the interaction) and precipitant drug (causes the interaction). There are two types of adverse DDIs: pharmacokinetic and pharmacodynamic. Pharmacokinetic DDIs are interactions that result in the concentration of the object drug. As an example, consider a patient taking both simvastatin and itraconazole. Itraconazole increases the concentration of simvastatin in the patient's blood, because it decreases its metabolism by inhibiting the enzyme CYP3A4.[65] Increased simvastatin concentration has been shown to increase the risk of adverse drug reactions (ADRs), like myopathy.[66] In addition to decreases in drug metabolism, other causes of increased drug concentration in the blood include increase in drug absorption and decrease in excretion. There are two general types of pharmacodynamic DDIs: (1) when one drug decreases the effects of a second drug on the body and (2) concurrent use of two drugs with additive effects (eg, drugs that may cause sedation, like cetirizine [an antihistamine] and amitriptyline [an antidepressant]).[67]

The US Food and Drug Administration (FDA) provides guidelines to drug sponsors and investigators for evaluating potential DDIs, starting first with *in vitro* studies, results that guide the *in vivo* studies that should be conducted.[68–70] Results from these studies are put on drug labels if the drug is approved. While these pre-market studies have decreased the incidence of DDIs, it is not feasible to test all the potential drug combinations that real-world patients use. To detect potential DDIs after a drug is approved, there are post-market surveillance systems, like the FDA Adverse Event Reporting System (FAERS). Systems like FAERS allow healthcare providers, companies, and patients to voluntarily report potential ADRs and DDIs. But, the voluntary reporting characteristic limits post-market surveillance systems to identify DDIs. EHR

data provide the opportunity to detect DDIs that may be missed in traditional pre-market studies and by post-market systems.

Existing methods to identify DDIs using EHR data

Since the availability of good quality and adequate amounts of longitudinal EHR data is very recent, researchers have developed methods to address key problems associated with using observational databases. These key problems are to limit the bias associated with confounding and detection of ADRs in EHRs.

To address confounding, researchers have used methods developed for observational analysis,[71] by calculating propensity scores to use for either matching, stratification, or use as covariates in regression models. The main problem is potential for false-positive findings due to unmeasured confounders. For example, researchers have evaluated different methods to control for unmeasured confounders in the evaluation of ADRs due to drug exposure.[72]

Researchers were initially focused on using EHR data to supplement existing post-market surveillance programs to monitor ADRs.[73] The basic study design is to detect drug exposure, which can be done using relatively structured data in patient medication lists. After drug exposed patients are found, the second task is to identify ADR events after patients are exposed to the drug(s) of interest. In EHRs, ADRs are detected using natural language processing (NLP) tools in unstructured clinical narratives,[74] which are then mapped to MedDRA terms.[73] Researchers then use statistical methods like disproportionality analysis to quantify the association between drug exposure and development of ADRs.[75] The comparison can be between two patient groups, where the cases could be patients exposed to the drug of interest and controls could be patients who were exposed to a drug with the same indication.

When researchers were able to demonstrate the use of EHR data for ADR surveillance, they then used EHR data to validate DDIs identified in post-market surveillance systems. Tatonetti et al developed a method to control for confounders in observational clinical data, like concomitant drugs and patient comorbidities.[76] In EHR data, they validated 47 drug class interactions identified using their method in FAERS. In a separate study, the same group also validated a novel DDI detected in FAERS, between paroxetine and pravastatin, which led to larger than expected increases in blood glucose levels in patients exposed concurrently to both drugs, compared to patients exposed to one of the drugs alone.[77] Interestingly, the glucose increase

was even more extreme in patients with diabetes. They also used EHR data to show a false-positive interaction between moxifloxacin and warfarin identified in FAERS, an interaction most likely confounded by kidney disease.[78]

Researchers have also developed methods to search the EHR to discover potentially novel DDIs. Iyer et al mined the EHR to validate and discover potential novel DDIs.[12] They used patient medication lists to infer drug exposure and developed a simple text annotation tool to search through the unstructured text in clinical narratives for ADRs, which were mapped to MedDRA terms. In their study, a potential DDI was defined as a *drug-drug-ADR.* They selected 14 ADRs to demonstrate their pipeline; the 14 ADRs were those that occurred frequently in their database. Potential DDIs were those for which when patients were exposed to two drugs concurrently, experienced higher odds of experiencing ADRs of interest compared to exposure to one of the drugs alone. They validated known DDIs and identified potential novel DDIs in two independent EHR databases.

Gap in knowledge for using EHR data to detect DDIs

No study has yet investigated the potential of information in EHR allergy lists to identify potential DDIs. Allergy lists are used to record both immune-related drug reactions and drug intolerance. These lists share a common semi-structured pattern across EHRs with two data types: a drug name (structured), and free text (unstructured) that allows healthcare providers to record the specific ADR experienced by the patient. Since allergy lists were required by EHR systems in the US to fulfill the requirements of Meaningful Use, methods using allergy list information to identify DDIs have high generalizability potential. In Chapter 3, I describe a systematic approach using EHR allergy lists to identify known DDIs and potential novel DDIs.

Drug Repurposing

In Chapter 4, I develop and evaluate an approach to identify drug repurposing candidates using public genomic data and validate candidates in the EHR. Here, I define drug repurposing as finding new indications for existing drugs. The motivation for drug repurposing is that drug development is expensive[79] and has a high failure rate.[80–82] Drug repurposing addresses the problems of high cost and failure rate in drug development.[83] Repurposing an existing drug saves money during clinical trial testing, as investigators can skip Phase 1[84], as the existing

drug has a well-characterized safety profile. Further, even in Phase 2 and 3 studies, adverse drug events can cause drugs to fail clinical trial testing.

There are many examples of successfully repurposed drugs. Etanercept, a drug originally developed for treating sepsis, was repurposed to treat rheumatoid arthritis (RA).[85] Bupropion was originally approved for depression which was then approved for smoking cessation.[86] Also more recently, bupropion was shown to be effective in treating methamphetamine use disorder.[87]

Repurposing drugs with retrospective observational analysis

In the past, drugs have been repurposed successfully following observations from retrospective observational analysis. For instance, rituximab was repurposed for RA. A patient with non-Hodgkin's lymphoma and RA was treated with rituximab, and rituximab was observed to improve the patient's RA symptoms.[88] Follow-up clinical trials showed that rituximab was effective in treating RA.[89–92]

The growth of EHR data led researchers to evaluate whether they could mine EHR data to identify drug repurposing candidates. For example, a previous study confirmed a finding that metformin was associated with reduced cancer mortality using EHR databases from VUMC and the Mayo clinic.[13] EHR data has also been used to develop data-driven approaches to identify novel non-cancer drugs that may improve cancer survival.[93] More recently, using a data-driven approach to repurpose drugs, structured lab values in the EHR (e.g., LDL-C measurements) were used to search for drugs that were correlated with low lab values.[94]

Repurposing drugs using human genetic data

In the past decade, researchers have developed methods to leverage human genomic data to identify drug repurposing candidates.[83,95,96] Supporting a genetics-based approach to discover drug repurposing candidates is that compared to drugs in Phase I clinical trials, there is a 4x enrichment of drugs that is supported by human genetics data.[97]

One such study developed a novel bioinformatics pipeline to augment the information from risk loci identified by GWAS to identify existing drugs as candidates for treating RA.[98] The authors first used cis-eQTL and functional annotation data to infer RA causal genes. They then searched for genes that directly interacted with the RA causal genes in protein-protein interaction

databases. The genes found in the databases were targets of approved RA drugs, demonstrating the feasibility of their pipeline. Applying the pipeline, they suggested palbociclib, an approved breast cancer drug, as a repurposing candidate for treating RA.

While genetic-target based search for finding drug repurposing candidates (i.e., targeting one gene by inhibiting one protein's activity) is a promising approach, an alternative method is signature-based discovery. The signature-based search, most often refers to gene expression signature, has its roots in systems biology. The idea is that for a phenotype of interest, there is a gene expression signature that represents the pathological mechanism that ultimately results in the phenotype. And that reversing the phenotype gene expression signature could potentially reverse the biological processes underlying the phenotype. Dudley et al[99] used publicly available gene expression data to identify small-molecule drugs as good candidates for treating inflammatory bowel disease. The best drug candidates were those that reversed the gene expression signature for inflammatory bowel disease. Using a rodent disease model, they validated an anticonvulsant, topiramate, as a repurposing candidate for inflammatory bowel disease.

Nonetheless, while gene expression signature-based drug repurposing is a promising approach, reading gene expression is currently much more expensive than genotyping. Recognizing this, researchers have leveraged S-PrediXcan with publicly available GWAS summary statistics in a signature-based approach to identify drug repurposing candidates. So et al performed such a study to find drugs for treating psychiatric disorders and Alzheimer's disease.[100] They used S-PrediXcan to estimate phenotype gene expression signatures and then searched in the Connectivity Map/L1000 databases to find drugs that reversed the S-PrediXcan phenotype gene expression signatures. With this approach, they replicated known approved drugs for the target psychiatric disorder phenotypes. They then validated potential novel drug repurposing candidates using evidence from the literature and clinical trials.

Gap in knowledge for using human 'omic data and clinical data in EHRs to identify drug repurposing candidates

While researchers have developed methods to identify drug repurposing candidates using large drug databases and S-PrediXcan phenotype transcriptomic signatures, they have not developed an approach to validate those candidates using real-world data contained in the EHRs. Further, no study has yet developed a generalizable approach to mine clinical data in EHRs to

systematically quantify treatment effects of drug candidates identified from ’omic analysis. Integrating these three orthogonal “big data” sources (estimated phenotype transcriptomic data, drug perturbation data, and clinical data in publicly available resources) has the potential to better prioritize drug repurposing candidates for clinical trial testing.

CHAPTER 3

DDIWAS: HIGH-THROUGHPUT ELECTRONIC HEALTH RECORD-BASED SCREENING OF DRUG-DRUG INTERACTIONS.

This manuscript has been accepted by the *Journal of the American Medical Informatics Association*:

Wu P, Nelson SD, Zhao J, Stone CA Jr, Feng Q, Chen Q, Larson EA, Li B, Cox NJ, Stein CM, Phillips EJ, Roden DM, Denny JC, Wei W-Q. 2021. DDIWAS: High-throughput electronic health record-based screening of drug-drug interactions. *J Am Med Inform Assoc*. doi:10.1093/jamia/ocab019

Abstract

Objective: We developed and evaluated Drug-Drug Interaction Wide Association Study (DDIWAS). This novel method detects potential drug-drug interactions (DDIs) by leveraging data from the electronic health record (EHR) allergy list.

Materials and Methods: To identify potential DDIs, DDIWAS scans for drug pairs that are frequently documented together on the allergy list. Using de-identified medical records, we tested 616 drugs for potential DDIs with simvastatin (a common lipid-lowering drug) and amlodipine (a common blood-pressure lowering drug). We evaluated the performance to rediscover known DDIs using existing knowledge bases and domain expert review. To validate potential novel DDIs, we manually reviewed patient charts and searched the literature.

Results: DDIWAS replicated 34 known DDIs. The positive predictive value to detect known DDIs was 0.85 and 0.86 for simvastatin and amlodipine, respectively. DDIWAS also discovered potential novel interactions between simvastatin-hydrochlorothiazide, amlodipine-omeprazole, and amlodipine-valacyclovir. A software package to conduct DDIWAS is publicly available.

Conclusions: In this proof-of-concept study, we demonstrate the value of incorporating information mined from existing allergy lists to detect DDIs in a real-world clinical setting. Since allergy lists are routinely collected in EHRs, DDIWAS has the potential to detect and validate DDI signals across institutions.

Introduction

Patients are taking more prescription drugs than ever to treat their chronic health conditions.[101] This rise in drug use increases their risk of developing drug-drug interactions (DDIs).[102] Patients experience DDIs when they concomitantly use an object drug (affected by the interaction) and a precipitant drug (causes the interaction). DDIs are responsible for >20% of adverse drug reactions (ADRs)[103] and for half of withdrawn drugs from the US market.[104]

DDIs can be recognized during drug development and clinical trials, but a lack of consensus for defining clinically-actionable DDIs remains.[105–108] Before a new drug is approved, potentially harmful DDIs are assessed using *in vitro* and *in vivo* methods. But, it is not feasible to test for all the possible interactions between the new drug and those prescribed to patients.[68] To identify

DDIs missed during drug development, healthcare providers can voluntarily report DDIs to post-market surveillance programs.[109,110] Yet, underreporting of DDI events can occur, as DDIs are hard to recognize and reporting events may not be the highest priority for healthcare providers. To complement post-market surveillance programs, researchers have developed methods to mine electronic health record (EHR) data for DDIs.[12,77] Implementing these methods across EHRs, however, remains challenging, because they are either purpose-built[111] or depend upon complex natural language processing (NLP).[12]

We developed Drug-Drug Interaction Wide Association Study (DDIWAS), a novel framework to identify potentially harmful DDIs by leveraging the EHR allergy list (**Figure 1A**). The allergy list is used by healthcare providers to document immune-mediated allergic drug reactions (e.g., penicillin anaphylaxis[112]) and drug intolerances (e.g., statin myopathy[113]) (**Figure 1B**). Allergy list entries also routinely contain only two data elements, the allergen (e.g., culprit drug's name) and reaction (e.g., "muscle cramp").[114] This standardized pattern shared among EHRs enables high-throughput DDI detection without sophisticated NLP. In this study, we assumed that a drug's appearance on the allergy list indicated that a drug-ADR occurred. With that assumption, we hypothesized that adversely interacting drugs would frequently be documented together on the allergy list. We only used allergy list data because EHR fragmentation can make it difficult to obtain accurate medication lists.[115]

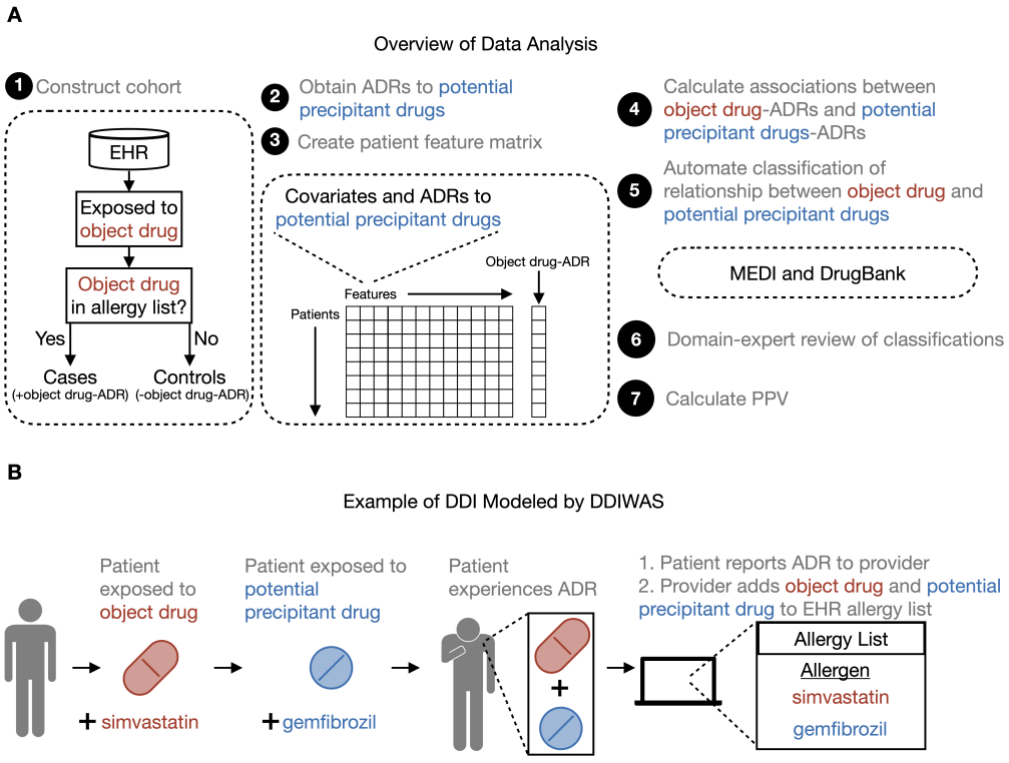


Figure 1. Overview of data analysis and example of DDIs modeled by DDIWAS.

(**A**) From a cohort of object drug-exposed patients, cases were those who had the object drug listed in their EHR allergy lists (+object drug-ADR), and controls were those who did not have the object drug documented on their allergy lists (-object drug-ADR). In this study, DDIWAS was applied on two object drugs, simvastatin and amlodipine. To search for potential precipitant drugs that increased the risk of object drug-ADRs, a systematic association test was performed using logistic regression. Potential precipitant drugs of interest were those that were positively associated with object-drug ADRs (logistic regression Bonferroni p-value < 0.05 and OR > 1). Using MEDI and DrugBank, the relationship between the object drug and potential precipitant drugs were determined. All object-potential precipitant drug relationships were manually reviewed by a domain expert (S.N., a pharmacist). PPV was then used to evaluate DDIWAS' ability to replicate known DDIs. See also **Supplementary Figure 1**. (**B**) In this example of a DDI modeled by DDIWAS, the object drug is simvastatin, and the potential precipitant drug is gemfibrozil. The patient develops an ADR after concurrently using simvastatin and gemfibrozil. At the next visit, the patient reports their ADR to their provider, who adds both drugs to the patient's allergy list. DDI: drug-drug interaction; DDIWAS: Drug-Drug Interaction Wide Association Study; EHR: electronic health record; ADR: adverse drug reaction; OR: odds ratio; MEDI: MEDication Indication resource; PPV: positive predictive value.

To start the DDIWAS pipeline, we first identified a cohort of object drug-exposed patients in a de-identified EHR database[3] (**Figure 1A**; **Supplementary Figure 1**). We then divided the patients into cases (+object drug-ADR, i.e., object drug on allergy list) and controls (-object drug-

ADR, i.e., object drug not on allergy list). We searched for potential precipitant drugs that were disproportionately co-documented with the object drug on patients' allergy lists. To measure DDIWAS' performance, we calculated a positive predictive value (PPV) using a gold standard reference comprised of MEDication Indication resource (MEDI),[23] DrugBank,[16] and domain expert review (**Supplementary Figure 2**). We validated DDIWAS by applying it on two common drugs, simvastatin and amlodipine.

Materials and Methods

Defining a drug-drug interaction (DDI)

In this study, a patient has experienced a DDI when the pharmacologic effects of two drugs overlap to produce an adverse outcome. When the object and precipitant drug interact, the patient experiences an ADR. The patient reports the ADR to their healthcare provider, who documents the adverse reaction in the patient's EHR by adding the object drug to the patient's allergy list. The provider does so because they believe that the ADR was most likely related to the patient's exposure to the object drug.[116] If the provider believes that the ADR was due to a DDI between the object and precipitant drug, then they may add both drugs to the patient's allergy list.

As a concrete example of how DDIWAS determines whether a potential DDI occurred using allergy list data, consider a DDI between simvastatin (the object drug) and gemfibrozil (the precipitant drug) (**Figure 1B**). A provider prescribes gemfibrozil to a patient already on simvastatin. At the subsequent visit, the provider learns that after starting gemfibrozil, the patient began experiencing muscle aches. Since muscle ache is a common ADR associated with simvastatin exposure,[117] the provider believes that a DDI between simvastatin and gemfibrozil occurred and adds both drugs to the patient's allergy list.

Study design

The study was reviewed and approved by the IRB at Vanderbilt University Medical Center (VUMC) (#180456). We used de-identified EHR data from VUMC. The EHR database maintains longitudinal clinical data for over 3.2 million unique patients from inpatient and outpatient encounters.[3] EHR data commonly includes diagnosis and procedure codes, medications, laboratory test results, unstructured clinical text, and demographics. We used EHR data from

outpatient visits from 1996-2020 and limited our analyses to adult patients between the ages of 18-90 years.

To demonstrate the feasibility of DDIWAS, we used it to identify DDIs for simvastatin and amlodipine, drugs that are commonly used with known precipitant drugs.[118] Simvastatin is one of the first-line therapies for hyperlipidemia and has a relatively increased frequency of myopathy at high doses.[119] Amlodipine is commonly used to treat hypertension and is known to inhibit CYP3A4,[117] a key enzyme involved in drug metabolism.

To identify drugs in the EHR, we used a standard terminology that formalizes all prescription drugs currently marketed in the US, RxNorm[14]. We used generic and brand names to first map drugs to RxNorm Concept Unique Identifiers (RxCUIs) and then to their respective drug ingredients, based on their relationships in RxNorm. For example, “Simvastatin” (RxCUI 36567) and “Zocor” (RxCUI 196503) were both mapped to the drug ingredient “simvastatin” (RxCUI 36567).

For each object drug, we started with a cohort of patients who had ≥1 exposure(s) to the object drug (**Figure 1A**; **Supplementary Figure 1A**). In this cohort, we defined cases as patients who had the object drug documented on their allergy lists (+object drug-ADR), and defined controls as patients who did not have the object drug listed on their allergy lists (-object drug-ADR).

For both cases and controls, we set the date on which object drug exposure occurred as the start of the observation period (T_0) (**Supplementary Figure 1B**). For cases, we set the date on which the object drug was first documented on their allergy list as the end of the observation period (T_e). We limited the duration of the observation period to twelve months, because we wanted to capture ADRs from both short and long object drug exposures.[120] If the observation period (T_e-T_0) was longer than twelve months, then we limited our analysis to the twelve month period prior to T_e. For controls, we set T_e as the date on which object drug exposure was last documented in their EHRs. If the observation period was longer than twelve months, we limited our analysis to the twelve month period after T_0.

We obtained potential precipitant drug-ADRs by extracting all drugs documented on the patient allergy lists during the observation period (**Supplementary Figure 1C**). We then mapped the potential precipitant drugs to their RxCUI ingredients (**Supplementary Figure 1D**). To obtain only ADRs potentially due to DDIs between object and potential precipitant drugs, we removed

drugs that were present on patient allergy lists prior to the start of the observation period. To prevent false-positive associations due to the absence of allergy list entries, we excluded controls who did not have any allergy list entries during the observation period (**Supplementary Figure 1E**).

Data preprocessing and association analysis

We created a patient feature matrix with each row representing one patient and with columns representing features (**Supplementary Figure 1F**). Features included covariates and potential precipitant drug-ADRs. The covariates were age, sex, race, duration of observation period, and number of unique drug ingredient exposures during the observation period. We encoded potential precipitant drug-ADRs as dichotomous variables. We then only tested potential precipitant drugs for which the 2x2 contingency table had ≥1 patients in each cell (**Supplementary Figure 1G**), because our goal was to identify drugs that increased the likelihood of object drug-ADRs.

To identify potential precipitant drugs that increased the risk of object drug-ADRs, we used the patient feature matrix to perform a systematic association study with logistic regression (**Supplementary Figure 1H**). For each patient, the dependent variable indicated whether the patient was a case (+object drug-ADR; object drug on allergy list) or control (-object drug-ADR; object drug not on allergy list). For each potential precipitant drug tested, the dichotomous independent variable indicated whether the drug was listed on each patient's allergy list. The logistic regression analysis was adjusted for the covariates described above. The outputs of the logistic regression analysis were association odds ratios (ORs) and p-values. Due to the limitations of logistic regression with rare events,[121] we used Firth regression for potential precipitant drugs that had <5 patients in each cell of the 2x2 contingency table. To account for multiple testing, we applied a Bonferroni correction with type I error rate set to 0.05. We considered a potential precipitant drug to have increased the risk of object drug-ADRs if the following conditions were met: (1) regression OR > 1 and (2) regression Bonferroni-corrected p-value < 0.05 (**Supplementary Figure 2**). Drugs that met these conditions indicated that patients with the potential precipitant drug listed on their allergy lists (+potential precipitant drug-ADR) were more likely to have the object drug listed as well (+object drug-ADR).

Labeling of DDIWAS output

We labeled the potential precipitant drugs meeting the two conditions, to help us interpret DDIWAS results, and to measure the method's ability to replicate known DDIs. We used three labels: "Exclude", "True-positive", and "False-positive" (**Supplementary Figure 2**). To automatically label these drugs, we leveraged MEDI[23] and DrugBank[16] resources. First, using MEDI and manual engineering, we tagged drugs as "Exclude" if they shared indications with the object drug. Indications were represented by the International Classification of Diseases Ninth Revision, Clinical Modification (ICD-9-CM) code(s)[17] most specific for the object drug. As an example, for the simvastatin experiment, we used the hyperlipidemia diagnosis codes, ICD-9-CM 272.4 "Other and unspecified hyperlipidemia" and ICD-9-CM 272.2 "Mixed hyperlipidemia".

Our gold standard reference for "True-positive" findings was DrugBank[16] followed by domain expert review. We used DrugBank to identify potential precipitant drugs that were known to interact adversely with the object drug of interest. We labeled potential precipitant drugs as "True-positive" if the DrugBank description indicated that for either the object or potential precipitant drug, drug metabolism decreased, serum concentration increased, drug absorption increased, drug elimination decreased, or concurrent use increased the risk of ADRs (e.g., rhabdomyolysis with simvastatin use). We then tagged the remaining unlabeled potential precipitant drugs as "False-positive". For final classification of object and potential precipitant drug pairs, all labels were manually reviewed by a domain expert (S.N., a pharmacist).

Measuring DDIWAS performance to replicate known DDIs

To quantify the performance of DDIWAS to replicate known DDIs, we used PPV (**Supplementary Figure 2**). PPV was calculated by dividing the number of "True-positive" drugs by the sum of "True-positive" and "False-positive" drugs. PPV represented the fraction of remaining drugs with previously reported DDIs with the object drug of interest.

Adjusting for potential confounders and sensitivity analysis

To identify associations that may have been confounded by indication(s) for each significantly associated potential precipitant drug, we independently adjusted the regression for indications represented by phecodes.[122] To select indications for each potential precipitant drug, we used the two ICD-9-CM codes with the highest prevalence from MEDI.[23] We mapped these ICD-9-CM codes to their respective phecodes using the ICD-9-CM to phecode v1.2 map. We then rolled up the mapped phecodes to their parent phecodes. Using the indication for digoxin as an

example, ICD-9-CM 427.31 “Atrial fibrillation” → phecode 427.21 “Atrial fibrillation” → phecode 427 “Cardiac dysrhythmias”. To obtain phecode indications for ICD-10-CM codes in our study cohort, we used the ICD-10-CM to phecode v1.2 (beta) map.[22]

We also conducted a sensitivity analysis by calculating PPVs and true-positive counts using minimum patient count thresholds of 1, 5, 10, and 20 patients, in each cell of the 2x2 contingency table (**Supplementary Figure 3**).

Data visualization

To present the results from our primary analysis, we used forest plots of regression OR (95% confidence interval [95% CI]) for all potential precipitant drugs that passed Bonferroni correction and OR > 1. For each drug, we show the OR (95% CI) from the logistic regression adjusted for baseline characteristics and a second regression with additional adjustment for each drug’s main indication(s) (**Figures 2A, 2B**).

Validation studies for potentially novel DDIs

We defined “False-positive” drugs as potentially novel DDIs (**Supplementary Figure 2**). To validate these potentially novel DDIs, we reviewed the clinical notes for ten randomly selected patients who DDIWAS labeled as (+object drug-ADR, +potential precipitant drug-ADR). If there were less than ten (+object drug-ADR, +potential precipitant drug-ADR) patients, we reviewed the clinical notes for all available patients. For each DDIWAS-labeled +drug-ADR patient, we reviewed their clinical notes to verify that the drug was intentionally added to their allergy lists. Each reviewed DDIWAS-labeled +drug-ADR patient was labeled “True-positive +drug-ADR” or “False-positive +drug-ADR”. A “True-positive +drug-ADR” patient was not exposed to the drug after the end of the observation period and/or for whom a provider mentioned the drug-ADR in additional EHR sections like “History of Present Illness” and “Assessment & Plan”. A “False-positive +drug-ADR” patient did not meet either criteria.

After reviewing patient charts, we were concerned about the remaining drugs that met our criteria for potentially novel DDIs, but whose associations with object drug-ADRs did not likely represent novel DDIs. Instead, the associations were more likely to be due to interactions between the object drug and other drugs; the other drugs were either commonly co-prescribed or combined with the drugs of concern. To address this problem, we adjusted the regressions for

additional potential precipitant drug-ADRs. For example, in the primary simvastatin DDIWAS analysis, the regression for triamterene was:

$$p(\text{+simvastatin-ADR}|\text{+triamterene-ADR}) \sim logit(\beta_0 + \beta_1 \text{triamterene-ADR} + [\text{baseline covariates}]) \quad (1)$$

We then adjusted the regression for hydrochlorothiazide (HCTZ)-ADRs, because HCTZ is frequently combined with triamterene:

$$p(\text{+simvastatin-ADR}|\text{+triamterene-ADR}) \sim logit(\beta_0 + \beta_1 \text{triamterene-ADR} + \beta_2 \text{HCTZ-ADR} + [\text{baseline covariates}]) \quad (2)$$

Results

Simvastatin DDIWAS

The simvastatin experiment (**Table 1**) had 85,873 controls (+simvastatin-exposed, -simvastatin-ADR) and 2,814 cases (+simvastatin-exposed, +simvastatin-ADR). Of the 282 potential precipitant drugs tested (**Supplementary Figure 4A**; **Supplementary Table 1**), thirteen increased the risk of simvastatin-ADRs (passing Bonferroni correction [$0.05/282 = 1.77 \times 10^{-4}$] with OR > 1; **Figure 2A**). To control for potential confounding by drug indications, we adjusted the regressions for potential precipitant drug indications and found that all thirteen associations remained significant (**Supplementary Table 2**). Eleven of the thirteen drugs were known to interact with simvastatin, including fenofibrate, gemfibrozil, niacin, and amlodipine. In DrugBank, the remaining two drugs not known to interact with simvastatin were HCTZ and triamterene.

Table 1: Simvastatin Patient-Level Characteristics

Characteristic	Controls (n = 85,873)	Cases (n = 2,814)	P
Female	0.52 (44,367)	0.56 (1,564)	<0.001
White	0.80 (68,900)	0.85 (2,384)	<0.001
Age, years	63 (54-71)	63 (54-70)	0.15
Observation period length, days	337 (15-365)	365 (111-365)	<0.001
Unique drug exposures, count	12 (7-19)	13 (8-25)	<0.001
Phecode 250.* (Diabetes mellitus)	0.19 (16,241)	0.25 (702)	<0.001
Phecode 272.* (Disorders of lipid metabolism)	0.34 (28,776)	0.74 (2,084)	<0.001
Phecode 401.* (Hypertensive disorder)	0.35 (30,253)	0.59 (1,654)	<0.001
Phecode 411.* (Myocardial infarction)	0.20 (17,384)	0.31 (859)	<0.001
Phecode 418.* (Chest pain)	0.13 (10,774)	0.18 (519)	<0.001
Phecode 743.* (Osteoporosis)	0.04 (3,280)	0.06 (172)	<0.001

For continuous variables, numbers represent median (interquartile range).
For dichotomous variables, numbers after proportions are counts.
P values indicate differences between cases and controls. For continuous variables, P values were calculated using Mann-Whitney test. For dichotomous variables, P values were calculated using χ^2 test. P < .05 was considered statistically significant.
*For phecodes, * means ≥1 digits or a period (e.g., phecode 401.* = phecodes 401, 401.1, 401.2, 401.21, 401.22, or 401.3).*

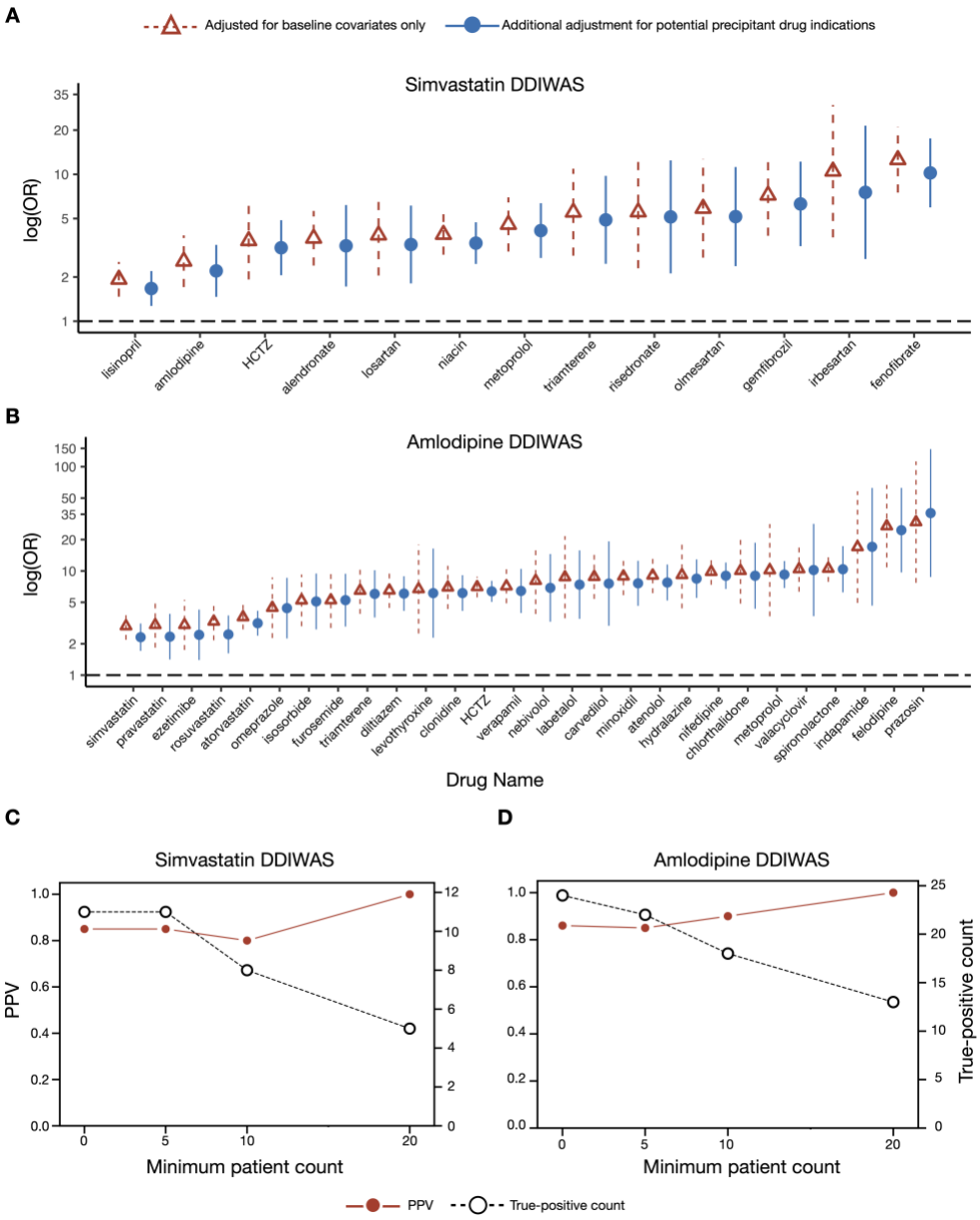


Figure 2. Forest plot of potential precipitant drugs associated with object drug ADRs and DDIWAS performance

(**A, B**) Forest plots summarizing the potential precipitant drugs that were significantly associated (logistic regression Bonferroni p-value < 0.05 and OR > 1) with (**A**) simvastatin- and (**B**) amlodipine-ADRs. On the horizontal axis, potential precipitant drugs are sorted from smallest to largest ORs. On the vertical axis, association ORs (95% CI) are plotted on a logarithmic scale. Red triangles with dashed lines represent values from logistic regressions adjusted for age, sex, race, length of observation period, and number of unique drug exposures for each patient. Blue circles with solid lines indicate values from logistic regressions with additional adjustment for potential precipitant drug indications. These values were from analyses with a minimum patient

count threshold of 1. Minimum patient count threshold refers to the number of patients required in each cell of the 2x2 contingency table (**Supplementary Figure 3**). The Bonferroni correction was 1.77×10^{-4} (0.05/282) for simvastatin and 1.49×10^{-4} (0.05/335) for amlodipine. See **Supplementary Table 2** for corresponding numbers. (**C, D**) PPV (left vertical axis) and true-positive count (right vertical axis) for (**C**) simvastatin and (**D**) amlodipine DDIWAS at minimum patient count thresholds of 1, 5, 10, and 20. True-positive count refers to the number of potential precipitant drugs (logistic regression Bonferroni p-value < 0.05 and OR > 1) that were known to interact with the object drug. See **Supplementary Table 3** for corresponding numbers. ADR: adverse drug reaction; DDIWAS: Drug-Drug Interaction Wide Association Study; PPV: positive predictive value; OR: odds ratio; 95% CI: 95% confidence interval; HCTZ: hydrochlorothiazide.

To examine the potential novel DDIs between simvastatin-HCTZ and simvastatin-triamterene, we manually reviewed clinical notes to verify that the drugs were intentionally listed on patient allergy lists (**Table 2**). The reviewed notes were from two types of patients: those who potentially experienced simvastatin-HCTZ DDIs, i.e., DDIWAS-labeled (+simvastatin-ADR, +HCTZ-ADR) and those who potentially experienced simvastatin-triamterene DDIs, i.e., DDIWAS-labeled (+simvastatin-ADR, +triamterene-ADR). All reviewed patients had the respective drugs intentionally listed on their allergy lists. We hypothesized that the triamterene association was confounded by HCTZ-ADRs, because all reviewed DDIWAS-labeled (+simvastatin-ADR, +triamterene-ADR) patients were exposed via a HCTZ/triamterene combination drug. Further, there were DDIWAS-labeled (+simvastatin-ADR, +HCTZ-ADR) patients who did not have triamterene on their allergy lists. To test our hypothesis, we adjusted the triamterene-ADR regression with HCTZ-ADRs and found that triamterene's association was no longer significant, while HCTZ's association remained significant (**Supplementary Table 4**).

Table 2: Validation analysis of potentially novel DDIs, manual chart review results

Object Drug	Potential Precipitant Drug	% TP drug-ADR (TP/number of patients reviewed)	Comments
simvastatin	HCTZ	100 (10/10)	
simvastatin	triamterene	100 (10/10)	All (+simvastatin-ADR, +triamterene-ADR) patients were exposed to triamterene using a HCTZ/triamterene combination drug
amlodipine	ezetimibe	90 (9/10)	Majority (8/10) of (+amlodipine-ADR, +ezetimibe-ADR) patients also had a statin drug on their allergy list. The single false-positive (+amlodipine-ADR, +ezetimibe-ADR) patient had neither the object nor potential precipitant drug on their allergy list.
amlodipine	levothyroxine	40 (2/5)	Reviewed all five available (+amlodipine-ADR, +levothyroxine-ADR) patients. Two false-positive (+amlodipine-ADR, +levothyroxine-ADR) patients had neither the object nor potential precipitant drug on their allergy lists. One false-positive (+amlodipine-ADR, +levothyroxine-ADR) patient did not have the potential precipitant drug on their allergy list.
amlodipine	valacyclovir	80 (4/5)	Reviewed all five available (+amlodipine-ADR, +valacyclovir-ADR) patients. The single false-positive (+amlodipine-ADR, +valacyclovir-ADR) patient had neither amlodipine nor valacyclovir on their allergy list.

Object Drug	Potential Precipitant Drug	% TP drug-ADR (TP/number of patients reviewed)	Comments
amlodipine	omeprazole	100 (10/10)	

True-positive patients were those for whom providers intentionally added both the object and potential precipitant drugs to their allergy lists.
DDIWAS: Drug-Drug-Interaction Wide Association Study; HCTZ: hydrochlorothiazide; ADR: adverse drug reaction; TP: True-positive.

Amlodipine DDIWAS

The amlodipine experiment (**Table 3**) had 83,732 controls (+amlodipine-exposed, -amlodipine-ADR) and 2,512 cases (+amlodipine-exposed, +amlodipine-ADR). Of the 335 potential precipitant drugs tested (**Supplementary Figure 4B**; **Supplementary Table 1**), 28 increased the risk of amlodipine-ADRs (passing Bonferroni correction [$0.05/335 = 1.49 \times 10^{-4}$] with OR > 1; **Figure 2B**). All associations remained significant after adjusting the regressions for potential precipitant drug indications (**Supplementary Table 2**). Twenty-four of the 28 drugs were known to interact with amlodipine, including prazosin, diltiazem, and verapamil. In DrugBank, there were four drugs not known to interact with amlodipine: levothyroxine, ezetimibe, omeprazole, and valacyclovir.

Table 3: Amlodipine Patient-Level Characteristics

Characteristic	Controls (n = 83,732)	Cases (n = 2,512)	P
Female	0.54 (45,315)	0.65 (1,637)	<0.001
White	0.75 (63,144)	0.83 (2,083)	<0.001
Age, years	63 (53-72)	65 (55-73)	<0.001
Observation period length, days	287 (14-365)	206 (38-365)	0.91
Unique drug exposures, count	12 (8-20)	13 (8-22)	<0.001
Phecodes 053.* (Herpes zoster)	2.93E-03 (245)	3.18E-03 (8)	0.81
Phecodes 054.* (Herpes simplex)	1.97E-03 (165)	2.39E-03 (6)	0.64
Phecodes 244.* (Hypothyroidism)	0.05 (4,558)	0.08 (201)	<0.001
Phecodes 250.* (Diabetes mellitus)	0.16 (13,144)	0.15 (365)	0.11
Phecodes 272.* (Disorders of lipid metabolism)	0.21 (17,898)	0.37 (927)	<0.001
Phecodes 300.* (Anxiety, phobic and dissociative disorders)	0.04 (3,325)	0.05 (114)	0.15
Phecodes 401.* (Hypertensive disorder)	0.43 (36,059)	0.68 (1,702)	<0.001
Phecodes 411.* (Myocardial infarction)	0.14 (11,898)	0.15 (371)	0.43
Phecodes 414.* (Other forms of chronic heart disease)	0.02 (1,968)	0.02 (46)	0.09
Phecodes 418.* (Chest pain)	0.11 (9,170)	0.14 (348)	<0.001
Phecodes 427.* (Cardiac dysrhythmias)	0.12 (10,096)	0.14 (355)	0.002
Phecodes 428.* (Congestive heart failure)	0.06 (4,609)	0.05 (138)	0.98

Characteristic	Controls (n = 83,732)	Cases (n = 2,512)	P
Phecodes 530.* (Esophageal disorders)	0.08 (6,706)	0.09 (228)	0.053
Phecodes 536.* (Disorders of function of stomach)	0.01 (858)	0.01 (18)	0.13

For continuous variables, numbers represent median (interquartile range).
For dichotomous variables, numbers after proportions are counts.
P values indicate differences between cases and controls. For continuous variables, P values were calculated using Mann-Whitney test. For dichotomous variables, P values were calculated using χ^2 test. P < .05 was considered statistically significant.
*For phecodes, * means ≥1 digits or a period (e.g., phecode 401.* = phecodes 401, 401.1, 401.2, 401.21, 401.22, or 401.3).*

To examine the potential novel DDIs between amlodipine and the four drugs, we manually reviewed clinical notes (**Table 2**). First, for levothyroxine, of the five available DDIWAS-labeled (+amlodipine-ADR, +levothyroxine-ADR) patients, two had both drugs listed on their allergy lists, one had only amlodipine listed, and two had neither drug listed. Of note, the two false-positive DDIWAS-labeled (+amlodipine-ADR, +levothyroxine-ADR) patients were taking both drugs during the observation period. Second, for ezetimibe, 90% (9/10) of DDIWAS-labeled (+amlodipine-ADR, +ezetimibe-ADR) patients had both drugs on their allergy lists. The single false-positive DDIWAS-labeled (+amlodipine-ADR, +ezetimibe-ADR) patient did not have either drug listed on their allergy list. Since ezetimibe is commonly used with statins to lower cholesterol, we then adjusted the ezetimibe regression for ADRs to common statins, simvastatin and atorvastatin; in this statin-ADRs adjusted regression, the association p-value for ezetimibe was no longer significant (p-value = 0.29; **Supplementary Table 4**) However, in this same adjusted ezetimibe regression, the association p-values for both statin-ADRs remained significant. Third, for omeprazole, all (10/10) DDIWAS-labeled (+amlodipine-ADR, +omeprazole-ADR) patients had both drugs documented on their allergy lists. Fourth, for valacyclovir, 4/5 DDIWAS-labeled (+amlodipine-ADR, +valacyclovir-ADR) patients had both drugs documented on their allergy lists.

Replication sensitivity analysis

To quantify the performance of DDIWAS to replicate known DDIs, we calculated the PPV for both simvastatin and amlodipine experiments at minimum patient count thresholds of 1, 5, 10, and 20 (**Supplementary Table 3**). In the simvastatin experiment, as thresholds increased, the PPV increased from 0.85 to 1.00, but the number of true-positive findings decreased from eleven to five potential precipitant drugs (**Figure 2C**). For amlodipine, as thresholds increased, the PPV increased from 0.86 to 1.00, but the number of true-positive findings decreased from 24 to thirteen (**Figure 2D**).

Discussion

DDIWAS is a high-throughput method to identify potential DDIs by mining the EHR allergy list. We used the method to identify potential DDIs for simvastatin and amlodipine. DDIWAS replicated known DDIs with a PPV of 0.85 and 0.86 for simvastatin and amlodipine, respectively. For both drugs, DDIWAS also detected potentially novel DDIs that were validated with manual

review of patient clinical notes. Our validation studies support potentially novel interactions between simvastatin-HCTZ, amlodipine-omeprazole, and amlodipine-valacyclovir.

Existing methods to mine EHR data have successfully replicated known DDIs,[12] but have limitations that prevent widespread adoption. First, the tools used to detect DDIs in EHRs are rarely publicly available. Second, even if they are available, these tools are often purpose-built advanced NLP or text annotation applications,[74] requiring users to perform substantial customization for use with external datasets.[123,124] In contrast, DDIWAS identifies DDI events using drug name recognition, a relatively simpler task than NLP-based detection of ADRs. Recognizing drug names is easier than detecting ADRs across health systems due to local documentation procedures that may lead to differences in how ADRs are represented in clinical narratives.[31,125] DDIWAS may be easier to implement in external databases, as it only searches for drug names in medication and allergy lists, EHR modules with smaller contextual variability than in clinical narratives. We anticipate that users will be able to apply DDIWAS to identify DDIs in their databases, without spending substantial time and resources to modify text annotation tools.

To test our approach to identify DDIs, we wanted to see whether we could replicate drugs known to interact with the object drugs, simvastatin and amlodipine. We found that 85% (35/41) of the significantly associated drugs (Bonferroni p-value < 0.05 and OR > 1) were known to interact with the object drugs. In the simvastatin analysis, we tested eight drugs that were recommended for inclusion in all clinical decision support (CDS) DDI alert systems.[126] These drugs were amiodarone, clarithromycin, diltiazem, erythromycin, fluconazole, ketoconazole, nefazodone, verapamil. Among these drugs, none were found to be significantly associated with simvastatin-ADRs. These “false-negative” findings could partially be attributed to intervention by the CDS alerts designed to reduce cases of clinically-significant DDIs.[127,128] Notably, drugs that were significantly associated with simvastatin-ADRs included niacin and warfarin. Although these drugs are known interact adversely with simvastatin, an expert committee recommended that alerts for these DDIs be deleted, because the therapeutic benefits of these drugs outweigh the risk of patient harm.[129] In this study’s amlodipine analysis, prazosin’s association had the largest effect size (**Figure 2B**; **Supplementary Table 1**). This finding is supported in the literature, as patients using both calcium-channel blockers (e.g., amlodipine) and alpha-1 blockers (e.g., prazosin) have been found to be at increased risk of developing hypotension.[130,131]

In addition to replicating known DDIs, DDIWAS also identified potentially novel DDIs. Our results suggest a potential novel simvastatin-HCTZ DDI. Out of the thirteen drugs significantly associated with simvastatin-ADRs, HCTZ and triamterene did not have previously reported DDIs with simvastatin. When the triamterene regression was adjusted for HCTZ-ADRs, HCTZ's association, but not triamterene's, was still significant at a Bonferroni p-value <0.05 (**Supplementary Table 4**). It has been shown that patients who concurrently used statins and HCTZ were at increased risk of adverse events, including chest pain, hyperglycemia, and muscle spasms.[76]. Additional evidence to support a simvastatin-HCTZ DDI can be found in DrugBank; rosuvastatin and pravastatin are predicted to decrease HCTZ excretion, suggesting a possible interaction between HCTZ and the statin drug class. Nonetheless, a biological mechanism to explain a simvastatin-HCTZ interaction remains to be explored.

DDIWAS found potentially novel amlodipine-DDIs with valacyclovir and omeprazole. A previous study has shown that patients exposed concurrently to amlodipine and valacyclovir were at increased risk of developing adverse outcomes like acute kidney failure, dysarthria, and dizziness.[76] The same study found that patients using both amlodipine and omeprazole were more likely to experience chest pain and dyspnea.[76] A pharmacogenomic study found that CYP2C19 intermediate metabolizers were more prone to developing amlodipine-omeprazole DDIs.[132] When exposed to both amlodipine and omeprazole, these patients experienced higher than expected drops in blood pressure. The authors proposed a mechanism in which elevated levels of omeprazole inhibits CYP3A metabolism of amlodipine, leading to lower blood pressures. Overall, results from the amlodipine experiments corroborate DDIWAS as an effective tool to detect potentially novel DDIs using real-world evidence in EHR data.

There are several limitations in this study. First, to detect DDIs, DDIWAS uses frequentist approaches assuming no prior information. If there is prior knowledge of a DDI, such as those derived from pharmacologic and/or pharmacokinetic studies, we can potentially improve DDIWAS using Bayesian approaches with prior probabilities determined from existing evidence[133,134]. Second, we only performed DDIWAS using a maximum observation window length of one year and did not examine other period lengths. Third, we assumed that a patient experienced an adverse outcome to a drug of interest if the drug was listed on the patient's allergy list. Even if a healthcare provider intentionally added a drug to a patient's allergy list, the patient still may not have truly experienced an ADR to the drug. Potential reasons for false-positive cases include unverified patient-reported ADRs,[135,136] disease exacerbation

presenting like an ADR, and variability among healthcare providers' abilities to identify the causal drug.[137] But, multiple studies have successfully used the allergy list to identify patients with ADRs.[138,139] Likewise, we found that the majority of the DDIWAS-labeled +drug-ADR patients reviewed truly had the drugs listed on their allergy lists (**Table 2**). The dependence on healthcare providers' abilities to correctly identify causal drugs also increases the probability of false-negative DDIs. For example, DDIWAS did not detect a well-known interaction between simvastatin and amiodarone.[113] It would be interesting to see whether using drug exposures from the medication list increases the sensitivity of DDIWAS to identify potential DDIs without sacrificing PPV. Fourth, in its current form, DDIWAS does not systematically adjust for combination drugs, which can confound the interpretation of associations. In the simvastatin experiment, our stratified analysis found that the simvastatin-triamterene association was confounded by patients taking HCTZ/triamterene combination drugs (**Table 2**; **Supplementary Table 4**). Drugs frequently co-prescribed can also contribute to false-positive findings. We found that the amlodipine-ezetimibe association was most likely confounded by interactions between amlodipine and statin drugs. A module to automatically adjust associations for combination drugs and drugs often used together is an opportunity for future development. Fifth, to maximize the transportability of DDIWAS, we did not use ADR information that was present in some allergy list entries. Using ADR information represented as unstructured text would likely require NLP expertise, as providers may describe the same ADRs differently (e.g., myopathy could be described as "muscle cramp", "myotoxicity" "muscle weakness"). Recently, Wang et al. developed a data-driven approach to help providers pick specific ADRs conditional on the drug selected in the allergy list.[114] Incorporating such approaches may increase the use of structured ADR entries, which could augment DDIWAS' ability to detect potential DDIs. Sixth, like other retrospective observational studies, we do not claim that these associations were caused by DDIs. Like previous studies,[12] our goal was to show that DDIWAS can generate DDI hypotheses that will require validation by follow-up studies. Last, while we applied DDIWAS to data from only one institution, users at external institutions that also organize their EHR data with the Observational Health Data Sciences and Informatics (OHDSI)/Observational Medical Outcomes Partnership (OMOP) Common Data Model[11] can apply DDIWAS to their dataset after making minor changes to the code that we have shared publicly.[140]

Conclusions

In summary, we developed and evaluated DDIWAS, a novel method that uses EHR allergy list entries to detect DDIs. DDIWAS replicated known DDIs and identified potentially novel DDIs. EHR-based methods like DDIWAS could complement existing tools to improve post-market surveillance of DDIs.

Acknowledgements

We would like to thank David Schlueter (National Institutes of Health) for his advice on statistical analysis. We would like to thank Raymond Wu, Aaron Lim (Vanderbilt University School of Medicine), Bryan Steitz (Vanderbilt University School of Medicine), and Vivian Siegel (Vanderbilt University Medical Center) for reading drafts of this manuscript.

Funding Statement

This work was supported by National Institutes of Health grant numbers T32GM007347, R01 LM010685, R01 HL133786, T15 LM007450, P50 GM115305, K12HS026395, R35 GM131770, and American Heart Association grant number 16SDG27490014. The data set used for the analyses described was obtained from Vanderbilt University Medical Center's resources, the Synthetic Derivative, which are supported by institutional funding and by the National Center for Advancing Translational Science grant number 2UL1 TR000445-06. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Competing Interests Statement

The authors have no competing interests to declare.

Contributorship Statement

Study initialization: P.W., J.C.D., and W-Q.W.

Study design: P.W., S.D.N., Q.F., Q.C., J.C.D., and W-Q.W.

Acquisition of data: P.W. and W-Q.W.

Analysis and interpretation of data: P.W., S.D.N., J.Z., C.S., Q.F., Q.C., E.A.L., B.L., N.C., C.M.S., E.J.P., D.M.R., J.C.D., and W-Q.W.

Drafting of the manuscript: P.W. and W-Q.W.; all authors contributed to refinement of the manuscript and approved the final manuscript.

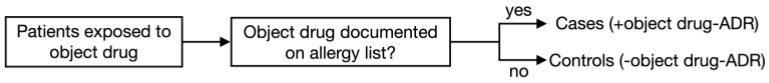
Grant holder: W-Q.W., E.J.P., D.M.R., and J.C.D.

Data Availability Statement

Due to patient privacy concerns, we are unable to share the EHR data used in this study. However, we have released the DDIWAS R package on GitHub (https://github.com/pwatrick/ddiwas) under an Apache License and archived version 0.1 on Zenodo.[140] We have written a tutorial to extract data from EHR databases organized using the OHDSI/OMOP Common Data Model[11], which can be found at https://pwatrick.github.io/ddiwas/articles/extract_ehr_data.html. We have also prepared a guide to process and analyze extracted EHR data, which can be found at https://pwatrick.github.io/ddiwas/articles/ddiwas_r_package_tutorial.html.

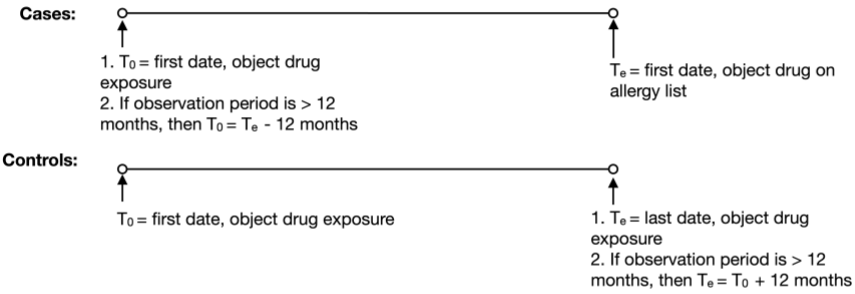
Supplementary Information

A Identify patients with outpatient exposure to object drug and define cases and controls



B Define observation period

T_0: start of observation
T_e: end of observation



C Get potential precipitant drugs from observation period and features (including covariates).

Potential precipitant drugs from allergy list	
id	RxCUI between T_0 and T_e

Features (including covariates): for each patient						
id	age at T_0	sex	race	days between T_0 and T_e	number of unique drug exposures	object drug-ADR

D Map potential precipitant drugs to ingredient(s). Remove drug ingredient(s) that were on allergy list prior to observation period.

ADRs table: RxCUI → RxCUI ingredient(s)	
id	RxCUI ingredient between T_0 and T_e

E During observation period, require patients to have ≥1 entries in EHR allergy list, including "No Known Allergies" and be between 18-90 years at start of observation period

F Prepare patient feature matrix

	covariates					potential precipitant drug ADRs				
id	age at T_0	sex	race	days between T_0 and T_e	number of unique drug exposures	m_1	m_2	…	m_i	object drug-ADR
1	55	0	1	365	20	1	0	…	0	1
2	40	1	0	93	12	0	1	…	0	0
3	63	1	1	30	10	0	0	…	1	0
…	…	…	…	…	…	…	…	…	…	…

medication list — allergy list

G Test only potential precipitant drugs with ≥1 patients in each cell of 2x2 contingency table

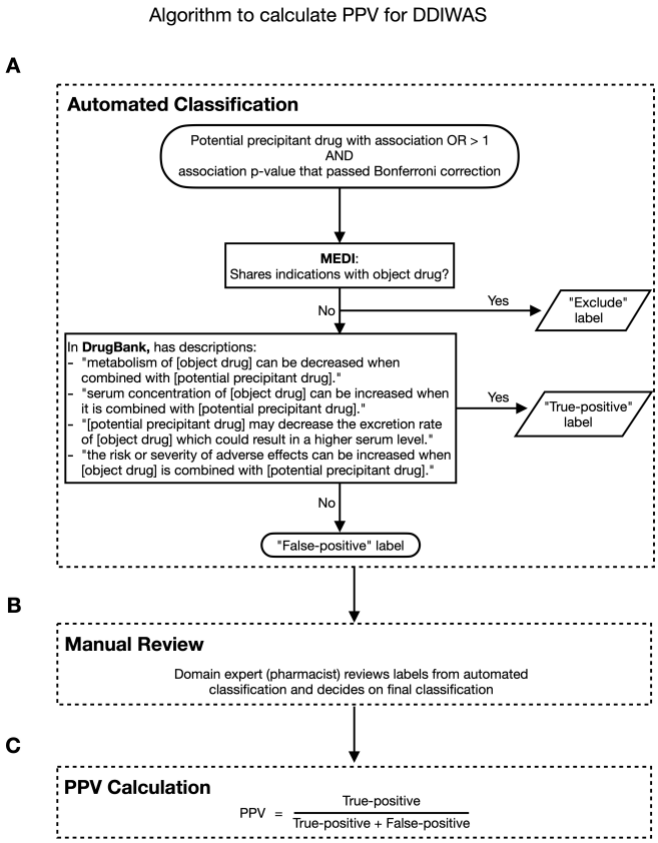
potential precipitant drug-ADR	object drug-ADR = 1	object drug-ADR = 0
1	≥1	≥1
0	≥1	≥1

H Run binary logistic regression to get odds ratios and p-values for association between potential precipitant drug of interest-ADR and object drug-ADR

$$p(+object\ drug-ADR|m_i) \sim$$
$$logit(\beta_0 + \beta_1 m_i + \beta_2 Age + \beta_3 Sex + \beta_4 Race + \beta_5 ObservationPeriod + \beta_6 NumDrugExposures)$$
$$where\ i = number\ of\ potential\ precipitant\ drugs\ tested$$

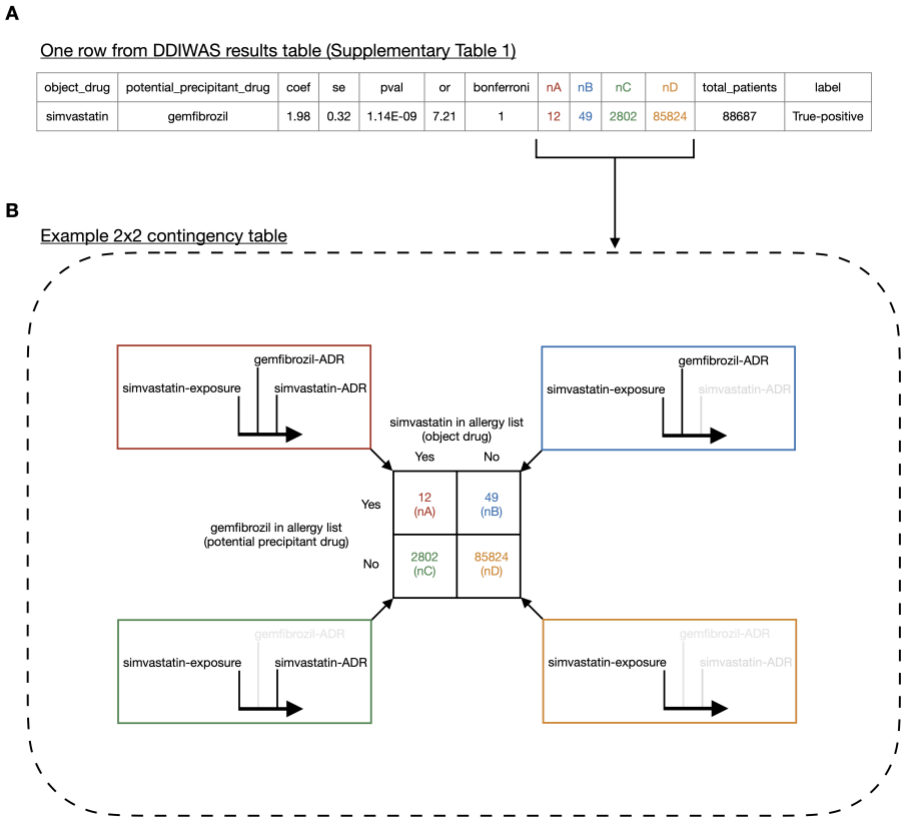
Supplementary Figure 1. Data processing and analysis overview.

(**A**) We identified patients who had ≥1 outpatient exposures to the object drug. Cases were patients who had the object drug listed in their EHR allergy list (+object drug-ADR), whereas controls were patients who did not (-object drug-ADR). (**B**) For each patient, we defined an observation period with a maximum length of 12 months. (**C**) We obtained the potential precipitant drugs that were first documented on each patient's allergy list during the defined observation period (+potential precipitant drug-ADR). The "object drug-ADR" column indicated whether each patient was case (= 1) or control (= 0). The "number of unique drug exposures" column indicated the number of unique drug ingredient(s) to which each patient was exposed during the observation period. We mapped the potential precipitant drugs from the allergy lists to RxCUIs using both brand and generic drug names, and then to their (**D**) RxCUI ingredient(s). (**E**) We removed patients who did not have ≥1 entries in their EHR allergy list during the observation period or were not between the 18-90 years at the start of the observation period. (**F**) We created a patient feature matrix with each row representing one patient and column representing a feature. For dichotomous features (sex, race, potential precipitant drug-ADRs, object drug-ADR), a value of 1 in a cell indicated that the feature was present in the medical record for that patient. (**G**) We only tested potential precipitant drugs if there were ≥1 patients in each cell of their respective 2x2 contingency tables. (**H**) To identify drugs that increased the risk of developing object drug-ADRs, we performed logistic regression with the dependent dichotomous variable indicating whether each patient experienced an object drug-ADR. The independent dichotomous variable (m_i) indicated whether a patient experienced ADR(s) to the potential precipitant drug being tested, adjusted for each patient's age at the start of the observation period, sex, race, number of unique drug ingredient exposures, and length of the observation period. The output from each regression were association OR and p-value. We used β_1 (beta-coefficient for the potential precipitant drug tested) to calculate the OR, which represented the degree of association between object drug-ADRs and potential precipitant drug-ADRs. Drugs that increased the risk of developing object drug-ADR(s) had association OR > 1 and p-value passing Bonferroni correction. EHR: electronic health record; ADR: adverse drug reaction; RxCUI: RxNorm concept unique identifier; OR: odds ratio.



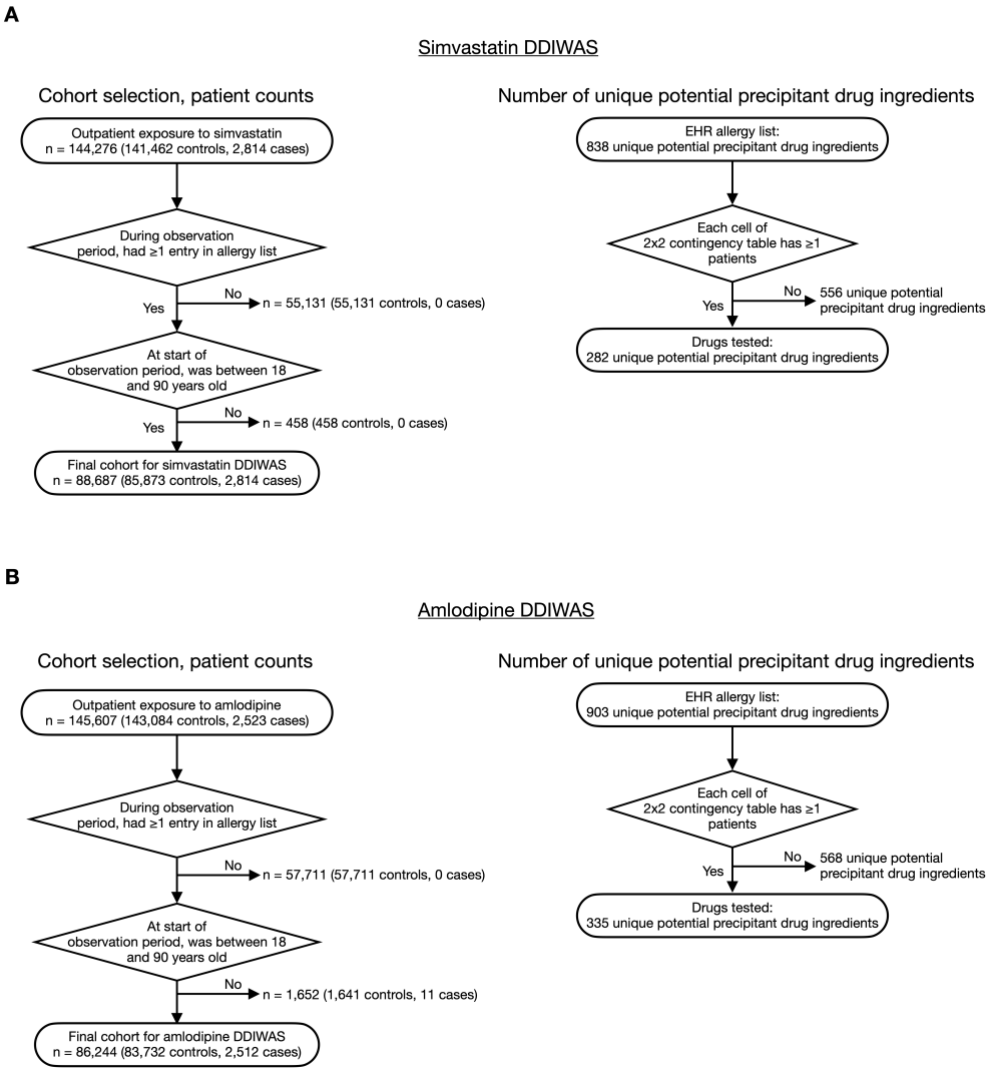
Supplementary Figure 2. Algorithm to calculate PPV for DDIWAS.

(**A**) We used an automated approach to classify the potential precipitant drugs that were significantly associated with object drug-ADRs. For the calculation of PPV, we used MEDI[23] to exclude drugs that shared main indication(s) with the object drug ("Exclude" label). To identify true-positive findings ("True-positive" label), we used DrugBank[16] as our gold standard reference. True-positive findings were potential precipitant drugs with DrugBank descriptions that predicted increased risk of adverse effect(s) when combined with the object drug. In the amlodipine DDIWAS, for example, prazosin was a true-positive finding, because in DrugBank, the "risk or severity of hypotension can be increased when Prazosin is combined with Amlodipine". We labeled potential precipitant drugs as false-positives ("False-positive" label) if they did not meet the criteria for true-positive findings. (**B**) A domain expert (S.N.) reviewed all labels and decided on the final classification for each drug. (**C**) We used the domain expert reviewed labels to calculate the PPV for each DDIWAS experiment. PPV: positive predictive value; DDIWAS: Drug-Drug Interaction Wide Association Study; MEDI: MEDication Indication resource; ADR: adverse drug reaction.



Supplementary Figure 3. Interpretation of DDIWAS results and 2x2 contingency table.

(**A**) One row from the simvastatin DDIWAS results table (**Supplementary Table 1**). In this analysis, simvastatin was the object drug and gemfibrozil was the potential precipitant drug. Columns “coef”, “se”, “pval”, and “or” contain the regression beta-coefficients, standard errors, p-values, and odds ratios, respectively for the potential precipitant drugs tested. Column “bonferroni” indicates whether the p-value for the potential precipitant drug tested passed Bonferroni correction. In this instance, documentation of gemfibrozil in the EHR allergy list increased a patients risk (OR > 1 and Bonferroni p-value < 0.05) of having simvastatin listed in their allergy list (i.e., patient potentially developed simvastatin-ADR). Column “label” indicates the final label for the object and potential precipitant drug pair after automated classification followed by domain expert review (**Supplementary Figure 2**). Columns “nA”, “nB”, “nC”, and “nD” correspond with the cells of (**B**) the 2x2 contingency table. The timelines illustrate the four DDIWAS-labeled types of patients in the 2x2 contingency table. Cell nA = number of patients who had both the object and potential precipitant drug listed on their allergy lists. Cell nB = number of patients who had the potential precipitant drug but not the object drug listed on their allergy lists. Cell nC = number of patients who had the object drug but not the potential precipitant drug listed on their allergy lists. Cell nD = number of patients who had neither drug listed on their allergy lists. DDIWAS: Drug-Drug Interaction Wide Association Study; ADR: adverse drug reaction; EHR: electronic health record; OR: odds ratio.



Supplementary Figure 4. Patient counts and number of unique potential precipitant drug ingredients.

(**A**) Simvastatin and (**B**) amlodipine DDIWAS. DDIWAS: Drug-Drug Interaction Wide Association Study; EHR: electronic health record.

Supplementary Table 1: Statistics for the primary simvastatin and amlodipine DDIWAS experiments.

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	pitavastatin	5.69	0.92	1.11E-16	297.29	1	10	1	2,804	85,872	88,687	Exclude
simvastatin	ezetimibe	5.28	0.16	2.72E-227	196.13	1	257	45	2,557	85,828	88,687	Exclude
simvastatin	pravastatin	3.98	0.16	4.74E-142	53.29	1	112	68	2,702	85,805	88,687	Exclude
simvastatin	colesevelam	3.82	0.43	1.26E-18	45.47	1	12	10	2,802	85,863	88,687	Exclude
simvastatin	fluvastatin	3.61	0.36	6.93E-24	37.09	1	17	15	2,797	85,858	88,687	Exclude
simvastatin	rosuvastatin	3.14	0.12	3.20E-143	23.07	1	121	162	2,693	85,711	88,687	Exclude
simvastatin	atorvastatin	2.85	0.08	1.43E-299	17.26	1	280	545	2,534	85,328	88,687	Exclude
simvastatin	fenofibrate	2.53	0.26	5.80E-22	12.57	1	21	50	2,793	85,823	88,687	True-positive
simvastatin	lovastatin	2.51	0.21	6.58E-34	12.26	1	34	81	2,780	85,792	88,687	Exclude
simvastatin	irbesartan	2.35	0.53	8.88E-06	10.50	1	5	13	2,809	85,860	88,687	True-positive
simvastatin	gemfibrozil	1.98	0.32	1.14E-09	7.21	1	12	49	2,802	85,824	88,687	True-positive
simvastatin	olmesartan	1.77	0.39	7.05E-06	5.86	1	8	37	2,806	85,836	88,687	True-positive
simvastatin	risedronate	1.71	0.45	1.42E-04	5.54	1	6	31	2,808	85,842	88,687	True-positive
simvastatin	triamterene	1.71	0.35	8.87E-07	5.52	1	10	52	2,804	85,821	88,687	False-positive

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	metoprolol	1.52	0.22	2.50E-12	4.56	1	25	165	2,789	85,708	88,687	True-positive
simvastatin	niacin	1.36	0.16	4.44E-17	3.90	1	44	349	2,770	85,524	88,687	True-positive
simvastatin	losartan	1.35	0.32	3.00E-05	3.86	1	11	82	2,803	85,791	88,687	True-positive
simvastatin	hydrochlorothiazide	1.30	0.22	2.63E-09	3.67	1	24	192	2,790	85,681	88,687	False-positive
simvastatin	alendronate	1.26	0.31	4.63E-05	3.53	1	12	93	2,802	85,780	88,687	True-positive
simvastatin	amlodipine	0.94	0.21	5.37E-06	2.56	1	26	285	2,788	85,588	88,687	True-positive
simvastatin	lisinopril	0.66	0.14	2.21E-06	1.93	1	57	853	2,757	85,020	88,687	True-positive
simvastatin	acetaminophen	-0.96	0.21	3.67E-06	0.38	1	24	1,800	2,790	84,073	88,687	NA
simvastatin	morphine	-1.42	0.23	8.71E-10	0.24	1	19	2,278	2,795	83,595	88,687	NA
simvastatin	codeine	-1.43	0.15	2.35E-21	0.24	1	45	5,387	2,769	80,486	88,687	NA
simvastatin	meperidine	-1.56	0.36	1.19E-05	0.21	1	8	1,103	2,806	84,770	88,687	NA
simvastatin	indapamide	4.10	1.20	3.45E-04	60.17	0	2	1	2,812	85,872	88,687	NA
simvastatin	cholestyramine resin	3.67	1.42	8.19E-03	39.21	0	1	1	2,813	85,872	88,687	NA
simvastatin	mexiletine	3.58	1.19	1.26E-03	35.91	0	2	1	2,812	85,872	88,687	NA
simvastatin	lactate	3.29	1.43	1.55E-02	26.94	0	1	1	2,813	85,872	88,687	NA
simvastatin	glucosamine	3.28	0.91	1.40E-03	26.54	0	2	3	2,812	85,870	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	rotigotine	3.27	1.42	1.55E-02	26.38	0	1	1	2,813	85,872	88,687	NA
simvastatin	urokinase	3.26	1.42	1.61E-02	26.17	0	1	1	2,813	85,872	88,687	NA
simvastatin	galantamine	3.20	1.41	1.73E-02	24.45	0	1	1	2,813	85,872	88,687	NA
simvastatin	fluvoxamine	3.14	1.42	1.89E-02	23.07	0	1	1	2,813	85,872	88,687	NA
simvastatin	water	2.93	1.20	2.19E-02	18.78	0	1	2	2,813	85,871	88,687	NA
simvastatin	denosumab	2.92	1.20	2.25E-02	18.60	0	1	2	2,813	85,871	88,687	NA
simvastatin	moexipril	2.87	1.10	2.19E-02	17.71	0	1	3	2,813	85,870	88,687	NA
simvastatin	estradiol	2.86	0.70	5.43E-04	17.53	0	3	6	2,811	85,867	88,687	NA
simvastatin	colestipol	2.86	0.73	6.38E-04	17.49	0	3	5	2,811	85,868	88,687	NA
simvastatin	apixaban	2.85	1.19	2.44E-02	17.35	0	1	2	2,813	85,871	88,687	NA
simvastatin	pentosan polysulfate	2.77	1.20	2.81E-02	15.98	0	1	2	2,813	85,871	88,687	NA
simvastatin	tapentadol	2.71	1.09	2.77E-02	15.10	0	1	3	2,813	85,870	88,687	NA
simvastatin	botulinum toxin type a	2.60	1.09	3.30E-02	13.50	0	1	3	2,813	85,870	88,687	NA
simvastatin	loperamide	2.55	1.12	3.75E-02	12.86	0	1	3	2,813	85,870	88,687	NA
simvastatin	miconazole	2.50	1.10	3.92E-02	12.18	0	1	3	2,813	85,870	88,687	NA
simvastatin	meclofenamate	2.46	1.10	4.19E-02	11.69	0	1	3	2,813	85,870	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	darifenacin	2.43	1.11	4.47E-02	11.32	0	1	3	2,813	85,870	88,687	NA
simvastatin	clorazepate	2.42	1.04	4.17E-02	11.27	0	1	4	2,813	85,869	88,687	NA
simvastatin	medroxyprogesterone	2.41	1.04	4.27E-02	11.14	0	1	4	2,813	85,869	88,687	NA
simvastatin	telmisartan	2.39	0.68	2.51E-03	10.95	0	3	8	2,811	85,865	88,687	NA
simvastatin	regular insulin, human	2.35	0.83	1.17E-02	10.45	0	2	5	2,812	85,868	88,687	NA
simvastatin	insulin, isophane	2.35	0.83	1.17E-02	10.45	0	2	5	2,812	85,868	88,687	NA
simvastatin	guanfacine	2.33	1.05	4.89E-02	10.30	0	1	4	2,813	85,869	88,687	NA
simvastatin	tegaserod	2.21	1.04	5.82E-02	9.11	0	1	4	2,813	85,869	88,687	NA
simvastatin	rabeprazole	2.09	0.74	1.78E-02	8.07	0	2	10	2,812	85,863	88,687	NA
simvastatin	teriparatide	2.08	1.02	7.10E-02	7.97	0	1	5	2,813	85,868	88,687	NA
simvastatin	nisoldipine	2.02	1.00	7.53E-02	7.53	0	1	5	2,813	85,868	88,687	NA
simvastatin	trandolapril	2.01	0.98	7.51E-02	7.45	0	1	6	2,813	85,867	88,687	NA
simvastatin	bisoprolol	2.00	0.98	7.59E-02	7.40	0	1	6	2,813	85,867	88,687	NA
simvastatin	torsemide	2.00	0.75	2.32E-02	7.37	0	2	9	2,812	85,864	88,687	NA
simvastatin	insulin detemir	1.98	0.74	2.37E-02	7.22	0	2	10	2,812	85,863	88,687	NA
simvastatin	choline	1.92	0.98	8.70E-02	6.80	0	1	6	2,813	85,867	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	candesartan	1.89	0.75	3.02E-02	6.63	0	2	10	2,812	85,863	88,687	NA
simvastatin	azelastine	1.88	0.74	2.92E-02	6.56	0	2	10	2,812	85,863	88,687	NA
simvastatin	fosinopril	1.83	0.61	1.14E-02	6.20	0	3	16	2,811	85,857	88,687	NA
simvastatin	insulin, aspart, human	1.79	1.03	1.12E-01	5.99	0	1	5	2,813	85,868	88,687	NA
simvastatin	chlorthalidone	1.79	0.97	1.04E-01	5.97	0	1	6	2,813	85,867	88,687	NA
simvastatin	cyproheptadine	1.77	0.95	1.05E-01	5.90	0	1	7	2,813	85,866	88,687	NA
simvastatin	phenyl salicylate	1.77	0.96	1.06E-01	5.87	0	1	7	2,813	85,866	88,687	NA
simvastatin	methylene blue	1.77	0.96	1.06E-01	5.87	0	1	7	2,813	85,866	88,687	NA
simvastatin	methenamine	1.77	0.96	1.06E-01	5.87	0	1	7	2,813	85,866	88,687	NA
simvastatin	finasteride	1.74	0.92	1.08E-01	5.69	0	1	9	2,813	85,864	88,687	NA
simvastatin	glyburide	1.74	0.53	5.97E-03	5.67	0	4	21	2,810	85,852	88,687	NA
simvastatin	pramipexole	1.73	0.61	1.59E-02	5.62	0	3	16	2,811	85,857	88,687	NA
simvastatin	benzoate	1.72	0.96	1.14E-01	5.57	0	1	7	2,813	85,866	88,687	NA
simvastatin	bumetanide	1.69	0.99	1.24E-01	5.40	0	1	6	2,813	85,867	88,687	NA
simvastatin	benzocaine	1.68	0.93	1.19E-01	5.35	0	1	9	2,813	85,864	88,687	NA
simvastatin	benzonatate	1.67	0.72	4.54E-02	5.33	0	2	13	2,812	85,860	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	testosterone	1.66	0.73	4.89E-02	5.25	0	2	12	2,812	85,861	88,687	NA
simvastatin	solifenacin	1.65	0.94	1.25E-01	5.21	0	1	8	2,813	85,865	88,687	NA
simvastatin	liraglutide	1.65	0.95	1.27E-01	5.18	0	1	7	2,813	85,866	88,687	NA
simvastatin	tadalafil	1.63	0.94	1.28E-01	5.12	0	1	8	2,813	85,865	88,687	NA
simvastatin	adalimumab	1.63	0.97	1.32E-01	5.10	0	1	6	2,813	85,867	88,687	NA
simvastatin	sildenafil	1.63	1.01	1.42E-01	5.09	0	1	6	2,813	85,867	88,687	NA
simvastatin	nefazodone	1.62	0.92	1.28E-01	5.07	0	1	10	2,813	85,863	88,687	NA
simvastatin	levalbuterol	1.60	0.96	1.38E-01	4.94	0	1	7	2,813	85,866	88,687	NA
simvastatin	salsalate	1.59	0.93	1.36E-01	4.89	0	1	9	2,813	85,864	88,687	NA
simvastatin	formaldehyde	1.58	0.92	1.36E-01	4.87	0	1	10	2,813	85,863	88,687	NA
simvastatin	cerivastatin	1.56	0.92	1.41E-01	4.76	0	1	10	2,813	85,863	88,687	NA
simvastatin	polyethylene glycol 3350	1.54	0.94	1.47E-01	4.67	0	1	8	2,813	85,865	88,687	NA
simvastatin	ketoconazole	1.52	0.94	1.52E-01	4.55	0	1	8	2,813	85,865	88,687	NA
simvastatin	glipizide	1.51	0.71	6.51E-02	4.54	0	2	14	2,812	85,859	88,687	NA
simvastatin	cetirizine	1.43	0.51	1.73E-02	4.18	0	4	32	2,810	85,841	88,687	NA
simvastatin	exenatide	1.38	0.44	1.83E-03	3.98	0	6	38	2,808	85,835	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	dexlansoprazole	1.38	0.92	1.83E-01	3.97	0	1	10	2,813	85,863	88,687	NA
simvastatin	diflunisal	1.37	0.92	1.85E-01	3.95	0	1	10	2,813	85,863	88,687	NA
simvastatin	mupirocin	1.37	0.92	1.87E-01	3.93	0	1	10	2,813	85,863	88,687	NA
simvastatin	brompheniramine	1.36	0.93	1.88E-01	3.91	0	1	9	2,813	85,864	88,687	NA
simvastatin	ipratropium	1.36	0.70	8.95E-02	3.90	0	2	17	2,812	85,856	88,687	NA
simvastatin	isosorbide	1.34	0.41	9.67E-04	3.81	0	7	51	2,807	85,822	88,687	NA
simvastatin	doxepin	1.34	0.90	1.93E-01	3.80	0	1	13	2,813	85,860	88,687	NA
simvastatin	atenolol	1.32	0.36	2.12E-04	3.73	0	9	72	2,805	85,801	88,687	NA
simvastatin	nebivolol	1.31	0.91	2.02E-01	3.71	0	1	10	2,813	85,863	88,687	NA
simvastatin	butalbital	1.29	0.91	2.06E-01	3.64	0	1	11	2,813	85,862	88,687	NA
simvastatin	ticlopidine	1.28	0.51	2.92E-02	3.60	0	4	38	2,810	85,835	88,687	NA
simvastatin	oxybutynin	1.28	0.69	1.06E-01	3.59	0	2	20	2,812	85,853	88,687	NA
simvastatin	omeprazole	1.26	0.43	3.70E-03	3.53	0	6	51	2,808	85,822	88,687	NA
simvastatin	flecainide	1.26	0.93	2.22E-01	3.52	0	1	9	2,813	85,864	88,687	NA
simvastatin	sitagliptin	1.25	0.58	6.14E-02	3.48	0	3	27	2,811	85,846	88,687	NA
simvastatin	spironolactone	1.24	0.38	1.03E-03	3.45	0	8	64	2,806	85,809	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	colchicine	1.24	0.70	1.18E-01	3.45	0	2	18	2,812	85,855	88,687	NA
simvastatin	methyldopa	1.24	0.91	2.25E-01	3.44	0	1	11	2,813	85,862	88,687	NA
simvastatin	dronedarone	1.22	0.90	2.28E-01	3.38	0	1	12	2,813	85,861	88,687	NA
simvastatin	esomeprazole	1.18	0.50	4.18E-02	3.25	0	4	44	2,810	85,829	88,687	NA
simvastatin	sotalol	1.17	0.58	7.62E-02	3.23	0	3	26	2,811	85,847	88,687	NA
simvastatin	montelukast	1.13	0.57	8.33E-02	3.10	0	3	33	2,811	85,840	88,687	NA
simvastatin	meclizine	1.10	0.68	1.52E-01	3.01	0	2	23	2,812	85,850	88,687	NA
simvastatin	clonidine	1.10	0.35	1.87E-03	3.00	0	9	82	2,805	85,791	88,687	NA
simvastatin	cimetidine	1.09	0.57	9.23E-02	2.98	0	3	35	2,811	85,838	88,687	NA
simvastatin	tolmetin	1.09	0.57	9.22E-02	2.98	0	3	37	2,811	85,836	88,687	NA
simvastatin	glimepiride	1.09	0.88	2.72E-01	2.96	0	1	16	2,813	85,857	88,687	NA
simvastatin	minocycline	1.08	0.57	9.44E-02	2.96	0	3	34	2,811	85,839	88,687	NA
simvastatin	benazepril	1.08	0.37	3.61E-03	2.95	0	8	82	2,806	85,791	88,687	NA
simvastatin	verapamil	1.07	0.43	1.28E-02	2.92	0	6	60	2,808	85,813	88,687	NA
simvastatin	sertraline	1.05	0.35	2.96E-03	2.84	0	9	92	2,805	85,781	88,687	NA
simvastatin	doxazosin	1.03	0.68	1.75E-01	2.81	0	2	26	2,812	85,847	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	minoxidil	1.03	0.68	1.78E-01	2.79	0	2	26	2,812	85,847	88,687	NA
simvastatin	triamcinolone	1.02	0.67	1.80E-01	2.76	0	2	27	2,812	85,846	88,687	NA
simvastatin	pantoprazole	1.00	0.57	1.20E-01	2.71	0	3	33	2,811	85,840	88,687	NA
simvastatin	fenoprofen	1.00	0.89	3.08E-01	2.71	0	1	15	2,813	85,858	88,687	NA
simvastatin	mometasone	1.00	0.88	3.08E-01	2.70	0	1	16	2,813	85,857	88,687	NA
simvastatin	ergotamine	0.99	0.88	3.09E-01	2.70	0	1	16	2,813	85,857	88,687	NA
simvastatin	insulin glargine	0.96	0.90	3.26E-01	2.62	0	1	12	2,813	85,861	88,687	NA
simvastatin	digoxin	0.96	0.68	2.03E-01	2.61	0	2	27	2,812	85,846	88,687	NA
simvastatin	sulindac	0.96	0.68	2.03E-01	2.61	0	2	26	2,812	85,847	88,687	NA
simvastatin	ranolazine	0.95	0.88	3.30E-01	2.57	0	1	16	2,813	85,857	88,687	NA
simvastatin	propranolol	0.93	0.67	2.15E-01	2.53	0	2	31	2,812	85,842	88,687	NA
simvastatin	allopurinol	0.92	0.43	3.21E-02	2.51	0	6	66	2,808	85,807	88,687	NA
simvastatin	amitriptyline	0.91	0.35	8.86E-03	2.50	0	9	103	2,805	85,770	88,687	NA
simvastatin	fluconazole	0.89	0.57	1.58E-01	2.44	0	3	38	2,811	85,835	88,687	NA
simvastatin	valdecoxib	0.89	0.57	1.58E-01	2.43	0	3	40	2,811	85,833	88,687	NA
simvastatin	enalapril	0.88	0.40	2.58E-02	2.41	0	7	82	2,807	85,791	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	rosiglitazone	0.88	0.57	1.62E-01	2.41	0	3	41	2,811	85,832	88,687	NA
simvastatin	donepezil	0.86	0.57	1.70E-01	2.37	0	3	41	2,811	85,832	88,687	NA
simvastatin	famotidine	0.85	0.88	3.73E-01	2.35	0	1	16	2,813	85,857	88,687	NA
simvastatin	linezolid	0.85	0.88	3.75E-01	2.34	0	1	18	2,813	85,855	88,687	NA
simvastatin	rifampin	0.85	0.68	2.56E-01	2.34	0	2	24	2,812	85,849	88,687	NA
simvastatin	metformin	0.83	0.23	3.75E-04	2.30	0	20	245	2,794	85,628	88,687	NA
simvastatin	escitalopram	0.83	0.49	1.32E-01	2.29	0	4	60	2,810	85,813	88,687	NA
simvastatin	triprolidine	0.83	0.88	3.86E-01	2.29	0	1	18	2,813	85,855	88,687	NA
simvastatin	propafenone	0.82	0.87	3.89E-01	2.27	0	1	18	2,813	85,855	88,687	NA
simvastatin	procaine	0.81	0.67	2.71E-01	2.25	0	2	31	2,812	85,842	88,687	NA
simvastatin	ropinirole	0.80	0.67	2.77E-01	2.23	0	2	27	2,812	85,846	88,687	NA
simvastatin	terfenadine	0.79	0.87	4.04E-01	2.20	0	1	20	2,813	85,853	88,687	NA
simvastatin	quinapril	0.76	0.46	1.03E-01	2.13	0	5	69	2,809	85,804	88,687	NA
simvastatin	terazosin	0.75	0.86	4.21E-01	2.13	0	1	23	2,813	85,850	88,687	NA
simvastatin	diclofenac	0.75	0.43	7.71E-02	2.12	0	6	79	2,808	85,794	88,687	NA
simvastatin	quetiapine	0.72	0.67	3.19E-01	2.06	0	2	34	2,812	85,839	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	albuterol	0.71	0.49	1.86E-01	2.04	0	4	64	2,810	85,809	88,687	NA
simvastatin	chlorpheniramine	0.71	0.87	4.48E-01	2.03	0	1	21	2,813	85,852	88,687	NA
simvastatin	cefadroxil	0.70	0.87	4.51E-01	2.02	0	1	21	2,813	85,852	88,687	NA
simvastatin	diltiazem	0.70	0.35	4.38E-02	2.01	0	9	126	2,805	85,747	88,687	NA
simvastatin	clopidogrel	0.68	0.27	1.11E-02	1.98	0	15	221	2,799	85,652	88,687	NA
simvastatin	estrogens, conjugated (usp)	0.67	0.86	4.70E-01	1.96	0	1	22	2,813	85,851	88,687	NA
simvastatin	fluticasone	0.67	0.46	1.51E-01	1.95	0	5	70	2,809	85,803	88,687	NA
simvastatin	thimerosal	0.66	0.87	4.76E-01	1.94	0	1	21	2,813	85,852	88,687	NA
simvastatin	sulfasalazine	0.65	0.87	4.83E-01	1.92	0	1	20	2,813	85,853	88,687	NA
simvastatin	alprazolam	0.64	0.66	3.75E-01	1.89	0	2	40	2,812	85,833	88,687	NA
simvastatin	ibandronate	0.63	0.87	4.95E-01	1.88	0	1	20	2,813	85,853	88,687	NA
simvastatin	budesonide	0.63	0.87	4.95E-01	1.88	0	1	21	2,813	85,852	88,687	NA
simvastatin	lansoprazole	0.62	0.66	3.83E-01	1.86	0	2	39	2,812	85,834	88,687	NA
simvastatin	paroxetine	0.61	0.49	2.51E-01	1.84	0	4	68	2,810	85,805	88,687	NA
simvastatin	fexofenadine	0.61	0.56	3.14E-01	1.83	0	3	55	2,811	85,818	88,687	NA
simvastatin	levothyroxine	0.59	0.86	5.23E-01	1.80	0	1	25	2,813	85,848	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	citalopram	0.58	0.49	2.70E-01	1.79	0	4	71	2,810	85,802	88,687	NA
simvastatin	cyclosporine	0.58	0.88	5.32E-01	1.79	0	1	19	2,813	85,854	88,687	NA
simvastatin	pioglitazone	0.54	0.49	3.05E-01	1.72	0	4	71	2,810	85,802	88,687	NA
simvastatin	oxcarbazepine	0.52	0.87	5.69E-01	1.68	0	1	21	2,813	85,852	88,687	NA
simvastatin	piroxicam	0.52	0.66	4.62E-01	1.68	0	2	44	2,812	85,829	88,687	NA
simvastatin	epinephrine	0.50	0.49	3.34E-01	1.66	0	4	82	2,810	85,791	88,687	NA
simvastatin	atropine	0.48	0.66	4.88E-01	1.62	0	2	44	2,812	85,829	88,687	NA
simvastatin	captopril	0.47	0.86	6.01E-01	1.60	0	1	25	2,813	85,848	88,687	NA
simvastatin	tizanidine	0.47	0.86	6.01E-01	1.60	0	1	27	2,813	85,846	88,687	NA
simvastatin	azathioprine	0.47	0.86	6.04E-01	1.59	0	1	26	2,813	85,847	88,687	NA
simvastatin	carvedilol	0.46	0.56	4.33E-01	1.59	0	3	61	2,811	85,812	88,687	NA
simvastatin	buspirone	0.45	0.85	6.18E-01	1.56	0	1	29	2,813	85,844	88,687	NA
simvastatin	clonazepam	0.40	0.86	6.57E-01	1.48	0	1	28	2,813	85,845	88,687	NA
simvastatin	hydralazine	0.39	0.66	5.69E-01	1.48	0	2	51	2,812	85,822	88,687	NA
simvastatin	nortriptyline	0.39	0.86	6.64E-01	1.47	0	1	28	2,813	85,845	88,687	NA
simvastatin	lorazepam	0.34	0.46	4.62E-01	1.40	0	5	101	2,809	85,772	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	ramipril	0.32	0.39	4.16E-01	1.37	0	7	152	2,807	85,721	88,687	NA
simvastatin	gabapentin	0.30	0.30	3.18E-01	1.34	0	12	253	2,802	85,620	88,687	NA
simvastatin	dipyridamole	0.29	0.85	7.37E-01	1.34	0	1	35	2,813	85,838	88,687	NA
simvastatin	ceftriaxone	0.27	0.49	5.90E-01	1.31	0	4	96	2,810	85,777	88,687	NA
simvastatin	loratadine	0.25	0.65	7.13E-01	1.28	0	2	56	2,812	85,817	88,687	NA
simvastatin	potassium	0.24	0.85	7.85E-01	1.27	0	1	36	2,813	85,837	88,687	NA
simvastatin	salmeterol	0.21	0.65	7.56E-01	1.23	0	2	53	2,812	85,820	88,687	NA
simvastatin	valsartan	0.19	0.55	7.39E-01	1.21	0	3	84	2,811	85,789	88,687	NA
simvastatin	warfarin	0.16	0.55	7.74E-01	1.17	0	3	87	2,811	85,786	88,687	NA
simvastatin	bupropion	0.16	0.42	7.04E-01	1.17	0	6	153	2,808	85,720	88,687	NA
simvastatin	gentamicin	0.16	0.85	8.55E-01	1.17	0	1	36	2,813	85,837	88,687	NA
simvastatin	trazodone	0.15	0.65	8.17E-01	1.17	0	2	59	2,812	85,814	88,687	NA
simvastatin	doxycycline	0.15	0.30	6.14E-01	1.16	0	12	296	2,802	85,577	88,687	NA
simvastatin	tamsulosin	0.14	0.84	8.65E-01	1.16	0	1	45	2,813	85,828	88,687	NA
simvastatin	etodolac	0.14	0.65	8.29E-01	1.15	0	2	65	2,812	85,808	88,687	NA
simvastatin	chlorpromazine	0.10	0.84	9.06E-01	1.10	0	1	39	2,813	85,834	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	venlafaxine	0.10	0.65	8.80E-01	1.10	0	2	65	2,812	85,808	88,687	NA
simvastatin	caffeine	0.09	0.65	8.88E-01	1.10	0	2	68	2,812	85,805	88,687	NA
simvastatin	mirtazapine	0.07	0.85	9.32E-01	1.08	0	1	36	2,813	85,837	88,687	NA
simvastatin	gatifloxacin	0.07	0.84	9.33E-01	1.07	0	1	39	2,813	85,834	88,687	NA
simvastatin	phenazopyridine	0.07	0.85	9.36E-01	1.07	0	1	38	2,813	85,835	88,687	NA
simvastatin	oxytetracycline	0.06	0.84	9.43E-01	1.06	0	1	40	2,813	85,833	88,687	NA
simvastatin	furosemide	0.05	0.66	9.35E-01	1.05	0	2	58	2,812	85,815	88,687	NA
simvastatin	nifedipine	0.05	0.46	9.18E-01	1.05	0	5	134	2,809	85,739	88,687	NA
simvastatin	timolol	0.04	0.84	9.61E-01	1.04	0	1	42	2,813	85,831	88,687	NA
simvastatin	amiodarone	0.02	0.55	9.71E-01	1.02	0	3	89	2,811	85,784	88,687	NA
simvastatin	duloxetine	0.01	0.55	9.79E-01	1.01	0	3	100	2,811	85,773	88,687	NA
simvastatin	hydroxyzine	0.01	0.65	9.83E-01	1.01	0	2	70	2,812	85,803	88,687	NA
simvastatin	dexamethasone	0.01	0.84	9.91E-01	1.01	0	1	41	2,813	85,832	88,687	NA
simvastatin	topiramate	-0.01	0.65	9.92E-01	0.99	0	2	76	2,812	85,797	88,687	NA
simvastatin	phenobarbital	-0.02	0.65	9.71E-01	0.98	0	2	77	2,812	85,796	88,687	NA
simvastatin	pregabalin	-0.03	0.36	9.40E-01	0.97	0	8	232	2,806	85,641	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	ranitidine	-0.08	0.84	9.23E-01	0.92	0	1	49	2,813	85,824	88,687	NA
simvastatin	cyclobenzaprine	-0.13	0.55	8.12E-01	0.88	0	3	114	2,811	85,759	88,687	NA
simvastatin	meloxicam	-0.21	0.54	6.82E-01	0.81	0	3	127	2,811	85,746	88,687	NA
simvastatin	varenicline	-0.23	0.84	7.78E-01	0.80	0	1	52	2,813	85,821	88,687	NA
simvastatin	lidocaine	-0.24	0.65	6.97E-01	0.79	0	2	94	2,812	85,779	88,687	NA
simvastatin	zolpidem	-0.25	0.45	5.88E-01	0.78	0	5	176	2,809	85,697	88,687	NA
simvastatin	hydroxychloroquine	-0.26	0.84	7.41E-01	0.77	0	1	50	2,813	85,823	88,687	NA
simvastatin	celecoxib	-0.29	0.36	4.16E-01	0.75	0	8	319	2,806	85,554	88,687	NA
simvastatin	bacitracin	-0.31	0.84	6.98E-01	0.74	0	1	57	2,813	85,816	88,687	NA
simvastatin	heparin	-0.32	0.55	5.31E-01	0.72	0	3	137	2,811	85,736	88,687	NA
simvastatin	azithromycin	-0.34	0.41	4.16E-01	0.71	0	6	243	2,808	85,630	88,687	NA
simvastatin	ampicillin	-0.34	0.48	4.49E-01	0.71	0	4	189	2,810	85,684	88,687	NA
simvastatin	rofecoxib	-0.34	0.48	4.49E-01	0.71	0	4	184	2,810	85,689	88,687	NA
simvastatin	levofloxacin	-0.36	0.23	1.24E-01	0.70	0	19	768	2,795	85,105	88,687	NA
simvastatin	moxifloxacin	-0.36	0.54	4.78E-01	0.70	0	3	144	2,811	85,729	88,687	NA
simvastatin	neomycin	-0.37	0.83	6.38E-01	0.69	0	1	64	2,813	85,809	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	fluoxetine	-0.37	0.83	6.31E-01	0.69	0	1	64	2,813	85,809	88,687	NA
simvastatin	tramadol	-0.40	0.34	2.35E-01	0.67	0	9	395	2,805	85,478	88,687	NA
simvastatin	pentazocine	-0.40	0.48	3.66E-01	0.67	0	4	194	2,810	85,679	88,687	NA
simvastatin	clindamycin	-0.43	0.45	3.41E-01	0.65	0	5	223	2,809	85,650	88,687	NA
simvastatin	guaifenesin	-0.44	0.83	5.63E-01	0.64	0	1	69	2,813	85,804	88,687	NA
simvastatin	fentanyl	-0.44	0.65	4.57E-01	0.64	0	2	101	2,812	85,772	88,687	NA
simvastatin	clarithromycin	-0.46	0.41	2.67E-01	0.63	0	6	269	2,808	85,604	88,687	NA
simvastatin	prochlorperazine	-0.49	0.48	2.71E-01	0.61	0	4	207	2,810	85,666	88,687	NA
simvastatin	pseudoephedrine	-0.53	0.64	3.69E-01	0.59	0	2	131	2,812	85,742	88,687	NA
simvastatin	oxycodone	-0.53	0.22	1.72E-02	0.59	0	21	1,028	2,793	84,845	88,687	NA
simvastatin	metronidazole	-0.58	0.54	2.34E-01	0.56	0	3	187	2,811	85,686	88,687	NA
simvastatin	ciprofloxacin	-0.60	0.27	2.66E-02	0.55	0	14	747	2,800	85,126	88,687	NA
simvastatin	nalbuphine	-0.61	0.83	4.12E-01	0.54	0	1	81	2,813	85,792	88,687	NA
simvastatin	hydrocodone	-0.61	0.20	2.07E-03	0.54	0	26	1,375	2,788	84,498	88,687	NA
simvastatin	ondansetron	-0.62	0.83	4.08E-01	0.54	0	1	69	2,813	85,804	88,687	NA
simvastatin	quinine	-0.62	0.64	2.83E-01	0.54	0	2	137	2,812	85,736	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	aspirin	-0.62	0.23	6.04E-03	0.54	0	20	1,137	2,794	84,736	88,687	NA
simvastatin	sulfamethoxazole	-0.66	0.27	1.44E-02	0.52	0	14	788	2,800	85,085	88,687	NA
simvastatin	erythromycin	-0.70	0.26	7.60E-03	0.50	0	15	866	2,799	85,007	88,687	NA
simvastatin	clavulanate	-0.74	0.38	5.34E-02	0.48	0	7	424	2,807	85,449	88,687	NA
simvastatin	trimethoprim	-0.76	0.28	6.72E-03	0.47	0	13	803	2,801	85,070	88,687	NA
simvastatin	nitrofurantoin	-0.76	0.45	9.23E-02	0.47	0	5	298	2,809	85,575	88,687	NA
simvastatin	amoxicillin	-0.78	0.27	4.13E-03	0.46	0	14	907	2,800	84,966	88,687	NA
simvastatin	ketorolac	-0.79	0.54	9.40E-02	0.45	0	3	224	2,811	85,649	88,687	NA
simvastatin	tetracycline	-0.80	0.38	3.72E-02	0.45	0	7	458	2,807	85,415	88,687	NA
simvastatin	indomethacin	-0.80	0.83	2.65E-01	0.45	0	1	98	2,813	85,775	88,687	NA
simvastatin	carbamazepine	-0.80	0.83	2.66E-01	0.45	0	1	93	2,813	85,780	88,687	NA
simvastatin	ibuprofen	-0.82	0.34	1.43E-02	0.44	0	9	621	2,805	85,252	88,687	NA
simvastatin	cefuroxime	-0.86	0.83	2.24E-01	0.42	0	1	102	2,813	85,771	88,687	NA
simvastatin	naproxen	-0.94	0.41	2.30E-02	0.39	0	6	456	2,808	85,417	88,687	NA
simvastatin	iodine	-0.94	0.36	8.57E-03	0.39	0	8	622	2,806	85,251	88,687	NA
simvastatin	hydromorphone	-0.95	0.45	3.50E-02	0.39	0	5	340	2,809	85,533	88,687	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
simvastatin	prednisone	-0.96	0.54	3.61E-02	0.38	0	3	266	2,811	85,607	88,687	NA
simvastatin	diazepam	-0.97	0.64	7.33E-02	0.38	0	2	187	2,812	85,686	88,687	NA
simvastatin	cephalexin	-1.01	0.30	9.07E-04	0.36	0	11	883	2,803	84,990	88,687	NA
simvastatin	butorphanol	-1.04	0.83	1.29E-01	0.35	0	1	126	2,813	85,747	88,687	NA
simvastatin	metoclopramide	-1.19	0.64	2.12E-02	0.30	0	2	232	2,812	85,641	88,687	NA
simvastatin	promethazine	-1.19	0.45	8.16E-03	0.30	0	5	450	2,809	85,423	88,687	NA
simvastatin	vancomycin	-1.24	0.64	1.56E-02	0.29	0	2	220	2,812	85,653	88,687	NA
simvastatin	propoxyphene	-1.26	0.41	2.22E-03	0.28	0	6	610	2,808	85,263	88,687	NA
simvastatin	povidone-iodine	-1.32	0.82	4.08E-02	0.27	0	1	158	2,813	85,715	88,687	NA
simvastatin	sulfur	-1.36	0.64	6.27E-03	0.26	0	2	301	2,812	85,572	88,687	NA
simvastatin	cefaclor	-1.44	0.64	3.20E-03	0.24	0	2	313	2,812	85,560	88,687	NA
simvastatin	diphenhydramine	-1.74	0.82	3.22E-03	0.17	0	1	245	2,813	85,628	88,687	NA
amlodipine	benazepril	4.22	0.13	9.89E-243	67.96	1	185	103	2,327	83,629	86,244	Exclude
amlodipine	prazosin	3.40	0.68	1.40E-05	30.04	1	4	5	2,508	83,727	86,244	True-positive
amlodipine	felodipine	3.29	0.46	1.38E-12	26.90	1	9	10	2,503	83,722	86,244	True-positive
amlodipine	olmesartan	2.97	0.20	1.18E-51	19.50	1	42	75	2,470	83,657	86,244	Exclude

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	aliskiren	2.94	0.38	5.83E-15	18.99	1	11	21	2,501	83,711	86,244	Exclude
amlodipine	indapamide	2.88	0.63	8.87E-05	17.78	1	4	7	2,508	83,725	86,244	True-positive
amlodipine	candesartan	2.80	0.42	2.72E-11	16.52	1	9	16	2,503	83,716	86,244	Exclude
amlodipine	irbesartan	2.78	0.29	1.79E-21	16.16	1	18	36	2,494	83,696	86,244	Exclude
amlodipine	valsartan	2.49	0.17	6.12E-46	12.02	1	46	128	2,466	83,604	86,244	Exclude
amlodipine	metoprolol	2.35	0.15	1.91E-56	10.52	1	62	194	2,450	83,538	86,244	True-positive
amlodipine	spironolactone	2.34	0.25	2.90E-20	10.40	1	21	67	2,491	83,665	86,244	True-positive
amlodipine	valacyclovir	2.32	0.52	8.22E-06	10.17	1	5	15	2,507	83,717	86,244	False-positive
amlodipine	chlorthalidone	2.30	0.37	4.28E-10	10.02	1	10	30	2,502	83,702	86,244	True-positive
amlodipine	nifedipine	2.28	0.15	1.66E-55	9.80	1	64	212	2,448	83,520	86,244	True-positive
amlodipine	doxazosin	2.26	0.36	5.55E-10	9.56	1	10	33	2,502	83,699	86,244	Exclude
amlodipine	minoxidil	2.21	0.38	4.05E-09	9.14	1	9	36	2,503	83,696	86,244	True-positive
amlodipine	atenolol	2.20	0.20	2.45E-28	9.05	1	33	119	2,479	83,613	86,244	True-positive
amlodipine	hydralazine	2.18	0.21	9.16E-25	8.88	1	29	108	2,483	83,624	86,244	True-positive
amlodipine	carvedilol	2.17	0.25	2.21E-18	8.77	1	21	79	2,491	83,653	86,244	True-positive
amlodipine	labetalol	2.17	0.46	2.97E-06	8.72	1	6	23	2,506	83,709	86,244	True-positive

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	losartan	2.14	0.14	6.55E-51	8.52	1	64	251	2,448	83,481	86,244	Exclude
amlodipine	nebivolol	2.09	0.38	2.68E-08	8.05	1	9	37	2,503	83,695	86,244	True-positive
amlodipine	clonidine	1.96	0.19	1.44E-24	7.12	1	34	155	2,478	83,577	86,244	True-positive
amlodipine	telmisartan	1.96	0.46	1.92E-05	7.07	1	6	25	2,506	83,707	86,244	Exclude
amlodipine	hydrochlorothiazide	1.95	0.12	1.18E-62	7.01	1	94	436	2,418	83,296	86,244	True-positive
amlodipine	verapamil	1.94	0.24	1.88E-15	6.95	1	21	94	2,491	83,638	86,244	True-positive
amlodipine	levothyroxine	1.90	0.50	1.45E-04	6.70	1	5	21	2,507	83,711	86,244	False-positive
amlodipine	diltiazem	1.87	0.20	2.17E-21	6.50	1	32	150	2,480	83,582	86,244	True-positive
amlodipine	triamterene	1.87	0.26	9.59E-13	6.47	1	18	88	2,494	83,644	86,244	True-positive
amlodipine	quinapril	1.67	0.29	9.10E-09	5.31	1	14	86	2,498	83,646	86,244	Exclude
amlodipine	isosorbide	1.66	0.31	1.33E-07	5.24	1	12	74	2,500	83,658	86,244	True-positive
amlodipine	furosemide	1.65	0.29	1.47E-08	5.22	1	14	83	2,498	83,649	86,244	True-positive
amlodipine	ramipril	1.55	0.24	4.70E-11	4.73	1	21	140	2,491	83,592	86,244	Exclude
amlodipine	omeprazole	1.49	0.34	1.31E-05	4.43	1	10	66	2,502	83,666	86,244	False-positive
amlodipine	lisinopril	1.43	0.08	5.04E-76	4.17	1	204	1,806	2,308	81,926	86,244	Exclude
amlodipine	atorvastatin	1.28	0.14	1.67E-20	3.60	1	60	560	2,452	83,172	86,244	True-positive

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	enalapril	1.22	0.29	3.23E-05	3.39	1	13	122	2,499	83,610	86,244	Exclude
amlodipine	rosuvastatin	1.19	0.21	2.14E-08	3.27	1	25	247	2,487	83,485	86,244	True-positive
amlodipine	ezetimibe	1.11	0.28	8.48E-05	3.03	1	14	138	2,498	83,594	86,244	False-positive
amlodipine	pravastatin	1.11	0.26	1.41E-05	3.03	1	17	182	2,495	83,550	86,244	True-positive
amlodipine	simvastatin	1.08	0.15	1.49E-12	2.95	1	48	525	2,464	83,207	86,244	True-positive
amlodipine	codeine	-1.12	0.14	2.17E-15	0.33	1	52	4,611	2,460	79,121	86,244	NA
amlodipine	isradipine	3.61	1.42	8.96E-03	37.06	0	1	1	2,511	83,731	86,244	NA
amlodipine	calcipotriene	3.56	1.42	1.00E-02	35.12	0	1	1	2,511	83,731	86,244	NA
amlodipine	gramicidin	3.53	1.00	9.44E-04	34.01	0	2	2	2,510	83,730	86,244	NA
amlodipine	cevimeline	3.51	1.46	1.22E-02	33.57	0	1	1	2,511	83,731	86,244	NA
amlodipine	naratriptan	3.49	1.42	1.10E-02	32.83	0	1	1	2,511	83,731	86,244	NA
amlodipine	vitamin b 12	3.40	1.44	1.35E-02	29.82	0	1	1	2,511	83,731	86,244	NA
amlodipine	folic acid	3.40	1.44	1.35E-02	29.82	0	1	1	2,511	83,731	86,244	NA
amlodipine	roflumilast	3.34	1.43	1.43E-02	28.36	0	1	1	2,511	83,731	86,244	NA
amlodipine	latanoprost	3.29	1.01	1.80E-03	26.78	0	2	2	2,510	83,730	86,244	NA
amlodipine	medroxyprogesterone	3.25	1.42	1.59E-02	25.83	0	1	1	2,511	83,731	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	pyridostigmine	3.25	1.20	1.33E-02	25.67	0	1	2	2,511	83,730	86,244	NA
amlodipine	methylphenidate	3.10	1.42	2.01E-02	22.27	0	1	1	2,511	83,731	86,244	NA
amlodipine	dexlansoprazole	3.08	0.87	2.19E-03	21.75	0	2	4	2,510	83,728	86,244	NA
amlodipine	febuxostat	3.07	0.91	2.41E-03	21.60	0	2	3	2,510	83,729	86,244	NA
amlodipine	azelastine	3.06	0.92	2.60E-03	21.27	0	2	3	2,510	83,729	86,244	NA
amlodipine	trihexyphenidyl	3.05	1.42	2.21E-02	21.04	0	1	1	2,511	83,731	86,244	NA
amlodipine	isopropyl alcohol	3.04	1.20	1.87E-02	20.83	0	1	2	2,511	83,730	86,244	NA
amlodipine	iodine povacrylex	3.04	1.20	1.87E-02	20.83	0	1	2	2,511	83,730	86,244	NA
amlodipine	epoetin alfa	2.99	1.20	2.00E-02	19.94	0	1	2	2,511	83,730	86,244	NA
amlodipine	rabeprazole	2.90	0.71	5.23E-04	18.09	0	3	6	2,509	83,726	86,244	NA
amlodipine	simethicone	2.87	1.21	2.46E-02	17.66	0	1	2	2,511	83,730	86,244	NA
amlodipine	nitroprusside	2.86	1.20	2.42E-02	17.54	0	1	2	2,511	83,730	86,244	NA
amlodipine	vardenafil	2.85	1.09	2.23E-02	17.32	0	1	3	2,511	83,729	86,244	NA
amlodipine	insulin, aspart, human	2.83	1.20	2.58E-02	16.89	0	1	2	2,511	83,730	86,244	NA
amlodipine	domperidone	2.82	1.23	2.73E-02	16.84	0	1	2	2,511	83,730	86,244	NA
amlodipine	insulin lispro	2.82	1.20	2.62E-02	16.79	0	1	2	2,511	83,730	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	dichloralphenazone	2.81	1.20	2.69E-02	16.53	0	1	2	2,511	83,730	86,244	NA
amlodipine	flecainide	2.73	0.86	4.86E-03	15.33	0	2	4	2,510	83,728	86,244	NA
amlodipine	loperamide	2.68	1.13	3.09E-02	14.62	0	1	3	2,511	83,729	86,244	NA
amlodipine	chlorzoxazone	2.67	1.10	2.97E-02	14.51	0	1	3	2,511	83,729	86,244	NA
amlodipine	moexipril	2.66	0.71	1.09E-03	14.35	0	3	6	2,509	83,726	86,244	NA
amlodipine	flunisolide	2.64	1.04	2.95E-02	14.01	0	1	4	2,511	83,728	86,244	NA
amlodipine	guanfacine	2.60	0.59	2.37E-04	13.48	0	4	10	2,508	83,722	86,244	NA
amlodipine	eplerenone	2.60	1.12	3.46E-02	13.45	0	1	3	2,511	83,729	86,244	NA
amlodipine	clidinium	2.59	0.80	6.08E-03	13.27	0	2	6	2,510	83,726	86,244	NA
amlodipine	cyclophosphamide	2.57	1.10	3.51E-02	13.02	0	1	3	2,511	83,729	86,244	NA
amlodipine	valganciclovir	2.52	1.10	3.82E-02	12.38	0	1	3	2,511	83,729	86,244	NA
amlodipine	glucosamine	2.51	0.83	7.83E-03	12.33	0	2	5	2,510	83,727	86,244	NA
amlodipine	isometheptene	2.50	1.10	3.91E-02	12.21	0	1	3	2,511	83,729	86,244	NA
amlodipine	clorazepate	2.49	1.10	4.00E-02	12.03	0	1	3	2,511	83,729	86,244	NA
amlodipine	estradiol	2.47	1.10	4.09E-02	11.88	0	1	3	2,511	83,729	86,244	NA
amlodipine	mirabegron	2.47	1.10	4.15E-02	11.80	0	1	3	2,511	83,729	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	amiloride	2.46	1.10	4.25E-02	11.66	0	1	3	2,511	83,729	86,244	NA
amlodipine	zonisamide	2.45	0.82	8.85E-03	11.59	0	2	5	2,510	83,727	86,244	NA
amlodipine	trifluoperazine	2.44	1.05	4.16E-02	11.49	0	1	4	2,511	83,728	86,244	NA
amlodipine	docetaxel	2.44	1.05	4.12E-02	11.45	0	1	4	2,511	83,728	86,244	NA
amlodipine	ethacrynate	2.38	1.10	4.75E-02	10.82	0	1	3	2,511	83,729	86,244	NA
amlodipine	ergotamine	2.33	1.04	4.87E-02	10.25	0	1	4	2,511	83,728	86,244	NA
amlodipine	perindopril	2.28	1.04	5.28E-02	9.73	0	1	4	2,511	83,728	86,244	NA
amlodipine	sodium sulfate	2.27	1.04	5.35E-02	9.64	0	1	4	2,511	83,728	86,244	NA
amlodipine	potassium chloride	2.27	1.04	5.35E-02	9.64	0	1	4	2,511	83,728	86,244	NA
amlodipine	sodium chloride	2.27	1.04	5.35E-02	9.64	0	1	4	2,511	83,728	86,244	NA
amlodipine	sodium bicarbonate	2.27	1.04	5.35E-02	9.64	0	1	4	2,511	83,728	86,244	NA
amlodipine	sildenafil	2.22	0.98	5.46E-02	9.21	0	1	6	2,511	83,726	86,244	NA
amlodipine	droperidol	2.20	0.97	5.53E-02	9.02	0	1	6	2,511	83,726	86,244	NA
amlodipine	salmon calcitonin	2.17	0.76	1.52E-02	8.79	0	2	8	2,510	83,724	86,244	NA
amlodipine	linezolid	2.17	0.63	4.17E-03	8.75	0	3	12	2,509	83,720	86,244	NA
amlodipine	bimatoprost	2.16	1.04	6.24E-02	8.70	0	1	4	2,511	83,728	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	bumetanide	2.13	0.77	1.75E-02	8.41	0	2	7	2,510	83,725	86,244	NA
amlodipine	solifenacin	2.13	0.62	4.51E-03	8.40	0	3	14	2,509	83,718	86,244	NA
amlodipine	acebutolol	2.10	1.00	6.62E-02	8.18	0	1	5	2,511	83,727	86,244	NA
amlodipine	polyethylene glycol 3350	2.10	1.00	6.66E-02	8.17	0	1	5	2,511	83,727	86,244	NA
amlodipine	trospium	2.10	1.04	6.91E-02	8.15	0	1	4	2,511	83,728	86,244	NA
amlodipine	tolterodine	2.08	0.55	1.63E-03	8.03	0	4	15	2,508	83,717	86,244	NA
amlodipine	fludrocortisone	2.08	1.00	6.91E-02	7.97	0	1	5	2,511	83,727	86,244	NA
amlodipine	estrogens, conjugated (usp)	2.07	0.63	5.58E-03	7.95	0	3	12	2,509	83,720	86,244	NA
amlodipine	desvenlafaxine	2.04	1.00	7.34E-02	7.67	0	1	5	2,511	83,727	86,244	NA
amlodipine	dicyclomine	1.99	0.54	2.29E-03	7.30	0	4	17	2,508	83,715	86,244	NA
amlodipine	bisoprolol	1.97	0.62	7.42E-03	7.18	0	3	15	2,509	83,717	86,244	NA
amlodipine	insulin detemir	1.93	0.94	8.14E-02	6.91	0	1	8	2,511	83,724	86,244	NA
amlodipine	travoprost	1.93	0.98	8.46E-02	6.89	0	1	6	2,511	83,726	86,244	NA
amlodipine	nadolol	1.93	0.63	8.96E-03	6.86	0	3	12	2,509	83,720	86,244	NA
amlodipine	pitavastatin	1.92	0.60	8.23E-03	6.84	0	3	18	2,509	83,714	86,244	NA
amlodipine	ketamine	1.91	0.97	8.68E-02	6.74	0	1	6	2,511	83,726	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	methyldopa	1.85	0.54	3.86E-03	6.37	0	4	18	2,508	83,714	86,244	NA
amlodipine	glyburide	1.84	0.71	3.02E-02	6.31	0	2	14	2,510	83,718	86,244	NA
amlodipine	levodopa	1.83	0.95	9.61E-02	6.25	0	1	7	2,511	83,725	86,244	NA
amlodipine	tacrolimus	1.83	0.54	4.19E-03	6.24	0	4	21	2,508	83,711	86,244	NA
amlodipine	ranolazine	1.83	0.49	2.05E-04	6.24	0	5	25	2,507	83,707	86,244	NA
amlodipine	canagliflozin	1.81	0.95	9.99E-02	6.09	0	1	7	2,511	83,725	86,244	NA
amlodipine	prasugrel	1.78	0.72	3.51E-02	5.95	0	2	13	2,510	83,719	86,244	NA
amlodipine	penicillamine	1.78	0.72	3.62E-02	5.91	0	2	12	2,510	83,720	86,244	NA
amlodipine	sevelamer	1.77	0.96	1.07E-01	5.88	0	1	7	2,511	83,725	86,244	NA
amlodipine	cisapride	1.69	0.95	1.19E-01	5.40	0	1	7	2,511	83,725	86,244	NA
amlodipine	insulin glargine	1.68	0.93	1.18E-01	5.39	0	1	9	2,511	83,723	86,244	NA
amlodipine	rizatriptan	1.68	0.94	1.18E-01	5.39	0	1	8	2,511	83,724	86,244	NA
amlodipine	formoterol	1.67	0.71	4.51E-02	5.32	0	2	14	2,510	83,718	86,244	NA
amlodipine	dipyridamole	1.67	0.60	1.84E-02	5.32	0	3	17	2,509	83,715	86,244	NA
amlodipine	metolazone	1.67	0.94	1.22E-01	5.29	0	1	8	2,511	83,724	86,244	NA
amlodipine	sulfadiazine	1.66	0.93	1.22E-01	5.25	0	1	9	2,511	83,723	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	nisoldipine	1.65	0.72	4.82E-02	5.21	0	2	12	2,510	83,720	86,244	NA
amlodipine	nafcillin	1.62	0.93	1.29E-01	5.06	0	1	9	2,511	83,723	86,244	NA
amlodipine	oseltamivir	1.62	0.72	5.15E-02	5.06	0	2	13	2,510	83,719	86,244	NA
amlodipine	tapentadol	1.62	0.93	1.30E-01	5.03	0	1	9	2,511	83,723	86,244	NA
amlodipine	orphenadrine	1.59	0.91	1.33E-01	4.91	0	1	11	2,511	83,721	86,244	NA
amlodipine	itraconazole	1.58	0.71	5.51E-02	4.87	0	2	15	2,510	83,717	86,244	NA
amlodipine	colestipol	1.57	0.94	1.41E-01	4.81	0	1	8	2,511	83,724	86,244	NA
amlodipine	cerivastatin	1.56	0.92	1.41E-01	4.75	0	1	10	2,511	83,722	86,244	NA
amlodipine	chlordiazepoxide	1.56	0.70	5.77E-02	4.74	0	2	16	2,510	83,716	86,244	NA
amlodipine	brinzolamide	1.56	0.95	1.45E-01	4.74	0	1	7	2,511	83,725	86,244	NA
amlodipine	cefpodoxime	1.55	0.94	1.44E-01	4.73	0	1	8	2,511	83,724	86,244	NA
amlodipine	testosterone	1.55	0.90	1.41E-01	4.70	0	1	12	2,511	83,720	86,244	NA
amlodipine	oxybutynin	1.52	0.59	2.78E-02	4.59	0	3	21	2,509	83,711	86,244	NA
amlodipine	formaldehyde	1.52	0.94	1.51E-01	4.57	0	1	8	2,511	83,724	86,244	NA
amlodipine	naloxone	1.51	0.93	1.51E-01	4.55	0	1	9	2,511	83,723	86,244	NA
amlodipine	rivaroxaban	1.49	0.70	6.65E-02	4.45	0	2	17	2,510	83,715	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	dabigatran etexilate	1.43	0.93	1.71E-01	4.19	0	1	9	2,511	83,723	86,244	NA
amlodipine	carbidopa	1.43	0.91	1.69E-01	4.17	0	1	11	2,511	83,721	86,244	NA
amlodipine	donepezil	1.43	0.52	1.79E-02	4.16	0	4	30	2,508	83,702	86,244	NA
amlodipine	budesonide	1.42	0.70	7.77E-02	4.15	0	2	17	2,510	83,715	86,244	NA
amlodipine	oxcarbazepine	1.41	0.58	3.84E-02	4.08	0	3	27	2,509	83,705	86,244	NA
amlodipine	thioridazine	1.40	0.91	1.75E-01	4.07	0	1	11	2,511	83,721	86,244	NA
amlodipine	nicotine	1.39	0.90	1.78E-01	4.00	0	1	13	2,511	83,719	86,244	NA
amlodipine	tiotropium	1.38	0.58	4.12E-02	3.98	0	3	26	2,509	83,706	86,244	NA
amlodipine	terazosin	1.35	0.57	4.50E-02	3.84	0	3	31	2,509	83,701	86,244	NA
amlodipine	nortriptyline	1.34	0.58	4.62E-02	3.83	0	3	27	2,509	83,705	86,244	NA
amlodipine	cetirizine	1.34	0.58	4.67E-02	3.81	0	3	28	2,509	83,704	86,244	NA
amlodipine	adalimumab	1.33	0.91	1.94E-01	3.80	0	1	11	2,511	83,721	86,244	NA
amlodipine	lovastatin	1.28	0.37	6.53E-04	3.59	0	8	74	2,504	83,658	86,244	NA
amlodipine	salmeterol	1.26	0.48	7.94E-03	3.53	0	5	42	2,507	83,690	86,244	NA
amlodipine	raloxifene	1.25	0.69	1.13E-01	3.49	0	2	17	2,510	83,715	86,244	NA
amlodipine	acyclovir	1.25	0.69	1.13E-01	3.48	0	2	20	2,510	83,712	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	neomycin	1.24	0.40	2.02E-03	3.45	0	7	61	2,505	83,671	86,244	NA
amlodipine	fosinopril	1.23	0.68	1.16E-01	3.42	0	2	22	2,510	83,710	86,244	NA
amlodipine	benzonatate	1.22	0.69	1.19E-01	3.40	0	2	20	2,510	83,712	86,244	NA
amlodipine	escitalopram	1.21	0.43	5.14E-03	3.35	0	6	54	2,506	83,678	86,244	NA
amlodipine	colesevelam	1.21	0.57	6.73E-02	3.35	0	3	31	2,509	83,701	86,244	NA
amlodipine	fluticasone	1.19	0.40	2.94E-03	3.29	0	7	64	2,505	83,668	86,244	NA
amlodipine	ibandronate	1.18	0.69	1.29E-01	3.27	0	2	19	2,510	83,713	86,244	NA
amlodipine	fenofibrate	1.17	0.37	1.81E-03	3.21	0	8	74	2,504	83,658	86,244	NA
amlodipine	propafenone	1.15	0.89	2.49E-01	3.17	0	1	13	2,511	83,719	86,244	NA
amlodipine	pantoprazole	1.15	0.57	7.91E-02	3.15	0	3	35	2,509	83,697	86,244	NA
amlodipine	metaxalone	1.15	0.68	1.38E-01	3.14	0	2	24	2,510	83,708	86,244	NA
amlodipine	propranolol	1.14	0.50	4.74E-02	3.13	0	4	43	2,508	83,689	86,244	NA
amlodipine	cyclosporine	1.11	0.68	1.47E-01	3.05	0	2	26	2,510	83,706	86,244	NA
amlodipine	lamotrigine	1.11	0.50	5.30E-02	3.04	0	4	45	2,508	83,687	86,244	NA
amlodipine	captopril	1.11	0.57	8.91E-02	3.02	0	3	33	2,509	83,699	86,244	NA
amlodipine	cefadroxil	1.10	0.89	2.67E-01	3.01	0	1	14	2,511	83,718	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	rifampin	1.08	0.68	1.58E-01	2.96	0	2	22	2,510	83,710	86,244	NA
amlodipine	potassium	1.08	0.68	1.58E-01	2.94	0	2	27	2,510	83,705	86,244	NA
amlodipine	tamsulosin	1.08	0.56	9.41E-02	2.94	0	3	45	2,509	83,687	86,244	NA
amlodipine	buspirone	1.08	0.68	1.60E-01	2.94	0	2	25	2,510	83,707	86,244	NA
amlodipine	alendronate	1.08	0.34	1.34E-03	2.93	0	10	88	2,502	83,644	86,244	NA
amlodipine	risedronate	1.06	0.47	2.49E-02	2.89	0	5	43	2,507	83,689	86,244	NA
amlodipine	sucralfate	1.05	0.90	2.90E-01	2.85	0	1	13	2,511	83,719	86,244	NA
amlodipine	cefprozil	1.04	0.89	2.89E-01	2.84	0	1	14	2,511	83,718	86,244	NA
amlodipine	fluoxetine	1.04	0.50	6.70E-02	2.83	0	4	48	2,508	83,684	86,244	NA
amlodipine	famotidine	1.03	0.88	2.92E-01	2.81	0	1	17	2,511	83,715	86,244	NA
amlodipine	citalopram	1.03	0.43	1.63E-02	2.80	0	6	63	2,506	83,669	86,244	NA
amlodipine	pioglitazone	1.02	0.50	7.07E-02	2.78	0	4	52	2,508	83,680	86,244	NA
amlodipine	sertraline	1.02	0.33	2.19E-03	2.77	0	10	110	2,502	83,622	86,244	NA
amlodipine	betamethasone	1.01	0.88	3.02E-01	2.74	0	1	15	2,511	83,717	86,244	NA
amlodipine	amiodarone	0.99	0.56	1.22E-01	2.68	0	3	43	2,509	83,689	86,244	NA
amlodipine	meloxicam	0.97	0.29	8.45E-04	2.64	0	13	155	2,499	83,577	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	theophylline	0.97	0.88	3.19E-01	2.63	0	1	18	2,511	83,714	86,244	NA
amlodipine	gemfibrozil	0.96	0.47	4.09E-02	2.60	0	5	59	2,507	83,673	86,244	NA
amlodipine	hydroxyzine	0.92	0.50	9.89E-02	2.51	0	4	50	2,508	83,682	86,244	NA
amlodipine	ipratropium	0.90	0.88	3.50E-01	2.46	0	1	16	2,511	83,716	86,244	NA
amlodipine	phenylephrine	0.88	0.88	3.58E-01	2.42	0	1	16	2,511	83,716	86,244	NA
amlodipine	alprazolam	0.88	0.56	1.60E-01	2.42	0	3	45	2,509	83,687	86,244	NA
amlodipine	valdecoxib	0.87	0.56	1.65E-01	2.39	0	3	43	2,509	83,689	86,244	NA
amlodipine	albuterol	0.86	0.50	1.19E-01	2.37	0	4	57	2,508	83,675	86,244	NA
amlodipine	ropinirole	0.85	0.67	2.52E-01	2.34	0	2	29	2,510	83,703	86,244	NA
amlodipine	minocycline	0.85	0.56	1.74E-01	2.34	0	3	48	2,509	83,684	86,244	NA
amlodipine	meclizine	0.85	0.88	3.75E-01	2.34	0	1	16	2,511	83,716	86,244	NA
amlodipine	levetiracetam	0.82	0.87	3.87E-01	2.28	0	1	20	2,511	83,712	86,244	NA
amlodipine	enoxaparin	0.80	0.87	3.98E-01	2.23	0	1	21	2,511	83,711	86,244	NA
amlodipine	infliximab	0.80	0.86	3.99E-01	2.22	0	1	23	2,511	83,709	86,244	NA
amlodipine	fexofenadine	0.79	0.56	2.01E-01	2.21	0	3	51	2,509	83,681	86,244	NA
amlodipine	venlafaxine	0.79	0.47	9.02E-02	2.20	0	5	66	2,507	83,666	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	allopurinol	0.79	0.46	9.03E-02	2.19	0	5	76	2,507	83,656	86,244	NA
amlodipine	cefuroxime	0.78	0.42	6.76E-02	2.17	0	6	84	2,506	83,648	86,244	NA
amlodipine	dorzolamide	0.75	0.87	4.25E-01	2.12	0	1	20	2,511	83,712	86,244	NA
amlodipine	bacitracin	0.74	0.56	2.29E-01	2.10	0	3	47	2,509	83,685	86,244	NA
amlodipine	povidone	0.74	0.87	4.33E-01	2.09	0	1	21	2,511	83,711	86,244	NA
amlodipine	fluconazole	0.73	0.67	3.15E-01	2.08	0	2	33	2,510	83,699	86,244	NA
amlodipine	methotrexate	0.73	0.49	1.79E-01	2.07	0	4	62	2,508	83,670	86,244	NA
amlodipine	leflunomide	0.72	0.87	4.40E-01	2.06	0	1	21	2,511	83,711	86,244	NA
amlodipine	sulfasalazine	0.71	0.87	4.48E-01	2.04	0	1	20	2,511	83,712	86,244	NA
amlodipine	montelukast	0.67	0.66	3.51E-01	1.96	0	2	36	2,510	83,696	86,244	NA
amlodipine	clonazepam	0.67	0.86	4.72E-01	1.95	0	1	24	2,511	83,708	86,244	NA
amlodipine	hydroxychloroquine	0.66	0.49	2.16E-01	1.94	0	4	64	2,508	83,668	86,244	NA
amlodipine	ranitidine	0.66	0.66	3.58E-01	1.93	0	2	39	2,510	83,693	86,244	NA
amlodipine	cyclobenzaprine	0.66	0.39	9.31E-02	1.93	0	7	116	2,505	83,616	86,244	NA
amlodipine	rosiglitazone	0.65	0.87	4.81E-01	1.92	0	1	21	2,511	83,711	86,244	NA
amlodipine	carbamazepine	0.64	0.49	2.32E-01	1.89	0	4	75	2,508	83,657	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	doxycycline	0.63	0.25	1.21E-02	1.88	0	17	269	2,495	83,463	86,244	NA
amlodipine	sulfacetamide	0.63	0.87	4.99E-01	1.87	0	1	21	2,511	83,711	86,244	NA
amlodipine	metformin	0.59	0.27	2.73E-02	1.80	0	15	268	2,497	83,464	86,244	NA
amlodipine	mycophenolate mofetil	0.59	0.66	4.08E-01	1.80	0	2	43	2,510	83,689	86,244	NA
amlodipine	lansoprazole	0.58	0.66	4.10E-01	1.79	0	2	43	2,510	83,689	86,244	NA
amlodipine	clopidogrel	0.57	0.33	7.95E-02	1.77	0	10	169	2,502	83,563	86,244	NA
amlodipine	paroxetine	0.52	0.55	3.79E-01	1.69	0	3	62	2,509	83,670	86,244	NA
amlodipine	baclofen	0.51	0.66	4.65E-01	1.67	0	2	45	2,510	83,687	86,244	NA
amlodipine	amitriptyline	0.51	0.42	2.27E-01	1.66	0	6	110	2,506	83,622	86,244	NA
amlodipine	niacin	0.50	0.27	6.17E-02	1.64	0	15	298	2,497	83,434	86,244	NA
amlodipine	duloxetine	0.48	0.39	2.19E-01	1.61	0	7	127	2,505	83,605	86,244	NA
amlodipine	gabapentin	0.48	0.26	6.41E-02	1.61	0	16	302	2,496	83,430	86,244	NA
amlodipine	ceftriaxone	0.48	0.46	3.00E-01	1.61	0	5	98	2,507	83,634	86,244	NA
amlodipine	chlorpromazine	0.47	0.85	6.04E-01	1.59	0	1	31	2,511	83,701	86,244	NA
amlodipine	exenatide	0.45	0.85	6.14E-01	1.57	0	1	32	2,511	83,700	86,244	NA
amlodipine	azathioprine	0.44	0.66	5.23E-01	1.56	0	2	47	2,510	83,685	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	esomeprazole	0.44	0.55	4.58E-01	1.55	0	3	64	2,509	83,668	86,244	NA
amlodipine	methadone	0.43	0.85	6.28E-01	1.54	0	1	30	2,511	83,702	86,244	NA
amlodipine	sulindac	0.43	0.86	6.33E-01	1.53	0	1	27	2,511	83,705	86,244	NA
amlodipine	triamcinolone	0.42	0.85	6.35E-01	1.53	0	1	30	2,511	83,702	86,244	NA
amlodipine	varenicline	0.42	0.85	6.38E-01	1.52	0	1	33	2,511	83,699	86,244	NA
amlodipine	polymyxin b	0.42	0.86	6.40E-01	1.52	0	1	27	2,511	83,705	86,244	NA
amlodipine	indomethacin	0.40	0.49	4.34E-01	1.49	0	4	96	2,508	83,636	86,244	NA
amlodipine	timolol	0.38	0.85	6.70E-01	1.46	0	1	30	2,511	83,702	86,244	NA
amlodipine	guaifenesin	0.37	0.55	5.22E-01	1.45	0	3	70	2,509	83,662	86,244	NA
amlodipine	fluvastatin	0.36	0.85	6.79E-01	1.44	0	1	33	2,511	83,699	86,244	NA
amlodipine	metoclopramide	0.36	0.30	2.22E-01	1.44	0	12	254	2,500	83,478	86,244	NA
amlodipine	mirtazapine	0.35	0.85	6.88E-01	1.43	0	1	30	2,511	83,702	86,244	NA
amlodipine	azithromycin	0.35	0.31	2.54E-01	1.42	0	11	240	2,501	83,492	86,244	NA
amlodipine	oxaprozin	0.35	0.85	6.94E-01	1.41	0	1	30	2,511	83,702	86,244	NA
amlodipine	cefdinir	0.31	0.65	6.45E-01	1.37	0	2	53	2,510	83,679	86,244	NA
amlodipine	bupropion	0.30	0.42	4.74E-01	1.35	0	6	142	2,506	83,590	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	brimonidine	0.28	0.85	7.46E-01	1.33	0	1	34	2,511	83,698	86,244	NA
amlodipine	gatifloxacin	0.25	0.85	7.71E-01	1.29	0	1	34	2,511	83,698	86,244	NA
amlodipine	etodolac	0.23	0.65	7.29E-01	1.26	0	2	57	2,510	83,675	86,244	NA
amlodipine	gentamicin	0.22	0.85	7.95E-01	1.25	0	1	35	2,511	83,697	86,244	NA
amlodipine	caffeine	0.20	0.65	7.62E-01	1.22	0	2	58	2,510	83,674	86,244	NA
amlodipine	carisoprodol	0.20	0.65	7.64E-01	1.22	0	2	62	2,510	83,670	86,244	NA
amlodipine	tizanidine	0.20	0.85	8.16E-01	1.22	0	1	37	2,511	83,695	86,244	NA
amlodipine	nitroglycerin	0.19	0.65	7.80E-01	1.20	0	2	67	2,510	83,665	86,244	NA
amlodipine	lorazepam	0.18	0.48	7.16E-01	1.20	0	4	119	2,508	83,613	86,244	NA
amlodipine	pregabalin	0.16	0.31	5.98E-01	1.18	0	11	266	2,501	83,466	86,244	NA
amlodipine	quinine	0.13	0.46	7.73E-01	1.14	0	5	131	2,507	83,601	86,244	NA
amlodipine	valproate	0.11	0.84	8.97E-01	1.12	0	1	46	2,511	83,686	86,244	NA
amlodipine	diclofenac	0.09	0.65	8.95E-01	1.09	0	2	67	2,510	83,665	86,244	NA
amlodipine	phenobarbital	0.08	0.65	9.02E-01	1.08	0	2	67	2,510	83,665	86,244	NA
amlodipine	quetiapine	0.07	0.84	9.37E-01	1.07	0	1	45	2,511	83,687	86,244	NA
amlodipine	dexamethasone	0.07	0.84	9.38E-01	1.07	0	1	42	2,511	83,690	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	nabumetone	0.06	0.84	9.48E-01	1.06	0	1	42	2,511	83,690	86,244	NA
amlodipine	warfarin	0.04	0.65	9.57E-01	1.04	0	2	73	2,510	83,659	86,244	NA
amlodipine	clarithromycin	0.03	0.34	9.22E-01	1.03	0	9	249	2,503	83,483	86,244	NA
amlodipine	clindamycin	0.01	0.36	9.81E-01	1.01	0	8	229	2,504	83,503	86,244	NA
amlodipine	phenazopyridine	-0.01	0.84	9.87E-01	0.99	0	1	40	2,511	83,692	86,244	NA
amlodipine	ondansetron	-0.03	0.55	9.61E-01	0.97	0	3	109	2,509	83,623	86,244	NA
amlodipine	loratadine	-0.07	0.84	9.31E-01	0.93	0	1	51	2,511	83,681	86,244	NA
amlodipine	sitagliptin	-0.07	0.84	9.29E-01	0.93	0	1	48	2,511	83,684	86,244	NA
amlodipine	piroxicam	-0.10	0.84	8.98E-01	0.90	0	1	48	2,511	83,684	86,244	NA
amlodipine	moxifloxacin	-0.14	0.48	7.70E-01	0.87	0	4	153	2,508	83,579	86,244	NA
amlodipine	naproxen	-0.18	0.32	5.71E-01	0.83	0	10	375	2,502	83,357	86,244	NA
amlodipine	phenytoin	-0.20	0.64	7.52E-01	0.82	0	2	96	2,510	83,636	86,244	NA
amlodipine	zolpidem	-0.20	0.48	6.61E-01	0.82	0	4	165	2,508	83,567	86,244	NA
amlodipine	methylprednisolone	-0.21	0.84	7.95E-01	0.81	0	1	53	2,511	83,679	86,244	NA
amlodipine	ciprofloxacin	-0.21	0.22	3.23E-01	0.81	0	22	818	2,490	82,914	86,244	NA
amlodipine	butorphanol	-0.22	0.54	6.79E-01	0.81	0	3	131	2,509	83,601	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	topiramate	-0.24	0.84	7.62E-01	0.79	0	1	57	2,511	83,675	86,244	NA
amlodipine	prednisone	-0.25	0.38	5.14E-01	0.78	0	7	266	2,505	83,466	86,244	NA
amlodipine	metronidazole	-0.28	0.45	5.41E-01	0.76	0	5	192	2,507	83,540	86,244	NA
amlodipine	nitrofurantoin	-0.28	0.36	4.38E-01	0.76	0	8	283	2,504	83,449	86,244	NA
amlodipine	midazolam	-0.29	0.83	7.10E-01	0.75	0	1	64	2,511	83,668	86,244	NA
amlodipine	levofloxacin	-0.30	0.23	1.92E-01	0.74	0	20	795	2,492	82,937	86,244	NA
amlodipine	sulfamethoxazole	-0.30	0.23	1.87E-01	0.74	0	20	820	2,492	82,912	86,244	NA
amlodipine	trazodone	-0.30	0.83	7.03E-01	0.74	0	1	61	2,511	83,671	86,244	NA
amlodipine	diphenhydramine	-0.31	0.38	4.12E-01	0.73	0	7	288	2,505	83,444	86,244	NA
amlodipine	fentanyl	-0.32	0.64	6.04E-01	0.73	0	2	106	2,510	83,626	86,244	NA
amlodipine	tramadol	-0.33	0.32	3.05E-01	0.72	0	10	441	2,502	83,291	86,244	NA
amlodipine	trimethoprim	-0.33	0.23	1.52E-01	0.72	0	19	801	2,493	82,931	86,244	NA
amlodipine	cortisone	-0.34	0.64	5.76E-01	0.71	0	2	104	2,510	83,628	86,244	NA
amlodipine	promethazine	-0.36	0.31	2.37E-01	0.70	0	11	473	2,501	83,259	86,244	NA
amlodipine	celecoxib	-0.36	0.38	3.44E-01	0.70	0	7	295	2,505	83,437	86,244	NA
amlodipine	aspirin	-0.39	0.22	6.87E-02	0.67	0	22	1,031	2,490	82,701	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	rofecoxib	-0.40	0.54	4.25E-01	0.67	0	3	157	2,509	83,575	86,244	NA
amlodipine	povidone-iodine	-0.44	0.64	4.62E-01	0.65	0	2	117	2,510	83,615	86,244	NA
amlodipine	lidocaine	-0.51	0.64	3.82E-01	0.60	0	2	120	2,510	83,612	86,244	NA
amlodipine	diazepam	-0.52	0.54	2.96E-01	0.59	0	3	170	2,509	83,562	86,244	NA
amlodipine	cephalexin	-0.54	0.25	3.18E-02	0.58	0	16	817	2,496	82,915	86,244	NA
amlodipine	sumatriptan	-0.55	0.83	4.67E-01	0.58	0	1	81	2,511	83,651	86,244	NA
amlodipine	ibuprofen	-0.55	0.34	1.03E-01	0.58	0	9	533	2,503	83,199	86,244	NA
amlodipine	morphine	-0.57	0.17	5.35E-04	0.56	0	38	2,078	2,474	81,654	86,244	NA
amlodipine	meperidine	-0.63	0.25	1.05E-02	0.53	0	17	948	2,495	82,784	86,244	NA
amlodipine	oxycodone	-0.63	0.25	1.26E-02	0.53	0	16	941	2,496	82,791	86,244	NA
amlodipine	hydromorphone	-0.63	0.41	1.25E-01	0.53	0	6	347	2,506	83,385	86,244	NA
amlodipine	acetaminophen	-0.69	0.20	3.93E-04	0.50	0	27	1,670	2,485	82,062	86,244	NA
amlodipine	clavulanate	-0.71	0.41	8.52E-02	0.49	0	6	368	2,506	83,364	86,244	NA
amlodipine	cefaclor	-0.74	0.48	7.93E-02	0.48	0	4	287	2,508	83,445	86,244	NA
amlodipine	hydrocodone	-0.75	0.23	9.29E-04	0.47	0	20	1,321	2,492	82,411	86,244	NA
amlodipine	heparin	-0.77	0.83	2.89E-01	0.46	0	1	107	2,511	83,625	86,244	NA

object drug	potential precipitant drug	coef	se	pval	or	bonf	nA	nB	nC	nD	total	label
amlodipine	pentazocine	-0.78	0.64	1.61E-01	0.46	0	2	161	2,510	83,571	86,244	NA
amlodipine	amoxicillin	-0.85	0.29	3.60E-03	0.43	0	12	874	2,500	82,858	86,244	NA
amlodipine	propoxyphene	-0.85	0.38	2.57E-02	0.43	0	7	496	2,505	83,236	86,244	NA
amlodipine	epinephrine	-0.88	0.83	2.15E-01	0.42	0	1	101	2,511	83,631	86,244	NA
amlodipine	ketorolac	-0.94	0.64	8.27E-02	0.39	0	2	208	2,510	83,524	86,244	NA
amlodipine	pseudoephedrine	-0.98	0.83	1.59E-01	0.38	0	1	124	2,511	83,608	86,244	NA
amlodipine	erythromycin	-1.02	0.34	2.35E-03	0.36	0	9	745	2,503	82,987	86,244	NA
amlodipine	prochlorperazine	-1.06	0.64	4.52E-02	0.35	0	2	217	2,510	83,515	86,244	NA
amlodipine	iodine	-1.21	0.45	7.16E-03	0.30	0	5	537	2,507	83,195	86,244	NA
amlodipine	tetracycline	-1.25	0.54	3.57E-03	0.29	0	3	377	2,509	83,355	86,244	NA
amlodipine	sulfur	-1.91	0.82	8.25E-04	0.15	0	1	310	2,511	83,422	86,244	NA

Columns “object drug” and “potential precipitant drug” indicate the ADRs for the object and potential precipitant drugs, respectively, used in the regression analysis.
Columns “coef”, “se”, “pval”, and “or” indicate the regression beta-coefficient, standard error, p-value, and odds ratio, respectively.
Column “bonf” indicates whether the p-value passed Bonferroni correction.
Columns “nA”, “nB”, “nC”, and “nD” are the patient counts in the 2x2 contingency table (**Supplementary Figure 3**). Column “total_patients” is the sum of columns “nA”, “nB”, “nC”, and “nD”.
Column “label” indicates the final label for the object and potential precipitant drug pair after automated classification followed by domain expert review (**Supplementary Figure 2**).
DDIWAS: Drug-Drug Interaction Wide Association Study; ADRs: adverse drug reactions.

Supplementary Table 2: Comparison of outputs from logistic regression adjusted for baseline covariates and with additional adjustment for potential precipitant drug indications.

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
simvastatin	alendronate	additional adjustment for potential precipitant drug indications	1.20	0.31	1.10E-04	3.33	12	93	2,802	85,780	88,687
simvastatin	alendronate	adjusted for baseline covariates only	1.26	0.31	4.63E-05	3.53	12	93	2,802	85,780	88,687
simvastatin	amlodipine	additional adjustment for potential precipitant drug indications	0.79	0.21	1.43E-04	2.20	26	285	2,788	85,588	88,687
simvastatin	amlodipine	adjusted for baseline covariates only	0.94	0.21	5.37E-06	2.56	26	285	2,788	85,588	88,687
simvastatin	fenofibrate	additional adjustment for potential precipitant drug indications	2.33	0.28	3.61E-17	10.23	21	50	2,793	85,823	88,687
simvastatin	fenofibrate	adjusted for baseline covariates only	2.53	0.26	5.80E-22	12.57	21	50	2,793	85,823	88,687
simvastatin	gemfibrozil	additional adjustment for potential precipitant drug indications	1.84	0.34	5.21E-08	6.29	12	49	2,802	85,824	88,687
simvastatin	gemfibrozil	adjusted for baseline covariates only	1.98	0.32	1.14E-09	7.21	12	49	2,802	85,824	88,687
simvastatin	hydrochlorothiazide	additional adjustment for potential precipitant drug indications	1.15	0.22	1.64E-07	3.17	24	192	2,790	85,681	88,687
simvastatin	hydrochlorothiazide	adjusted for baseline covariates only	1.30	0.22	2.63E-09	3.67	24	192	2,790	85,681	88,687
simvastatin	irbesartan	additional adjustment for potential precipitant drug indications	2.02	0.53	1.50E-04	7.54	5	13	2,809	85,860	88,687

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
simvastatin	irbesartan	adjusted for baseline covariates only	2.35	0.53	8.88E-06	10.50	5	13	2,809	85,860	88,687
simvastatin	lisinopril	additional adjustment for potential precipitant drug indications	0.51	0.14	2.31E-04	1.67	57	853	2,757	85,020	88,687
simvastatin	lisinopril	adjusted for baseline covariates only	0.66	0.14	2.21E-06	1.93	57	853	2,757	85,020	88,687
simvastatin	losartan	additional adjustment for potential precipitant drug indications	1.18	0.33	2.83E-04	3.26	11	82	2,803	85,791	88,687
simvastatin	losartan	adjusted for baseline covariates only	1.35	0.32	3.00E-05	3.86	11	82	2,803	85,791	88,687
simvastatin	metoprolol	additional adjustment for potential precipitant drug indications	1.42	0.22	8.57E-11	4.14	25	165	2,789	85,708	88,687
simvastatin	metoprolol	adjusted for baseline covariates only	1.52	0.22	2.50E-12	4.56	25	165	2,789	85,708	88,687
simvastatin	niacin	additional adjustment for potential precipitant drug indications	1.22	0.17	1.86E-13	3.40	44	349	2,770	85,524	88,687
simvastatin	niacin	adjusted for baseline covariates only	1.36	0.16	4.44E-17	3.90	44	349	2,770	85,524	88,687
simvastatin	olmesartan	additional adjustment for potential precipitant drug indications	1.64	0.40	3.50E-05	5.16	8	37	2,806	85,836	88,687
simvastatin	olmesartan	adjusted for baseline covariates only	1.77	0.39	7.05E-06	5.86	8	37	2,806	85,836	88,687
simvastatin	risedronate	additional adjustment for potential precipitant drug indications	1.64	0.45	2.95E-04	5.13	6	31	2,808	85,842	88,687
simvastatin	risedronate	adjusted for baseline covariates only	1.71	0.45	1.42E-04	5.54	6	31	2,808	85,842	88,687

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
simvastatin	triamterene	additional adjustment for potential precipitant drug indications	1.59	0.35	6.12E-06	4.90	10	52	2,804	85,821	88,687
simvastatin	triamterene	adjusted for baseline covariates only	1.71	0.35	8.87E-07	5.52	10	52	2,804	85,821	88,687
amlodipine	atenolol	additional adjustment for potential precipitant drug indications	2.05	0.20	5.80E-24	7.75	33	119	2,479	83,613	86,244
amlodipine	atenolol	adjusted for baseline covariates only	2.20	0.20	2.45E-28	9.05	33	119	2,479	83,613	86,244
amlodipine	atorvastatin	additional adjustment for potential precipitant drug indications	1.15	0.14	2.82E-16	3.15	60	560	2,452	83,172	86,244
amlodipine	atorvastatin	adjusted for baseline covariates only	1.28	0.14	1.67E-20	3.60	60	560	2,452	83,172	86,244
amlodipine	carvedilol	additional adjustment for potential precipitant drug indications	2.03	0.25	1.30E-15	7.61	21	79	2,491	83,653	86,244
amlodipine	carvedilol	adjusted for baseline covariates only	2.17	0.25	2.21E-18	8.77	21	79	2,491	83,653	86,244
amlodipine	chlorthalidone	additional adjustment for potential precipitant drug indications	2.20	0.37	3.55E-09	9.00	10	30	2,502	83,702	86,244
amlodipine	chlorthalidone	adjusted for baseline covariates only	2.30	0.37	4.28E-10	10.02	10	30	2,502	83,702	86,244
amlodipine	clonidine	additional adjustment for potential precipitant drug indications	1.80	0.20	2.42E-20	6.07	34	155	2,478	83,577	86,244
amlodipine	clonidine	adjusted for baseline covariates only	1.96	0.19	1.44E-24	7.12	34	155	2,478	83,577	86,244

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
amlodipine	diltiazem	additional adjustment for potential precipitant drug indications	1.81	0.20	1.86E-19	6.14	32	150	2,480	83,582	86,244
amlodipine	diltiazem	adjusted for baseline covariates only	1.87	0.20	2.16E-21	6.50	32	150	2,480	83,582	86,244
amlodipine	ezetimibe	additional adjustment for potential precipitant drug indications	0.89	0.28	1.71E-03	2.44	14	138	2,498	83,594	86,244
amlodipine	ezetimibe	adjusted for baseline covariates only	1.11	0.28	8.48E-05	3.03	14	138	2,498	83,594	86,244
amlodipine	felodipine	additional adjustment for potential precipitant drug indications	3.20	0.48	1.77E-11	24.62	9	10	2,503	83,722	86,244
amlodipine	felodipine	adjusted for baseline covariates only	3.29	0.46	1.38E-12	26.90	9	10	2,503	83,722	86,244
amlodipine	furosemide	additional adjustment for potential precipitant drug indications	1.66	0.30	2.53E-08	5.25	14	83	2,498	83,649	86,244
amlodipine	furosemide	adjusted for baseline covariates only	1.65	0.29	1.47E-08	5.22	14	83	2,498	83,649	86,244
amlodipine	hydralazine	additional adjustment for potential precipitant drug indications	2.13	0.22	8.38E-23	8.44	29	108	2,483	83,624	86,244
amlodipine	hydralazine	adjusted for baseline covariates only	2.18	0.21	9.15E-25	8.88	29	108	2,483	83,624	86,244
amlodipine	hydrochlorothiazide	additional adjustment for potential precipitant drug indications	1.85	0.12	6.92E-55	6.37	94	436	2,418	83,296	86,244
amlodipine	hydrochlorothiazide	adjusted for baseline covariates only	1.95	0.12	1.18E-62	7.01	94	436	2,418	83,296	86,244

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
amlodipine	indapamide	additional adjustment for potential precipitant drug indications	2.84	0.66	1.89E-05	17.06	4	7	2,508	83,725	86,244
amlodipine	indapamide	adjusted for baseline covariates only	2.83	0.63	6.90E-06	16.95	4	7	2,508	83,725	86,244
amlodipine	isosorbide	additional adjustment for potential precipitant drug indications	1.63	0.31	2.32E-07	5.09	12	74	2,500	83,658	86,244
amlodipine	isosorbide	adjusted for baseline covariates only	1.66	0.31	1.33E-07	5.24	12	74	2,500	83,658	86,244
amlodipine	labetalol	additional adjustment for potential precipitant drug indications	2.02	0.48	2.08E-05	7.57	6	23	2,506	83,709	86,244
amlodipine	labetalol	adjusted for baseline covariates only	2.17	0.46	2.97E-06	8.72	6	23	2,506	83,709	86,244
amlodipine	levothyroxine	additional adjustment for potential precipitant drug indications	1.81	0.50	3.07E-04	6.13	5	21	2,507	83,711	86,244
amlodipine	levothyroxine	adjusted for baseline covariates only	1.90	0.50	1.45E-04	6.70	5	21	2,507	83,711	86,244
amlodipine	metoprolol	additional adjustment for potential precipitant drug indications	2.22	0.15	2.70E-48	9.24	62	194	2,450	83,538	86,244
amlodipine	metoprolol	adjusted for baseline covariates only	2.35	0.15	1.91E-56	10.52	62	194	2,450	83,538	86,244
amlodipine	minoxidil	additional adjustment for potential precipitant drug indications	2.00	0.39	2.11E-07	7.40	9	36	2,503	83,696	86,244
amlodipine	minoxidil	adjusted for baseline covariates only	2.21	0.38	4.05E-09	9.14	9	36	2,503	83,696	86,244

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
amlodipine	nebivolol	additional adjustment for potential precipitant drug indications	1.93	0.38	4.21E-07	6.89	9	37	2,503	83,695	86,244
amlodipine	nebivolol	adjusted for baseline covariates only	2.09	0.38	2.68E-08	8.05	9	37	2,503	83,695	86,244
amlodipine	nifedipine	additional adjustment for potential precipitant drug indications	2.20	0.15	2.21E-49	9.00	64	212	2,448	83,520	86,244
amlodipine	nifedipine	adjusted for baseline covariates only	2.28	0.15	1.65E-55	9.80	64	212	2,448	83,520	86,244
amlodipine	omeprazole	additional adjustment for potential precipitant drug indications	1.48	0.34	1.46E-05	4.40	10	66	2,502	83,666	86,244
amlodipine	omeprazole	adjusted for baseline covariates only	1.49	0.34	1.31E-05	4.43	10	66	2,502	83,666	86,244
amlodipine	pravastatin	additional adjustment for potential precipitant drug indications	0.85	0.26	9.38E-04	2.34	17	182	2,495	83,550	86,244
amlodipine	pravastatin	adjusted for baseline covariates only	1.11	0.26	1.41E-05	3.03	17	182	2,495	83,550	86,244
amlodipine	prazosin	additional adjustment for potential precipitant drug indications	3.58	0.72	6.58E-07	35.93	4	5	2,508	83,727	86,244
amlodipine	prazosin	adjusted for baseline covariates only	3.38	0.68	7.69E-07	29.43	4	5	2,508	83,727	86,244
amlodipine	rosuvastatin	additional adjustment for potential precipitant drug indications	0.90	0.21	2.45E-05	2.46	25	247	2,487	83,485	86,244
amlodipine	rosuvastatin	adjusted for baseline covariates only	1.19	0.21	2.14E-08	3.27	25	247	2,487	83,485	86,244

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
amlodipine	simvastatin	additional adjustment for potential precipitant drug indications	0.84	0.15	5.56E-08	2.31	48	525	2,464	83,207	86,244
amlodipine	simvastatin	adjusted for baseline covariates only	1.08	0.15	1.49E-12	2.95	48	525	2,464	83,207	86,244
amlodipine	spironolactone	additional adjustment for potential precipitant drug indications	2.34	0.26	2.80E-19	10.39	21	67	2,491	83,665	86,244
amlodipine	spironolactone	adjusted for baseline covariates only	2.34	0.25	2.89E-20	10.40	21	67	2,491	83,665	86,244
amlodipine	triamterene	additional adjustment for potential precipitant drug indications	1.80	0.27	1.48E-11	6.03	18	88	2,494	83,644	86,244
amlodipine	triamterene	adjusted for baseline covariates only	1.87	0.26	9.59E-13	6.47	18	88	2,494	83,644	86,244
amlodipine	valacyclovir	additional adjustment for potential precipitant drug indications	2.32	0.52	8.05E-06	10.21	5	15	2,507	83,717	86,244
amlodipine	valacyclovir	adjusted for baseline covariates only	2.32	0.52	8.22E-06	10.17	5	15	2,507	83,717	86,244
amlodipine	verapamil	additional adjustment for potential precipitant drug indications	1.86	0.25	7.53E-14	6.43	21	94	2,491	83,638	86,244
amlodipine	verapamil	adjusted for baseline covariates only	1.94	0.24	1.88E-15	6.95	21	94	2,491	83,638	86,244

Table is a companion to **Figure 2** and shows only drugs with Bonferroni p-value < 0.05 and OR > 1.
See **Supplementary Table 1** for description of columns.
The “model” column indicates whether the results are from the regression adjusted for baseline covariates only or with additional adjustment for potential precipitant drug indications.

Supplementary Table 3: PPVs for Simvastatin and Amlodipine DDIWAS

Object Drug	Minimum Patient Count Thresholds	TPs	FPs	PPV
simvastatin	1	11	2	0.85
simvastatin	5	11	2	0.85
simvastatin	10	8	2	0.80
simvastatin	20	5	0	1.00
amlodipine	1	24	4	0.86
amlodipine	5	22	4	0.85
amlodipine	10	18	2	0.90
amlodipine	20	13	0	1.00

Minimum patient count thresholds were the number of patients required in each cell of the 2x2 contingency table.
TPs: True-positives. Number of significantly associated potential precipitant drugs that had known DDIs with the object drug.
FPs: False-positives. Number of significantly associated potential precipitant drugs that did not have previously reported DDIs with the object drug.
PPV: Positive predictive value.

Supplementary Table 4: Comparison of results from baseline regression and regression with additional adjustment for potential precipitant drug-ADRs.

object drug	potential precipitant drug	model	coef	se	pval	or	nA	nB	nC	nD	total patients
amlodipine	atorvastatin	adjusted for baseline covariates only	1.28	0.14	1.67E-20	3.60	60	560	2,452	83,172	86,244
amlodipine	atorvastatin	additional adjustment for ADRs to simvastatin and atorvastatin	1.10	0.15	6.44E-14	2.99	60	560	2,452	83,172	86,244
amlodipine	ezetimibe	adjusted for baseline covariates only	1.11	0.28	8.48E-05	3.03	14	138	2,498	83,594	86,244
amlodipine	ezetimibe	additional adjustment for ADRs to simvastatin and atorvastatin	0.32	0.31	2.89E-01	1.38	14	138	2,498	83,594	86,244
amlodipine	simvastatin	adjusted for baseline covariates only	1.08	0.15	1.49E-12	2.95	48	525	2,464	83,207	86,244
amlodipine	simvastatin	additional adjustment for ADRs to simvastatin and atorvastatin	0.78	0.17	3.08E-06	2.18	48	525	2,464	83,207	86,244
simvastatin	hydrochlorothiazide	adjusted for baseline covariates only	1.30	0.22	2.63E-09	3.67	24	192	2,790	85,681	88,687
simvastatin	hydrochlorothiazide	additional adjustment for ADRs to HCTZ	1.07	0.27	7.63E-05	2.91	24	192	2,790	85,681	88,687
simvastatin	triamterene	adjusted for baseline covariates only	1.71	0.35	8.87E-07	5.52	10	52	2,804	85,821	88,687
simvastatin	triamterene	additional adjustment for ADRs to HCTZ	0.76	0.43	7.56E-02	2.14	10	52	2,804	85,821	88,687

See **Supplementary Table 1** for description of columns.
The “model” column indicates whether the results are from the regression adjusted for baseline covariates only or regression with additional adjustment for potential precipitant drug-ADRs.
HCTZ: hydrochlorothiazide; ADRs: adverse drug reactions.

CHAPTER 4

COMBINING TRANSCRIPTOMIC DATA AND ELECTRONIC HEALTH RECORDS TO REPURPOSE DRUGS

Abstract

Discovering novel uses for existing drugs or drug repurposing reduces time, costs, and the risk of failure associated with new drug development. We present a generalizable high-throughput approach to identify and validate drug repurposing candidates, which integrates human transcriptomes, drug perturbation data, and clinical data. We applied our method to find potential repurposing candidates for two complex phenotypes, hyperlipidemia and hypertension. We screened >21,000 drugs and replicated seven and ten approved drugs for hyperlipidemia and hypertension, respectively. We also identified existing drugs with repurposing potential for both phenotypes. We found three drugs with significant therapeutic effects on clinically-relevant biomarkers in two independent electronic health record databases, including the All of Us data set. Our approach allows researchers to integrate multiple large publicly available biomedical data sets to repurpose drugs across the human phenome.

Introduction

New drug development is expensive, often fails, and takes a long time. Drug repurposing aims to address these problems by finding new indications for approved drugs.[83] Repurposing existing drugs reduces the risk of failure due to well-characterized safety profiles, lowers development costs, and shortens the overall duration by bypassing safety studies. A challenge to drug repurposing has been the lack of systematic approaches to identify promising drug candidates.

To address this problem, researchers developed methods to identify drug repurposing candidates by mining clinical data in electronic health records (EHRs) and publicly available transcriptomic data. Many successfully repurposed drugs, like rituximab[88] were identified by retrospective clinical analysis. To apply retrospective analysis at scale, researchers developed methods to systematically search EHR data for drug repurposing candidates. They have used these data mining approaches to identify existing drugs as new treatments for cancer[13,93] and neurodegenerative diseases.[141] Likewise, other groups have leveraged public transcriptomic

data to repurpose existing drugs. They reason that drugs which reverse a disease phenotype's transcriptomic signature will reverse the mechanisms causing the disease. Using signature matching, researchers have replicated drug indications[142] and discovered new uses for existing drugs.[99]

These studies demonstrated the feasibility of using EHR and transcriptomic data to repurpose drugs, but technological barriers limited the generalizability of these methods. For EHR mining studies, programs written to extract data from one EHR system could not be used without labor and time-intensive changes. These changes were required due to proprietary EHR data structures between institutions. However, standard frameworks, like the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) allow researchers to use programs with only minor changes. Likewise, signature matching approaches were limited by the high costs to generate transcriptomic data for human phenotypes. Thus, for phenotypes without human data, researchers used data generated from pre-clinical animal studies. Recent methods[47] to estimate transcriptomic signatures for human diseases provide a rich resource for signature matching approaches.[100,143]

Here, we describe a high-throughput approach integrating genetically regulated transcriptomic signatures, drug perturbation studies, and routinely collected clinical data to identify and validate drug repurposing candidates (Figure 1). The three major steps of the approach include 1) estimating the transcriptomic signatures associated with a target phenotype using genome-wide association study (GWAS) summary statistics and S-PrediXcan,[47] 2) searching the Integrative Library of Integrated Network-Based Cellular Signatures (iLINCS)[62,63,144] database for drugs that reversed that signal, and 3) Validating promising drug repurposing candidates using real-world clinical data within a local EHR and in the All of Us data set.[9,145] We applied our method to find potential repurposing candidates for two complex phenotypes, hyperlipidemia and hypertension. Transcriptomes, drug perturbation data, and the All of Us data set are all publicly available. We have also shared scripts designed to extract data from the widely used OMOP CDM.

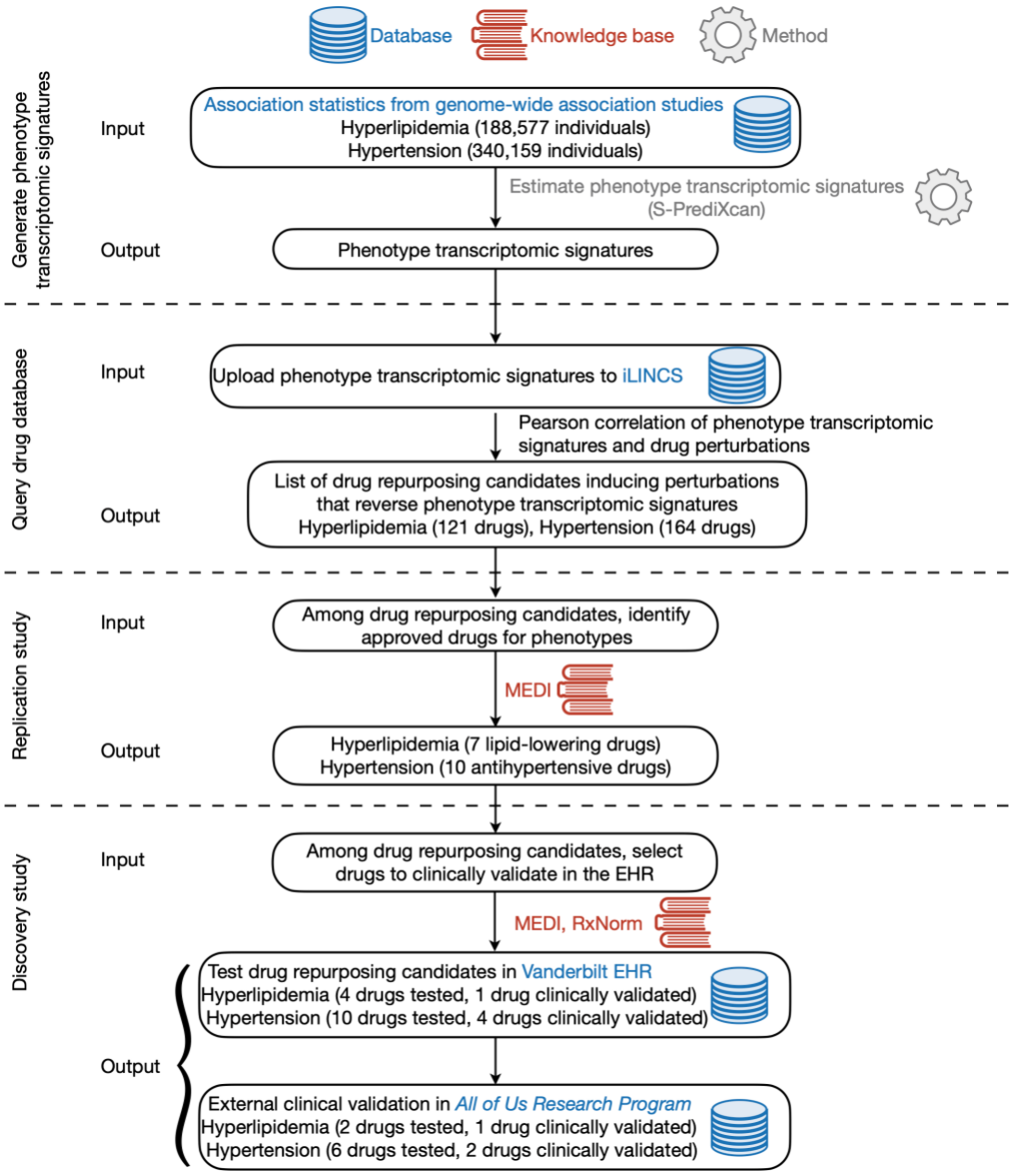


Figure 1. Study design and workflow

A transcriptomic signature for each phenotype was estimated using S-PrediXcan and GWAS summary statistics. For each phenotype's transcriptomic signature, genes were sorted from the most upregulated to the most downregulated. The top upregulated and downregulated genes were then uploaded to the iLINCS drug database. From iLINCs, an initial list of drug repurposing candidates was obtained; drugs in this list induced perturbations that reversed the S-PrediXcan estimated phenotype transcriptomic signatures. In this set of drug repurposing candidates, a replication study was then conducted by looking for approved drugs for the target phenotypes. In the discovery study, potentially novel drug repurposing candidates were clinically validated using real-world data from two independent EHR databases. MEDI: MEDication Indication; GWAS: genome-wide association study; iLINCS: Integrative Library of Integrated Network-Based Cellular Signature; EHRs: electronic health records.

Results

Phenotype transcriptomic signatures and drug database query

For hyperlipidemia and hypertension phenotypes, we searched a public database[146,147] for genetically regulated transcriptomic signatures. The database's authors estimated these transcriptomic signatures using GWAS summary statistics[148,149] and S-PrediXcan.[47,48] We uploaded the S-PrediXcan estimated transcriptomic signatures (Supplementary Tables 1-2 , Supplementary Figure 1) to iLINCS. In iLINCS, we found 121 and 164 potential drug repurposing candidates with perturbations reversing the transcriptomic signatures for hyperlipidemia (Supplementary Table 3) and hypertension (Supplementary Table 4), respectively.

Replication study

We validated our estimated signature matching approach by searching the iLINCS lists for drugs approved for treating hyperlipidemia and hypertension. For hyperlipidemia, we found seven lipid-lowering drugs, including multiple statins and fibrates (Supplementary Table 3). For hypertension, we found ten antihypertensive drugs, including three diuretics (bendroflumethiazide, ethacrynate, spironolactone), one vasodilator (iloprost), one calcium channel blocker (amlodipine), three beta-blockers (carvedilol, esmolol, and nadolol), and two angiotensin-converting enzyme (ACE) inhibitors (benazepril and captopril) (Supplementary Table 4).

Discovery study

We next performed a discovery study, applying our method to find existing drugs not approved for treating the phenotypes of interest with clinical validation in Vanderbilt's EHR database. To identify drug repurposing candidates from the iLINCS lists for clinical validation, we used a two-stage filtering approach (Supplementary Figure 2). In the first stage, we leveraged structured data to automate the filtering process. We excluded drugs if they were not prescribable, were approved for treating the phenotype of interest (i.e., approved lipid-lowering and antihypertensive drugs), and induced perturbations positively correlated with the estimated phenotype transcriptomic signature. In the second stage, we manually reviewed the remaining

candidates to exclude non-systemic drugs, drugs more commonly used for acute conditions, drugs with toxicity issues (e.g., chemotherapeutic agents), and drugs that were not taken by enough patients for clinical validation studies. For hyperlipidemia, this resulted in four drug repurposing candidates to clinically validate in the EHR: alendronate, megestrol, quinapril, and lisinopril. Lisinopril was not in the hyperlipidemia iLINCS list; we added the drug to the clinical validation list, as quinapril and lisinopril are both angiotensin-converting enzyme (ACE) inhibitors, but lisinopril is more commonly prescribed. For hypertension, we validated ten drug repurposing candidates with the largest sample sizes in the Vanderbilt EHR. These drugs were tacrolimus, budesonide, digoxin, pioglitazone, fluoxetine, haloperidol, sertraline, estradiol, escitalopram, and atorvastatin.

To validate these candidates, we used a self-controlled case series (SCCS) study design. As shown Figure 2, consider the clinical validation study of lisinopril as a drug repurposing candidate for hyperlipidemia. In this experiment, we calculated the low-density lipoprotein cholesterol (LDL-C) decrease due to lisinopril exposure. For each patient, we defined an observation period comprising two parts, a baseline period (prior to lisinopril exposure) and treatment period (post lisinopril exposure). The baseline and treatment periods were separated by the index date, which we defined as the first date each patient was exposed to lisinopril. We then calculated the outpatient median LDL-C values for both baseline and treatment periods, respectively. We used a paired one-tailed t-test to determine whether patients experienced statistically significant decreases of LDL-C after being exposed to lisinopril.

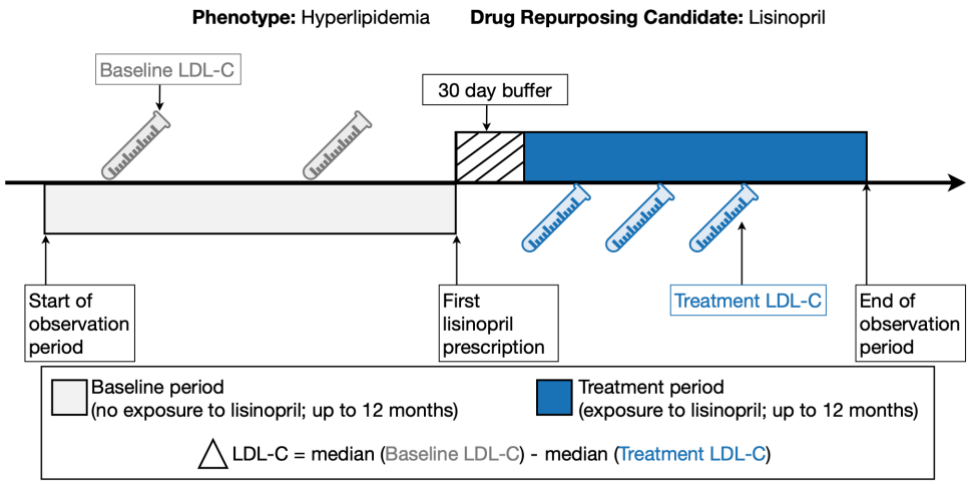


Figure 2. Clinical validation study design

Illustration of self-controlled case series study design used for clinical validation studies in EHRs. The phenotype of interest is hyperlipidemia, and the drug repurposing candidate is lisinopril. The outcome is the change in median LDL-C from baseline after exposure to lisinopril. LDL-C: low-density lipoprotein cholesterol; EHRs: electronic health records.

For the hyperlipidemia clinical validation study, we quantified the effects of five drugs on LDL-C levels. In addition to the four drug repurposing candidates from iLINCS, we added the most common lipid-lowering drug, simvastatin as a positive control. There were 6,305 patients on simvastatin, 620 on alendronate, 36 on megestrol, 170 on quinapril, and 2,447 on lisinopril in Vanderbilt's EHR (Supplementary Figure 3). These patients were not exposed to known lipid-lowering drugs during the baseline and treatment periods. For the simvastatin positive control experiment, patients were not exposed to other known lipid-lowering drugs. The sociodemographic characteristics for these patients are shown in Supplementary Table 5. As expected, we saw significant decreases in LDL-C levels in patients exposed to simvastatin (-31.30 mg/dl, $P < 2.2 \times 10^{-16}$) (Figure 3a, Table 1). Out of the four candidates tested, we found only lisinopril had a significant effect on decreasing LDL-C levels (-1.18 mg/dl, $P = 5.88 \times 10^{-3}$) after Bonferroni correction. Using a raw $P < 0.05$, patients exposed to quinapril also experienced decreases in LDL-C (-3.65 mg/dl, $P = 0.028$).

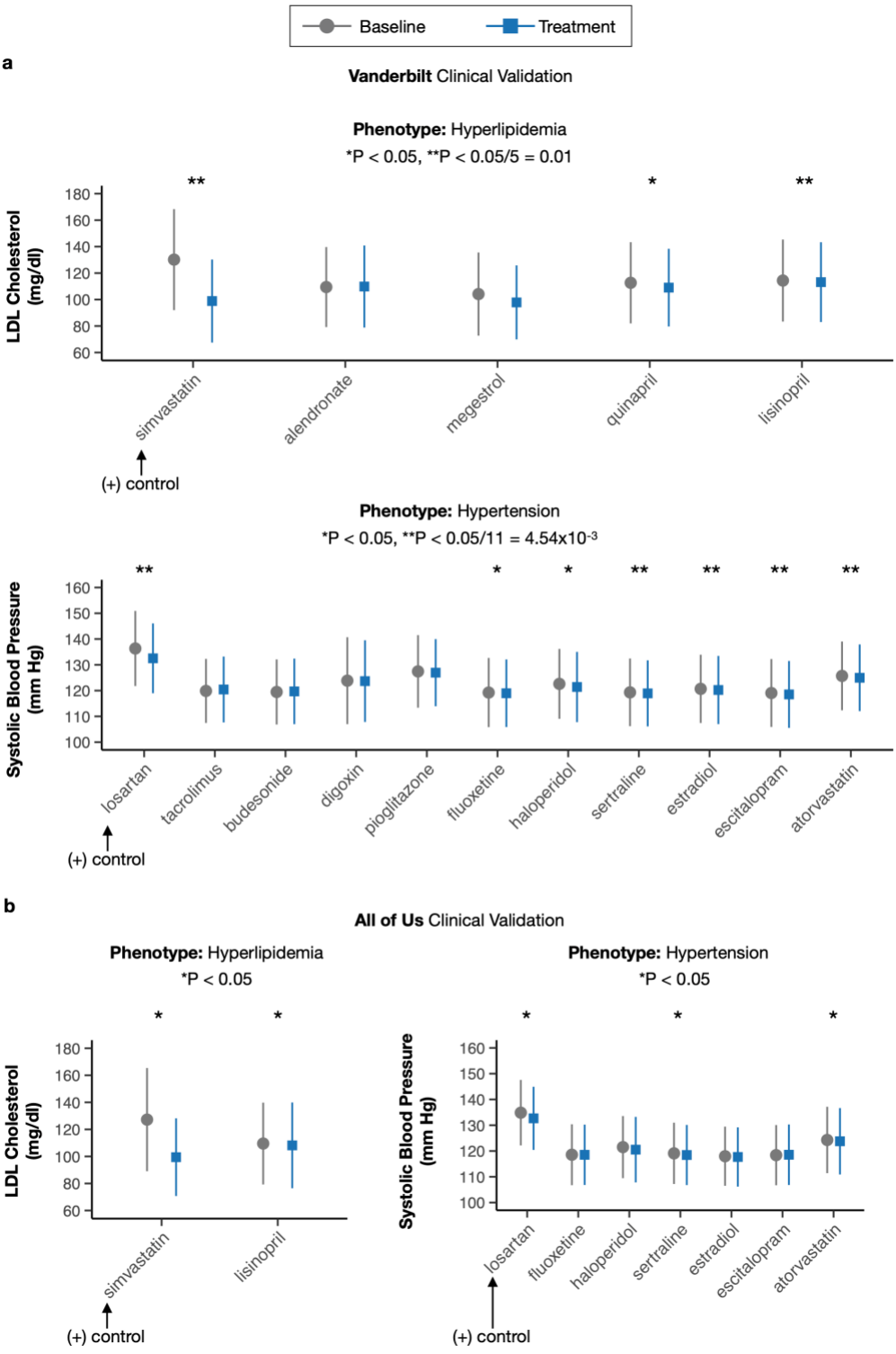

Baseline
Treatment
a
Vanderbilt Clinical Validation
Phenotype: Hyperlipidemia
*P < 0.05, **P < 0.05/5 = 0.01
LDL Cholesterol (mg/dl)
simvastatin
alendronate
megestrol
quinapril
lisinopril
(+) control
Phenotype: Hypertension
*P < 0.05, **P < 0.05/11 = 4.54x10-3
Systolic Blood Pressure (mm Hg)
losartan
tacrolimus
budesonide
digoxin
pioglitazone
fluoxetine
haloperidol
sertraline
estradiol
escitalopram
atorvastatin
(+) control
b
All of Us Clinical Validation
Phenotype: Hyperlipidemia
*P < 0.05
Phenotype: Hypertension
*P < 0.05

Figure 3. EHR validation study results for drug repurposing candidates.

Means of median biomarker values are shown. Circles indicate baseline values, squares indicate treatment values, and error bars represent 95% confidence intervals. P-values were calculated using paired one-tailed t-tests. The null hypothesis was that the drugs had no effect on or increased the biomarker, and the alternative hypothesis was that the drugs decreased the biomarker. The biomarker for hyperlipidemia was LDL-C. The biomarker for hypertension was systolic blood pressure. **a**, Vanderbilt clinical validation study results. **b**, All of Us clinical validation study results. EHR: electronic health record; LDL-C: low-density lipoprotein cholesterol.

Table 1: Vanderbilt clinical validation study results.

Phenotype	Drug	N	Biomarker	Baseline	Treatment	Change	P	Bonferroni
Hyperlipidemia	Simvastatin[#]	6,305	LDL-C (mg/dl)	130.18 (38.16)	98.89 (31.38)	-31.30 (36.45)	< 2.2E-16	Yes
	Alendronate	620		109.45 (30.23)	109.86 (31.02)	0.41 (22.42)	0.68	No
	Megestrol	36		104.15 (31.42)	97.86 (27.94)	-6.29 (30.46)	0.11	No
	Quinapril	170		112.67 (30.68)	109.02 (29.36)	-3.65 (24.83)	0.028	No
	Lisinopril	2,447		114.36 (30.98)	113.18 (30.15)	-1.18 (23.19)	5.88E-03	Yes
Hypertension	Losartan[#]	3,759	SBP (mm Hg)	136.34 (14.57)	132.52 (13.55)	-3.82 (14.88)	< 2.2E-16	Yes
	Tacrolimus	527		119.87 (12.42)	120.43 (12.77)	0.56 (11.55)	0.87	No
	Budesonide	4,204		119.47 (12.63)	119.69 (12.72)	0.22 (11.12)	0.90	No
	Digoxin	220		123.85 (16.85)	123.68 (15.85)	-0.17 (15.07)	0.43	No
	Pioglitazone	581		127.44 (14.08)	126.95 (13.02)	-0.49 (13.09)	0.18	No
	Fluoxetine	7,724		119.24 (13.42)	118.97 (13.12)	-0.27 (12.08)	0.03	No
	Haloperidol	441		122.61 (13.55)	121.39 (13.63)	-1.22 (14.14)	0.036	No
	Sertraline	14,168		119.32 (13.13)	118.92 (12.82)	-0.40 (11.75)	2.73E-05	Yes
	Estradiol	9,794		120.67 (13.24)	120.23 (13.22)	-0.43 (11.87)	1.52E-04	Yes
	Escitalopram	12,535		119.06 (13.17)	118.52 (12.99)	-0.55 (11.79)	1.03E-07	Yes
	Atorvastatin	8,027		125.69 (13.34)	124.97 (12.94)	-0.73 (12.71)	1.58E-07	Yes

Change was calculated by subtracting median measurements during treatment period from that of baseline periods.
For LDL-C, values are mean (SD) LDL-C plasma levels, mg/dl. For SBP, values are mean (SD), mm Hg.
pvalues were calculated using one-tailed paired t-test.
Bonferroni: hyperlipidemia = 0.05/5 = 0.1, hypertension = 0.05/11 = 4.54×10^{-3}.
[#] = Positive Control.
LDL-C: low-density lipoprotein cholesterol; SBP: systolic blood pressure.

For hypertension, we quantified the effects of eleven drugs on the biomarker, systolic blood pressure (SBP). In addition to the ten drugs from the iLINCS drug list, we added losartan as a positive control. There were 3,759 patients on losartan, 527 on tacrolimus, 4,204 on budesonide, 220 on digoxin, 581 on pioglitazone, 7,724 on fluoxetine, 441 on haloperidol, 14,168 on sertraline, 9,787 on estradiol, 12,535 on escitalopram, and 8,027 on atorvastatin from Vanderbilt's EHR (Supplementary Figure 3). These patients were not exposed to known antihypertensive drugs during baseline and treatment periods. For the losartan positive control experiment, patients were not exposed to other known antihypertensive drugs. The sociodemographic characteristics for the cohorts are shown in Supplementary Table 5. For these patients, we extracted SBP measurements both at baseline and after treatment. As expected in the positive control experiment, we saw a significant decrease in SBP after patients were exposed to losartan (-3,82 mm Hg, $P < 2.2x10^{-16}$) (Figure 3a, Table 1). Out of the ten drug repurposing candidates tested, four drugs had significant effects on decreasing SBP with Bonferroni correction: sertraline (-0.40 mm Hg, $P = 2.73x10^{-5}$), estradiol (-0.43 mm Hg, $P = 1.52x10^{-4}$), escitalopram (-0.55 mm Hg, $P = 1.03x10^{-7}$), and atorvastatin (-0.73 mm Hg, $P = 1.58x10^{-7}$) (Figure 3a, Table 1). Using a raw $P < 0.05$, patients experienced decreases in SBP after being exposed to fluoxetine (-0.27 mm Hg, P = 0.03) and haloperidol (-1.22 mm Hg, P = 0.036).

To confirm our observations in the clinical validation studies, we performed external replication studies using data from All of Us. We performed replication studies for drugs with $P < 0.05$ in the Vanderbilt EHR clinical validation experiments. For hyperlipidemia, there were 1526 patients on simvastatin (positive control) and 899 patients on lisinopril (Supplementary Figure 4, Supplementary Table 5). There were only 12 patients in the quinapril cohort, so we did not include the results in this analysis. We replicated the LDL-C lowering effects for both simvastatin (-27.78 mg/dl, $P < 2.2x10^{-16}$) and lisinopril (-1.34 mg/dl, P = 0.041) (Figure 3b, Table 2). For hypertension, there were 328 patients on losartan (positive control), 839 on fluoxetine, 203 on haloperidol, 1094 on sertraline, 838 on estradiol, 709 on escitalopram, and 1842 on atorvastatin (Supplementary Figure 4, Supplementary Table 5). We replicated the SBP lowering effects for losartan (-2.20 mm Hg, $P = 2.10x10^{-3}$), sertraline (-0.65 mm Hg, P = 0.019), and atorvastatin (-0.51 mm Hg, P = 0.031) (Figure 3b, Table 2).

Table 2: All of Us clinical validation study results.

Phenotype	Drug	N	Biomarker	Baseline	Treatment	Change	P	Validated
Hyperlipidemia	Simvastatin[#]	1,526	LDL-C (mg/dl)	127.23 (38.16)	99.45 (28.74)	-27.78 (34.32)	< 2.2E-16	Yes
	Lisinopril	899		109.54 (30.27)	108.19 (31.72)	-1.34 (23.06)	0.041	Yes
Hypertension	Losartan[#]	328	SBP (mm Hg)	134.86 (12.71)	132.67 (12.26)	-2.20 (13.79)	2.10E-03	Yes
	Fluoxetine	839		118.55 (11.82)	118.52 (11.71)	-0.03 (10.28)	0.47	No
	Haloperidol	203		121.51 (12.05)	120.52 (12.72)	-0.99 (11.04)	0.10	No
	Sertraline	1,094		119.09 (11.89)	118.44 (11.67)	-0.65 (10.28)	0.019	Yes
	Estradiol	838		117.98 (11.48)	117.69 (11.49)	-0.29 (10.17)	0.20	No
	Escitalopram	709		118.39 (11.68)	118.54 (11.75)	0.15 (10.73)	0.65	No
	Atorvastatin	1,842		124.28 (12.87)	123.77 (12.89)	-0.51 (11.65)	0.031	Yes

Change was calculated by subtracting median measurements during treatment period from that of baseline periods.
For LDL-C, values are mean (SD) LDL-C plasma levels, mg/dl. For SBP, values are mean (SD), mm Hg.
pvalues were calculated using one-tailed paired t-test.
Bonferroni: hyperlipidemia = 0.05/5 = 0.1, hypertension = 0.05/11 = 4.54×10^{-3}.
[#] = Positive Control.
LDL-C: low-density lipoprotein cholesterol; SBP: systolic blood pressure.

Discussion

In this study, we described a high-throughput approach integrating estimated phenotype transcriptomic signatures and real-world data in EHRs to systematically discover and clinically validate potential drug repurposing candidates. We found that patients exposed to lisinopril experienced significant decreases in LDL-C levels. We also found that patients exposed to sertraline, escitalopram, estradiol, and atorvastatin experienced lower SBP. We also validated the LDL-C lowering effects for lisinopril and SBP lowering effects for sertraline and atorvastatin in the *All of Us* data set, with similar effect sizes. These effects have been observed in the published literature.[150,151] These results suggest that the increasing amount of publicly available omics and EHR data can be leveraged for drug repurposing studies.

The main contribution of this study is an end-to-end pipeline using large publicly available human phenotype transcriptomic data and clinical data in EHRs. Only recently has it been able to combine these two powerful systematic approaches of using transcriptomic and clinical data, as methods and frameworks have been developed to overcome technological barriers. Methods like S-PrediXcan have made it possible to obtain genetically regulated transcriptomic signatures for 44 tissues for a large amount of human diseases with GWAS summary statistics available. Due to the still relatively high cost of RNA-seq experiments and difficulty with obtaining tissue biopsies of living patients (e.g., brain), estimated transcriptomic data provides a rich resource for signature matching drug repurposing. Further, these estimated transcriptomic data addresses the hurdle of obtaining transcriptomic data from traditional sources due to patient privacy concerns.[152] For clinical data, widely-adopted standards to structure data, like the OMOP CDM, increase the generalizability of EHR data mining studies and promotes reproducibility. Further, initiatives like the All of Us Research Program, allows researchers working at institutions without access to EHR data to conduct drug repurposing studies for their phenotype of interest.

There are two key limitations of this study. First, the iLINCS database has drug signatures from a variety of cell types, but much of the data comes from cancer cell lines and primary non-human cells, which could respond biologically differently than primary human cells. Thus, our initial query of iLINCS could be biased due to these in vitro experiments in non-primary human cells. As future similar drug repositories are generated, potentially from induced pluripotent stem cells and organoids,[153] increasingly more accurate drug perturbed transcriptomic data can be

generated. Second, we tackled “low-hanging” phenotypes in the EHR, phenotypes with common quantifiable biomarkers as the primary outcomes. But, many important diseases are not as easily measurable as these traits, so deep phenotyping methods, especially in the space of tracking progress over time[154] could further improve this method’s generalizability.

In conclusion, we demonstrated as a proof-of-concept a high-throughput generalizable approach to identify drug repurposing candidates using genome-wide transcriptomic signatures. Novel drug repurposing candidates were clinically validated in real-world patients in two independent EHR databases.

Methods

Estimation of phenotype transcriptomic signatures

We used publicly available transcriptomic signatures[155–157] estimated by S-PrediXcan;[47,48] this method computes genome-wide transcriptomic signatures for target phenotypes using GWAS summary statistics (Figure 1). S-PrediXcan was trained using Genotype-Tissue Expression (GTEx),[49] a data set with genotypes linked to RNA-seq data for 44 human tissues. The transcriptomic signature for hyperlipidemia was estimated using the whole blood elastic net model (tissue = “TW_Whole_Blood_Elastic_Net_0.5”)[158] and GWAS summary statistics from the Global Lipids Genetics Consortium with 188,577 individuals (phenotype = “GLGC_Mc_LDL”).[148,159] The file was downloaded from “https://s3.amazonaws.com/imlab-open/Data/MetaXcan/results/metaxcan_results_database_v0.1.tar.gz”. The transcriptomic signature for hypertension (phenotype = “Systolic blood pressure, automated reading”) was estimated using an aggregate tissue model and GWAS summary statistics from a UK Biobank study with 340,159 individuals.[8,149] The file “smultixcan_4080_raw_ccn30.tsv.gz”, was downloaded from https://uchicago.box.com/shared/static/vket4ickq7qt3sj8dy3mv8zsr1our3xd.gz.

Querying iLINCS database for drug repurposing candidates

We obtained the initial list of drug repurposing candidates from iLINCS.[144] This database hosts drug perturbations (i.e., transcriptomic data) from *in vitro* experiments using a variety of cells including human cancer cell lines[63] and primary rat hepatocytes.[64,160] Overall, iLINCS contains expression measurements for 74,201 genes from 21,299 small molecules.[144]

To query iLINCS for drug repurposing candidates, we searched for drugs that reversed the S-PrediXcan estimated phenotype transcriptomic signatures. For hyperlipidemia, we sorted the genes from most upregulated to most downregulated, using z-scores. We then selected the top fifty most upregulated and downregulated genes, for a total of 100 genes. For hypertension, transcriptomic signatures from "smultixcan_4080_raw_ccn30.tsv.gz" for genes in "suppl_table_S1-significant_gene_trait_associations.xlsx",[156] (trait = "4080_raw-Systolic_blood_pressure_automated_reading") were used. These genes were identified as likely causal genes for SBP. For both phenotypes, the transcriptomic signatures were uploaded to iLINCS. To measure the similarity between S-PrediXcan estimated phenotype transcriptomic signatures and drug perturbed transcriptomic signatures in iLINCS, either a weighted Pearson correlation[144] or moderated z-scores were used.[63] Promising drug repurposing candidates were those with perturbations that reversed the phenotype transcriptomic signature estimated by S-PrediXcan (i.e., negative correlation coefficient/concordance) with a $P < 0.05$ for hyperlipidemia and $P < 0.001$ for hypertension.

For hyperlipidemia, we obtained drug repurposing candidates from DrugMatrix signatures. The DrugMatrix data set contains differential gene expression and p-values for ~13,000 genes.[64,161] We used this set of drugs for hyperlipidemia because it contained data from primary liver tissue, which is one of the major tissues regulating LDL-C levels. For hypertension, we obtained drug repurposing candidates from LINCS chemical perturbagen experiments. The LINCS data set contains transcriptomic signatures from the L1000 project,[63] mainly from cell lines. We used the LINCS data set for hypertension because the drug list contained more known antihypertensive drugs than other data sets in iLINCS.

Replication of approved drugs for target phenotypes

To demonstrate the feasibility of using S-PrediXcan estimated phenotype transcriptomic signatures to identify drug repurposing candidates, we searched for known approved drugs for the target phenotypes (e.g., statin drugs for LDL-C) in the iLINCS drug lists. We mapped the drugs from iLINCS to active pharmaceutical ingredients using RxNorm. RxNorm is a standardized terminology that links drugs to concepts, unique terms that represent therapeutically equivalent medications.[14,162] To identify drugs approved for treating the target phenotypes, we used the MEDication Indication high-precision subset (MEDI-HPS) knowledge base.[23] MEDI-HPS links drug ingredients to phenotypes represented as International

Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes. To identify drug ingredients approved for treating hyperlipidemia, we used ICD-9-CM codes 272.0 “Pure hypercholesterolemia”, 272.2 “Mixed hyperlipidemia”, and 272.4 “Other and unspecified hyperlipidemia”. To identify drug ingredients approved treating hypertension, we used ICD-9-CM code 401.9 “Hypertension NOS”. We then manually reviewed the drug lists and added drugs that were approved after MEDI-HPS was released (e.g., PCSK9 antibodies for hyperlipidemia).[163]

Discovery study: selection of drug repurposing candidates for clinical validation in the EHR

We used a two-stage filtering approach to select drugs from iLINCS to validate in the EHR. In the first automated stage, we used structured data contained in the iLINCS drug lists. From the iLINCS drug lists, we mapped drugs to their ingredients in RxNorm, excluded drugs if they were not prescribable using RxNorm (flag CVF = 4096), were approved for treating the phenotype of interest using MEDI-HPS, or were incubated for more than one day in iLINCS experiments. We then removed drugs that induced transcriptomic signatures iLINCS, that positively correlated with the S-PrediXcan estimated phenotype transcriptomic signatures. There were drugs that in certain conditions (variable incubation time and drug concentration) induced perturbations that reversed the phenotype transcriptomic signatures (i.e., were negatively correlated) and in other conditions did not (i.e., were positively correlated). We reasoned that drug perturbations that positively correlated with S-PrediXcan phenotype transcriptomic signatures, would simulate the phenotype; for instance, a drug that increased a patient’s blood pressure may be predicted to induce perturbations positively correlated with the transcriptomic signature for hypertension. For the last step in the first automated stage, we then corrected for multiple testing by excluding drugs with gene-set *P* that did not pass Bonferroni correction.

In the second stage of the filtering process, we manually inspected the remaining candidates to decide whether we should validate those drugs in the EHR. We excluded drugs that are commonly used in a non-systemic form (e.g., topical drugs), those taken for short-term (e.g., antibiotics), drugs recommended to be taken “as needed” (e.g., benzodiazepines), toxicity (e.g., NSAIDS, chemotherapeutic agents). We used these exclusion criteria, because we wanted to see the effects of chronic exposure of drugs on the target phenotypes. Last, we excluded drugs with low patient numbers in the Vanderbilt EHR to maximize statistical power.

Discovery study: clinical validation of drug repurposing candidates using real-world clinical data in the EHR

Drug repurposing candidates were tested for their effects on the target phenotype using a de-identified copy of Vanderbilt's EHR. The Vanderbilt EHR has longitudinal data for 3.2 million patients including billing codes, lab values, and medication exposure information.

To validate drug repurposing candidates in the EHR, we used a SCCS study design (Figure 2).[164] Using SCCS allowed us to reduce the bias due to confounders. Minimizing bias allowed us to be more confident in the effects of the drug repurposing candidates on the target phenotype in our validation EHR studies. We designed the SCCS study by creating an observation window with two periods: baseline and treatment period. The index date was the first date of exposure to the drug repurposing candidates of interest. The baseline period was the time before the index date, with a maximum length of one year before. After the index date, we required a minimum period of thirty days in the treatment period, the time after the start of the index date with a maximum length of one year.

For hyperlipidemia, we used change in outpatient LDL-C levels as the biomarker of interest to quantify each drug's therapeutic effects. For hypertension, we used change in outpatient SBP measurements. We selected these biomarkers because they are easily quantifiable clinical variables in the EHR and have been shown to be important for predicting risk of cardiovascular disease in The Framingham Study.[165] We chose to measure changes in outpatient biomarkers, as the values of the biomarkers during inpatient stays can be dramatically affected by acute disease processes related to inpatient admissions.

For the hyperlipidemia clinical validation studies, we tested the potential LDL-C lowering effects of five drugs: simvastatin (positive control), megestrol, alendronate, quinapril, and lisinopril. Although lisinopril did not appear among the top hits in the transcriptomic signature comparison analysis, we tested the drug because it belongs to the same drug class (angiotensin-converting enzyme [ACE] inhibitor) as quinapril, but is much more frequently prescribed.

For each drug repurposing candidate, we identified a cohort of adults (≥ 18 years and < 90 years) with outpatient exposure to the drug being tested. We estimated the treatment effect of the drugs on LDL-C in a cohort of patients without exposure to lipid-lowering drugs. This study design allowed us to measure the lipid-lowering treatment effect of the drug candidates.

The treatment effect was defined as the between-group difference. To measure the treatment effect of the drugs on LDL-C, we calculated the difference between the median LDL-C levels during baseline period and the levels during the treatment period. Outliers were removed (defined as 1.5 x interquartile range, outside first and third quartiles). We reported the mean and standard deviation (SD) for median LDL-C plasma levels for both periods and for LDL-C change. We used a one-tailed paired t-test to measure each drug's effect on LDL-C plasma levels with a type-1 error rate set to the Bonferroni-corrected value of 0.05. The Bonferroni correction was set to 0.05/5 = 0.01. The null hypothesis was that the drug had no effect or increased LDL-C plasma levels, and the alternative hypothesis was that the drug decreased LDL-C levels. We reported the one-tailed p-value from this analysis.

For the hypertension clinical validation studies, we validated the potential SBP lowering effects of eleven drugs: losartan (positive control), tacrolimus, budesonide, digoxin, pioglitazone, fluoxetine, haloperidol, sertraline, estradiol, escitalopram, and atorvastatin. Treatment effect was measured in a cohort of patients without exposure to known antihypertensive drugs during the observation period. The Bonferroni correction was set to $0.05/11 = 4.54 \times 10^{-3}$.

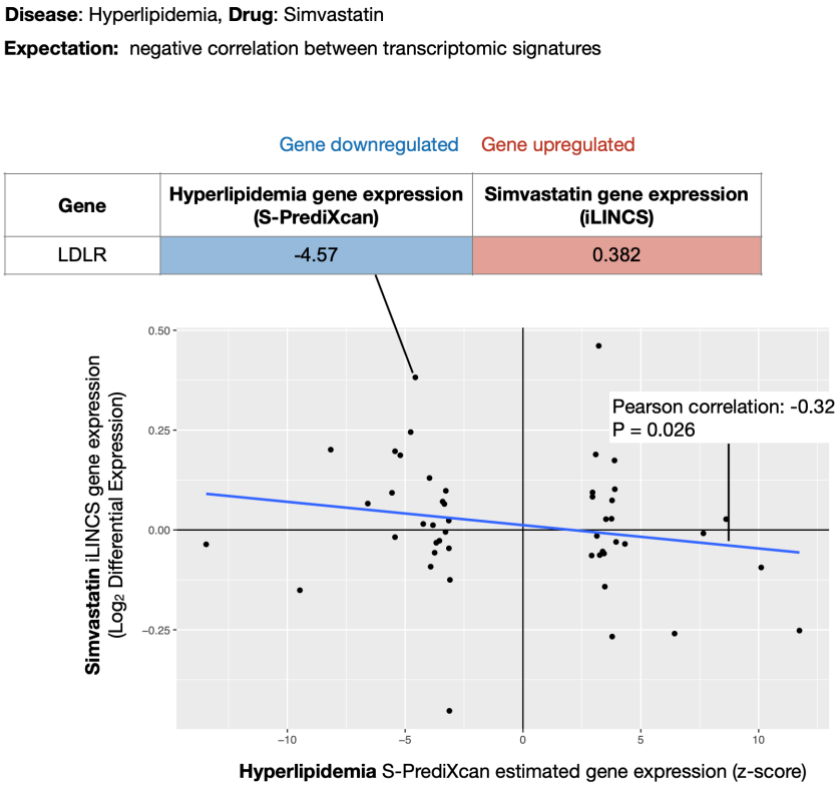
Discovery study: external clinical validation of drug repurposing candidates in the All of Us data set

We performed external clinical validation studies using the All of Us data set. The All of Us data set is a unique resource with health data from a diverse group of participants, with >50% of participants as members of racial and ethnic minorities, and >80% from underrepresented groups in biomedical research. The data set currently contains >370,000 diverse participants and EHRs for >236,000 participants. Analyses were performed in All of Us data set, v4 during the beta testing phase of the program, which began in May 2020.[166] We tested drugs with $P < 0.05$ in the Vanderbilt clinical validation analyses and used the same positive controls.

Code availability

We have made publicly available code to query EHR databases and for data processing and analysis (https://github.com/pwatrick/DrugRepurposingToolKit).

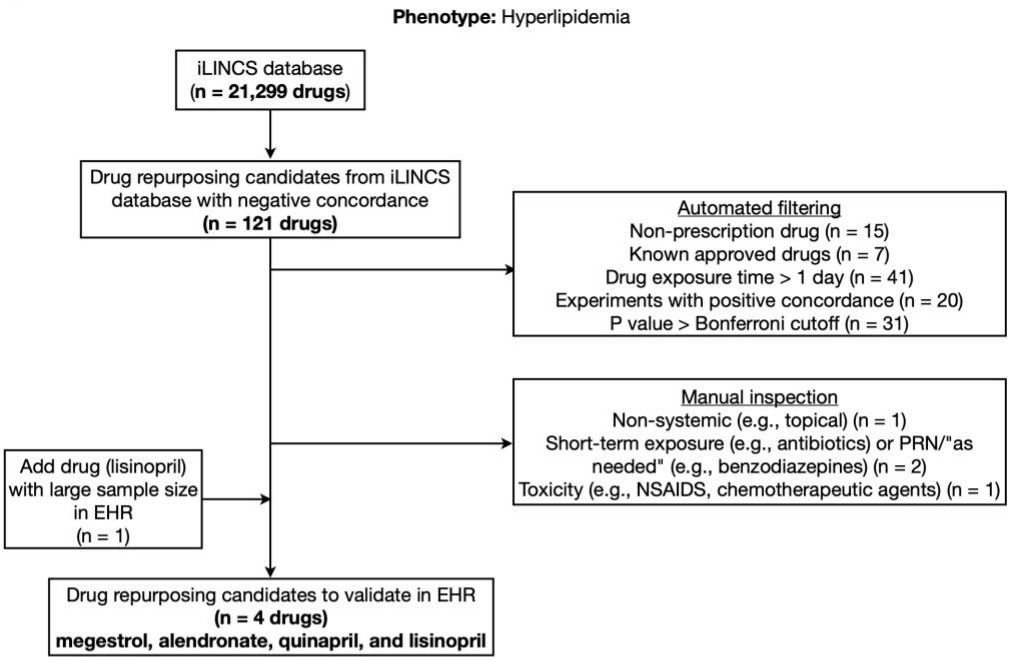
Supplementary Material



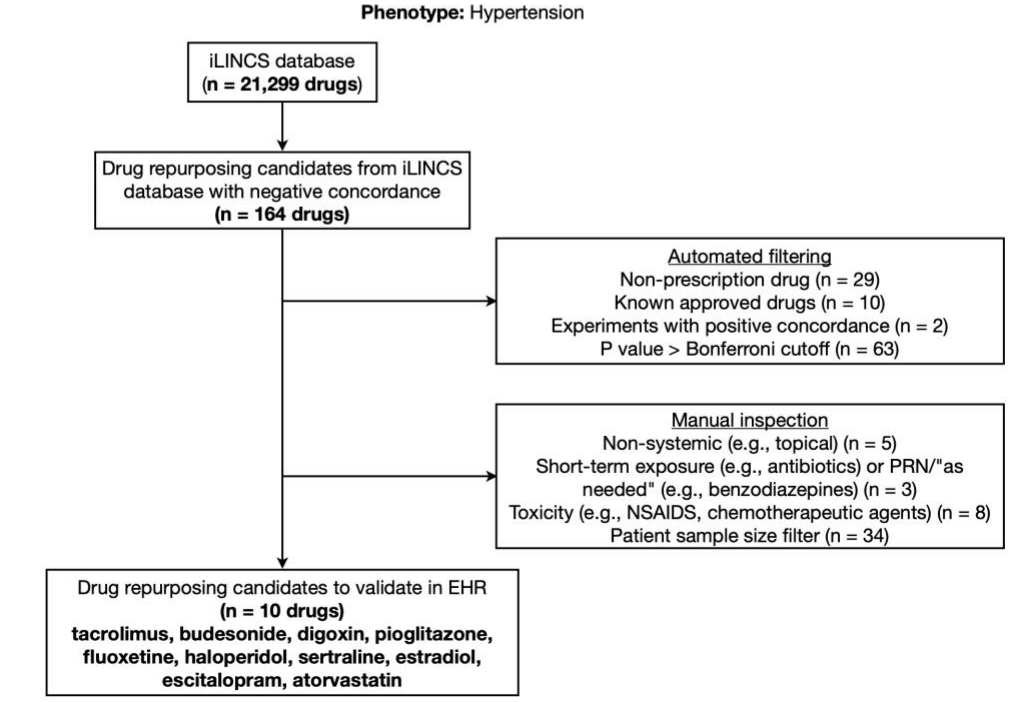
Supplementary Figure 1. Example of drug perturbation reversal of phenotype transcriptomic signature estimated by S-PrediXcan.

Each point represents one gene. Since simvastatin is a known lipid-lowering drug, simvastatin induced transcriptomic signature was predicted to reverse the S-PrediXcan estimated transcriptomic signature for hyperlipidemia. The blue line indicates the expected negative correlation between S-PrediXcan estimated hyperlipidemia transcriptomic signature (horizontal axis) and iLINCS simvastatin induced transcriptomic signature (vertical axis). As expected, the *LDLR* gene was downregulated in patients with hyperlipidemia and upregulated in simvastatin perturbation experiments. iLINCS: Integrative Library of Integrated Network-Based Cellular Signature.

a

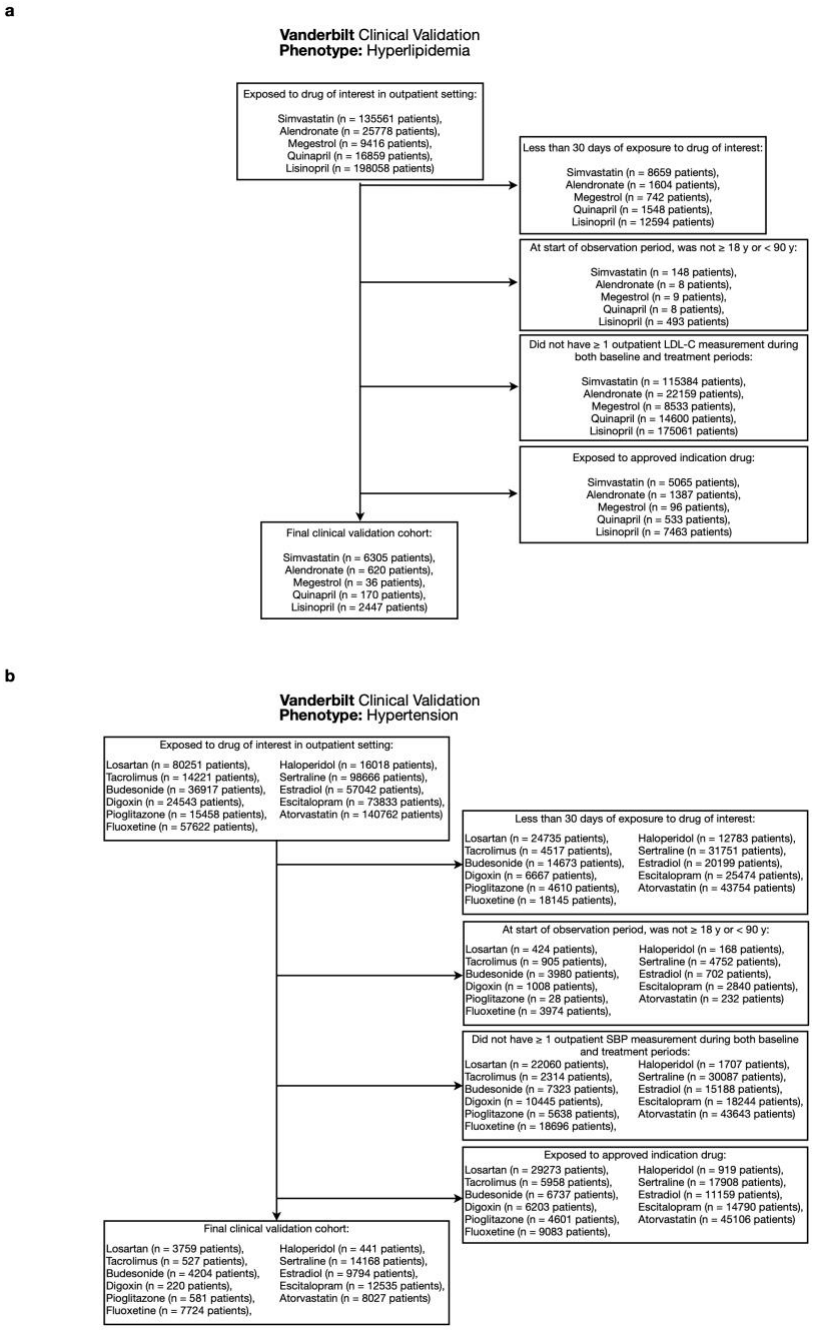

Phenotype: Hyperlipidemia
iLINCS database
(n = 21,299 drugs)
Drug repurposing candidates from iLINCS database with negative concordance
(n = 121 drugs)
Automated filtering
Non-prescription drug (n = 15)
Known approved drugs (n = 7)
Drug exposure time > 1 day (n = 41)
Experiments with positive concordance (n = 20)
P value > Bonferroni cutoff (n = 31)
Manual inspection
Non-systemic (e.g., topical) (n = 1)
Short-term exposure (e.g., antibiotics) or PRN/"as needed" (e.g., benzodiazepines) (n = 2)
Toxicity (e.g., NSAIDS, chemotherapeutic agents) (n = 1)
Add drug (lisinopril) with large sample size in EHR
(n = 1)
Drug repurposing candidates to validate in EHR
(n = 4 drugs)
megestrol, alendronate, quinapril, and lisinopril

b


Phenotype: Hypertension
iLINCS database
(n = 21,299 drugs)
Drug repurposing candidates from iLINCS database with negative concordance
(n = 164 drugs)
Automated filtering
Non-prescription drug (n = 29)
Known approved drugs (n = 10)
Experiments with positive concordance (n = 2)
P value > Bonferroni cutoff (n = 63)
Manual inspection
Non-systemic (e.g., topical) (n = 5)
Short-term exposure (e.g., antibiotics) or PRN/"as needed" (e.g., benzodiazepines) (n = 3)
Toxicity (e.g., NSAIDS, chemotherapeutic agents) (n = 8)
Patient sample size filter (n = 34)
Drug repurposing candidates to validate in EHR
(n = 10 drugs)
tacrolimus, budesonide, digoxin, pioglitazone, fluoxetine, haloperidol, sertraline, estradiol, escitalopram, atorvastatin

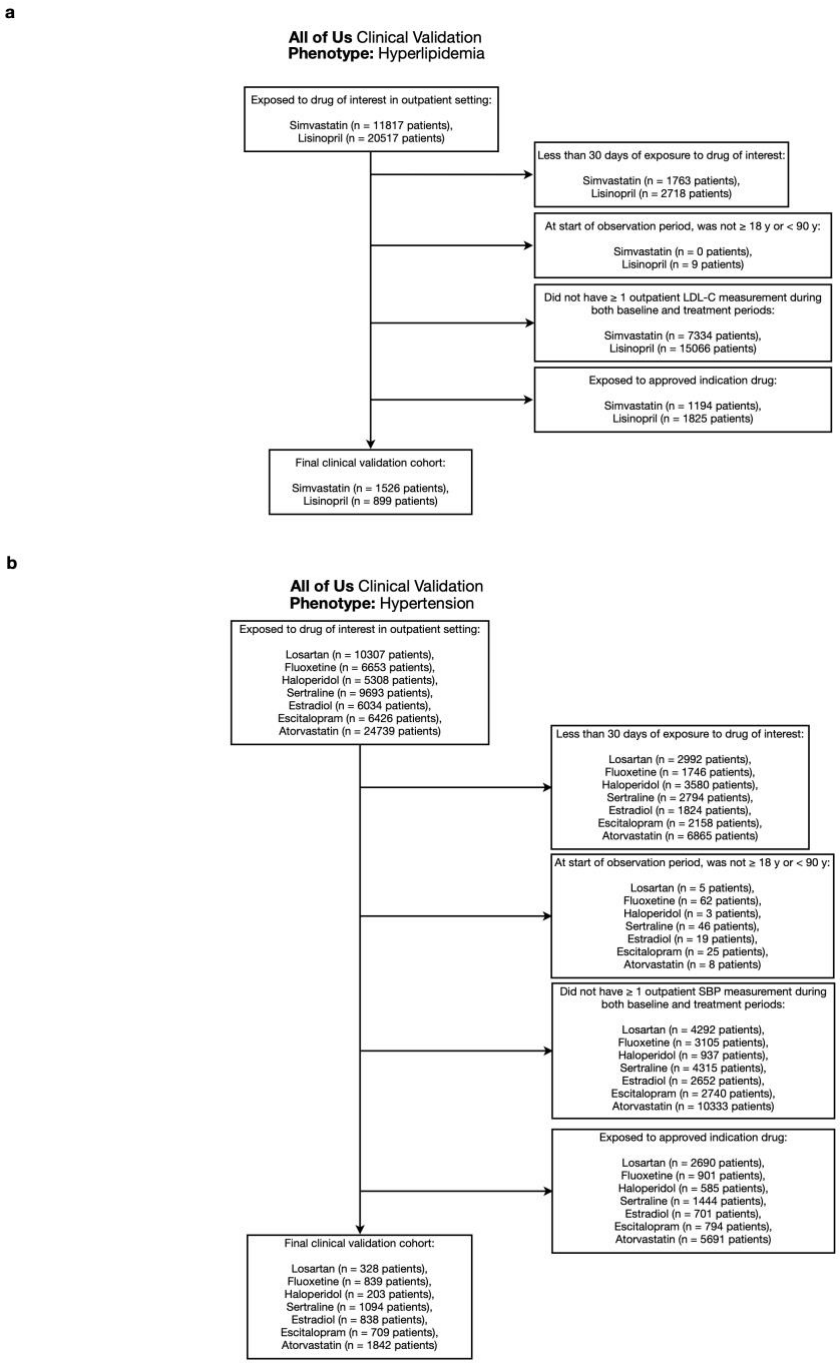
Supplementary Figure 2. Selection of drug repurposing candidates to clinically validate in EHR.

a, Selection of drug repurposing candidates from the iLINCS database to validate in the EHR, for hyperlipidemia. In addition to the three drugs from iLINCS (megestrol, alendronate, quinapril), lisinopril was also tested. Like quinapril, lisinopril is an ACE inhibitor, but is much more commonly prescribed. **b**, Selection of drug repurposing candidates from iLINCS database to validate in the EHR, for hypertension. EHR: electronic health record; iLINCS: Integrative Library of Integrated Network-Based Cellular Signatures; NSAIDs: Nonsteroidal anti-inflammatory drugs; ACE: angiotensin-converting enzyme.



Supplementary Figure 3. Cohort selection for Vanderbilt clinical validation studies.

Cohort selection for a hyperlipidemia and b hypertension clinical validation in Vanderbilt's EHR. EHR: electronic health record; LDL-C: low-density lipoprotein cholesterol; SBP: systolic blood pressure.



Supplementary Figure 4. Cohort selection for All of Us clinical validation studies.

Cohort selection for **a** hyperlipidemia and **b** hypertension clinical validation in All of Us data set. LDL-C: low-density lipoprotein cholesterol; SBP: systolic blood pressure.

Supplementary Table 1: S-PrediXcan estimated transcriptomic signatures used to query iLINCS for hyperlipidemia drug repurposing candidates.

phenotype	gene_name	zscore	pvalue
Hyperlipidemia	GEMIN7	11.73	8.83E-32
Hyperlipidemia	CELSR2	10.11	5.05E-24
Hyperlipidemia	MAU2	8.62	6.74E-18
Hyperlipidemia	UBXN1	8.55	1.23E-17
Hyperlipidemia	PUM2	7.66	1.92E-14
Hyperlipidemia	PVRL2	6.43	1.25E-10
Hyperlipidemia	VASP	6.08	1.17E-09
Hyperlipidemia	QPCTL	5.83	5.53E-09
Hyperlipidemia	TMEM161A	5.34	9.54E-08
Hyperlipidemia	OPA3	4.81	1.53E-06
Hyperlipidemia	DPP3	4.33	1.51E-05
Hyperlipidemia	ROM1	3.95	7.79E-05
Hyperlipidemia	COL10A1	3.90	9.74E-05
Hyperlipidemia	HSPA1L	3.90	9.82E-05
Hyperlipidemia	CCS	3.88	1.03E-04
Hyperlipidemia	UEVLD	3.84	1.25E-04
Hyperlipidemia	MCM6	3.79	1.53E-04
Hyperlipidemia	GDI2	3.78	1.56E-04
Hyperlipidemia	DARS	3.76	1.71E-04
Hyperlipidemia	TOM1	3.75	1.77E-04
Hyperlipidemia	GALK1	3.75	1.80E-04
Hyperlipidemia	NUDCD3	3.74	1.83E-04
Hyperlipidemia	PMF1	3.66	2.53E-04
Hyperlipidemia	ALDH2	3.56	3.70E-04
Hyperlipidemia	KPNA1	3.52	4.29E-04
Hyperlipidemia	MPI	3.50	4.58E-04
Hyperlipidemia	HLA-DQA1	3.50	4.61E-04
Hyperlipidemia	OASL	3.47	5.13E-04
Hyperlipidemia	SPCS1	3.44	5.80E-04
Hyperlipidemia	GFRA2	3.38	7.24E-04

phenotype	gene_name	zscore	pvalue
Hyperlipidemia	HLA-DQB1	3.33	8.78E-04
Hyperlipidemia	LDLRAP1	3.29	9.85E-04
Hyperlipidemia	M6PR	3.29	1.01E-03
Hyperlipidemia	SMPD2	3.26	1.12E-03
Hyperlipidemia	JUND	3.22	1.30E-03
Hyperlipidemia	MAPRE3	3.20	1.37E-03
Hyperlipidemia	SCAMP2	3.14	1.71E-03
Hyperlipidemia	TRIM22	3.13	1.77E-03
Hyperlipidemia	TPM3	3.12	1.80E-03
Hyperlipidemia	WDR25	3.10	1.91E-03
Hyperlipidemia	PYGB	3.09	1.98E-03
Hyperlipidemia	SLC25A28	3.09	2.01E-03
Hyperlipidemia	TMED1	3.03	2.45E-03
Hyperlipidemia	PPP5C	3.02	2.52E-03
Hyperlipidemia	ARSA	3.02	2.56E-03
Hyperlipidemia	DDX11	3.01	2.63E-03
Hyperlipidemia	UQCC1	2.95	3.15E-03
Hyperlipidemia	PDZK1IP1	2.95	3.17E-03
Hyperlipidemia	CSNK1D	2.91	3.56E-03
Hyperlipidemia	LSM7	2.91	3.64E-03
Hyperlipidemia	PSRC1	-33.97	7.06E-253
Hyperlipidemia	ECSIT	-15.55	1.63E-54
Hyperlipidemia	S1PR5	-13.44	3.31E-41
Hyperlipidemia	TMEM258	-12.26	1.41E-34
Hyperlipidemia	GATAD2A	-9.47	2.83E-21
Hyperlipidemia	FADS2	-8.16	3.40E-16
Hyperlipidemia	HPR	-7.13	1.03E-12
Hyperlipidemia	GRINA	-6.81	9.85E-12
Hyperlipidemia	PLEC	-6.59	4.54E-11
Hyperlipidemia	HP	-5.56	2.72E-08
Hyperlipidemia	CKM	-5.43	5.62E-08
Hyperlipidemia	ICAM1	-5.43	5.64E-08

phenotype	gene_name	zscore	pvalue
Hyperlipidemia	RHD	-5.35	8.58E-08
Hyperlipidemia	EVI5	-5.21	1.92E-07
Hyperlipidemia	C19orf66	-4.77	1.88E-06
Hyperlipidemia	C6orf106	-4.59	4.47E-06
Hyperlipidemia	GSTM4	-4.59	4.51E-06
Hyperlipidemia	LDLR	-4.57	4.93E-06
Hyperlipidemia	MAP3K11	-4.34	1.42E-05
Hyperlipidemia	C19orf60	-4.23	2.31E-05
Hyperlipidemia	TBKBP1	-4.04	5.31E-05
Hyperlipidemia	EIF3G	-3.97	7.27E-05
Hyperlipidemia	ARID1A	-3.92	8.96E-05
Hyperlipidemia	PIGV	-3.84	1.21E-04
Hyperlipidemia	RPS6	-3.82	1.33E-04
Hyperlipidemia	C10orf88	-3.79	1.53E-04
Hyperlipidemia	C2	-3.75	1.77E-04
Hyperlipidemia	CTC1	-3.72	1.97E-04
Hyperlipidemia	HLA-DRA	-3.68	2.34E-04
Hyperlipidemia	KCNC4	-3.61	3.06E-04
Hyperlipidemia	HECTD4	-3.60	3.19E-04
Hyperlipidemia	ZNF668	-3.56	3.72E-04
Hyperlipidemia	TAF6L	-3.55	3.90E-04
Hyperlipidemia	IQGAP2	-3.46	5.43E-04
Hyperlipidemia	DBN1	-3.44	5.73E-04
Hyperlipidemia	SMG5	-3.44	5.87E-04
Hyperlipidemia	BACE1	-3.41	6.52E-04
Hyperlipidemia	DEF6	-3.37	7.55E-04
Hyperlipidemia	TMEM50A	-3.33	8.56E-04
Hyperlipidemia	WARS	-3.29	1.01E-03
Hyperlipidemia	ARCN1	-3.27	1.06E-03
Hyperlipidemia	DDB1	-3.23	1.25E-03
Hyperlipidemia	HLA-C	-3.19	1.41E-03
Hyperlipidemia	RAB3IL1	-3.15	1.63E-03

phenotype	gene_name	zscore	pvalue
Hyperlipidemia	COQ9	-3.15	1.64E-03
Hyperlipidemia	COPB1	-3.14	1.68E-03
Hyperlipidemia	ASPSCR1	-3.12	1.79E-03
Hyperlipidemia	PEX6	-3.12	1.82E-03
Hyperlipidemia	LYRM9	-3.11	1.87E-03
Hyperlipidemia	PAQR6	-3.10	1.94E-03

Supplementary Table 2: S-PrediXcan estimated transcriptomic signatures used to query iLINCS for hypertension drug repurposing candidates.

phenotype	gene_name	zscore	pvalue
Hypertension	FES	-9.49	1.32E-38
Hypertension	CLCN6	-8.84	4.61E-62
Hypertension	C20orf187	-8.18	2.89E-16
Hypertension	RP3-473L9.4	-6.86	4.91E-13
Hypertension	NDUFS3	-6.51	2.47E-20
Hypertension	SLC5A11	-6.36	3.72E-14
Hypertension	HELLS	-5.53	4.48E-13
Hypertension	TBX2	-5.51	1.16E-12
Hypertension	AGBL2	-5.07	4.06E-17
Hypertension	CHP1	-4.60	5.47E-10
Hypertension	CFDP1	-4.52	8.73E-11
Hypertension	ARHGAP42	-4.41	2.28E-35
Hypertension	ACADVL	-4.33	5.78E-12
Hypertension	FHL5	-4.10	2.22E-10
Hypertension	HFE	-4.05	3.30E-10
Hypertension	RP11-103J8.2	-3.68	1.89E-15
Hypertension	ENPEP	-3.27	1.09E-10
Hypertension	TMEM170A	-2.91	1.29E-11
Hypertension	ITGB5	-2.77	2.30E-10
Hypertension	HIC1	-2.52	3.20E-18
Hypertension	NT5C2	-2.37	1.07E-31
Hypertension	ATP2B1-AS1	-2.26	7.00E-26
Hypertension	TNNT3	-2.17	1.16E-24
Hypertension	CYP2C9	-2.01	2.96E-11
Hypertension	ZNF827	-1.99	5.16E-11
Hypertension	LMAN1L	-1.95	3.63E-15
Hypertension	TMEM133	-1.69	3.25E-34
Hypertension	PLCE1	-1.64	1.86E-21
Hypertension	SLC4A7	-1.59	6.53E-13
Hypertension	C5orf47	-1.58	5.16E-11

phenotype	gene_name	zscore	pvalue
Hypertension	AP3D1	-1.56	1.17E-12
Hypertension	CLDN7	-1.54	8.20E-11
Hypertension	ADAM11	-1.44	2.63E-12
Hypertension	LRP4	-1.16	4.51E-12
Hypertension	INA	-1.08	1.00E-20
Hypertension	BCAR1	-1.04	8.37E-11
Hypertension	COX14	-1.01	2.46E-10
Hypertension	MAPK4	-0.97	4.50E-14
Hypertension	NICN1	-0.86	4.83E-10
Hypertension	FERMT2	-0.81	5.19E-12
Hypertension	NPR3	-0.65	5.72E-51
Hypertension	CYP2C19	-0.62	3.21E-14
Hypertension	YBX2	-0.59	1.12E-10
Hypertension	CYP2C18	-0.38	3.84E-13
Hypertension	CPLX3	-0.30	1.11E-12
Hypertension	CERS5	-0.27	6.35E-11
Hypertension	TNFSF13	-0.23	1.63E-10
Hypertension	ADAMTS8	-0.20	3.12E-15
Hypertension	MYOZ1	-0.16	1.32E-11
Hypertension	GUCY1A3	-0.14	4.03E-24
Hypertension	SYNPO2L	-0.12	1.00E-10
Hypertension	WWP2	-0.08	2.38E-10
Hypertension	AGT	-0.05	3.88E-13
Hypertension	ZFYVE1	0.13	3.08E-10
Hypertension	CIB4	0.16	8.28E-22
Hypertension	SETBP1	0.16	1.35E-11
Hypertension	NFE2L1	0.27	6.23E-11
Hypertension	MYH7	0.36	1.72E-10
Hypertension	ARHGAP24	0.66	5.66E-15
Hypertension	KCNK3	0.69	6.70E-25
Hypertension	TNS2	0.78	1.00E-11
Hypertension	FBN2	1.18	7.48E-12

phenotype	gene_name	zscore	pvalue
Hypertension	CACNB2	1.20	3.80E-17
Hypertension	ATP2B1	1.35	1.51E-23
Hypertension	CCDC71L	1.36	5.30E-18
Hypertension	C17orf82	1.44	1.66E-11
Hypertension	LINC01358	1.48	1.01E-11
Hypertension	NGF	1.60	6.23E-12
Hypertension	HOXA10	1.70	7.99E-11
Hypertension	DLG4	1.90	2.72E-12
Hypertension	INO80	1.95	7.18E-11
Hypertension	MADD	2.06	1.58E-22
Hypertension	ABHD17C	2.10	3.04E-11
Hypertension	ADRB1	2.25	9.58E-14
Hypertension	FGF5	2.30	2.21E-39
Hypertension	EFEMP1	2.30	4.65E-10
Hypertension	CYP1A1	2.33	3.34E-12
Hypertension	PRR33	2.51	1.93E-23
Hypertension	VARS	2.68	2.30E-10
Hypertension	GPER1	2.71	3.88E-12
Hypertension	LCORL	3.08	2.39E-10
Hypertension	SHBG	3.32	8.50E-12
Hypertension	TCEA2	3.44	4.02E-12
Hypertension	BAG6	3.46	4.53E-12
Hypertension	SLC35E2	3.94	4.47E-10
Hypertension	AGAP5	4.01	3.27E-10
Hypertension	ULK3	4.02	1.24E-12
Hypertension	PRRC2A	4.07	1.69E-12
Hypertension	IGFBP3	4.88	2.66E-11
Hypertension	FN1	4.92	1.01E-12
Hypertension	CSK	4.98	8.62E-14
Hypertension	CEP68	5.22	1.97E-11
Hypertension	SIPA1	5.25	1.35E-10
Hypertension	SLC2A4	5.29	7.31E-15

phenotype	gene_name	zscore	pvalue
Hypertension	RERE	5.62	3.29E-12
Hypertension	LSP1	5.64	2.14E-23
Hypertension	CBX1	5.80	1.52E-10
Hypertension	FURIN	5.80	7.94E-25
Hypertension	CDC16	5.97	5.46E-10
Hypertension	CTB-30L5.1	8.37	1.25E-49
Hypertension	NMT1	9.30	4.43E-25

Supplementary Table 3: iLINCS drug repurposing candidate list for hyperlipidemia.

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_2102	butenafine	150 mg/kg	liver	1 d	-0.54	8.19E-05
Hyperlipidemia	DM_939	lorazepam	2000 mg/kg	liver	1 d	-0.54	1.36E-07
Hyperlipidemia	DM_3106	fenofibrate	215 mg/kg	kidney	3 d	-0.48	5.81E-04
Hyperlipidemia	DM_4565	quinapril	2500 mg/kg	kidney	1 d	-0.48	6.19E-04
Hyperlipidemia	DM_435	clotrimazole	60 uM	primary rat hepatocytes	1 d	-0.47	5.86E-06
Hyperlipidemia	DM_3566	ketorolac	48 mg/kg	liver	1 d	-0.47	8.13E-04
Hyperlipidemia	DM_743	gefitinib	116 mg/kg	liver	5 d	-0.47	7.55E-06
Hyperlipidemia	DM_4398	phenylbutazone	368 mg/kg	kidney	5 d	-0.46	1.04E-03
Hyperlipidemia	DM_5060	tolazamide	1500 mg/kg	kidney	3 d	-0.45	1.32E-03
Hyperlipidemia	DM_2710	diazepam	710 mg/kg	kidney	5 d	-0.45	1.34E-03
Hyperlipidemia	DM_2757	diethylstilbestrol	2.8 mg/kg	kidney	1 d	-0.45	1.39E-03
Hyperlipidemia	DM_1057	nevirapine	250 uM	primary rat hepatocytes	1 d	-0.45	1.93E-05
Hyperlipidemia	DM_1123	oxfendazole	1500 mg/kg	liver	1 d	-0.45	1.95E-05
Hyperlipidemia	DM_759	gentamicin	2900 uM	primary rat hepatocytes	1 d	-0.45	2.00E-05
Hyperlipidemia	DM_2817	dipyridamole	750 mg/kg	liver	3 d	-0.44	1.62E-03
Hyperlipidemia	DM_2427	cisplatin	2 mg/kg	kidney	3 d	-0.44	1.62E-03
Hyperlipidemia	DM_2650	dactinomycin	0.06 mg/kg	bone marrow	5 d	-0.43	2.42E-03
Hyperlipidemia	DM_3498	irinotecan	5 mg/kg	bone marrow	3 d	-0.43	2.48E-03
Hyperlipidemia	DM_5079	tramadol	114 mg/kg	brain	5 d	-0.42	2.86E-03
Hyperlipidemia	DM_5207	vecuronium	0.05 mg/kg	liver	1 d	-0.41	3.80E-03
Hyperlipidemia	DM_4420	pioglitazone	1500 mg/kg	liver	5 d	-0.41	3.95E-03
Hyperlipidemia	DM_84	alendronate	138 mg/kg	heart	1 d	-0.41	1.22E-04
Hyperlipidemia	DM_2374	cholecalciferol	8 mg/kg	kidney	5 d	-0.41	4.24E-03
Hyperlipidemia	DM_559	doxorubicin	1.5 uM	primary rat hepatocytes	1 d	-0.41	1.31E-04
Hyperlipidemia	DM_969	megestrol	132 mg/kg	liver	1 d	-0.40	1.36E-04
Hyperlipidemia	DM_5078	tramadol	114 mg/kg	brain	3 d	-0.40	4.42E-03
Hyperlipidemia	DM_3984	moxonidine	17 mg/kg	heart	5 d	-0.40	4.55E-03
Hyperlipidemia	DM_1595	acyclovir	330 mg/kg	liver	1 d	-0.40	4.83E-03
Hyperlipidemia	DM_1374	valproate	1340 mg/kg	liver	5 d	-0.40	1.67E-04
Hyperlipidemia	DM_4711	sertraline	23 uM	primary rat hepatocytes	1 d	-0.40	4.95E-03
Hyperlipidemia	DM_1775	antipyrine	1500 mg/kg	kidney	5 d	-0.40	5.22E-03

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_4304	paroxetine	30 uM	primary rat hepatocytes	1 d	-0.40	5.36E-03
Hyperlipidemia	DM_2949	ergocalciferol	15 mg/kg	kidney	3 d	-0.39	6.08E-03
Hyperlipidemia	DM_3741	loratadine	62.25 uM	primary rat hepatocytes	1 d	-0.39	6.68E-03
Hyperlipidemia	DM_4583	raloxifene	650 mg/kg	liver	5 d	-0.39	6.69E-03
Hyperlipidemia	DM_1794	aspirin	167 mg/kg	liver	3 d	-0.38	7.05E-03
Hyperlipidemia	DM_4110	nitrofurantoin	76 mg/kg	liver	5 d	-0.38	7.31E-03
Hyperlipidemia	DM_2129	calcitriol	0.04 mg/kg	kidney	5 d	-0.38	7.46E-03
Hyperlipidemia	DM_1762	amprenavir	600 mg/kg	kidney	5 d	-0.38	7.48E-03
Hyperlipidemia	DM_3406	ibuprofen	90 mg/kg	liver	3 d	-0.38	7.78E-03
Hyperlipidemia	DM_1052	neomycin	56 mg/kg	kidney	5 d	-0.38	3.71E-04
Hyperlipidemia	DM_4375	phenelzine	27 mg/kg	brain	5 d	-0.37	8.73E-03
Hyperlipidemia	DM_2302	chlorambucil	0.6 mg/kg	liver	0.25 d	-0.37	1.01E-02
Hyperlipidemia	DM_3284	gentian violet	18 mg/kg	liver	3 d	-0.37	1.01E-02
Hyperlipidemia	DM_4443	pramoxine	526 mg/kg	heart	5 d	-0.37	1.05E-02
Hyperlipidemia	DM_653	fenofibrate	43 mg/kg	liver	1 d	-0.36	6.72E-04
Hyperlipidemia	DM_2371	cholecalciferol	8 mg/kg	kidney	3 d	-0.36	1.12E-02
Hyperlipidemia	DM_754	gemfibrozil	700 mg/kg	liver	3 d	-0.36	6.94E-04
Hyperlipidemia	DM_4963	terbinafine	2000 mg/kg	liver	5 d	-0.36	1.13E-02
Hyperlipidemia	DM_3918	miconazole	200 mg/kg	liver	3 d	-0.36	1.14E-02
Hyperlipidemia	DM_1667	alprazolam	115 mg/kg	liver	1 d	-0.36	1.35E-02
Hyperlipidemia	DM_1143	pemoline	70 mg/kg	liver	1 d	-0.36	8.57E-04
Hyperlipidemia	DM_1629	alendronate	138 mg/kg	heart	1 d	-0.35	1.34E-02
Hyperlipidemia	DM_4413	pioglitazone	1500 mg/kg	liver	1 d	-0.35	1.34E-02
Hyperlipidemia	DM_3196	flurbiprofen	10 mg/kg	kidney	1 d	-0.35	2.02E-02
Hyperlipidemia	DM_811	ifosfamide	143 mg/kg	liver	3 d	-0.35	1.13E-03
Hyperlipidemia	DM_970	megestrol	132 mg/kg	liver	5 d	-0.35	1.13E-03
Hyperlipidemia	DM_4065	nicotine	75 mg/kg	brain	5 d	-0.35	1.53E-02
Hyperlipidemia	DM_4196	norethindrone	375 mg/kg	liver	5 d	-0.35	1.54E-02
Hyperlipidemia	DM_941	lorazepam	2000 mg/kg	liver	5 d	-0.35	1.23E-03
Hyperlipidemia	DM_670	finasteride	800 mg/kg	liver	5 d	-0.35	1.30E-03
Hyperlipidemia	DM_4662	rosiglitazone	10 mg/kg	liver	0.25 d	-0.34	1.66E-02
Hyperlipidemia	DM_2622	cyproterone	2500 mg/kg	liver	5 d	-0.34	1.68E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_3490	iproniazid	46 mg/kg	brain	0.25 d	-0.34	1.68E-02
Hyperlipidemia	DM_2368	cholecalciferol	8 mg/kg	kidney	1 d	-0.34	1.72E-02
Hyperlipidemia	DM_1287	spironolactone	300 mg/kg	liver	5 d	-0.34	1.57E-03
Hyperlipidemia	DM_2517	clotrimazole	60 uM	primary rat hepatocytes	1 d	-0.34	1.85E-02
Hyperlipidemia	DM_1812	atorvastatin	300 mg/kg	liver	1 d	-0.34	1.89E-02
Hyperlipidemia	DM_2442	citalopram	40 mg/kg	liver	3 d	-0.34	1.92E-02
Hyperlipidemia	DM_4636	rofecoxib	1550 mg/kg	liver	3 d	-0.34	1.93E-02
Hyperlipidemia	DM_4878	sulfisoxazole	2500 mg/kg	liver	1 d	-0.34	2.08E-02
Hyperlipidemia	DM_3091	fenbendazole	375 mg/kg	liver	1 d	-0.34	1.98E-02
Hyperlipidemia	DM_3366	hydroquinone	800 mg/kg	kidney	1 d	-0.34	1.98E-02
Hyperlipidemia	DM_2265	celecoxib	100 uM	primary rat hepatocytes	1 d	-0.33	2.00E-02
Hyperlipidemia	DM_436	clotrimazole	89 mg/kg	liver	1 d	-0.33	1.87E-03
Hyperlipidemia	DM_2148	captopril	1750 mg/kg	kidney	0.25 d	-0.33	2.02E-02
Hyperlipidemia	DM_4524	promazine	100 mg/kg	liver	0.25 d	-0.33	2.39E-02
Hyperlipidemia	DM_1803	aspirin	375 mg/kg	kidney	5 d	-0.33	2.10E-02
Hyperlipidemia	DM_3119	fenoprofen	52 mg/kg	kidney	1 d	-0.33	2.14E-02
Hyperlipidemia	DM_291	calcitriol	0.04 mg/kg	kidney	1 d	-0.33	2.09E-03
Hyperlipidemia	DM_4050	nevirapine	200 mg/kg	liver	3 d	-0.33	2.16E-02
Hyperlipidemia	DM_1928	benzoate	1700 mg/kg	liver	5 d	-0.33	2.18E-02
Hyperlipidemia	DM_4011	naproxen	134 mg/kg	kidney	3 d	-0.33	2.20E-02
Hyperlipidemia	DM_3394	ibuprofen	275 mg/kg	liver	1 d	-0.33	2.23E-02
Hyperlipidemia	DM_115	altretamine	13 mg/kg	liver	3 d	-0.33	2.36E-03
Hyperlipidemia	DM_5155	valdecoxib	404 mg/kg	kidney	3 d	-0.33	2.32E-02
Hyperlipidemia	DM_3740	loperamide	47 mg/kg	intestine	3 d	-0.33	2.34E-02
Hyperlipidemia	DM_3254	gemfibrozil	100 mg/kg	kidney	1 d	-0.33	2.73E-02
Hyperlipidemia	DM_1804	aspirin	375 mg/kg	liver	5 d	-0.33	2.41E-02
Hyperlipidemia	DM_484	cyproterone	2500 mg/kg	liver	5 d	-0.33	2.54E-03
Hyperlipidemia	DM_1313	sulpiride	667 uM	primary rat hepatocytes	1 d	-0.32	2.63E-03
Hyperlipidemia	DM_2508	clopidogrel	400 mg/kg	bone marrow	1 d	-0.32	2.67E-02
Hyperlipidemia	DM_3097	fenofibrate	250 uM	primary rat hepatocytes	1 d	-0.32	2.56E-02
Hyperlipidemia	DM_5128	troglitazone	1200 mg/kg	liver	1 d	-0.32	2.57E-02
Hyperlipidemia	DM_4740	simvastatin	160 uM	primary rat hepatocytes	1 d	-0.32	2.58E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_3444	imatinib	150 mg/kg	liver	1 d	-0.32	2.59E-02
Hyperlipidemia	DM_1808	atorvastatin	250 uM	primary rat hepatocytes	1 d	-0.32	2.60E-02
Hyperlipidemia	DM_866	kanamycin	6584 uM	primary rat hepatocytes	1 d	-0.32	3.00E-03
Hyperlipidemia	DM_1235	risperidone	22 mg/kg	heart	1 d	-0.32	3.04E-03
Hyperlipidemia	DM_2649	dactinomycin	0.06 mg/kg	bone marrow	3 d	-0.32	2.72E-02
Hyperlipidemia	DM_524	diazepam	710 mg/kg	liver	3 d	-0.32	3.30E-03
Hyperlipidemia	DM_2404	cisapride	250 mg/kg	intestine	3 d	-0.32	2.82E-02
Hyperlipidemia	DM_1197	progesterone	164 mg/kg	liver	1 d	-0.32	3.31E-03
Hyperlipidemia	DM_1262	sertraline	23 uM	primary rat hepatocytes	1 d	-0.32	3.32E-03
Hyperlipidemia	DM_3794	mefenamate	93 mg/kg	liver	0.25 d	-0.32	2.83E-02
Hyperlipidemia	DM_1582	acetazolamide	250 mg/kg	liver	5 d	-0.32	2.85E-02
Hyperlipidemia	DM_816	imatinib	15 mg/kg	heart	1 d	-0.32	3.45E-03
Hyperlipidemia	DM_4243	ondansetron	84 mg/kg	intestine	1 d	-0.31	2.92E-02
Hyperlipidemia	DM_2433	cisplatin	2 mg/kg	kidney	5 d	-0.31	2.93E-02
Hyperlipidemia	DM_1239	rofecoxib	111 uM	primary rat hepatocytes	1 d	-0.31	3.87E-03
Hyperlipidemia	DM_2628	cytarabine	487 mg/kg	liver	1 d	-0.31	3.13E-02
Hyperlipidemia	DM_4716	sibutramine	50 mg/kg	brain	3 d	-0.31	3.16E-02
Hyperlipidemia	DM_3423	ifosfamide	11000 uM	primary rat hepatocytes	1 d	-0.31	3.16E-02
Hyperlipidemia	DM_2863	doxorubicin	0.65 mg/kg	liver	0.25 d	-0.31	3.16E-02
Hyperlipidemia	DM_1914	benzethonium	138 mg/kg	heart	3 d	-0.31	3.21E-02
Hyperlipidemia	DM_2432	cisplatin	2 mg/kg	bone marrow	5 d	-0.31	3.24E-02
Hyperlipidemia	DM_4606	rifabutin	1500 mg/kg	liver	3 d	-0.31	3.30E-02
Hyperlipidemia	DM_3567	ketorolac	48 mg/kg	liver	3 d	-0.31	3.39E-02
Hyperlipidemia	DM_4992	thiabendazole	10 mg/kg	liver	1 d	-0.31	3.48E-02
Hyperlipidemia	DM_284	busulfan	500 uM	primary rat hepatocytes	1 d	-0.31	4.77E-03
Hyperlipidemia	DM_3198	flurbiprofen	10 mg/kg	kidney	5 d	-0.30	3.54E-02
Hyperlipidemia	DM_2406	cisapride	250 mg/kg	intestine	5 d	-0.30	3.56E-02
Hyperlipidemia	DM_5223	verapamil	108 mg/kg	heart	3 d	-0.30	3.61E-02
Hyperlipidemia	DM_4446	pravastatin	1200 mg/kg	liver	1 d	-0.30	3.68E-02
Hyperlipidemia	DM_2725	diclofenac	10 mg/kg	liver	0.25 d	-0.30	3.88E-02
Hyperlipidemia	DM_4305	paroxetine	104 mg/kg	brain	0.25 d	-0.30	3.91E-02
Hyperlipidemia	DM_703	fluphenazine	2.5 mg/kg	liver	3 d	-0.30	5.81E-03

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_4108	nitrofurantoin	76 mg/kg	liver	1 d	-0.30	3.93E-02
Hyperlipidemia	DM_2792	digoxin	0.26 mg/kg	liver	0.25 d	-0.30	3.97E-02
Hyperlipidemia	DM_4396	phenylbutazone	368 mg/kg	kidney	1 d	-0.30	4.06E-02
Hyperlipidemia	DM_2419	cisplatin	1.17 mg/kg	liver	0.25 d	-0.30	4.10E-02
Hyperlipidemia	DM_2504	clonazepam	2500 mg/kg	liver	5 d	-0.30	4.13E-02
Hyperlipidemia	DM_1257	salicylamide	1300 mg/kg	liver	1 d	-0.30	6.34E-03
Hyperlipidemia	DM_3756	lovastatin	1500 mg/kg	liver	1 d	-0.29	4.30E-02
Hyperlipidemia	DM_3314	griseofulvin	75 uM	primary rat hepatocytes	1 d	-0.29	4.33E-02
Hyperlipidemia	DM_868	ketoconazole	90 uM	primary rat hepatocytes	1 d	-0.29	6.95E-03
Hyperlipidemia	DM_3850	methotrexate	27 mg/kg	kidney	1 d	-0.29	4.47E-02
Hyperlipidemia	DM_4607	rifabutin	1500 mg/kg	kidney	5 d	-0.29	4.48E-02
Hyperlipidemia	DM_2054	bromfenac	5 mg/kg	liver	5 d	-0.29	4.52E-02
Hyperlipidemia	DM_3890	metoclopramide	185 mg/kg	intestine	1 d	-0.29	4.53E-02
Hyperlipidemia	DM_1919	benzethonium	30 mg/kg	heart	5 d	-0.29	4.54E-02
Hyperlipidemia	DM_3232	gabapentin	500 uM	primary rat hepatocytes	1 d	-0.29	4.61E-02
Hyperlipidemia	DM_137	amitriptyline	160 mg/kg	kidney	1 d	-0.29	7.61E-03
Hyperlipidemia	DM_1611	acyclovir	980 mg/kg	bone marrow	5 d	-0.29	4.64E-02
Hyperlipidemia	DM_3244	gallamine	2.5 mg/kg	heart	3 d	-0.29	4.68E-02
Hyperlipidemia	DM_434	clotrimazole	52 mg/kg	liver	1 d	-0.29	7.79E-03
Hyperlipidemia	DM_807	ifosfamide	17 mg/kg	liver	1 d	-0.29	8.07E-03
Hyperlipidemia	DM_666	fenofibrate	430 mg/kg	liver	5 d	-0.29	8.26E-03
Hyperlipidemia	DM_5265	zileuton	450 mg/kg	kidney	3 d	-0.29	4.89E-02
Hyperlipidemia	DM_4624	ritonavir	1200 mg/kg	kidney	3 d	-0.29	4.93E-02
Hyperlipidemia	DM_5127	troglitazone	100 mg/kg	liver	1 d	-0.29	4.93E-02
Hyperlipidemia	DM_3197	flurbiprofen	10 mg/kg	kidney	3 d	-0.29	4.93E-02
Hyperlipidemia	DM_1195	procarbazine	54 mg/kg	liver	5 d	-0.29	8.54E-03
Hyperlipidemia	DM_808	ifosfamide	143 mg/kg	liver	0.25 d	-0.28	9.19E-03
Hyperlipidemia	DM_580	econazole	43 mg/kg	liver	5 d	-0.28	9.26E-03
Hyperlipidemia	DM_362	chlorambucil	0.6 mg/kg	liver	0.25 d	-0.28	9.73E-03
Hyperlipidemia	DM_1279	sotalol	2000 mg/kg	liver	3 d	-0.28	1.02E-02
Hyperlipidemia	DM_1266	simvastatin	160 uM	primary rat hepatocytes	1 d	-0.28	1.12E-02
Hyperlipidemia	DM_1352	tretinoin	156 uM	primary rat hepatocytes	1 d	-0.27	1.29E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_299	captopril	1750 mg/kg	kidney	5 d	-0.27	1.46E-02
Hyperlipidemia	DM_667	fenofibrate	43 mg/kg	liver	5 d	-0.26	1.54E-02
Hyperlipidemia	DM_707	fluvastatin	5 mg/kg	liver	1 d	-0.26	1.61E-02
Hyperlipidemia	DM_66	acetaminophen	486 mg/kg	kidney	3 d	-0.26	1.71E-02
Hyperlipidemia	DM_700	fluphenazine	45 uM	primary rat hepatocytes	1 d	-0.26	1.74E-02
Hyperlipidemia	DM_1378	valsartan	30 mg/kg	heart	0.25 d	-0.26	1.82E-02
Hyperlipidemia	DM_576	econazole	334 mg/kg	liver	0.25 d	-0.26	1.88E-02
Hyperlipidemia	DM_775	haloperidol	77 uM	primary rat hepatocytes	1 d	-0.25	1.99E-02
Hyperlipidemia	DM_1311	sulindac	23 mg/kg	kidney	5 d	-0.25	2.20E-02
Hyperlipidemia	DM_532	diclofenac	3.5 mg/kg	kidney	5 d	-0.25	2.23E-02
Hyperlipidemia	DM_323	carboplatin	5 mg/kg	kidney	5 d	-0.25	2.25E-02
Hyperlipidemia	DM_330	carmustine	4 mg/kg	liver	3 d	-0.25	2.25E-02
Hyperlipidemia	DM_427	clonazepam	2500 mg/kg	liver	1 d	-0.25	2.33E-02
Hyperlipidemia	DM_699	fluphenazine	2.5 mg/kg	liver	1 d	-0.25	2.35E-02
Hyperlipidemia	DM_462	cyclophosphamide	1320 uM	primary rat hepatocytes	1 d	-0.25	2.36E-02
Hyperlipidemia	DM_867	ketoconazole	227 mg/kg	liver	1 d	-0.25	2.39E-02
Hyperlipidemia	DM_199	azithromycin	225 mg/kg	kidney	5 d	0.25	2.29E-02
Hyperlipidemia	DM_734	gefitinib	116 mg/kg	liver	1 d	0.25	2.21E-02
Hyperlipidemia	DM_949	lovastatin	450 mg/kg	kidney	1 d	0.25	2.13E-02
Hyperlipidemia	DM_613	erythromycin	1500 mg/kg	liver	5 d	0.25	2.13E-02
Hyperlipidemia	DM_392	cholecalciferol	8 mg/kg	liver	3 d	0.25	2.03E-02
Hyperlipidemia	DM_1201	progesterone	11.3 mg/kg	liver	5 d	0.25	1.93E-02
Hyperlipidemia	DM_1124	oxybutynin	230 mg/kg	heart	3 d	0.26	1.92E-02
Hyperlipidemia	DM_857	isotretinoin	13 mg/kg	liver	0.25 d	0.26	1.72E-02
Hyperlipidemia	DM_1186	prednisone	68 mg/kg	heart	1 d	0.27	1.42E-02
Hyperlipidemia	DM_192	azathioprine	54 mg/kg	kidney	3 d	0.27	1.41E-02
Hyperlipidemia	DM_742	gefitinib	116 mg/kg	kidney	5 d	0.27	1.31E-02
Hyperlipidemia	DM_663	fenofibrate	215 mg/kg	heart	5 d	0.27	1.17E-02
Hyperlipidemia	DM_715	fluvastatin	5 mg/kg	liver	5 d	0.28	1.12E-02
Hyperlipidemia	DM_1259	salicylic acid	223 mg/kg	liver	1 d	0.28	1.08E-02
Hyperlipidemia	DM_302	carbamazepine	490 mg/kg	liver	5 d	0.28	1.02E-02
Hyperlipidemia	DM_942	losartan	1000 mg/kg	heart	1 d	0.28	1.02E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_1337	testosterone	375 mg/kg	liver	5 d	0.28	9.42E-03
Hyperlipidemia	DM_1260	salicylic acid	223 mg/kg	liver	5 d	0.28	8.83E-03
Hyperlipidemia	DM_1862	azathioprine	20 mg/kg	spleen	3 d	0.28	4.99E-02
Hyperlipidemia	DM_814	imatinib	150 mg/kg	kidney	1 d	0.29	8.55E-03
Hyperlipidemia	DM_4987	theophylline	225 mg/kg	heart	1 d	0.29	4.94E-02
Hyperlipidemia	DM_4228	olanzapine	23 mg/kg	heart	1 d	0.29	4.93E-02
Hyperlipidemia	DM_3471	indomethacin	9.6 mg/kg	liver	3 d	0.29	4.83E-02
Hyperlipidemia	DM_4299	papaverine	69 mg/kg	heart	1 d	0.29	4.81E-02
Hyperlipidemia	DM_738	gefitinib	116 mg/kg	kidney	3 d	0.29	8.09E-03
Hyperlipidemia	DM_817	imatinib	15 mg/kg	kidney	1 d	0.29	8.07E-03
Hyperlipidemia	DM_3870	methyldopa	325 mg/kg	heart	1 d	0.29	4.72E-02
Hyperlipidemia	DM_3983	moxonidine	17 mg/kg	heart	3 d	0.29	4.59E-02
Hyperlipidemia	DM_1370	valproate	1000 uM	primary rat hepatocytes	1 d	0.29	7.40E-03
Hyperlipidemia	DM_1963	betamethasone	79 mg/kg	liver	1 d	0.29	4.53E-02
Hyperlipidemia	DM_5013	thioguanine	12 mg/kg	liver	0.25 d	0.29	4.52E-02
Hyperlipidemia	DM_4893	sulindac	132 mg/kg	liver	3 d	0.29	4.44E-02
Hyperlipidemia	DM_3656	leflunomide	30 mg/kg	liver	0.25 d	0.29	4.43E-02
Hyperlipidemia	DM_1233	rifabutin	1500 mg/kg	liver	5 d	0.29	7.05E-03
Hyperlipidemia	DM_938	loratadine	2000 mg/kg	heart	5 d	0.29	6.86E-03
Hyperlipidemia	DM_2767	diethylstilbestrol	280 mg/kg	liver	3 d	0.29	4.55E-02
Hyperlipidemia	DM_2182	carboplatin	14 mg/kg	kidney	1 d	0.29	4.26E-02
Hyperlipidemia	DM_4512	progesterone	148 uM	primary rat hepatocytes	1 d	0.29	4.24E-02
Hyperlipidemia	DM_457	cortisone	206 mg/kg	liver	5 d	0.30	6.31E-03
Hyperlipidemia	DM_3260	gemfibrozil	700 mg/kg	liver	0.25 d	0.30	4.08E-02
Hyperlipidemia	DM_3157	fludrocortisone	125 mg/kg	kidney	3 d	0.30	4.04E-02
Hyperlipidemia	DM_4946	teicoplanin	41 mg/kg	bone marrow	3 d	0.30	4.00E-02
Hyperlipidemia	DM_1967	betamethasone	79 mg/kg	liver	5 d	0.30	3.96E-02
Hyperlipidemia	DM_4090	nimodipine	1100 mg/kg	heart	3 d	0.30	3.96E-02
Hyperlipidemia	DM_176	atorvastatin	2.5 mg/kg	liver	1 d	0.30	5.82E-03
Hyperlipidemia	DM_3492	iproniazid	46 mg/kg	heart	3 d	0.30	3.92E-02
Hyperlipidemia	DM_1965	betamethasone	79 mg/kg	liver	3 d	0.30	3.92E-02
Hyperlipidemia	DM_928	lomustine	4.2 mg/kg	liver	1 d	0.30	5.73E-03

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_562	doxorubicin	3 mg/kg	liver	1 d	0.30	5.54E-03
Hyperlipidemia	DM_787	hydrocortisone	56 mg/kg	liver	1 d	0.30	5.52E-03
Hyperlipidemia	DM_4506	procarbazine	54 mg/kg	liver	5 d	0.30	3.79E-02
Hyperlipidemia	DM_4006	naproxen	10 mg/kg	kidney	0.25 d	0.30	3.78E-02
Hyperlipidemia	DM_4889	sulindac	23 mg/kg	kidney	0.25 d	0.30	3.77E-02
Hyperlipidemia	DM_1727	amitraz	75 mg/kg	heart	3 d	0.30	3.71E-02
Hyperlipidemia	DM_3864	methotrexate	27 mg/kg	spleen	3 d	0.30	3.67E-02
Hyperlipidemia	DM_4630	rofecoxib	775 mg/kg	kidney	1 d	0.30	3.61E-02
Hyperlipidemia	DM_648	fenbendazole	375 mg/kg	liver	1 d	0.30	4.98E-03
Hyperlipidemia	DM_164	aspirin	35 mg/kg	heart	3 d	0.30	4.94E-03
Hyperlipidemia	DM_4494	procarbazine	54 mg/kg	liver	1 d	0.30	3.51E-02
Hyperlipidemia	DM_3082	famciclovir	1200 mg/kg	liver	3 d	0.31	3.50E-02
Hyperlipidemia	DM_3531	isotretinoin	125 mg/kg	liver	5 d	0.31	3.48E-02
Hyperlipidemia	DM_4418	pioglitazone	300 mg/kg	liver	3 d	0.31	3.41E-02
Hyperlipidemia	DM_584	enoxacin	100 mg/kg	liver	0.25 d	0.31	4.53E-03
Hyperlipidemia	DM_736	gefitinib	58 mg/kg	kidney	1 d	0.31	4.52E-03
Hyperlipidemia	DM_1170	phenytoin	572 mg/kg	heart	5 d	0.31	4.35E-03
Hyperlipidemia	DM_4279	oxybutynin	230 mg/kg	heart	5 d	0.31	3.30E-02
Hyperlipidemia	DM_985	methimazole	100 mg/kg	liver	3 d	0.31	4.06E-03
Hyperlipidemia	DM_4030	neomycin	56 mg/kg	kidney	5 d	0.31	3.17E-02
Hyperlipidemia	DM_611	erythromycin	1500 mg/kg	liver	1 d	0.31	3.96E-03
Hyperlipidemia	DM_3412	idarubicin	1.5 uM	primary rat hepatocytes	1 d	0.31	3.13E-02
Hyperlipidemia	DM_1182	pramoxine	526 mg/kg	liver	3 d	0.31	3.75E-03
Hyperlipidemia	DM_4610	rifampin	99 mg/kg	kidney	1 d	0.31	3.01E-02
Hyperlipidemia	DM_1174	pioglitazone	1500 mg/kg	liver	3 d	0.31	3.69E-03
Hyperlipidemia	DM_3867	methotrexate	0.3 mg/kg	liver	3 d	0.31	2.99E-02
Hyperlipidemia	DM_3377	hydroxyurea	59 mg/kg	liver	0.25 d	0.31	2.99E-02
Hyperlipidemia	DM_3191	fluphenazine	22 mg/kg	liver	3 d	0.31	2.97E-02
Hyperlipidemia	DM_300	carbamazepine	490 mg/kg	liver	1 d	0.32	3.34E-03
Hyperlipidemia	DM_1007	mifepristone	300 mg/kg	liver	1 d	0.32	3.27E-03
Hyperlipidemia	DM_987	methotrexate	0.3 mg/kg	liver	1 d	0.32	3.27E-03
Hyperlipidemia	DM_3529	isotretinoin	125 mg/kg	liver	3 d	0.32	2.77E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_1342	thioguanine	12 mg/kg	liver	1 d	0.32	3.14E-03
Hyperlipidemia	DM_2879	doxorubicin	4 mg/kg	kidney	5 d	0.32	2.70E-02
Hyperlipidemia	DM_1286	spironolactone	300 mg/kg	kidney	5 d	0.32	3.09E-03
Hyperlipidemia	DM_3183	fluoxetine	52 mg/kg	kidney	5 d	0.32	2.68E-02
Hyperlipidemia	DM_3671	leflunomide	60 mg/kg	spleen	5 d	0.32	2.66E-02
Hyperlipidemia	DM_3178	fluoxetine	52 mg/kg	kidney	1 d	0.32	2.65E-02
Hyperlipidemia	DM_5021	thioguanine	12 mg/kg	liver	5 d	0.32	2.60E-02
Hyperlipidemia	DM_2866	doxorubicin	3 mg/kg	heart	3 d	0.32	2.54E-02
Hyperlipidemia	DM_592	epirubicin	2.7 mg/kg	heart	3 d	0.32	2.69E-03
Hyperlipidemia	DM_482	cyproterone	2500 mg/kg	liver	0.25 d	0.32	2.57E-03
Hyperlipidemia	DM_994	metoprolol	120 mg/kg	heart	1 d	0.33	2.51E-03
Hyperlipidemia	DM_5179	valproate	1500 mg/kg	heart	3 d	0.33	2.39E-02
Hyperlipidemia	DM_1118	omeprazole	435 uM	primary rat hepatocytes	1 d	0.33	2.47E-03
Hyperlipidemia	DM_4899	sulindac	132 mg/kg	liver	5 d	0.33	2.35E-02
Hyperlipidemia	DM_190	azathioprine	160 mg/kg	liver	3 d	0.33	2.34E-03
Hyperlipidemia	DM_591	epirubicin	2.7 mg/kg	liver	1 d	0.33	2.31E-03
Hyperlipidemia	DM_2198	carboplatin	6 mg/kg	kidney	3 d	0.33	2.16E-02
Hyperlipidemia	DM_367	chlorambucil	0.6 mg/kg	liver	5 d	0.33	2.07E-03
Hyperlipidemia	DM_2533	clozapine	95 mg/kg	heart	3 d	0.33	2.08E-02
Hyperlipidemia	DM_426	clomipramine	115 mg/kg	liver	3 d	0.33	1.93E-03
Hyperlipidemia	DM_2314	chlorambucil	4.5 mg/kg	spleen	5 d	0.33	2.03E-02
Hyperlipidemia	DM_5186	valproate	850 mg/kg	kidney	3 d	0.33	2.01E-02
Hyperlipidemia	DM_801	idarubicin	0.625 mg/kg	heart	1 d	0.33	1.85E-03
Hyperlipidemia	DM_1148	pentoxifylline	1170 mg/kg	thigh muscle	3 d	0.34	1.72E-03
Hyperlipidemia	DM_1115	omeprazole	365 uM	primary rat hepatocytes	1 d	0.34	1.59E-03
Hyperlipidemia	DM_3440	ifosfamide	143 mg/kg	heart	5 d	0.34	1.81E-02
Hyperlipidemia	DM_4923	tacrolimus	134 mg/kg	heart	3 d	0.34	1.76E-02
Hyperlipidemia	DM_702	fluphenazine	2.5 mg/kg	liver	0.25 d	0.34	1.47E-03
Hyperlipidemia	DM_2434	cisplatin	2 mg/kg	liver	5 d	0.34	1.75E-02
Hyperlipidemia	DM_4452	pravastatin	1200 mg/kg	thigh muscle	5 d	0.34	1.74E-02
Hyperlipidemia	DM_2921	enoxacin	750 mg/kg	liver	3 d	0.34	1.70E-02
Hyperlipidemia	DM_3064	etoposide	188 mg/kg	liver	5 d	0.34	1.70E-02

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_2449	citric acid	3000 mg/kg	liver	3 d	0.34	1.70E-02
Hyperlipidemia	DM_698	fluphenazine	22 mg/kg	liver	1 d	0.34	1.32E-03
Hyperlipidemia	DM_1723	amiodarone	147 mg/kg	kidney	5 d	0.35	1.62E-02
Hyperlipidemia	DM_3189	fluphenazine	2.5 mg/kg	liver	0.25 d	0.35	1.71E-02
Hyperlipidemia	DM_4587	ramipril	620 uM	primary rat hepatocytes	1 d	0.35	1.54E-02
Hyperlipidemia	DM_2505	clonidine	6 mg/kg	heart	1 d	0.35	1.46E-02
Hyperlipidemia	DM_4278	oxybutynin	230 mg/kg	heart	3 d	0.35	1.41E-02
Hyperlipidemia	DM_705	flurbiprofen	10 mg/kg	kidney	3 d	0.35	9.74E-04
Hyperlipidemia	DM_4394	phentolamine	44 mg/kg	heart	3 d	0.35	1.35E-02
Hyperlipidemia	DM_2814	diphenidol	300 mg/kg	heart	3 d	0.36	1.32E-02
Hyperlipidemia	DM_140	amlodipine	0.2 mg/kg	liver	0.25 d	0.36	9.07E-04
Hyperlipidemia	DM_1749	amoxapine	313 mg/kg	heart	5 d	0.36	1.29E-02
Hyperlipidemia	DM_1716	amiodarone	147 mg/kg	kidney	1 d	0.36	1.28E-02
Hyperlipidemia	DM_1185	prednisolone	184 mg/kg	liver	1 d	0.36	8.64E-04
Hyperlipidemia	DM_2685	dexamethasone	1 mg/kg	liver	3 d	0.36	1.23E-02
Hyperlipidemia	DM_4728	sildenafil	14.6 mg/kg	liver	3 d	0.36	1.29E-02
Hyperlipidemia	DM_612	erythromycin	1500 mg/kg	liver	3 d	0.36	7.45E-04
Hyperlipidemia	DM_208	benzethonium	138 mg/kg	heart	3 d	0.36	7.28E-04
Hyperlipidemia	DM_1903	benazepril	1750 mg/kg	kidney	1 d	0.36	1.14E-02
Hyperlipidemia	DM_3959	mitoxantrone	2 mg/kg	heart	3 d	0.36	1.14E-02
Hyperlipidemia	DM_1391	vinorelbine	1.5 mg/kg	heart	3 d	0.36	6.58E-04
Hyperlipidemia	DM_69	acetazolamide	250 mg/kg	liver	0.25 d	0.37	5.35E-04
Hyperlipidemia	DM_1014	mifepristone	3 mg/kg	liver	5 d	0.37	5.19E-04
Hyperlipidemia	DM_1857	azathioprine	54 mg/kg	liver	0.25 d	0.37	9.49E-03
Hyperlipidemia	DM_678	fludrocortisone	125 mg/kg	kidney	1 d	0.37	4.86E-04
Hyperlipidemia	DM_1144	pemoline	833 uM	primary rat hepatocytes	1 d	0.37	4.81E-04
Hyperlipidemia	DM_3170	fluocinolone	2.5 mg/kg	liver	3 d	0.37	8.98E-03
Hyperlipidemia	DM_3968	modafinil	17.5 mg/kg	liver	3 d	0.37	8.97E-03
Hyperlipidemia	DM_383	chlorpromazine	43.9 uM	primary rat hepatocytes	1 d	0.37	4.53E-04
Hyperlipidemia	DM_3982	moxonidine	17 mg/kg	heart	1 d	0.38	8.39E-03
Hyperlipidemia	DM_458	cromolyn	1500 mg/kg	kidney	5 d	0.38	4.09E-04
Hyperlipidemia	DM_4667	rosiglitazone	1800 mg/kg	thigh muscle	3 d	0.38	7.96E-03

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_1171	phenytoin	572 mg/kg	thigh muscle	5 d	0.38	3.81E-04
Hyperlipidemia	DM_3386	hydroxyurea	400 mg/kg	liver	5 d	0.38	7.90E-03
Hyperlipidemia	DM_2842	doxepin	147 mg/kg	heart	5 d	0.38	7.76E-03
Hyperlipidemia	DM_588	enrofloxacin	2000 mg/kg	kidney	3 d	0.38	3.67E-04
Hyperlipidemia	DM_4805	spironolactone	300 mg/kg	heart	5 d	0.38	7.65E-03
Hyperlipidemia	DM_2694	dexfenfluramine	29 mg/kg	kidney	1 d	0.38	7.58E-03
Hyperlipidemia	DM_3899	metronidazole	1500 mg/kg	liver	3 d	0.38	7.46E-03
Hyperlipidemia	DM_2267	celecoxib	400 mg/kg	heart	3 d	0.38	7.43E-03
Hyperlipidemia	DM_1001	miconazole	920 mg/kg	liver	0.25 d	0.38	3.29E-04
Hyperlipidemia	DM_2830	doxapram	20 mg/kg	heart	1 d	0.38	7.15E-03
Hyperlipidemia	DM_4591	ranitidine	1500 mg/kg	heart	3 d	0.38	7.07E-03
Hyperlipidemia	DM_195	azathioprine	54 mg/kg	liver	5 d	0.39	2.96E-04
Hyperlipidemia	DM_3194	fluphenazine	22 mg/kg	liver	5 d	0.39	6.68E-03
Hyperlipidemia	DM_3049	etoposide	100 mg/kg	liver	0.25 d	0.39	6.10E-03
Hyperlipidemia	DM_752	gemfibrozil	700 mg/kg	liver	0.25 d	0.39	2.35E-04
Hyperlipidemia	DM_438	clotrimazole	52 mg/kg	liver	0.25 d	0.39	2.34E-04
Hyperlipidemia	DM_2454	clarithromycin	56 mg/kg	liver	0.25 d	0.39	5.96E-03
Hyperlipidemia	DM_4535	propranolol	175 mg/kg	heart	1 d	0.39	5.82E-03
Hyperlipidemia	DM_441	clotrimazole	52 mg/kg	liver	3 d	0.40	1.96E-04
Hyperlipidemia	DM_582	enoxacin	100 mg/kg	liver	1 d	0.40	1.73E-04
Hyperlipidemia	DM_278	buspirone	196 mg/kg	heart	3 d	0.40	1.45E-04
Hyperlipidemia	DM_2365	chlorzoxazone	763 mg/kg	liver	3 d	0.41	3.93E-03
Hyperlipidemia	DM_831	imatinib	15 mg/kg	liver	5 d	0.41	1.08E-04
Hyperlipidemia	DM_2236	carvedilol	2000 mg/kg	heart	3 d	0.41	3.71E-03
Hyperlipidemia	DM_4543	propylene glycol	2000 mg/kg	liver	5 d	0.41	3.62E-03
Hyperlipidemia	DM_1851	azathioprine	0.5 uM	primary rat hepatocytes	1 d	0.42	3.30E-03
Hyperlipidemia	DM_860	itraconazole	10 uM	primary rat hepatocytes	1 d	0.42	8.03E-05
Hyperlipidemia	DM_361	chlorambucil	1540 uM	primary rat hepatocytes	1 d	0.42	6.91E-05
Hyperlipidemia	DM_283	busulfan	250 uM	primary rat hepatocytes	1 d	0.42	6.79E-05
Hyperlipidemia	DM_158	aspirin	167 mg/kg	liver	1 d	0.42	6.04E-05
Hyperlipidemia	DM_354	cerivastatin	7 mg/kg	kidney	1 d	0.43	4.28E-05
Hyperlipidemia	DM_3530	isotretinoin	13 mg/kg	liver	3 d	0.43	2.06E-03

phenotype	signatureid	drug_name	concentration	tissue	time	concordance	pvalue
Hyperlipidemia	DM_437	clotrimazole	178 mg/kg	liver	0.25 d	0.44	2.62E-05
Hyperlipidemia	DM_2506	clonidine	6 mg/kg	heart	3 d	0.45	1.52E-03
Hyperlipidemia	DM_4045	nevirapine	250 uM	primary rat hepatocytes	1 d	0.45	1.50E-03
Hyperlipidemia	DM_2796	digoxin	11 mg/kg	heart	5 d	0.45	1.42E-03
Hyperlipidemia	DM_233	betamethasone	79 mg/kg	liver	1 d	0.46	1.06E-05
Hyperlipidemia	DM_829	imatinib	15 mg/kg	heart	5 d	0.46	1.06E-05
Hyperlipidemia	DM_328	carmustine	4 mg/kg	liver	0.25 d	0.46	9.68E-06
Hyperlipidemia	DM_3970	modafinil	17.5 mg/kg	liver	5 d	0.46	9.21E-04
Hyperlipidemia	DM_3784	maprotiline	380 mg/kg	heart	1 d	0.46	8.89E-04
Hyperlipidemia	DM_3182	fluoxetine	52 mg/kg	heart	5 d	0.47	8.35E-04
Hyperlipidemia	DM_1548	aceclofenac	9 mg/kg	kidney	3 d	0.47	8.18E-04
Hyperlipidemia	DM_2187	carboplatin	6 mg/kg	liver	1 d	0.47	6.81E-04
Hyperlipidemia	DM_936	loratadine	2000 mg/kg	heart	1 d	0.48	3.13E-06
Hyperlipidemia	DM_5161	valproate	1500 mg/kg	heart	1 d	0.49	4.31E-04
Hyperlipidemia	DM_336	carvedilol	2000 mg/kg	liver	5 d	0.52	3.55E-07
Hyperlipidemia	DM_282	busulfan	125 uM	primary rat hepatocytes	1 d	0.53	1.75E-07

Supplementary Table 4: iLINCS drug repurposing candidate list for hypertension.

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	vorinostat	-	22.25	5.15E-110
Hypertension	niclosamide	-	14.00	7.35E-45
Hypertension	valdecoxib	-	8.17	1.53E-16
Hypertension	crizotinib	-	7.98	7.53E-16
Hypertension	guanethidine	+	7.60	1.48E-14
Hypertension	perphenazine	-	7.39	7.41E-14
Hypertension	thiothixene	-	7.23	2.41E-13
Hypertension	troglitazone	-	7.18	3.49E-13
Hypertension	sirolimus	-	6.95	1.87E-12
Hypertension	thioridazine	-	6.94	1.94E-12
Hypertension	terfenadine	-	6.93	2.09E-12
Hypertension	azacitidine	+	6.70	1.04E-11
Hypertension	trifluprомazine	-	6.68	1.22E-11
Hypertension	doxorubicin	-	6.65	1.48E-11
Hypertension	imatinib	-	6.61	1.97E-11
Hypertension	chlorpromazine	-	6.56	2.75E-11
Hypertension	captopril	-	6.31	1.38E-10
Hypertension	olaparib	-	5.98	1.10E-09
Hypertension	ticagrelor	+	5.98	1.14E-09
Hypertension	anagrelide	+	5.87	2.12E-09
Hypertension	cabozantinib	-	5.80	3.33E-09
Hypertension	edaravone	+	5.75	4.45E-09
Hypertension	ouabain	-	5.74	4.67E-09
Hypertension	idelalisib	-	5.73	4.96E-09
Hypertension	levonorgestrel	+	5.55	1.45E-08
Hypertension	oxymetholone	-	5.50	1.91E-08
Hypertension	valrubicin	-	5.48	2.14E-08
Hypertension	cobimetinib	+	5.42	3.01E-08
Hypertension	benzoate	-	5.31	5.42E-08
Hypertension	thiostrepton	-	5.26	7.18E-08
Hypertension	griseofulvin	+	5.25	7.42E-08

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	rivaroxaban	-	5.24	8.14E-08
Hypertension	sirolimus	-	5.21	9.39E-08
Hypertension	bosentan	+	5.15	1.27E-07
Hypertension	amiodarone	-	5.12	1.55E-07
Hypertension	fluphenazine	-	5.09	1.78E-07
Hypertension	ruxolitinib	-	5.03	2.51E-07
Hypertension	procyclidine	-	5.01	2.77E-07
Hypertension	perindopril	+	5.00	2.83E-07
Hypertension	lacosamide	+	4.94	3.88E-07
Hypertension	carbinoxamine	-	4.85	6.14E-07
Hypertension	tacalcitol	+	4.84	6.37E-07
Hypertension	fluoxetine	-	4.81	7.72E-07
Hypertension	fluorouracil	-	4.81	7.72E-07
Hypertension	halofantrine	+	4.80	7.83E-07
Hypertension	gemcitabine	-	4.79	8.27E-07
Hypertension	caffeine	-	4.71	1.21E-06
Hypertension	amisulpride	+	4.69	1.36E-06
Hypertension	rosiglitazone	-	4.67	1.54E-06
Hypertension	pioglitazone	-	4.65	1.69E-06
Hypertension	clonazepam	+	4.62	1.93E-06
Hypertension	troleandomycin	-	4.58	2.31E-06
Hypertension	iloprost	+	4.57	2.43E-06
Hypertension	escitalopram	-	4.53	2.91E-06
Hypertension	terbinafine	+	4.48	3.73E-06
Hypertension	melatonin	-	4.47	3.85E-06
Hypertension	gemifloxacin	+	4.47	3.96E-06
Hypertension	haloperidol	-	4.46	4.02E-06
Hypertension	nandrolone	+	4.45	4.24E-06
Hypertension	trovafloxacin	+	4.41	5.20E-06
Hypertension	tiaprofenate	+	4.41	5.22E-06
Hypertension	beclomethasone	-	4.40	5.50E-06
Hypertension	digoxin	-	4.38	5.97E-06

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	tranylcypromine	-	4.37	6.29E-06
Hypertension	ethotoin	-	4.37	6.32E-06
Hypertension	megestrol	-	4.36	6.47E-06
Hypertension	nadolol	-	4.35	6.68E-06
Hypertension	nilotinib	-	4.33	7.63E-06
Hypertension	iloprost	-	4.32	7.79E-06
Hypertension	diethylstilbestrol	-	4.29	8.95E-06
Hypertension	azithromycin	-	4.25	1.05E-05
Hypertension	lomefloxacin	+	4.21	1.25E-05
Hypertension	estradiol	-	4.21	1.26E-05
Hypertension	pilocarpine	+	4.13	1.81E-05
Hypertension	primidone	+	4.10	2.10E-05
Hypertension	nelarabine	+	4.08	2.27E-05
Hypertension	minoxidil	+	4.01	3.09E-05
Hypertension	rimantadine	+	3.97	3.57E-05
Hypertension	midodrine	-	3.92	4.45E-05
Hypertension	dihydroergotamine	-	3.90	4.76E-05
Hypertension	hydrocortisone	-	3.88	5.21E-05
Hypertension	meclizine	+	3.84	6.24E-05
Hypertension	homatropine	+	3.84	6.27E-05
Hypertension	budesonide	-	3.83	6.33E-05
Hypertension	atorvastatin	-	3.82	6.71E-05
Hypertension	osimertinib	-	3.81	7.08E-05
Hypertension	loperamide	-	3.80	7.09E-05
Hypertension	calcitriol	+	3.80	7.25E-05
Hypertension	mifepristone	-	3.80	7.32E-05
Hypertension	loteprednol etabonate	-	3.79	7.57E-05
Hypertension	linezolid	-	3.78	7.74E-05
Hypertension	medrysone	-	3.78	7.87E-05
Hypertension	lisinopril	+	3.76	8.33E-05
Hypertension	gefitinib	-	3.76	8.45E-05
Hypertension	sertraline	-	3.75	8.80E-05

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	isoniazid	+	3.71	1.05E-04
Hypertension	everolimus	-	3.70	1.06E-04
Hypertension	mianserin	-	3.70	1.08E-04
Hypertension	tacrolimus	-	3.70	1.09E-04
Hypertension	valproate	+	3.69	1.11E-04
Hypertension	pyridoxine	-	3.67	1.21E-04
Hypertension	erlotinib	+	3.66	1.26E-04
Hypertension	ifosfamide	+	3.66	1.26E-04
Hypertension	norgestrel	+	3.66	1.26E-04
Hypertension	miglitol	-	3.64	1.36E-04
Hypertension	spironolactone	-	3.63	1.41E-04
Hypertension	rasagiline	+	3.62	1.45E-04
Hypertension	fenoldopam	+	3.59	1.62E-04
Hypertension	dichlorphenamide	+	3.59	1.64E-04
Hypertension	pimozide	-	3.58	1.74E-04
Hypertension	rizatriptan	+	3.56	1.83E-04
Hypertension	iopanoic acid	-	3.55	1.91E-04
Hypertension	selamectin	-	3.52	2.16E-04
Hypertension	toremifene	-	3.48	2.55E-04
Hypertension	paricalcitol	+	3.47	2.65E-04
Hypertension	ambrisentan	+	3.46	2.69E-04
Hypertension	dorzolamide	+	3.45	2.78E-04
Hypertension	cytarabine	-	3.45	2.80E-04
Hypertension	telbivudine	+	3.40	3.36E-04
Hypertension	thioguanine	+	3.38	3.67E-04
Hypertension	amoxapine	-	3.37	3.76E-04
Hypertension	febuxostat	-	3.35	4.04E-04
Hypertension	acyclovir	+	3.34	4.17E-04
Hypertension	probenecid	-	3.31	4.64E-04
Hypertension	methylprednisolone	-	3.30	4.82E-04
Hypertension	lovastatin	-	3.30	4.84E-04
Hypertension	palbociclib	-	3.29	5.00E-04

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	thiotepa	-	3.29	5.02E-04
Hypertension	perampanel	+	3.27	5.41E-04
Hypertension	dactinomycin	-	3.26	5.63E-04
Hypertension	mecamylamine	+	3.25	5.82E-04
Hypertension	thalidomide	-	3.24	5.90E-04
Hypertension	zolpidem	+	3.22	6.34E-04
Hypertension	sirolimus	-	3.21	6.57E-04
Hypertension	ipratropium	-	3.20	6.89E-04
Hypertension	loracarbef	-	3.18	7.28E-04
Hypertension	enoxacin	-	3.18	7.33E-04
Hypertension	tegaserod	-	3.17	7.52E-04
Hypertension	cyclobenzaprine	-	3.17	7.53E-04
Hypertension	protriptyline	-	3.17	7.71E-04
Hypertension	piperidolate	-	3.15	8.08E-04
Hypertension	sulpiride	+	3.15	8.20E-04
Hypertension	inositol	-	3.14	8.37E-04
Hypertension	topiramate	+	3.14	8.53E-04
Hypertension	trifluoperazine	-	3.14	8.58E-04
Hypertension	oxaprozin	-	3.13	8.63E-04
Hypertension	piretanide	-	3.12	9.07E-04
Hypertension	nitazoxanide	-	3.12	9.07E-04
Hypertension	naftifine	+	3.12	9.12E-04
Hypertension	salmeterol	-	3.11	9.27E-04
Hypertension	vilazodone	+	3.11	9.27E-04
Hypertension	olmesartan	+	3.11	9.39E-04
Hypertension	amlodipine	-	3.10	9.53E-04
Hypertension	dexfenfluramine	+	3.06	1.10E-03
Hypertension	phentermine	-	3.05	1.14E-03
Hypertension	bicalutamide	+	3.05	1.14E-03
Hypertension	famciclovir	+	3.05	1.15E-03
Hypertension	simvastatin	-	3.05	1.16E-03
Hypertension	fluticasone	+	3.03	1.21E-03

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	rilpivirine	+	3.03	1.22E-03
Hypertension	alectinib	+	3.03	1.24E-03
Hypertension	zeranol	-	3.02	1.25E-03
Hypertension	ipratropium	-	3.02	1.25E-03
Hypertension	sunitinib	+	3.01	1.30E-03
Hypertension	paroxetine	-	3.00	1.34E-03
Hypertension	celecoxib	-	3.00	1.35E-03
Hypertension	dipyridamole	-	3.00	1.35E-03
Hypertension	sulfacetamide	-	3.00	1.36E-03
Hypertension	ziprasidone	+	3.00	1.37E-03
Hypertension	efavirenz	-	2.99	1.39E-03
Hypertension	lapatinib	-	2.98	1.43E-03
Hypertension	desloratadine	-	2.96	1.51E-03
Hypertension	penicillin v	+	2.96	1.55E-03
Hypertension	sorafenib	-	2.94	1.66E-03
Hypertension	tacrine	+	2.93	1.70E-03
Hypertension	dichloroacetate	+	2.91	1.80E-03
Hypertension	altrenogest	+	2.91	1.80E-03
Hypertension	levothyroxine	+	2.91	1.82E-03
Hypertension	cefuroxime	+	2.90	1.88E-03
Hypertension	bendroflumethiazide	-	2.89	1.90E-03
Hypertension	medetomidine	-	2.89	1.90E-03
Hypertension	nimodipine	-	2.88	1.99E-03
Hypertension	sildenafil	+	2.88	2.00E-03
Hypertension	clomipramine	-	2.86	2.09E-03
Hypertension	raloxifene	-	2.86	2.10E-03
Hypertension	benzethonium	-	2.86	2.12E-03
Hypertension	acamprosate	+	2.85	2.17E-03
Hypertension	diphenidol	-	2.85	2.18E-03
Hypertension	fluvastatin	-	2.85	2.20E-03
Hypertension	propantheline	-	2.84	2.23E-03
Hypertension	amoxicillin	+	2.84	2.28E-03

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	fluocinonide	-	2.83	2.32E-03
Hypertension	regorafenib	-	2.83	2.32E-03
Hypertension	tenoxicam	+	2.83	2.34E-03
Hypertension	danazol	-	2.83	2.34E-03
Hypertension	phenazopyridine	+	2.82	2.37E-03
Hypertension	clenbuterol	+	2.81	2.46E-03
Hypertension	atomoxetine	+	2.81	2.48E-03
Hypertension	altrenogest	+	2.80	2.54E-03
Hypertension	fexofenadine	-	2.79	2.62E-03
Hypertension	aminolevulinic acid	+	2.78	2.70E-03
Hypertension	methenamine	-	2.77	2.82E-03
Hypertension	nabumetone	-	2.76	2.87E-03
Hypertension	sorbitol	+	2.76	2.90E-03
Hypertension	duloxetine	-	2.75	2.94E-03
Hypertension	trametinib	+	2.75	3.01E-03
Hypertension	pralatrexate	+	2.75	3.02E-03
Hypertension	chlorotrianisene	-	2.74	3.04E-03
Hypertension	pentobarbital	+	2.74	3.10E-03
Hypertension	tretinoin	-	2.73	3.12E-03
Hypertension	aceclofenac	-	2.73	3.18E-03
Hypertension	praziquantel	+	2.71	3.35E-03
Hypertension	amitriptyline	-	2.71	3.36E-03
Hypertension	ethacrynate	-	2.69	3.52E-03
Hypertension	vecuronium	+	2.69	3.53E-03
Hypertension	carvedilol	-	2.69	3.59E-03
Hypertension	papaverine	-	2.68	3.64E-03
Hypertension	propafenone	-	2.68	3.64E-03
Hypertension	dexamethasone	-	2.68	3.66E-03
Hypertension	oxandrolone	-	2.68	3.67E-03
Hypertension	levofloxacin	-	2.68	3.69E-03
Hypertension	indomethacin	-	2.67	3.79E-03
Hypertension	ribavirin	-	2.66	3.85E-03

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	phensuximide	-	2.66	3.87E-03
Hypertension	triamcinolone	-	2.65	3.97E-03
Hypertension	epirubicin	-	2.65	4.00E-03
Hypertension	meptazinol	+	2.64	4.17E-03
Hypertension	guanadrel	+	2.64	4.18E-03
Hypertension	tamoxifen	+	2.63	4.29E-03
Hypertension	sunitinib	-	2.62	4.44E-03
Hypertension	novobiocin	+	2.61	4.51E-03
Hypertension	glyburide	-	2.61	4.52E-03
Hypertension	apazone	-	2.60	4.67E-03
Hypertension	metronidazole	+	2.60	4.72E-03
Hypertension	cerivastatin	-	2.58	4.89E-03
Hypertension	raltegravir	+	2.58	4.95E-03
Hypertension	brompheniramine	-	2.57	5.04E-03
Hypertension	tolnaftate	+	2.56	5.17E-03
Hypertension	paroxetine	-	2.56	5.20E-03
Hypertension	ropinirole	+	2.56	5.22E-03
Hypertension	tigecycline	-	2.56	5.26E-03
Hypertension	thonzonium	-	2.56	5.29E-03
Hypertension	amcinonide	-	2.56	5.30E-03
Hypertension	phenylephrine	-	2.55	5.32E-03
Hypertension	nifedipine	+	2.55	5.34E-03
Hypertension	pyrimethamine	-	2.55	5.44E-03
Hypertension	chlorambucil	-	2.54	5.52E-03
Hypertension	phenytoin	-	2.54	5.59E-03
Hypertension	griseofulvin	-	2.53	5.73E-03
Hypertension	benzthiazide	+	2.53	5.76E-03
Hypertension	oxprenolol	-	2.52	5.89E-03
Hypertension	benazepril	-	2.51	5.96E-03
Hypertension	liothyronine	-	2.51	6.00E-03
Hypertension	bisoprolol	+	2.49	6.31E-03
Hypertension	fomepizole	+	2.49	6.39E-03

phenotype	drug_name	correlation	zscore	pvalue
Hypertension	dabrafenib	+	2.48	6.64E-03
Hypertension	diazoxide	-	2.47	6.70E-03
Hypertension	ethosuximide	+	2.47	6.75E-03
Hypertension	fenbufen	-	2.47	6.83E-03
Hypertension	dinoprost	-	2.46	6.91E-03
Hypertension	mibefradil	-	2.46	6.94E-03
Hypertension	biperiden	-	2.45	7.15E-03
Hypertension	fenofibrate	+	2.44	7.41E-03
Hypertension	esmolol	-	2.43	7.65E-03
Hypertension	artemether	-	2.42	7.66E-03
Hypertension	ethamivan	-	2.42	7.78E-03

Supplementary Table 5: Demographic Characteristics of clinical validation study cohorts.

Source	Phenotype	Drugs	N	Female - no. (%)	White - no. (%)	Age - yr	Observation Period Length - d	Treatment Period Length - d
Vanderbilt	Hyperlipidemia	Simvastatin	6,305	3201 (50.8)	5022 (79.7)	56.5 (12.8)	710 (62.6)	345 (62.6)
		Alendronate	620	496 (80.0)	528 (85.2)	60.4 (13.7)	710 (57.6)	345 (57.6)
		Megestrol	36	17 (47.2)	20 (55.6)	58.3 (13.0)	648 (108)	283 (108)
		Quinapril	170	76 (44.7)	129 (75.9%)	55.9 (13.2)	699 (67.6)	334 (67.6)
		Lisinopril	2,447	1216 (49.7)	1820 (74.4)	52.1 (14.1)	704 (69.8)	339 (69.8)
	Hypertension	Losartan	3,759	2065 (54.9)	3221 (85.7)	55.7 (13.4)	648 (118)	283 (118)
		Tacrolimus	527	324 (61.5)	449 (85.2)	43.5 (15.3)	638 (120)	273 (120)
		Budesonide	4,204	2825 (67.2)	3693 (87.8)	42.7 (15.3)	637 (121)	272 (121)
		Digoxin	220	126 (57.3)	200 (90.9)	61.0 (16.8)	619 (133)	254 (133)
		Pioglitazone	581	277 (47.7)	455 (78.3)	50.8 (13.3)	620 (128)	255 (128)
		Fluoxetine	7,724	6318 (81.8)	6765 (87.6)	39.1 (14.3)	632 (122)	267 (123)
		Haloperidol	441	244 (55.3)	328 (74.4)	40.6 (15.0)	609 (130)	243 (130)
		Sertraline	14,168	10872 (76.7)	12196 (86.1)	38.8 (14.8)	629 (124)	264 (124)
		Estradiol	9,794	9772 (99.8)	8765 (89.5)	50.0 (12.4)	659 (113)	293 (113)
		Escitalopram	12,535	9464 (75.5)	11094 (88.5)	39.9 (14.7)	631 (123)	265 (123)

Source	Phenotype	Drugs	N	Female - no. (%)	White - no. (%)	Age - yr	Observation Period Length - d	Treatment Period Length - d
		Atorvastatin	8,027	4179 (52.1)	7057 (87.9)	54.6 (12.3)	662 (110)	296 (110)
All of Us	Hyperlipidemia	Simvastatin	1,526	727 (47.6)	990 (64.9)	56.6 (9.60)	717 (50.8)	351 (50.8)
		Lisinopril	899	422 (46.9)	499 (55.5)	54.1 (11.3)	712 (59.1)	347 (59.1)
	Hypertension	Losartan	328	189 (57.6)	216 (65.9)	60.0 (11.5)	686 (94.8)	321 (94.8)
		Fluoxetine	839	640 (76.3)	575 (68.5)	42.1 (13.2)	670 (105)	305 (105)
		Haloperidol	203	156 (76.8)	143 (70.4)	48.7 (14.7)	665 (101)	299 (101)
		Sertraline	1,094	841 (76.9)	739 (67.6)	42.6 (14.3)	665 (109)	300 (109)
		Estradiol	838	831 (99.2)	692 (82.6)	51.6 (13.1)	690 (87.6)	325 (87.5)
		Escitalopram	709	536 (75.6)	481 (67.8)	43.9 (14.2)	653 (115)	287 (115)
		Atorvastatin	1,842	1056 (57.3)	1153 (62.6)	54.8 (13.1)	696 (82.0)	331 (82.0)

CHAPTER 5

SUMMARY

Summary of findings

In this dissertation, I developed and evaluated two novel methods to detect DDI signals and identify drug repurposing candidates. In Chapter 3, I demonstrated that information from the allergy list can be used to detect DDI signals, using DDIWAS. Because the allergy list is a common module in the EHR, there is a high potential that researchers at other institutions will be interested in applying DDIWAS to identify DDI signals in their database. In the Biomedical Informatics literature, the allergy list has been relatively underused as a source of information for pharmacovigilance. DDIWAS was applied to identify DDIs for two common drugs, simvastatin and amlodipine. For both drugs, DDIWAS replicated known DDIs and identified potential novel DDIs. For Chapter 3's project, I developed software tools to implement these two methods and released them as R packages in publicly available GitHub repositories. The DDIWAS R package can be found at "https://github.com/pwatrick/ddiwas".

In Chapter 4, I demonstrated a method that integrates multiple sources of publicly available biomedical data to identify and validate drug repurposing candidates. Because of its use of publicly available data sources and open-source software tools, this method can be considered portable. First, this method uses S-PrediXcan (an open-source software tool) and publicly available GWAS summary statistics to estimate human phenotype transcriptomic signatures. Second, this method searches in iLINCS (a publicly available database) for drugs with perturbation signatures that reverse the phenotype transcriptomic signatures estimated using S-PrediXcan. Third, this method validates drug repurposing candidates using clinical data stored in EHR databases organized using the OHDSI/OMOP CDM. I was able to easily and quickly replicate the clinical validation study in the *All of Us* dataset with minor code modifications, because the dataset and the VUMC SD both use the OHDSI/OMOP CDM format. The *All of Us* program is in its initial phase of recruiting when these experiments were conducted, but I successfully replicated the treatment effects of several drug repurposing candidates with similar effect sizes. As of March 2021, the NIH *All of Us* Research Program database is available to researchers from 183 institutions in the US. They can use an R package that I have developed,

which can be found at "https://github.com/pwatrick/DrugRepurposingToolKit". The NIH *All of Us* Research Program workbench was in beta when these experiments were conducted, and only demonstration projects focused on replicating known findings[145] and predictive models[167] have been published. To my knowledge, this is the first study introducing a novel method for biomedical discovery, performed in the NIH *All of Us* Research Program database.

Limitations

The two methods have limitations. The main limitation for DDIWAS is that we did not demonstrate portability of the method in an external database. The challenge to replicating DDIWAS in an external database is that there is not a consensus approach for representing allergy list information in the most recent version of the OHDSI/OMOP CDM (version 6.0; https://ohdsi.github.io/CommonDataModel/cdm60.html). In the VUMC SD, allergy list information is found in the `NOTE` table by string matching `ALLERGY` in the `NOTE_SOURCE_VALUE` column. However, allergy list information may not be found using a similar query in an external OHDSI/OMOP CDM database, because `NOTE_SOURCE_VALUE` is not a required column. Discussions related to representing allergy list information is still ongoing in the OHDSI community.[168]

For the drug repurposing method described in Chapter 4, the main limitation is that I only applied the pipeline to two phenotypes, hyperlipidemia and hypertension. These two phenotypes were chosen because the goal was to demonstrate proof-of-concept, of using longitudinal clinical data in the EHR to clinically validate drug repurposing candidates. To validate drug repurposing candidates in the EHR required large patient cohorts and common biomarker measurements. LDL-C and SBP measurements are relatively common because they are recorded as part of preventive care and the clinical intake process, respectively. A large starting cohort with common biomarkers maximized the statistical power to measure the treatment effects of drug repurposing candidates.

Future Directions

There are several potential avenues to build upon these methods. The performance of DDIWAS to detect DDI signals can potentially be improved by incorporating drug doses in the model. Performance can potentially improve because higher drug doses increase the likelihood of known DDIs.[113] Historically, it has been difficult to obtain accurate drug dose information, but

novel methods have been recently been developed to extract longitudinal drug doses accurately from clinical notes.[169]

Approaches similar to DDIWAS, using information from the allergy list, can potentially be useful for identifying not only DDIs but also ADRs attributed to new classes of therapeutics, for which the pharmacological properties we are just beginning to understand. Many DDIs and ADRs for small-molecule drugs can be investigated using *in vitro* and *in vivo* studies, as suggested by the US FDA.[69] But, these tests are only possible due to decades of research that have improved our understanding of the pharmacokinetic and pharmacodynamic properties of prescribed small-molecule drugs.[68] With each passing year, new classes of "biologic" treatments, like cell-,[170] antibody-,[171] and gene-based therapeutics are approved, therapeutics for which we are just starting to understand the causal biological mechanisms.[172] Information from the allergy list could potentially be used to detect ADR and DDI cases involving these new therapeutics, and results from these studies could lead to the design of experiments to elucidate the underlying mechanisms and eventual development of pre-clinical tests that can be used to assess risk of DDIs and ADRs prior to drug approval.

Another potential question that may be worth exploring is to understand the signals captured by DDIWAS. For example, DDIWAS could be capturing latent DDI signals, potentially missed by existing methods in the literature. Drugs entered in the allergy could perhaps represent those drugs providers think are responsible for a patient's reported symptoms. This could explain the relatively high performance of DDIWAS of capturing known DDIs, as providers may be more likely to suspect certain well-known DDIs if they learned about them through training or were contained in the CDS knowledge base of their EHR system. On the other hand, DDIWAS could also be capturing DDIs across temporal space, in that the allergy list entries could be entered by different providers at different visits. Future work could explore this temporal aspect to better understand the latent signals captured by DDIWAS.

The modular nature of the drug repurposing method described in Chapter 4 provides several opportunities for future investigation. First, the estimation of phenotype transcriptomic signature can be optimized, as new tools are published, which improve upon S-PrediXcan's ability to infer gene expression. For instance, researchers recently found that detection of known gene expression-phenotype associations is improved by integrating gene expression information from multiple tissues.[173] Further, methods that incorporate epigenetic data have been shown to

improve the prediction accuracy of gene expression for tissues with small sample sizes (eg, brain).[174] Future research may consider experiments to see whether these methods lead to closer approximations of the “true” transcriptomic signature for phenotypes. Second, the large number of publicly available GWAS summary statistics datasets (869 and 7,221 unique human phenotypes available in the GWAS catalog[175] and UK Biobank,[149] respectively), provides researchers with the opportunity to apply this method to find and clinically validate drug repurposing candidates for human diseases, especially for those with few effective treatments. Third, future work may consider developing methods to better infer the molecular signature of phenotypes through the incorporation of new data types, like single-cell gene expression experiments.[176]

Fourth, databases with drug perturbation data that serve as improved proxies for human biological drug response can further improve this approach to identify drug repurposing candidates. The iLINCS database serves as a valuable resource for perturbation transcriptomic signatures, but the majority of data are derived from experiments using immortalized human cell lines and non-human primary tissues. If similar drug perturbation datasets are released from experiments using human organoids[177] or induced pluripotent stem cells (iPSCs),[178,179] which serve as closer proxies than cell lines and non-human tissues, the signature matching approach to identify drug repurposing candidates can potentially improve.

Fifth, the results from studies in Chapter 4, provide further support for the importance of investing in projects focused on developing digital biomarkers for diseases that do not have reliable biomarkers as proxies for disease prognosis, ie, data that represents whether a patients disease is worsening or improving. Such approaches using mobile devices for digital phenotyping is especially promising for diseases without objective lab measurements, like psychiatric disorders[180] and Parkinson’s disease.[181,182] The NIH *All of Us* database is already integrating clinical data outside of EHRs, from wearable platforms Fitbit and Apple HealthKit.[145] As large, integrated clinical datasets like the NIH *All of Us* recruit more patients, this approach to clinically validating drug candidates can improve.

Both approaches can improve as future tools are developed to better understand biological mechanisms affecting human health across space and time. There are now methods available to detect genome-wide gene expression in tissues,[183] and to track the dynamics of gene expression changes across time to understand cell lineage development.[184] The next phase

of research in translational biomedical informatics may focus on the incorporation of longitudinal information across an individual's life (re: pediatric health data),[9], the transformation unstructured EHR data (majority of EHR data), the integration EHR data with other data sources (genetic data, consumer mobile health data, environmental data) answer questions pertinent to human health, and inclusion of biomedical data from diverse populations.[185] Such work will allow us to better understand the mechanisms underlying genotype-phenotype relationships, with the ultimate goal of developing improved treatments for human diseases.

REFERENCES

1 Kulikowski CA, Shortliffe EH, Currie LM, *et al.* AMIA board white paper: Definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc* 2012;**19**:931–8. doi:10.1136/amiajnl-2012-001053

2 Johnson KB, Patel NR. Biomedical informatics and health information technology: A critical, pragmatic collaboration for clinical transformation. *J Gen Intern Med* Published Online First: April 2020. doi:10.1007/s11606-020-05833-y

3 Roden DM, Pulley JM, Basford MA, *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;**84**:362–9. doi:10.1038/clpt.2008.89

4 Ritchie MD, Denny JC, Crawford DC, *et al.* Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010;**86**:560–72. doi:10.1016/j.ajhg.2010.03.003

5 Denny JC, Ritchie MD, Basford MA, *et al.* PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010;**26**:1205–10. doi:10.1093/bioinformatics/btq126

6 Kho AN, Pacheco JA, Peissig PL, *et al.* Electronic medical records for genetic research: Results of the eMERGE consortium. *Sci Transl Med* 2011;**3**:79re1. doi:10.1126/scitranslmed.3001807

7 Denny JC, Crawford DC, Ritchie MD, *et al.* Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: Using electronic medical records for genome- and phenome-wide studies. *Am J Hum Genet* 2011;**89**:529–42. doi:10.1016/j.ajhg.2011.09.008

8 Sudlow C, Gallacher J, Allen N, *et al.* UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;**12**:e1001779. doi:10.1371/journal.pmed.1001779

9 All of Us Research Program Investigators, Denny JC, Rutter JL, *et al.* The ‘all of us’ research program. *N Engl J Med* 2019;**381**:668–76. doi:10.1056/NEJMsr1809937

10 Stang PE, Ryan PB, Racoosin JA, *et al.* Advancing the science for active surveillance: Rationale and design for the observational medical outcomes partnership. *Ann Intern Med* 2010;**153**:600–6. doi:10.7326/0003-4819-153-9-201011020-00010

11 Overhage JM, Ryan PB, Reich CG, *et al.* Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;**19**:54–60. doi:10.1136/amiajnl-2011-000376

12 Iyer SV, Harpaz R, LePendu P, *et al.* Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc* 2014;**21**:353–62. doi:10.1136/amiajnl-2013-001612

13 Xu H, Aldrich MC, Chen Q, *et al.* Validating drug repurposing signals using electronic health records: A case study of metformin associated with reduced cancer mortality. *J Am Med Inform Assoc* 2015;**22**:179–91. doi:10.1136/amiajnl-2014-002649

14 Nelson SJ, Zeng K, Kilbourne J, *et al.* Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Inform Assoc* 2011;**18**:441–8. doi:10.1136/amiajnl-2011-000116

15 WHOCC. Purpose of the ATC/DDD system. https://www.whocc.no/atc_ddd_methodology/purpose_of_the_atc_ddd_system/

16 Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82. doi:10.1093/nar/gkx1037

17 Steindel SJ. International classification of diseases, 10th edition, clinical modification and procedure coding system: Descriptive overview of the next generation HIPAA code sets. *J Am Med Inform Assoc* 2010;**17**:274–82. doi:10.1136/jamia.2009.001230

18 Topaz M, Shafran-Topaz L, Bowles KH. ICD-9 to ICD-10: Evolution, revolution, and current debates in the united states. *Perspect Health Inf Manag* 2013;**10**:1d.https://www.ncbi.nlm.nih.gov/pubmed/23805064

19 Wei W-Q, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Med* 2015;**7**:41. doi:10.1186/s13073-015-0166-y

20 Wei W-Q, Teixeira PL, Mo H, *et al.* Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *J Am Med Inform Assoc* 2016;**23**:e20–7. doi:10.1093/jamia/ocv130

21 Hripcsak G, Levine ME, Shang N, *et al.* Effect of vocabulary mapping for conditions on phenotype cohorts. *J Am Med Inform Assoc* 2018;**25**:1618–25. doi:10.1093/jamia/ocy124

22 Wu P, Gifford A, Meng X, *et al.* Mapping ICD-10 and ICD-10-CM codes to phecodes: Workflow development and initial evaluation. *JMIR Med Inform* 2019;**7**:e14325. doi:10.2196/14325

23 Wei W-Q, Cronin RM, Xu H, *et al.* Development and evaluation of an ensemble resource linking medications to their indications. *J Am Med Inform Assoc* 2013;**20**:954–61. doi:10.1136/amiajnl-2012-001431

24 Regenstrief Institute. Logical observation identifiers names and codes. http://archive.today/2021.02.07-221944/https://loinc.org/get-started/

25 Abhyankar S, Demner-Fushman D, McDonald CJ. Standardizing clinical laboratory data for secondary use. *J Biomed Inform* 2012;**45**:642–50. doi:10.1016/j.jbi.2012.04.012

26 Bodenreider O. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Res* 2004;**32**:D267–70. doi:10.1093/nar/gkh061

27 Humphreys BL, Del Fiol G, Xu H. The UMLS knowledge sources at 30: Indispensable to current research and applications in biomedical informatics. *J Am Med Inform Assoc* 2020;**27**:1499–501. doi:10.1093/jamia/ocaa208

28 Hripcsak G, Duke JD, Shah NH, *et al.* Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers. *Stud Health Technol Inform* 2015;**216**:574–8.https://www.ncbi.nlm.nih.gov/pubmed/26262116

29 Hripcsak G, Ryan PB, Duke JD, *et al.* Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci U S A* 2016;**113**:7329–36. doi:10.1073/pnas.1510502113

30 Vashisht R, Jung K, Schuler A, *et al.* Association of hemoglobin A1c levels with use of sulfonylureas, dipeptidyl peptidase 4 inhibitors, and thiazolidinediones in patients with type 2 diabetes treated with metformin: Analysis from the observational health data sciences and informatics initiative. *JAMA Netw Open* 2018;**1**:e181755. doi:10.1001/jamanetworkopen.2018.1755

31 Rosenbloom ST, Carroll RJ, Warner JL, *et al.* Representing knowledge consistently across health systems. *Yearb Med Inform* 2017;**26**:139–47. doi:10.15265/IY-2017-018

32 Jha AK, DesRoches CM, Campbell EG, *et al.* Use of electronic health records in US hospitals. *N Engl J Med* 2009;**360**:1628–38.https://www.nejm.org/doi/full/10.1056/nejmsa0900592

33 Gold M, McLAUGHLIN C. Assessing HITECH implementation and lessons: 5 years later. *Milbank Q* 2016;**94**:654–87. doi:10.1111/1468-0009.12214

34 Adler-Milstein J, Jha AK. HITECH act drove large gains in hospital electronic health record adoption. *Health Aff* 2017;**36**:1416–22. doi:10.1377/hlthaff.2016.1651

35 Office of the National Coordinator for Health Information Technology. Percent of hospitals, by type, that possess certified health IT. 2018.https://dashboard.healthit.gov/quickstats/pages/certified-electronic-health-record-technology-in-hospitals.php

36 Sheikh A, Cornford T, Barber N, *et al.* Implementation and adoption of nationwide electronic health records in secondary care in england: Final qualitative results from prospective national evaluation in ‘early adopter’ hospitals. *BMJ* 2011;**343**.https://www.bmj.com/content/343/bmj.d6054

37 Office of the National Coordinator for Health Information Technology. ’Non-federal acute care hospital electronic health record adoption,’ health IT Quick-Stat #47. 2017.dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php.

38 Unlu G, Qi X, Gamazon ER, *et al.* Phenome-based approach identifies RIC1-linked mendelian syndrome through zebrafish models, biobank associations and clinical studies. *Nat Med* 2020;**26**:98–109. doi:10.1038/s41591-019-0705-y

39 Rogers JR, Lee J, Zhou Z, *et al.* Contemporary use of real-world data for clinical trial conduct in the united states: A scoping review. *J Am Med Inform Assoc* Published Online First: November 2020. doi:10.1093/jamia/ocaa224

40 Sboner A, Mu XJ, Greenbaum D, *et al.* The real cost of sequencing: Higher than you think! *Genome Biol* 2011;**12**:125. doi:10.1186/gb-2011-12-8-125

41 Muir P, Li S, Lou S, *et al.* The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol* 2016;**17**:53. doi:10.1186/s13059-016-0917-0

42 Greene CS, Krishnan A, Wong AK, *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* 2015;**47**:569–76. doi:10.1038/ng.3259

43 Hekselman I, Yeger-Lotem E. Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nat Rev Genet* 2020;**21**:137–50. doi:10.1038/s41576-019-0200-9

44 Duffy Á, Verbanck M, Dobbyn A, *et al.* Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Sci Adv* 2020;**6**. doi:10.1126/sciadv.abb6242

45 Mak IW, Evaniew N, Ghert M. Lost in translation: Animal models and clinical trials in cancer treatment. *Am J Transl Res* 2014;**6**:114–8.https://www.ncbi.nlm.nih.gov/pubmed/24489990

46 Beura LK, Hamilton SE, Bi K, *et al.* Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* 2016;**532**:512–6. doi:10.1038/nature17655

47 Barbeira AN, Dickinson SP, Bonazzola R, *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018;**9**:1825. doi:10.1038/s41467-018-03621-1

48 Gamazon ER, Wheeler HE, Shah KP, *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015;**47**:1091–8. doi:10.1038/ng.3367

49 GTEx Consortium. The Genotype-Tissue expression (GTEx) project. *Nat Genet* 2013;**45**:580–5. doi:10.1038/ng.2653

50 Chee M, Yang R, Hubbell E, *et al.* Accessing genetic information with high-density DNA arrays. *Science* 1996;**274**:610–4. doi:10.1126/science.274.5287.610

51 Ishkanian AS, Malloff CA, Watson SK, *et al.* A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* 2004;**36**:299–303. doi:10.1038/ng1307

52 Hinds DA, Stuve LL, Nilsen GB, *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;**307**:1072–9. doi:10.1126/science.1105436

53 Gresham D, Dunham MJ, Botstein D. Comparing whole genomes using DNA microarrays. *Nat Rev Genet* 2008;**9**:291–302. doi:10.1038/nrg2335

54 Willer CJ, Sanna S, Jackson AU, *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008;**40**:161–9. doi:10.1038/ng.76

55 Consortium IH, Others. A haplotype map of the human genome. *Nature* 2005;**437**:1299.https://www.ncbi.nlm.nih.gov/pmc/articles/pmc1880871/

56 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.* A global reference for human genetic variation. *Nature* 2015;**526**:68–74. doi:10.1038/nature15393

57 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;**48**:1279–83. doi:10.1038/ng.3643

58 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43. doi:10.1038/s41586-020-2308-7

59 GTEx Consortium. Human genomics. The Genotype-Tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 2015;**348**:648–60. doi:10.1126/science.1262110

60 GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–30. doi:10.1126/science.aaz1776

61 Keenan AB, Jenkins SL, Jagodnik KM, *et al.* The library of integrated Network-Based cellular signatures NIH program: System-Level cataloging of human cells response to perturbations. *Cell Syst* 2018;**6**:13–24. doi:10.1016/j.cels.2017.11.001

62 Lamb J, Crawford ED, Peck D, *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929–35. doi:10.1126/science.1132939

63 Subramanian A, Narayan R, Corsello SM, *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452.e17. doi:10.1016/j.cell.2017.10.049

64 Svoboda DL, Saddler T, Auerbach SS. An overview of national toxicology program's toxicogenomic applications: DrugMatrix and ToxFX. Challenges and Advances in Computational Chemistry and Physics. 2019;141–57. doi:10.1007/978-3-030-16443-0_8

65 Prueksaritanont T, Vega JM, Zhao J, *et al.* Interactions between simvastatin and troglitazone or pioglitazone in healthy subjects. *J Clin Pharmacol* 2001;**41**:573–81. doi:10.1177/00912700122010311

66 Backman JT, Kyrklund C, Kivistö KT, *et al.* Plasma concentrations of active simvastatin acid are increased by gemfibrozil. *Clin Pharmacol Ther* 2000;**68**:122–9. doi:10.1067/mcp.2000.108507

67 Nelson SD, LaFleur J, Hunter E, *et al.* Identifying and communicating clinically meaningful Drug-Drug interactions. *J Pharm Pract* 2016;**29**:110–5. doi:10.1177/0897190014544793

68 Huang S-M, Temple R, Throckmorton DC, *et al.* Drug interaction studies: Study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther* 2007;**81**:298–304. doi:10.1038/sj.clpt.6100054

69 U.S. Food and Drug Administration. Clinical drug interaction studies — cytochrome P450 enzyme- and Transporter-Mediated drug interactions guidance for industry. http://archive.today/2021.02.26-201919/https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-drug-interaction-studies-cytochrome-p450-enzyme-and-transporter-mediated-drug-interactions

70 U.S. Food and Drug Administration. In vitro drug interaction studies — cytochrome P450 enzyme- and Transporter-Mediated drug interactions guidance for industry. http://archive.today/2021.02.26-201903/https://www.fda.gov/regulatory-information/search-fda-guidance-documents/vitro-drug-interaction-studies-cytochrome-p450-enzyme-and-transporter-mediated-drug-interactions

71 D'Agostino RB Jr, D'Agostino RB Sr. Estimating treatment effects using observational data. *JAMA* 2007;**297**:314–6. doi:10.1001/jama.297.3.314

72 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018;**47**:2005–14. doi:10.1093/ije/dyy120

73 Harpaz R, Vilar S, Dumouchel W, *et al.* Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013;**20**:413–9. doi:10.1136/amiajnl-2012-000930

74 Lependu P, Iyer SV, Fairon C, *et al.* Annotation analysis for testing drug safety signals using unstructured clinical notes. *J Biomed Semantics* 2012;**3 Suppl 1**:S5. doi:10.1186/2041-1480-3-S1-S5

75 LePendu P, Iyer SV, Bauer-Mehren A, *et al.* Pharmacovigilance using clinical notes. *Clin Pharmacol Ther* 2013;**93**:547–55. doi:10.1038/clpt.2013.47

76 Tatonetti NP, Ye PP, Daneshjou R, *et al.* Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**:125ra31. doi:10.1126/scitranslmed.3003377

77 Tatonetti NP, Denny JC, Murphy SN, *et al.* Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology & Therapeutics* 2011;**90**:133–42. doi:10.1038/clpt.2011.83

78 Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *J Am Med Inform Assoc* 2012;**19**:79–85. doi:10.1136/amiajnl-2011-000214

79 DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* 2016;**47**:20–33. doi:10.1016/j.jhealeco.2016.01.012

80 Hay M, Thomas DW, Craighead JL, *et al.* Clinical development success rates for investigational drugs. *Nat Biotechnol* 2014;**32**:40–51. doi:10.1038/nbt.2786

81 Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics* 2019;**20**:273–86. doi:10.1093/biostatistics/kxx069

82 Hwang TJ, Carpenter D, Lauffenburger JC, *et al.* Failure of investigational drugs in Late-Stage clinical development and publication of trial results. *JAMA Intern Med* 2016;**176**:1826–33. doi:10.1001/jamainternmed.2016.6008

83 Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: Progress, challenges and recommendations. *Nat Rev Drug Discov* 2019;**18**:41–58. doi:10.1038/nrd.2018.168

84 Ashburn TT, Thor KB. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;**3**:673–83. doi:10.1038/nrd1468

85 Kingsmore KM, Grammer AC, Lipsky PE. Drug repurposing to improve treatment of rheumatic autoimmune inflammatory diseases. *Nat Rev Rheumatol* 2020;**16**:32–52. doi:10.1038/s41584-019-0337-0

86 Ferry L, Johnston JA. Efficacy and safety of bupropion SR for smoking cessation: Data from clinical trials and five years of postmarketing experience. *Int J Clin Pract* 2003;**57**:224–30.https://www.ncbi.nlm.nih.gov/pubmed/12723728

87 Trivedi MH, Walker R, Ling W, *et al.* Bupropion and naltrexone in methamphetamine use disorder. *N Engl J Med* 2021;**384**:140–53. doi:10.1056/NEJMoa2020214

88 Protheroe A, Edwards JC, Simmons A, *et al.* Remission of inflammatory arthropathy in association with anti-CD20 therapy for non-hodgkin's lymphoma. *Rheumatology* 1999;**38**:1150–2. doi:10.1093/rheumatology/38.11.1150

89 Keystone E, Burmester GR, Furie R, *et al.* Improvement in patient-reported outcomes in a rituximab trial in patients with severe rheumatoid arthritis refractory to anti-tumor necrosis factor therapy. *Arthritis Rheum* 2008;**59**:785–93. doi:10.1002/art.23715

90 Cohen SB, Emery P, Greenwald MW, *et al.* Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum* 2006;**54**:2793–806. doi:10.1002/art.22025

91 Edwards JC, Cambridge G. Sustained improvement in rheumatoid arthritis following a protocol designed to deplete B lymphocytes. *Rheumatology* 2001;**40**:205–11. doi:10.1093/rheumatology/40.2.205

92 Edwards JCW, Szczepanski L, Szechinski J, *et al.* Efficacy of b-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. *N Engl J Med* 2004;**350**:2572–81. doi:10.1056/NEJMoa032534

93 Wu Y, Warner JL, Wang L, *et al.* Discovery of noncancer drug effects on survival in electronic health records of patients with cancer: A new paradigm for drug repurposing. *JCO Clin Cancer Inform* 2019;**3**:1–9. doi:10.1200/CCI.19.00001

94 Kim D-H, Lee J-E, Kim Y-G, *et al.* High-Throughput algorithm for discovering new drug indications by utilizing Large-Scale electronic medical record data. *Clin Pharmacol Ther* 2020;**108**:1299–307. doi:10.1002/cpt.1980

95 Robinson JR, Denny JC, Roden DM, *et al.* Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clin Transl Sci* 2018;**11**:112–22. doi:10.1111/cts.12522

96 Pulley JM, Rhoads JP, Jerome RN, *et al.* Using what we already have: Uncovering new drug repurposing strategies in existing omics data. *Annu Rev Pharmacol Toxicol* 2020;**60**:333–52. doi:10.1146/annurev-pharmtox-010919-023537

97 Nelson MR, Tipney H, Painter JL, *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;**47**:856–60. doi:10.1038/ng.3314

98 Okada Y, Wu D, Trynka G, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;**506**:376–81. doi:10.1038/nature12873

99 Dudley JT, Sirota M, Shenoy M, *et al.* Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med* 2011;**3**:96ra76. doi:10.1126/scitranslmed.3002648

100 So H-C, Chau CK-L, Chiu W-T, *et al.* Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci* 2017;**20**:1342–9. doi:10.1038/nn.4618

101 Kantor ED, Rehm CD, Haas JS, *et al.* Trends in prescription drug use among adults in the united states from 1999-2012. *JAMA* 2015;**314**:1818–31. doi:10.1001/jama.2015.13766

102 Johnell K, Klarin I. The relationship between number of drugs and potential drug-drug interactions in the elderly: A study of over 600,000 elderly patients from the swedish prescribed drug register. *Drug Saf* 2007;**30**:911–8. doi:10.2165/00002018-200730100-00009

103 Magro L, Moretti U, Leone R. Epidemiology and characteristics of adverse drug reactions caused by drug-drug interactions. *Expert Opin Drug Saf* 2012;**11**:83–94. doi:10.1517/14740338.2012.631910

104 Huang S-M, Lesko LJ. Drug-drug, drug-dietary supplement, and drug-citrus fruit and other food interactions: What have we learned? *J Clin Pharmacol* 2004;**44**:559–69. doi:10.1177/0091270004265367

105 Olvey EL, Clauschee S, Malone DC. Comparison of critical Drug–Drug interaction listings: The department of veterans affairs medical system and standard reference compendia. *Clin Pharmacol Ther* 2010;**87**:48–51. doi:10.1038/clpt.2009.198

106 Phansalkar S, Desai A, Choksi A, *et al.* Criteria for assessing high-priority drug-drug interactions for clinical decision support in electronic health records. *BMC Med Inform Decis Mak* 2013;**13**:65. doi:10.1186/1472-6947-13-65

107 Vitry AI. Comparative assessment of four drug interaction compendia. *Br J Clin Pharmacol* 2007;**63**:709–14. doi:10.1111/j.1365-2125.2006.02809.x

108 Fung KW, Kapusnik-Uner J, Cunningham J, *et al.* Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. *J Am Med Inform Assoc* 2017;**24**:806–12. doi:10.1093/jamia/ocx010

109 Food US, Administration D, Others. Guidance for industry: Drug interaction studies—study design, data analysis, implications for dosing, and labeling recommendations. *Center for Drug Evaluation and Research (CDER)* 2012;1–75.

110 Center for Drug Evaluation, Research. Questions and answers on FDA's adverse event reporting system (FAERS). 2019.http://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers

111 Lorberbaum T, Sampson KJ, Chang JB, *et al.* Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation. *J Am Coll Cardiol* 2016;**68**:1756–64. doi:10.1016/j.jacc.2016.07.761

112 Warrington R, Silviu-Dan F. Drug allergy. *Allergy Asthma Clin Immunol* 2011;**7 Suppl 1**:S10. doi:10.1186/1710-1492-7-S1-S10

113 Newman CB, Preiss D, Tobert JA, *et al.* Statin safety and associated adverse events: A scientific statement from the american heart association. *Arterioscler Thromb Vasc Biol* 2019;**39**:e38–81. doi:10.1161/ATV.0000000000000073

114 Wang L, Blackley SV, Blumenthal KG, *et al.* A dynamic reaction picklist for improving allergy reaction documentation in the electronic health record. *J Am Med Inform Assoc* 2020;**27**:917–23. doi:10.1093/jamia/ocaa042

115 Wei W-Q, Leibson CL, Ransom JE, *et al.* Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc* 2012;**19**:219–24. doi:10.1136/amiajnl-2011-000597

116 McConeghy KW, Caffrey AR, Morrill HJ, *et al.* Are non-allergic drug reactions commonly documented as medication 'allergies'? A national cohort of veterans' admissions from 2000 to 2014. *Pharmacoepidemiol Drug Saf* 2017;**26**:472–6. doi:10.1002/pds.4134

117 Wiggins BS, Saseen JJ, Page RL 2nd, *et al.* Recommendations for management of clinically significant Drug-Drug interactions with statins and select agents used in patients with cardiovascular disease: A scientific statement from the american heart association. *Circulation* 2016;**134**:e468–95. doi:10.1161/CIR.0000000000000456

118 Qato DM, Wilder J, Schumm LP, *et al.* Changes in prescription and Over-the-Counter medication and dietary supplement use among older adults in the united states, 2005 vs 2011. *JAMA Intern Med* 2016;**176**:473–82. doi:10.1001/jamainternmed.2015.8581

119 Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH) Collaborative Group, Armitage J, Bowman L, *et al.* Intensive lowering of LDL cholesterol with 80 mg versus 20 mg simvastatin daily in 12,064 survivors of myocardial infarction: A double-blind randomised trial. *Lancet* 2010;**376**:1658–69. doi:10.1016/S0140-6736(10)60310-8

120 Molokhia M, McKeigue P, Curcin V, *et al.* Statin induced myopathy and myalgia: Time trend analysis and comparison of risk associated with statin class from 1991-2006. *PLoS One* 2008;**3**:e2522. doi:10.1371/journal.pone.0002522

121 Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med* 2002;**21**:2409–19. doi:10.1002/sim.1047

122 Denny JC, Bastarache L, Ritchie MD, *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013;**31**:1102–10. doi:10.1038/nbt.2749

123 Chapman WW, Nadkarni PM, Hirschman L, *et al.* Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions. *J Am Med Inform Assoc* 2011;**18**:540–3. doi:10.1136/amiajnl-2011-000465

124 Zheng K, Vydiswaran VGV, Liu Y, *et al.* Ease of adoption of clinical natural language processing software: An evaluation of five systems. *J Biomed Inform* 2015;**58 Suppl**:S189–96. doi:10.1016/j.jbi.2015.07.008

125 Harpaz R, Callahan A, Tamang S, *et al.* Text mining for adverse drug events: The promise, challenges, and state of the art. *Drug Saf* 2014;**37**:777–90. doi:10.1007/s40264-014-0218-z

126 Phansalkar S, Sijs H van der, Tucker AD, *et al.* Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. *J Am Med Inform Assoc* 2013;**20**:489–93. doi:10.1136/amiajnl-2012-001089

127 Sutton RT, Pincock D, Baumgart DC, *et al.* An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digit Med* 2020;**3**:17. doi:10.1038/s41746-020-0221-y

128 Helmons PJ, Suijkerbuijk BO, Nannan Panday PV, *et al.* Drug-drug interaction checking assisted by clinical decision support: A return on investment analysis. *J Am Med Inform Assoc* 2015;**22**:764–72. doi:10.1093/jamia/ocu010

129 Phansalkar S, Desai AA, Bell D, *et al.* High-priority drug-drug interactions for use in electronic health records. *J Am Med Inform Assoc* 2012;**19**:735–43. doi:10.1136/amiajnl-2011-000612

130 Elliott HL, Meredith PA, Campbell L, *et al.* The combination of prazosin and verapamil in the treatment of essential hypertension. *Clin Pharmacol Ther* 1988;**43**:554–60. doi:10.1038/clpt.1988.72

131 Lenz ML, Pool JL, Laddu AR, *et al.* Combined terazosin and verapamil therapy in essential hypertension. Hemodynamic and pharmacokinetic interactions. *Am J Hypertens* 1995;**8**:133–45. doi:10.1016/0895-7061(94)00162-5

132 Dorofeeva MN, Shikh EV, Sizova ZM, *et al.* Antihypertensive effect of amlodipine in Co-Administration with omeprazole in patients with hypertension and Acid-Related disorders: Cytochrome P450-Associated aspects. *Pharmgenomics Pers Med* 2019;**12**:329–39. doi:10.2147/PGPM.S217725

133 Almenoff JS, DuMouchel W, Kindman LA, *et al.* Disproportionality analysis using empirical bayes data mining: A tool for the evaluation of drug interactions in the post-marketing setting. *Pharmacoepidemiol Drug Saf* 2003;**12**:517–21. doi:10.1002/pds.885

134 Steeg E van de, Venhorst J, Jansen HT, *et al.* Generation of bayesian prediction models for OATP-mediated drug-drug interactions based on inhibition screen of OATP1B1, OATP1B1∗15 and OATP1B3. *Eur J Pharm Sci* 2015;**70**:29–36. doi:10.1016/j.ejps.2015.01.004

135 Trubiano JA, Adkinson NF, Phillips EJ. Penicillin allergy is not necessarily forever. *JAMA* 2017;**318**:82–3. doi:10.1001/jama.2017.6510

136 Stone CA Jr, Trubiano J, Coleman DT, *et al.* The challenge of de-labeling penicillin allergy. *Allergy* Published Online First: May 2019. doi:10.1111/all.13848

137 Grizzle AJ, Hines LE, Malone DC, *et al.* Testing the face validity and inter-rater agreement of a simple approach to drug-drug interaction evidence assessment. *J Biomed Inform* 2020;**101**:103355. doi:10.1016/j.jbi.2019.103355

138 Krebs K, Bovijn J, Zheng N, *et al.* Genome-wide study identifies association between HLA-B∗55:01 and Self-Reported penicillin allergy. *Am J Hum Genet* 2020;**107**:612–21. doi:10.1016/j.ajhg.2020.08.008

139 Wiley LK, Moretz JD, Denny JC, *et al.* Phenotyping adverse drug reactions: Statin-Related myotoxicity. *AMIA Jt Summits Transl Sci Proc* 2015;**2015**:466–70.https://www.ncbi.nlm.nih.gov/pubmed/26306287

140 Wu P. Pwatrick/ddiwas: First commit. 2020. doi:10.5281/zenodo.4251662

141 Paik H, Chung A-Y, Park H-C, *et al.* Repurpose terbutaline sulfate for amyotrophic lateral sclerosis using electronic medical records. *Sci Rep* 2015;**5**:8580. doi:10.1038/srep08580

142 Sirota M, Dudley JT, Kim J, *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med* 2011;**3**:96ra77. doi:10.1126/scitranslmed.3001318

143 Zhang W, Voloudakis G, Rajagopal VM, *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* 2019;**10**:3834. doi:10.1038/s41467-019-11874-7

144 Pilarczyk M, Najafabadi MF, Kouril M, *et al.* Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINCS. bioRxiv. 2019;826271. doi:10.1101/826271

145 Ramirez AH, Sulieman L, Schlueter DJ, *et al.* The all of us research program: Data quality, utility, and diversity. Public and Global Health. 2020. doi:10.1101/2020.05.29.20116905

146 Im HK. MetaXcan results. https://s3.amazonaws.com/imlab-open/Data/MetaXcan/results/metaxcan_results_database_v0.1.tar.gz

147 Im HK. S-PrediXcan results. Diagnoses - main ICD10: I10 essential (primary) hypertension. https://uchicago.box.com/shared/static/6tdiyksvxcm2nxjiml14deqiz1r6kqp7.bz2

148 Willer CJ, Schmidt EM, Sengupta S, *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**45**:1274–83. doi:10.1038/ng.2797

149 Neale BM. Neale lab - UK biobank GWAS results. 2020.http://www.nealelab.is/uk-biobank/

150 Grimm RH Jr, Flack JM, Grandits GA, *et al.* Long-term effects on plasma lipids of diet and drugs to treat hypertension. Treatment of mild hypertension study (TOMHS) research group. *JAMA* 1996;**275**:1549–56. doi:10.1001/jama.1996.03530440029033

151 Strazzullo P, Kerry SM, Barbato A, *et al.* Do statins reduce blood pressure?: A meta-analysis of randomized, controlled trials. *Hypertension* 2007;**49**:792–8. doi:10.1161/01.HYP.0000259737.43916.42

152 Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. *Nat Rev Genet* 2014;**15**:409–21. doi:10.1038/nrg3723

153 Bock C, Boutros M, Camp JG, *et al.* The organoid cell atlas. *Nat Biotechnol* 2021;**39**:13–7. doi:10.1038/s41587-020-00762-x

154 Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS One* 2013;**8**:e66341. doi:10.1371/journal.pone.0066341

155 Im HK. PhenomeXcan. 2020.http://apps.hakyimlab.org/phenomexcan/

156 Pividori M, Rajagopal PS, Barbeira A, *et al.* PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci Adv* 2020;**6**. doi:10.1126/sciadv.aba2083

157 Barbeira AN, Bonazzola R, Gamazon ER, *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Cold Spring Harbor Laboratory. 2020;814350. doi:10.1101/814350

158 Im HK. Im lab's PredictDB data repository. http://predictdb.org/

159 Abecasis G. Global lipids genetics consortium results. http://csg.sph.umich.edu/willer/public/lipids2013/

160 Ganter B, Tugendreich S, Pearson CI, *et al.* Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J Biotechnol* 2005;**119**:219–44. doi:10.1016/j.jbiotec.2005.03.022

161 DrugMatrix. https://ntp.niehs.nih.gov/results/toxfx/index.html

162 Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform* 2018;**27**:129–39. doi:10.1055/s-0038-1667077

163 Stein EA, Mellis S, Yancopoulos GD, *et al.* Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N Engl J Med* 2012;**366**:1108–18. doi:10.1056/NEJMoa1105803

164 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: An alternative to standard epidemiological study designs. *BMJ* 2016;**354**:i4515. doi:10.1136/bmj.i4515

165 Kannel WB, Gordon T, Schwartz MJ. Systolic versus diastolic blood pressure and risk of coronary heart disease. The framingham study. *Am J Cardiol* 1971;**27**:335–46. doi:10.1016/0002-9149(71)90428-0

166 Denny J. All of us research program begins beta testing of data platform. 2020.http://archive.today/2021.02.23-233456/https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-begins-beta-testing-data-platform

167 Baxter SL, Saseendrakumar BR, Paul P, *et al.* Predictive analytics for glaucoma using data from the all of us research program. *Am J Ophthalmol* Published Online First: January 2021. doi:10.1016/j.ajo.2021.01.008

168 NOTE NLP table #85. 2017.http://archive.today/2020.10.03-114046/https://github.com/OHDSI/CommonDataModel/issues/85

169 McNeer E, Beck C, Weeks HL, *et al.* Building longitudinal medication dose data using medication information extracted from clinical notes in electronic health records. *J Am Med Inform Assoc* Published Online First: December 2020. doi:10.1093/jamia/ocaa291

170 Ellebrecht CT, Bhoj VG, Nace A, *et al.* Reengineering chimeric antigen receptor T cells for targeted therapy of autoimmune disease. *Science* 2016;**353**:179–84. doi:10.1126/science.aaf6756

171 Johnson DB, Balko JM, Compton ML, *et al.* Fulminant myocarditis with combination immune checkpoint blockade. *N Engl J Med* 2016;**375**:1749–55. doi:10.1056/NEJMoa1609214

172 Martins F, Sofiya L, Sykiotis GP, *et al.* Adverse effects of immune-checkpoint inhibitors: Epidemiology, management and surveillance. *Nat Rev Clin Oncol* 2019;**16**:563–80. doi:10.1038/s41571-019-0218-0

173 Barbeira AN, Pividori M, Zheng J, *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet* 2019;**15**:e1007889. doi:10.1371/journal.pgen.1007889

174 Zhou D, Jiang Y, Zhong X, *et al.* A unified framework for joint-tissue transcriptome-wide association and mendelian randomization analysis. *Nat Genet* 2020;**52**:1239–46. doi:10.1038/s41588-020-0706-2

175 MacArthur J, Bowler E, Cerezo M, *et al.* The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res* 2017;**45**:D896–901. doi:10.1093/nar/gkw1133

176 Jagadeesh KA, Dey KK, Montoro DT, *et al.* Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. Cold Spring Harbor Laboratory. 2021;2021.03.19.436212. doi:10.1101/2021.03.19.436212

177 Takahashi T. Organoids for drug discovery and personalized medicine. *Annu Rev Pharmacol Toxicol* 2019;**59**:447–62. doi:10.1146/annurev-pharmtox-010818-021108

178 Williams G, Gatt A, Clarke E, *et al.* Drug repurposing for alzheimer's disease based on transcriptional profiling of human iPSC-derived cortical neurons. *Transl Psychiatry* 2019;**9**:220. doi:10.1038/s41398-019-0555-x

179 Rowe RG, Daley GQ. Induced pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Genet* 2019;**20**:377–88. doi:10.1038/s41576-019-0100-z

180 Insel TR. Digital phenotyping: Technology for a new science of behavior. *JAMA* 2017;**318**:1215–6. doi:10.1001/jama.2017.11295

181 Powers R, Etezadi-Amoli M, Arnold EM, *et al.* Smartwatch inertial sensors continuously monitor real-world motor fluctuations in parkinson's disease. *Sci Transl Med* 2021;**13**. doi:10.1126/scitranslmed.abd7865

182 Zhan A, Mohan S, Tarolli C, *et al.* Using smartphones and machine learning to quantify parkinson disease severity: The mobile parkinson disease score. *JAMA Neurol* 2018;**75**:876–80. doi:10.1001/jamaneurol.2018.0809

183 Rodriques SG, Stickels RR, Goeva A, *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;**363**:1463–7. doi:10.1126/science.aaw1219

184 Setty M, Tadmor MD, Reich-Zeliger S, *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat Biotechnol* 2016;**34**:637–45. doi:10.1038/nbt.3569

185 Wojcik GL, Graff M, Nishimura KK, *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 2019;**570**:514–8. doi:10.1038/s41586-019-1310-4